Data-driven Ultrasound Localization Microscopy using Deep Learning

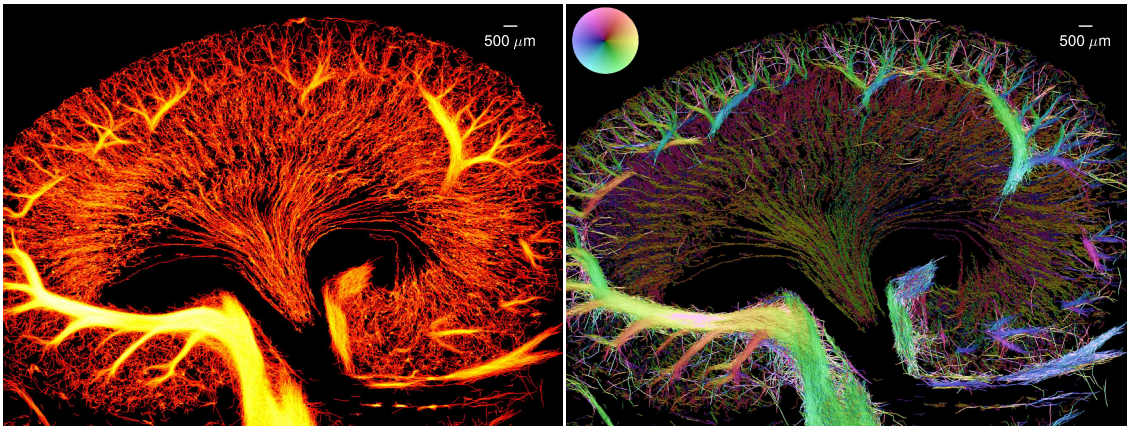
Youn, Jihwan

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

DTU Health Tech
Department of Health Technology

Ph.D. Thesis



Data-driven Ultrasound Localization Microscopy using Deep Learning

Author: Jihwan Youn

Supervised by: Prof. Jørgen Arendt Jensen, Ph.D., Dr. Techn.
Co-supervised by: Assoc. Prof. Matthias Bo Stuart, Ph.D.

Technical University of Denmark, Kgs. Lyngby, Denmark, 2021

Cover image: Super-resolution ultrasound image of a rat kidney. A microbubble intensity map (left) and a track image (right) obtained from the estimated microbubble (MB) positions using a convolutional neural network. The figure is from (Youn et al. 2021).

Center for Fast Ultrasound imaging (CFU)
DTU Health Tech
Department of Health Technology
Technical University of Denmark
Ørsteds Plads
Building 349
2800 Kgs. Lyngby, Denmark

Preface

This Ph.D. thesis has been submitted to the Department of Health Technology at Technical University of Denmark (DTU) in partial fulfillment of the requirements for acquiring the Ph.D. degree. The research providing the foundation for this thesis has been conducted for three years, from May 15th, 2018 to May 14th, 2021, at the Center for Fast Ultrasound Imaging (CFU), Department of Health Technology. The project has been supervised by Prof. Jørgen Arendt Jensen, Ph.D., Dr. Techn., and co-supervised by Assoc. Prof. Matthias Bo Stuart, Ph.D. The project was financially supported in part by the Fondation Idella.

For the past three years, I had the opportunity of attending three IEEE International Ultrasonics Symposiums (two in Kobe, Japan and Glasgow, Scotland, UK, and one virtually due to COVID-19), the 2019 Artimino Conference on Medical Ultrasound Technology in Nijmegen, the Netherlands, and the 26th European Symposium on Ultrasound Contrast Imaging virtually. I also had the privilege to participate in the Nordic Probabilistic AI School, Trondheim, Norway, and I had a chance to conduct external research for three months at Eindhoven University of Technology, Eindhoven, the Netherlands, where I focused on model-based neural networks under the supervision of Asst. Prof. Ruud J. G. van Sloun, Ph.D. Those experiences gave me chances to meet colleagues having similar research interests, and I was able to broaden my knowledge and perspective in medical ultrasound imaging and deep learning.

During the winter semesters in 2019 and 2020, I worked as a teaching assistant in the course of Medical Imaging Systems. That allowed improving my ability to explain scientific concepts succinctly to students considering their backgrounds and knowledge. In addition, I had the pleasure to co-supervise a student, Emil Kristiansen, for his bachelor's thesis: "Tracking of Targets Using Deep Learning."

I won first prize in the competition of artificial intelligence with microbubbles at the 26th European symposium on Ultrasound Contrast Imaging with the work of "Task-adaptive Beamforming for Microbubble Localization using Deep Learning."

Jihwan Youn
Kgs. Lyngby, Denmark
May 2021

Contents

Summary

Ultrasound localization microscopy (ULM) can surpass the spatial resolution limit of conventional ultrasound imaging by accumulating centroids of MBs injected into the bloodstream in one image frame. However, there is a trade-off between the resolution and data acquisition time. For accurate localization, low concentrations of diluted MBs are commonly used. That limits the number of detectable MBs, and the long data acquisition time is thus required. This Ph.D. project aims to localize high concentrations of MBs using data-driven deep learning methods.

Initially, localizing scatterers from radiofrequency (RF) channel data has been investigated since point spread functions (PSFs) of closely spaced scatterers overlap each other in beamformed ultrasound images. Convolutional neural networks (CNNs) were trained with simulated ultrasound data and non-overlapping Gaussian confidence maps for stable training. The performance was evaluated in the simulated test data and phantom measurements, showing that scatterers closer than the resolution limit of delay-and-sum (DAS) beamforming can be localized.

Next, a sub-pixel localization method using a CNN on the beamformed ultrasound images has been studied. Sub-pixel localization was achieved by fitting Gaussians in the extended non-overlapping Gaussian confidence maps. That allows utilizing computational resources efficiently as no additional upsampling is required. In a phantom experiment at a high MB concentration, the sub-pixel CNN localization method resolved a pair of channels spaced 22 µm away while centroid detection failed. Sub-pixel CNN localization was also tested on *in vivo* data and resulted in estimated MBs spaced closer than a wavelength.

Lastly, model-based neural networks for ULM have been investigated. The model-based neural networks are designed based on mathematical foundations. Hence, compared with the model-agnostic data-driven methods, fewer learning parameters are required. The few learning parameters allow a short training time with a small number of data, good generalization, and fast inference speed. Deep unfolded ULM, which localizes the overlapping MBs using a model-based neural network, has shown comparable results to the fully data-driven methods on simulated and measured data. In addition, task-adaptive beamforming for MB localization has been investigated. By jointly optimizing a deep beamformer and localization network, ultrasound images tailored for MB localization were able to be obtained, and thus, the performance of deep unfolded ULM increased.

x

Resumé

ULM kan overgå grænsen for rumlig opløsning ved konventionel ultralydsbilleddannelse ved at akkumulere centroider af mikrobobler (MB'er), der injiceres i blodbanen. Dog er der en afvejning mellem rumlig opløsning og dataindsamlingstid. Ofte anvendes lave koncentrationer af MB'er for at opnå nøjagtig lokalisering. Det begrænser antallet af detekterbare MB'er, og den lange dataindsamlingstid kræves derfor. Dette Ph.D. projekt sigter mod at lokalisere høje koncentrationer af MB'er ved hjælp af datadrevne deep learning metoder.

Først er lokalisering af scatterers fra radiofrekvens-kanaldata blevet undersøgt, da punktspredningsfunktioner (PSF'er) af scatterers tæt på hinanden overlapper hinanden i beamformede ultralydsbilleder. Convolutional Neural Networks (CNN'er) er blevet trænet med simulerede ultralydsdata, og ikke-overlappende Gaussiske confidence maps for mere stabil træning. Ydeevnen blev evalueret i de simulerede testdata og fantommålinger, der viser, at scatterers, der er tættere på hinanden end grænsen for rumlig opløsning ved DAS beamforming, kan lokaliseres.

Dernæst er en sub-pixel lokaliseringsmetode, ved anvendelse af en CNN på de beamformede ultralydsbilleder, blevet undersøgt. Sub-pixel-lokaliseringen blev opnået ved at passe Gauss kurver på de udvidede ikke-overlappende Gaussiske confidence maps. Dette gør det muligt at udnytte beregningsressourcer effektivt, da der ikke kræves yderligere opsampling. I et fantomeksperiment med en høj MB-koncentration kunne sub-pixel CNN-lokaliseringsmetoden bestemme to kanaler, der var anbragt 22 µm væk, mens centroid-detektion mislykkedes. Sub-pixel CNN-lokalisering blev også testet på '*in vivo*-data og viste estimerede MB'er, der var anbragt tættere på hinanden end bølgelængden.

Endelig er model-based neural networks til ULM blevet undersøgt. De model-based neural networks er designet ud fra matematiske fundamenter. Derfor kræves der færre indlæringsparametre sammenlignet med de model-agnostiske datadrevne metoder. De få læringsparametre tillader en kort træningstid med et lille antal data, god generalisering og hurtig inferenshastighed. Deep unfolded ULM, som lokaliserer overlappende MB'er ved hjælp af et modelbaseret neuralt netværk, har vist sammenlignelige resultater med de fuldt datadrevne metoder på simulerede og målte data. Derudover er task-adaptive beamforming til MB-lokalisering blevet undersøgt. Ved samtidigt at optimere en deep beamformer og et lokaliseringsnetværk kan ultralydsbilleder, der er skræddersyet til MB-lokalisering, opnås, og dermed øges ydeevnen for deep unfolded ULM.

Acknowledgments

I would like to express my gratitude to my supervisors Prof. Jørgen Arendt Jensen and Assoc. Prof. Matthias Bo Staurt for all of their guidance, support, and patience during my Ph.D. study over the past three years. I will never forget your motivation that there is always a way to solve a problem, but we just do not know it yet. With your supervision and intellectual insight, I could have completed my Ph.D. study successfully.

I thank the collaborators: Prof. Yonina Eldar, Asst. Prof. Ruud J. G. van Sloun, and Mr. Ben Luijten, during my external stay in technical university of Eindhoven, the Netherlands. It was a great pleasure for me to experience and learn new perspectives to my work and broaden my intelligence.

I would like to thank my colleagues at the center for fast ultrasound, technical university of Denmark. I really enjoyed my office life with you professionally and personally: Mikkel for his knowledge in ultrasound, Rasmus for helping me being familiar with the Danish culture, Kseniya for sharing useful information to live in Denmark as an expatriate, Sigrid for help with microbubble physics and English writing skills, Iman for help with processing tracking, Isabella for the positive vibes, and Lasse for help with running jobs on the cluster. I also thank Senior Researcher Borislav Gueorgiev Tomov, Ph.D., for his technical assistance with SARUS.

Personally, I thank my family in Korea for supporting me all the time when I was frustrated and suffered.

Lastly, I would like to thank my girlfriend Youngmi, who has always been on my side and believed in me with love.

List of Figures

List of Tables

List of Algorithms

Abbreviations

ABLE Adaptive Beamforming by deep LEarning.

AM amplitude modulation.

B-mode brightness mode.

CEUS contrast-enhanced ultrasound.

CNN convolutional neural network.

DAS delay-and-sum.

DI differential imaging.

FISTA fast iterative shrinkage-thresholding algorithm.

FLOP floating point operation.

FWHM full width at half maximum.

ISTA iterative shrinkage-thresholding algorithm.

MB microbubble.

MSE mean squared error.

PI pulse inversion.

PSF point spread function.

RAdam Rectified Adam.

ReLU rectified linear unit.

RF radiofrequency.

ROI region of interest.

SARUS synthetic aperture real-time ultrasound system.

SNR signal-to-noise ratio.

SVD singular value decomposition.

ULM ultrasound localization microscopy.

VFI vector flow imaging.

Part I

Introduction

CHAPTER 1

Introduction

1.1 Medical Ultrasound Imaging

The use of ultrasound for medical imaging was proposed over 70 years ago (Edler and Hertz 1954; Howry and Bliss 1952; Wild 1950), and medical ultrasound imaging is now one of the most widely used imaging modalities. It is safe, non-invasive, low-cost, and does not employ ionizing radiation. Ultrasound scanners are portable and stream images in real-time so it is easily accessible and effective at the bedside.

Medical ultrasound relies on the pulse-echo principle, assuming that the speed of sound is constant. Ultrasound probes can convert mechanical vibration to electrical energy or vice versa, i.e., transmit and receive ultrasound waves. Transmitted ultrasound beams by the probes are backscattered when they face the acoustic impedance changes in tissue. The changes in acoustic impedance along the beams can be measured from the received echoes with their time of flight and amplitude. Brightness mode (B-mode) images are formed by collecting such information in a 2-D plane where their brightness represents the backscattering amplitude at the pixel location. The B-mode images allow investigating internal body structures and, in clinics, they are used to diagnose diseases, monitor treatments, examine fetuses in the womb, guide biopsies.

Medical ultrasound can visualize blood flow in Doppler mode, as well as anatomical structures in B-mode. Spectral Doppler measures the blood flow in a position over time, and color Doppler overlays color-coded velocity information in a 2-D region on top of B-mode images. The velocity is estimated by measuring the phase shift of the echoes induced by moving scatterers away from or towards the ultrasound probe. In practice, the Doppler mode images help physicians diagnose vascular diseases, examine heart valve function, and determine if a patient is in good condition for angioplasty.

The velocity estimation in Doppler mode is operator-dependent since the estimation accuracy depends on the flow angle which the operator determines. Vector flow imaging (VFI) has been proposed to overcome the limitation using speckle tracking (Bohs et al. 1993), transverse oscillation (Jensen and Munk 1998), and vector Doppler (Dunmire et al. 2000). Those techniques are angle-independent by providing the velocity estimations in all directions in a 2-D plane. Hence, more consistent velocity estimation is available regardless of the operators. VFI using transverse oscillation is already implemented in commercial ultrasound systems bk5000 (BK Medical, Herlev, Denmark).

Despite the advances in ultrasound imaging technology, it has been challenging to visualize microvasculature due to the limited spatial resolution of ultrasound until the

advent of ultrasound localization microscopy (ULM) (Christensen-Jeffries et al. 2015; Couture et al. 2011; Errico et al. 2015; O'Reilly and Hynynen 2013; Siepmann et al. 2011; Viessmann et al. 2013). ULM is one of the super-resolution imaging methods that can break the diffraction limit. By pinpointing individual microbubbles (MBs) injected into the bloodstream and superimposing their centroids in one image frame, sub-wavelength imaging can be achieved. ULM enables mapping microvascular networks composed of microvessels spaced closer than the resolution limit of conventional ultrasound imaging. Its resulting super-resolution images can be used for the diagnosis of the early-stage cancer (Lin et al. 2017), ischemic kidney disease (Andersen et al. 2020), and diabetes (Ghosh et al. 2019), as well as functional ultrasound (Deffieux et al. 2018).

1.2 Motivation

ULM has shown great potential as a breakthrough in super-resolution ultrasound imaging. However, long data acquisition time is one of the major limitations that hinder ULM to be employed in practice. Generally, MB localization is performed on the beamformed ultrasound images, which are diffraction limited. The standard MB localization methods, e.g., centroid detection or Gaussian fitting, are not able to correctly localize overlapping point spread functions (PSFs). Therefore, for accurate localization, low concentrations of diluted MBs are commonly used to avoid the overlapping PSFs as much as possible. But this also limits the number of detectable MBs in an image frame, and as a result, the long data acquisition time is required to map the entire target structures.

Recently, deep learning has had a profound impact on processing complex information and making associated decisions. By constructing deep neural networks with a lot of learning parameters and training them with a large amount of data, unprecedented improvements have been achieved in many various areas, e.g., image classification (He, Zhang, et al. 2016a,b; Krizhevsky, Sutskever, and Hinton 2012), object detection (He, Gkioxari, et al. 2017; Huang et al. 2017; Redmon and Farhadi 2018), semantic segmentation (Chen et al. 2018; Ronneberger, Fischer, and Brox 2015; Zhao et al. 2017), single-image super-resolution (Ledig et al. 2017; Lim et al. 2017), natural language processing (Brown et al. 2020; Devlind et al. 2019), and image generation (Goodfellow et al. 2014; Karras et al. 2018; Radford, Metz, and Chintala 2016).

Correspondingly, deep learning techniques have been applied to ultrasound imaging applications. In beamforming, image contrast was improved by suppressing off-axis scattering (Luchies and Byram 2018) and reducing speckle noise (Hyun et al. 2019). A content-adaptive beamformer that estimates beamforming weights and produces high-quality ultrasound images was proposed in (Luijten et al. 2020). Radiofrequency (RF) channel data sub-sampling was suggested to reduce the data rate without losing image quality (Huijben et al. 2020; Khan, Huh, and Ye 2020; Yoon et al. 2018). A robust PCA-based neural network was presented to perform clutter filtering in (Solomon et al. 2020).

In this Ph.D. project, it is hypothesized that the data-driven deep learning techniques can solve the trade-off problem between the resolution and data acquisition time in ULM. Therefore, deep learning-based MB localization methods, especially for conditions with overlapping PSFs, have been studied. Ultimately, this project aims to localize high concentrations of MBs with a high localization accuracy, so that the data acquisition time can be shortened without sacrificing the resolution of ULM, by localizing more MBs in an image frame.

1.3 Scientific Contribution

A list of published papers and papers in preparation during this Ph.D. project is shown below. The listed papers can be found in the appendix.

- **Paper 1**
  Jihwan Youn, Martin Lind Ommen, Matthias Bo Stuart, Erik Vilain Thomsen, Niels Bent Larsen, Jørgen Arendt Jensen,
  "Multiple Point Target Detection and Localization using Deep Learning,"
  In *IEEE Int. Ultrason. Symp.*, pp. 1937-1940, 2019.

- **Paper 2**
  Jihwan Youn, Martin Lind Ommen, Matthias Bo Stuart, Erik Vilain Thomsen, Niels Bent Larsen, Jørgen Arendt Jensen,
  "Detection and Localization of Ultrasound Scatterers Using Convolutional Neural Networks,"
  In *IEEE Trans. Med. Imag.*, pp. 3855-3867, 2020.

- **Paper 3**
  Jihwan Youn, Iman Taghavi, Martin Lind Ommen, Mikkel Schou, Matthias Bo Stuart, Erik Vilain Thomsen, Niels Bent Larsen, Jørgen Arendt Jensen,
  "Sub-pixel Accuracy Microbubble Localization using Convolutional Neural Networks,"
  In *preparation*.

- **Paper 4**
  Jihwan Youn, Ben Luijten, Matthias Bo Stuart, Yonina C. Eldar, Ruud J. G. van Sloun, Jørgen Arendt Jensen,
  "Deep Learning Models for Fast Ultrasound Localization Microscopy,"
  In *IEEE Int. Ultrason. Symp.*, pp. 1-4, 2020

- **Paper 5**
  Jihwan Youn, Ben Luijten, Matthias Bo Stuart, Yonina C. Eldar, Ruud J. G. van Sloun, Jørgen Arendt Jensen,
  "Model-based Deep Learning on Ultrasound Channel Data for Fast Ultrasound

Localization Microscopy,"
In *preparation*.

1.4 Outline

In the following chapter, ULM is introduced. And then, the main part comes, which consists of two parts: fully data-driven methods and model-based data-driven methods, where the main contributions are presented. Finally, the conclusion and outlook are given. A brief description of each chapter is as follows.

Chapter 2 explains the basic concepts of ULM and reviews each step of the ULM pipeline. And the limitations of current ULM processes and potential solutions are discussed.

Part I Fully data-driven methods

Chapter 3 presents ultrasound scatterer localization from RF channel data using convolutional neural networks (CNNs) without explicit beamforming. Fabrication of 3-D printed PEDGA phantoms containing ultrasound scatterers and validation of the localization method on the phantoms are introduced. This chapter is based on Paper 1 (Youn, Ommen, et al. 2019) and Paper 2 (Youn, Ommen, et al. 2020).

Chapter 4 describes MB localization on the beamformed ultrasound images. Sub-pixel localization is achieved unlike other deep learning methods by Gaussian fitting. A 3-D printed phantom having pairs of closely spaced channels is employed to compare different localization methods. *In vivo* validation is performed at various MB concentrations to assess the ability to localize the overlapping MBs. This chapter is based on Paper 3 (Youn, Taghavi, et al. 2021).

Part II Model-based data-driven methods

Chapter 5 explains a kind of model-based neural network, deep unfolded neural networks, that solves sparse recovery problems. And deep unfolded ULM that performs MB localization using the model-based network is introduced. The localization performance of deep unfolded ULM is compared against model-agnostic data-driven methods on simulated test data and the phantom measurements acquired in Chapter 4. This chapter is based on Paper 4 (Youn, Luijten, et al. 2020).

Chapter 6 investigates task-adaptive beamforming for MB localization. The beamformer and localization network is trained jointly in an end-to-end fashion. The resulting

beamformed images become favorable to the downstream localization task. This chapter is based on Paper 5 (Youn, Luijten, et al. 2021).

Part III Conclusion

Chapter 7 summarizes this Ph.D. project and provides an outlook on ULM using deep learning techniques.

CHAPTER 2

Ultrasound Localization Microscopy

In this chapter, a brief overview of ULM is presented. The steps of the ULM pipeline are reviewed. In addition, the limitations of current ULM methods and potential solutions are discussed.

2.1 Introduction

There have been efforts to increase the spatial resolution of ultrasound imaging since 1979 (Ikeda, Sato, and Suzuki 1979). The most straightforward way of increasing the resolution would be transmitting high frequency ultrasound beams (Lockwood et al. 1996). The higher frequencies give shorter wavelengths, so better resolution can be achieved as the resolution is proportional to the wavelength. For example, a 15 MHz frequency ultrasound gives resolution below 100 µm, assuming the speed of sound is around 1540 m/s. However, it is still diffraction limited, and the penetration depth becomes more constrained as the frequency increases. Super-resolution ultrasound imaging aims to separate targets placed closer than the resolution limit of ultrasound by diffraction, i.e., half of a wavelength (Christensen-Jeffries, Couture, et al. 2020).

In 2006, super-resolution microscopy using fluorescence sources was proposed, surpassing the diffraction of light in optics (Betzig et al. 2006; Hess, Girirajan, and Mason 2006; Rust, Bates, and Zhuang 2006). The basic concept is based on the fact that the center positions of the isolated sources can be determined with a localization precision higher than the wavelength. By activating a sub-set of the fluorescence sources, the interference among the signals triggered by the activated sources can be avoided. The super-resolution images with the resolution of several nanometers can be reconstructed by collecting the centroids of the sources over a large number of frames in an image frame. Eric Betzig, Stefan Hell, and William E. Moerner won the Nobel Prize in Chemistry 2014 for the development of super-resolved fluorescence microscopy.

Inspired by the fluorescence localization microscopy, ULM has been introduced in (Couture et al. 2011). The fluorescence sources were replaced by ultrasound contrast agents and ultrasound waves were employed instead of light. The resulting ULM image can achieve resolution improvements by a factor of 10 (Christensen-Jeffries, Couture, et al. 2020). The capability of ULM has been extensively investigated on *in-vitro* (Viessmann et al. 2013) and *in-vivo* data (Christensen-Jeffries, Browning, et al. 2015; Errico et al. 2015; M. A. O'Reilly and Hynynen 2013). Fig. 2.1 shows an overview of the ULM

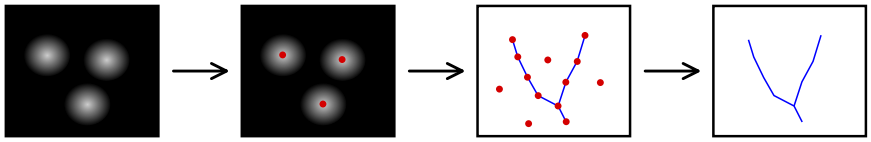


Figure 2.1: An overview of the ULM pipeline. (a) is MB signals isolated from tissue signals. (b) is localization of individual MBs on the image obtained in (a). (c) is tracking on the estimated MB positions over multiple consecutive frames to remove false estimation and provide velocity information, and (d) is the final ULM image.

pipeline, and the detail of each step is presented in the following sections.

2.2 Microbubble Data Acquisition

Firstly, the scattered sound from MBs need to be separated from the surrounding tissue; otherwise, it is difficult to identify and localize individual MBs in the B-mode images. The MB signal separation can be performed by utilizing their non-linear property or capturing the movement of MBs across multiple frames.

MBs show non-linear behavior when they are insonified by ultrasound beams. Contrast-enhanced ultrasound (CEUS), e.g., pulse inversion (PI) (Simpson, Chin, and Burns 1999) or amplitude modulation (AM) (Brock-Fischer, Poland, and Rafter 1996; Mor-Avi et al. 2001) can isolate the non-linear signals and visualize MBs effectively. CEUS imaging involves multiple transmissions and the summation of adjacent frames in time. For PI, the fundamental signals are removed and the even-ordered harmonics are captured by summing a positive and a negative transmission frames. Therefore, the non-linear signals from MBs can be clearly separated from the background tissue. Furthermore, it does not rely on the movement of MBs, therefore, slow moving or stationary MBs can be detected. AM also employs the non-linear behavior of MBs, but it maintains the fundamental signals as well as the harmonics. Hence, more sensitivity can be achieved compared to PI because AM preserves the non-linear component of the fundamental frequency and suffer less attenuation. Alternatively, a combination of PI and AM (PIAM) (Eckersley, Chin, and Burns 2005) can be considered to further increase the sensitivity.

On the contrary, differential imaging (DI) (Desailly et al. 2013) and singular value decomposition (SVD) filtering (Demene et al. 2015) do not employ the non-linear property of MBs but filter out stationary echoes to separate MB signals. DI subtracts two adjacent frames, which leaves the signals incurred by movement between the frames. It is simple and easy to implement, however, not robust to tissue motion caused by heartbeat and breathing. Also, slow moving or stationary MBs are not well detected. SVD represents a stack of ultrasound images with singular vectors and their corresponding singular values describing the temporal consistency. The singular vectors with small singular values

represent moving objects, i.e., moving MBs. By reconstructing ultrasound images with the singular vectors with low singular values, MB signals can be extracted. SVD filtering is often more robust than DI and does not need to sacrifice frame rate, yet there are several limitations from the practical point of view. For example, SVD is computationally expensive, so real-time streaming is almost impossible. And it is not straightforward to find the optimal singular value to determine MB signals and its performance is also highly dependent on such decision parameters. More detailed comparison of PI, DI, and SVD for MB signal separation can be found in (Brown et al. 2019).

2.3 Localization

MB localization can be performed either on RF channel data or beamformed ultrasound images. On the RF channel data, the vertex is found by parabola fitting (Couture et al. 2011; Desailly et al. 2013). On the beamformed images, the MB positions can be found by intensity-weighted centroid detection using image moments (Christensen-Jeffries, Browning, et al. 2015; Siepmann et al. 2011; Viessmann et al. 2013), fitting a Gaussian function to the local peak and its neighboring pixels, or deconvolution (Errico et al. 2015).

ULM maps target structures indirectly by collecting the centers of MBs injected into the bloodstream. Therefore, accurate localization is essential to achieve high-resolution images. Overlapping PSFs make the aforementioned localization methods inaccurate, so it matters to ensure that individual MBs are well isolated without interference. Accordingly, low concentrations of diluted MBs are commonly employed (Christensen-Jeffries, Browning, et al. 2015; M. A. O'Reilly and Hynynen 2013; Viessmann et al. 2013) and interfering signals due to closely spaced MBs are rejected (Christensen-Jeffries, Browning, et al. 2015; M. A. O'Reilly and Hynynen 2013). However, by reducing the concentrations of MBs, the number of detectable MBs becomes limited, which eventually requires long data acquisition time.

Besides the overlapping PSFs, the image quality also affects on the localization performance. Advanced imaging sequences, e.g., synthetic aperture imaging with diverging (Jensen, Nikolov, et al. 2006) or plane waves (Tanter and Fink 2014), adaptive beamforming (Diamantis et al. 2018), filtering for noise reduction (Song, Trzasko, et al. 2017) can improve MB localization by offering the ultrasound images with higher resolution, contrast, frame rates, and signal-to-noise ratios (SNRs).

2.4 Motion Correction

During *in-vivo* measurements, it is inevitable to avoid motion artifacts by the subject and operator. Considering the resolution of ULM can be several micrometers and the long data acquisition time is required, proper motion compensation can improve the image quality of ULM enormously.

A simple way of removing the motion artifacts is excluding image frames affected by the breathing motion (Christensen-Jeffries, Browning, et al. 2015; F. Lin et al. 2017), however, it is not always applicable. Rigid motion of tissue is often estimated by phase correlation (Song, Trzasko, et al. 2017) or spatial correlation on the B-mode images (Foiret et al. 2017; Taghavi, Andersen, et al. 2021), and more advanced non-rigid motion correction is also available (Harput, Christensen-Jeffries, Brown, et al. 2018).

2.5 Tracking

Tracking is a process of associating estimated MBs temporally over frames and improves the image quality. Intravascular MBs move along the blood flow inside vessels. By taking the temporal correlation of MB movements into account, tracks that give a partial view of the vascular structures can be found. Tracking allows suppressing erroneous MB estimations by removing not associated MBs and short tracks. Moreover, tracking offers velocity information, i.e., the direction and speed of blood flow in the resolution of a few micrometers. The velocity information is important quantities for physicians. And attached microvessels can be distinguished based on the blood flow direction, which cannot be achieved in MB intensity images.

There are elementary tracking methods, e.g., cross correlation (Christensen-Jeffries, Browning, et al. 2015), nearest neighbor (Errico et al. 2015), and Hungarian algorithm (Song, Manduca, et al. 2018). More advanced methods utilizing the Markov chain (Ackermann and Schmitz 2016) or Kalman filtering (Solomon et al. 2019; Taghavi, Schou, et al. 2020) have been investigated.

2.6 Discussion

In this chapter, a general overview of ULM has been reviewed. Using the fact that the isolated single sources can be localized with a sub-wavelength precision, ULM is able to break the resolution limit by accumulating the centers of estimated MBs in an image frame. Also, the use of ULM to various clinical applications is expected (Andersen et al. 2020; Ghosh et al. 2019; C. Lin, Chang, and Chuang 2016; Opacic et al. 2018; Siepmann et al. 2011).

Nevertheless, there are several limitations that make ULM challenging in practice. One is the long data acquisition. Several minutes of ultrasound scans required for ULM are not realistic. On top of that, ultrasound data from the long scanning are likely exposed to more motion artifacts, which potentially results in motion error accumulation. Super-resolution imaging at high concentrations of MB have been studied by exploiting sparse recovery methods (Bar-Zion, Solomon, et al. 2018; Bar-Zion, Tremblay-Darveau, et al. 2016; Solomon et al. 2019). Recently, deep learning methods have been suggested for localizing the overlapping MBs (van Sloun, Solomon, et al. 2021; Youn, Ommen, et al. 2020) or directly estimate tracks (Milecki et al. 2021). For more efficient computation

and generalization, model-based neural networks, embedding prior knowledge into the network architecture, have been applied to MB localization (van Sloun, Cohen, and Eldar 2020; Youn, Luijten, et al. 2020, 2021).

The processing chain of ULM consists of several steps, and the performance of each step is interdependent. Considering there are so many factors affecting ULM, e.g., the SNR, contrast agents, motion, and imaging sequence, a proper way of validating each step of ULM processing is necessary. In general, it is challenging to evaluate *in vivo* results due to the absence of ground truth. To circumvent this problem, the ULM methods were compared with other modalities such as micro-CT (Zhu et al. 2019) or optical microscopy (Christensen-Jeffries, Browning, et al. 2015). Cross-modality validation offers the consistency among different modalities yet the accuracy cannot be measured since all the modalities have their own uncertainties. Alternatively, 3-D PEGDA printed phantoms can be employed for the validation (Ommen et al. 2021). Unlike other flow tube phantoms used in (Harput, Christensen-Jeffries, Ramalli, et al. 2020; Viessmann et al. 2013), more complex structures can be designed by users and fabricated precisely. The use of phantoms for validation is shown in Chapter 3, 4, and 5.

Lastly, there are inherent problems of imaging 3-D structures in the 2-D planes using 1-D array probes. A 2-D ultrasound image is essentially an integration over the elevation beam profile, so ambiguity exists in the data along the elevation direction. This can degrade the localization performance, especially on complex *in vivo* measurements, and the resulting out-of-plane motion cannot be compensated for. These problems can be dealt with 3-D ULM using 2-D array probes such as fully-addressed matrix array probes (Heiles et al. 2019; Provost et al. 2014) or row-column addressed matrix probes (Jensen, Ommen, et al. 2020).

Part II

Fully data-driven methods

CHAPTER 3

Localization on Radiofrequency Channel Data

In this chapter, scatterer localization directly on RF channel data using CNN is given. Ultrasound data generation with the designed imaging sequence and CNN architecture along with non-overlapping Gaussian confidence maps are explained. The methods are evaluated on the simulated test data and phantom measured data. This chapter is based on Paper 1 (Youn, Ommen, et al. 2019) and Paper 2 (Youn, Ommen, et al. 2020).

3.1 Introduction

3.1.1 Motivation

The standard MB localization methods for ULM are limited by ultrasound diffraction since they are performed on conventional delay-and-sum (DAS) beamformed images. In general, overlapping PSFs cannot be easily localized by the standard methods and induce wrong estimates. Here, it is hypothesized that performing localization directly on RF channel data can localize the scatterers spaced closer than the resolution limit of conventional ultrasound imaging.

It has been shown that localization on RF channel data is available by fitting a parabola to the echo from an isolated scatterer and finding its summit (Couture et al. 2011; Desailly et al. 2013). However, the parabola fitting is not suitable for localizing high-density scatterers since it is not straightforward to separate the echoes from multiple closely spaced scatterers. Therefore, a data-driven localization method using CNNs is proposed. CNNs can model complex non-linear mappings by a series of convolution operations and non-linear functions. The mapping from the RF channel data to scatterer positions is estimated using CNNs without beamforming.

3.1.2 Problem Definition

Let us consider received ultrasound RF channel data $\boldsymbol{x} \in \mathbb{R}^{N_a \times N_l \times N_t}$ which are induced by scatterers placed at $\boldsymbol{p} \in \mathbb{R}^{N_s \times 2}$, where N_a is the number of samples in the axial direction, N_l is the number of active elements in reception, N_t is the number of transmission events, N_s is the number of scatterers, and 2 is the number of the spatial dimensions, i.e., the lateral and axial directions. To localize the scatterers from the RF channel data, a mapping $f\colon \mathbb{R}^{N_a \times N_l \times N_t} \to \mathbb{R}^{N_s \times 2}$ satisfying

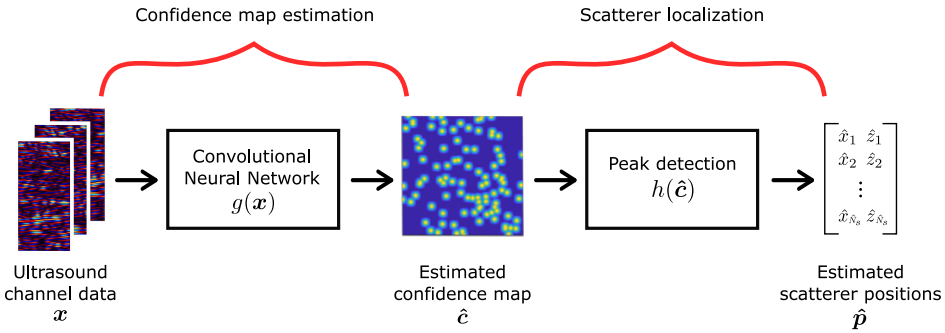


Figure 3.1: Overview of scatterer localization from ultrasound RF channel data. A two-stage process was adopted to handle the varying number of scatterers. A CNN formed a confidence map from ultrasound RF channel data and scatterer localization using local peak detection was followed. The illustration is modified from Paper 2 (Youn, Ommen, et al. 2020).

$$\boldsymbol{p} = f\left(\boldsymbol{x}\right) \tag{3.1}$$

needs to be found.

The number of scatterers N_s are different depending on the input ultrasound images, so the mapping f should be able to manage the varying number of scatterers. Fully-connected neural networks are not suitable for this application as their output size is commonly fixed. Therefore, a two-stage process is used to handle the varying number of scatterers by reformulating the mapping f as follows:

$$\boldsymbol{p} = f\left(\boldsymbol{x}\right) = h\left(g\left(\boldsymbol{x}\right)\right) = h\left(\boldsymbol{c}\right). \tag{3.2}$$

The function $g\colon \mathbb{R}^{N_a \times N_l \times N_t} \to \mathbb{R}^{N_z \times N_x}$ produces a confidence map $\boldsymbol{c} \in \mathbb{R}^{N_z \times N_x}$, where N_z and N_x are the number of samples in the axial and lateral directions, respectively. The confidence map represents the spatial domain of a region of interest (ROI) whose pixel value indicates the confidence of scatterer presence in the corresponding pixel location. The mapping $h\colon \mathbb{R}^{N_z \times N_x} \to \mathbb{R}^{N_s \times 2}$ locates scatterers from the confidence map. The confidence map estimation, i.e., the mapping g, was modeled by a fully CNN, and scatterer localization, i.e., the mapping h, was implemented by local peak detection. The overview of the method is illustrated in Fig. 3.1.

3.2 Ultrasound Data Generation

CNNs require a large amount of training data with labels, i.e., true scatterer positions. However, it is extremely difficult to acquire measured data with ground truth for these

Table 3.1: Ultrasound RF channel data simulation parameters

Category	Parameter	Value
Transducer	Transmission frequency	$5.2\,\mathrm{MHz}$
	Pitch	$0.20\,\mathrm{mm}$
	Element width	$0.18\,\mathrm{mm}$
	Element height	$6\,\mathrm{mm}$
	Number of elements	192
Imaging	Number of TX elements	32
	Number of RX elements (N_l)	64
	Steering angles	$-15°$, $0°$, $15°$
Environment	Speed of sound (c)	$1480\,\mathrm{m/s}$
	Field II sampling frequency	$120\,\mathrm{MHz}$
	RF channel data sampling frequency	$29.6\,\mathrm{MHz}$
Scatterer	Number of scatterers (N_s)	$20 \cdot i, \forall i \in \{1, 2, \ldots, 10\}$
	Lateral range	$(-3.2, 3.2)\,\mathrm{mm}$
	Axial range	$(14.8, 21.2)\,\mathrm{mm}$

kinds of works. Alternatively, ultrasound RF channel data were simulated for training, validation, and evaluation. The simulation was performed in Field II pro with the parameter values in Table 3.1. It is important to simulate the channel data as close to measured data as possible; otherwise, the trained CNN will suffer a generalization problem. For this reason, the impulse response of a commercial ultrasound probe used for experiments was measured (Tomov et al. 2018) and applied in the simulation (Jensen 2016).

One image frame was generated by placing point scatterers randomly in the $6.4\,\mathrm{mm} \times 6.4\,\mathrm{mm}$ region and simulating three steered plane waves. Common plane wave imaging employs all the elements both in transmit and receive (Tanter and Fink 2014). Here, however, to insonify the ROI only, different sub-apertures were defined in transmission using 32 elements for each steering angle, as shown in Fig. 3.2. For receiving backscattered signals, 64 elements in the center of the probe were used to reduce data rate.

The raw RF channel data consisted of parabolic wavefronts, as shown in Fig. 3.3(a), which makes the confidence map estimation complicated. To ease the problem by making wavefronts more like straight lines, a pre-processing was applied to the channel data by delaying the signal according to the following time-of-flight,

$$\tau_i(x, z) = \left(\sqrt{(x - x_i)^2 + z^2} + z \right) / c, \tag{3.3}$$

where τ_i is the time-of-flight of the i-th transmission, (x, z) is the data point, x_i is the center of the i-th transmission aperture, and c is the speed of sound. After the pre-processing, the wavefronts became more like lines, as shown in Fig. 3.3(b). This pre-processing has improved the CNN performance on validation data. Note that the

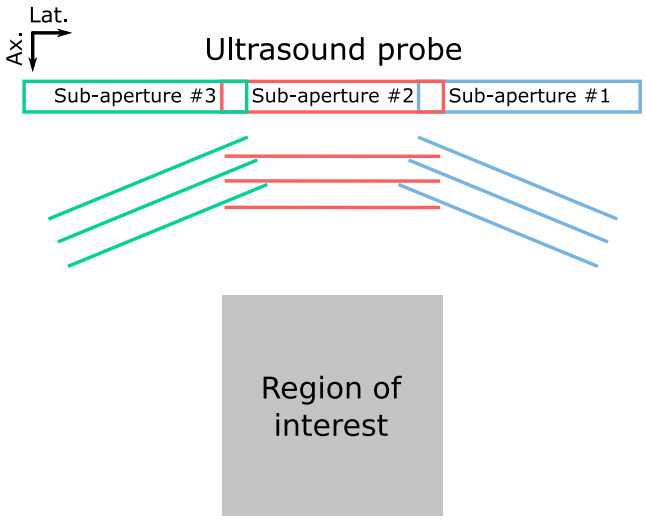


Figure 3.2: Illustration of transmission scheme. For one image frame, three plane waves were transmitted. To insonify the ROI only, different sub-apertures were defined using 32 elements for the steered ultrasound beam transmissions. The illustration is modified from Paper 2 (Youn, Ommen, et al. 2020).

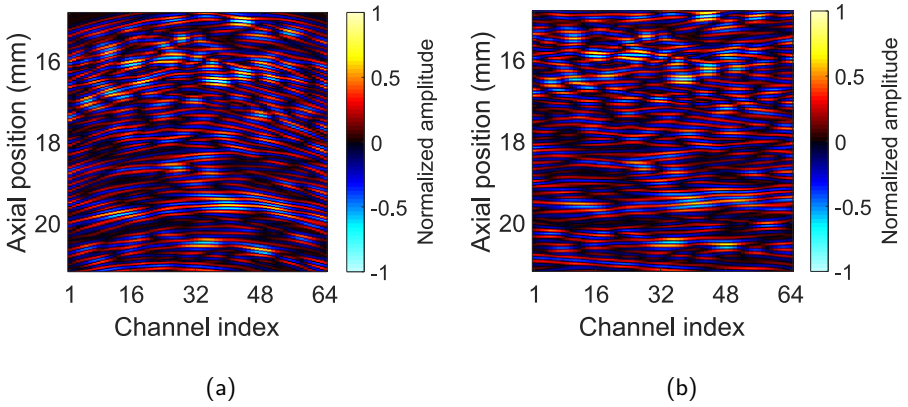


Figure 3.3: Example of simulated RF channel data. (a) is simulated raw RF channel data and (b) is delayed RF channel data. Note that the delay here is different from the delay for beamforming. The figure is modified from Paper 2 (Youn, Ommen, et al. 2020).

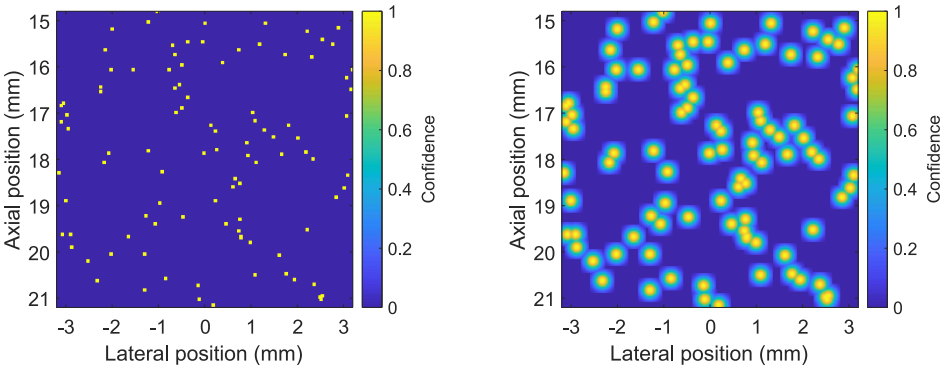


Figure 3.4: Comparison of 2-D binary and non-overlapping Gaussian confidence maps. (a) is the binary confidence map that is so sparse that large enough gradients for stable training cannot be provided during gradient descent-based optimization. (b) is the non-overlapping Gaussian confidence map that provides large gradients for stable training by Gaussian filtering while being able to recover closely spaced scatterers correctly thanks to the maximum operation.

delay for this pre-processing is performed in the channel data domain and different from the delay for beamforming.

A high sampling frequency was used in Field II pro to avoid numerical errors that can perturb the accuracy of the simulation. The proposed CNN, which will be introduced in Section 3.4, was designed that the input and output data have the same number of samples along the axial direction. The RF channel data were, therefore, downsampled to match the size of the data along the axial direction with that of confidence maps, i.e., $N_a = N_z$. Pixel size of the confidence maps effectively defined the final sampling frequency of the channel data.

3.3 Confidence Map

3.3.1 Non-overlapping Gaussian Confidence Map

Confidence maps represent the presence of scatterers in the spatial domain. The simplest confidence map is a binary confidence map whose pixel value is 1 if a scatterer is present in the corresponding pixel location, and 0 otherwise, as shown in Fig. 3.4(a). Initially, CNNs were trained using the binary confidence maps, however, the estimated confidence maps by the trained CNNs were always 0 irrespective of input channel data. The CNNs

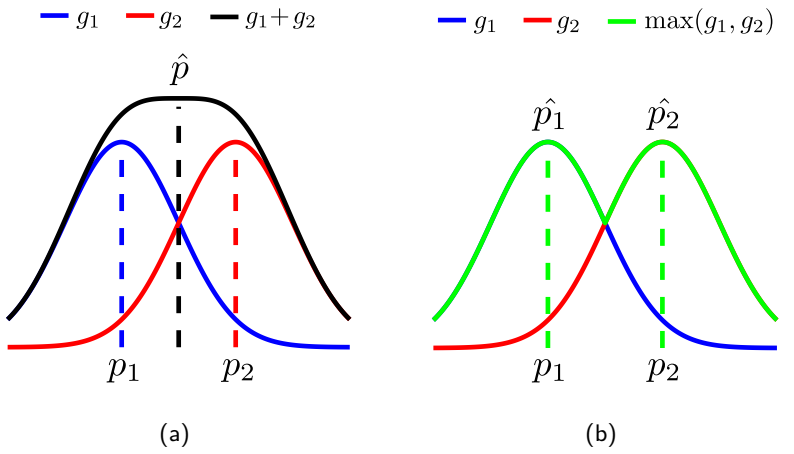


Figure 3.5: Comparison of 1-D Gaussian and non-overlapping Gaussian confidence maps. There are two scatterers at p_1 and p_2, and g_1 and g_2 are Gaussians applied to their positions, respectively. The black curve in (a) is the Gaussian confidence map created by the summation of g_1 and g_2. The green curve in (b) is the non-overlapping Gaussian confidence map created by the maximum of g_1 and g_2. In (b), two scatterers $\hat{p}_1$ and $\hat{p}_2$ can be estimated at correct positions from the confidence map, however, in (a), one scatterer $\hat{p}$ is found at a wrong position. The figure is modified from Paper 2 (Youn, Ommen, et al. 2020)

were optimized by the gradient descent algorithm, but the binary confidence maps were too sparse to provide large enough gradients for stable training; therefore, the CNNs were trained to output 0's. Advanced losses such as weighted cross entropy (Ronneberger, Fischer, and Brox 2015), jaccard loss (Jaccard 1912), or focal loss (Lin et al. 2017) were also considered, but did not work.

To relax the sparsity of the binary confidence maps while being able to localize closely spaced scatterers correctly from confidence maps, a non-overlapping Gaussian confidence map was proposed, as shown in Fig. 3.4(b). It has been reported that stable training is available by applying a Gaussian filter to sparse labels (Gomariz et al. 2019; Nehme et al. 2018; van Sloun et al. 2021). But such naïve Gaussian filtering occurs a problem similar to overlapping PSFs in high-density scatterer localization as the resulting confidence map after Gaussian filtering is essentially a summation of the Gaussians at each scatterer position. Let us consider an 1-D example, where two scatterers are placed closer than the full width at half maximum (FWHM) of the Gaussian filter. Then their positions cannot be recovered from the confidence map because a single peak will appear between two scatterers, as shown in Fig. 3.5(a).

Fortunately, the confidence maps are labels that can be controlled unlike the overlapping PSFs which are physical effect. Therefore, the non-overlapping Gaussian confidence map was suggested to avoid the overlaps among the Gaussians. It is created by applying the Gaussian filter to each scatterer position separately and taking the maximum of all the Gaussians. Now, in the non-overlapping Gaussian confidence map, the two closely spaced scatterers can be separated correctly, as shown in Fig. 3.5(b). The operation of creating the non-overlapping Gaussian confidence maps is non-linear and takes more time compared to simple Gaussian filtering. However, it needs to be run only one time when preparing for training data, and the peaks correspond to the scatterer positions even when the scatterers are spaced closer than the FWHM of the Gaussian.

The standard deviation of the Gaussian filter is a hyper-parameter that determines the smoothness of Gaussians. To maximize the localization performance, it is important to find an optimal standard deviation. For example, a large standard deviation will make localization of closely spaced scatterers difficult. On the other hand, a small standard deviation will not provide large enough gradients for stable training. For this work, the standard deviation of 5 pixels was chosen through validation. The scatterer positions were quantized and represented in the confidence map with respect to the pixel coordinates with the pixel size of $25\,\mu\text{m}$ ($\approx \lambda/10$).

3.3.2 Scatterer Localization on Confidence Map

In the estimated confidence maps $\hat{c}$, scatterers need to be localized. Ideally, pixels containing scatterers and their neighboring pixels are supposed to follow the Gaussian function, as shown in Fig. 3.4(b) and 3.5(b). Based on this fact, localization was implemented by finding local peaks in the confidence maps. As localization is performed on the estimated confidence maps, there can be unwanted local peaks that do not correspond to true scatterers. To avoid such wrong estimations, the peaks having confidences higher than a certain value were accepted. The threshold value of 0.9 was chosen heuristically.

3.4 Convolutional Neural Network

3.4.1 Network Architecture

An encoder-decoder structured CNN has been designed to reconstruct confidence maps from ultrasound RF channel data. The input and output are not in the same domain unlike common fully CNN applications such as semantic segmentation (Badrinarayanan, Kendall, and Cipolla 2017; Ronneberger, Fischer, and Brox 2015) single-image super-resolution(J. Kim, J. K. Lee, and K. M. Lee 2016; Ledig et al. 2017; Lim et al. 2017). The input is in the channel data domain, but the output is in the ultrasound image domain; therefore, it can also be interpreted that the CNN performs beamforming implicitly in the confidence map estimation. In the encoding path, features are extracted from the channel data and represented in the latent space. In the decoding path, the corresponding confidence maps are produced from the extracted features.

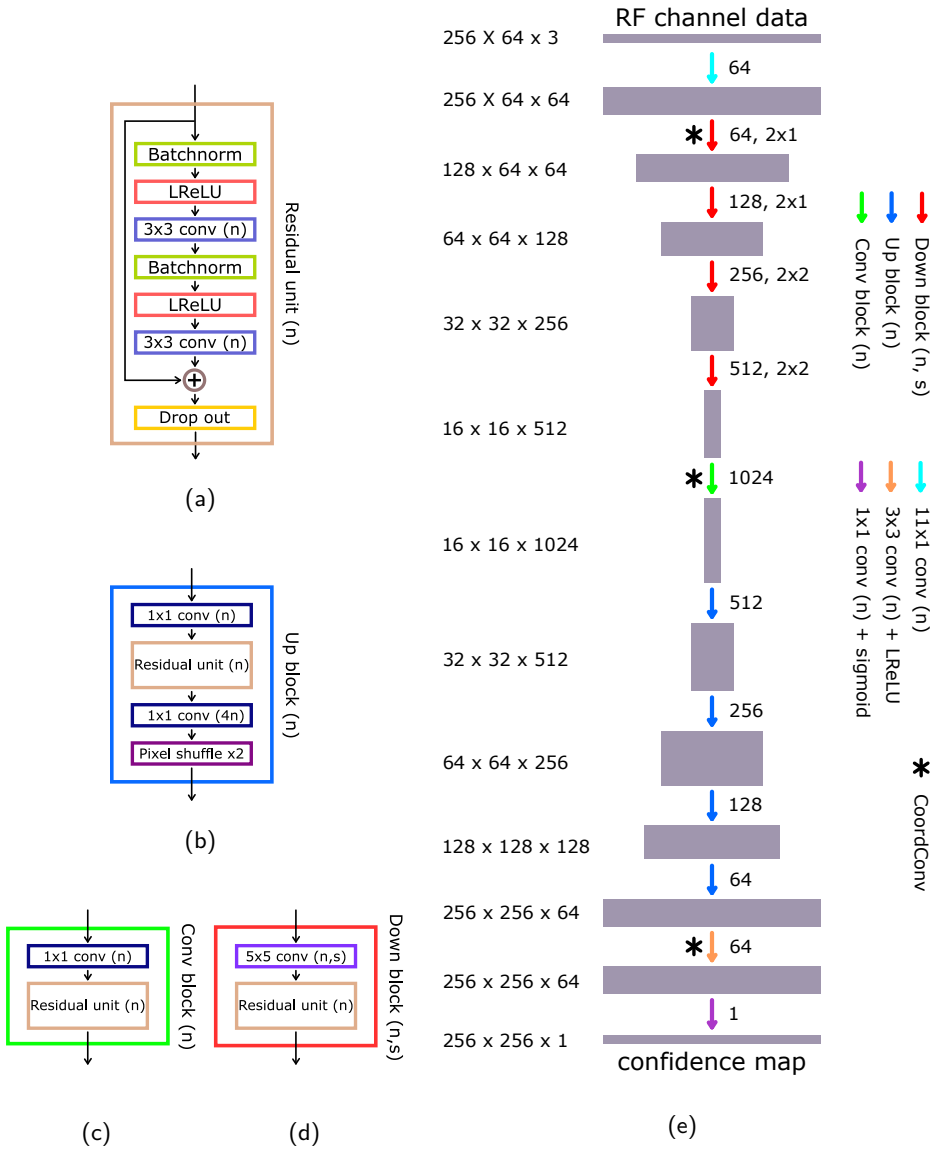


Figure 3.6: Proposed CNN architecture and its blocks. (a) is the modified pre-activation residual unit, (b) is the *down-block*, (c) is the *conv-block*, (d) is the *up-block*, and (e) is the network architecture. In (e), the number of kernels (n) and stride (s), if necessary, are presented for each block next to the arrows. The asterisk represents that CoordConv (Liu et al. 2018) was applied to its first convolution layer. The feature size was given in the form of (height $\times$ width $\times$ kernel). This illustration is modified from Paper 1 (Youn, Ommen, et al. 2019) and Paper 2 (Youn, Ommen, et al. 2020)

A detailed illustration of the proposed CNN architecture is shown in Fig. 3.6. The major parts of the network were composed of four *down-blocks* (Fig. 3.6(c)), one *conv-block* (Fig. 3.6(d)), and four *up-blocks* (Fig. 3.6(b)), and each block was based on pre-activation residual units (He et al. 2016b) (Fig. 3.6(a)). The residual units have advantages over conventional convolution and rectified linear unit (ReLU) (Nair and Hinton 2010) layers in the sense that they allow training deeper neural networks and, as a result, achieving better performance. In theory, the deeper networks can achieve better performance as they can represent more complex non-linear functions. However, in practice, training the deeper networks is challenging due to the gradient vanishing problem, and the network performance gets saturated and degrades as increasing the depth. The residual units ease the optimization problem by learning residual mappings instead of direct mappings; therefore, the deeper networks can be trained effectively (He et al. 2016a,b).

The original pre-activation unit does not include dropout (Srivastava et al. 2014), but that was necessary to apply the proposed network to measured ultrasound data due to the generalization problem. It was considered to put dropout layers after each convolution layer or between two convolution layers in the residual unit as proposed in (Zagoruyko and Komodakis 2016). However, it was found that putting the dropout layers after the element-wise summation achieves the best performance through validation. Additionally, batch normalization was added to stabilize training by suppressing the internal covariate shift and gain the regularization effect (Ioffe and Szegedy 2015). Lim *et al*. argued that batch normalization layers need to be removed as they remove range flexibility by normalizing the features for single-image super-resolution (Lim et al. 2017). Nonetheless, batch normalization was essential for the confidence map estimation here. Presumably, the internal covariate shift is more drastic when the discrepancy between the input, i.e., channel data, and the output data, i.e., confidence map, is large.

Before the encoding path, an 11×1 convolution layer was placed to extract per-channel features. In the encoding path, the *down-blocks* downsampled the features using strided convolution. Downsampling helps decrease the number of parameters and provides different receptive fields without changing the kernel size. Max pooling or average pooling are generally used for downsampling, but the spatial information can be lost. Therefore, strided convolution was chosen to keep the spatial information. In the decoding path, the *up-blocks* upsampled the downsampled features to the size of the confidence maps. For effective and efficient upsampling, pixel shuffle was employed (Shi et al. 2016). Lastly, two convolution layers were placed to refine the confidence maps. For activation, Leaky ReLUs (Maas, Hannun, and Ng 2013) were selected to avoid the dying ReLU problem and Sigmoid was used in the output layer to force to the output values to be $[0, 1]$.

For the confidence map estimation, the large receptive field was required. The backscattered signal by a single scatterer appears across all the channels at several depths, although the pre-processing introduced in Section 3.2 makes wavefronts more like straight lines. Hence, local information in the RF channel data is not enough to localize a scatterer, and the proposed CNN was established by four *down-* and four *up-blocks*. Skip connection is a widely used technique to transfer the spatial information over the convolution layers

(Drozdzal et al. 2016; Ronneberger, Fischer, and Brox 2015). However, skip connections were not able to be implemented in the proposed network since training failed when they are included. The features extracted from the channel data in the encoding path are not correlated to the reconstruction of the confidence maps; therefore, the skip connections deterred training. Alternatively, CoordConv (Liu et al. 2018) was utilized for certain convolution layers, which is indicated in Fig. 3.6(e).

3.4.2 Training Detail

Training was performed by optimizing the mean squared error (MSE) between true and estimated confidence maps,

$$\mathcal{L}_{\text{MSE}}\left(\boldsymbol{x}, \boldsymbol{c}; g\right) = \frac{1}{N}\sum_{i=1}^{N}\left\|\boldsymbol{c}_i - g\left(\boldsymbol{x}_i; \theta\right)\right\|_F^2, \tag{3.4}$$

where x_i and c_i are the i-th ultrasound RF channel data and corresponding confidence map, g is the proposed CNN with learning parameters θ, N is the number of samples, and $\|\cdot\|_F$ is the Frobenius norm.

Training and validation data were simulated at four different scatterer densities of $0.49\,\mathrm{mm}^{-2}$, $0.98\,\mathrm{mm}^{-2}$, $2.44\,\mathrm{mm}^{-2}$, and $4.88\,\mathrm{mm}^{-2}$ by changing the number of scatterers for one image frame in the simulation. For each scatterer density, the number of training and validation data were 10 240 and 1280 frames, respectively. The network parameters were initialized by the orthogonal initialization (Saxe, McClelland, and Ganguli 2013) and the ADAM (Kingma and Ba 2015) optimizer was employed with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-7}$. The network was trained for 800 epochs with the batch size of 32. For the first 600 epochs, training was performed using the training data simulated at the scatterer density of $2.44\,\mathrm{mm}^{-2}$ with the learning rate of 10^{-4}, and the learning rate was halved every 100 epochs. And then, the training continued for 200 epochs using all the training data, where the learning rate was 10^{-5} and it was halved every 50 epochs. The two-phase training was more efficient in terms of convergence, i.e., requiring fewer iterations to the solution, compared with training a CNN using all the training data from scratch. The CNNs were implemented using Tensorflow (Abadi et al. 2011) in Python, and training took approximately 40 hours in a server equipped with a NVIDIA TESLA V100 16 GB PCIe graphics card.

Having more training data is preferable because the CNNs can learn more diverse data distributions. The correspondence between RF channel data and confidence maps are valid after being flipped along the lateral direction. Therefore, for data augmentation, the training data were horizontally flipped at random during training. Additionally, perturbing the training data allows the CNNs to be robust to the noise and improves their generalizability. So, additive white Gaussian noise was added for generalization. Especially, the Gaussian noise was added during training to provide independent noise at each training iteration. The dropout rate was set to 0.3. Lastly, the RF channel data

and confidence maps were normalized by their maximum values, and thus their ranges became $[-1, 1]$ and $[0, 1]$.

3.5 Simulation Experiment

The capability of the trained CNN was assessed on simulated test data. The test data were generated at 10 scatterer densities from $0.49\,\mathrm{mm}^{-2}$ to $4.88\,\mathrm{mm}^{-2}$ to validate the performance at different scatterer densities, i.e., different degrees of overlaps. For each scatterer density, 3840 data were generated using the parameter values in Table 3.1.

3.5.1 Evaluation Metric

To quantify the localization performance, the estimated scatterers need to be associated with true scatterers to decide whether an estimated scatterer is positive or negative detection. Simply finding the closest true scatterer given an estimated scatterer can encounter problematic situations where one true scatterer is associated with multiple estimated scatterers. Therefore, a bi-directional matching process was proposed, inspired by the left-right consistency check in computer vision (Chang, Chatterjee, and Kube 1991; Fua 1993). A detailed procedure is described in Algorithm 3.1. The bi-directional matching process satisfies the uniqueness constraint, i.e., a true scatterer is either matched with only one estimated scatterer or not matched.

Algorithm 3.1: Data association for determining positive or negative detection.

Input: True scatterer positions $\boldsymbol{p} \in \mathbb{R}^{N_s \times 2}$ and estimated scatterer positions $\hat{\boldsymbol{p}} \in \mathbb{R}^{\hat{N}_s \times 2}$
Output: A vector represents positive or negative detection $\boldsymbol{y} \in \mathbb{R}^{\hat{N}_s \times 1}$

$\boldsymbol{y} \leftarrow \mathbf{0} \in \mathbb{R}^{\hat{N}_s \times 1}$ // Initialization
$D \leftarrow \left\{ (d_{ij}) \in \mathbb{R}^{N_s \times \hat{N}_s} \,\middle|\, d_{ij} = \|p_i - \hat{p}_j\|_2 \right\}$ // Pairwise distance
for $j = 1$ to $\hat{N}_s$ **do**
  $\hat{i} \leftarrow \arg\min D_{*j}$
  if $j = \arg\min D_{\hat{i}*}$ **then**
    if $\frac{(p_{\hat{i}1} - \hat{p}_{j1})^2}{(\mathrm{FWHM}_x/2)^2} + \frac{(p_{\hat{i}2} - \hat{p}_{j2})^2}{(\mathrm{FWHM}_z/2)^2} < 1$ **then** // Localization error
      $a_j \leftarrow 1$ // Positive detection
    else
      $a_j \leftarrow 0$ // Negative detection
    end if
  else
    $a_j \leftarrow 0$ // Negative detection
  end if
end for

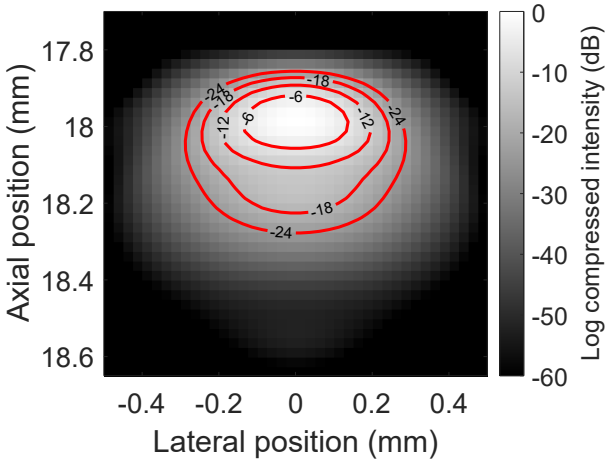


Figure 3.7: A simulated PSF at the center of the ROI. The −6 dB contour can be approximated to an ellipse.

Basically, an estimated scatterer is decided to be a positive detection when it is exclusively matched with a true scatterer while the localization error is smaller than a certain localization error. The acceptable localization error can be related to the target resolution of ULM, and the half of the FWHM was set to the criteria, which was defined from a simulated PSF at the center of the ROI, as shown in Fig. 3.7. The FWHM in the lateral and axial directions were 265 µm (0.93λ) and 140 µm (0.49λ), and the −6 dB contour, i.e., FWHM, can be approximated to an ellipse. Therefore, the half of FWHM was modeled as an ellipse whose major axis is half of the lateral FWHM (FWHM_x) and minor axis is half of the axial FWHM (FWHM_z).

After matching the estimated scatterers with the true scatterers, the performance was assessed by precision and recall, localization precision, and resolved rate. Precision and recall were defined by

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{3.5}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{3.6}$$

where TP is the number of true positives, i.e., correct estimations, FP is the number of false positives, i.e., wrong estimations, and FN is the number of false negatives, i.e., missed scatterers. Localization precision was measured by the standard deviation of localization errors of positive detections in the lateral and axial directions. The spatial resolution refers to the ability of how closely spaced targets can be separated. It was

measured statistically by finding pairs of two isolated true scatterers and checking whether they were matched with the estimated scatterers, i.e., correctly detected, or not. If both scatterers were detected, the pair was regarded as a resolved case. If only one of them was detected, the pair was regarded as a not resolved case. And the case that none of them were detected was not considered. And then, the resolved rate was calculated by

$$\text{Resolved rate} = \frac{N_{\text{res}}}{N_{\text{res}+N_{\text{non-res}}}}, \tag{3.7}$$

where N_{res} is the number of resolved pairs and $N_{\text{non-res}}$ is the number of non-resolved pairs.

3.5.2 Result

The proposed method was compared with peak detection and deconvolution. Peak detection was performed by finding local peaks on the DAS beamformed and compounded images. Deconvolution was performed on the same images using Richardson–Lucy (RL) deconvolution (Lucy 1974; Richardson 1972) with the PSF simulated at the center of the ROI in Fig 3.7 as a reference. The localization results by different methods on a test image frame are shown in Fig. 3.8. The isolated scatterers were well localized by all the methods. On the other hand, the closely spaced scatterers were localized correctly by the proposed method but not by peak detection and deconvolution.

Precision and recall, and localization precision at various scatterer densities are shown in Fig. 3.9 and Fig. 3.10, respectively. As the scatterer density increases, the performance degraded for all the methods because more overlapping PSFs appear. However, for all the metrics, the proposed method outperformed peak detection and deconvolution. Also, the proposed method kept relatively high performance at high scatterer densities, showing that it can handle a certain degree of overlaps. Additionally, the lateral localization precision was worse than the axial localization precision for all the methods, showing that localization along the lateral direction is typically more difficult in ultrasound data for the CNN-based proposed method, as well as the standard methods such as peak detection and deconvolution.

Fig. 3.11 visualizes 2-D histograms of the resolved rate in the $20\,\mu\text{m} \times 20\,\mu\text{m}$ grids. The blue curve indicates the theoretical resolution limit estimated by the FWHM of the PSF simulated in the center of the ROI in Fig 3.7. The resolution limit, i.e., $-6\,\text{dB}$ contour, was assumed to be an ellipse whose major axis is the FWHM_x, 265 µm (0.93λ), and minor axis is the FWHM_z, 140 µm (0.49λ). It is clearly shown that the proposed method can separate the scatterers spaced closer than the resolution limit compared to peak detection and deconvolution. The mean resolved rate was measured in the region under the green curves, i.e., two scatterers placed closer than the resolution limit. The proposed method achieved the mean resolved rate of 0.67, but peak detection and deconvolution achieved the mean resolved rate of 0.17 and 0.11, respectively.

Deconvolution has been employed for MB localization (Couture et al. 2011; Siepmann et al. 2011) and Yu *et al.* have demonstrated that sub-wavelength localization using

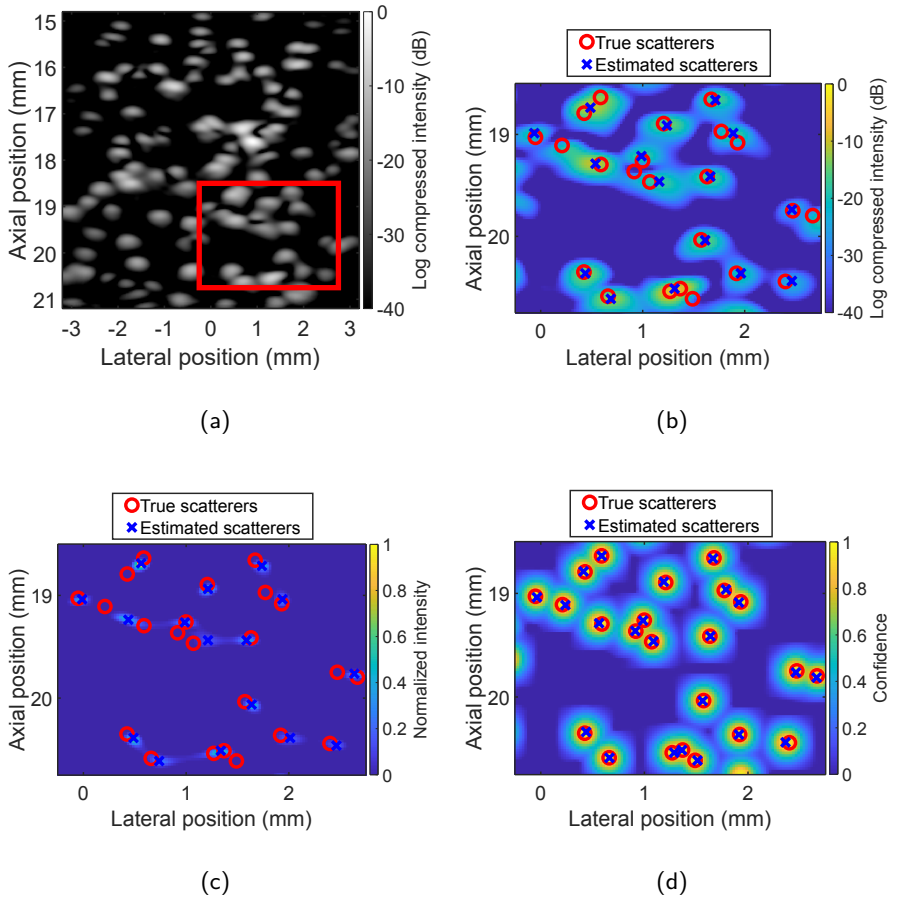


Figure 3.8: Scatterer localization on a simulated test frame. (a) is the B-mode image beamformed by DAS and compounded. (b)-(d) are the localization results by different methods in the red rectangle region in (a). (b) is peak detection, (c) is deconvolution, and (d) is the proposed method.

deconvolution on simulated data (Yu, Lavery, and K. Kim 2018). Even so, the performance of peak detection surpassed that of deconvolution in all aspects. RL deconvolution, the chosen deconvolution method, is an iterative algorithm that deblurs images using a reference PSF; therefore, PSFs in the image need to be spatially stationary. However, in the given ultrasound images, the PSF was spatially varied more dynamically than usual plane wave images since the designed imaging sequence constrained the aperture

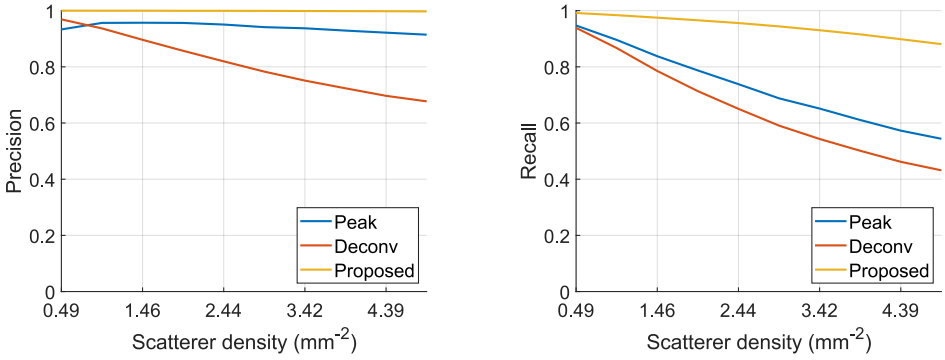


Figure 3.9: Precision and recall on simulated test data by peak detection, deconvolution, and the proposed method. (a) is precision and (b) is recall. The figure is modified from Paper 1 (Youn, Ommen, et al. 2019) and Paper 2 (Youn, Ommen, et al. 2020).

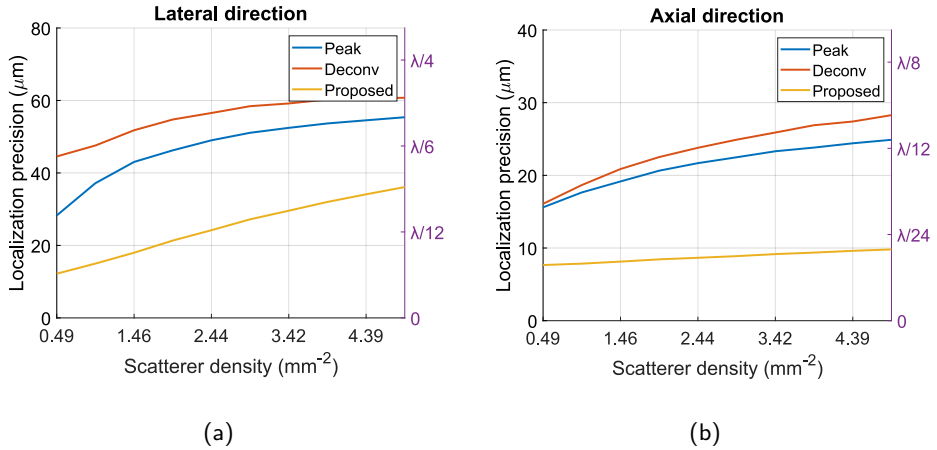


Figure 3.10: Localization precision on simulated test data by peak detection, deconvolution, and the proposed method. (a) is the lateral localization precision and (b) is the axial localization precision.

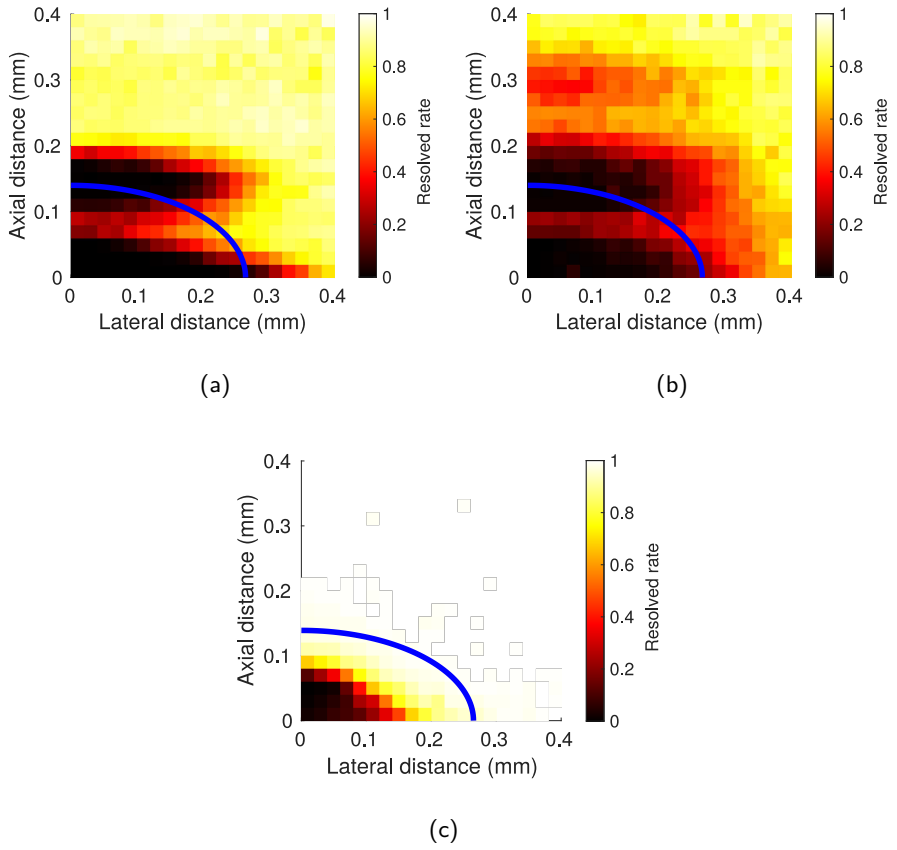


Figure 3.11: 2-D histograms of resolved rate calculated in a $20\,\mu\mathrm{m} \times 20\,\mu\mathrm{m}$ grid. (a) is peak detection, (b) is deconvolution, and (c) is the proposed method. The figure is modified from Paper 2 (Youn, Ommen, et al. 2020).

size in transmitting and receiving ultrasound beams. As a result, deconvolution was not suitable due to the highly variant PSFs; nonetheless, it cannot be generalized that peak detection is superior to deconvolution for localization since the designed imaging sequence is somewhat of an exception.

3.6 Phantom Experiment

3-D printed phantoms were fabricated and scanned to validate the proposed method on measured data and compare it with peak detection. Deconvolution was not considered as its performance on the simulated test data was not as good as peak detection due to the exceptional imaging sequence, and extensive parameter tuning was required in the measured data.

3.6.1 Phantom Fabrication

Two PEGDA 700 g/mol hydrogel phantoms (Fig. 3.12(a)) that contained water-filled cavities inside were fabricated (Ommen et al. 2019, 2021). The cavities act as scatterers because the acoustic property of water and the phantom medium are different. The size of the cavities was $45\,\mu\mathrm{m} \times 45\,\mu\mathrm{m}$ in the imaging plane and 1 mm in the elevation direction. As it was assumed that the scatterers are infinitesimally small points in the simulation, the cavities were made as small as possible in the imaging plane. Contrarily, they were relatively long in the elevation direction to maximize the backscattering energy.

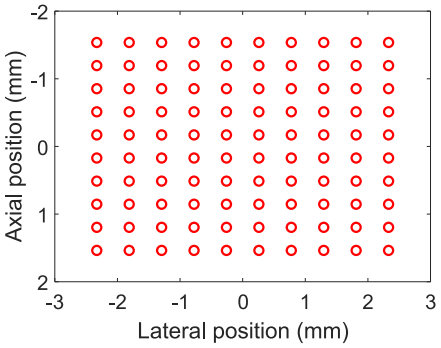
For the first phantom, 100 cavities were placed uniformly in a 10×10 grid at a spacing of 518 µm laterally and 342 µm axially, as shown in Fig. 3.12(b). The purpose of this *uniform* phantom was to validate whether the scatterers are placed as designed and the trained CNN can be generalized to the measured data. Hence, the spacing among the cavities was set to be larger than the resolution limit of the DAS beamforming, i.e., FWHM. On the other hand, for the second phantom, 100 cavities were placed randomly with a minimum spacing of 190 µm among the scatterers, as illustrated in Fig. 3.12(c). This *random* phantom was made to check if the proposed method can localize closely spaced scatterers. The minimum spacing among cavities was introduced due to the limit in the voxel size of the 3-D printer.

3.6.2 Experiment Setup

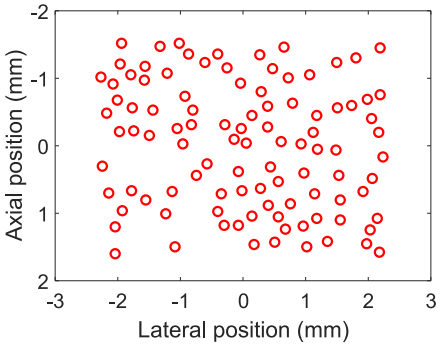
The 3-D printed phantoms were scanned with a commercial 192-element linear array probe whose specification followed the parameter values in Table 3.1. The same imaging sequence with the simulation was implemented, and raw RF channel data were acquired in synthetic aperture real-time ultrasound system (SARUS).

The experiment setup is given in Fig. 3.13. The whole setup was placed on a optical table to control the disturbance caused by vibration, as seen in Fig. 3.13(a). The probe was fixated to the probe fixture. The phantom was submerged in a water tank and the water tank was placed on the motion stage, as illustrated in Fig. 3.13(b). The motion stage can rotate around the z-direction and translate in the x- and y-directions. So, the phantom was able to be aligned in the imaging plane by the motion stage. And then, 33 frames were obtained by translating the phantom in a step of 50 µm in the x-direction, i.e., lateral direction, between the frames.

(a)

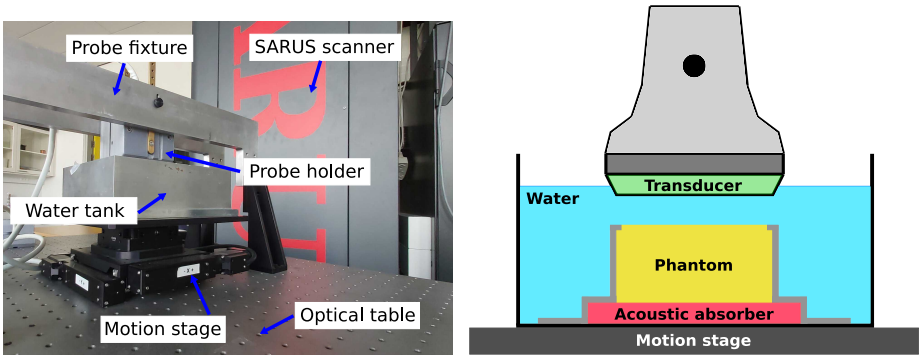


(b)



(c)

Figure 3.12: 3-D printed phantom and scatterer positions. (a) is the picture of a 3-D printed PEGDA hydrogel phantom. (b) is the scatterer placement of the *uniform* phantom and (c) is the scatterer placement of the *random* phantom. The picture is modified from Paper 2 (Youn, Luijten, et al. 2020)



(a) (b)

Figure 3.13: Phantom experiment setup. (a) is the picture of the experiment setup and (b) is the illustration of the experiment setup. The figure was modified from Paper 2 (Youn, Ommen, et al. 2020).

3.6.3 Training Data Modification

The initially trained CNN did not work properly on the phantom measurements due to the generalization problem, as shown in Fig. 3.14(a). Even though the cavities were made as small as possible, that was still not enough for the CNN to recognize individual scatterers correctly since they were modeled by infinitesimally small point scatterers in the simulation.

Accordingly, the training data were updated by considering the physical aspects of the cavities. A simplified 1-D illustration of scattering at a cavity in the phantom along the axial direction is shown in Fig. 3.15. Firstly, the scatterers were modeled by two point scatterers because scattering happens twice. One is when an ultrasound beam goes into the cavity and the other is when the beam comes out of the cavity. Furthermore, the sign of the first scattering amplitude was changed since the acoustic impedance is higher in the phantom than water, so the first scattering experiences the phase reversal. In the updated training data, the scatterer positions of the original training data were used to maintain consistency.

For the phantom experiment, a new CNN was trained from scratch with the updated training data. It was considered to apply transfer learning to the CNN trained with the original training data, but that did not solve the generalization problem. The newly trained CNN estimated confidence maps more accurately and individual scatterers were able to be identified, as shown in Fig. 3.14(b).

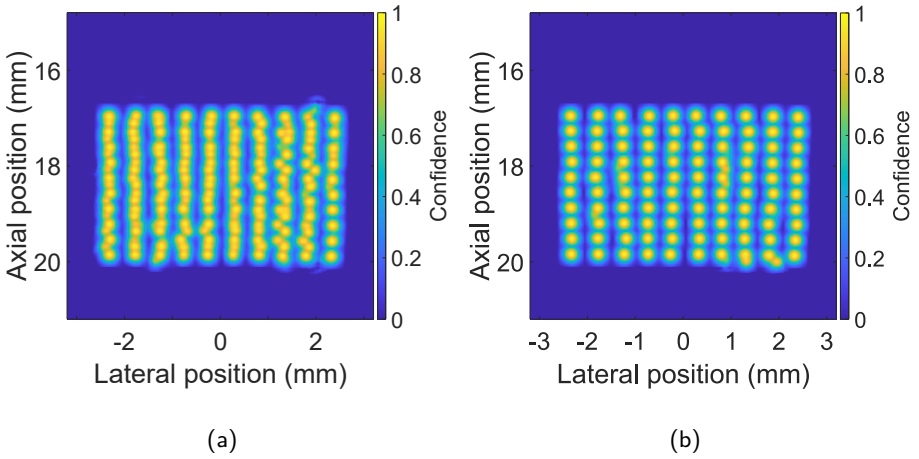


Figure 3.14: Confidence map estimation on phantom measured RF channel data. (a) is before and (b) is after the training data modification.

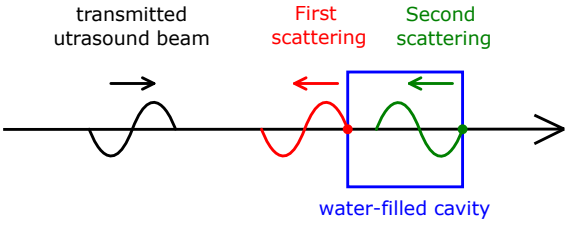


Figure 3.15: Simplified 1-D illustration of scattering in a cavity of the phantom in the axial direction. Two scattering happens. One is when an ultrasound beam goes into the cavity and the other is when the beam comes out of the cavity. The first scattering experiences phase reversal as the acoustic impedance is higher in the phantom medium than in water.

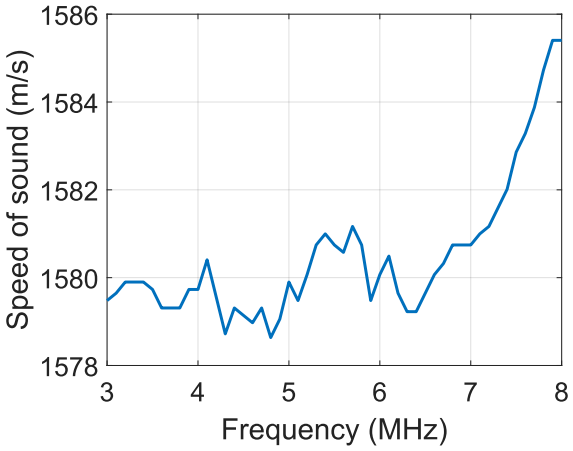


Figure 3.16: Measured speed of sound in the 3-D printed phantom at various frequencies.

3.6.4 Depth Correction

The speed of sound in the phantom was not identical to that in water. The measured speed of sound in the phantom medium is shown in Fig. 3.16. It changed depending on the frequency of the ultrasound beam and was faster than the speed of sound in water. For this reason, the estimated scatterers appeared at the shallower depths than the designed positions. So, the different speed of sound was compensated by correcting the axial positions of the estimated scatterers as follows:

$$\hat{z}^* = (\hat{z} - d_{\text{pht}}) \cdot \frac{c_{\text{water}}}{c_{\text{pht}}} + d_{\text{pht}}, \tag{3.8}$$

where $\hat{z}$ and $\hat{z}^*$ are the estimated axial position of a scatterer before and after the depth correction, c_{water} and c_{pht} are the speed of sound in water and in the phantom, and d_{pht} is the depth of the upper surface of the phantom.

3.6.5 Results

Fig. 3.17 shows scatterer localization on one of the phantom measured data by peak detection and the proposed method. On the *uniform* phantom, both methods successfully localized scatterers accurately. However, it was difficult to localize the closely spaced scatterers on the *random* phantom, which agrees with the simulation results.

Precision, recall, and localization precision are given in Table 3.2. The proposed method achieved slightly lower precision due to few false estimations, but the localization precision was better than centroid detection on the *uniform* phantom. On the *random* phantom, the proposed method achieved better precision and recall by localizing the closely spaced scatterers. For localization, the proposed method achieved worse lateral

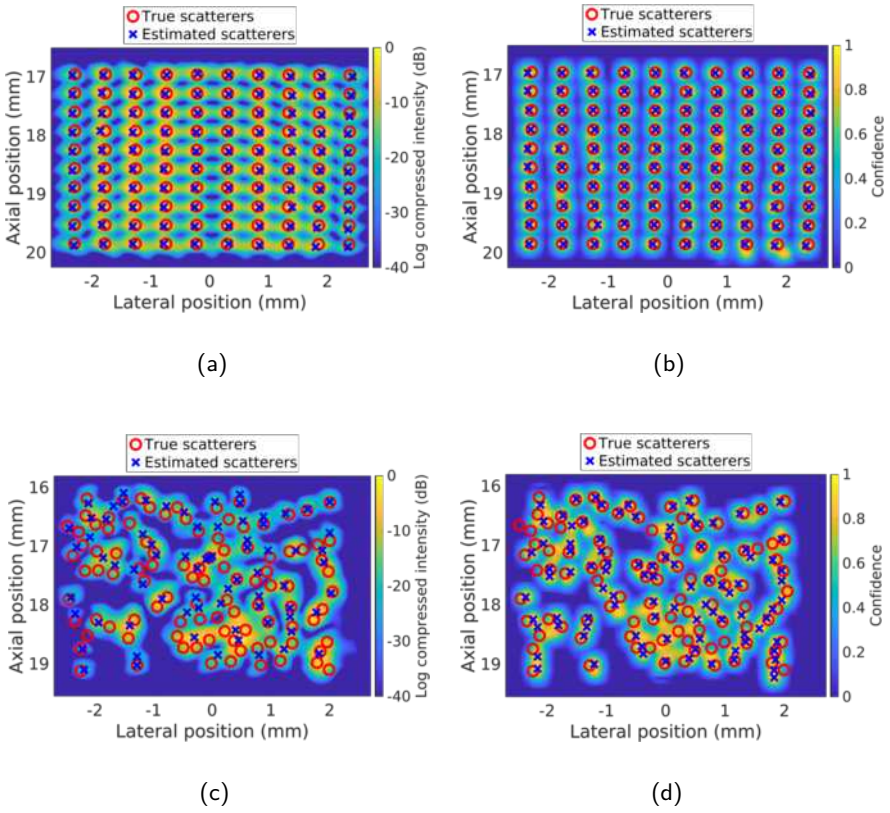


Figure 3.17: Confidence map estimation and scatterer localization. The first row shows the results on the *uniform* phantom by (a) peak detection and (b) the proposed method. The second row shows the results on the *random* phantom by (c) peak detection and (d) the proposed method. The figure is modified from Paper 1 (Youn, Ommen, et al. 2019) and Paper 2 (Youn, Ommen, et al. 2020).

Table 3.2: Precision, recall, and localization precision on the phantom measured data.

Phantom	Method	Precision	Recall	Localization Precision	
				Lateral	Axial
Uniform	Peak detection	1.00	1.00	39.09 µm	18.40 µm
	Proposed	0.98	1.00	36.30 µm	14.69 µm
Random	Peak detection	0.49	0.32	51.96 µm	28.75 µm
	Proposed	0.59	0.63	61.53 µm	16.89 µm

localization precision but better axial localization precision compared with centroid detection.

3.7 Discussion

Scatterer localization directly from RF channel data using a CNN is presented. The CNN estimated non-overlapping confidence maps, and scatterers were localized from the confidence maps. The simulation results showed that the proposed method could localize the scatterers spaced closer than the resolution of conventional ultrasound imaging. In the phantom results, for the trivial case, i.e., *uniform* phantom, centroid detection and the proposed method showed similar results by localizing all the scatterers. However, on the *random* phantom, where some of the scatterers were spaced closer than the resolution limit of ultrasound, the proposed method showed better performance than centroid detection apart from the lateral localization precision. This shows that the proposed localization method can be employed to localize more MBs at high concentrations of MBs, and to potentially shorten the data acquisition time of ULM.

Deep-ULM can localize overlapping PSFs on beamformed images (van Sloun et al. 2021) unlike the proposed method performing localization on the RF channel data. The proposed method was compared with deep-ULM after recalculating recall and localization error in Fig. 3.18, following the method van Sloun *et al.* used to make the results in Fig. 2 in (van Sloun et al. 2021). Both methods showed good performance at high densities, however, the proposed method achieved slightly better performance. Specifically, deep-ULM recovered roughly $2.10\,\mathrm{mm}^{-2}$ at the density of $3.53\,\mathrm{mm}^{-2}$ when the proposed method recovered $3.00\,\mathrm{mm}^{-2}$ at the density of $3.42\,\mathrm{mm}^{-2}$. The median localization error of deep-ULM was approximately $\lambda/12$ over all the scatterer densities, but the proposed method achieved smaller errors than $\lambda/12$. Yet, it is difficult to conclude that the proposed method outperforms deep-ULM since the evaluation was not performed on the same test data. This comparison, however, shows the potential of the methods localizing scatterers

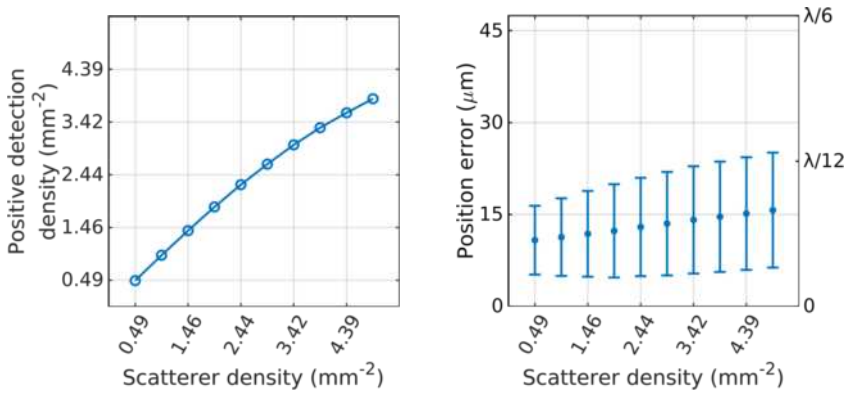


Figure 3.18: Recalculated recall and localization error of the proposed method to compare it with deep-ULM: (a) Positive detection density and (b) median position error with bars representing the standard deviation at different scatterer densities.

directly in the RF channel data.

The CNN was able to be generalized to the phantom measurements by considering the physical aspects of the cavities and phantom medium in the simulation, although training was performed using the simulated ultrasound data. Even so, there were still discrepancies between the simulated data and real-world data, and that resulted in the performance degradation in the phantom experiments compared with the simulation results. In *in vivo* scenarios, the discrepancies will be larger due to the variations of echoes by artifacts such as reverberation, attenuation, refraction, electrical noise, various physical properties of MBs, and tissue. Therefore, further performance degradation is expected when applying the method for MB localization. To handle those problems, a more sophisticated simulation to cover possible *in vivo* variations of channel data is necessary. At the same time, deep learning models that have better generalizability need to be investigated to increase the localization performance on the measured data and overcome the expected limits on the *in vivo* data.

CHAPTER 4

Localization on Beamformed Ultrasound Data

This chapter introduces a CNN-based MB localization method on beamformed ultrasound data. Especially, sub-pixel localization using non-overlapping confidence maps is presented. The method is evaluated on measured MB data from a 3-D printed phantom and animal experiments, as well as simulated data. This chapter is based on the Paper 3 (Youn, Taghavi, et al. 2021).

4.1 Introduction

4.1.1 Motivation

In Chapter 3, it has been discussed that the CNN-based localization method on ultrasound channel data suffers generalization problems due to the limited simulation accuracy, though that has high potential for localizing overlapping PSFs. Also, the channel data are not easily accessible, and streaming the channel data is challenging due to the high data rates. Such problems can be alleviated by localizing MBs on beamformed ultrasound data. There have been efforts to use deep learning techniques for super-resolution imaging at high concentrations of MB. Localization of overlapping MBs using CNNs was first proposed in (van Sloun et al. 2021), and improvement by adopting a different network architecture was reported in (X. Liu et al. 2020). The use of 3-D CNNs on a stack of beamformed ultrasound images has been suggested to utilize the temporal correlation of MB echoes for more effective localization (Brown, Ghosh, and Hoyt 2020) or to estimate track images directly without localization (Milecki et al. 2021). Nonetheless, all of them localize MBs in the pixel coordinates, so the localization accuracy is constrained by pixel size. To address this problem, additional upsampling layers are included in the networks (X. Liu et al. 2020; van Sloun et al. 2021), or the networks were applied after upsampling the beamformed ultrasound images. In this chapter, a sub-pixel localization method using a CNN and non-overlapping Gaussian confidence maps is discussed, which performs sub-pixel localization in the same image resolution of the input without additional upsampling.

4.1.2 Problem Definition

Let us consider a beamformed ultrasound image $\boldsymbol{x} \in \mathbb{R}^{N_z \times N_x}$ which are induced by MBs located at $\boldsymbol{p} \in \mathbb{R}^{N_{mb} \times 2}$, where N_z and N_x are the number of pixels in the axial and lateral

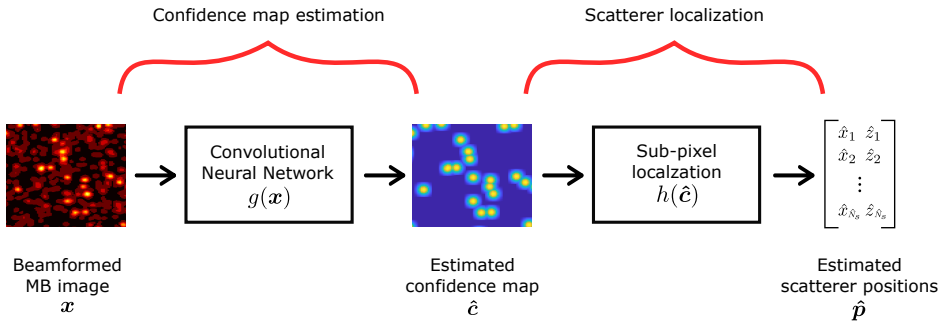


Figure 4.1: Overview of scatterer localization from beamformed MB images. A two-stage process was adopted similarly to Chapter 3, as shown in Fig 3.1, to achieve sub-pixel localization as well as to handle the varying number of MBs. The CNN formed a confidence map from the MB image and sub-pixel localization was followed. The illustration is modified from Paper 3 (Youn, Taghavi, et al. 2021).

directions, N_{mb} is the number of MBs, and 2 is the number of spatial dimensions, i.e., the lateral and axial directions. An beamformed ultrasound image can be approximated by

$$\boldsymbol{x} = \sum_{i=1}^{N_{mb}} \text{PSF}(\boldsymbol{p_i}) * \delta(\boldsymbol{p_i}) + \boldsymbol{n}, \tag{4.1}$$

where $\text{PSF}(\boldsymbol{p_i})$ is the PSF at the i-th MB position, δ is the Dirac delta function, and $\boldsymbol{n}$ is the noise. To locate the MBs from the ultrasound image, a mapping $f\colon \mathbb{R}^{N_z \times N_x} \to \mathbb{R}^{N_{mb} \times 2}$ that estimates $\boldsymbol{p}$ from $\boldsymbol{x}$,

$$\boldsymbol{p} = f(\boldsymbol{x}), \tag{4.2}$$

needs to be found.

The mapping f was modeled similarly to Chapter 3 as a two-stage process as follows:

$$\boldsymbol{p} = f(\boldsymbol{x}) = h(g(\boldsymbol{x})) = h(\boldsymbol{c}). \tag{4.3}$$

The two-stage process allowed sub-pixel localization as well as handling the varying number of MBs depending on the input image. The function $g\colon \mathbb{R}^{N_z \times N_x} \to \mathbb{R}^{N_z \times N_x}$ was a CNN that estimates confidence maps $\boldsymbol{c} \in \mathbb{R}^{N_z \times N_x}$ from the ultrasound images $\boldsymbol{x} \in \mathbb{R}^{N_z \times N_x}$ and sub-pixel localization on the confidence maps was performed by the function $h\colon \mathbb{R}^{N_z \times N_x} \to \mathbb{R}^{N_{mb} \times 2}$, which will be introduced in Section 4.5.2.

4.2 Imaging Sequence

Ultrasound data were acquired by the commercial ultrasound system bk5000 (BK Medical, Herlev, Denmark) using the 150-element linear array probe X18L5s (BK Medical, Herlev,

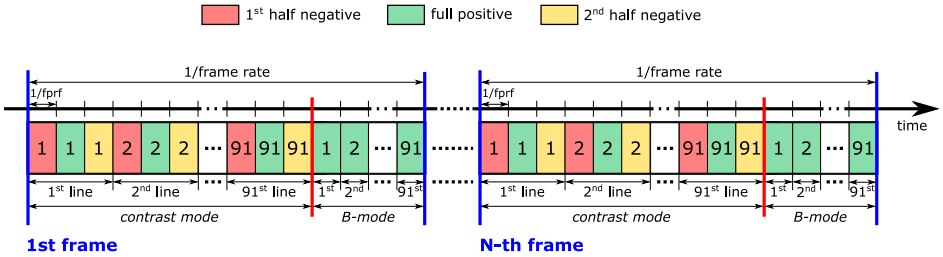


Figure 4.2: Overview of imaging sequence implemented in the bk5000 scanner (BK Medical, Herlev, Denmark). The sequence was composed of contrast mode and B-mode. Conventional focused beam transmissions using a sliding aperture with 91 sub-apertures was employed. In contrast mode, the CEUS imaging was achieved by the amplitude modulation scheme with three transmissions per sub-aperture: one full positive and two half negative transmissions. In B-mode, one full positive transmission was employed per sub-aperture. For both contrast mode and B-mode, a total of 364 transmission events were required for one cycle. The numbers inside transmission events correspond to the sub-aperture index. The illustration was modified from Paper 3

Denmark). In research ultrasound systems, e.g., SARUS (Jensen, Holten-Lund, et al. 2013) and vantage systems (Verasonics Inc., Redmond, WA, USA), imaging sequences can be customized and raw channel data are accessible. However, in the employed commercial scanner, the imaging parameters that can be modified were limited and only beamformed ultrasound data were accessible.

The imaging sequence of the scanner is illustrated in Fig. 4.2. It consisted of contrast mode and B-mode. For both modes, the conventional focused beam was transmitted using 91 sub-apertures with a sliding aperture of 25 elements. In contrast mode, CEUS was achieved by amplitude modulation (Mor-Avi et al. 2001) with three transmissions in each sub-aperture, i.e., one full positive and two half negative transmissions. The contrast-mode separated MB signals using the non-linear behavior of MBs and resulted in MB images for localization. After that, the B-mode sequence followed with one full positive transmission. The B-mode images were used for motion compensation in animal study. For one image frame cycle, $91 \times 3 + 91 = 364$ transmissions were involved. The frame rate was 53.85 Hz with the pulse repetition frequency f_{prf} of 19.6 kHz.

4.3 Microbubble Data Generation

RF channel data were simulated in Field II pro (Jensen 1996, 2014; Jensen and Svendsen 1992) to generate MB data for training, validation, and evaluation. In the simulation, MBs were modeled as point scatterers since the diameters of MBs are much smaller than the diffraction limit of ultrasound. Specifically, SonoVue (Bracco Imaging, Milan, Italy)

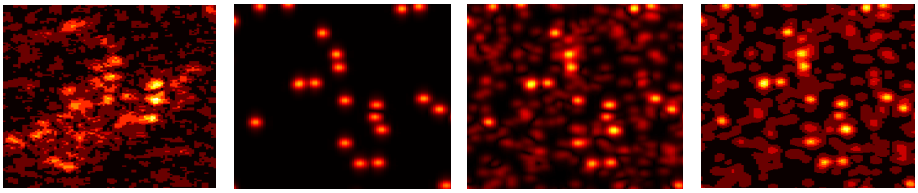
Table 4.1: Field II simulation parameters.

	Parameter	Value
Transducer	Number of elements	150
	Pitch	0.16 mm
	Element height	3.40 mm
	Element width	0.15 mm
	Elevation focus	20 mm
Imaging	Transmission pulse frequency	6 MHz
	Number of transmission pulse cycle	2
	Number of active elements in transmission	25
	Wave type	Focused beam
	Focal depth in transmission	0.01 mm
	Apodization in transmission	Boxcar window
Beamforming	Method	Delay-and-sum
	F-number	1
	Apodization in reception	Gaussian window
	Pixel size	48 µm axially 79 µm laterally
	Region of interest	(0.02, 20.01) mm axially (−10.72, 10.64) mm laterally
Environment	Speed of sound	1540 m/s
	Field II sampling freq.	350 MHz

has the mean diameter of 2.5 µm (Schneider 1999), which was the employed contrast agent. Also, the non-linear behavior of MBs was not considered to simplify the simulation model. The simulation parameter values in Table 4.1 were chosen following the scanner configuration and the probe specification that were used for ULM experiments.

In measured data, weak scattering originated from not rejected stationary echoes, out-of-plane MBs, and the low SNR appeared in the background, as shown in Fig. 4.3(a). CNNs can handle such noise as long as it is reflected in the training data. Hence, the noise was included in the simulation by adding another point scatterers that have 4 times smaller scattering amplitudes than MBs.

To generate one MB image frame, the RF channel data were simulated by placing point scatterers randomly in the ROI. The simulated RF channel data were then beamformed by DAS (Thurstone and Ramm 1974) (Fig 4.3(b)), the weak scattering noise was added (Fig 4.3(c)), and envelope detection was performed using the Hilbert transform. Lastly, the image was quantized to be matched with the measured data because the measured MB data had few intensity levels. The quantization process was applied so that isolated MBs



(a) (b) (c) (d)

Figure 4.3: An example of measured and simulated CEUS MB images. (a) is the measured image, (b) is the simulated image only with MBs, (c) is the simulated MB image with the noise, and (d) is the final simulated MB image after the quantization.

have five-level intensities, as shown in Fig 4.3(d), which was defined empirically from the measured data.

4.4 Convolutional Neural Network

4.4.1 Network Architecture

Fig. 4.4 shows the proposed CNN architecture that was constructed based on U-Net (Ronneberger, Fischer, and Brox 2015) and pre-activation residual units (He et al. 2016). It was composed of three *down*-blocks, one *conv*-block, and three *up*-blocks (Youn, Ommen, et al. 2019, 2020), which are presented in Section 3.4. Localization on RF channel data in Chapter 3 required a large receptive field as the network implicitly performed beamforming along with localization. On the other hand, for localization on beamformed ultrasound data, having three pooling and unpooling layers were enough to achieve good confidence map estimation. The input data were already beamformed; therefore, MB positions could be determined by locally extracted features.

Both the proposed CNN and deep-ULM (van Sloun et al. 2021) adopted the encoder-decoder structure. Nevertheless, the proposed method can achieve sub-pixel localization, so the localization process is performed in the same image resolution to the input MB image. However, for deep-ULM, localization is available only in the pixel coordinates without sub-pixel accuracy. Therefore, it has additional upsampling layers to maximize localization accuracy, which increases computational complexity.

4.4.2 Training Detail

The proposed CNN was trained to obtain the mapping that returns confidence maps $\boldsymbol{c}$ given ultrasound images $\boldsymbol{x}$. Training was performed by optimizing the difference between true and estimated confidence maps captured by the MSE:

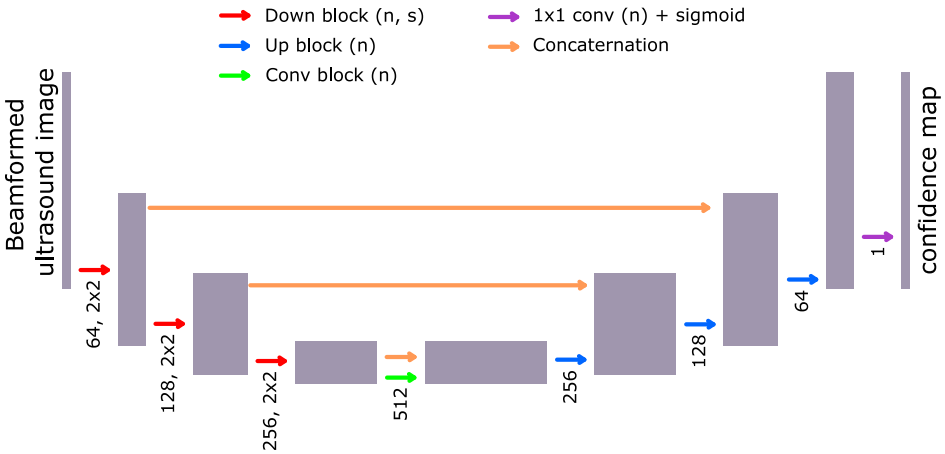


Figure 4.4: Proposed network architecture based on U-Net (Ronneberger, Fischer, and Brox 2015) and pre-activation residual units (He et al. 2016). The number of kernels (n) and stride (s) are indicated next to the arrows. The details about *down*-, *conv*-, and *up*-blocks (Youn, Ommen, et al. 2019, 2020) can be found in Section 3.4. The illustration is modified from Paper 3 (Youn, Taghavi, et al. 2021).

$$\mathcal{L}_{\text{MSE}}\left(\boldsymbol{x}, \boldsymbol{c}; g\right) = \frac{1}{N}\sum_{i=1}^{N}\left\|\boldsymbol{c}_i - g\left(\boldsymbol{x}_i; \theta\right)\right\|_F^2, \tag{4.4}$$

where $\boldsymbol{x}_i$ and $\boldsymbol{c}_i$ are the i-th MB image and corresponding confidence map, g is the proposed CNN with learning parameters θ, N is the number of samples, and $\|\cdot\|_F$ is the Frobenius norm.

The learning parameters were optimized using Rectified Adam (RAdam) (L. Liu et al. 2020) and Lookahead (Zhang et al. 2019) after being initialized by orthogonal initialization (Saxe, McClelland, and Ganguli 2013). RAdam provides training stability at the beginning of training (L. Liu et al. 2020) and Lookahead provides training stability during the rest (Zhang et al. 2019). So, the combination of them is known to stabilize the whole training and converge to the solution with fewer iterations than Adam (Kingma and Ba 2015). The learning rate was set to 0.0001 and halved every 200 epochs. The proposed CNN was implemented using Tensorflow (Abadi et al. 2011) in Python and a server equipped with a NVIDIA TESLA V100 16 GB PCIe graphics card was employed for training, which took approximately 24 hours for 1000 epochs.

The MB images and weak scattering images, i.e., noise, were simulated separately. An MB image was generated by placing 400 scatterers and a weak scattering image was generated by placing 4000 scatterers in the ROI. The scattering amplitudes among MBs

were assumed to be the same, and those of weak scatterers were 4 times smaller than the MB scattering.

A training image frame was formed by selecting one MB image and one weak scattering image randomly during training and summing them up to provide more diverse training data. That allows training the network on the noise independent to the MBs images. And then, the training frame was randomly cropped to be a size of 128×128. Although the the same number of MBs were used for the MB image simulation, the MB overlaps and MB densities in the cropped region were different since the ROI was large. For data augmentation, the training frame was flipped in the lateral direction at random. Finally, the input MB image and corresponding confidence maps were normalized to be in the range of $[0, 1]$.

To monitor training and select hyper-parameters, validation was conducted. For validation data, 128 MB images and 128 weak scattering images were generated in the same way as the training data.

4.5 Confidence Map and Sub-pixel Localization

4.5.1 Non-overlapping Gaussian Confidence Map

In confidence maps, the pixel values indicate the confidences of MB presence in each pixel position. The higher the confidence is, the higher the change that an MB exists in the pixel. By training CNNs to learn the confidence map and localizing MBs in the confidence map, varying numbers of MBs depending on the input ultrasound image can be dealt with. Especially, non-overlapping Gaussian confidence map has been proposed to provide large gradients for stable training without losing positions of closely spaced targets (Youn, Ommen, et al. 2019, 2020).

Previously, target positions were quantized based on pixel size and Gaussians were defined in the discrete image grid. To achieve sub-pixel localization, the non-overlapping Gaussian confidence maps were extended by defining the Gaussians in the continuous domain. The confidence maps were then constructed by sampling the maximum of the Gaussians in the image coordinates. A detailed process of implementing the non-overlapping Gaussian confidence maps is described in Algorithm 4.1.

4.5.2 Sub-pixel Localization

Local peaks and their surrounding pixels in the extended non-overlapping Gaussian confidence maps follow the Gaussian functions thanks to the maximum operation in the confidence map generation. Also, the Gaussians are defined in the continuous spatial domain. Therefore, sub-pixel localization can be achieved by applying Gaussian fitting to the local peaks and their surrounding pixels and taking the centers of the fitted Gaussians. The sub-pixel localization scheme in a confidence map is presented in Algorithm 4.2.

Algorithm 4.1: Non-overlapping Gaussian confidence map implementation.

Input: MB positions $\boldsymbol{p} \in \mathbb{R}^{N_s \times 2}$, confidence map pixel coordinates $\boldsymbol{p}^{img} \in \mathbb{R}^{N_z \times N_x}$, and a covariance matrix $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_z^2 & 0 \\ 0 & \sigma_x^2 \end{pmatrix}$, where σ_z and σ_x are the standard deviations along the z and x directions.

Output: A non-overlapping Gaussian confidence map $\boldsymbol{c} \in \mathbb{R}^{N_z \times N_x}$

Let us consider a 2-D Gaussian function

$$\mathcal{G}\left(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}\right) = \exp\left\{-\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)^{\intercal} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\right\},$$

where $\boldsymbol{\mu} = (\mu_z, \mu_x)$.

for $k = 1$ to N_s **do**

  $\boldsymbol{c}^k \leftarrow \left\{ \left(c_{ij}^k\right) \in \mathbb{R}^{N_z \times N_x} \,\middle|\, c_{ij}^k = \mathcal{G}(\boldsymbol{p}_{ij}^{img}; \boldsymbol{p}_{k*}, \boldsymbol{\Sigma}) \right\}$ // k-th Gaussian

end for

$\boldsymbol{c} \leftarrow \left\{ (c_{ij}) \in \mathbb{R}^{N_z \times N_x} \,\middle|\, c_{ij} = \max_{k \in [1, N_s]} c_{ij}^k \right\}$. // Maximum of Gaussians

Algorithm 4.2: Sub-pixel localization from a confidence map.

Input: A non-overlapping Gaussian confidence map $\boldsymbol{c} \in \mathbb{R}^{N_z \times N_x}$ and confidence map pixel coordinates $\boldsymbol{p}^{img} \in \mathbb{R}^{N_z \times N_x}$.

Output: Estimated MB positions $\hat{\boldsymbol{p}}^{mb} \in \mathbb{R}^{\hat{N}_{mb} \times 2}$.

$\hat{\boldsymbol{p}}^{mb} \leftarrow \{\ \}$

for $i = 2$ to $N_z - 1$ **do** // Local peak search

  for $j = 2$ to $N_x - 1$ **do**

    if $c_{i,j} = \max\{c_{i-1,j}, c_{i,j-1}, c_{i,j}, c_{i+1,j}, c_{i,j+1}\}$ **then**

      $\hat{p} \leftarrow fitGaussian\left(i, j, \boldsymbol{c}, \boldsymbol{p}^{img}\right)$ // Gaussian fitting

      $\hat{\boldsymbol{p}}^{mb}$.insert($\hat{p}$)

    end if

  end for

end for

The Gaussian fitting can be solved analytically. Let us consider N data samples $\{(y_i; x_{i1}, x_{i2})\}_{i=1}^{N}$ that follow a 2-D Gaussian function

$$y = \exp\left\{-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)\right\}, \tag{4.5}$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2)$ is the mean, i.e., center, and $\boldsymbol{\sigma} = (\sigma_1, \sigma_2)$ is the standard deviation of the Gaussian function. Then, the following equation can be acquired by taking natural logarithms in equation (4.5),

$$\ln y = ax_1^2 + bx_2^2 + cx_1 + dx_2 + e, \tag{4.6}$$

where $a = -1/2\sigma_1^2$, $b = -1/2\sigma_2^2$ $c = \mu_/\sigma_1^2$, $d = \mu_2/\sigma_2^2$, and $e = -\left(x_1^2/2\sigma_1^2 + x_2^2/2\sigma_2^2\right)$. A linear regression can be formalized by the data samples and equation (4.6):

$$\begin{pmatrix} x_{11}^2 & x_{12}^2 & x_{11} & x_{12} & 1 \\ x_{21}^2 & x_{22}^2 & x_{21} & x_{22} & 1 \\ & & \vdots & & \\ x_{N1}^2 & x_{N2}^2 & x_{N1} & x_{N2} & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \\ e \end{pmatrix} = \begin{pmatrix} \ln y_1 \\ \ln y_2 \\ \vdots \\ \ln y_N \end{pmatrix}. \tag{4.7}$$

The analytic solution of equation 4.7 can be found by

$$\begin{pmatrix} a \\ b \\ c \\ d \\ e \end{pmatrix} = \begin{pmatrix} x_{11}^2 & x_{12}^2 & x_{11} & x_{12} & 1 \\ x_{21}^2 & x_{22}^2 & x_{21} & x_{22} & 1 \\ & & \vdots & & \\ x_{N1}^2 & x_{N2}^2 & x_{N1} & x_{N2} & 1 \end{pmatrix}^{-1} \begin{pmatrix} \ln y_1 \\ \ln y_2 \\ \vdots \\ \ln y_N, \end{pmatrix}, \tag{4.8}$$

and the center of the Gaussian can be estimated by

$$\mu_1 = -c/2a \quad \text{and} \quad \mu_2 = -d/2b. \tag{4.9}$$

As seen from equation 4.8, at least five data points are necessary to estimate the peak of a 2-D Gaussian function. So, the Gaussian function was fitted to the local peak and its 4 neighboring pixels in the estimated confidence map.

4.6 Simulation Experiment

In this section, (1) sub-pixel localization given true confidence maps and (2) the performance of the trained CNN were assessed on the simulated test data. The test data were generated at 9 different MB densities from $0.3\,\mathrm{mm}^{-2}$ to $4.9\,\mathrm{mm}^{-2}$ in a 128×128 region. At an MB density, 128 frames were generated, which have similar degrees of MB overlaps. The 128×128 region was selected randomly for each frame to take the

spatially varying PSFs into account. Unlike the training data, the MB images and weak scattering images were simulated simultaneously for the test data. The number of MBs was changed depending on the target MB density, but the number of weak scatterers was fixed, assuming the noise is independent to the number of MBs.

4.6.1 Evaluation Metrics

For the sub-pixel localization evaluation, the estimated MBs were obtained by performing localization in the true non-overlapping Gaussian confidence maps since the purpose was to evaluate the sub-pixel localization process against localization in the pixel coordinates. Localization precision was calculated by

$$\text{localization precision} = 2\sqrt{2\ln 2}\,\sigma, \tag{4.10}$$

where σ is the standard deviation of localization errors between true and estimated MBs. Equation 4.10 is the FWHM of a Gaussian function. Therefore, the localization error will be the resolution of reconstructed ULM images if the localization errors follow a Gaussian distribution.

For the trained CNN assessment, precision, recall or reconstructed MB density, and localization precision (equation 4.10) in the lateral and axial directions were measured. Precision and recall are

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{4.11}$$

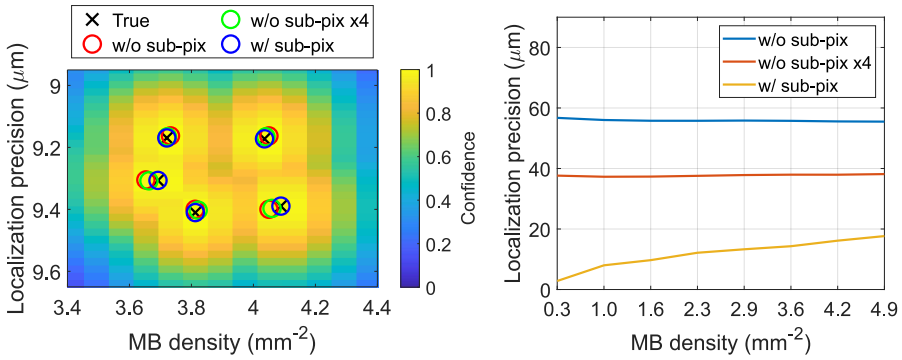
$$\text{Recall} = \frac{TP}{TP + FN}, \tag{4.12}$$

and the reconstructed MB density $\hat{d}_{mb}$ is

$$\hat{d}_{mb} = \text{recall} \times d_{mb}, \tag{4.13}$$

where TP is the number of true positives (correct MB localization), FP is the number of false positives (wrong MB localization), FN is the number of false negatives (missed MBs), and d_{mb} is the true MB density.

Unlike the sub-pixel localization evaluation, it is necessary to match the true MBs with the estimated MBs for the CNN evaluation to determine the wrong localization and missed MBs. As stated in Section 3.5.1, simply matching an estimated MB with the nearest true MB has a problem that a true MB can be matched with several estimated MBs (Youn, Ommen, et al. 2020). The bi-directional matching was suggested in Section 3.5.1, but it is computationally expensive. Thus, the matching problem was formulated as a linear assignment problem, and the MATLAB (MathWorks, MA, USA) built-in function *matchpairs* (Duff and Koster 2001) was used to solve the optimization problem. By doing so, the same solution can be obtained in a much shorter computation time. The cost matrix of the linear assignment problem was defined by pairwise distances between the true and estimated MBs. In addition, the cost of not matching, i.e., the cost of not assigning an



(a) (b)

Figure 4.5: Comparison of localization in the pixel coordinates and sub-pixel localization. Sub-pixel localization (*w/ sub-pix*) was achieved in the true confidence maps using the Gaussian fitting presented in Section 4.5.2. Localization without sub-pixel accuracy was performed by quantizing the true MB positions to the input image grid (*w/ sub-pix*) and the 4 times higher resolution image grid than the input (*w/ sub-pix ×4*). (a) is an example of a true confidence map with the true and estimated MB positions. (b) is the localization precision of the different methods at various MB densities.

estimated MB to a true MB, was set to reject assignments with large localization errors. The estimated MBs with large localization errors are essentially wrong localization. The cost of not matching was $\lambda/5$ (49 µm) for precision and recall, and $\lambda/2$ (123 µm) for localization precision.

4.6.2 Result

Fig.4.5 shows a comparison of localization in the pixel coordinates and sub-pixel localization using true MBs or true confidence maps. Sub-pixel localization (*w/ sub-pix*) was achieved using the aforementioned Gaussian fitting in Section 4.5.2. Localization without sub-pixel accuracy was performed by quantizing the true MB positions to the input image grid (*w/ sub-pix*) and to the 4 times higher resolution image grid than the input image (*w/ sub-pix ×4*). The 4 times higher resolution image grid was selected because Liu *et al.* has reported that the additional sampling in the network with a factor of 4 shows good balance between training stability and localization accuracy empirically (X. Liu et al. 2020).

Understandably, *w/o sub-pix ×4* showed better localization precision than *w/o sub-pix* as its quantization errors are lower thanks to the smaller pixel size. Nonetheless, *w/o sub-pix ×4* was not as good as *w/ sub-pix*, though *w/ sub-pix* had the larger pixel size, as

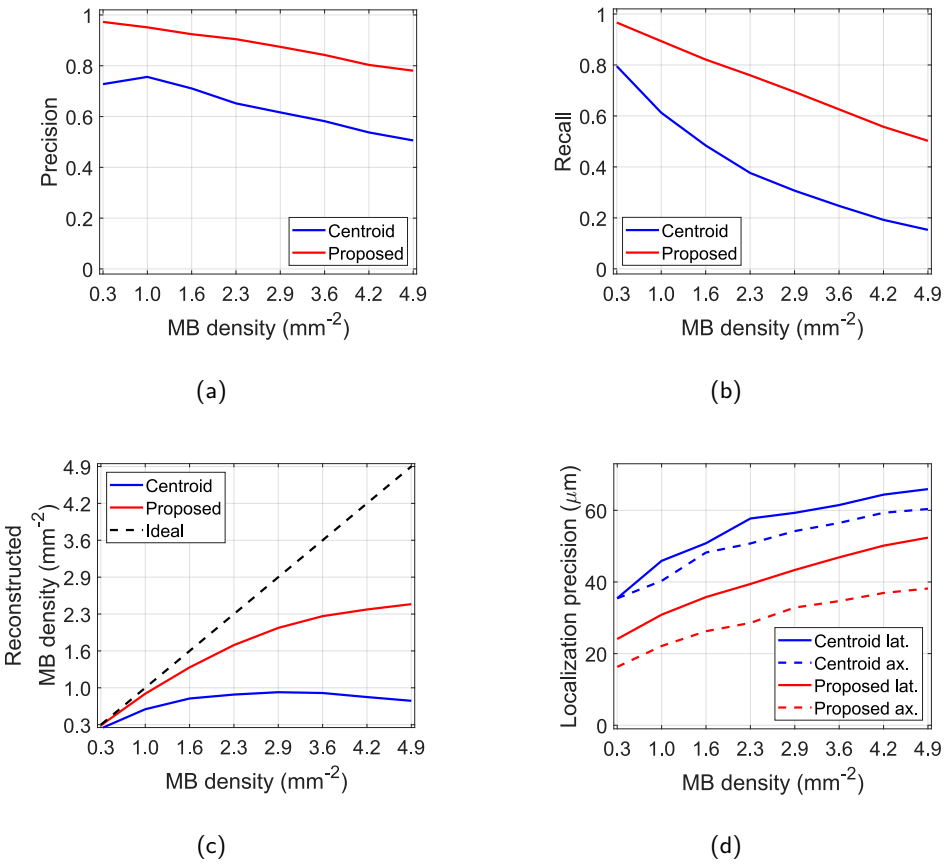


Figure 4.6: Comparison of localization capability between centroid detection and the proposed method on test data simulated at various MB densities. (a) is precision, (b) is recall, (c) is the reconstructed MB density, and (d) is localization precision in the lateral and axial directions. The figure is modified from Paper 3 (Youn, Taghavi, et al. 2021)

shown in Fig. 4.5(b). Sub-pixel localization achieved more than 2 times better localization precision than localizing MBs in the pixel coordinates in the given MB densities when the true MB positions and true confidence maps are available.

Notably, the localization precision of *w/ sub-pix* degraded as the MB density increased, indicating that *w/ sub-pix* was affected by the MB density. The reason is that the local peaks and their neighboring pixels are getting away from the Gaussian function as more overlapping MBs appear. On the other hand, the quantization error is not affected by

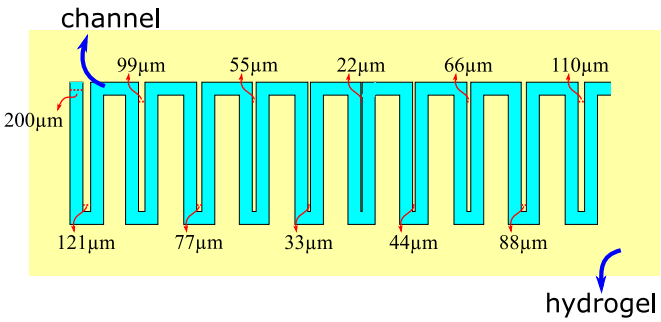


Figure 4.7: Illustration of 3-D printed channel phantom. The channel was bent a 90° angle several times and fashioned ten local pairs of closely spaced channels with various spacing. The illustration was from (Youn, Taghavi, et al. 2021)

the MB density, so the localization precision of *w/o sub-pix* and *w/o sub-pix* $\times 4$ were consistent across different MB densities.

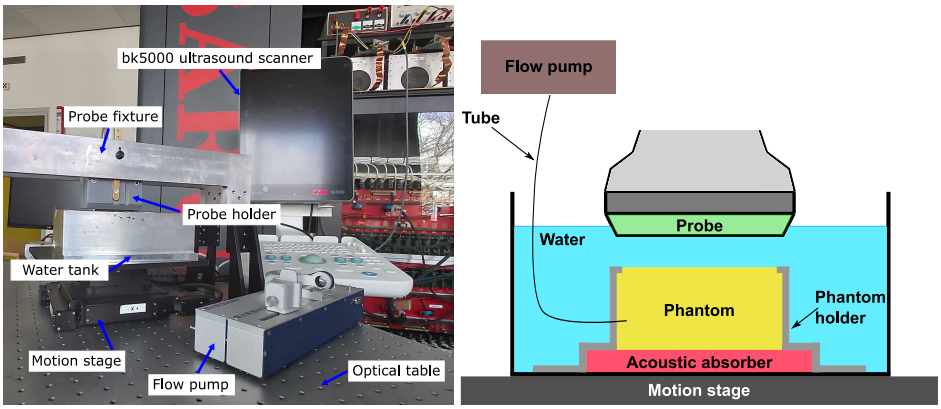
Fig. 4.6 shows the comparison of the proposed CNN localization method against centroid detection on DAS beamformed images. As centroid detection cannot handle overlapping PSF localization, the proposed method achieved better precision, recall, and localization precision at high MB densities. The proposed method also outperformed centroid detection at low MB densities where the overlapping PSFs are less likely to appear, showing that the isolated MBs can also be localized more accurately.

Without the ability of overlapping PSF localization, the number of MBs can be localized was limited even though the MB density increased. In Fig. 4.6(b), recall of centroid detection decreased as the MB increased, and this resulted in the saturation of reconstructed MB density, as shown in Fig. 4.6(c). The reconstructed MB density of centroid detection reached a peak of around 1.0 mm^{-2} at the MB density of 2.9 mm^{-2} and started to decrease. The reconstructed MB density of the proposed method, however, kept increasing.

4.7 Phantom Experiment

4.7.1 Phantom Fabrication

A PEGDA 700 g/mol hydrogel phantom (Ommen et al. 2019, 2021) that embeds a channel inside was fabricated. Contrary to the phantoms in Section 3.6.1 that have cavities acting as scatterers, this phantom has the empty channel in a plane, so MBs can be infused into it and the structure of the channel can be imaged by ULM. An illustration of the phantom is shown in Fig. 4.7. The channel whose diameter is 200 µm was bent at a 90° angle several times and fashioned ten pairs of closely spaced channels. To evaluate the limit of different localization methods, the spacing of each pair was varied from 22 µm to 121 µm. The



(a) (b)

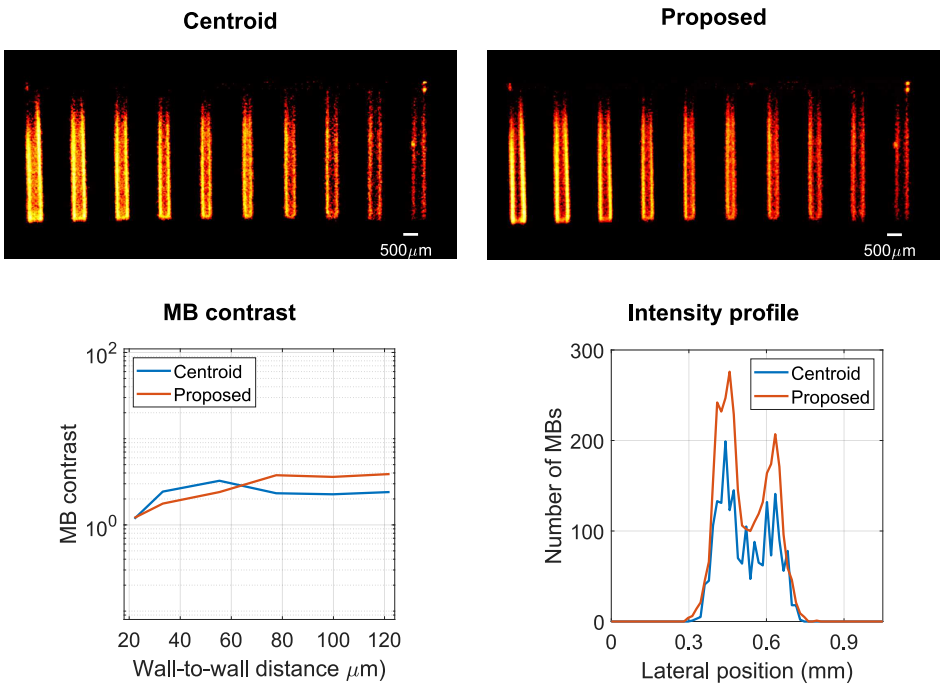
Figure 4.8: Phantom experiment setup. (a) is the picture of the experiment setup and (b) is the illustration of the experiment setup. The figure is modified from Paper 3 (Youn, Taghavi, et al. 2021).

spacing decreased and increased again from the left to right to maintain the stability of the phantom during 3-D printing and minimize unexpected 3-D printing errors.

4.7.2 Experiment Setup

The 3-D printed phantom was scanned with the X18L5s linear array probe (BK Medical, Herlev, Denmark) whose specification followed the parameter values in Table 4.1. The MB images were acquired in the commercial ultrasound system bk5000 (BK Medical, Herlev, Denmark) by the imaging sequence introduced in Section 4.2.

The scanning was performed using the experiment setup shown in Fig.4.8. The setup was installed on the optical table that absorbs and dissipates vibration. The probe was fixated to a probe fixture using a probe holder, and the phantom was submerged and fixed in a water tank. The water tank is then laid on the motion stage, and the phantom was aligned so that the channel can be positioned in the imaging plane. For ultrasound contrast agents, SonoVue (Bracco Imaging, Milan, Italy) was diluted and injected to the channel of the phantom by a syringe which was controlled by a flow pump to keep the MB concentration uniform by applying a constant flow. The phantom was measured at 2 MB concentrations. One was 1:40 dilution (*low* concentration) and the other was 1:20 dilution (*high* concentration). The volume flow rate was fixed to $1\,\mu\mathrm{L}/\mathrm{min}$.



(a)

Figure 4.9: Phantom measurement results. (a) and (b) include the ULM reconstruction by centroid detection (top left) and the proposed method (top right), MB contrast at each pair (bottom left), and the lateral intensity profile at the closest pair (bottom right). (a) is the result at *low* MB concentration with 3000 frames and (b) is the result at the *high* MB concentration with 800 frames. The figure is modified from Paper 3 (Youn, Taghavi, et al. 2021)

4.7.3 Evaluation Metric

In phantom measurements, true MB positions are unknown, therefore, the evaluation metrics suggested for the simulation experiments are not applicable. Instead, the structure of the channel is known, so a new metric, MB contrast ratio, was calculated by checking whether the estimated MBs are inside or outside of the channel. The MB contrast ratio C_{mb} is given by

$$\frac{N_{mb,ch}\,(A_{tot} - A_{ch})}{(N_{mb,tot} - N_{mb,ch})\,A_{ch}}, \tag{4.14}$$

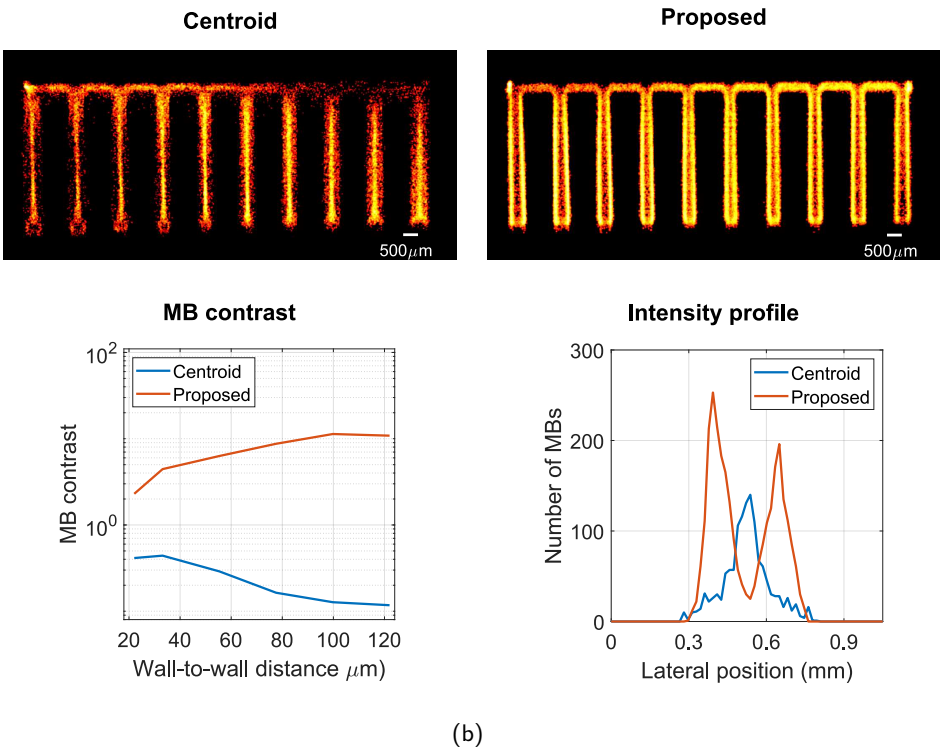


Figure 4.9: Phantom measurement results. (a) and (b) include the ULM reconstruction by centroid detection (top left) and the proposed method (top right), MB contrast at each pair (bottom left), and the lateral intensity profile at the closest pair (bottom right). (a) is the result at *low* MB concentration with 3000 frames and (b) is the result at the *high* MB concentration with 800 frames. The figure is modified from Paper 3 (Youn, Taghavi, et al. 2021)

where $N_{mb,tot}$ and A_{tot} are the total number of MBs in an region and the area of the region, and $N_{mb,tot}$ and A_{ch} are the number of MBs inside channels and the area of channels in the region. Having an MB contrast ratio higher than 1 is a necessary condition for a pair of channels being resolved.

4.7.4 Result

The phantom experiment results at the *low* MB concentration with 3000 frames and at the *high* MB concentration with 800 frames are shown in Fig. 4.9. At the *low* concentration,

for both methods, all the pairs were well separated, and the MB contrast ratio satisfied the necessary condition to resolve the channels, i.e., higher than 1. The lateral intensity profile also shows that both methods successfully resolved the most closely spaced pair whose wall-to-wall distance is 22 µm. On the other hand, at the *high* MB concentration, the proposed method still resolved all the pairs clearly, but centroid detection failed as it cannot localize the overlapping PSFs accurately, as shown in Fig. 4.10(b). The MB contrast ratio of centroid detection was lower than 1 at all the pairs, and the lateral profile also showed one peak in the middle of the channels. On the other hand, the proposed method achieved the MB contrast ratio higher than 1 and showed clear separation of the closest pair.

The *high* concentration result of the proposed method was comparable to the *low* concentration result with 3.8 times fewer frames. This shows that the potential of the proposed method for shortening the data acquisition time of ULM by employing high concentrations of MBs. The average number of the estimated MBs per frame for the proposed method was 32 at the *low* and 124 at the *high* concentration. When the concentration was doubled from the *low* to *high* concentration, the number of localized MBs increased by a factor of 3.9. The average number of the estimated MBs per frame for centroid detection was also increased from 19 to 40, but the estimated MBs were mostly found outside the channel, i.e., wrong estimations.

4.8 Animal Experiment

The kidney of a healthy male Sprague-Dawley rat was scanned following the protocols approved by the Danish National Animal Experiments Inspectorate. The procedures were conducted at University of Copenhagen and the details can be found in (Youn, Taghavi, et al. 2021). The scan was performed for 9 minutes at 4 MB concentrations using the same model of the ultrasound system and probe as the phantom experiments, as shown in Table 4.2. The MB concentration was assumed to be linearly proportional to the MB dilution and volume flow rate.

After scanning, localization was performed in MB images and motion correction was applied to the estimated MB positions. The 2-D cross-correlation in small patches between the reference and a current frame was calculated to estimate the rigid motion in the B-mode images (Taghavi, Andersen, Hoyos, Nielsen, et al. 2021). Lastly, tracking was applied on the motion corrected MBs using the hierarchical Kalman filtering suggested in (Taghavi, Andersen, Hoyos, Schou, et al. 2020). The hierarchical Kalman filtering has multiple motion and noise models unlike normal Kalman filter-based MB tracking (Tang et al. 2020). Therefore, different models can be adapted depending on the velocity of the MBs for more effective tracking.

For *in vivo* measurements, tracking is essential since the chance of wrong estimations appearing is higher due to the complexity of target structures, poor SNR, and ultrasound artifacts. The wrong estimations can be removed effectively by taking the temporal

Table 4.2: MB concentrations for animal experiments.

Experiment	MB dilution	Flow rate	MB concentration (a.u.)
Scenario 1	1:20	85 µL/min	4
Scenario 2	1:20	170 µL/min	9
Scenario 3	1:10	170 µL/min	17
Scenario 4	1:5	170 µL/min	34

correlation of the estimated MBs into account. The tracking stage also provides velocity information which allows separating mingled microvessels with different flow directions (Couture et al. 2018). Hence, a proper tracking process provides better ULM image quality. For the animal experiments, ground truth is unknown, so it is assumed that track samples, i.e., the samples consist of the tracks, are correct and the results were assessed based on the number of track samples and their distances in a frame.

4.8.1 Result

The reconstructed ULM images from the rat kidney measurements are shown in Fig. 4.10. In *scenario 1*, both methods resulted in similar high-resolution microvascular images. In *scenario 2*, more microvessels were highlighted and the proposed method achieved a little brighter image than centroid detection by localizing more MBs. Centroid detection started to fail in *scenario 3* by losing many microvessels and failed in *scenario 4* apart from the inner medulla region, where the local MB concentration was relatively low compared to the other regions. On the contrary, the proposed method achieved good image quality in *scenario 3* and the overall shape of the kidney was perceptible with clear large vessels in *scenario 4*, although it also mostly failed to reconstruct microvessels except in some inner medulla regions.

To analyze the effect of the MB concentrations locally, three regions were selected from the inner medulla, outer medulla, and cortex. The selected regions are highlighted as blue rectangles in Fig 4.10. Fig. 4.11 shows the ULM reconstruction and the number of track samples by centroid detection and the proposed method in the selected regions at the different MB concentrations. In the inner medulla, a similar trend was observed for both methods. The number of track samples increased up to *scenario 3* and decreased in *scenario 4*. In the outer medulla and cortex, more MBs were localized by the proposed method at all the MB concentrations. Furthermore, the number of track samples for the proposed method peaked at the higher MB concentrations than centroid detection.

The capability of separating closely spaced MBs was investigated implicitly by measuring the smallest pairwise distances among track samples in a frame over the 9 minute measurements. Fig. 4.12 shows the normalized counts of the smallest pairwise distances on the *scenario 2* rat data as a histogram with bins of 50 µm. The normalized counts were acquired by dividing the counts by the total number of counts. For centroid

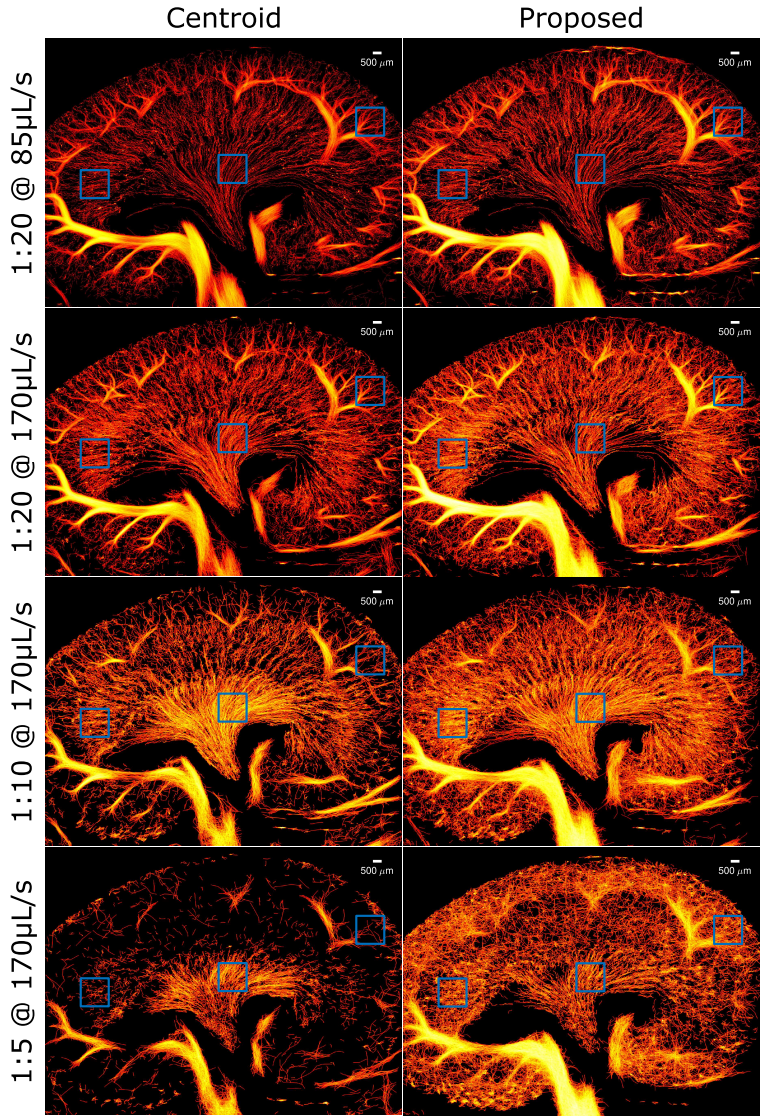


Figure 4.10: ULM reconstruction from the rat measurements by centroid detection and the proposed method at 4 MB concentrations. The figure is modified from Paper 3 (Youn, Taghavi, et al. 2021)

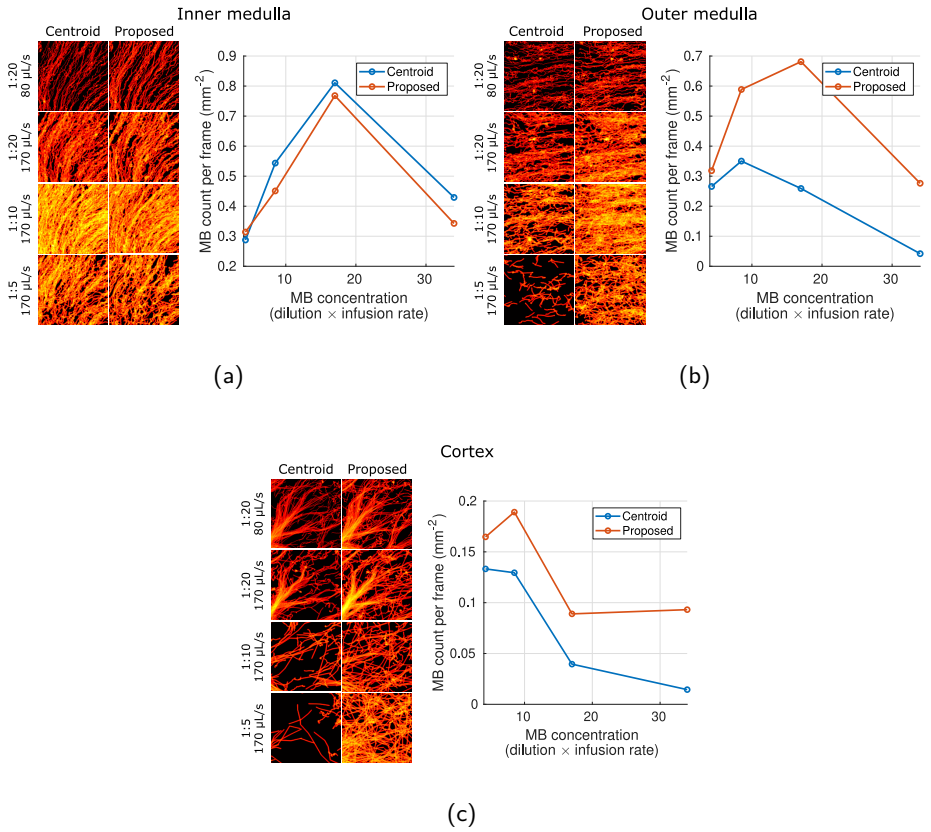


Figure 4.11: Rat experiment results in the three selected regions for local analysis. The selected regions are highlighted as blue rectangles in Fig.4.10. The ULM reconstruction and the number of track samples by centroid detection and the proposed method at different MB concentrations are shown in (b) for the inner medulla, (c) for the outer medulla, and (d) for the cortex. The figure is modified from Paper 3 (Youn, Taghavi, et al. 2021).

detection, there were not track samples closer than 250 µm ($\approx \lambda$), represented as a red vertical dashed line. However, for the proposed method, 8% of the estimated track samples were closer than 250 µm ($\approx \lambda$), showing that the proposed method can localize MBs closer than the resolution limit of ultrasound.

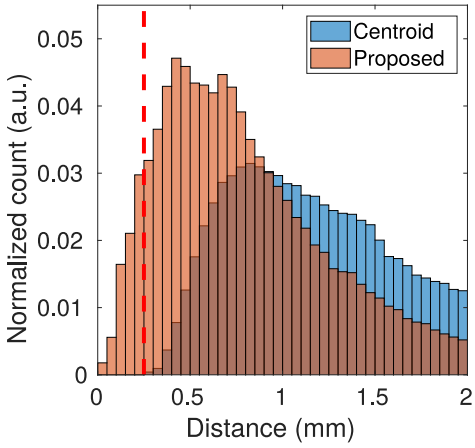


Figure 4.12: Histogram of the smallest pairwise distances among track samples in a frame over the 9 minute measurements on the *scenario 2* rat data. The normalized counts were acquired by dividing the counts by the total number of counts. The red vertical dashed line represents 250 µm ($\approx \lambda$). The figure is from Paper 3 (Youn, Taghavi, et al. 2021).

4.9 Computation Complexity

The proposed method utilizes computational resources efficiently by virtue of the sub-pixel accuracy processing as additional upsampling layers are not necessary and localization is performed on the same image resolution as the input image. The computational complexity of deep-ULM (van Sloun et al. 2021), mSPCN-ULM (X. Liu et al. 2020), and the proposed method given ultrasound images with a size of 786 × 272, the image size of the phantom and animal measurements, were investigated. The number of model parameters and the number of floating point operations (FLOPs) were calculated manually. The process time was measured for one image frame by repeating inference for 1000 times. Lastly, the maximum available batch size, i.e., the number of image frames that can be processed in a single iteration, was measured by increasing the batch size in powers of 2 until running out of GPU memory. For the computational complexity evaluation, the PC equipped with a NVIDIA Titan V graphics card was used, and the results are shown in Table 4.3. Note that the number of FLOPs, process time, and maximum batch size depend on the input image size, while the number of parameters does not.

Deep-ULM and the proposed method have similar encoder-decoder architecture, thereby both require similar number of parameters and FLOPs. However, the proposed method was faster by a factor of 2.3 for processing one image frame since the additional upsampling was not necessary. Also, the proposed method was able to process with roughly 2^3 times more images in a batch. Considering current ULM processing is mostly performed off-line, larger batch size is beneficial as more image frames can be processed

Table 4.3: Comparison of computational complexity given an ultrasound image with a size of 768×272.

Model	Number of parameters	Number of FLOPs	Process time	Maximum batch size
Deep-ULM	5.9×10^6	29×10^9	355 ms	2^3
mSPCN-ULM	0.4×10^6	63×10^9	158 ms	2^6
Proposed	5.8×10^6	23×10^9	156 ms	2^6

in parallel. Contrarily, the number of parameters for mSPCN-ULM was much less because mSPCN-ULM adopts a ResNet style architecture. Nonetheless, the number of FLOPs was much larger since the size of feature maps are kept in the same image resolution as the input before the additional upsampling layers due to the lack of pooling and unpooling operations. Hence, the proposed method achieved comparable process time to mSPCN-ULM with 15 times more model parameters.

4.10 Discussion

In this chapter, a CNN-based localization method that can handle overlapping MBs with sub-pixel accuracy has been introduced. The sub-pixel accuracy was achieved by learning the non-overlapping Gaussian confidence maps and applying Gaussian fitting to the local peaks in the confidence maps. The method was evaluated on the simulation data, phantom measurements, and animal measurements at various MB concentrations. The results showed that the proposed method can separate closely spaced MBs that cannot be separated by centroid detection. In the phantom experiments, the proposed method successfully resolved the pair of channels whose wall-to-wall distance is 22 µm at a high MB concentration when centroid detection failed. And, in the *in vivo* measurements, the proposed method was able to detect more MBs, and MBs closer than 250 µm were separated, which was not achieved by centroid detection.

The proposed method performs localization explicitly at high MB concentrations, unlike some other works that directly produce super-resolved images at high MB concentrations (Bar-Zion et al. 2016; Milecki et al. 2021). Therefore, velocity information, i.e., the magnitude and direction of blood flow can be obtained through tracking, as shown in Fig. 4.13. The track images provide clinical quantities that can be used by clinicians to diagnose diseases, as well as better image quality by filtering out wrong estimations, and the ultimate resolution by separating attached microvessels using their flow directions, which cannot be distinguished in the intensity images.

The effective MB concentrations in local regions are determined by perfusion, vessel size, and microvascular structures, as well as infused MB concentrations, under *in vivo* scenarios. The trends of the number of track samples were locally different, although

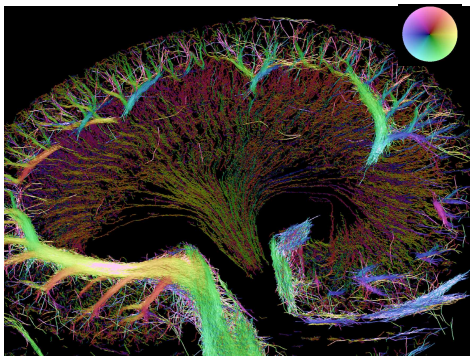


Figure 4.13: The ULM track image generated from the *scenario 2* rat kidney measurement using the proposed localization method and hierarchical Kalman filter. The color wheel on the top right corner represents the magnitude and direction of the velocity. The figure is from Paper 3 (Youn, Taghavi, et al. 2021).

the infused MB concentrations were identical, as shown in Fig. 4.11. For example, the effective MB concentration was higher in the cortex than the inner medulla so, the number of track samples peaked at different scenarios, i.e., infused MB concentrations. The degree of overlaps that can be handled by the proposed method limited. Therefore, ULM at high MB concentrations can give different image qualities depending on the target structures, and the MB concentrations should be selected based on the region of interests and the applications.

Part III

Model-based data-driven methods

CHAPTER 5

Deep Unfolded Ultrasound Microscopy Localization

In this chapter, MB localization using one of the model-based neural networks that embeds a sparse prior, a deep unfolded network, is presented. The model-based network is designed on mathematical formulations, and can thus achieve similar performance to fully data-driven methods with much fewer learning parameters. The model-based data-driven method is assessed on simulated test data and phantom measurements. This chapter is based on Paper 4 (Youn, Luijten, et al. 2020).

5.1 Introduction

Generalization is critical when applying machine learning models to real-world applications. Deep neural networks are model-agnostic so, a mapping from the input to the output is learned fully from given training data using a lot of learning parameters. So, it is difficult to achieve good generalization when the real-world data does not follow the training data distributions. On the contrary, model-based neural networks are designed based on mathematical structures using prior information and underlying domain knowledge. Therefore, the model-based neural networks require much fewer learning parameters and can achieve better generalization even when only a limited number of training data is available.

Deep unfolded networks are one kind of model-based network that solves sparse recovery (Monga, Li, and Eldar 2020), and MB localization for ULM can be defined as a sparse recovery problem (Eldar 2015). In this chapter, MB localization using a deep unfolded network, deep unfolded ULM (van Sloun, Cohen, and Eldar 2020), is introduced, and its performance is compared with centroid detection and other fully-data driven methods.

5.2 Deep Unfolded ULM

5.2.1 Sparse Recovery

Deep unfolded ULM solves MB localization as a sparse recovery problem which can be

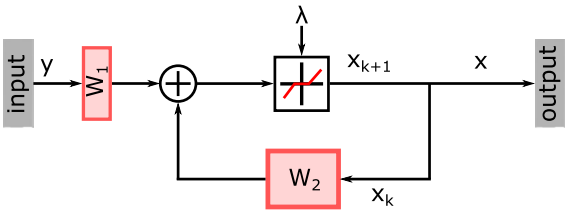


Figure 5.1: Illustration of the ISTA.

formalized as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \tag{5.1}$$

where $\mathbf{y}$ is the upsampled MB ultrasound image, $\mathbf{A}$ is the PSF model, $\mathbf{x}$ is the MB distribution in a high-resolution image grid, and $\mathbf{n}$ is the noise. The MB distribution $\mathbf{x}$ can be assumed to be sparse since the MB positions are represented in the high-resolution image grid. By exploiting the sparse prior, an optimization problem with the the ℓ_1-regularization to find the solution $\hat{\mathbf{x}}$ can be defined by

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1, \tag{5.2}$$

where λ is the regularization coefficient that controls the sparsity of $\mathbf{x}$.

5.2.2 Iterative Shrinkage-thresholding Algorithm

The problem (5.2) can be optimized iteratively using the proximal gradient methods such as the fast iterative shrinkage-thresholding algorithm (FISTA) (Beck and Teboulle 2009). The proximal form of the gradient descent of (5.2) is

$$\mathbf{x}^{k+1} = \text{prox}_{\lambda\|\cdot\|_1}\left(\mathbf{x}^k - \mu\mathbf{A}^\intercal\left(\mathbf{A}\mathbf{x}^k - \mathbf{y}\right)\right), \tag{5.3}$$

where k is the step, μ is the step size, and $\text{prox}_{\lambda\|\cdot\|_1}(\mathbf{x}) = \text{sign}(\mathbf{x})\max(|\mathbf{x}| - \lambda, 0)$ is the proximal operator of the ℓ_1-norm. Equation (5.3) can be simplified by

$$\mathbf{x}^{k+1} = \text{prox}_{\lambda\|\cdot\|_1}\left(\mathbf{W}_1\mathbf{y} + \mathbf{W}_2\mathbf{x}^k\right), \tag{5.4}$$

where $\mathbf{W}_1 = \mu\mathbf{A}^\intercal$ and $\mathbf{W}_2 = \mathbf{I} - \mu\mathbf{A}^\intercal\mathbf{A}$. An overview of the iterative shrinkage-thresholding algorithm (ISTA) scheme is illustrated in Fig. 5.1.

5.2.3 Deep Unfolded Network

Proximal gradient-based methods such as the FISTA (Beck and Teboulle 2009) require many iterations, so it often takes a long time to converge to a solution. Additionally, the solution is highly dependent on the optimization parameters such as the step size μ, the regularization coefficient λ, and the PSF model $\mathbf{A}$; therefore, careful tuning is necessary.

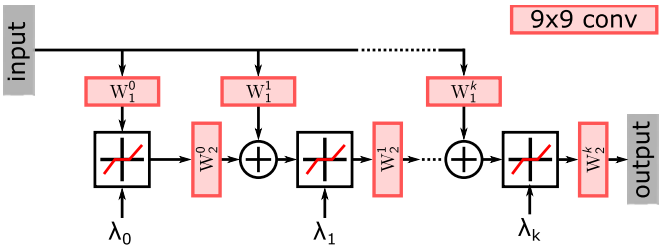


Figure 5.2: Illustration of deep unfolded ULM constructed by unfolding the iteration part of the ISTA in Fig. 5.1. The image is modified from Paper 4 (Youn, Luijten, et al. 2020).

To cope with those limitations, deep unfolded networks (Monga, Li, and Eldar 2020) have been proposed, where the ISTA is learned from data (Gregor and LeCun 2010). The deep unfolded network is formed by unfolding the iteration part of the ISTA and creating a K-layer network with learning parameters $\mathbf{W}_1^k$, $\mathbf{W}_2^k$, and λ_k, as shown in Fig. 5.2. Essentially, the deep unfolded network is a K-iteration ISTA, where the optimization parameters are learnable at each layer from the training data.

Deep unfolded ULM employs deep unfolded networks for MB localization. Hence, it is fast as the iteration part in the ISTA is removed, and it does not require parameter tuning since the optimization parameters are embedded in the network architecture, so the optimal parameter values can be learned from the data during training. Deep unfolded ULM can, therefore, achieve more robust MB localization by learning more diverse PSF models from the training data, compared with the ISTA. Also, much fewer learning parameters are necessary than fully data-driven methods, which allows better generalizability and efficient computation in training and inference.

The sub-pixel localization scheme in Chapter 4 was considered, but the non-overlapping Gaussians were not able to be reconstructed effectively by the deep unfolded networks. Therefore, the MB positions were quantized and represented in a $\lambda/16$ grid, i.e., the binary confidence map, and the input MB data were upsampled by a factor of 4 before being fed to the network. The learning parameters were defined by 9×9 convolutions, so each convolution kernel covers an area of $0.6\lambda \times 0.6\lambda$. The networks were trained by minimizing the MSE loss between true and estimated MB positions using the ADAM optimizer (Kingma and Ba 2015). In the loss function, a 2-D Gaussian is applied to the true MB positions to ensure training stability as follows:

$$\mathcal{L}\left(\mathbf{x}, \mathbf{y}; \theta, \sigma\right) = \frac{1}{N} \sum_{i=1}^{N} \| G\left(\mathbf{y}_i; \sigma\right) - f\left(\mathbf{x}_i; \theta\right) \|_F^2, \tag{5.5}$$

where N is the number of samples, G is the 2-D Gaussian smoothing with a standard deviation of σ, $f\left(\cdot; \theta\right)$ represents the neural network with learning parameters θ, and

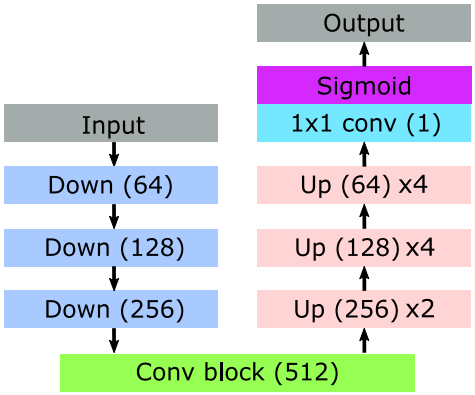


Figure 5.3: Deep-ULM: an encoder-decoder structure CNN that is compared with deep unfolded ULM. The details on the *down-block*, *conv-block*, and *up-block* can be found in (Youn, Ommen, et al. 2020). The upsampling factor of the first *up-block* was 2, but those of the second and third *up-blocks* were 4 to localize MBs in a higher-resolution image grid. The illustration is modified from Paper 4 (Youn, Luijten, et al. 2020)

$\|\cdot\|_F$ is the Frobenius norm. The standard deviation of the Gaussian filter was set to 1 pixel.

5.3 Simulation Experiment

In the simulation experiments, deep unfolded ULM was compared with standard ULM, i.e., centroid detection, and deep-ULM on two test sets. One test set was generated by simulating ultrasound data with randomly placed scatterers in the ROI at different MB densities, where each frame was simulated independently. The other test set was simulated consecutively with the scatterers flowing along a pair of parallel channels whose flow directions are opposite to each other.

Deep-ULM is a fully data-driven method that uses an encoder-decoder structure CNN. The encoder-decoder structure has been widely used for various computer vision and image processing problems such as image segmentation (Badrinarayanan, Kendall, and Cipolla 2017; Ronneberger, Fischer, and Brox 2015) and image generation (Isola et al. 2016). To compare the performance of the methods, deep-ULM was also trained with the same training data. The network architecture of deep-ULM is shown in Fig. 5.3. It was composed of three *down-blocks*, one *conv-block*, and three *up-blocks* (Youn, Ommen, et al. 2020), and the details of each block are presented in Section 3.4. In the encoding, feature maps were downsampled by a factor of 2. In the decoding path, they were upsampled by a factor of 2 in the first *up-block* but by a factor of 4 in the second and third *up-blocks* to localize the MBs in a 4 times higher image grid. The sub-pixel localization scheme

Table 5.1: Field II simulation parameters

	Parameter	Value
Transducer	Transmit frequency	6.9 MHz
	Pitch	30 mm
	Element height	5 mm
	Element width	27 mm
	Number of elements	128
Imaging	Wave type	Plane
	Steering angles	$2 \cdot i^\circ, i \in \{-5, \ldots, 5\}$
	F#	0.5
	# of elements in TX	128
	Apodization in TX	Hann window
	Apodization in RX	Hann window
Environment	Speed of sound	1480 m/s
	Field II sampling frequency	180 MHz

using the non-overlapping Gaussians was not considered for a fair comparison since deep unfolded ULM performed localization in the pixel coordinates without sub-pixel accuracy.

5.3.1 Ultrasound Data Generation

MB data were simulated in Field II pro (Jensen 1996, 2014; Jensen and Svendsen 1992) using plane wave imaging (Tanter and Fink 2014) which allows higher frame rates than conventional line-by-line focused beam transmission. The simulation parameter values are given in Table 5.1. RF channel data were simulated using a 128-element linear array probe with randomly placed scatterers in the ROI. For one image frame, 11 steered plane waves were simulated with a single cycle sinusoid at the frequency of 6.9 MHz. To form an ultrasound image, DAS beamforming with a dynamic apodization and coherent compounding were applied to the simulated RF channel data. The beamforming was performed in a $\lambda/4$ grid, and 256 image frames were generated for the training set.

5.3.2 Result

The evaluation on the randomly placed scatterer test set shows the performance of the localization methods at different MB densities. For the assessment, precision, recall, and the median localization errors were calculated. Precision and recall are defined by

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{5.6}$$

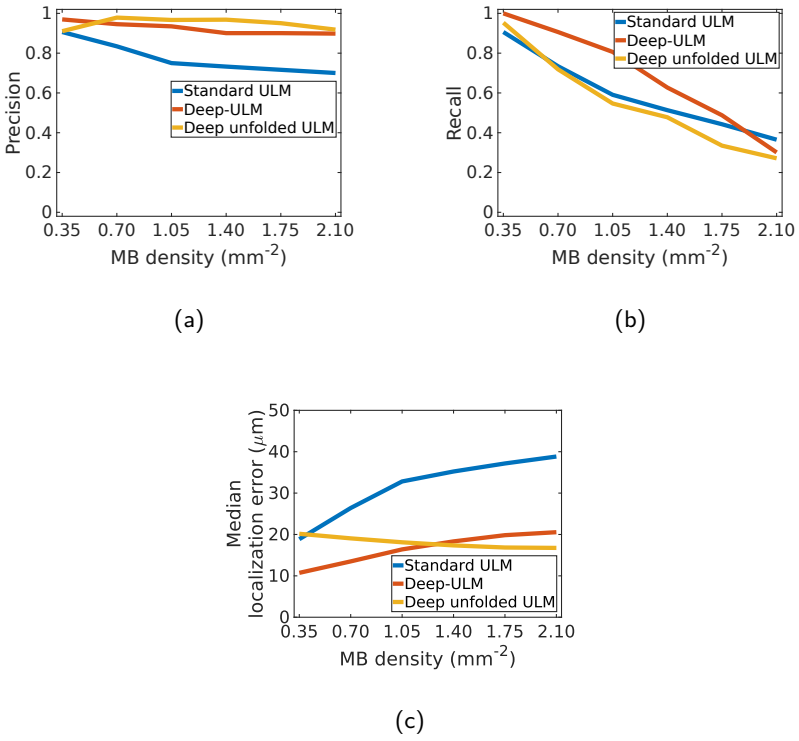


Figure 5.4: Comparison of localization methods at different MB densities. (a) is precision, (b) is recall, and (c) is median localization error. The figure is from Paper 4 (Youn, Luijten, et al. 2020).

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{5.7}$$

where TP is the number of true positive, FP is the number of false positive, and FN is the number of false negative. The true and estimated MBs were matched using the method explained in Section 4.6.1, allowing a localization error of 50 µm (0.23λ).

Fig. 5.4 shows the results of standard ULM, deep-ULM, and deep unfolded ULM. Centroid detection cannot localize overlapping PSFs; therefore the performance of it degraded as the MB density increased. A similar trend was observed for deep-ULM, but its performance was much better than centroid detection since the overlapping PSFs could be handled. Deep unfolded ULM achieved comparable precision and localization uncertainty to deep-ULM, though the recall was not as good as deep unfolded ULM.

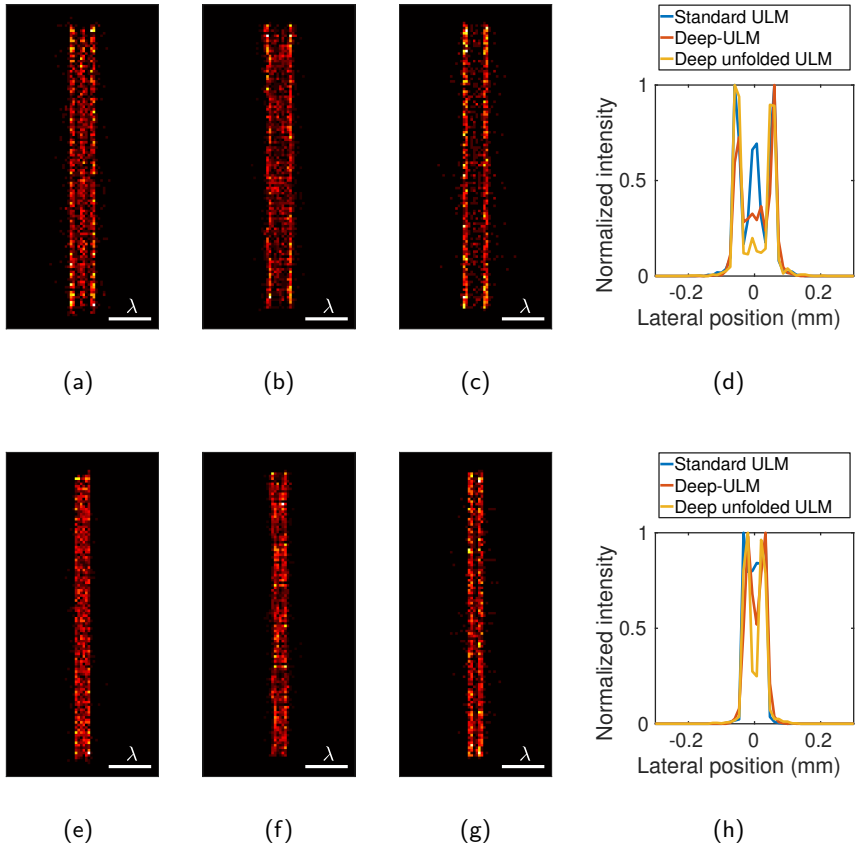


Figure 5.5: Comparison of the methods on the parallel channel simulation data. (a) - (d) are the results of channels separated by $\lambda/2$ and (e) - (h) are the results of channels separated by $\lambda/4$, where (a), (e) are stand ULM, (b), (f) are deep-ULM, (c), (g) are deep unfolded ULM, and (d), (h) are the lateral intensity profile of each method. The figure is from Paper 4 (Youn, Luijten, et al. 2020).

This result shows that deep-ULM can achieve better performance on the data drawn from the same data distribution as the training data, i.e., randomly placed scatterer data, by exploiting a larger number of learning parameters.

To evaluate the methods in more realistic situations, 1024 consecutive frames were simulated with scatterers flowing along a pair of channels separated by $\lambda/2$ and $\lambda/4$. Fig 5.5 shows the ULM reconstruction and intensity profile in the lateral direction. It

was challenging for standard ULM to localize MBs correctly since the channels were spaced closer than the resolution limit of ultrasound imaging. In the reconstructed ULM images by centroid detection, wrong MB estimations were observed in the middle of the channels, where MBs are not supposed to be localized. The intensity profile of both data-driven methods, i.e., deep-ULM and deep unfolded ULM, showed two peaks where the channels were placed and lower intensities in the middle. Comparing the deep learning methods, deep-ULM achieved better imaging quality with fewer false estimations and clearly resolved the channels with better localization precision. This shows better generalizability of deep unfolded ULM to various data distributions that are not seen during training, which is consistent with (Dardikman-Yoffe and Eldar 2020; Monga, Li, and Eldar 2020).

5.4 Phantom Result

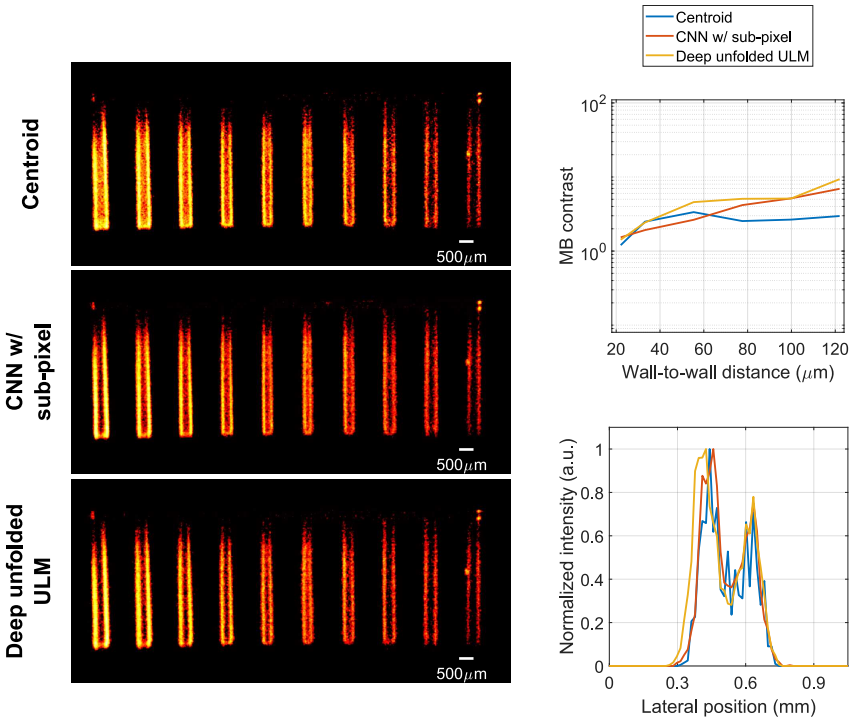
To see how it works on measured data, deep unfolded ULM was applied to the phantom measurements acquired in Section 4.7. For that, a new deep unfolded network was trained using the training data used for training the sub-pixel localization CNN in Chapter 4. The simulation parameters can be found in Table 4.1. As deep unfolded ULM localizes MBs in the pixel coordinates without sub-pixel accuracy, the input image was interpolated in the $\lambda/4$ grid and the MB positions were quantized and represented in the $\lambda/16$ grid.

The ULM reconstruction, MB contrast ratio, and lateral intensity profile at the most closely spaced pair, i.e., the sixth pair from the left, are shown in Fig. 5.6. At the *low* MB concentration (Fig. 5.6(a)), deep unfolded ULM as well as other methods resolved all the pairs clearly, showing that deep unfolded ULM can be generalized to the measured data. At the *high* MB concentration, deep unfolded ULM resolved all the channels successfully while achieving comparable performance to the sub-pixel localization CNN when centroid detection failed. This demonstrates that deep unfolded ULM can localize the overlapping PSFs effectively on the measured data.

It is surprising as deep unfolded ULM requires much fewer learning parameters than the CNN method. The number of learning parameters for deep unfolded ULM was 1735 when that for the sub-pixel localization CNN was about 5.8×10^6. Accordingly, the number of operations was also smaller. To process a 128×128 image, the FLOPs necessary for the CNN model was 12 305 787 392, however, that for deep unfolded ULM was 57 016 320.

5.5 Discussion

MB localization using the deep unfolded network, i.e., deep unfolded ULM, has been presented and evaluated on the simulated and measured data. By learning the optimization parameters from the training data, the deep unfolded network solves the sparse coding problem more effectively and efficiently without iterations and parameter tuning. There-

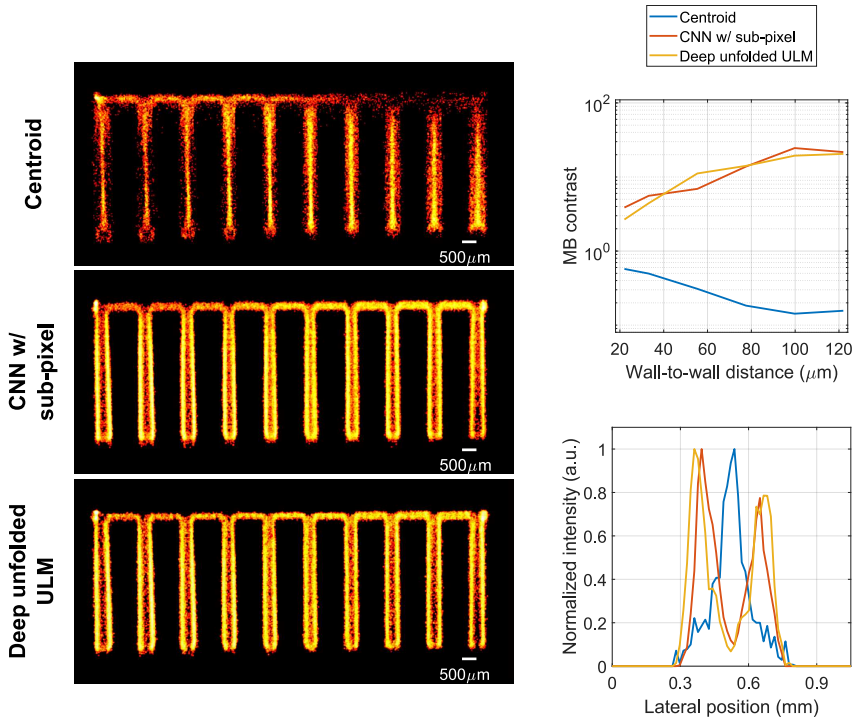


(a)

Figure 5.6: The results on the phantom measurements by centroid detection, the sub-pixel localization CNN which was introduced in Chapter 4, and deep unfolded ULM on the phantom measurements in Section 4.7. The ULM reconstruction, MB contrast ratio, and lateral intensity in the most closely spaced are shown at the (a) *low* concentration and (b) *high* concentration.

fore, deep unfolded ULM can localize MBs more robustly than the ISTA-based methods such as the FISTA.

Compared to the fully data-driven CNN, deep unfolded ULM required fewer learning parameters by a factor of about 3000 thanks to its model-based approach, which allows better generalizability. On the simulated parallel tube data, deep unfolded ULM achieved better imaging quality by localizing MBs more accurately with fewer wrong estimations. Deep unfolded ULM was also well generalized to the phantom measurement with comparable results to the sub-pixel localization CNN. Under the better generalizability,



(b)

Figure 5.6: The results on the phantom measurements by centroid detection, the sub-pixel localization CNN which was introduced in Chapter 4, and deep unfolded ULM on the phantom measurements in Section 4.7. The ULM reconstruction, MB contrast ratio, and lateral intensity in the most closely spaced are shown at the (a) *low* concentration and (b) *high* concentration.

deep unfolded ULM will possibly be able to achieve more robust MB localization than deep-ULM on *in vivo* measurements.

Deep unfolded ULM required fewer FLOPs by a factor of about 200, leading to faster inference. Currently, most of ULM processing is performed off-line due to the high computational complexity such as beamforming, MB localization, motion correction, and tracking. Deep unfolded ULM may open the possibility of real-time ULM by reducing the processing time for localization as well as shorten the data acquisition time by localizing high concentrations of MBs.

CHAPTER 6

Task-adaptive Beamforming and Localization

The performance of MB localization in beamformed ultrasound images is bounded by the adequacy of the beamforming stage even though recent deep learning methods can localize overlapping MBs. This chapter introduces a model-based neural network that performs both beamforming and localization. By doing so, the beamforming stage can be optimized for the subsequent task, i.e., MB localization, and the localization performance can be improved. This chapter is based on Paper 5 (Youn, Luijten, et al. 2021).

6.1 Introduction

Several deep learning methods have been proposed to localize overlapping MBs on beamformed ultrasound images (Brown, Ghosh, and Hoyt 2020; Liu et al. 2020; Milecki et al. 2021; van Sloun, Cohen, and Eldar 2020; van Sloun, Solomon, et al. 2021; Youn, Taghavi, et al. 2021). For those methods, DAS beamforming is commonly used as it is efficient and effective. However, the ability of the deep learning-based localization methods is potentially limited by the adequacy of the beamforming stage as DAS is devised for investigating anatomical structures.

To push this boundary, task-adaptive beamforming is proposed by jointly optimizing a deep neural beamformer and localization network using Adaptive Beamforming by deep LEarning (ABLE) (Luijten et al. 2020) and deep unfolded ULM (van Sloun, Solomon, et al. 2021). ABLE is a model-based neural network that performs content-adaptive beamforming and results in high-quality ultrasound images. By placing ABLE before deep unfolded ULM and training the network as a whole, in an end-to-end fashion, adaptive beamforming weights tailored for the downstream task, i.e., MB localization by deep unfolded ULM, can be obtained. The images beamformed by ABLE ease the downstream localization problem, and thus, the localization performance of deep unfolded ULM can be improved.

6.2 Ultrasound Data Generation

Ultrasound data were simulated in Field II pro (Jensen 1996, 2014; Jensen and Svendsen 1992) using the same imaging sequence introduced in Section 5.3.1. The parameters for

Table 6.1: Field II simulation parameters

	Parameter	Value
Transducer	Transmit frequency	6.9 MHz
	Pitch	30 mm
	Element height	5 mm
	Element width	27 mm
	Number of elements	128
Imaging	Wave type	Plane
	Steering angles	$2 \cdot i^{\circ}, i \in \{-5, \ldots, 5\}$
	F#	0.5
	# of elements in TX	128
	Apodization in TX	Hann window
	Apodization in RX	Hann window
Environment	Speed of sound	1480 m/s
	Field II sampling frequency	180 MHz

the simulation are presented in Table 6.1. One image frame was simulated with randomly placed point scatterers and 11 plane waves. The simulated RF channel data were then delayed on a $\lambda/4$ grid but not summed along the channel directions. Therefore, the data size of one image frame was M $\times$ N $\times$ C $\times$ 11, where M and N are the numbers of image points in the axial and lateral directions, C is the number of the transducer channels, and 11 is the number of transmission events. For training, 768 frames were generated.

6.3 Network Architecture

The overview of the proposed network is shown in Fig. 6.1(a). The network performing task-adaptive beamforming and MB localization is designed by incorporating ABLE (Luijten et al. 2020) into deep unfolded ULM (van Sloun, Cohen, and Eldar 2020). By training the beamformer and localization networks as an end-to-end fashion, beamforming weights that are optimal for the downstream task, i.e., MB localization by deep unfolded ULM, can be learned. In this network, the beamforming part firstly estimates the task-adaptive beamforming weights from delayed but not summed RF channel data. And then, the estimated weights are applied to the channel data, and the beamformed images are compounded. Finally, MB localization is performed on the image beamformed and compounded by the network.

Fig. 6.1(b) shows the beamforming part of the network for one transmission event. A 5 $\times$ 5 convolution layer is placed before ABLE to offer a larger receptive field, so

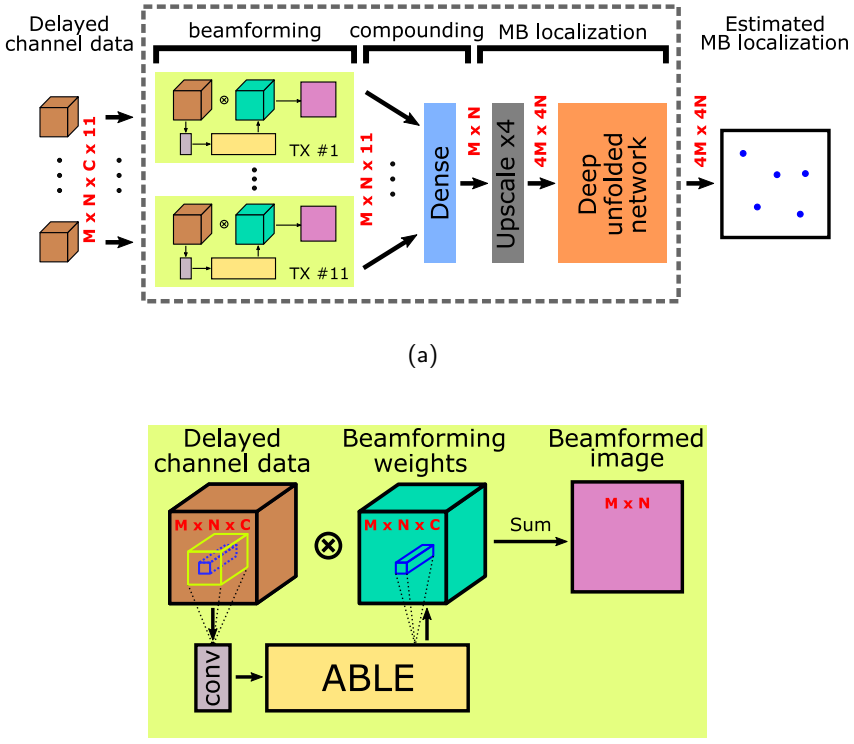


Figure 6.1: A schematic overview of the proposed network. (a) shows the whole pipeline and (b) shows the beamforming process for one transmit event. The proposed network takes delayed RF channel data as input and performs beamforming. Here, optimal apodization weights for the downstream task (i.e., MB localization) are learned by ABLE(Luijten et al. 2020). This is why the method is referred to as task-adaptive beamforming. After that, beamformed signals from each transmit are compounded using a dense layer, and MBs are localized using deep unfolded ULM (van Sloun, Cohen, and Eldar 2020; Youn, Luijten, et al. 2020) in the image beamformed and compounded by the network. The red text represents data size. The illustration is modified from Paper 5 (Youn, Luijten, et al. 2021).

that the subsequent beamformer network, i.e., ABLE, can consider neighboring pixels when estimating the beamforming weights at a pixel position. Further details of ABLE can be found in (Luijten et al. 2020). Distinct convolution layers and ABLE networks were defined for each transmission event, i.e., 11 convolution layers and ABLE networks. The beamformed images were then compounded by a dense layer that learns a weighted summation. After the dense layer, a beamformed and compounded ultrasound image whose size is M $\times$ N can be obtained.

Deep unfolded ULM cannot localize MBs with sub-pixel accuracy. To minimize the quantization error of the estimated MB positions, MB localization was performed in a $\lambda/16$ image grid after upsampling the beamformed images by a factor of 4. Deep unfolded ULM was explained in Section 5.2 and the details can be in (Monga, Li, and Eldar 2020; van Sloun, Solomon, et al. 2021). A 10-layer deep unfolded network composed of 9×9 convolutions was used for localization.

The proposed network was trained by minimizing the MSE loss with the 2-D Gaussian smoothing as follows:

$$\mathcal{L}\left(\mathbf{x}, \mathbf{y}; \theta, \sigma\right) = \frac{1}{n} \sum_{i=1}^{n} \| G\left(\mathbf{y}_i; \sigma\right) - f\left(\mathbf{x}_i; \theta\right) \|_F^2, \tag{6.1}$$

where $\mathbf{x}_i$ is the i-th delayed RF channel data in the $\lambda/4$ grid, $\mathbf{y}_i$ is the i-th MB positions represented in the $\lambda/16$ grid, n is the number of data, G is the 2-D Gaussian filtering with the standard deviation of σ, $f\left(\cdot; \theta\right)$ is the neural network function with learning parameters θ, and $\|\cdot\|_F$ is the Frobenius norm.

The standard deviation of the Gaussian filter was initially set to 4 pixels and training was performed with 1000 epochs, where the initial learning rate was 0.0001 and it was halved every 200 epochs. After that, the network was further trained with the standard deviation of 1 pixel for 200 epochs to acquire sharper MB distributions, where the learning rate was 0.0001 and it was halved every 40 epochs.

6.4 Simulation Result

The task-adaptive beamforming and localization method was evaluated on simulated test data and compared against centroid detection and deep unfolded ULM. Deep unfolded ULM was trained with the images obtained by beamforming the channel data used for training the proposed method. Beamforming was performed by DAS with a dynamic apodization where the F# was 0.5.

The mapping from RF channel data to MB positions is learned in an end-to-end fashion, although the proposed neural network is constructed by combining two networks. Therefore, the beamforming network is tailored for the downstream task, i.e., MB localization. Fig. 6.2 shows a DAS beamformed image and a network beamformed image before MB localization on a test data. The network beamformed image in Fig. 6.2(b)

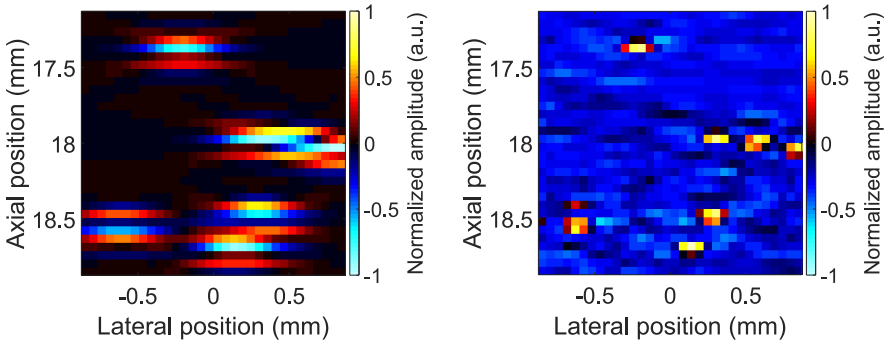


Figure 6.2: A comparison of (a) DAS beamformed image with a dynamic apodization where the F# is 0.5 and (b) task-adaptive beamforming result which was jointly trained with deep unfolded ULM. The task-adaptive beamforming achieved sharper peaks at MB positions. The image is from Paper 5 (Youn, Luijten, et al. 2021).

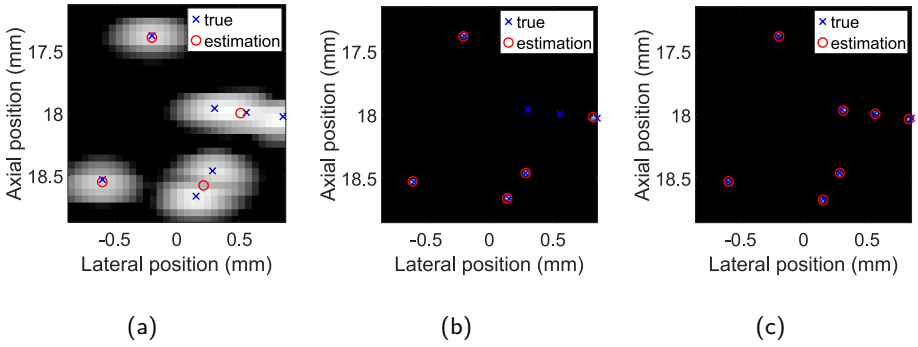
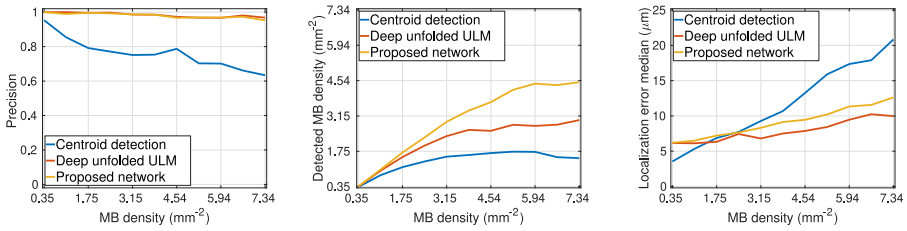


Figure 6.3: A comparison of different method localization results on the same test data used in Fig. 6.2. (a) is centroid detection on the DAS beamformed and envelope detected image, (b) is deep unfolded ULM on the DAS beamformed RF image, and (c) is the result of the jointly optimized task-adaptive beamforming and localization network.



(a) (b) (c)

Figure 6.4: A comparison of centroid detection, deep unfolded ULM, and task-adaptive beamforming and localization. (a) is precision, (b) is detected MB density, and (c) is median localization error at different MB densities.

resulted in sharper peaks at MB positions compared to the DAS beamformed image in Fig. 6.2(a).

Fig. 6.3 shows a comparison of the localization results on the same test data used in Fig. 6.2. Fig. 6.4 shows precision, detected MB density, and median localization error, following the methods presented in Section 5.3, at different MB densities. Centroid detection on the DAS beamformed image (Fig. 6.3(a)) could localize isolated MBs accurately but failed to localize overlapping PSFs and produced a wrong estimation. For the result of deep unfolded ULM on the DAS beamformed RF image (Fig. 6.3(b)), two MBs were missed, but other MBs were accurately localized without producing the wrong estimations. This explains higher precision of deep unfolded ULM than centroid detection, as shown in Fig. 6.4(a). The task-adaptive beamforming and localization network estimated all the MBs at the correction positions. For deep unfolded ULM, the degree of overlaps can be handled was limited due to DAS beamforming. However, for the proposed method, the beamforming part already made localization easier by making sharper peaks at the MB positions, as shown in Fig. 6.2(b). Therefore, it was able to localize all the MBs successfully. This is also shown in Fig. 6.4(b) that the detected MB density of the proposed method converged to a higher density than deep unfolded ULM. Deep unfolded ULM and the proposed method showed comparable localization precision, as shown in Fig. 6.4(c).

6.5 Discussion

A beamformer and localization network which can localize high concentrations of MBs is proposed. The network is constructed by incorporating ABLE into deep unfolded ULM. Hence, adaptive beamforming weights optimal for deep unfolded ULM to locate MBs was able to be learned. The beamforming network was not optimized for perceptually

appealing images for humans, but to ease the subsequent MB localization. The beamformed results showed sharper peaks at the MB positions than DAS. By localizing MBs in the sharper images, the proposed network achieved higher recall than deep unfolded ULM while keeping similar precision and localization error to deep unfolded ULM. The model-based beamformer and localization network can possibly be used to reduce the data acquisition time of ULM by localizing more MBs accurately, whilst employing high concentrations of MBs.

In this chapter, the task-adaptive beamforming scheme was used for MB localization, however, its application is not limited to localization. By combining the beamforming network with another network performing some task, the adaptive beamforming weights dedicated to the given task can be learned. Examples of such tasks include clutter removal, artifact suppression, and diagnosis. It may open up a new approach of accommodating beamforming for the downstream task.

Part IV

Conclusion

CHAPTER 7

Conclusion

The goal of this Ph.D. project is to localize more MBs by employing high concentrations of MBs to decrease data acquisition time of ULM. The standard localization methods suffer when the targets are closer than the resolution limit and overlapping PSFs appear. Several data-driven deep learning methods have been investigated to localize the overlapping PSFs accurately and robustly.

Fully Data-driven Methods

Localization directly on RF channel data using an encoder-decoder structure CNNs was studied and presented in Chapter 3. The training was performed with simulated ultrasound channel data. Non-overlapping confidence map has been proposed to provide large gradients for stable training without losing closely spaced target positions. In the simulation results, the CNN localization method outperformed peak detection and deconvolution methods by localizing scatterers placed closer than the resolution of ultrasound. Also, the comparison of the proposed method with deep-ULM has shown the great potential of localizing scatterers on the RF channel data without explicit beamforming. For the assessment on measured data, 3-D printed scatterer phantoms were fabricated. And the training data were updated following the physical properties of the scatterers inside the phantoms due to the generalization problem. The CNN trained with the updated data was able to identify scatterers, however, its performance on the phantom measurements was not as good as on the simulated data.

In Chapter 4, the CNN localization method on the beamformed ultrasound data was investigated since it was difficult to simulate the channel data accurately and the accessibility to the channel data is limited. Especially, sub-pixel accuracy was achieved by extending the non-overlapping Gaussian confidence maps in the continuous spatial domain and applying Gaussian fitting to the local peaks in the estimated confidence maps. This method does not require additional upsampling, unlike other deep learning-based methods; therefore, the computation was faster and required less GPU memory. The sub-pixel CNN method was evaluated in a 3-D printed channel phantom, where the dimensions of the channel was known. At a high MB concentration, the sub-pixel CNN method successfully resolved the pair of channels whose wall-to-wall distance is 22 µm when centroid detection failed. Furthermore, in the *in vivo* measurements, more MBs were estimated by localizing MBs spaced closer than 250 µm ($\approx \lambda$), which was not available by centroid detection.

Model-based Data-driven Methods

Deep unfolded ULM that solves the MB localization problem using deep unfolded networks, which are one kind of model-based neural network that can solve the sparse recovery problem, has been investigated in Chapter 5. Deep unfolded ULM required much fewer learning parameters thanks to its model-based approach. Therefore, it was able to achieve better generalizability to out of training data distributions compared with deep-ULM, the model-agnostic fully data-driven method. In addition, the training of deep unfolded ULM took a short time and its inference was fast by virtue of the fewer learning parameters. On the phantom measurements obtained in Chapter 4, similar performance to the sub-pixel CNN localization method was achieved by deep unfolded ULM. Considering its better generalizability, more robust MB localization is expected on *in vivo* measurements.

To further employ the model-based network approach in the ultrasound data processing chain, the network that performs beamforming and MB localization simultaneously was investigated in Chapter 6. The task-adaptive beamforming network was constructed by combining two model-based networks that estimate content-adaptive beamforming weights (ABLE) and localize MBs (deep unfolded ULM). The beamforming and localization network was trained as an end-to-end fashion; hence, the beamforming part was optimized to learn the optimal weights for the downstream localization task from training data. The ultrasound images beamformed by the task-adaptive beamformer showed sharper peaks at the MB positions than the DAS beamformed ultrasound images. The sharper peaks eased the downstream localization problem, and thus, more MBs were localized at high MB densities. At the MB density of 7.34 mm^{-2}, the beamforming and localization method reconstructed 4.48 mm^{-2} when centroid detection and deep unfolded ULM reconstructed 1.47 mm^{-2} and 2.98 mm^{-2}, respectively.

7.1 Perspective and Outlook

From this Ph.D. project, it has been shown that deep learning-based data-driven localization methods outperform other localization methods for localizing overlapping high concentrations of MBs. For applying the deep learning methods to real-world applications, more efficient and generalized models for ultrasound signal processing need to be investigated. CNNs showed descent performance, however, by employing the model-based deep learning, deep unfolded ULM achieved comparable performance with 3000 times fewer learning parameters and 200 times faster inference, compared to deep-ULM, a fully data-driven method. For the moment, most ULM processing is performed off-line, as it takes a long time. The advanced deep learning models will be able to decrease the processing time and help ULM operate in real time.

This project mostly focused on applying deep learning to MB localization, but it can be extended to other ULM processing. The task-adaptive beamforming and localization has already shown that the localization performance can be improved by incorporating

the beamforming stage into the localization network, although it is still in its early stage and further development needs to follow. Specifically, deep learning can be applied to consider the temporal correlation for clutter rejection, motion correction, and tracking. Each step of ULM that determines the final image quality is interdependent. Ultimately, deep learning models that perform the whole ULM chain need to be developed to optimize each step jointly.

Lastly, there are fundamental limitations of 2-D ULM for scanning 3-D structures. The 2-D ultrasound image is an integration over the elevation beam profile, which can potentially degrade localization. Also, it is difficult to capture the out-of-plane motion for motion correction. There have been 3-D ULM using fully-addressed matrix probes or row-column addressed matrix probes. Accordingly, the deep learning methods need be extended for 3-D ULM.

Bibliography

References from Chapter 1

Andersen, S. B., I. Taghavi, C. A. V. Hoyos, S. B. Søgaard, F. Gran, L. Lonn, K. L. Hansen, J. A. Jensen, M. B. Nielsen, and C. M. Sørensen (2020). "Super-Resolution Imaging with Ultrasound for Visualization of the Renal Microvasculature in Rats Before and After Renal Ischemia: A Pilot Study". In: *Diagnostics* 10.11, p. 862 (cit. on p. 4).

Bohs, L. N., B. H. Friemal, B. A. McDermott, and G. E. Trahey (1993). "A real-time system for quantifying and displaying two-dimensional velocities using ultrasound". In: *Ultrasound Med. Biol.* 19, pp. 751–761 (cit. on p. 3).

Brown, T. B., B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, and T. H (2020). "Language Models are Few-Shot Learners". In: *arXiv:2005.14165v4 [cs.CL]* (cit. on p. 4).

Chen, L., Y. Zhu, G. Papandreou, F. Schroff, and H. Adam (2018). "Encoder-decoder with atrous separable convolution for semantic image segmentation". In: *Eur. Conf. Computer Vision*, pp. 801–818 (cit. on p. 4).

Christensen-Jeffries, K., R. J. Browning, M. Tang, C. Dunsby, and R. J. Eckersley (Feb. 2015). "In Vivo Acoustic Super-Resolution and Super-Resolved Velocity Mapping Using Microbubbles". In: *IEEE Trans. Med. Imag.* 34.2, pp. 433–440 (cit. on p. 4).

Couture, O., B. Besson, G. Montaldo, M. Fink, and M. Tanter (2011). "Microbubble ultrasound super-localization imaging (MUSLI)". In: *Proc. IEEE Ultrason. Symp.* Pp. 1285–1287 (cit. on p. 4).

Deffieux, T., C. Demene, M. Pernot, and M. Tanter (2018). "Functional ultrasound neuroimaging: a review of the preclinical and clinical state of the art". In: *Curr. Opin. Neurol.* 50, pp. 128–135 (cit. on p. 4).

Devlind, J., M.-W. Chang, K. Lee, and K. Toutanov (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". In: *arXiv:1810.04805v2 [cs.CL]* (cit. on p. 4).

Dunmire, B., K. W. Beach, K.-H. Labs., M. Plett, and D. E. Strandness (2000). "Cross-beam vector Doppler ultrasound for angle independent velocity measurements". In: *Ultrasound Med. Biol.* 26, pp. 1213–1235 (cit. on p. 3).

Edler, I. and C. H. Hertz (1954). "The use of ultrasonic reflectoscope for the continuous recording of the movement of heart walls". In: *Kungl. Fysiogr. Sällskap. i Lund Föhandl* 24, pp. 40–58 (cit. on p. 3).

Errico, C., J. Pierre, S. Pezet, Y. Desailly, Z. Lenkei, O. Couture, and M. Tanter (Nov. 2015). "Ultrafast ultrasound localization microscopy for deep super-resolution vascular imaging". In: *Nature* 527, pp. 499–502 (cit. on p. 4).

Ghosh, D., J. Peng, K. Brown, S. Sirsi, C. Mineo, P. W. Shaul, and K. Hoyt (2019). "Super-Resolution Ultrasound Imaging of Skeletal Muscle Microvascular Dysfunction in an Animal Model of Type 2 Diabetes". In: *J. Ultrasound Med.* 38.10, pp. 2589–2599 (cit. on p. 4).

Goodfellow, I. J., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio (2014). "Generative Adversarial Networks". In: *arXiv:1406.2661v1 [stat.ML]* (cit. on p. 4).

He, K., G. Gkioxari, P. Dollár, and R. Girshick (2017). "Mask R-CNN". In: *IEEE Int. Conf. Computer Vision*, pp. 2980–2988 (cit. on p. 4).

He, K., X. Zhang, S. Ren, and J. Sun (2016a). "Deep Residual Learning for Image Recognition". In: *IEEE Conf. Computer Vision and Pattern Recognition*, pp. 770–778 (cit. on p. 4).

— (2016b). "Identity Mappings in Deep Residual Networks". In: *Eur. Conf. Computer Vision*, pp. 630–645 (cit. on p. 4).

Howry, D. H. and W. R. Bliss (1952). "Ultrasonic visualization of soft tissue structures of the body". In: *J. Lab. Clin. Med.* 40, pp. 579–592 (cit. on p. 3).

Huang, G., Z. Liu, L. v. d. Maaten, and K. Q. Weinberger (2017). "Densely connected convolutional networks". In: *IEEE Conf. Computer Vision and Pattern Recognition*, pp. 2261–2269 (cit. on p. 4).

Huijben, I. A. M., B. S. Veeling, K. Janse, M. Mischi, and R. J. G. van Sloun (2020). "Learning Sub-Sampling and Signal Recovery with Applications in Ultrasound Imaging". In: *IEEE Trans. Med. Imag.* 39.12, pp. 3955–3966 (cit. on p. 4).

Hyun, D., L. L. Brickson, K. T. Looby, and J. J. Dahl (2019). "Beamforming and speckle reduction using neural networks". In: *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.* 66.5, pp. 898–910 (cit. on p. 4).

Jensen, J. A. and P. Munk (1998). "A New Method for Estimation of Velocity Vectors". In: *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.* 45.3, pp. 837–851 (cit. on p. 3).

Karras, T., T. Aila, S. Laine, and J. Lehtinen (2018). "Progressive Growing of GANs for Improved Quality, Stability, and Variation". In: *Int. Conf. Learning Representations* (cit. on p. 4).

Khan, S., J. Huh, and J. C. Ye (2020). "Adaptive and Compressive Beamforming Using Deep Learning for Medical Ultrasound". In: *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.* 67.8, pp. 1558–1572 (cit. on p. 4).

Krizhevsky, A., I. Sutskever, and G. E. Hinton (2012). "ImageNet Classification with Deep Convolutional Neural Networks". In: *Neural Information Processing Systems*, pp. 1097–1105 (cit. on p. 4).

Ledig, C., L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, and Z. Wang (2017). "Photo-Realistic Single Image Super-Resolution

Using a Generative Adversarial Network". In: *IEEE Conf. Computer Vision and Pattern Recognition*, pp. 105–114 (cit. on p. 4).

Lim, B., S. Son, H. Kim, S. Nah, and K. M. Lee (2017). "Enhanced Deep Residual Networks for Single Image Super-Resolution". In: *IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1132–1140 (cit. on p. 4).

Lin, F., S. E. Shelton, D. Espindola, J. D. Rojas, G. Pinton, and P. A. Dayton (2017). "3-D Ultrasound Localization Microscopy for Identifying Microvascular Morphology Features of Tumor Angiogenesis at a Resolution Beyond the Diffraction Limit of Conventional Ultrasound". In: *Theranostics* 7.1, pp. 196–204. DOI: 10.7150/thno.16899 (cit. on p. 4).

Luchies, A. C. and B. C. Byram (2018). "Suppressing off-axis scattering using deep neural networks". In: *Proc. SPIE Med. Imag.* Vol. 10580, pp. 10580-10580–8 (cit. on p. 4).

Luijten, B., R. Cohen, F. J. De Bruijn, H. A. W. Schmeitz, M. Mischi, Y. C. Eldar, and R. J. G. Van Sloun (2020). "Adaptive Ultrasound Beamforming using Deep Learning". In: *IEEE Trans. Med. Imag.* 39.12, pp. 3967–3978 (cit. on p. 4).

O'Reilly, M. A. and K. Hynynen (2013). "A super-resolution ultrasound method for brain vascular mapping". In: *Med. Phys.* 40.11, pp. 110701–7. DOI: 10.1118/1.4823762 (cit. on p. 4).

Radford, A., L. Metz, and S. Chintala (2016). "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks". In: *Int. Conf. Learning Representations* (cit. on p. 4).

Redmon, J. and A. Farhadi (2018). "Yolov3: An incremental improvement". In: *arXiv:1804.02767v1 [cs.CV]* (cit. on p. 4).

Ronneberger, O., P. Fischer, and T. Brox (2015). "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241 (cit. on p. 4).

Siepmann, M., G. Schmitz, J. Bzyl, M. Palmowski, and F. Kiessling (2011). "Imaging tumor vascularity by tracing single microbubbles". In: *Proc. IEEE Ultrason. Symp.*, pp. 6293297, 1906–1908. DOI: 10.1109/ULTSYM.2011.0476 (cit. on p. 4).

Solomon, O., R. Cohen, Y. Zhang, Y. Yang, Q. He, J. Luo, R. J. G. van Sloun, and Y. C. Eldar (2020). "Deep unfolded robust PCA with application to clutter suppression in ultrasound". In: *IEEE Trans. Med. Imag.* 39.4, pp. 1051–1063 (cit. on p. 4).

Viessmann, O. M., R. J. Eckersley, K. Christensen-Jeffries, M. X. Tang, and C. Dunsby (2013). "Acoustic super-resolution with ultrasound and microbubbles". In: *Phys. Med. Biol.* 58, pp. 6447–6458 (cit. on p. 4).

Wild, J. J. (1950). "The use of ultrasonic pulses for the measurement of biologic tissues and the detection of tissue density changes". In: *Surgery* 27, pp. 183–188 (cit. on p. 3).

Yoon, Y. H., S. Khan, J. Huh, and J. C. Ye (2018). "Efficient B-mode Ultrasound Image Reconstruction from Sub-sampled RF Data using Deep Learning". In: *IEEE Trans. Med. Imag.* (cit. on p. 4).

Youn, J., B. Luijten, M. B. Stuart, Y. C. Eldar, R. J. G. van Sloun, and J. A. Jensen (2020). "Deep Learning Models for Fast Ultrasound Localization Microscopy". In: *Proc. IEEE Ultrason. Symp.* Pp. 1–4 (cit. on p. 6).

Youn, J., B. Luijten, M. B. Stuart, Y. C. Eldar, R. J. G. van Sloun, and J. A. Jensen (2021). "Model-based Deep Learning on Ultrasound Channel Data for Fast Ultrasound Localization Microscopy". In: *In preparation* (cit. on p. 7).

Youn, J., M. L. Ommen, M. B. Stuart, E. V. Thomsen, N. B. Larsen, and J. A. Jensen (2019). "Ultrasound Multiple Point Target Detection and Localization using Deep Learning". In: *Proc. IEEE Ultrason. Symp.* Pp. 1937–1940 (cit. on p. 6).

— (2020). "Detection and Localization of Ultrasound Scatterers Using Convolutional Neural Networks". In: *IEEE Trans. Med. Imag.* 39.12, pp. 3855–3867 (cit. on p. 6).

Youn, J., I. Taghavi, M. L. Ommen, M. Schou, M. B. Stuart, E. V. Thomsen, N. B. Larsen, and J. A. Jensen (2021). "Sub-pixel Accuracy Microbubble Localization using Convolutional Neural Networks". In: *In preparation* (cit. on p. 6).

Zhao, H., J. Shi, X. Qi, X. Wang, and J. Jia (2017). "Pyramid scene parsing network". In: *IEEE Conf. Computer Vision and Pattern Recognition*, pp. 2881–2890 (cit. on p. 4).

References from Chapter 2

Ackermann, D. and G. Schmitz (Jan. 2016). "Detection and Tracking of Multiple Microbubbles in Ultrasound B-Mode Images". In: *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.* 63.1, pp. 72–82 (cit. on p. 12).

Andersen, S. B., I. Taghavi, C. A. V. Hoyos, S. B. Søgaard, F. Gran, L. Lonn, K. L. Hansen, J. A. Jensen, M. B. Nielsen, and C. M. Sørensen (2020). "Super-Resolution Imaging with Ultrasound for Visualization of the Renal Microvasculature in Rats Before and After Renal Ischemia: A Pilot Study". In: *Diagnostics* 10.11, p. 862 (cit. on p. 12).

Bar-Zion, A., O. Solomon, C. Tremblay-Darveau, D. Adam, and Y. C. Eldar (2018). "SUSHI: Sparsity-based ultrasound super-resolution hemodynamic imaging". In: *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.* 65.12, pp. 2365–2380 (cit. on p. 12).

Bar-Zion, A., C. Tremblay-Darveau, O. Solomon, D. Adam, and Y. C. Eldar (2016). "Fast vascular ultrasound imaging with enhanced spatial resolution and background rejection". In: *IEEE Trans. Med. Imag.* 36.1, pp. 169–180 (cit. on p. 12).

Betzig, E., G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess (2006). "Imaging intracellular fluorescent proteins at nanometer resolution". In: *Science* 313.5793, pp. 1642–1645 (cit. on p. 9).

Brock-Fischer, G. A., M. D. Poland, and P. G. Rafter (1996). *Means for increasing sensitivity in non linear ultrasound systems*. US Patent (5577505) (cit. on p. 10).

Brown, J., K. Christensen-Jeffries, S. Harput, G. Zhang, J. Zhu, C. Dunsby, M.-X. Tang, and R. J. Eckersley (2019). "Investigation of Microbubbles Detection Methods for

Super-Resolution Imaging of Microvasculature". In: *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.* 66.4, pp. 676–691 (cit. on p. 11).

Christensen-Jeffries, K., R. J. Browning, M. Tang, C. Dunsby, and R. J. Eckersley (Feb. 2015). "In Vivo Acoustic Super-Resolution and Super-Resolved Velocity Mapping Using Microbubbles". In: *IEEE Trans. Med. Imag.* 34.2, pp. 433–440 (cit. on pp. 9, 11–13).

Christensen-Jeffries, K., O. Couture, P. A. Dayton, Y. C. Eldar, K. Hynynen, F. Kiessling, M. O'Reilly, G. F. Pinton, G. Schmitz, M. Tang, et al. (2020). "Super-resolution ultrasound imaging". In: *Ultrasound Med. Biol.* 46.4, pp. 865–891 (cit. on p. 9).

Couture, O., B. Besson, G. Montaldo, M. Fink, and M. Tanter (2011). "Microbubble ultrasound super-localization imaging (MUSLI)". In: *Proc. IEEE Ultrason. Symp.* Pp. 1285–1287 (cit. on pp. 9, 11).

Demene, C., T. Deffieux, M. Pernot, B.-F. Osmanski, V. Biran, J.-L. Gennisson, L.-A. Sieu, A. Bergel, S. Franqui, J.-M. Correas, I. Cohen, O. Baud, and M. Tanter (2015). "Spatiotemporal clutter filtering of ultrafast ultrasound data highly increases Doppler and fUltrasound sensitivity". In: *IEEE Trans. Med. Imag.* 34.11, pp. 2271–2285. DOI: 10.1109/TMI.2015.2428634 (cit. on p. 10).

Desailly, Y., O. Couture, M. Fink, and M. Tanter (2013). "Sono-activated ultrasound localization microscopy". In: *Appl. Phys. Lett.* 103.17, p. 174107 (cit. on pp. 10, 11).

Diamantis, K., T. Anderson, M. B. Butler, C. A. V. Hoyos, J. A. Jensen, and V. Sboros (2018). "Resolving Ultrasound Contrast Microbubbles using Minimum Variance Beamforming". In: *IEEE Trans. Med. Imag.* DOI: 10.1109/TMI.2018.2859262 (cit. on p. 11).

Eckersley, R. J., C. T. Chin, and P. N. Burns (2005). "Optimising phase and amplitude modulation schemes for imaging microbubble contrast agents at low acoustic power". In: *Ultrasound Med. Biol.* 31.2, pp. 213–219 (cit. on p. 10).

Errico, C., J. Pierre, S. Pezet, Y. Desailly, Z. Lenkei, O. Couture, and M. Tanter (Nov. 2015). "Ultrafast ultrasound localization microscopy for deep super-resolution vascular imaging". In: *Nature* 527, pp. 499–502 (cit. on pp. 9, 11, 12).

Foiret, J., H. Zhang, T. Ilovitsh, L. Mahakian, S. Tam, and K. W. Ferrara (2017). "Ultrasound localization microscopy to image and assess microvasculature in a rat kidney". In: *Scientific Reports* 7.1, 13662:1–12. DOI: 10.1038/s41598-017-13676-7 (cit. on p. 12).

Ghosh, D., J. Peng, K. Brown, S. Sirsi, C. Mineo, P. W. Shaul, and K. Hoyt (2019). "Super-Resolution Ultrasound Imaging of Skeletal Muscle Microvascular Dysfunction in an Animal Model of Type 2 Diabetes". In: *J. Ultrasound Med.* 38.10, pp. 2589–2599 (cit. on p. 12).

Harput, S., K. Christensen-Jeffries, J. Brown, Y. Li, K. J. Williams, A. H. Davies, R. J. Eckersley, C. Dunsby, and M. Tang (2018). "Two-Stage Motion Correction for Super-Resolution Ultrasound Imaging in Human Lower Limb". In: *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.* 65.5, pp. 803–814. DOI: 10.1109/TUFFC.2018.2824846 (cit. on p. 12).

Harput, S., K. Christensen-Jeffries, A. Ramalli, J. Brown, J. Zhu, G. Zhang, C. H. Leow, M. Toulemonde, E. Boni, P. Tortoli, R. J. Eckersley, C. Dunsby, and M. Tang (Feb. 2020). "3-D Super-Resolution Ultrasound Imaging With a 2-D Sparse Array". In: *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.* 67.2, pp. 269–277 (cit. on p. 13).

Heiles, B., M. Correia, V. Hingot, M. Pernot, J. Provost, M. Tanter, and O. Couture (Sept. 2019). "Ultrafast 3D Ultrasound Localization Microscopy Using a 32 x 32 Matrix Array". In: *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.* 38.9, pp. 2005–2015 (cit. on p. 13).

Hess, S. T., T. P. K. Girirajan, and M. D. Mason (2006). "Ultra-high resolution imaging by fluorescence photoactivation localization microscopy". In: *Biophysical Journal* 91.11, pp. 4258–4272 (cit. on p. 9).

Ikeda, O., T. Sato, and K. Suzuki (1979). "Super-resolution imaging system using waves with a limited frequency bandwidth". In: *J. Acoust. Soc. Am.*, pp. 75–81 (cit. on p. 9).

Jensen, J. A., S. Nikolov, B. Tomov, F. Gran, M. Hansen, and T. V. Hansen (2006). *Specification of SARUS: the Synthetic Aperture Real-time Ultrasound Scanner*. Tech. rep. Ørsted • DTU, Technical University of Denmark (cit. on p. 11).

Jensen, J. A., M. L. Ommen, S. H. Øygard, M. Schou, T. Sams, M. B. Stuart, C. Beers, E. V. Thomsen, N. B. Larsen, and B. G. Tomov (2020). "Three-Dimensional Super Resolution Imaging using a Row-Column Array". In: *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.* 67.3, pp. 538–546. DOI: 10.1109/TUFFC.2019.2948563 (cit. on p. 13).

Lin, C., Y. Chang, and L. Chuang (2016). "Early detection of diabetic kidney disease: Present limitations and future perspectives". In: *World Journal of Diabetes* 7.14, pp. 290–301. DOI: 10.4239/wjd.v7.i14.290 (cit. on p. 12).

Lin, F., S. E. Shelton, D. Espindola, J. D. Rojas, G. Pinton, and P. A. Dayton (2017). "3-D Ultrasound Localization Microscopy for Identifying Microvascular Morphology Features of Tumor Angiogenesis at a Resolution Beyond the Diffraction Limit of Conventional Ultrasound". In: *Theranostics* 7.1, pp. 196–204. DOI: 10.7150/thno.16899 (cit. on p. 12).

Lockwood, G. R., P.-C. Li, M. O'Donnell, and F. S. Foster (1996). "Optimizing the Radiation Pattern of Sparse Periodic Linear Arrays". In: *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.* 43, pp. 7–14 (cit. on p. 9).

Milecki, L., J. Poée, H. Belgharbi, C. Bourquin, R. Damseh, P. Delafontaine-Martel, F. Lesage, M. Gasse, and J. Provost (2021). "A Deep Learning Framework for Spatiotemporal Ultrasound Localization Microscopy". In: *IEEE Trans. Med. Imag.* early access (cit. on p. 12).

Mor-Avi, V., E. G. Caiani, K. A. Collins, C. E. Korcarz, J. E. Bednarz, and R. M. Lang (2001). "Combined assessment of myocardial perfusion and regional left ventricular function by analysis of contrast-enhanced power modulation images". In: *Circulation* 104.3, pp. 352–357. DOI: 10.1161/01.CIR.104.3.352 (cit. on p. 10).

O'Reilly, M. A. and K. Hynynen (2013). "A super-resolution ultrasound method for brain vascular mapping". In: *Med. Phys.* 40.11, pp. 110701–7. DOI: 10.1118/1.4823762 (cit. on pp. 9, 11).

Ommen, M. L., M. Schou, C. Beers, J. A. Jensen, N. B. Larsen, and E. V. Thomsen (2021). "3D Printed Calibration Micro-Phantoms for Super-Resolution Ultrasound Imaging Validation". In: *Ultrasonics*. Accepted manuscript (cit. on p. 13).

Opacic, T., S. Dencks, B. Theek, M. Piepenbrock, D. Ackermann, A. Rix, T. Lammers, E. Stickeler, S. Delorme, G. Schmitz, and F. Kiessling (2018). "Motion model ultrasound localization microscopy for preclinical and clinical multiparametric tumor characterization". In: *Nat. comm.* 9.1, 1527:1–13. DOI: 10.1038/s41467-018-03973-8 (cit. on p. 12).

Provost, J., C. Papadacci, J. E. Arango, M. Imbault, M. Fink, J. L. Gennisson, M. Tanter, and M. Pernot (2014). "3-D ultrafast ultrasound imaging in vivo". In: *Phys. Med. Biol.* 59.19, pp. L1–L13 (cit. on p. 13).

Rust, M. J., M. Bates, and X. Zhuang (2006). "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)". In: *Nat. Methods* 3.10, pp. 793–795 (cit. on p. 9).

Siepmann, M., G. Schmitz, J. Bzyl, M. Palmowski, and F. Kiessling (2011). "Imaging tumor vascularity by tracing single microbubbles". In: *Proc. IEEE Ultrason. Symp.*, pp. 6293297, 1906–1908. DOI: 10.1109/ULTSYM.2011.0476 (cit. on pp. 11, 12).

Simpson, D. H., C. T. Chin, and P. N. Burns (1999). "Pulse inversion Doppler: a new method for detecting nonlinear echoes from microbubble contrast agents". In: *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.* 46.2, pp. 372–382 (cit. on p. 10).

Solomon, O., R. J. G. van Sloun, H. Wijkstra, M. Mischi, and Y. C. Eldar (2019). "Exploiting flow dynamics for super-resolution in contrast-enhanced ultrasound". In: *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.* 60.10, pp. 1573–1586 (cit. on p. 12).

Song, P., A. Manduca, J. D. Trzasko, R. E. Daigle, and S. Chen (2018). "On the Effects of Spatial Sampling Quantization in Super-Resolution Ultrasound Microvessel Imaging". In: *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.* 65.12, pp. 2264–2276 (cit. on p. 12).

Song, P., J. D. Trzasko, A. Manduca, R. Huang, R. Kadirvel, D. F. Kallmes, and S. Chen (2017). "Improved super-resolution ultrasound microvessel imaging with spatiotemporal nonlocal means filtering and bipartite graph-based microbubble tracking". In: *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.* 65.2, pp. 149–167 (cit. on pp. 11, 12).

Taghavi, I., S. B. Andersen, C. A. V. Hoyos, M. B. Nielsen, C. M. Sørensen, and J. A. Jensen (2021). "In Vivo Motion Correction in Super Resolution Imaging of Rat Kidneys". submitted (cit. on p. 12).

Taghavi, I., M. Schou, S. B. Andersen, C. Hoyos, F. Gran, M. B. Nielsen, C. M. Sørensen, and J. A. Jensen (2020). "In vivo Ultrasound Super-resolution Imaging with Motion correction and Robust Tracking". Submitted (cit. on p. 12).

Tanter, M. and M. Fink (Jan. 2014). "Ultrafast imaging in biomedical ultrasound". In: *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.* 61.1, pp. 102–119. DOI: 10.1109/TUFFC.2014.6689779 (cit. on p. 11).

van Sloun, R. J. G., R. Cohen, and Y. C. Eldar (2020). "Deep Learning in Ultrasound Imaging". In: *IEEE Proc.* 108.1, pp. 11–29 (cit. on p. 13).

van Sloun, R. J. G., O. Solomon, M. Bruce, Z. Z. Khaing, H. Wijkstra, Y. C. Eldar, and M. Mischi (2021). "Super-resolution Ultrasound Localization Microscopy through Deep Learning". In: *IEEE Trans. Med. Imag.* 40.3, pp. 829–839 (cit. on p. 12).

Viessmann, O. M., R. J. Eckersley, K. Christensen-Jeffries, M. X. Tang, and C. Dunsby (2013). "Acoustic super-resolution with ultrasound and microbubbles". In: *Phys. Med. Biol.* 58, pp. 6447–6458 (cit. on pp. 9, 11, 13).

Youn, J., B. Luijten, M. B. Stuart, Y. C. Eldar, R. J. G. van Sloun, and J. A. Jensen (2020). "Deep Learning Models for Fast Ultrasound Localization Microscopy". In: *Proc. IEEE Ultrason. Symp.* Pp. 1–4 (cit. on p. 13).

Youn, J., B. Luijten, M. B. Stuart, Y. C. Eldar, R. J. G. van Sloun, and J. A. Jensen (2021). "Model-based Deep Learning on Ultrasound Channel Data for Fast Ultrasound Localization Microscopy". In: *In preparation* (cit. on p. 13).

Youn, J., M. L. Ommen, M. B. Stuart, E. V. Thomsen, N. B. Larsen, and J. A. Jensen (2020). "Detection and Localization of Ultrasound Scatterers Using Convolutional Neural Networks". In: *IEEE Trans. Med. Imag.* 39.12, pp. 3855–3867 (cit. on p. 12).

Zhu, J., E. M. Rowland, S. Harput, K. Riemer, C. H. Leow, B. Clark, K. Cox, A. Lim, K. Christensen-Jeffries, G. Zhang, J. Brown, C. Dunsby, R. J. Eckersley, P. D. Weinberg, and M.-X. Tang (2019). "3D Super-Resolution US Imaging of Rabbit Lymph Node Vasculature in Vivo by Using Microbubbles". In: *Radiology* 291.3, pp. 642–650. DOI: 10.1148/radiol.2019182593 (cit. on p. 13).

References from Chapter 3

Abadi, M., A. Agarwal, P.Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. G. andI. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, and M. K (2011). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. URL: https://www.tensorflow.org/ (cit. on p. 26).

Badrinarayanan, V., A. Kendall, and R. Cipolla (2017). "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation". In: *V. Badrinarayanan and A. Kendall and R. Cipolla* 39.12, pp. 2481–2495 (cit. on p. 23).

Chang, C., S. Chatterjee, and P. R. Kube (1991). "On an analysis of static occlusion in stereo vision". In: *IEEE Conf. Computer Vision and Pattern Recognition*, pp. 722–723 (cit. on p. 27).

Couture, O., B. Besson, G. Montaldo, M. Fink, and M. Tanter (2011). "Microbubble ultrasound super-localization imaging (MUSLI)". In: *Proc. IEEE Ultrason. Symp.* Pp. 1285–1287 (cit. on pp. 17, 29).

Desailly, Y., O. Couture, M. Fink, and M. Tanter (2013). "Sono-activated ultrasound localization microscopy". In: *Appl. Phys. Lett.* 103.17, p. 174107 (cit. on p. 17).

Drozdzal, M., E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal (2016). "The Importance of Skip Connections in Biomedical Image Segmentation". In: *arXiv:1608.04117v2 [cs.CV]* (cit. on p. 26).

Fua, P. (1993). "A parallel stereo algorithm that produces dense depth maps and preserves image features". In: *Mach. Vis. Appl.* 6.1, pp. 35–49 (cit. on p. 27).

Gomariz, A., W. Li, E. Ozkan, C. Tanner, and O. Goksel (2019). "Siamese Networks With Location Prior for Landmark Tracking in Liver Ultrasound Sequences". In: *Proc. IEEE Int. Symp. Biomed. Imag.* Pp. 1757–1760 (cit. on p. 22).

He, K., X. Zhang, S. Ren, and J. Sun (2016a). "Deep Residual Learning for Image Recognition". In: *IEEE Conf. Computer Vision and Pattern Recognition*, pp. 770–778 (cit. on p. 25).

— (2016b). "Identity Mappings in Deep Residual Networks". In: *Eur. Conf. Computer Vision*, pp. 630–645 (cit. on p. 25).

Ioffe, S. and C. Szegedy (2015). "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift". In: *Int. Conf. Machine Learning*, pp. 448–456 (cit. on p. 25).

Jaccard, P. (1912). "The distribution of the flora in the alpine zone". In: *New phytologist* 11.2, pp. 37–50 (cit. on p. 22).

Jensen, J. A. (2016). "Safety Assessment of Advanced Imaging Sequences, II: Simulations". In: *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.* 63.1, pp. 120–127 (cit. on p. 19).

Kim, J., J. K. Lee, and K. M. Lee (2016). "Accurate Image Super-Resolution Using Very Deep Convolutional Networks". In: *IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1646–1654 (cit. on p. 23).

Kingma, D. and L. Ba (2015). "ADAM: A Method for Stochastic Optimization". In: *arXiv:1412.6980 [cs.LG]* (cit. on p. 26).

Ledig, C., L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, and Z. Wang (2017). "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network". In: *IEEE Conf. Computer Vision and Pattern Recognition*, pp. 105–114 (cit. on p. 23).

Lim, B., S. Son, H. Kim, S. Nah, and K. M. Lee (2017). "Enhanced Deep Residual Networks for Single Image Super-Resolution". In: *IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1132–1140 (cit. on pp. 23, 25).

Lin, F., S. E. Shelton, D. Espindola, J. D. Rojas, G. Pinton, and P. A. Dayton (2017). "3-D Ultrasound Localization Microscopy for Identifying Microvascular Morphology Features of Tumor Angiogenesis at a Resolution Beyond the Diffraction Limit of

Conventional Ultrasound". In: *Theranostics* 7.1, pp. 196–204. DOI: 10.7150/thno.16899 (cit. on p. 22).

Liu, R., J. Lehman, P. Molino, F. P. Such, E. Frank, A. Sergeev, and J. Yosinski (2018). "An intriguing failing of convolutional neural networks and the CoordConv solution". In: *Neural Information Processing Systems*, pp. 9605–9616 (cit. on pp. 24, 26).

Lucy, L. (1974). "An iterative technique for the rectification of observed distributions". In: *Astron. J.* 79, pp. 745–754 (cit. on p. 29).

Maas, A. L., A. Y. Hannun, and A. Y. Ng (2013). "Rectifier Nonlinearities Improve Neural Network Acoustic Models". In: *ICML Workshop on Deep Learning for Audio, Speech, and Language Processing* (cit. on p. 25).

Nair, V. and G. Hinton (2010). "Rectified Linear Units Improve Restricted Boltzmann Machines". In: *Int. Conf. Machine Learning*, pp. 807–814 (cit. on p. 25).

Nehme, E., L. E. Weiss, T. Michaeli, and Y. Shechtman (Apr. 2018). "Deep-STORM: super-resolution single-molecule microscopy by deep learning". In: *Optica* 5.4, pp. 458–464 (cit. on p. 22).

Ommen, M. L., M. Schou, C. Beers, J. A. Jensen, N. B. Larsen, and E. V. Thomsen (2019). "3D Printed Calibration Micro-phantoms for Validation of Super-Resolution Ultrasound Imaging". In: *Proc. IEEE Ultrason. Symp.* Pp. 1–4 (cit. on p. 33).

— (2021). "3D Printed Calibration Micro-Phantoms for Super-Resolution Ultrasound Imaging Validation". In: *Ultrasonics*. Accepted manuscript (cit. on p. 33).

Richardson, W. H. (Jan. 1972). "Bayesian-Based Iterative Method of Image Restoration∗". In: *J. Opt. Soc. Am.* 62.1, pp. 55–59 (cit. on p. 29).

Ronneberger, O., P. Fischer, and T. Brox (2015). "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241 (cit. on pp. 22, 23, 26).

Saxe, A. M., J. L. McClelland, and S. Ganguli (2013). "Exact solutions to the nonlinear dynamics of learning in deep linear neural networks". In: *arXiv:1312.6120v3 [cs.NE]* (cit. on p. 26).

Shi, W., J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang (2016). "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network". In: *IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1874–1883 (cit. on p. 25).

Siepmann, M., G. Schmitz, J. Bzyl, M. Palmowski, and F. Kiessling (2011). "Imaging tumor vascularity by tracing single microbubbles". In: *Proc. IEEE Ultrason. Symp.*, pp. 6293297, 1906–1908. DOI: 10.1109/ULTSYM.2011.0476 (cit. on p. 29).

Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov (2014). "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *J. Mach. Learn. Res.* 15, pp. 1929–1958 (cit. on p. 25).

Tanter, M. and M. Fink (Jan. 2014). "Ultrafast imaging in biomedical ultrasound". In: *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.* 61.1, pp. 102–119. DOI: 10.1109/TUFFC.2014.6689779 (cit. on p. 19).

Tomov, B. G., S. E. Diederichsen, E. V. Thomsen, and J. A. Jensen (2018). "Characterization of medical ultrasound transducers". In: *Proc. IEEE Ultrason. Symp.* Pp. 1–4 (cit. on p. 19).

van Sloun, R. J. G., O. Solomon, M. Bruce, Z. Z. Khaing, H. Wijkstra, Y. C. Eldar, and M. Mischi (2021). "Super-resolution Ultrasound Localization Microscopy through Deep Learning". In: *IEEE Trans. Med. Imag.* 40.3, pp. 829–839 (cit. on pp. 22, 39).

Youn, J., B. Luijten, M. B. Stuart, Y. C. Eldar, R. J. G. van Sloun, and J. A. Jensen (2020). "Deep Learning Models for Fast Ultrasound Localization Microscopy". In: *Proc. IEEE Ultrason. Symp.* Pp. 1–4 (cit. on p. 34).

Youn, J., M. L. Ommen, M. B. Stuart, E. V. Thomsen, N. B. Larsen, and J. A. Jensen (2019). "Ultrasound Multiple Point Target Detection and Localization using Deep Learning". In: *Proc. IEEE Ultrason. Symp.* Pp. 1937–1940 (cit. on pp. 17, 24, 31, 38).

— (2020). "Detection and Localization of Ultrasound Scatterers Using Convolutional Neural Networks". In: *IEEE Trans. Med. Imag.* 39.12, pp. 3855–3867 (cit. on pp. 17, 18, 20, 22, 24, 31, 32, 35, 38).

Yu, J., L. Lavery, and K. Kim (2018). "Super-resolution ultrasound imaging method for microvasculature in vivo with a high temporal accuracy". In: *Scientific reports* 8.1, pp. 1–11 (cit. on p. 30).

Zagoruyko, S. and N. Komodakis (2016). "Wide Residual Networks". In: *arXiv:1605.07146v4 [cs.CV]* (cit. on p. 25).

References from Chapter 4

Abadi, M., A. Agarwal, P.Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. G. andI. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, and M. K (2011). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. URL: `https://www.tensorflow.org/` (cit. on p. 46).

Bar-Zion, A., C. Tremblay-Darveau, O. Solomon, D. Adam, and Y. C. Eldar (2016). "Fast vascular ultrasound imaging with enhanced spatial resolution and background rejection". In: *IEEE Trans. Med. Imag.* 36.1, pp. 169–180 (cit. on p. 62).

Brown, K. G., D. Ghosh, and K. Hoyt (2020). "Deep learning of spatiotemporal filtering for fast super-resolution ultrasound imaging". In: *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.* PP.99, pp. 1–1. DOI: `10.1109/tuffc.2020.2988164` (cit. on p. 41).

Couture, O., V. Hingot, B. Heiles, P. Muleki-Seya, and M. Tanter (2018). "Ultrasound Localization Microscopy and Super-Resolution: A State of the Art". In: *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.* 65.8, pp. 1304–1320. DOI: `10.1109/TUFFC.2018.2850811` (cit. on p. 58).

Duff, I. S. and J. Koster (2001). “On Algorithms For Permuting Large Entries to the Diagonal of a Sparse Matrix”. In: *SIAM J. Matrix Anal. Appl.* 22.4, pp. 973–996 (cit. on p. 50).

He, K., X. Zhang, S. Ren, and J. Sun (2016). “Identity Mappings in Deep Residual Networks”. In: *Eur. Conf. Computer Vision*, pp. 630–645 (cit. on pp. 45, 46).

Jensen, J. A. (1996). “Field: A Program for Simulating Ultrasound Systems”. In: *Med. Biol. Eng. Comp.* 10th Nordic-Baltic Conference on Biomedical Imaging, Vol. 4, Supplement 1, Part 1, pp. 351–353 (cit. on p. 43).

— (2014). “A Multi-threaded Version of Field II”. In: *Proc. IEEE Ultrason. Symp.* Pp. 2229–2232 (cit. on p. 43).

Jensen, J. A., H. Holten-Lund, R. T. Nilsson, M. Hansen, U. D. Larsen, R. P. Domsten, B. G. Tomov, M. B. Stuart, S. I. Nikolov, M. J. Pihl, Y. Du, J. H. Rasmussen, and M. F. Rasmussen (2013). “SARUS: A Synthetic Aperture Real-time Ultrasound System”. In: *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.* 60.9, pp. 1838–1852 (cit. on p. 43).

Jensen, J. A. and N. B. Svendsen (1992). “Calculation of Pressure Fields from Arbitrarily Shaped, Apodized, and Excited Ultrasound Transducers”. In: *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.* 39.2, pp. 262–267 (cit. on p. 43).

Kingma, D. and L. Ba (2015). “ADAM: A Method for Stochastic Optimization”. In: *arXiv:1412.6980 [cs.LG]* (cit. on p. 46).

Liu, L., H. Jiang, P. He, W. Chen, X. Liu, J. Gao, and J. Han (2020). “On the Variance of the Adaptive Learning Rate and Beyond”. In: *arXiv:1908.03265v3 [cs.LG]* (cit. on p. 46).

Liu, X., T. Zhou, M. Lu, Y. Yang, Q. He, and J. Luo (2020). “Deep Learning for Ultrasound Localization Microscopy”. In: *IEEE Trans. Med. Imag.* 39.10, pp. 3064–3078. DOI: 10.1109/tmi.2020.2986781 (cit. on pp. 41, 51, 61).

Milecki, L., J. Poée, H. Belgharbi, C. Bourquin, R. Damseh, P. Delafontaine-Martel, F. Lesage, M. Gasse, and J. Provost (2021). “A Deep Learning Framework for Spatiotemporal Ultrasound Localization Microscopy”. In: *IEEE Trans. Med. Imag.* early access (cit. on pp. 41, 62).

Mor-Avi, V., E. G. Caiani, K. A. Collins, C. E. Korcarz, J. E. Bednarz, and R. M. Lang (2001). “Combined assessment of myocardial perfusion and regional left ventricular function by analysis of contrast-enhanced power modulation images”. In: *Circulation* 104.3, pp. 352–357. DOI: 10.1161/01.CIR.104.3.352 (cit. on p. 43).

Ommen, M. L., M. Schou, C. Beers, J. A. Jensen, N. B. Larsen, and E. V. Thomsen (2019). “3D Printed Calibration Micro-phantoms for Validation of Super-Resolution Ultrasound Imaging”. In: *Proc. IEEE Ultrason. Symp.* Pp. 1–4 (cit. on p. 53).

— (2021). “3D Printed Calibration Micro-Phantoms for Super-Resolution Ultrasound Imaging Validation”. In: *Ultrasonics*. Accepted manuscript (cit. on p. 53).

Ronneberger, O., P. Fischer, and T. Brox (2015). “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241 (cit. on pp. 45, 46).

Saxe, A. M., J. L. McClelland, and S. Ganguli (2013). “Exact solutions to the nonlinear dynamics of learning in deep linear neural networks”. In: *arXiv:1312.6120v3 [cs.NE]* (cit. on p. 46).

Schneider, M. (1999). “Characteristics of SonoVue”. In: *Echocardiography* 16.1, pp. 743–746 (cit. on p. 44).

Taghavi, I., S. B. Andersen, C. A. V. Hoyos, M. B. Nielsen, C. M. Sørensen, and J. A. Jensen (2021). “In Vivo Motion Correction in Super Resolution Imaging of Rat Kidneys”. submitted (cit. on p. 57).

Taghavi, I., S. B. Andersen, C. A. V. Hoyos, M. Schou, S. H. Øygard, F. Gran, K. L. Hansen, C. M. Sørensen, M. B. Nielsen, M. B. Stuart, and J. A. Jensen (2020). “Tracking Performance in Ultrasound Super-Resolution Imaging”. In: *Proc. IEEE Ultrason. Symp.* Pp. 1–4 (cit. on p. 57).

Tang, S., P. Song, J. D. Trzasko, M. Lowerison, C. Huang, P. Gong, U. Lok, A. Manduca, and S. Chen (2020). “Kalman Filter–Based Microbubble Tracking for Robust Super-Resolution Ultrasound Microvessel Imaging”. In: *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.* 67.9, pp. 1738–1751 (cit. on p. 57).

Thurstone, F. L. and O. T. von Ramm (1974). “A new ultrasound imaging technique employing two-dimensional electronic beam steering”. In: *Acoustical Holography*. Ed. by P. S. Green. Vol. 5. New York: Plenum Press, pp. 249–259 (cit. on p. 44).

van Sloun, R. J. G., O. Solomon, M. Bruce, Z. Z. Khaing, H. Wijkstra, Y. C. Eldar, and M. Mischi (2021). “Super-resolution Ultrasound Localization Microscopy through Deep Learning”. In: *IEEE Trans. Med. Imag.* 40.3, pp. 829–839 (cit. on pp. 41, 45, 61).

Youn, J., M. L. Ommen, M. B. Stuart, E. V. Thomsen, N. B. Larsen, and J. A. Jensen (2019). “Ultrasound Multiple Point Target Detection and Localization using Deep Learning”. In: *Proc. IEEE Ultrason. Symp.* Pp. 1937–1940 (cit. on pp. 45–47).

— (2020). “Detection and Localization of Ultrasound Scatterers Using Convolutional Neural Networks”. In: *IEEE Trans. Med. Imag.* 39.12, pp. 3855–3867 (cit. on pp. 45–47, 50).

Youn, J., I. Taghavi, M. L. Ommen, M. Schou, M. B. Stuart, E. V. Thomsen, N. B. Larsen, and J. A. Jensen (2021). “Sub-pixel Accuracy Microbubble Localization using Convolutional Neural Networks”. In: *In preparation* (cit. on pp. 41, 42, 46, 52–57, 59–61, 63).

Zhang, M., J. Lucas, J. Ba, and G. E. Hinton (2019). “Lookahead Optimizer: k steps forward, 1 step back”. In: *Neural Information Processing Systems*. Vol. 32, pp. 9597–9608 (cit. on p. 46).

References from Chapter 5

Badrinarayanan, V., A. Kendall, and R. Cipolla (2017). "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation". In: *V. Badrinarayanan and A. Kendall and R. Cipolla* 39.12, pp. 2481–2495 (cit. on p. 70).

Beck, A. and M. Teboulle (2009). "A fast iterative shrinkage-thresholding algorithm for linear inverse problems". In: *SIAM J. Imaging Sci.* 2.1, pp. 183–202 (cit. on p. 68).

Dardikman-Yoffe, G. and Y. C. Eldar (2020). "Learned SPARCOM: Unfolded Deep Super-Resolution Microscopy". In: *Opt. Express* 28.19, pp. 27736–27763 (cit. on p. 74).

Eldar, Y. C. (2015). *Sampling Theory: Beyond Bandlimited Systems*. Cambridge University Press (cit. on p. 67).

Gregor, K. and Y. LeCun (2010). "Learning fast approximations of sparse coding". In: *Int. Conf. Machine Learning*, pp. 399–406 (cit. on p. 69).

Isola, P., J. Zhu, T. Zhou, and A. A. Efros (2016). "Image-to-Image Translation with Conditional Adversarial Networks". In: *arXiv:1611.07004v3 [cs.CV]* (cit. on p. 70).

Jensen, J. A. (1996). "Field: A Program for Simulating Ultrasound Systems". In: *Med. Biol. Eng. Comp.* 10th Nordic-Baltic Conference on Biomedical Imaging, Vol. 4, Supplement 1, Part 1, pp. 351–353 (cit. on p. 71).

— (2014). "A Multi-threaded Version of Field II". In: *Proc. IEEE Ultrason. Symp.* Pp. 2229–2232 (cit. on p. 71).

Jensen, J. A. and N. B. Svendsen (1992). "Calculation of Pressure Fields from Arbitrarily Shaped, Apodized, and Excited Ultrasound Transducers". In: *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.* 39.2, pp. 262–267 (cit. on p. 71).

Kingma, D. and L. Ba (2015). "ADAM: A Method for Stochastic Optimization". In: *arXiv:1412.6980 [cs.LG]* (cit. on p. 69).

Monga, V., Y. Li, and Y. C. Eldar (2020). "Algorithm Unrolling: Interpretable, Efficient Deep Learning for Signal and Image Processing". In: *arXiv:1912.10557v3 [eess.IV]* (cit. on pp. 67, 69, 74).

Ronneberger, O., P. Fischer, and T. Brox (2015). "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241 (cit. on p. 70).

Tanter, M. and M. Fink (Jan. 2014). "Ultrafast imaging in biomedical ultrasound". In: *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.* 61.1, pp. 102–119. DOI: 10.1109/TUFFC.2014.6689779 (cit. on p. 71).

van Sloun, R. J. G., R. Cohen, and Y. C. Eldar (2020). "Deep Learning in Ultrasound Imaging". In: *IEEE Proc.* 108.1, pp. 11–29 (cit. on p. 67).

Youn, J., B. Luijten, M. B. Stuart, Y. C. Eldar, R. J. G. van Sloun, and J. A. Jensen (2020). "Deep Learning Models for Fast Ultrasound Localization Microscopy". In: *Proc. IEEE Ultrason. Symp.* Pp. 1–4 (cit. on pp. 67, 69, 70, 72, 73).

Youn, J., M. L. Ommen, M. B. Stuart, E. V. Thomsen, N. B. Larsen, and J. A. Jensen (2020). "Detection and Localization of Ultrasound Scatterers Using Convolutional Neural Networks". In: *IEEE Trans. Med. Imag.* 39.12, pp. 3855–3867 (cit. on p. 70).

References from Chapter 6

Brown, K. G., D. Ghosh, and K. Hoyt (2020). "Deep learning of spatiotemporal filtering for fast super-resolution ultrasound imaging". In: *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.* PP.99, pp. 1–1. DOI: 10.1109/tuffc.2020.2988164 (cit. on p. 77).

Jensen, J. A. (1996). "Field: A Program for Simulating Ultrasound Systems". In: *Med. Biol. Eng. Comp.* 10th Nordic-Baltic Conference on Biomedical Imaging, Vol. 4, Supplement 1, Part 1, pp. 351–353 (cit. on p. 77).

— (2014). "A Multi-threaded Version of Field II". In: *Proc. IEEE Ultrason. Symp.* Pp. 2229–2232 (cit. on p. 77).

Jensen, J. A. and N. B. Svendsen (1992). "Calculation of Pressure Fields from Arbitrarily Shaped, Apodized, and Excited Ultrasound Transducers". In: *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.* 39.2, pp. 262–267 (cit. on p. 77).

Liu, X., T. Zhou, M. Lu, Y. Yang, Q. He, and J. Luo (2020). "Deep Learning for Ultrasound Localization Microscopy". In: *IEEE Trans. Med. Imag.* 39.10, pp. 3064–3078. DOI: 10.1109/tmi.2020.2986781 (cit. on p. 77).

Luijten, B., R. Cohen, F. J. De Bruijn, H. A. W. Schmeitz, M. Mischi, Y. C. Eldar, and R. J. G. Van Sloun (2020). "Adaptive Ultrasound Beamforming using Deep Learning". In: *IEEE Trans. Med. Imag.* 39.12, pp. 3967–3978 (cit. on pp. 77–80).

Milecki, L., J. Poée, H. Belgharbi, C. Bourquin, R. Damseh, P. Delafontaine-Martel, F. Lesage, M. Gasse, and J. Provost (2021). "A Deep Learning Framework for Spatiotemporal Ultrasound Localization Microscopy". In: *IEEE Trans. Med. Imag.* early access (cit. on p. 77).

Monga, V., Y. Li, and Y. C. Eldar (2020). "Algorithm Unrolling: Interpretable, Efficient Deep Learning for Signal and Image Processing". In: *arXiv:1912.10557v3 [eess.IV]* (cit. on p. 80).

van Sloun, R. J. G., R. Cohen, and Y. C. Eldar (2020). "Deep Learning in Ultrasound Imaging". In: *IEEE Proc.* 108.1, pp. 11–29 (cit. on pp. 77–79).

van Sloun, R. J. G., O. Solomon, M. Bruce, Z. Z. Khaing, H. Wijkstra, Y. C. Eldar, and M. Mischi (2021). "Super-resolution Ultrasound Localization Microscopy through Deep Learning". In: *IEEE Trans. Med. Imag.* 40.3, pp. 829–839 (cit. on pp. 77, 80).

Youn, J., B. Luijten, M. B. Stuart, Y. C. Eldar, R. J. G. van Sloun, and J. A. Jensen (2020). "Deep Learning Models for Fast Ultrasound Localization Microscopy". In: *Proc. IEEE Ultrason. Symp.* Pp. 1–4 (cit. on p. 79).

— (2021). "Model-based Deep Learning on Ultrasound Channel Data for Fast Ultrasound Localization Microscopy". In: *In preparation* (cit. on pp. 77, 79, 81).

Youn, J., I. Taghavi, M. L. Ommen, M. Schou, M. B. Stuart, E. V. Thomsen, N. B. Larsen, and J. A. Jensen (2021). “Sub-pixel Accuracy Microbubble Localization using Convolutional Neural Networks”. In: *In preparation* (cit. on p. 77).

Paper 1

Ultrasound Multiple Point Target Detection and Localization using Deep Learning

Jihwan Youn, Martin Lind Ommen, Matthias Bo Stuart, Erik Vilain Thomsen, Niels Bent Larsen, Jørgen Arendt Jensen

Published in:
Proceedings of the IEEE International Ultrasonic Symposium

Document Version:
Published

DOI:
10.1109/ULTSYM.2019.8925885

Ultrasound Multiple Point Target Detection and Localization using Deep Learning

Jihwan Youn, Martin Lind Ommen, Matthias Bo Stuart, Erik Vilain Thomsen,
Niels Bent Larsen, Jørgen Arendt Jensen

Department of Health Technology, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

***Abstract*—Super-resolution imaging (SRI) can achieve sub-wavelength resolution by detecting and tracking intravenously injected microbubbles (MBs) over time. However, current SRI is limited by long data acquisition times since the MB detection still relies on diffraction-limited conventional ultrasound images. This limits the number of detectable MBs in a fixed time duration. In this work, we propose a deep learning-based method for detecting and localizing high-density multiple point targets from radio frequency (RF) channel data. A Convolutional Neural Network (CNN) was trained to return confidence maps given RF channel data, and the positions of point targets were estimated from the confidence maps. RF channel data for training and evaluation were simulated in Field II by placing point targets randomly in the region of interest and transmitting three steered plane waves. The trained CNN achieved a precision and recall of 0.999 and 0.960 on a simulated test dataset. The localization errors after excluding outliers were within $\pm\,46\,\mu$m and $\pm\,27\,\mu$m in the lateral and axial directions. A scatterer phantom was 3-D printed and imaged by the Synthetic Aperture Real-time Ultrasound System (SARUS). On measured data, a precision and recall of 0.976 and 0.998 were achieved, and the localization errors after excluding outliers were within $\pm\,101\,\mu$m and $\pm\,75\,\mu$m in the lateral and axial directions. We expect that this method can be extended to highly concentrated microbubble (MB) detection in order to accelerate SRI.**

I. Introduction

Super-resolution imaging (SRI), often referred to as ultrasound localization microscopy (ULM), has demonstrated that it is possible to surpass the diffraction limit of conventional ultrasound imaging. Microvessels laying closer than a half-wavelength apart have been resolved by deploying microbubbles (MBs) as a contrast agent and using SRI [1]–[5]. The centroids of individual MBs can be easily found as MB echoes are much stronger than surrounding tissues when insonified, and their sizes are much smaller than a wavelength. Sub-wavelength imaging is achieved by accumulating the detected MB positions over time, revealing the fine structure of the microvasculature.

The MB detection in SRI, however, is still diffraction-limited because it is performed in conventional ultrasound images which are commonly formed by delay-and-sum (DAS) beamforming [6]. For accurate and reliable detection and localization, the MBs need to be more than a wavelength apart to avoid the overlaps of MB point spread functions (PSFs). Diluted concentrations of MBs are commonly used to satisfy this criteria as the behavior of MBs is hard to control. The number of detectable MBs, therefore, is constrained and this leads to very long data acquisition times in order to map the entire microvasculature.

In this work, we propose a deep learning-based method for detecting and localizing multiple ultrasound point targets. The method especially aims to identify high-density point targets whose PSFs are overlapping, by feeding radio frequency (RF) channel data directly as input. A fully convolutional neural network (CNN) was designed to return 2-D confidence maps given RF channel data. The pixel values of the confidence maps correspond to the confidence of point targets existing in the pixels. The point target positions were extracted from the confidence maps by identifying local maxima. The CNN was trained and evaluated using simulated RF channel data. To further investigate the method on measured data, a phantom experiment was performed using a 3-D printed PEGDA $700\,\mathrm{g/mol}$ hydrogel phantom [7].

II. Method

A. Simulated Dataset

1) RF channel data: The Field II ultrasound simulation program [8], [9] was used to simulate RF channel data for generating a training and a test datasets. The datasets were composed of a certain number of frames. One frame was created by transmitting three steered plane waves after placing 100 point targets randomly within a region of $6.4 \times 6.4\,\mathrm{mm}^2$ (an average target density of $2.44\,\mathrm{mm}^{-2}$) where the center was $18\,\mathrm{mm}$ away from a transducer. The transducer was modeled after a commercial 192-element linear array, and the measured impulse response [10], [11] was applied to make the RF data as close to real measured data as possible. The parameters used in simulation are listed in Table I.

The simulated raw RF data were not beamformed but delayed, based on the time-of-flight calculated by

$$\tau_i(x,z) = \left(\sqrt{(x-x_i)^2 + z^2} + z\right)/c \qquad (1)$$

where τ_i is the time-of-flight of the i-th transmission, (x, z) is the point, x_i is the center of the i-th transmission aperture, and c is the speed of sound. The delayed RF data were then sampled to have the same number of samples as that of confidence maps along the axial direction. The size of resulting RF data for one frame was $256 \times 64 \times 3$.

TABLE I
RF CHANNEL DATA SIMULATION PARAMETERS

Category	Parameter	Value
Transducer	Center frequency	$5.2\,\mathrm{MHz}$
	Pitch	$0.20\,\mathrm{mm}$
	Element width	$0.18\,\mathrm{mm}$
	Element height	$6\,\mathrm{mm}$
	Number of elements	192
Imaging	Number of TX elements	32
	Number of RX elements	64
	Steering angles	$-15°$, $0°$, $15°$
Environment	Speed of sound	$1480\,\mathrm{m/s}$
	Field II sampling frequency	$120\,\mathrm{MHz}$
	RF data sampling frequency	$29.6\,\mathrm{MHz}$
Scatterer	Number of scatterers	100
	Lateral position range	$(-3.2, 3.2)\,\mathrm{mm}$
	Axial position range	$(14.8, 21.2)\,\mathrm{mm}$

2) Confidence Map: Non-overlapping Gaussian confidence maps were used as labels for training CNNs. Initially, binary confidence maps were created, where pixel values of one indicated a point target and the remaining pixel values were zero. A 21×21 Gaussian filter with a standard deviation of six was then applied at each point target position in the binary confidence maps. The filter values from the targets will be overlapped when some targets are closer than a half of the filter size in the confidence maps. In that case, the maximum value at each pixel location was taken. This maintained local maxima at target positions as opposed to the overlapping PSFs of DAS beamforming, and enabled the CNN to resolve targets closer than the diffraction limit.

The pixel size of the confidence maps was set to $25\,\mu\mathrm{m}$, and the image size of them became 256×256, given the pixel size and the region of interest.

B. Convolutional Neural Network

1) Network Architecture: The proposed CNN is adapted from U-Net [12] which has an encoder-decoder structure. The feature maps are downsampled while the number of feature maps increases in the encoding path. Then, the feature maps are upsampled to their original size while the number of feature maps decreases in the decoding path. U-Net has a large receptive field, an effective input size that is covered by a convolution operation in an unit, for the sake of this structure. This is beneficial because a partial view of RF data is not enough to determine point target existence.

A detailed CNN architecture is illustrated in Fig. 1. Convolution and rectified linear unit (ReLU) layers in U-Net were replaced with pre-activation residual units (Fig. 1a) [13]. The pre-activation residual units ease optimization problem by introducing shortcuts, thereby improving performance. The proposed CNN (Fig. 1e) mainly consisted of four *down-blocks* (Fig. 1b), one *conv-block* (Fig. 1c), and four *up-blocks* (Fig. 1d). The skip-connections in U-Net was removed since it hindered the training. Instead, CoordConv [14] was added to transfer spatial information over convolution layers. Dropout [15] was attached after the shortcut in residual blocks for regularization. For pooling and unpooling, strided convolution and pixel shuffle [16] were chosen, respectively. Leaky rectified linear units (Leaky ReLU) [17] were applied as non-linear activation to avoid dying ReLU problem causing nonactivated units.

2) Training Details: The CNN was trained by minimizing the mean squared error (MSE) between true confidence maps and CNN outputs. The training dataset consisted of a total of $10,240$ frames. The kernel weights were initialized with orthogonal initialization [18] and optimized with ADAM [19] by setting $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-7}$. The initial learning rate was 10^{-4} and it was halved at every 100 epoch while limiting the minimum learning rate to 10^{-6}. The number of epochs was 600 and the mini-batch size was 32.

C. 3-D Printed Scatterer Phantom

A PEGDA $700\,\mathrm{g/mol}$ hydrogel scatterer phantom [7] was 3-D printed to investigate the proposed method on measured data. The phantom contained water-filled cavities which acted as scatterers. A total of 100 scatterers were placed on a 10×10 grid with a spacing of $518\,\mu\mathrm{m}$ in the lateral direction and $342\,\mu\mathrm{m}$ in the axial direction, as illustrated in Fig. 2.

The 3-D printed phantom was scanned by the Synthetic Aperture Real-time Ultrasound System (SARUS) [20] to acquire RF channel data. The same imaging scheme and transducer described in Table I were used. The phantom was placed on a motion stage and scanned at different positions by moving the motion stage at a step of $50\,\mu\mathrm{m}$ in the lateral direction. A total of 33 frames were obtained.

III. RESULTS

A. Simulation Experiment

The trained CNN was initially evaluated on a simulated test dataset. It was simulated in the same way as the training dataset in Field II, and consisted of 3,840 frames. In Fig. 3, the result of applying the CNN method to a test frame is compared with simply using the conventional DAS beamforming on the same frame. The CNN method was able to identify highly concentrated point targets while the DAS beamforming failed due to the overlapping PSFs. Full width at half maximum (FWHM) of the DAS beamforming at a depth of 18 mm was $387\,\mu\mathrm{m}$ ($1.36\ \lambda$) in the lateral direction and $140\,\mu\mathrm{m}$ ($0.49\ \lambda$) in the axial direction.

The CNN's capability to detect and localize point targets were quantitatively evaluated. Detection was measured by precision and recall that are defined by

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

where TP is the number of true positives, FP is the number of false positives, and FN is the number of false negatives. The positive and negative detections were determined by comparing estimated target positions with true target positions based on their pair-wise distances. The CNN method achieved

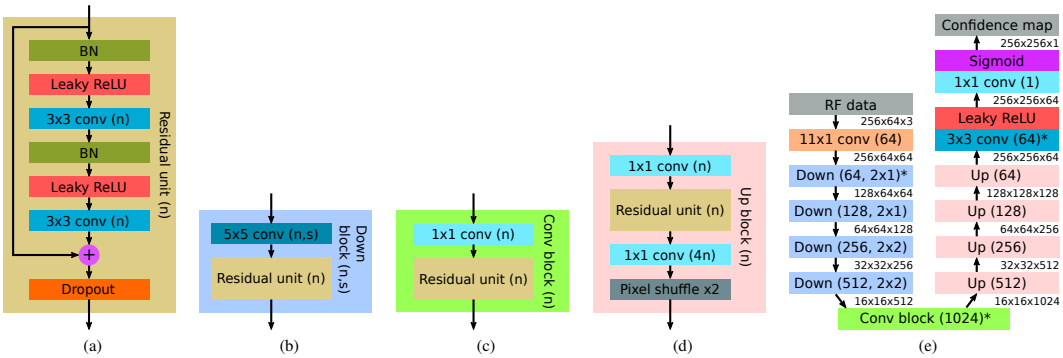


Fig. 1. The proposed CNN architecture and its components. (a) residual unit, (b) down-block, (c) conv-block, (d) up-block, and (e) the network overview. n and s in the parenthesis are the number of kernels and stride. The asterisk in (e) indicates that its first convolution in the block is CoordConv. The three numbers between blocks in (e) represent feature map size in the order of height, width, and the number of feature maps.

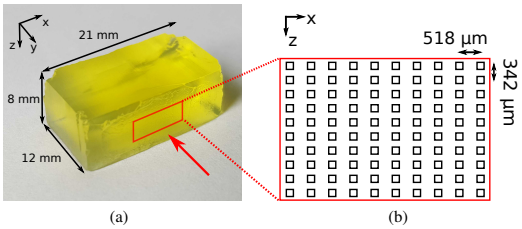


Fig. 2. Fabricated 3-D scatterer phantom: (a) photograph of the phantom and (b) 100 scatterers placed in a 10×10 grid.

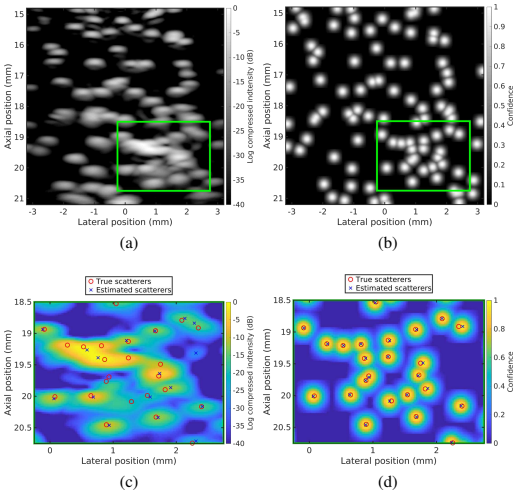


Fig. 3. Comparison of point target detection between DAS beamforming and CNN on a simulated test data using three steered plane wave transmissions. (a) DAS beamformed B-mode image, (b) confidence map returned from CNN, (c) true and estimated scatterer positions in the green square region of (a), and (d) true and estimated scatterer positions in the green square region of (b)

a precision and recall of 0.999 and 0.960, while DAS beamforming achieved a precision and recall of 0.986 and 0.756.

Localization uncertainties in the lateral and axial position were calculated using the positive detections, and is illustrated using a box-and-whisker plot in Fig. 4a. The bottom and top edges of the blue box indicate the 25th (q_1) and 75th percentiles (q_3) and the center red edge indicates the median. The vertically extended line from the box (whisker) indicates the range of inliers which are smaller than $q_3 + 1.5 \times (q_3 - q_1)$ and greater than $q_1 - 1.5 \times (q_3 - q_1)$. The inliers were within $\pm 46\,\mu\text{m}$ (0.16λ) in the lateral direction and $\pm 27\,\mu\text{m}$ (0.09λ) in the axial direction.

B. Phantom Experiment

The CNN trained for the simulation experiment was not effective on the measured data because the scatterers in the phantom are not infinitesimally small point targets. The ultrasound beam is actually scattered twice at each scatterer in the phantom. Therefore, the RF data in the training dataset were simulated a second time by modeling a target using two points. In addition, the first scattering was phase reversed since the acoustic impedance is higher in the phantom than in the water inside the targets.

A new CNN was trained using the modified training dataset, and it successfully identified scatterers from the measured data as shown in Fig. 5. The achieved precision and recall were 0.976 and 0.998. The inliers were within $\pm 101\,\mu\text{m}$ (0.33λ) in the lateral direction and $\pm 75\,\mu\text{m}$ (0.25λ) in the axial direction, as illustrated in Fig. 4b.

IV. CONCLUSION

A CNN-based ultrasound multiple point target detection and localization method was demonstrated. The CNN was trained

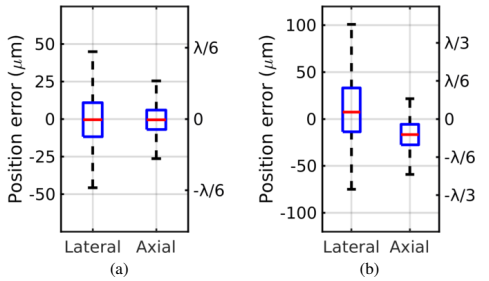


Fig. 4. Localization uncertainty in the lateral and axial direction measured (a) on the simulated test dataset and (b) on the measured phantom data.

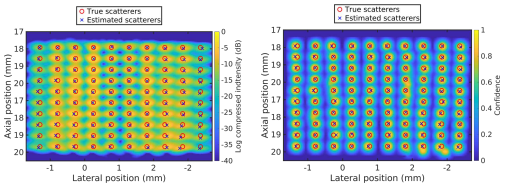


Fig. 5. Comparison of scatterer detection between DAS beamforming and CNN on phantom data using three steered plane wave transmissions. (a) DAS beamformed B-mode image and (b) confidence map returned from CNN with true and estimated scatterer positions

to learn a mapping from RF channel data to non-overlapping Gaussian confidence maps, and point target positions were estimated from the confidence maps by identifying local maxima. The non-overlapping Gaussian confidence maps were introduced to relax the sparsity of binary confidence maps while maintaining local maxima as target positions. The CNN method resolved point targets closer than the diffraction limit, whereas DAS beamforming failed as shown in Fig. 3.

It is also shown that the CNN method is applicable to real-world data, as well as simulated data, through the phantom experiment. It is notable that the training was performed solely using simulated data because it is nearly impossible to obtain a large number of measurements with ground truth for these kinds of work. It was also imperative to employ the measured impulse response and model targets following realistic physical modeling in the simulation.

We expect that this method can be extended to MB detection and potentially shorten the data acquisition time of SRI by detecting a greater number of MBs in a shorter amount of time.

Acknowledgment

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan V Volta GPU used for this research.

References

[1] O. Couture, B. Besson, G. Montaldo, M. Fink, and M. Tanter, "Microbubble ultrasound super-localization imaging (MUSLI)," in *Proc. IEEE Ultrason. Symp.*, 2011, pp. 1285–1287.

[2] O. M. Viessmann, R. J. Eckersley, K. C. Jeffries, M. X. Tang, and C. Dunsby, "Acoustic super-resolution with ultrasound and microbubbles," *Phys. Med. Biol.*, vol. 58, pp. 6447–6458, 2013.

[3] M. A. O'Reilly and K. Hynynen, "A super-resolution ultrasound method for brain vascular mapping," *Med. Phys.*, vol. 40, no. 11, pp. 110 701–7, 2013.

[4] C. Errico, J. Pierre, S. Pezet, Y. Desailly, Z. Lenkei, O. Couture, and M. Tanter, "Ultrafast ultrasound localization microscopy for deep super-resolution vascular imaging," *Nature*, vol. 527, pp. 499–502, November 2015.

[5] K. Christensen-Jeffries, R. J. Browning, M. Tang, C. Dunsby, and R. J. Eckersley, "In vivo acoustic super-resolution and super-resolved velocity mapping using microbubbles," *IEEE Trans. Med. Imag.*, vol. 34, no. 2, pp. 433–440, February 2015.

[6] F. L. Thurstone and O. T. von Ramm, "A new ultrasound imaging technique employing two-dimensional electronic beam steering," in *Acoustical Holography*, P. S. Green, Ed., vol. 5. New York: Plenum Press, 1974, pp. 249–259.

[7] M. L. Ommen, M. Schou, R. Zhang, C. A. V. Hoyos, J. A. Jensen, N. B. Larsen, and E. V. Thomsen, "3D printed flow phantoms with fiducial markers for super-resolution ultrasound imaging," in *Proc. IEEE Ultrason. Symp.*, 2018, pp. 1–4.

[8] J. A. Jensen and N. B. Svendsen, "Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 39, no. 2, pp. 262–267, 1992.

[9] J. A. Jensen, "Field: A program for simulating ultrasound systems," *Med. Biol. Eng. Comp.*, vol. 10th Nordic-Baltic Conference on Biomedical Imaging, Vol. 4, Supplement 1, Part 1, pp. 351–353, 1996.

[10] ——, "Safety assessment of advanced imaging sequences, II: Simulations," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 63, no. 1, pp. 120–127, 2016.

[11] B. G. Tomov, S. E. Diederichsen, E. V. Thomsen, and J. A. Jensen, "Characterization of medical ultrasound transducers," in *Proc. IEEE Ultrason. Symp.*, 2018, pp. 1–4.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Eur. Conf. Computer Vision*, 2016, pp. 630–645.

[14] R. Liu, J. Lehman, P. Molino, F. P. Such, E. Frank, A. Sergeev, and J. Yosinski, "An intriguing failing of convolutional neural networks and the coordconv solution," in *Neural Information Processing Systems*, 2018, pp. 9605–9616.

[15] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014.

[16] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2016, pp. 1874–1883.

[17] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *ICML Workshop on Deep Learning for Audio, Speech, and Language Processing*, 2013.

[18] A. M. Saxe, J. L. McClelland, and S. Ganguli, "Exact solutions to the nonlinear dynamics of learning indeep linear neural networks," `arXiv:1312.6120v3 [cs.NE]`, 2013.

[19] D. Kingma and L. Ba, "Adam: A method for stochastic optimization," `arXiv:1412.6980 [cs.LG]`, 2015.

[20] J. A. Jensen, H. Holten-Lund, R. T. Nilsson, M. Hansen, U. D. Larsen, R. P. Domsten, B. G. Tomov, M. B. Stuart, S. I. Nikolov, M. J. Pihl, Y. Du, J. H. Rasmussen, and M. F. Rasmussen, "SARUS: A synthetic aperture real-time ultrasound system," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 60, no. 9, pp. 1838–1852, 2013.

Paper 2

Detection and Localization of Ultrasound Scatterers Using Convolutional Neural Networks

Jihwan Youn, Martin Lind Ommen, Matthias Bo Stuart, Erik Vilain Thomsen, Niels Bent Larsen, Jørgen Arendt Jensen

Name of journal in:
IEEE Transactions on Medical Imaging

Document Version:
Published

DOI:
10.1109/TMI.2020.3006445

Detection and Localization of Ultrasound Scatterers Using Convolutional Neural Networks

Jihwan Youn, Martin Lind Ommen, Matthias Bo Stuart, Erik Vilain Thomsen, Niels Bent Larsen, Jørgen Arendt Jensen, *Fellow, IEEE*

***Abstract*—Delay-and-sum (DAS) beamforming is unable to identify individual scatterers when their density is so high that their point spread functions overlap each other. This paper proposes a convolutional neural network (CNN)-based method to detect and localize high-density scatterers, some of which are closer than the resolution limit of DAS beamforming. A CNN was designed to take radio frequency channel data and return non-overlapping Gaussian confidence maps. The scatterer positions were estimated from the confidence maps by identifying local maxima. On simulated test sets, the CNN method with three plane waves achieved a precision of 1.00 and a recall of 0.91. Localization uncertainties after excluding outliers were $\pm 46\,\mu m$ (outlier ratio: 4%) laterally and $\pm 26\,\mu m$ (outlier ratio: 1%) axially. To evaluate the proposed method on measured data, two phantoms containing cavities were 3-D printed and imaged. For phantom study, training data were modified according to the physical properties of the phantoms and a new CNN was trained. On an uniformly spaced scatterer phantom, a precision of 0.98 and a recall of 1.00 were achieved with the localization uncertainties of $\pm 101\,\mu m$ (outlier ratio: 1%) laterally and $\pm 37\,\mu m$ (outlier ratio: 1%) axially. On a randomly spaced scatterer phantom, a precision of 0.59 and a recall of 0.63 were achieved. The localization uncertainties were $\pm 132\,\mu m$ (outlier ratio: 0%) laterally and $\pm 44\,\mu m$ with a bias of $22\,\mu m$ (outlier ratio: 0%) axially. This method can potentially be extended to detect highly concentrated microbubbles in order to shorten data acquisition times of super-resolution ultrasound imaging.**

***Index Terms*—high-density scatterers, convolutional neural network, super-resolution ultrasound imaging, ultrasound localization microscopy**

I. Introduction

DELAY-AND-SUM (DAS) beamforming [1] is simple and effective for B-mode image generation, but the spatial resolution is limited by wave diffraction. The resolution of conventional ultrasound imaging depends on wavelength, f-number, and excitation pulse bandwidth. Recently, ultrasound localization microscopy (ULM) and the resulting super-resolution ultrasound imaging (SRUS) were devised to overcome the diffraction limit [2]–[6]. The microvasculature, composed of vessels that are separated by less than a half-wavelength, was mapped by deploying microbubbles (MBs) as contrast agents. SRUS can be achieved by detecting and tracking the centroids of individual MBs over time.

ULM-based SRUS, however, requires long data acquisition times since the MB detection still relies on conventional ultrasound images. The ultrasound images are generally DAS beamformed and diffraction-limited as a consequence. Therefore, the MB concentration should be low to avoid overlapping point spread functions (PSFs) for accurate and reliable MB detection and localization. This constrains the number of detectable MBs in a frame, and it leads to long data acquisition times for mapping the entire target structure.

A novel method is proposed in this paper to detect and localize high-density scatterers by using convolutional neural networks (CNNs). Deep learning has had a profound impact on processing complex data and making associated decisions. By training deep neural networks with a large number of examples, impressive improvements were achieved in various challenging problems such as image classification [7]–[10], object detection [11], [12], semantic segmentation [13]–[15], and single-image super-resolution [16], [17]. It would be nearly impossible to attain such improvements using traditional logic programming or model-based approaches. The same principles can be applicable to ultrasound signals. It is hypothesized that a data-driven CNN-based method can identify scatterers laying closer than the resolution limit of DAS beamforming directly from radio frequency (RF) channel data.

In optics, where localization microscopy was firstly proposed [18]–[20], several studies have been conducted to incorporate deep learning in super-resolution localization microscopy [21]–[23]. These studies used CNNs to localize fluorescent molecules and showed that deep learning-based methods can drastically reduce data acquisition times and data processing times while achieving state-of-the-art performance.

Similar attempts also exist in SRUS. Van Sloun *et al.* [24] proposed Deep-ULM that outputs high-resolution images where the pixel values correspond to scattering intensities, given image patches of contrast-enhanced ultrasound (CEUS) acquisitions. This is similar to our approach in the sense that it handles high-density scatterer detection using CNNs, but Deep-ULM takes beamformed signals as input, whereas the proposed method only uses RF channel data without beamforming. Allman *et al.* [25] tried to locate and classify sources and artifacts from pre-beamformed photoacoustic channel data using Faster R-CNN [26] with VGG16 [27]. However, only up to 10 sources were considered, and classification for artifact removal is not necessary for scatterer detection.

Deep learning techniques have been applied to achieve better ultrasound image quality. A fully connected neural network beamformer improved image contrast by suppressing off-axis scattering [28]. Hyun *et al.* [29] proposed a CNN beamformer that reduces speckle and eventually enhances contrast while

This work was supported in part by the Fondation Idella.

The authors are with the Department of Health Technology, Technical University of Denmark, 2800 Lyngby, Denmark (email: jihyoun@dtu.dk).

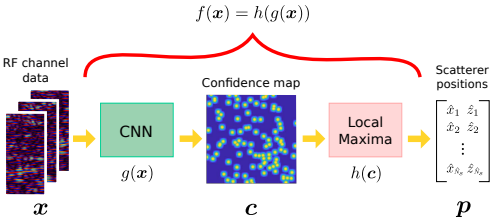


Fig. 1. Overview of the proposed scatterer detection and localization method.

preserving resolution. Generative Adversarial Network (GAN) [30], an architecture that generates output following the same distribution as training data, were applied to improve image quality without sacrificing frame rate. Multi-focus line-by-line images were synthesized from single-focus line-by-line images [31] and image quality comparable to using thirty one plane waves was achieved using three plane waves [32].

In this work, CNNs were trained to learn a mapping from RF channel data to confidence maps, and scatterer positions were then estimated from the confidence maps by identifying local maxima. The RF channel data were directly fed to the CNNs without beamforming to avoid the information loss caused by overlapping PSFs. The potential of the CNN-based method using RF channel data has been shown in [33]. Previously, however, the training was performed at a fixed scatterer density and its performance was not fully investigated. In this paper, two CNNs were trained and evaluated using simulated RF channel data generated using one plane wave or three plane waves. The training sets were generated using four different scatterer densities and the test sets were generated using ten different scatterer densities. The evaluation was performed with respect to three criteria, which are detection, localization, and resolution. Additionally, two phantoms with water-filled cavities were 3-D printed and imaged to examine the feasibility of the CNN method on measured data. Lastly, a comparison of the proposed method to Deep-ULM is discussed.

II. METHODS

Consider RF channel data $\boldsymbol{x} \in \mathbb{R}^{N_a \times N_l \times N_t}$ induced by scatterers $\boldsymbol{p} \in \mathbb{R}^{N_s \times 2}$, where N_a is the number of samples along the axial direction, N_l is the number of active elements of a transducer in reception, N_t is the number of transmissions, N_s is the number of scatterers, and 2 is the number of spatial dimensions (in the lateral and axial positions). The nonlinear mapping $f\colon \mathbb{R}^{N_a \times N_l \times N_t} \to \mathbb{R}^{N_s \times 2}$ needs to be found to estimate scatterer positions from the RF channel data, which satisfies

$$\boldsymbol{p} = f(\boldsymbol{x}). \tag{1}$$

Here N_s varies depending on the given RF channel data $\boldsymbol{x}$, so the mapping f needs to adjust N_s adaptively, but this is not straightforward. Therefore, the mapping f is decomposed into two functions g and h to handle the varying N_s. The mapping $g\colon \mathbb{R}^{N_a \times N_l \times N_t} \to \mathbb{R}^{N_h \times N_w}$ forms a confidence map $\boldsymbol{c} \in \mathbb{R}^{N_h \times N_w}$, where N_h and N_w are the number of samples in the axial and lateral directions, respectively. The confidence map $\boldsymbol{c}$ represents a region of interest (ROI) where the pixel values indicate confidences of scatterer presence in each pixel. The mapping $h\colon \mathbb{R}^{N_h \times N_w} \to \mathbb{R}^{N_s \times 2}$ detects and locates scatterers from the confidence map. The mapping in (1) can be rewritten using g and h as follows:

$$\begin{aligned}\boldsymbol{p} &= f(\boldsymbol{x}) \\ &= h(g(\boldsymbol{x})) = h(\boldsymbol{c}),\end{aligned} \tag{2}$$

where

$$\boldsymbol{c} = g(\boldsymbol{x}). \tag{3}$$

The overview of the proposed method is illustrated in Fig. 1. The mapping g was modeled by a fully CNN and the mapping h corresponded to local maxima identification with thresholding. The RF channel data simulation and confidence map generation are explained in Section II-A and II-B, respectively. The architecture of the proposed CNN is introduced in Section II-C. Scatterer detection from the confidence maps is explained in II-D and the phantom fabrication is described in Section II-E. A baseline method for comparison is introduced in Section II-F.

TABLE I
RF CHANNEL DATA SIMULATION PARAMETERS

Category	Parameter	Value
Transducer	Transmission frequency	5.2 MHz
	Pitch	0.20 mm
	Element width	0.18 mm
	Element height	6 mm
	Number of elements	192
Imaging	Number of TX elements	32
	Number of RX elements (N_l)	64
	Steered angles	$-15°$, $0°$, $15°$
Environment	Speed of sound (c)	1480 m/s
	Field II sampling frequency	120 MHz
	RF data sampling frequency	29.6 MHz
Scatterer	Number of scatterers (N_s)	$20 \cdot i$, $\forall i \in \{1, 2, \ldots, 10\}$
	Lateral position range	$(-3.2,\ 3.2)$ mm
	Axial position range	$(14.8,\ 21.2)$ mm

A. RF Channel Data Simulation

Field II pro [34]–[36] was used to simulate RF channel data to generate data sets for training, validation, and evaluation. The parameters for the simulation are listed in Table I. The transducer was modeled after a commercial 5.2 MHz 192-element linear array transducer, and a measured impulse response [37] was applied to make the simulated RF channel data as close to measured data as possible [38].

For each frame, a certain number of point scatterers were placed randomly within a region of 6.4 mm × 6.4 mm where the center of the region was 18 mm away from the transducer, and three steered plane waves were transmitted using 32 elements. All the simulated scatterers had the same scattering intensity. Motion and flow were not considered, therefore, the scatterers used in each frame were static in the three plane wave transmissions and the scatterer positions were independent between frames. The aperture was shifted for each steered angle to insonify only the ROI, as shown in Fig. 2. The

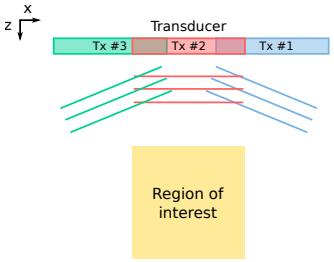


Fig. 2. Illustration of the imaging scheme. Scatterers were placed in the region of interest, and three steered plane waves were transmitted for each frame. The aperture was shifted to insonify only the region of interest.

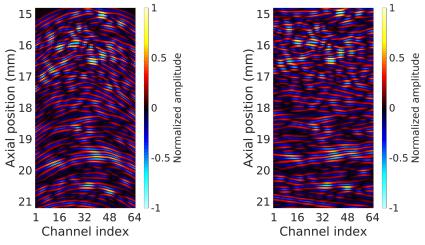


Fig. 3. An example of simulated RF channel data with one plane wave without steering. (a) is simulated raw RF channel data and (b) is delayed RF channel data. Note that the delay here is different from the delay for beamforming.

elements used in transmission were the 105th to the 136th ($-15°$), the 81st to the 112nd ($0°$), and the 57th to the 88th ($15°$) elements. Backscattered waves were received with 64 elements in the center of the transducer.

The simulated RF channel data were not beamformed but delayed based on the time-of-flight calculated as

$$\tau_i(x,z) = \left(\sqrt{(x-x_i)^2 + z^2} + z\right)/c. \tag{4}$$

Here τ_i is the time-of-flight of the i-th transmission, (x, z) is the data point, x_i is the center of the i-th transmission aperture, and c is the speed of sound. This preprocessing helped the CNN solve the problem by making wavefronts appear more like straight lines, instead of parabolas, as shown in Fig. 3, so it is different from the delay for beamforming.

The input and output of the proposed CNN were required to have the same number of samples along the axial direction. Therefore, the delayed RF channel data were re-sampled to match the same number of samples as confidence maps along the axial direction ($N_a = N_h$). Essentially, the sampling frequency of the RF channel data was determined by the pixel size of the confidence maps, and N_a was determined by the sampling frequency and the ROI. After preprocessing, the size of RF channel data $\boldsymbol{x}$ for one frame was $256 \times 64 \times 3$ before being fed to the CNNs.

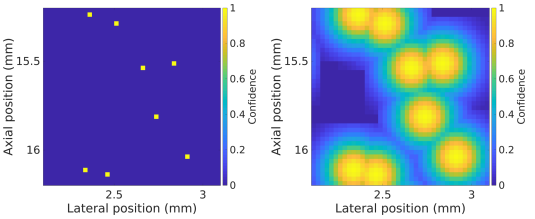


Fig. 4. An example of cropped confidence maps. (a) is a binary confidence map and (b) is a non-overlapping Gaussian confidence map created from (a).

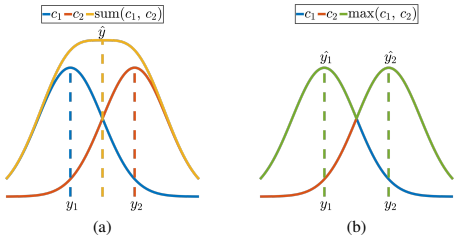


Fig. 5. A comparison of 1-D Gaussian confidence maps created by (a) summation and (b) maximum operation. There are two scatterers y_1 and y_2, and c_1 and c_2 are their confidence maps, respectively. The yellow line in (a) is the sum of c_1 and c_2. The green line in (b) is the maximum of c_1 and c_2. In (a), one scatterer $\hat{y}$ is found at a wrong position, whereas in (b), two scatterers $\hat{y}_1$ and $\hat{y}_2$ can be recovered at correct positions in the confidence map.

B. Non-overlapping Gaussian Confidence Map

Initially, binary confidence maps were created, where pixel values indicated presence (1) or absence (0) of a scatterer in the corresponding location, as shown in Fig. 4a. However, CNNs were not able to be trained using such confidence maps because most of their pixel values were zero. The sparse confidence maps provided small gradients during optimization and made the CNNs prone to converging to the wrong optimal solutions, returning only zero confidence maps regardless of input.

A non-overlapping Gaussian confidence map (Fig. 4b) was proposed to solve the imbalance problem of the binary confidence maps. Applying 2-D Gaussian filtering to sparse labels can improve training stability and guide CNNs to correct solutions [21], [24], [39]. But simply applying 2-D Gaussian filtering is problematic because the scatterer positions cannot be recovered in the confidence maps when the scatterers are closely spaced, as shown in Fig. 5a. To keep peaks at scatterer positions in the confidence maps, the Gaussian filter was applied one by one at each scatterer position in the binary confidence maps. Notably, when the Gaussian filter values induced by different scatterers were overlapped, the maximum values were taken instead of summation. By doing so, clearly separated peaks can be obtained at the true scatterer positions, as shown in Fig. 5b.

The parameters for non-overlapping Gaussian confidence maps are listed in Table II. The 2-D Gaussian filter is defined

TABLE II
CONFIDENCE MAP PARAMETERS

Parameter	Value
Pixel size	$25\,\mu\mathrm{m}$
Confidence map size ($N_h \times N_w$)	256×256
Gaussian filter size	21 pixels
Gaussian filter standard deviation	5 pixels

by

$$G(u, v; \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{u^2+v^2}{2\sigma^2}}, \tag{5}$$

where u and v are the pixel distances from the scatterer position in the lateral and axial directions, respectively, and σ is the standard deviation. The filter size was fixed to $4\sigma+1$ and the standard deviation was chosen by cross-validation among 3, 5, and 7 pixels. The scatterer positions were quantized according to pixel size since the confidence maps are on discrete grids. Here the pixel size was set to $25\,\mu\mathrm{m}$ ($\approx \lambda/10$); the lateral and axial localization uncertainties are $\pm 12.5\,\mu\mathrm{m}$ in an ideal situation. The confidence map size was 256×256 ($N_w = N_h = 256$) given the pixel size and the area of the ROI.

C. Convolutional Neural Network Architecture

The proposed CNN has an encoder-decoder structure with pooling and unpooling, similar to U-Net [13] but without skip connections. The encoder-decoder structure was adopted to transform the input in the channel data domain to the confidence map in the ultrasound image domain. In the encoding path, information is extracted from the RF channel data, and in the decoding path, the confidence maps are reconstructed based on the extracted information.

The overview of the CNN architecture and its components are shown in Fig. 6. It mainly consists of four *down-blocks*, one *conv-block*, and four *up-blocks*. In the *down-blocks*, the feature map size is decreased by strided convolution to reduce the amount of parameters, and in the *up-blocks*, the feature map size is increased to the confidence map size by pixel shuffle [40]. An 11×1 convolution layer prior to the encoding path extracts per-channel features, and two convolution layers after the decoding path refine the feature maps and return the confidence maps.

The pre-activation residual units [9] (Fig. 6a) were used instead of common convolution and rectified linear unit (ReLU) layers to improve the network performance. Batch normalization (BN) in the residual units helped ease the optimization, limited covariate shift, and had the effect of regularization [41]. Dropout [42] was additionally attached after the shortcut for further regularization. Leaky ReLU [43] and Sigmoid were chosen as non-linear activation. CoordConv [44] was added to transfer spatial information over convolution layers.

The same CNN architecture was used for both one and three plane wave data. For three plane waves, the preprocessed RF channel data from each transmission in a frame were stacked along the third dimension before applied to a CNN.

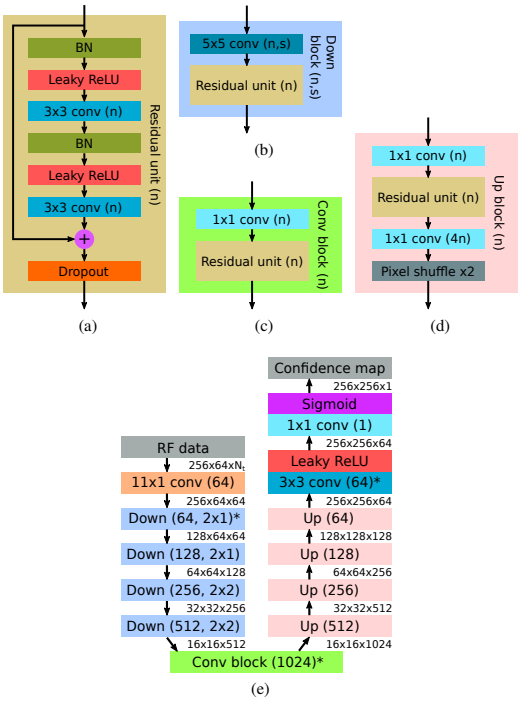


Fig. 6. The proposed CNN architecture and its components: (a) residual unit, (b) down-block, (c) conv-block, (d) up-block, and (e) the network overview. The n and s in the parenthesis are the number of kernels and stride. In (e), the sets of three numbers are the feature map size between two blocks, and the asterisk indicates that CoordConv was applied at the first convolution in the block.

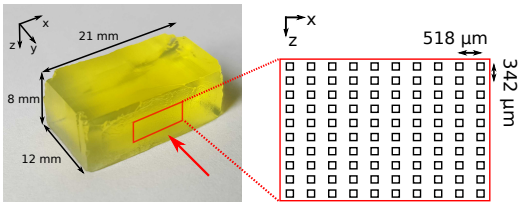
D. Scatterer Detection from Confidence Maps

The scatterer positions can be found by locating the pixels whose confidences are one in the true confidence map $\boldsymbol{c}$. However, the estimated confidence map $\hat{\boldsymbol{c}} = g(\boldsymbol{x})$ acquired from a trained CNN is an approximation of $\boldsymbol{c}$. It is not guaranteed that the confidences are one where scatterers are located in $\hat{\boldsymbol{c}}$. Therefore, the algorithm relied on the fact that pixels containing scatterers are local peaks. The scatterer positions were recovered by finding the local maxima whose confidence is larger than a certain decision value. The chosen decision value was 0.9 in this work.

E. Phantom Fabrication

Two PEGDA $700\,\mathrm{g/mol}$ hydrogel phantoms were 3-D printed [45], [46] to assess the CNN method on measured data. The phantoms contained water-filled cavities which acted as scatterers. The volume of each cavity was $45\,\mu\mathrm{m} \times 1000\,\mu\mathrm{m} \times 45\,\mu\mathrm{m}$. The cavities were designed to be elongated in the elevation direction to increase the intensity of received signals.

In the first phantom, 100 cavities were placed on a 10×10 grid with a spacing of $518\,\mu\mathrm{m}$ in the lateral direction and $342\,\mu\mathrm{m}$ in the axial direction, as illustrated in Fig. 7. This grid scatterer phantom had the spacing larger than the resolution



(a) (b)

Fig. 7. Fabricated 3-D phantom with uniformly spaced cavities: (a) photograph of the phantom and (b) 100 cavities placed on a 10×10 grid.

limit of DAS to show that the CNN method works on measured data. The second phantom, on the other hand, had 100 cavities randomly distributed with a minimum spacing of $190\,\mu\mathrm{m}$ to demonstrate that the CNN method can resolve targets closer than the conventional resolution limit. The minimum spacing between cavities were constrained due to the cavity size and the 3-D printer voxel size.

F. Baseline Method

Local peak detection on the beamformed images was chosen as a baseline method for comparison. RF channel data were DAS beamformed in the region of interest with the same pixel size as the confidence map, and, for three plane wave transmissions, beamformed images in a frame were coherently compounded [47]. The baseline method detected and located scatterers in the envelope detected and log-compressed B-mode images with a dynamic range of $40\,\mathrm{dB}$. The B-mode images were smoothed to avoid more than one pixel corresponding to a peak, and scatterer positions were estimated by finding local maxima.

Deconvolution using an estimated PSF is one of the commonly used techniques for microbubble localization [5]. However, it was not considered in this work since its performance has been found to be sensitive to parameters when the PSFs were highly overlapped, and the spatially varying PSF of ultrasound imaging resulted in imprecise scatterer localization.

III. Experiments

A. Training Details

CNNs, which correspond to the mapping g in (2), were trained to return the corresponding confidence map $\boldsymbol{c}_i$ given RF channel data $\boldsymbol{x}_i$ by minimizing the mean squared error (MSE), given by

$$\mathcal{L}_{\mathrm{MSE}}\left(\boldsymbol{x}_i, \boldsymbol{c}_i; g\right) = \frac{1}{N} \sum_{i=1}^{N} \left\| \boldsymbol{c}_i - g\left(\boldsymbol{x}_i\right) \right\|_F^2, \tag{6}$$

where N is the number of samples and $\|\cdot\|_F$ is the Frobenius norm.

One data set consisted of frames simulated at the same scatterer density, and four training sets and four validation sets were generated at the scatterer densities of $0.49\,\mathrm{mm}^{-2}$, $0.98\,\mathrm{mm}^{-2}$, $2.44\,\mathrm{mm}^{-2}$, and $4.88\,\mathrm{mm}^{-2}$, i.e., the numbers of scatterers were 20, 40, 100, and 200 in one frame, respectively. Each training set and validation set had 10 240 and 1280 frames, respectively.

The kernel weights were initialized by orthogonal initialization [48] and optimized with ADAM [49] by setting $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-7}$. Firstly, the training was performed using only the training set at the scatterer density of $2.44\,\mathrm{mm}^{-2}$. The initial learning rate was 10^{-4} and it was halved every 100 epochs. After 600 epochs, the learning rate was set to 10^{-5} and the training continued using all the training sets while the learning rate was halved every 50 epochs. The mini-batch size was 32, and each batch was composed of frames from all four training sets after 600 epochs. The CNN was implemented in Python using Tensorflow [50], and were trained on a server equipped with a NVIDIA TESLA V100 16 GB PCIe graphics card. The total number of training epochs was 800, and the training took approximately 40 hours.

During training, the RF channel data and confidence maps were flipped along the lateral direction at random with a probability of 0.5 to augment the training sets. White Gaussian noise was added to the RF channel data for generalization along with BN and dropout. The signal-to-noise ratio after noise addition was $6\,\mathrm{dB}$, and the dropout rate was 0.3. The RF channel data and confidence maps were then normalized to be in the range $[-1, 1]$ and $[0, 1]$, respectively. Validation was performed every epoch to monitor the training, and also for cross-validation to choose hyper-parameters.

For both simulation and phantom experiment, two CNNs were trained and compared: one CNN acting on the data from one plane wave ($0°$) and the other CNN acting on the data from three plane waves ($-15°$, $0°$, $15°$).

B. Simulation Experiment

The CNNs were evaluated on simulated test sets firstly. One test set consisted of 3840 frames simulated at the same scatterer density, and ten test sets were created at scatterer densities from $0.49\,\mathrm{mm}^{-2}$ to $4.88\,\mathrm{mm}^{-2}$ by varying the number of scatterers from 20 to 200 with intervals of 20. The parameters in Table I were used again, apart from the number of scatterers. The evaluation was performed on the frames simulated at various scatterer densities to evaluate how the performance changes over different scatterer densities and how well the CNNs were generalized in terms of scatterer density.

C. 3-D Printed Phantom Experiment

1) RF Channel Data Acquisition: The 3-D printed phantoms were scanned using the $5.2\,\mathrm{MHz}$ 192-element linear array transducer which has the same parameters as in Table I. The raw RF channel data were acquired by the synthetic aperture real-time ultrasound system (SARUS) experimental ultrasound scanner [51]. The same imaging scheme and processing as in the simulation were applied.

The experimental setup is shown in Fig. 8. The transducer was fixed, and a water tank containing the phantom was placed on a motion stage. The phantom was aligned with the transducer by the motion stage, capable of translating

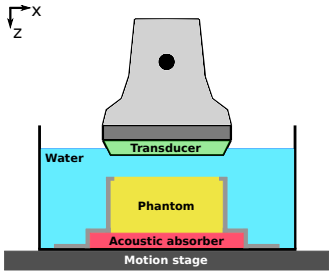


Fig. 8. Illustration of the experimental setup for phantom measurement.

in the x- and y-axis, and rotating around the z-axis. During measurement, the motion stage was translated along the x-axis in steps of $50\,\mu\text{m}$ between frames, and 33 frames were acquired for each phantom experiment.

2) Training Set Modification: The training sets were modified and new CNNs were trained from scratch for the phantom experiment. Transfer learning was also considered but it did not show as good performance as training from scratch. In the simulation, it was assumed that scatterers were infinitesimally small points. However, the cavities in the phantoms were squares, as shown in Fig. 7b, if the elevation direction is ignored. Scattering, therefore, happens twice at each cavity: once when a wave goes into the cavity and the other when the wave comes out of the cavity. Additionally, the first scattering experiences a phase reversal because the acoustic impedance of the phantoms is higher than that of water.

RF channel data for training were accordingly re-simulated by modeling each scatterer using two points separated by the cavity size axially and with a phase reversal. To remain consistent, the same scatterer positions of the original training set were used.

3) Depth Correction: The speed of sound in the phantoms is higher than in water. The axial positions of the estimated scatterers were corrected to compensate for the different speed of sound in the phantoms by

$$\hat{z}^* = (\hat{z} - d_{\text{pht}}) \cdot \frac{c_{\text{water}}}{c_{\text{pht}}} + d_{\text{pht}}, \tag{7}$$

where $\hat{z}$ and $\hat{z}^*$ are the axial position before and after correction, c_{water} and c_{pht} are the speed of sound in water and in the phantoms, respectively, and d_{pht} is the distance from the transducer to the surface of the phantoms.

D. Evaluation Metrics

Three evaluation criteria were considered to assess the CNNs: detection, localization, and resolution. The positive and negative detections were determined by pairing estimated scatterers with true scatterers based on their pair-wise distances, as stated in Algorithm 1. Namely, to be a positive detection, an estimated scatterer should be exclusively matched with a true scatterer within a certain localization precision. This localization precision can be translated to the target resolution of ULM without tracking. It was set to be half of the full width at half maximum (FWHM) in this work. Specifically, an

Algorithm 1 Algorithm for determining positive or negative detections

Input: $\boldsymbol{p} \in \mathbb{R}^{N_s \times 2}$ and $\hat{\boldsymbol{p}} \in \mathbb{R}^{\hat{N}_s \times 2}$, where $\boldsymbol{p}$ is true scatterer positions and $\hat{\boldsymbol{p}}$ is estimated scatterer posions
Output: Positive or negative detection $\boldsymbol{a} \in \mathbb{R}^{\hat{N}_s \times 1}$

$\boldsymbol{a} \leftarrow \boldsymbol{0} \in \mathbb{R}^{\hat{N}_s \times 1}$
$D \leftarrow \left\{ (d_{ij}) \in \mathbb{R}^{N_s \times \hat{N}_s} \,\middle|\, d_{ij} = \|p_i - \hat{p}_j\|_2 \right\}$
for $j = 1$ to $\hat{N}_s$ **do**
  $\hat{i} \leftarrow \arg\min D_{*,j}$
  if $j = \arg\min D_{\hat{i},*}$ **and** $\frac{\left(p_{\hat{i}1} - \hat{p}_{j1}\right)^2}{(\text{FWHM}_x/2)^2} + \frac{\left(p_{\hat{i}2} - \hat{p}_{j2}\right)^2}{(\text{FWHM}_z/2)^2} < 1$ **then**
    $a_j \leftarrow 1$
  else
    $a_j \leftarrow 0$
  end if
end for

ellipse whose major axis and minor axis were half of FWMH_x and half of FWMH_z, respectively, was used as the desired localization precision, where FWMH_x is the lateral FWHM and FWMH_z is the axial FWHM. This bi-directional matching process was extended from the left-right consistency check [52], [53] for stereo matching in computer vision. It conforms to the uniqueness constraint; one true scatterer can be paired with at most one estimated scatterer.

Detection capability was assessed by quantifying wrong detections and missed detections using precision, recall, and F_1 score, which are defined as follows:

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{8}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{9}$$

and

$$\text{F}_1 \text{ score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{10}$$

where TP is the number of true positives (correct detections), FP is the number of false positives (wrong detections), and FN is the number of false negatives (missed detections).

Localization uncertainties were measured by calculating the lateral and axial position errors. Only positive detections were considered for the localization assessment.

Spatial resolution, meaning the ability to separate two points that are close together, was investigated statistically. For two isolated true scatterers, it was checked whether they were detected. A pair of scatterers was set to *resolved* if both scatterers were detected. It was set to *non-resolved* if only one of them was detected. And it was not considered if none of them were detected, as this would be a detection problem. The resolved rates were calculated in $20\,\mu\text{m} \times 20\,\mu\text{m}$ bins by

$$\text{Resolved rate} = \frac{N_{\text{res}}}{N_{\text{res}} + N_{\text{non-res}}}, \tag{11}$$

where N_{res} is the number of resolved pairs and $N_{\text{non-res}}$ is the number of non-resolved pairs in a bin.

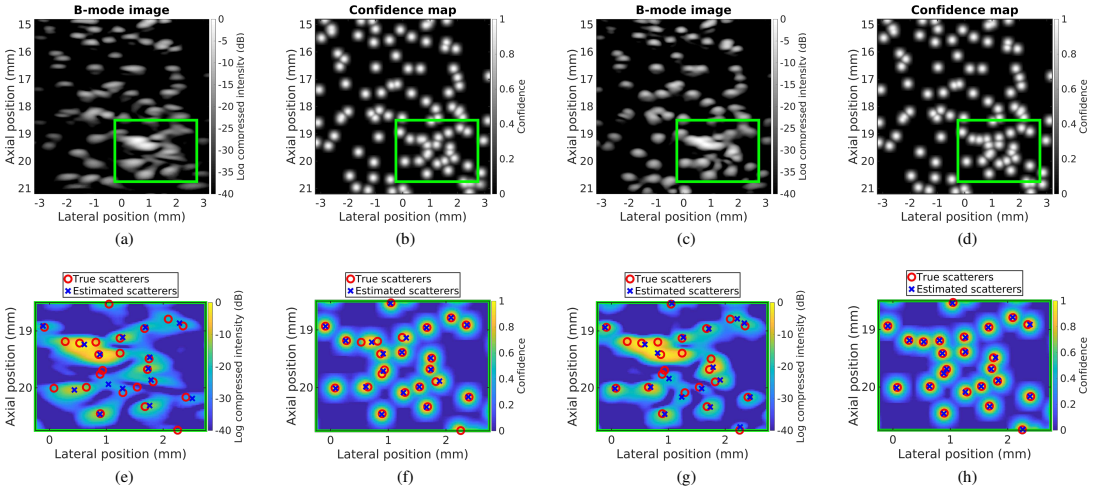


Fig. 9. A comparison of scatterer detection between baseline method and CNN method on a simulated test frame. (a) and (c) are DAS beamformed B-mode images with one and three plane waves, respectively. (b) and (d) are estimated confidence maps by CNNs with one and three plane waves, respectively. (e) - (h) show true scatterers and estimated scatterers from their corresponding results above in the same column in the green box region.

TABLE III
PRECISION, RECALL, AND F_1 SCORE COMPARISON ON THE SIMULATED TEST SETS

Method	One plane wave			Three plane waves		
	Precision	Recall	F_1	Precision	Recall	F_1
Baseline	0.83	0.51	0.63	0.93	0.62	0.75
CNN	0.99	0.83	0.90	1.00	0.91	0.96

IV. RESULTS

The CNN method results on the simulated data and the measured data of the 3-D printed phantoms presented in this Section. Quantitative evaluation comparing one plane wave and three plane waves was performed as specified in Section III-D. The results of the baseline method on the same test data are also presented for comparison.

A. Simulation Experiment

The qualitative comparison between the baseline and CNN methods is shown in Fig 9. The proposed CNN method successfully detected and localized high-density scatterers when the baseline method failed due to overlapping PSFs.

The detection results on the simulated test sets are shown in Table III. The CNN method achieved the better precision, recall, and F_1 score for both one and three plane transmissions. Also, when the higher number of transmissions was involved, the detection performance was improved for both methods. The detection capabilities over different scatterer densities were investigated, as shown in Fig. 10. The recalls dropped as the scatterer density increased while the precisions were relatively kept high. In addition, the recalls of the baseline method decreased more drastically as the scatterer density increased, which led to the lower F_1 scores.

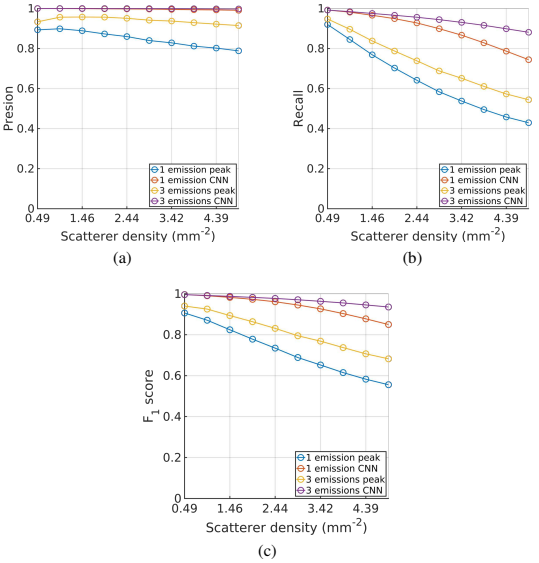


Fig. 10. Detection capabilities of the baseline and CNN methods over different scatterer densities on the simulated test sets with one and three plane waves: (a) precision, (b) recall, and (c) F_1 score.

The comparison of localization uncertainties between the baseline and CNN methods on the simulated test sets are presented in Fig. 11, using box-and-whisker plots along with violin plots. The bottom and top edges of the blue boxes

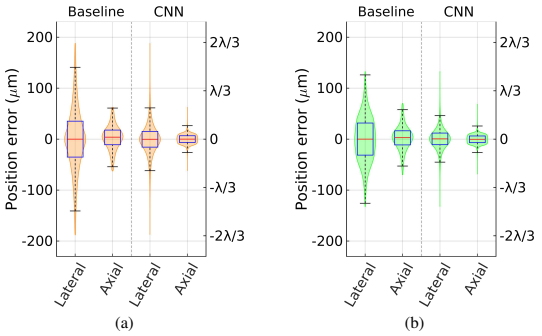


Fig. 11. Localization uncertainties of baseline and CNN methods on the simulated test sets. (a) and (b) are the results with one plane wave and three plane waves, respectively.

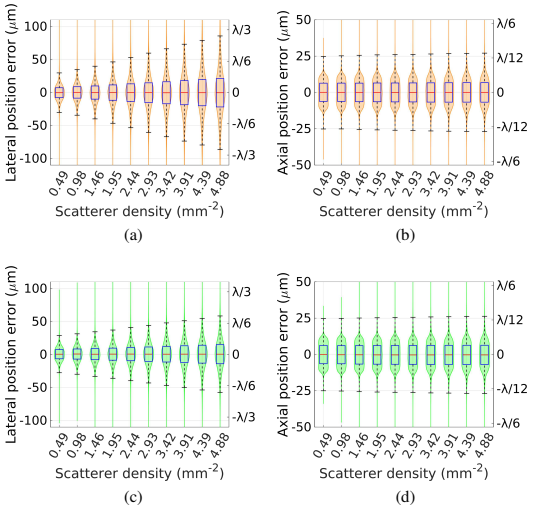


Fig. 12. Localization uncertainties of the CNN method on the simulated test sets at different scatterer densities: the lateral position errors with (a) one plane wave and (b) three plane waves, and the axial position errors with (c) one plane wave and (d) three plane waves.

indicate the 25 th (q_1) and 75 th percentiles (q_3), and the center red lines indicate the medians. The whiskers, vertically extended lines from the boxes, indicate the range of values except outliers, which are greater than $q_3 + 1.5 \times (q_3 - q_1)$ or less than $q_1 - 1.5 \times (q_3 - q_1)$. The violin plots were overlaid as shaded area to demonstrate the error distribution directly. For both methods, the lateral position error was higher than the axial position error, and the CNN method achieved clearly better localization than the baseline method. For the most part the medians were very close to zero, indicating that the scatterer position estimation was unbiased in both directions. The localization was also improved when more plane waves were transmitted. Localization uncertainties of the CNN method at different scatterer densities are shown in Fig. 12. Neither the

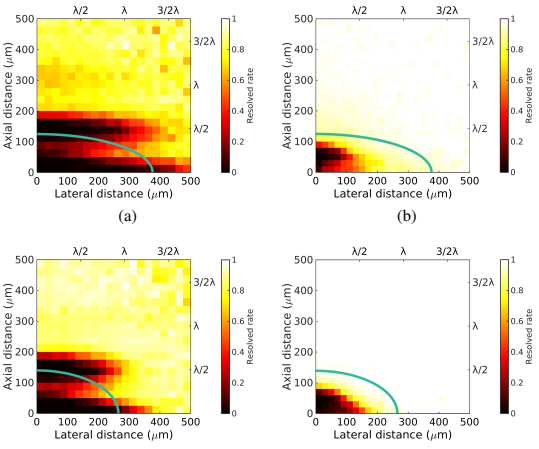


Fig. 13. Resolved rate of (a), (c) baseline methods and (b), (d) CNN methods on the simulated test sets where (a) and (b) are with one plane wave and (c) and (d) are with three plane waves. The green lines represent the theoretical resolution limit of DAS beamforming.

TABLE IV
PRECISION, RECALL, AND F_1 SCORE COMPARISON ON THE PHANTOM TEST SETS

Phantom	Method	One plane wave			Three plane waves		
		Precision	Recall	F_1	Precision	Recall	F_1
Grid	Baseline	0.82	0.41	0.54	1.00	1.00	1.00
	CNN	0.89	0.22	0.35	0.98	1.00	0.98
Random	Baseline	0.47	0.23	0.31	0.49	0.32	0.39
	CNN	0.53	0.37	0.44	0.59	0.63	0.61

scatterer density nor the number of transmissions had much impact on the axial position errors. The lateral position errors, on the other hand, gradually increased as the scatterer density increased.

The 2-D histograms in Fig. 13 show the resolved rates of two isolated scatterers measured in $20\,\mu\text{m} \times 20\,\mu\text{m}$ bins. The green lines represent the theoretical resolution limit of DAS beamformed images, assuming that the 6 dB contour of a PSF is an ellipse. The FWHM was measured on a simulated PSF in the center of the ROI. For one plane wave, the FWHM was $376\,\mu\text{m}$ ($1.32\,\lambda$) laterally and $125\,\mu\text{m}$ ($0.44\,\lambda$) axially. For three plane waves, the FWHM was $265\,\mu\text{m}$ ($0.93\,\lambda$) laterally and $140\,\mu\text{m}$ ($0.49\,\lambda$) axially. The resolution results show that the CNN method can resolve scatterers closer than the DAS limit. The mean resolved rates in the area under the green line for the baseline and CNN methods were 0.16 and 0.68 with one plane wave, and 0.17 and 0.67 with three plane waves, respectively.

B. 3-D Printed Phantom Experiment

For the phantom study, CNNs were applied to measured data without evaluation on simulated test data. The qualitative

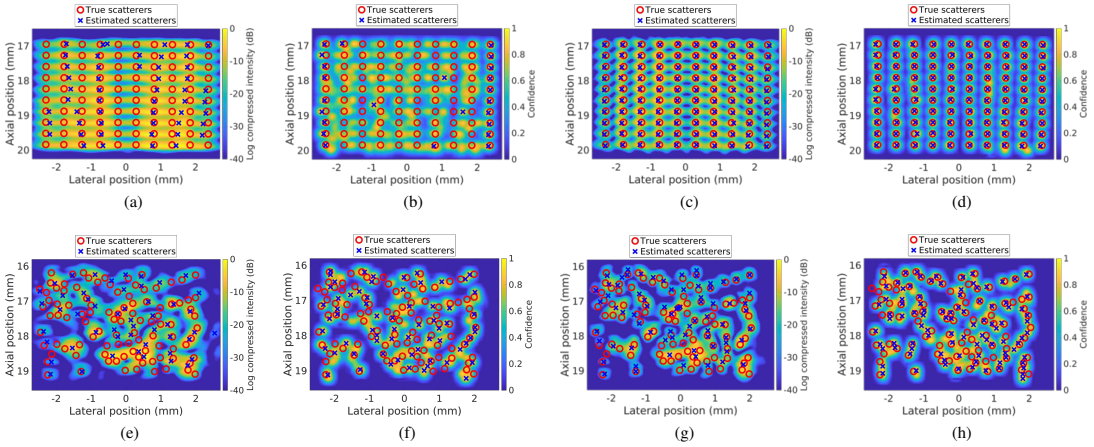


Fig. 14. A comparison of scatterer detection between baseline method and CNN method on phantom measured frames. (a) - (d) are results of the grid phantom and (e) - (h) are results of the random phantom. B-mode images with (a), (e) one plane wave and (c), (g) three plane waves and confidence maps with (b), (f) one plane wave and (d), (h) three plane waves are shown with true scatterers and estimated scatterers.

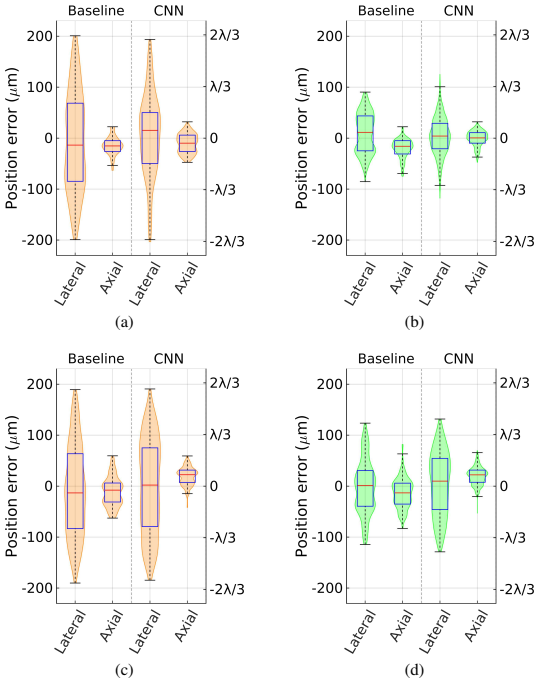


Fig. 15. Localization uncertainties of baseline and CNN methods on phantom measured data: (a) and (b) are results on the grid scatterer phantom with one and three plane waves, respectively. (c) and (d) are results on the random scatterer phantom with one and three plane waves, respectively.

results of the baseline and CNN methods on the grid and random scatterer phantoms are presented in Fig. 14 and their quantitative comparison is shown in Table IV and Fig. 15.

With one plane wave, side lobe level was high, and side lobes were added up when the scatterers were placed in a grid. Therefore, the DAS beamforming was unable to identify individual scatterers of the grid phantom properly, as shown in Fig. 14a. The CNN method also achieved poor detection with one plane wave on the grid phantom, as shown in Fig. 14b. The CNN was not generalized sufficiently to handle regularly placed scatterers as the training frames were generated by placing scatterers randomly. Most of the scatterers in the first and the last columns were correctly detected, but the other scatterers were missed. Thus, the precision was higher than the baseline but the recall was lower. On the contrary, with three plane waves, the baseline method found all the scatterers without any false detection. The CNN method also achieved comparable detection results with three plane waves, showing that more transmissions for a frame helped generalization of the CNN. For localization, the CNN method showed slightly smaller uncertainties except the axial localization with one plane wave.

On the random scatterer phantom, the CNN method achieved better detection for both one and three plane waves. For localization, the CNN method showed smaller axial uncertainties but little higher lateral uncertainties. With three plane waves, the detection and localization were improved but, in general, it was more challenging to identify scatterers for both methods on the random scatterer phantom.

V. Discussion

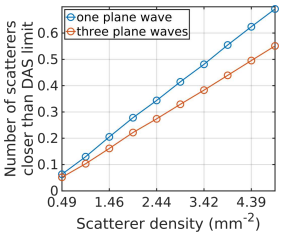
A CNN-based scatterer detection and localization method is presented. Instead of end-to-end training, the CNNs were trained to learn the mapping from RF channel data to non-overlapping Gaussian confidence maps, and scatterers were

detected and localized from the confidence maps by looking for local maxima. This two-step framework made it possible to handle varying numbers of scatterers (N_s). By obtaining non-overlapping Gaussian confidence maps from RF channel data without beamforming, it was able to identify high concentrations of scatterers which cannot be separated by conventional ultrasound imaging due to the overlapping PSFs. This method also has an advantage of fast processing by exploiting GPU computation. The proposed CNN implicitly included beamforming since it is a mapping from the channel domain to the ultrasound image domain, which is a bottleneck of current ultrasound imaging. For the CNNs, processing time for a frame was $16\,\mathrm{ms}$ on average in a PC equipped with a NVIDIA Titan V graphics card.

It was essential to use non-overlapping Gaussian confidence maps to make training work. Binary confidence maps were initially used to train CNNs with advanced loss functions such as weighted cross entropy [13], jaccard loss [54], or focal loss [55], as well as simple loss functions such as MSE or mean absolute error, but all of them failed. The binary confidence maps were too sparse to be handled by simply manipulating the loss function. However, non-overlapping Gaussian confidence maps relaxed the sparsity of the binary confidence maps while being able to recover scatterer positions by taking the maximum of overlapping Gaussians. Therefore, the larger gradients were provided during training and the CNNs were able to be guided to the correct solutions stably.

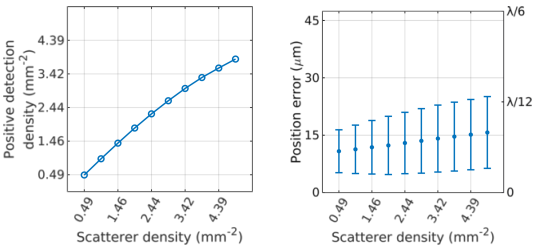
The training was firstly performed in the training set at the scatterer density of $2.44\,\mathrm{mm}^{-2}$, and was further performed on the whole training sets later. Interestingly, the CNNs trained at the scatterer density of $2.44\,\mathrm{mm}^{-2}$ were already well generalized at the scatterer densities higher than $2.44\,\mathrm{mm}^{-2}$. On the other hand, the CNNs achieved poor precision and localization at the lower scatterer densities as two Gaussian peaks appeared laterally near a true scatterer position in the confidence maps. Therefore, the training sets had more frames at the lower scatterer densities. It was also investigated to train CNNs using the whole training sets from the beginning of the training but the proposed way was more efficient; CNNs converged to the solutions with fewer iterations.

The delayed RF signal induced by a scatterer lies across all the channels and at several depths depending on the lateral location of the scatterer. Hence, large receptive fields were required for a CNN, so four *down* and four *up blocks* were used. We tried to incorporate skip connections into the proposed CNN by, if necessary, applying upsampling to the feature maps in the contracting path to match the size of their corresponding feature maps in the expanding path. For image segmentation, the skip connections play an important role to recover lost spatial information during downsampling [13], [56]. The resulting reconstructed images have more fine details and, as a result, provide better localized semantic segmentation. However, the skip connections hindered successful training for the task in this paper and the CNNs learned zero confidence maps. We presume that the feature maps extracted from RF channel data in the contracting path are not directly related to the reconstruction of confidence maps, unlike image segmentation. Instead, CoordConv [44] was applied to cope with the spatial information loss. The CNNs with CoordConv localized non-overlapping Gaussians more precisely and achieved the better recall and localization precision on the validation sets.



(a)

Fig. 16. The average numbers of scatterers closer than the theoretical resolution limit of DAS beamforming given a scatterer at different scatterer densities in the simulated test sets.



(a) (b)

Fig. 17. Recall and localization precision re-calculated to compare CNN method to Deep-ULM: (a) Positive detection density and (b) median of Euclidean position errors with one standard deviation bars at different scatterer densities.

On the simulated test sets, the proposed method outperformed the baseline method. The performance drop was much more severe for the baseline method at high scatterer densities, where the more scatterers were placed with in the resolution limit. Fig.16 shows the average numbers of scatterers within the FWHM (the $6\,\mathrm{dB}$ ellipse contour) given a scatterer in the simulated test sets.

Deep-ULM is another CNN-based method which localizes high-density targets from beamformed images that contain overlapping PSFs. To compare the proposed method with Deep-ULM, the recall and localization errors were recalculated following the method which van Sloun *et al.* used to generate the results in the supplementary Fig. 1 in [24]. The threshold value for determining positive detection was $\lambda/7$ and Euclidean distances between the true and estimated scatterers were calculated. The evaluation results depend on the threshold value. As it increases, recall improves while localization precision degrades. The threshold $\lambda/7$ was chosen following [24] for a fair comparison. The results are presented in Fig. 17. Both methods showed good performance at high densities but the proposed method achieved slightly better recall and localization precision. Deep-ULM recovered roughly $1.80\,\mathrm{mm}^{-2}$, while the proposed method recovered $2.26\,\mathrm{mm}^{-2}$

at the density of $2.44\,\mathrm{mm}^{-2}$, and Deep-ULM recovered roughly $2.10\,\mathrm{mm}^{-2}$ at the density of $3.53\,\mathrm{mm}^{-2}$ when the proposed method recovered $3.00\,\mathrm{mm}^{-2}$ at the density of $3.42\,\mathrm{mm}^{-2}$. The median of Euclidean errors of Deep-ULM was approximately $\lambda/12$ but the proposed method achieved smaller errors than that. It is difficult to conclude that the proposed method outperforms Deep-ULM since the evaluation was not performed on the same test data. This, however, shows the potential of the methods directly employing RF channel data.

To assess the proposed method for real world applications, two 3-D printed phantoms were imaged. One of the benefits of using the 3-D printed phantoms is that true scatterer positions and the dimensions of the phantom and scatterers (cavities) are known. It was important to modify the scatterers in the training sets to match the cavity dimensions. The CNNs trained for the simulation experiment failed on the measured data, showing too many false positive detections axially. However, the CNNs trained with the modified training sets successfully identified scatterers to some extent except some scatterers on the grid phantom when one plane wave was transmitted as seen in Fig. 14b. It is notable that this was achieved only with the simulated training data, since it is extremely difficult to obtain sufficient training data with ground truth for these kinds of experiments.

The phantom experiments show that the CNN method is transferable to measured data by modeling scatterers properly in the training data simulation. The baseline method performed slightly better for the most trivial case, namely the grid scatterer phantom with three plane waves, but the CNN method performed better on the random scatterer phantom. Even so, the CNN method on the random scatterer phantom presented a relatively large number of false positives compared to the simulation results. This could be because of the discrepancy between the training (simulated) data and the test (phantom) data. There are factors not considered in the simulation such as attenuation, different scattering intensities of the cavities, and different speed of sound in the phantom medium. Moreover, a further degradation of the performance is expected on *in vivo* data since the discrepancy between the training data and the *in vivo* data would become larger due to scatterer response variations, refraction, reverberation artifacts, etc. A more versatile simulation using various parameters to cover possible *in vivo* variations of RF channel data and a more generalized CNN model could increase the CNN method performance on the measured phantom data and overcome the potential limits in *in vivo* scenarios.

The proposed method gives 2-D images using a 1-D transducer. This limits the view of target structure along the elevation direction. The 3-D printed phantoms are essentially 2-D phantoms which have elongated cavities and the dimension along the elevation direction was not captured in the results. This limitation can be solved by using 2-D transducers such as fully addressed transducers or row-column addressed transducers.

Several problems are expected to occur if the CNN method is applied to MB detection for SRUS. MBs are not static but move with different velocities depending on the vessel size. This should be considered during training data generation. Also, it is important to model MBs properly in simulations since their sizes and other physical properties vary. It was necessary to remodel scatterers following the real physical structure for the phantom experiment. This is expected to be an important factor when applying the CNN method on measured MB signals.

Background scattering from tissue was not dealt with here since this work focused on a proof-of-concept of CNNsâĂŹ ability to detect and localize high concentrations of scatteres from RF channel data. For *in-vivo* scenarios, the tissue signals may hinder the CNN method, so a way of rejecting them without hurting the performance of CNNs needs to be investigated. For example, clutter filtering based on singular value decomposition (SVD) or contrast-enhanced ultrasound (CEUS) imaging such as pulse inversion [57] or amplitude modulation [58] can be applied. However, the drawbacks of such methods are that it is difficult to find an optimal singular value for SVD to separate MB signals, and the CEUS imaging limits the frame-rate. In addition, both methods have a chance to distort the signals from the MBs, which would make the detected MB signals different from the data used for training. Alternatively, another neural network such as CORONA [59] can be deployed, which is a Robust PCA-based unfolded neural network that performs clutter filtering. By incorporating CORONA with the proposed CNN method, clutter filtering and MB localization can be learned simultaneously.

Lastly, further research on the optimal imaging scheme and scalability of CNN is required. Plane waves were used to support the hypothesis in a small region. In practice, however, a larger field of view is needed. Also, the more correlated data are available, the better estimation can be achieved. The CNNs with three plane waves achieved better performance than the CNN with one plane wave in all evaluation criteria, but this increases the required GPU memory. In addition, the imaging scheme would affect the capability of the CNN method and plane waves might not be the optimal choice. It is necessary to examine how other imaging schemes, such as focused or defocused waves affect the CNN method, or a new imaging scheme could be developed.

VI. Conclusion

The CNN-based scatterer detection and localization method is presented. CNNs were trained to return non-overlapping Gaussian confidence maps from simulated RF channel data, and the scatterer positions were estimated from the confidence maps. The simulation results show that the proposed method can identify high-density scatterers successfully even when some of them are closer than the resolution limit of conventional ultrasound imaging. It is also shown that the CNN method can be applied to real measured data by modeling scatterers following the true scatterer structure. The CNN method can potentially be extended to replace DAS beamforming for high concentration MB detection and thus reduce the long data acquisition times of SRUS using ULM.

ACKNOWLEDGMENT

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan V graphics card used for this research.

REFERENCES

[1] F. L. Thurstone and O. T. von Ramm, "A new ultrasound imaging technique employing two-dimensional electronic beam steering," in *Acoustical Holography*, P. S. Green, Ed., vol. 5. New York: Plenum Press, 1974, pp. 249–259.

[2] O. Couture, B. Besson, G. Montaldo, M. Fink, and M. Tanter, "Microbubble ultrasound super-localization imaging (MUSLI)," in *Proc. IEEE Ultrason. Symp.*, 2011, pp. 1285–1287.

[3] O. M. Viessmann, R. J. Eckersley, K. Christensen-Jeffries, M. X. Tang, and C. Dunsby, "Acoustic super-resolution with ultrasound and microbubbles," *Phys. Med. Biol.*, vol. 58, pp. 6447–6458, 2013.

[4] M. A. O'Reilly and K. Hynynen, "A super-resolution ultrasound method for brain vascular mapping," *Med. Phys.*, vol. 40, no. 11, pp. 110 701–7, 2013.

[5] C. Errico, J. Pierre, S. Pezet, Y. Desailly, Z. Lenkei, O. Couture *et al.*, "Ultrafast ultrasound localization microscopy for deep super-resolution vascular imaging," *Nature*, vol. 527, pp. 499–502, November 2015.

[6] K. Christensen-Jeffries, R. J. Browning, M. Tang, C. Dunsby, and R. J. Eckersley, "In vivo acoustic super-resolution and super-resolved velocity mapping using microbubbles," *IEEE Trans. Med. Imag.*, vol. 34, no. 2, pp. 433–440, February 2015.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Neural Information Processing Systems*, 2012, pp. 1097–1105.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[9] ——, "Identity mappings in deep residual networks," in *Eur. Conf. Computer Vision*, 2016, pp. 630–645.

[10] G. Huang, Z. Liu, L. v. d. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2017, pp. 2261–2269.

[11] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *IEEE Int. Conf. Computer Vision*, 2017, pp. 2980–2988.

[12] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," `arXiv:1804.02767v1 [cs.CV]`, 2018.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.

[14] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2017, pp. 2881–2890.

[15] L. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Eur. Conf. Computer Vision*, 2018, pp. 801–818.

[16] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2017, pp. 105–114.

[17] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2017, pp. 1132–1140.

[18] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino *et al.*, "Imaging intracellular fluorescent proteins at nanometer resolution," *Science*, vol. 313, no. 5793, pp. 1642–1645, 2006.

[19] S. T. Hess, T. P. K. Girirajan, and M. D. Mason, "Ultra-high resolution imaging by fluorescence photoactivation localization microscopy," *Biophysical Journal*, vol. 91, no. 11, pp. 4258–4272, 2006.

[20] M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)," *Nature methods*, vol. 3, no. 10, pp. 793–795, 2006.

[21] E. Nehme, L. E. Weiss, T. Michaeli, and Y. Shechtman, "Deep-STORM: super-resolution single-molecule microscopy by deep learning," *Optica*, vol. 5, no. 4, pp. 458–464, Apr 2018.

[22] W. Ouyang, A. Aristov, M. Lelek, X. Hao, and C. Zimmer, "Deep learning massively accelerates super-resolution localization microscopy," *Nature biotechnology*, 2018.

[23] N. Boyd, E. Jonas, H. Babcock, and B. Recht, "Deeploco: Fast 3d localization microscopy using neural networks," `bioRxiv 267096`, 2018.

[24] R. J. G. van Sloun, O. Solomon, M. Bruce, Z. Z. Khaing, H. Wijkstra, Y. C. Eldar *et al.*, "Super-resolution ultrasound localization microscopy through deep learning," `arXiv:1804.07661v2 [eess.SP]`, 2018.

[25] D. Allman, A. Reiter, and M. A. L. Bell, "Photoacoustic source detection and reflection artifact removal enabled by deep learning," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1464–1477, 2018.

[26] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2017.

[27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Int. Conf. Learning Representations*, 2015.

[28] A. C. Luchies and B. C. Byram, "Deep neural networks for ultrasound beamforming," *IEEE Trans. Med. Imag.*, vol. 37, no. 9, pp. 2010–2021, 2018.

[29] D. Hyun, L. L. Brickson, K. T. Looby, and J. J. Dahl, "Beamforming and speckle reduction using neural networks," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 66, no. 3, pp. 898–910, 2019.

[30] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair *et al.*, "Generative adversarial nets," in *Neural Information Processing Systems*, 2014, pp. 2672–2680.

[31] S. Goudarzi, A. Asif, and H. Rivaz, "Multi-focus ultrasound imaging using generative adversarial networks," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2019, pp. 1118–1121.

[32] X. Zhang, J. Li, Q. He, H. Zhang, and J. Luo, "High-quality reconstruction of plane-wave imaging using generative adversarial network," in *Proc. IEEE Ultrason. Symp.*, 2018, pp. 1–4.

[33] J. Youn, M. L. Ommen, M. B. Stuart, E. V. Thomsen, N. B. Larsen, and J. A. Jensen, "Ultrasound multiple point target detection and localization using deep learning," in *Proc. IEEE Ultrason. Symp.*, 2019, pp. 1937–1940.

[34] J. A. Jensen and N. B. Svendsen, "Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 39, no. 2, pp. 262–267, 1992.

[35] J. A. Jensen, "Field: A program for simulating ultrasound systems," *Med. Biol. Eng. Comp.*, vol. 10th Nordic-Baltic Conference on Biomedical Imaging, Vol. 4, Supplement 1, Part 1, pp. 351–353, 1996.

[36] ——, "A multi-threaded version of Field II," in *Proc. IEEE Ultrason. Symp.* IEEE, 2014, pp. 2229–2232.

[37] B. G. Tomov, S. E. Diederichsen, E. V. Thomsen, and J. A. Jensen, "Characterization of medical ultrasound transducers," in *Proc. IEEE Ultrason. Symp.*, 2018, pp. 1–4.

[38] J. A. Jensen, "Safety assessment of advanced imaging sequences, II: Simulations," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 63, no. 1, pp. 120–127, 2016.

[39] A. Gomariz, W. Li, E. Ozkan, C. Tanner, and O. Goksel, "Siamese networks with location prior for landmark tracking in liver ultrasound sequences," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2019, pp. 1757–1760.

[40] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop *et al.*, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2016, pp. 1874–1883.

[41] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Int. Conf. Machine Learning*, 2015, pp. 448–456.

[42] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014.

[43] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *ICML Workshop on Deep Learning for Audio, Speech, and Language Processing*, 2013.

[44] R. Liu, J. Lehman, P. Molino, F. P. Such, E. Frank, A. Sergeev *et al.*, "An intriguing failing of convolutional neural networks and the coordconv solution," in *Neural Information Processing Systems*, 2018, pp. 9605–9616.

[45] M. L. Ommen, M. Schou, R. Zhang, C. A. V. Hoyos, J. A. Jensen, N. B. Larsen *et al.*, "3D printed flow phantoms with fiducial markers for super-resolution ultrasound imaging," in *Proc. IEEE Ultrason. Symp.*, 2018, pp. 1–4.

[46] M. L. Ommen, M. Schou, C. Beers, J. A. Jensen, N. B. Larsen, and E. V. Thomsen, "3D printed calibration micro-phantoms for validation

of super-resolution ultrasound imaging," in *Proc. IEEE Ultrason. Symp.*, 2019, pp. 1212–1215.

[47] G. Montaldo, M. Tanter, J. Bercoff, N. Benech, and M. Fink, "Coherent plane-wave compounding for very high frame rate ultrasonography and transient elastography," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 56, no. 3, pp. 489–506, March 2009.

[48] A. M. Saxe, J. L. McClelland, and S. Ganguli, "Exact solutions to the nonlinear dynamics of learning in deep linear neural networks," arXiv:1312.6120v3 [cs.NE], 2013.

[49] D. Kingma and L. Ba, "Adam: A method for stochastic optimization," arXiv:1412.6980 [cs.LG], 2015.

[50] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2011, software available from tensorflow.org. [Online]. Available: https://www.tensorflow.org/

[51] J. A. Jensen, H. Holten-Lund, R. T. Nilsson, M. Hansen, U. D. Larsen, R. P. Domsten *et al.*, "SARUS: A synthetic aperture real-time ultrasound system," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 60, no. 9, pp. 1838–1852, 2013.

[52] C. Chang, S. Chatterjee, and P. R. Kube, "On an analysis of static occlusion in stereo vision," in *IEEE Conf. Computer Vision and Pattern Recognition*, 1991, pp. 722–723.

[53] P. Fua, "A parallel stereo algorithm that produces dense depth maps and preserves image features," *Mach. Vis. Appl.*, vol. 6, no. 1, pp. 35–49, 1993.

[54] P. Jaccard, "The distribution of the flora in the alpine zone," *New phytologist*, vol. 11, no. 2, pp. 37–50, 1912.

[55] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *IEEE Int. Conf. Computer Vision*, 2017, pp. 2999–3007.

[56] M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, "The importance of skip connections in biomedical image segmentation," arXiv:1608.04117v2 [cs.CV], 2016.

[57] D. H. Simpson, C. T. Chin, and P. N. Burns, "Pulse inversion Doppler: a new method for detecting nonlinear echoes from microbubble contrast agents," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 46, no. 2, pp. 372–382, 1999.

[58] V. Mor-Avi, E. G. Caiani, K. A. Collins, C. E. Korcarz, J. E. Bednarz, and R. M. Lang, "Combined assessment of myocardial perfusion and regional left ventricular function by analysis of contrast-enhanced power modulation images," *Circulation*, vol. 104, no. 3, pp. 352–357, 2001.

[59] O. Solomon, R. Cohen, Y. Zhang, Y. Yang, Q. He, J. Luo *et al.*, "Deep unfolded robust PCA with application to clutter suppression in ultrasound," *IEEE Trans. Med. Imag.*, vol. 39, no. 4, pp. 1051–1063, 2020.

Paper 3

Sub-pixel Accuracy Microbubble Localization using Convolutional Neural Networks

Jihwan Youn, Iman Taghavi, Martin Lind Ommen, Mikkel Schou, Matthias Bo Stuart, Erik Vilain Thomsen, Niels Bent Larsen, Jørgen Arendt Jensen

Name of journal in:
In preparation

Document Version:
In preparation

Sub-pixel Accuracy Microbubble Localization using Convolutional Neural Networks

Jihwan Youn, *Graduate Student Memeber, IEEE*, Iman Taghavi, Martin Lind Ommen, Mikkel Schou, Matthias Bo Stuart*Memeber, IEEE*, Erik Vilain Thomsen, Niels Bent Larsen, and Jørgen Arendt Jensen, *Fellow, IEEE*

***Abstract*— Localizing more microbubbles (MBs) using high concentrations of MBs is preferred as the data acquisition time of ultrasound localization microscopy can be shortened. However, it is challenging for standard methods to localize overlapping point spread functions (PSFs). Recently, several deep learning methods have been proposed to localize the overlapping PSFs, but lack the ability to achieve sub-pixel accuracy localization. This work proposes a way of achieving sub-pixel MB localization with the overlapping PSFs using convolutional neural networks by finding interpolated peaks of Gaussians. On simulated test data, the proposed method achieved precision and recall of 0.93 and 0.83 with localization precision of 35.09 µm laterally and 25.29 µm axially, when centroid detection achieved precision and recall of 0.77 and 0.53 with localization precision of 48.65 µm laterally and 43.13 µm axially. To validate the method on measured data, a phantom that embeds a channel was 3-D printed and scanned with a high concentration of MBs injected into the channel. The proposed method reconstructed the channel successfully with MB contrast of 10.62, a ratio of MBs inside the channel to all the estimated MBs per unit area, whereas centroid detection failed with MB contrast of 0.34. Finally, the proposed method was applied to measurements of a rat kidney at various MB concentrations. In the inner medulla, both methods showed similar results, however, in the outer medulla and cortex, the proposed method was able to achieve higher detection counts of MBs than centroid detection.**

***Index Terms*— Convolutional neural network, localization of high concentration microbubbles, sub-pixel microbubble localization, super-resolution ultrasound imaging, ultrasound localization microscopy**

I. Introduction

SPATIAL resolution of conventional ultrasound systems is limited by wave diffraction. The lateral and axial resolutions are determined by the wavelength, aperture size, pulse length, and imaging sequence, and cannot commonly surpass a half-wavelength. Ultrasound localization microscopy (ULM) is one of the super-resolution ultrasound imaging methods that can break the resolution limit of conventional ultrasound imaging [1]–[6]. ULM can reconstruct an image of microvasculature with a sub-wavelength resolution by localizing microbubbles (MBs), injected into the bloodstream over time, and accumulating their centroids in an image frame.

This work was supported in part by the Fondation Idella.

The authors are with the Department of Health Technology, Technical University of Denmark, 2800 Lyngby, Denmark (email: jihyoun@dtu.dk).

MBs are ultrasound contrast agents that show non-linear behavior when insonified by ultrasound beams. Contrast-enhanced ultrasound (CEUS) imaging such as amplitude modulation [7] or pulse inversion [8] can separate MB signals from stationary echoes by utilizing the non-linearity. Also, MBs can flow through capillaries thanks to their size of several micrometers. Such properties made MBs appropriate for ULM to map the fine structures in the microvasculature. ULM is expected to be practical in clinics for the diagnosis of early-stage cancer [9], ischemic kidney disease [10], and diabetes [11], as well as functional ultrasound [12].

The image quality, i.e., contrast and resolution, of ULM is essentially determined by localization accuracy of estimated MBs [13]. The more accurate MB localization is, the higher resolution ULM can achieve. Centroid detection or Gaussian fitting are commonly used for MB localization [2]–[5], [14]. However, since such methods are poor at localizing overlapping MB point spread functions (PSFs), diluted low concentrations of MBs are commonly employed to avoid the overlaps. As a result, a long data acquisition time is required to fully image the target structure, as the number of MBs that can be localized in a fixed time duration becomes limited.

Lately, deep learning has been applied to ultrasound imaging applications such as beamforming to suppress off-axis scattering [15], reduce speckle noise [16], and calculate content-adaptive weights [17]. There have also been efforts to reconstruct ultrasound images from sub-sampled radiofrequency (RF) data without sacrificing image quality [18]–[20] and to perform clutter filtering using a Robust PCA-based neural network [21]. Correspondingly, deep learning for MB localization has been investigated to deal with the trade-off between the localization accuracy and data acquisition time. Deep-ULM [22] and mSPCN-ULM [23] localized high concentrations of MBs from beamformed ultrasound images. To consider temporal correlation, 3-D convolutional neural networks (CNNs) were applied to a stack of ultrasound images over time as a spatiotemporal filter to remove tissue signals and localize MBs effectively [24] or obtain MB tracks directly without localization [25]. Deep unfolded ULM [26], a model-based neural network [27], has been suggested to improve generalization using a sparsity prior while keeping comparable performance to the fully data-driven methods [28]. However, the aforementioned methods perform localization in the pixel coordinates without sub-pixel accuracy, which

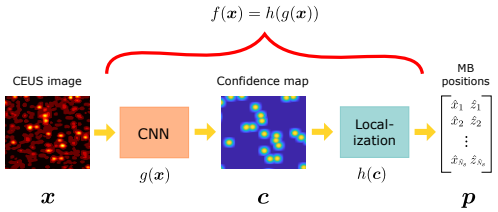


Fig. 1. An overview of the proposed method. A CNN takes a CEUS image $\boldsymbol{x}$ as input and returns a confidence map $\boldsymbol{c}$. The MB positions $\boldsymbol{p}$ are then estimated from the confidence map with sub-pixel accuracy.

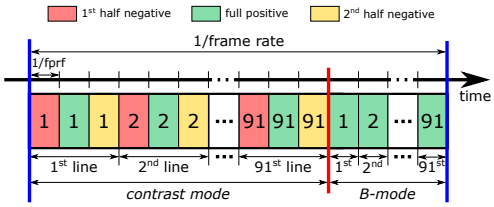


Fig. 2. An overview of B-mode and CEUS imaging sequence for one cycle. Conventional focused beam transmissions using a sliding aperture with 91 sub-apertures was employed. In *contrast mode*, the CEUS imaging was achieved by the amplitude modulation scheme with three transmissions per sub-aperture: one full positive and two half negative transmissions. In *B-mode*, one full positive transmission was employed per sub-aperture. For both *constrast mode* and *B-mode*, a total of 364 transmission events were required for one cycle. The numbers inside transmission events correspond to the sub-aperture index.

means that achievable localization accuracy, i.e., resolution of ULM, is constrained by pixel size. Therefore, the deep learning methods either have additional upsampling layers in their network architecture [22], [23] or are applied to upsampled ultrasound images [24]–[26], [28].

This work proposes a sub-pixel accuracy MB localization method using CNNs that can localize high concentrations of MBs. Specifically, the proposed CNN was trained to return a non-overlapping Gaussian confidence map [29] from an ultrasound image, and sub-pixel localization was performed by applying Gaussian fitting on the peaks in the confidence map. Additional upsampling is not necessary unlike other deep learning methods since sub-pixel localization is available, therefore, computational resources can be managed more efficiently. Also, it is more flexible in a way that the ULM images can be reconstructed in image grids of any pixel size. The proposed CNN was designed based on U-Net [30] with pre-activation residual blocks [31], and training set was generated using Field II pro ultrasound simulation [32]–[34]. In simulation experiments, localization accuracy of the trained network was assessed at various MB densities. Then, phantom experiments were performed using a 3-D printed phantom at two MB concentrations, showing that the generalizability of the CNNs to measured ultrasound data and the ability of the proposed method at a high MB concentration. Finally, the proposed method was validated on *in-vivo* data from animal experiments at 4 different MB concentrations.

II. Methods

Let us consider a CEUS image $\boldsymbol{x} \in \mathbb{R}^{N_z \times N_x}$ induced by MBs located at $\boldsymbol{p} \in \mathbb{R}^{N_{mb} \times 2}$, where N_z and N_x are the number of image samples along the lateral and axial directions, N_{mb} is the number of MBs, and 2 is the number of spatial dimensions (i.e., the lateral and axial positions). The CEUS image can be expressed by

$$\boldsymbol{x} = \sum_{i=1}^{N_{mb}} \mathrm{PSF}(\boldsymbol{p_i}) * \delta(\boldsymbol{p_i}) + \boldsymbol{n}, \tag{1}$$

where $\mathrm{PSF}(\boldsymbol{p_i})$ is the PSF at the i-th MB position, δ is the Dirac delta function, and $\boldsymbol{n}$ is the noise. The goal is to find a mapping $f\colon \mathbb{R}^{N_z \times N_x} \to \mathbb{R}^{N_{mb} \times 2}$ that recovers $\boldsymbol{p}$ from $\boldsymbol{x}$:

$$\boldsymbol{p} = f(\boldsymbol{x}). \tag{2}$$

In this work, the mapping f is composed of two functions g and h to handle sub-pixel localization and varying numbers of MBs depending on the input image. The mapping $g\colon \mathbb{R}^{N_z \times N_x} \to \mathbb{R}^{N_z \times N_x}$ is a CNN that returns a confidence map $\boldsymbol{c} \in \mathbb{R}^{N_z \times N_x}$ and $h\colon \mathbb{R}^{N_z \times N_x} \to \mathbb{R}^{N_{mb} \times 2}$ detects and localizes MBs with sub-pixel accuracy from the confidence map.

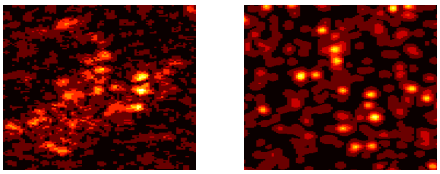
The pipeline of the proposed method is shown in Fig. 1. The method estimates a confidence map from a CEUS image using a CNN, i.e., the mapping g, and localizes MBs with sub-pixel accuracy on the confidence map, i.e., the mapping h. An imaging sequence for CEUS and RF channel data simulation for training, validation, and test sets are explained in Section II-A and II-B, respectively. The CNN architecture is presented in Section II-C. Sub-pixel localization in the confidence map is introduced in Section II-D, and 3-D phantom fabrication for validation is described in Section II-E.

A. Imaging Sequence

A commercial ultrasound system bk5000 (BK Medical, Herlev, Denmark) was used to acquire ultrasound data. Research ultrasound scanners allow to customize the imaging sequence and acquire raw channel data, e.g., synthetic aperture real-time ultrasound system (SARUS) [35] and vantage systems (Verasonics Inc., Redmond, WA, USA). However, in commercial scanners, it is not easy to modify imaging parameters, and the scanners commonly return beamformed ultrasound data only. The imaging sequence implemented in the scanner is illustrated in Fig. 2. The sequence employs focused beam transmissions using a sliding aperture with 91 sub-apertures for both *contrast mode* and *B-mode*. CEUS imaging was achieved using the amplitude modulation [7] scheme in *contrast mode* to isolate non-linear MB signals and reject tissue signals using three transmissions per sub-aperture, i.e., one full positive and two half negative transmissions. A *B-mode* sequence followed using one transmission per sub-aperture. There were 91 sub-apertures, therefore, the total number of transmission events was $91 \times 3 + 91 = 364$. The pulse repetition frequency f_{prf} was 19.6 kHz, which resulted in a frame rate of 53.85 Hz.

TABLE I
FIELD II SIMULATION PARAMETERS

	Parameter	Value
Transducer	# of elements	150
	Pitch	0.16 mm
	Element height	3.40 mm
	Element width	0.15 mm
	Elevation focus	20 mm
Imaging	TX pulse frequency	6 MHz
	# of TX pulse cycle	2
	# of TX elements	25
	Wave type	Focused beam
	Focal depth in TX	10 mm
	Apodization in TX	Boxcar window
Beamforming	Method	Delay-and-sum
	F-number	1
	Apodization in RX	Gaussian window
	Pixel size	24 μm axially
		79 μm laterally
	Region of interest	[0.02, 20.01] mm ax.
		[−10.72, 10.64] mm lat.
Environment	Speed of sound	1540 m/s
	Field II sampling freq.	350 MHz



(a) (b)

Fig. 3. An example of (a) measured and (b) simulated CEUS MB images.

B. Data Generation

Ultrasound data were generated by simulating RF channel data in Field II pro [32]–[34] and beamforming them following the parameter values listed in Table I. It is difficult to obtain true MB positions in measured data, especially when there are many overlapping PSFs, therefore, the simulated data were used for training, validation, and evaluation. The simulation parameter values were chosen from the commercial scanner setting and transducer specification that were used for the ULM experiments.

The diameters of MBs are much smaller than the diffraction limit of ultrasound. The mean diameter of SonoVue (Bracco Imaging, Milan, Italy), the contrast agent used in this work, is 2.5 μm [36] when the wavelengths of typical ultrasound pulses are several hundreds of micrometers. Thus, point scatterers were used to simulate MBs, and the non-linear behavior of MBs was not considered to simplify the simulation model. In measured data, weak scattering was observed apart from MB signals, caused mainly by not rejected stationary echoes, out-of-plane MBs, and low singal-to-noise ratio (SNR) due to electronic noise, as shown in Fig. 3a. To take such noise into consideration, a different kind of point scatterers with a 4 times smaller scattering amplitude than MBs were added. The weak scatterers represent the noise in the measurement

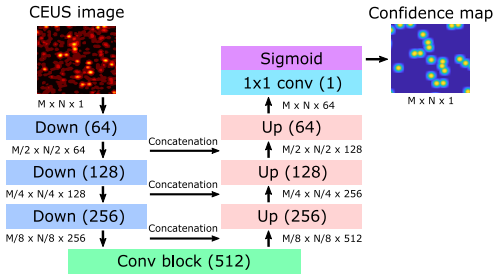


Fig. 4. The proposed U-Net style CNN architecture. The number in the parenthesis are the number of kernels of the corresponding block, and the sets of three numbers represent data size.

and they are not supposed to be localized as MBs.

For an image frame, RF channel data were simulated by placing point scatterers randomly in the region of interest. The simulated RF channel data were then beamformed by delay-and-sum [37], and envelope detection was performed using the Hilbert transform. To further mimic measured data, pixel values of the resulting image was quantized, so that a peak of an isolated PSF has five levels of pixel values. This quantization scheme was chosen empirically from measured CEUS images of MBs. A simulated image frame is shown in Fig. 3b.

C. Convolutional Neural Network

A U-Net [30] style CNN was adopted to estimate a confidence map from a CEUS image, as shown in Fig. 4. The proposed CNN architecture is similar to deep-ULM [22] in the sense that both have an encoder-decoder structure. However, deep-ULM localizes MBs in the pixel coordinates without sub-pixel accuracy, so it requires additional upsampling layers to increase localization accuracy. On the other hand, the proposed method can achieve sub-pixel localization without additional upsampling layers.

The proposed network consists of three *down-blocks*, one *conv-block*, and three *up-blocks*. Those blocks use the pre-activation residual unit [31] to improve the network performance. A detailed description of each block can be found in [29]. In the encoding path, features are extracted from the CEUS image at different scales by the *down-blocks*. In the decoding path, the corresponding confidence map is reconstructed from the representation in the latent space by the *up-blocks*. Skip connections are implemented as concatenation.

D. Confidence Map and Sub-pixel Accuracy Localization

A confidence map represents confidences of MB presence in each pixel location by its pixel values. Localizing MBs in the confidence map allows to handle varying numbers of MBs depending on the input ultrasound image. Especially, non-overlapping Gaussian confidence map has been proposed to localize closely spaced scatterers by identifying local maxima, while providing large gradients for stable training [29]. In the previous work, however, Gaussians were defined in the discrete

image grid, so localization was performed with pixel accuracy. In this work, for sub-pixel localization, Gaussians were created in the continuous domain and sampled according to the image grid coordinates, as described in Algorithm 1. By doing so, not only closely spaced MBs can be localized, but also sub-pixel localization can be achieved. An example of a non-overlapping Gaussian confidence map is shown in Fig. 5a.

Algorithm 1 Confidence map generation

Input: Microbubble positions $\boldsymbol{p}^{mb} \in \mathbb{R}^{N_{mb}\times 2}$, image pixel positions $\boldsymbol{p}^{img} \in \mathbb{R}^{N_z\times N_x}$, and a covariance matrix $\boldsymbol{\Sigma} = \begin{pmatrix}\sigma_z^2 & 0\\ 0 & \sigma_x^2\end{pmatrix}$, where σ_z and σ_x are the standard deviations along the z and x directions.

Output: A non-overlapping Gaussian confidence map $\boldsymbol{c} \in \mathbb{R}^{N_z\times N_x}$

Let's consider a normalized 2-D Gaussian function

$$\mathcal{N}(\boldsymbol{p};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \exp\left\{-\frac{1}{2}(\boldsymbol{p}-\boldsymbol{\mu})^{\intercal}\boldsymbol{\Sigma}^{-1}(\boldsymbol{p}-\boldsymbol{\mu})\right\},$$

where $\boldsymbol{\mu} = (\mu_z, \mu_x)$.

for $k = 1$ to N_{mb} **do**

  $\boldsymbol{c}^k \leftarrow \left\{(c^k_{i,j}) \in \mathbb{R}^{N_z\times N_x} \,\middle|\, c^k_{i,j} = \mathcal{N}(\boldsymbol{p}^{img}_{i,j}; \boldsymbol{p}^{mb}_{k,*}, \boldsymbol{\Sigma})\right\}$

end for

$\boldsymbol{c} \leftarrow \left\{(c_{i,j}) \in \mathbb{R}^{N_z\times N_x} \,\middle|\, c_{i,j} = \max_{k\in[1,N_{mb}]} c^k_{i,j}\right\}$.

In the non-overlapping Gaussian confidence map, pixel values around a local peak follows a Gaussian function thanks to the maximum operation in the generation of the confidence map. Based on this fact, a MB position can be localized by fitting a Gaussian function to a local maximum and its neighboring pixels in the confidence map. Essentially, the center of the Gaussian can be estimated in the continuous spatial domain, which corresponds to the sub-pixel position of a MB. The procedure of sub-pixel localization in a confidence map is described in Algorithm 2.

Algorithm 2 MB localization from a confidence map

Input: A non-overlapping Gaussian confidence map $\boldsymbol{c} \in \mathbb{R}^{N_z\times N_x}$ and image pixel positions $\boldsymbol{p}^{img} \in \mathbb{R}^{N_z\times N_x}$.

Output: Estimated MB positions $\hat{\boldsymbol{p}}^{mb} \in \mathbb{R}^{N_{mb}\times 2}$

$\hat{\boldsymbol{p}}^{mb} \leftarrow \{\,\}$

for $i = 2$ to $N_z - 1$ **do**

  for $j = 2$ to $N_x - 1$ **do**

    if $c_{i,j} = \max\{c_{i-1,j}, c_{i,j-1}, c_{i,j}, c_{i+1,j}, c_{i,j+1}\}$ **then**

      $\hat{p} \leftarrow fitGaussian\left(i, j, \boldsymbol{c}, \boldsymbol{p}^{img}\right)$

      $\hat{\boldsymbol{p}}^{mb}$.insert($\hat{p}$)

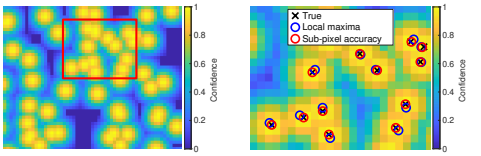
    end if

  end for

end for

The Gaussian fitting can be performed as follows. Let us consider N data points $\{(y_i, x_{i1}, x_{i2})\}_{i=1}^{N}$ that follow an 2-D Gaussian function

$$y = \exp\left\{-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)\right\}, \quad (3)$$



(a) (b)

Fig. 5. An example of the non-overlapping Gaussian confidence map with true and estimated scatterer positions.

where $\boldsymbol{\mu} = (\mu_1, \mu_2)$ is the center and $\boldsymbol{\sigma} = (\sigma_1, \sigma_2)$ is the standard deviation of the Gaussian function. By taking natural logarithms in (3), the following can be obtained,

$$\ln y = ax_1^2 + bx_2^2 + cx_1 + dx_2 + e, \quad (4)$$

where $a = \frac{-1}{2\sigma_1^2}$, $b = \frac{-1}{2\sigma_2^2}$ $c = \frac{\mu_1}{\sigma_1^2}$, $d = \frac{\mu_2}{\sigma_2^2}$, and $e = -\left(\frac{x_1^2}{2\sigma_1^2} + \frac{x_2^2}{2\sigma_2^2}\right)$. By using the data points and (4), a linear regression can be formalized as follows,

$$\begin{pmatrix} x_{11}^2 & x_{12}^2 & x_{11} & x_{12} & 1\\ x_{21}^2 & x_{22}^2 & x_{21} & x_{22} & 1\\ \cdots & & & & \\ x_{N1}^2 & x_{N2}^2 & x_{N1} & x_{N2} & 1 \end{pmatrix}\begin{pmatrix} a\\ b\\ c\\ d\\ e\end{pmatrix} = \begin{pmatrix}\ln y_1\\ \ln y_2\\ \cdots\\ \ln y_N\end{pmatrix}, \quad (5)$$

the analytic solution can be found as follows,

$$\begin{pmatrix} a\\ b\\ c\\ d\\ e\end{pmatrix} = \begin{pmatrix} x_{11}^2 & x_{12}^2 & x_{11} & x_{12} & 1\\ x_{21}^2 & x_{22}^2 & x_{21} & x_{22} & 1\\ \cdots & & & & \\ x_{N1}^2 & x_{N2}^2 & x_{N1} & x_{N2} & 1 \end{pmatrix}^{-1}\begin{pmatrix}\ln y_1\\ \ln y_2\\ \cdots\\ \ln y_N,\end{pmatrix}, \quad (6)$$

and the center of the Gaussian can be found as

$$\mu_1 = -c/2a \quad \text{and} \quad \mu_2 = -d/2b. \quad (7)$$

To estimate the peak of a 2-D Gaussian function, at least five data points are necessary. In this work, the local maximum along with its 4 adjacent pixel values, i.e., 5 pixels, were used to fit a Gaussian function.

E. Phantom Fabrication

For the validation of ULM on measured data, a PEGDA 700 g/mol phantom was 3-D printed using stereolithography [38]. The phantom had a channel with a diameter of 200 µm, where MBs can be injected through a syringe. The channel was designed to be bent 90 degrees multiple times on an imaging plane, as illustrated in Fig 6. This design resulted pairs of parallel channels where, in each pair, the MBs flowed in opposite directions to each other. The wall-to-wall spacing between the pairs was varied from 22 to 120 µm to compare various ULM methods at different spacing. From left to right, the spacing decreased from 121 µm to 22 µm and again increased to 110 µm to maintain the stability during 3-D printing.

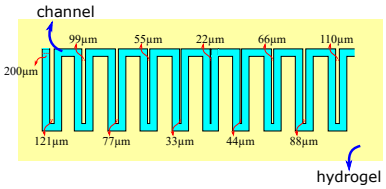


Fig. 6. Layout of the 3-D printed phantom embedding a channel to validate ULM.

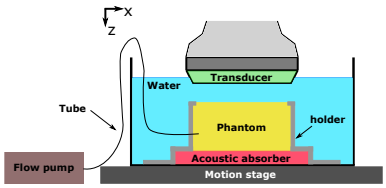


Fig. 7. Illustration of phantom experiment setup

III. Experiments

A. Training Detail

A CNN, the mapping g with its learning parameters θ, was trained to obtain the model that returns the confidence map $\boldsymbol{c}$ given the ultrasound image $\boldsymbol{x}$. For training, the difference between true and estimated confidence maps was captured by the mean squared error (MSE) which is given by

$$\mathcal{L}_{\mathrm{MSE}}\left(\boldsymbol{x}, \boldsymbol{c}; g\right) = \frac{1}{N}\sum_{i=1}^{N}\left\|\boldsymbol{c}_i - g\left(\boldsymbol{x}_i; \theta\right)\right\|_F^2, \tag{8}$$

where N is the number of samples and $\|\cdot\|_F$ is the Frobenius norm.

The learning parameters were initialized by orthogonal initialization [39] and updated using the Rectified Adam (RAdam) [40] and LookAHead [41] optimizer. The learning rate was initially set to 0.0001 and halved every 200 epochs. The CNN model was implemented using Tensorflow [42] in Python. A server equipped with a NVIDIA TESLA V100 16 GB PCIe graphics card was used for training. The total number of training epochs was 1000 and the training took approximately 24 hours.

The training set consisted of 3840 MB images and 3840 weak scattering images, i.e., noise, that were simulated separately. One MB image was simulated with 400 point scatterers, and one weak scattering image was simulated with 4000 point scatterers. All MB scattering had the same amplitude and the weak scattering amplitudes were a quarter of that. During training, one ultrasound image frame was formed by selecting one MB image and one weak scattering image randomly from the training set. And then, the frame was randomly cropped to a size of 128×128 and was flipped along the lateral direction with a probability of 0.5 for data augmentation. The degree of the MB overlap and essential MB density in the cropped region varied due to the large initial imaging region even though the MB images were simulated with the same number of MBs. The ultrasound image and confidence map were lastly normalized to be in $[0, 1]$.

Validation was performed at the end of every epoch to monitor the training and choose hyper-parameters. The validation set was created in the same way as the training set with 128 MB images and 128 noise images.

B. Simulation Experiment

The trained CNN was firstly evaluated on simulated test sets. Each test set was composed of 128 frames that were simulated at the same MB density and a total of 9 test sets were created by varying the MB density from $0.3\,\mathrm{mm}^{-2}$ to $4.9\,\mathrm{mm}^{-2}$. Unlike the training data that cover the whole imaging region, the test data were simulated in a 128×128 region, so that the similar degree of the MB overlap can appear at the same MB density. The 128×128 region was selected randomly for each data since the PSFs vary spatially. The number of MBs were changed depending on the test set, but the number of weak scatterers was not.

C. Phantom Experiment

The 3-D printed phantom was measured using the bk5000 ultrasound system and X18L5s linear array transducer (BK Medical, Herlev, Denmark). The imaging sequence of the scanner is shown in Fig. 2, and the specification of the transducer follows the parameter values in Table I. The phantom was submerged in a water tank and fixed by a 3-D printed holder, as illustrated in Fig. 7. The water tank was placed on a motion stage that helps align the channel inside the phantom in the imaging plane. SonoVue (Bracco Imaging SpA, Milan, Italy) was diluted into two concentrations: 1:40 (low) and 1:20 (high), and then injected into the channel using a syringe and a syringe pump neMESYS 290N (Centoni, Thuringia, Germany) with a constant volume flow rate of $1\,\mu\mathrm{L/s}$.

D. Animal Experiment

Animal experiment was performed on a healthy male Sprague-Dawley rat according to the protocols approved by the Danish National Animal Experiments Inspectorate. The procedures were conducted at the University of Copenhagen, following all local ethical standards. The ethical policy adheres to that of the National Institutes of Health. The animal was housed in an animal facility under the supervision of trained animal caretakers at the Department of Experimental Medicine, University of Copenhagen.

Prior to ultrasound scans, the rat was anesthetized with $5\,\%$ isoflurane and placed on a heating pad ($37\,^{\circ}\mathrm{C}$) to keep body temperature. Through tracheotomy, a mechanical ventilator (Ugo Basile, Gemonio, Italy) was connected to the rat to control respiration with a cycle of $72\,\mathrm{breaths/min}$ and anesthesia was maintained with $1{\sim}2\,\%$ isoflurane. Jugular vein was catheterized to provide $0.85\,\mathrm{mg/mL}$ cisatracurium (Nimbex; GlaxoSmithKline, Brentford, UK) at $20\,\mu\mathrm{L/min}$. The arterial blood pressure was monitored in the left carotid artery using a pressure transducer P23Db (Gould Statham Instr. Inc., CA, USA). The left kidney was exposed through laparotomy with

TABLE II
MB CONCENTRATIONS FOR ANIMAL EXPERIMENTS.

Experiment	MB dilution	Volume flow rate	MB concentration
Scenario 1	1:20	85 μL/s	4
Scenario 2	1:20	170 μL/s	9
Scenario 3	1:10	170 μL/s	17
Scenario 4	1:5	170 μL/s	34

TABLE III
COMPARISONS OF PRECISION, RECALL, AND LOCALIZATION PRECISION ON SIMULATED TEST SETS.

Method	Precision	Recall	localization precision (lat./ax.)
Centroid	0.77	0.53	48.65 μm / 43.13 μm
Proposed	**0.93**	**0.83**	**35.09** μm / **25.29** μm

the rat in the supine position. The diaphragm was pulled cranially with a retractor to further expose the kidney and reduce respiratory motion. The rat kidney was scanned at 4 different MB concentrations by adjusting MB dilution and infusion volume flow rate, as shown in Table II. The MB concentration was assumed to be linearly proportional to MB dilution and volume flow rate. The scan for each concentration lasted for 9 minutes, and the rat was euthanized in anesthesia after the experiments.

On the animal measurements, motion correction was applied using B-mode images [43]. Local motion was estimated by dividing the B-mode images into small patches and co-registering them to a reference image frame using 2-D cross-correlation. The locally estimated motion was then interpolated to a finer grid, and its inverse transform was applied to the estimated MB positions for motion correction.

E. Evaluation Metrics

For the simulation experiments, evaluation was performed by calculating precision, recall or reconstructed MB density, and localization precision in the lateral and axial directions. Precision P, recall R, and reconstructed MB density $\hat{d}_{mb}$ are given by

$$P = \frac{TP}{TP + FP}, \quad R = \frac{TP}{TP + FN}, \text{ and} \tag{9}$$

$$\hat{d}_{mb} = R \times d_{mb}, \tag{10}$$

where TP is the number of true positives (correct MB localization), FP is the number of false positives (wrong MB localization), FN is the number of false negatives (missed MBs), and d_{mb} is the true MB density. Localization precision σ_{loc} was measured by

$$\sigma_{loc} = 2\sqrt{2 \ln 2}\,\sigma, \tag{11}$$

where σ is the standard deviation of distance errors between true and estimated MBs. The localization precision (11) is the full width at half maximum (FWHM) of a Gaussian, therefore, it can be translated to the resolution of ULM, assuming that the localization errors follow a Gaussian distribution.

To determine whether the estimated localization is correct or wrong, true and estimated MBs need to be paired. Simply finding the nearest true MB given an estimated MB is problematic since that can lead to a situation where one true MB is paired with more than one estimated MB [29]. Therefore, the matching problem was formulated as a linear assignment problem and solved by the built-in function *matchpairs* in MATLAB (MathWorks, MA, USA) [44]. A cost matrix was defined by pairwise distances of all possible assignments between the true and estimated MBs, and the optimal matches that give the minimum cost assignment were obtained. The threshold of not matching was also employed to reject the assignments whose assignment costs (i.e., distance) are higher than the threshold. In this work, the threshold was set to $\lambda/5$ (49 μm) for precision, recall, and reconstructed MB density and $\lambda/2$ (123 μm) for localization precision.

For the assessment of the phantom measurements, the above metrics are not appropriate due to the lack of true MB positions. Instead, the dimensions of channels inside the 3-D phantom where MBs are injected were used. The ratio of the number of MBs per unit area inside channels to the number of MBs per unit area outside channels, termed as MB contrast ratio, was calculated for evaluation in the following way:

$$CR_{mb} = \frac{N_{mb,ch}\,(A_{tot} - A_{ch})}{(N_{mb,tot} - N_{mb,ch})\,A_{ch}}, \tag{12}$$

where $N_{mb,tot}$ and A_{tot} are the total number of MBs in an image region and the area of the region, and $N_{mb,tot}$ and A_{ch} are the number of MBs inside channels and the area of channels in the region.

For the animal experiment, it is difficult to evaluate the results due to the absence of ground truth. Hence, tracking was applied by Kalman filtering on the estimated MBs and the assessment was performed on the MBs that contribute to tracks, which will be referred to as track samples. Specifically, unlike single model Kalman filtering such as [45], the Kalman filter was implemented in a hierarchical way by exploiting multiple models for each velocity range. The concept of the hierarchical Kalman filter was discussed in [46]. Tracking is an important step along with localization in ULM [5], [6], which filters out wrong localization by taking temporal correlation into account, thereby improving image quality and providing velocity estimation of microvessels [47]. The number of the track samples and the distance between the closest track samples at each frame were calculated, assuming track samples are correct localization.

IV. RESULTS

A. Simulation Experiment

Localization capability of centroid detection and the proposed method on the simulated test sets is shown in Table III. The proposed method achieved better performance for all metrics than centroid detection. The performance of each method at different MB densities is also shown in Fig. 8. In general, localization performance deteriorated as the MB density increased regardless of the methods. The proposed method, however, suffered less than centroid detection by its ability to localize overlapping MBs.

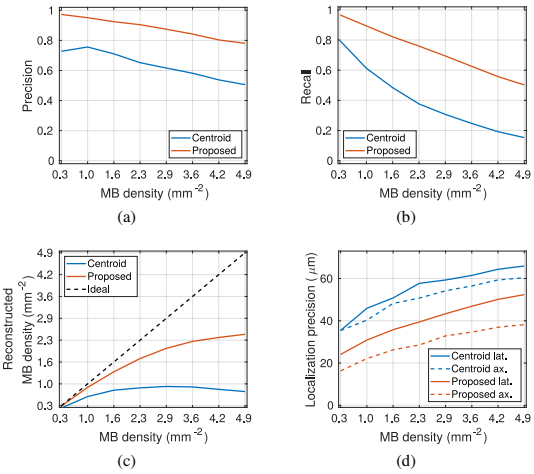


Fig. 8. Localization capability of centroid detection and the proposed method over different MB densities on the simulated test sets: (a) precision, (b) recall, (c) reconstructed MB density, and (d) localization precision in the lateral and axial directions.

Localizing more MBs within an image frame allows to shorten the data acquisition time of ULM. However, simply increasing MB concentrations will not help centroid detection shorten the data acquisition time as the reconstructed MB density peaked to $0.9\,\mathrm{mm}^{-2}$ at the MB density of $2.9\,\mathrm{mm}^{-2}$ and started to decrease. However, the reconstructed MB density of the proposed method kept increasing and converged to $2.4\,\mathrm{mm}^{-2}$. The higher reconstructed MB density, seen in Fig. 8c, essentially indicates the proposed method can map the target structure with fewer image frames by localizing more MBs.

B. Phantom Experiment

The ULM results of the 3-D printed phantom measurements at two different MB concentrations are shown in Table IV and Fig. 9. For ULM image reconstruction, 3000 frames for the *low* concentration and 800 frames for the *high* concentration were used. Even though a constant concentration of MBs were infused, the concentration decreased inside the phantom channel as they flowed from left to right since more MBs were disrupted as exposed by more ultrasound beams. The MB contrast ratio was, therefore, calculated on the first 6 pairs of channels from the left to measure it under similar MB concentrations.

Centroid detection and the proposed method showed similar results at the *low* concentration (1:40). Both resolved well all the pairs of channels in the phantom, as shown in Fig. 9a (first row). The mean MB contrast (Table IV) and the MB contrast ratio at each pair (Fig. 9b) were higher than 1, which is the necessary condition for resolving a pair of channels. The lateral intensity profile of the most closely spaced pair (the sixth pair from the left) in Fig. 9c confirmed that the pair with the wall-to-wall distance of $22\,\mu\mathrm{m}$ was clearly separated

TABLE IV

COMPARISONS OF THE NUMBER OF ESTIMATED MBS PER FRAME AND MEAN MB CONTRAST RATIO OVER ALL THE PAIRS OF CHANNELS ON THE PHANTOM MEASUREMENTS.

Concentration	Method	# estimated MBs per frame	mean MB contrast ratio
Low (1:40)	Centroid	19	3.67
	Proposed	**32**	**3.75**
High (1:20)	Centroid	40	0.34
	Proposed	**124**	**10.62**

by both methods. Nonetheless, the proposed method localized more MBs by a factor of 1.7 per frame, which agrees with the higher recall or reconstructed MB density at a low MB density in the simulation.

At the *high* concentration (1:20), the proposed method still resolved all the channels clearly, but centroid detection failed, as shown in Fig. 9a (second row). Centroid detection resulted in single channels for all the pairs of channels. The mean MB contrast ratio of centroid detection was 0.34 and the MB contrast ratio at each pair were below 1, which is the sufficient condition that a pair of channels is not resolved. The lateral intensity profile in Fig. 9c (second row) clearly shows a high peak in the center of the channels, where no MBs were supposed to be localized. However, the proposed method achieved MB contrast ratio higher than 1 at all the pairs, and the most closely spaced channels were also well resolved as shown in Fig. 9c (second row). Compared to the *low* concentration result, the proposed method localized more MBs by a factor of 2.8 per frame at the *high* concentration, and it achieved a comparable ULM image with 3.8 times less image frames.

C. Animal Experiment

The ULM images of a rat kidney at 4 different MB concentrations using centroid detection and the proposed method are shown in Fig. 11a. Both methods produced high-resolution images by resolving the microvasculature of the kidney, and the results were consistent with each other in *scenario 1*. In *scenario 2*, when the MB concentration was doubled, more microvessels were developed by both methods, however, the proposed method achieved mostly brighter intensity by localizing more MBs from the same data. As the MB concentration further increased, centroid detection started to fail in *scenario 3* by losing many microvessels and failed to reconstruct most microvessels except the inner medulla region in *scenario 4*. On the other hand, the proposed method still showed a decent result in *scenario 3* though some vessels in the cortex area were missed. In *scenario 4*, the proposed method also failed to reconstruct microvessels properly except some region in the inner medulla, but the overall shape was still perceptible and relatively large vessels were visible.

The ability of how closely spaced MBs can be resolved has implicitly been investigated by measuring the smallest pairwise distances among track samples in a frame. The histogram in Fig. 10 shows the normalized counts of the smallest pairwise distances on the *scenario 2* rat data in bins

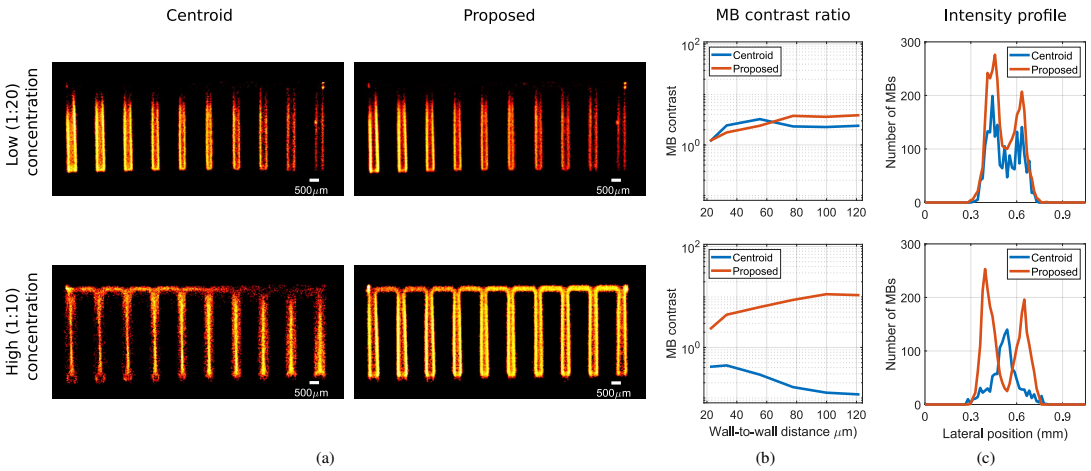


Fig. 9. Comparison of ULM results on 3-D phantom measurements at a low (first row) and a high (second row) MB concentrations. (a) is the ULM reconstruction by centroid detection (first column) and the proposed method (second column). (b) is the MB contrast ratio at the first 6 pairs of channels from the left. (c) shows the lateral intensity profile of the most closely spaced pair, the sixth from the left in (a).

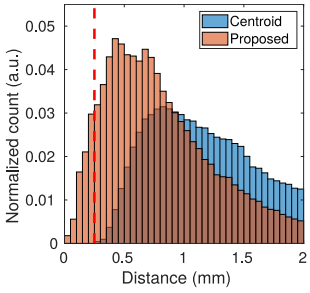


Fig. 10. A histogram that shows how closely spaced MBs were able to be localized by centroid detection and the proposed method. The minimum of pairwise distances were found in each frame in the rat measurements of *scenario 2*. The red dashed line represents 250 μm ($\approx \lambda$).

of $50\,\mu\mathrm{m}$. The counts were normalizing by the total number of counts, so the normalized count does not reflect the real number of counts but the ratio of the counts in each method. For the proposed method, $8\,\%$ of the estimated track samples were closer than $250\,\mu\mathrm{m}$ ($\approx \lambda$), represented as a red vertical dashed line, while there were no such track samples for centroid detection.

Three $1.6\,\mathrm{mm} \times 1.6\,\mathrm{mm}$ regions were selected from the inner medulla, outer medulla, and cortex to further analyze the effect of the MB concentrations locally. The selected regions are highlighted as blue rectangles in Fig. 11a. The local ULM results and the number of track samples in the regions at the different MB concentrations are shown in Fig. 11b, 11c, and 11d. For inner medulla (Fig. 11b), both methods showed similar trend when the MB concentration increased. The number of track samples kept increasing up to *scenario 3* and dropped in *scenario 4*. For the outer medulla (Fig. 11c) and cortex (Fig. 11d), the proposed method localized more MBs over all MB concentrations. In addition, as the MB concentration increased, the proposed method acquired more track samples up to *scenario 3* and *scenario 2* for the outer medulla and cortex, respectively, where the number of track samples by centroid detection started decreasing.

V. Discussion

In this work, a sub-pixel MB localization method that can handle overlapping MBs using CNNs has been proposed. The CNN was trained to learn non-overlapping Gaussian confidence maps, instead of MB positions, from CEUS images. And then, the sub-pixel localization was performed by applying Gaussian fitting to the local peaks in the confidence maps. The method was evaluated on simulation data, phantom measurements, and animal measurements at various MB concentrations, showing that the proposed method can separate MBs that were spaced closer than the resolution of conventional ultrasound imaging without being limited to the input image pixel size.

The CNN for the proposed method needs the capability of reconstructing the non-overlapping Gaussian maps properly for accurate localization. To select an appropriate deep learning model, validation was performed on the U-Net and ResNet style architecture. And the U-Net style CNN showed better confidence map estimation on the validation set, resulting in more accurate localization. It has been reported in [23] that the ResNet style architecture showed improvement of localization in the pixel coordinates. However, for sub-pixel localization via the non-overlapping Gaussians, it is complex to reconstruct the confidence maps, so the U-Net style network worked better thanks to its encoder-decoder structure.

The proposed method utilizes computational resources efficiently by virtue of the sub-pixel accuracy processing as

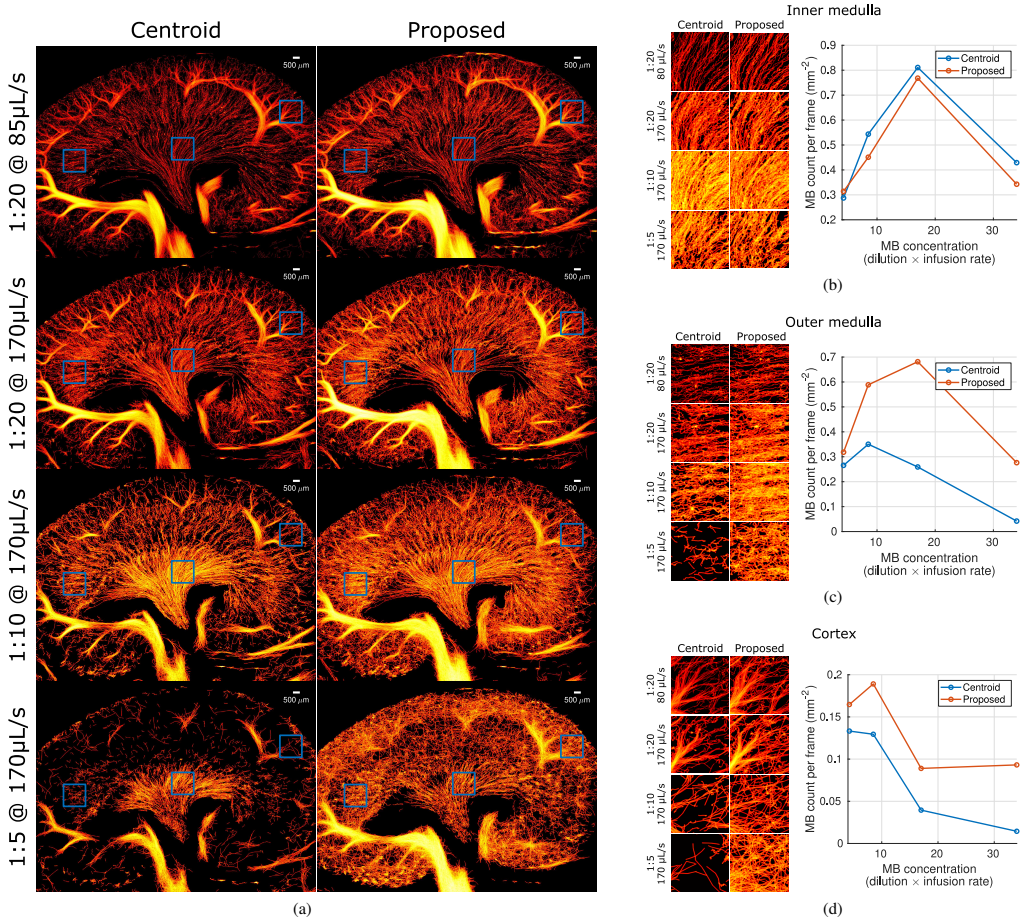


Fig. 11. Comparison of ULM results on the rat kidney measurements. (a) is the ULM reconstruction at 4 different MB concentrations by centroid detection (first column) and the proposed method (second column). Three 1.6 mm × 1.6 mm regions were selected from the inner medulla, outer medulla, and cortex for local analysis, which were highlighted by blue rectangles in (a). The local ULM image results and the number of track samples over different MB concentrations are shown in (b) for the inner medulla, (c) for the outer medulla, and (d) for the cortex.

additional upsampling layers are not necessary and localization is performed on the same image resolution to the input image. The computational complexity of Deep-ULM, mSPCN-ULM, and the proposed method given ultrasound images with a size of 786×272, the same image size with the phantom and animal measurements, were investigated. The number of model parameters and the number of floating point operations (FLOPs) were calculated manually. The process time was measured for one image frame by repeating inference for 1000 times. The maximum available batch size, which determines the number of image frames can be processed in a single iteration, was obtained by increasing the batch size in powers of two until running out of GPU memory. The PC equipped with a NVIDIA Titan V graphics card was used for the computational complexity evaluation, and the results are shown in Table V. Deep-ULM and the proposed method have similar architecture except the additional upsampling layers, thereby both require similar number of parameters and FLOPs. However, the proposed method was faster by a factor of 2.3 for processing one image frame and can deal with roughly 2^3 times more images in a batch. Considering current ULM processing is mostly performed off-line, larger batch size is beneficial as it allows to process more image frames in parallel. Contrarily, the number of parameters for mSPCN-ULM was much less because mSPCN-ULM follows a ResNet style architecture. Nonetheless, the number of FLOPs was much larger since the size of feature maps are kept in the same image resolution with the input before the additional upsampling layers due to the lack of pooling and unpooling operations. The proposed method achieved comparable process

TABLE V
COMPARISON OF COMPUTATIONAL COMPLEXITY GIVEN AN ULTRASOUND IMAGE WITH A SIZE OF $\mathbf{768 \times 272}$.

Model	Number of parameters	Number of FLOPs	Process time	Maximum batch size
Deep-ULM	5.9×10^6	29×10^9	355 ms	2^3
mSPCN-ULM	0.4×10^6	63×10^9	158 ms	2^6
Proposed	5.8×10^6	23×10^9	156 ms	2^6

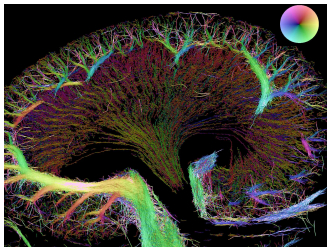


Fig. 12. ULM track image generated from the rat kidney measurement of *scenario 3* using the proposed localization method and the hierarchical Kalman filter. The color wheel on the top right corner represents the magnitude and direction of the velocity.

time to mSPCN-ULM with 15 times more model parameters. Note that the number of FLOPs, process time, and maximum batch size depend on the input image size, while the number of parameters does not.

Unlike [48] and [25] that produce super-resolved images directly, the proposed method performs localization explicitly at high concentrations of MBs. The access to the estimated MB positions allows to make tracks out of them by taking temporal consistency into account. An example of the track image from the rat kidney measurement in *Scenario 2* the *middle* concentration using the proposed localization method is shown in Fig. 12. Tracking helps filter out wrong localization, separates entangled microvessels which is impossible in MB intensity images, and, more importantly, provides velocity information. The velocity information is one of the useful ULM quantities that helps doctors diagnose diseases in clinics [49].

The phantom experiment results clearly showed that the proposed method could localize closely spaced MBs on the measured data when centroid detection could not. The phantom channel whose wall-to-wall distance varied from 22 to 121 µm was successfully reconstructed at the *high* MB concentration. The image quality at the *high* MB concentration was comparable to that at the *low* MB concentration in terms of the number of localized MBs and MB contrast ratio with 6 times less image frames. It also looked promising that the proposed method can localize the overlapping PSFs on the *in-vivo* data, as shown in Fig. 10. Yet, it is uncertain if the track samples are true MBs, so further investigation will be necessary, e.g., cross-modality validation for a concrete conclusion.

In *in-vivo* scenarios, perfusion, vessel size, and microvascular structure, as well as infused MB concentrations and MB disruption affect on the actual MB concentration in local regions. Even for the same localization method, the trend of the number of track samples was different in each local region, as shown in Fig. 11b, 11c, and 11d. The actual MB concentration was higher in the cortex than in the inner medulla even though the infused MB concentration is the same. This explains why the number of track samples started to decrease at different MB concentrations as the degree of overlapping that the proposed method can handle is limited, although it can localize overlapping PSFs. Therefore, ULM at a high MB concentration can give a different image quality depending on the target structure, and the concentration should be selected based on the region of interest and the application.

The imaging sequence used in this work was limited and not optimal to achieve good transmission focusing, a high frame rate, and a high SNR. An advanced imaging sequence e.g., synthetic aperture imaging using diverging wave [50] or plane wave [51] would improve localization results dramatically by offering a better image quality. Furthermore, 2-D ULM using 1-D array probes has a problem that the elevation direction is not recognized. The 2-D ultrasound images are essentially an integration over the elevation beam profile, determined by elevational resolution. This is fine for simple-structured targets such as the channel of the 3-D printed phantom in Fig 6 as the channel was virtually a 2-D structure well aligned in the imaging plane using the motion stage. However, *in-vivo* targets have much more complicated structure. There are microvessels in the out of imaging plane direction and microvessels flowing in different directions can lie on top of each other in the elevation direction, which hinder accurate localization, motion correction, and tracking. Those limitations can be solved by 3-D ULM using 2-D array probes such as fully-addressed matrix array probes [52], [53] or row-column (RC) addressed matrix array probes [54]. Especially, the RC probes only require $2N$ connections compared to N^2 of the fully-addressed probes. The proposed method can be extended for 3-D data by implementing the CNN and non-overlapping Gaussian confidence maps in 3-D. Therefore, it is expected that 3-D ULM using a synthetic aperture sequence with the RC probe will allow the proposed method to achieve better localization performance by removing the ambiguity of the data in the elevation direction, and as a result, high-fidelity ULM reconstruction.

VI. CONCLUSION

A sub-pixel MB localization method using a CNN has been proposed. The CNN was trained to learn the mapping from a CEUS image to a non-overlapping Gaussian confidence map and sub-pixel localization was performed by applying Gaussian fitting on the local peaks. The method was evaluated on the simulation data, phantom measurements, and animal measurements, showing overlapping PSFs spaced closer than the ultrasound resolution limit can separated. This method can achieve ULM at a higher MB concentration with a shorter data acquisition time, and this will potentially help making ULM more feasible in clinics.

ACKNOWLEDGMENT

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan V Volta GPU used for this research.

REFERENCES

[1] O. Couture, B. Besson, G. Montaldo, M. Fink, and M. Tanter, "Microbubble ultrasound super-localization imaging (MUSLI)," in *Proc. IEEE Ultrason. Symp.*, 2011, pp. 1285–1287.

[2] M. Siepmann, G. Schmitz, J. Bzyl, M. Palmowski, and F. Kiessling, "Imaging tumor vascularity by tracing single microbubbles," *Proc. IEEE Ultrason. Symp.*, pp. 6 293 297, 1906–1908, 2011.

[3] O. M. Viessmann, R. J. Eckersley, K. Christensen-Jeffries, M. X. Tang, and C. Dunsby, "Acoustic super-resolution with ultrasound and microbubbles," *Phys. Med. Biol.*, vol. 58, pp. 6447–6458, 2013.

[4] M. A. O'Reilly and K. Hynynen, "A super-resolution ultrasound method for brain vascular mapping," *Med. Phys.*, vol. 40, no. 11, pp. 110 701–7, 2013.

[5] K. Christensen-Jeffries, R. J. Browning, M. Tang, C. Dunsby, and R. J. Eckersley, "In vivo acoustic super-resolution and super-resolved velocity mapping using microbubbles," *IEEE Trans. Med. Imag.*, vol. 34, no. 2, pp. 433–440, February 2015.

[6] C. Errico, J. Pierre, S. Pezet, Y. Desailly, Z. Lenkei, O. Couture *et al.*, "Ultrafast ultrasound localization microscopy for deep super-resolution vascular imaging," *Nature*, vol. 527, pp. 499–502, November 2015.

[7] V. Mor-Avi, E. G. Caiani, K. A. Collins, C. E. Korcarz, J. E. Bednarz, and R. M. Lang, "Combined assessment of myocardial perfusion and regional left ventricular function by analysis of contrast-enhanced power modulation images," *Circulation*, vol. 104, no. 3, pp. 352–357, 2001.

[8] D. H. Simpson, C. T. Chin, and P. N. Burns, "Pulse inversion Doppler: a new method for detecting nonlinear echoes from microbubble contrast agents," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 46, no. 2, pp. 372–382, 1999.

[9] F. Lin, S. E. Shelton, D. Espindola, J. D. Rojas, G. Pinton, and P. A. Dayton, "3-D ultrasound localization microscopy for identifying microvascular morphology features of tumor angiogenesis at a resolution beyond the diffraction limit of conventional ultrasound," *Theranostics*, vol. 7, no. 1, pp. 196–204, 2017.

[10] S. B. Andersen, I. Taghavi, C. A. V. Hoyos, S. B. Søgaard, F. Gran, L. Lonn *et al.*, "Super-resolution imaging with ultrasound for visualization of the renal microvasculature in rats before and after renal ischemia: A pilot study," *Diagnostics*, vol. 10, no. 11, p. 862, 2020.

[11] D. Ghosh, J. Peng, K. Brown, S. Sirsi, C. Mineo, P. W. Shaul *et al.*, "Super-resolution ultrasound imaging of skeletal muscle microvascular dysfunction in an animal model of type 2 diabetes," *J. Ultrasound Med.*, vol. 38, no. 10, pp. 2589–2599, 2019.

[12] T. Deffieux, C. Demene, M. Pernot, and M. Tanter, "Functional ultrasound neuroimaging: a review of the preclinical and clinical state of the art," *Curr. Opin. Neurol.*, vol. 50, pp. 128–135, 2018.

[13] K. Christensen-Jeffries, O. Couture, P. A. Dayton, Y. C. Eldar, K. Hynynen, F. Kiessling *et al.*, "Super-resolution ultrasound imaging," *Ultrasound Med. Biol.*, vol. 46, no. 4, pp. 865–891, 2020.

[14] P. Song, A. Manduca, J. D. Trzasko, R. E. Daigle, and S. Chen, "On the effects of spatial sampling quantization in super-resolution ultrasound microvessel imaging," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 65, no. 12, pp. 2264–2276, 2018.

[15] A. C. Luchies and B. C. Byram, "Deep neural networks for ultrasound beamforming," *IEEE Trans. Med. Imag.*, vol. 37, no. 9, pp. 2010–2021, 2018.

[16] D. Hyun, L. L. Brickson, K. T. Looby, and J. J. Dahl, "Beamforming and speckle reduction using neural networks," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 66, no. 5, pp. 898–910, 2019.

[17] B. Luijten, R. Cohen, F. J. De Bruijn, H. A. W. Schmeitz, M. Mischi, Y. C. Eldar *et al.*, "Adaptive ultrasound beamforming using deep learning," *IEEE Trans. Med. Imag.*, vol. 39, no. 12, pp. 3967–3978, 2020.

[18] Y. H. Yoon, S. Khan, J. Huh, and J. C. Ye, "Efficient b-mode ultrasound image reconstruction from sub-sampled rf data using deep learning," *IEEE Trans. Med. Imag.*, 2018.

[19] I. A. M. Huijben, B. S. Veeling, K. Janse, M. Mischi, and R. J. G. van Sloun, "Learning sub-sampling and signal recovery with applications in ultrasound imaging," *IEEE Trans. Med. Imag.*, vol. 39, no. 12, pp. 3955–3966, 2020.

[20] S. Khan, J. Huh, and J. C. Ye, "Adaptive and compressive beamforming using deep learning for medical ultrasound," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 67, no. 8, pp. 1558–1572, 2020.

[21] O. Solomon, R. Cohen, Y. Zhang, Y. Yang, Q. He, J. Luo *et al.*, "Deep unfolded robust PCA with application to clutter suppression in ultrasound," *IEEE Trans. Med. Imag.*, vol. 39, no. 4, pp. 1051–1063, 2020.

[22] R. J. G. van Sloun, O. Solomon, M. Bruce, Z. Z. Khaing, H. Wijkstra, Y. C. Eldar *et al.*, "Super-resolution ultrasound localization microscopy through deep learning," *IEEE Trans. Med. Imag.*, vol. 40, no. 3, pp. 829–839, 2021.

[23] X. Liu, T. Zhou, M. Lu, Y. Yang, Q. He, and J. Luo, "Deep learning for ultrasound localization microscopy," *IEEE Trans. Med. Imag.*, vol. 39, no. 10, pp. 3064–3078, 2020.

[24] K. G. Brown, D. Ghosh, and K. Hoyt, "Deep learning of spatiotemporal filtering for fast super-resolution ultrasound imaging," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. PP, no. 99, pp. 1–1, 2020.

[25] L. Milecki, J. Poée, H. Belgharbi, C. Bourquin, R. Damseh, P. Delafontaine-Martel *et al.*, "A deep learning framework for spatiotemporal ultrasound localization microscopy," *IEEE Trans. Med. Imag.*, 2021, early access.

[26] R. J. G. van Sloun, R. Cohen, and Y. C. Eldar, "Deep learning in ultrasound imaging," *IEEE Proc.*, vol. 108, no. 1, pp. 11–29, 2020.

[27] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," `arXiv:1912.10557v3 [eess.IV]`, 2020.

[28] J. Youn, B. Luijten, M. B. Stuart, Y. C. Eldar, R. J. G. van Sloun, and J. A. Jensen, "Deep learning models for fast ultrasound localization microscopy," in *Proc. IEEE Ultrason. Symp.*, 2020, pp. 1–4.

[29] J. Youn, M. L. Ommen, M. B. Stuart, E. V. Thomsen, N. B. Larsen, and J. A. Jensen, "Detection and localization of ultrasound scatterers using convolutional neural networks," *IEEE Trans. Med. Imag.*, vol. 39, no. 12, pp. 3855–3867, 2020.

[30] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Eur. Conf. Computer Vision*, 2016, pp. 630–645.

[32] J. A. Jensen and N. B. Svendsen, "Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 39, no. 2, pp. 262–267, 1992.

[33] J. A. Jensen, "Field: A program for simulating ultrasound systems," *Med. Biol. Eng. Comp.*, vol. 10th Nordic-Baltic Conference on Biomedical Imaging, Vol. 4, Supplement 1, Part 1, pp. 351–353, 1996.

[34] ——, "A multi-threaded version of Field II," in *Proc. IEEE Ultrason. Symp.*, 2014, pp. 2229–2232.

[35] J. A. Jensen, H. Holten-Lund, R. T. Nilsson, M. Hansen, U. D. Larsen, R. P. Domsten *et al.*, "SARUS: A synthetic aperture real-time ultrasound system," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 60, no. 9, pp. 1838–1852, 2013.

[36] M. Schneider, "Characteristics of sonovue," *Echocardiography*, vol. 16, no. 1, pp. 743–746, 1999.

[37] F. L. Thurstone and O. T. von Ramm, "A new ultrasound imaging technique employing two-dimensional electronic beam steering," in *Acoustical Holography*, P. S. Green, Ed., vol. 5. New York: Plenum Press, 1974, pp. 249–259.

[38] M. L. Ommen, M. Schou, C. Beers, J. A. Jensen, N. B. Larsen, and E. V. Thomsen, "3d printed calibration micro-phantoms for super-resolution ultrasound imaging validation," *Ultrasonics*, 2021, accepted manuscript.

[39] A. M. Saxe, J. L. McClelland, and S. Ganguli, "Exact solutions to the nonlinear dynamics of learning in deep linear neural networks," `arXiv:1312.6120v3 [cs.NE]`, 2013.

[40] L. Liu, H. Jiang, P. He, W. Chen, X. Liu, J. Gao *et al.*, "On the variance of the adaptive learning rate and beyond," `arXiv:1908.03265v3 [cs.LG]`, 2020.

[41] M. Zhang, J. Lucas, J. Ba, and G. E. Hinton, "Lookahead optimizer: k steps forward, 1 step back," in *Neural Information Processing Systems*, vol. 32, 2019, pp. 9597–9608.

[42] M. Abadi, A. Agarwal, P.Barham, E. Brevdo, Z. Chen, C. Citro *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2011, software available from tensorflow.org. [Online]. Available: https://www.tensorflow.org/

[43] I. Taghavi, S. B. Andersen, C. A. V. Hoyos, M. B. Nielsen, C. M. Sørensen, and J. A. Jensen, "In vivo motion correction in super resolution imaging of rat kidneys," 2021, manuscript submitted for publication.

[44] I. S. Duff and J. Koster, "On algorithms for permuting large entries to the diagonal of a sparse matrix," *SIAM J. Matrix Anal. Appl.*, vol. 22, no. 4, pp. 973–996, 2001.
[45] S. Tang, P. Song, J. D. Trzasko, M. Lowerison, C. Huang, P. Gong *et al.*, "Kalman filter–based microbubble tracking for robust super-resolution ultrasound microvessel imaging," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 67, no. 9, pp. 1738–1751, 2020.
[46] I. Taghavi, S. B. Andersen, C. A. V. Hoyos, M. Schou, S. H. Øygard, F. Gran *et al.*, "Tracking performance in ultrasound super-resolution imaging," in *Proc. IEEE Ultrason. Symp.*, 2020, pp. 1–4.
[47] O. Couture, V. Hingot, B. Heiles, P. Muleki-Seya, and M. Tanter, "Ultrasound localization microscopy and super-resolution: A state of the art," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 65, no. 8, pp. 1304–1320, 2018.
[48] A. Bar-Zion, C. Tremblay-Darveau, O. Solomon, D. Adam, and Y. C. Eldar, "Fast vascular ultrasound imaging with enhanced spatial resolution and background rejection," *IEEE Trans. Med. Imag.*, vol. 36, no. 1, pp. 169–180, 2016.
[49] T. Opacic, S. Dencks, B. Theek, M. Piepenbrock, D. Ackermann, A. Rix *et al.*, "Motion model ultrasound localization microscopy for preclinical and clinical multiparametric tumor characterization," *Nat. comm.*, vol. 9, no. 1, pp. 1527:1–13, 2018.
[50] J. A. Jensen, S. Nikolov, K. L. Gammelmark, and M. H. Pedersen, "Synthetic aperture ultrasound imaging," *Ultrasonics*, vol. 44, pp. e5–e15, 2006.
[51] M. Tanter and M. Fink, "Ultrafast imaging in biomedical ultrasound," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 61, no. 1, pp. 102–119, January 2014.
[52] J. Provost, C. Papadacci, J. E. Arango, M. Imbault, M. Fink, J. L. Gennisson *et al.*, "3-D ultrafast ultrasound imaging in vivo," *Phys. Med. Biol.*, vol. 59, no. 19, pp. L1–L13, 2014.
[53] B. Heiles, M. Correia, V. Hingot, M. Pernot, J. Provost, M. Tanter *et al.*, "Ultrafast 3d ultrasound localization microscopy using a 32 x 32 matrix array," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 38, no. 9, pp. 2005–2015, September 2019.
[54] J. A. Jensen, M. L. Ommen, S. H. Øygard, M. Schou, T. Sams, M. B. Stuart *et al.*, "Three-dimensional super resolution imaging using a row-column array," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 67, no. 3, pp. 538–546, 2020.

Paper 4

Deep Learning Models for Fast Ultrasound Localization Microscopy

Jihwan Youn, Ben Luijten, Matthias Bo Stuart, Yonina C. Eldar, Ruud J. G. van Sloun, Jørgen Arendt Jensen

Published in:
Proceedings of the IEEE International Ultrasonic Symposium

Document Version:
Published

DOI:
10.1109/IUS46767.2020.9251561

Deep Learning Models for Fast Ultrasound Localization Microscopy

1st Jihwan Youn
Dept. of Health Technology
Technical University of Denmark
Lyngby, Denmark
jihyoun@dtu.dk

2nd Ben Luijten
Dept. of Electrical Engineering
Eindhoven University of Technology
& Philips Research
Eindhoven, The Netherlands
w.m.b.luijten@tue.nl

3rd Matthias Bo Stuart
Dept. of Health Technology
Technical University of Denmark
Lyngby, Denmark
mbst@dtu.dk

4th Yonina C. Eldar
Dept. of Computer Science
and Applied Mathematics
Weizmann Institute of Science
Rehovot, Israel
yonina.eldar@weizmann.ac.il

5th Ruud J. G. van Sloun
Dept. of Electrical Engineering
Eindhoven University of Technology
& Philips Research
Eindhoven, The Netherlands
r.j.g.v.sloun@tue.nl

6th Jørgen Arendt Jensen
Dept. of Health Technology
Technical University of Denmark
Lyngby, Denmark
jaje@dtu.dk

***Abstract*—Ultrasound localization microscopy (ULM) can surpass the resolution limit of conventional ultrasound imaging. However, a trade-off between resolution and data acquisition time is introduced. For microbubble (MB) localization, centroid detection is commonly used. Therefore, low-concentrations of MBs are required to avoid overlapping point spread functions (PSFs), leading to a long data acquisition time due to the limited number of detectable MBs in an image frame. Recently, deep learning-based MB localization methods across high-concentration regimes have been proposed to shorten the data acquisition time. In this work, a data-driven encoder-decoder convolutional neural network (deep-ULM) and a model-based deep unfolded network embedding a sparsity prior (deep unfolded ULM) are analyzed in terms of localization accuracy and computational complexity. The results of simulated test data showed that both deep learning methods could handle overlapping PSFs better than centroid detection. Additionally, thanks to its model-based approach, deep unfolded ULM needed much fewer learning parameters and was computationally more efficient, and consequently achieved better generalizability than deep-ULM. It is expected that deep unfolded ULM will be more robust *in-vivo*.**

***Index Terms*—deep unfolded network, high-concentration microbubble localization, model-based neural network, super-resolution ultrasound imaging, ultrasound localization microscopy**

I. Introduction

Ultrasound localization microscopy (ULM) has shown great potential as a breakthrough in super-resolution ultrasound imaging (SRUS) by imaging microvasculature whose vessels are spaced closer than the resolution limit of conventional ultrasound imaging [1]–[6]. ULM is achieved by localizing gas-filled microbubbles (MBs) that are injected into the blood-stream and accumulating their centroids from multiple frames in an image. The resulting super-resolution images can be used to diagnose early-stage cancer [7], ischemic kidney disease [8], and diabetes [9].

The fidelity of ULM depends on the number of detected MBs and their localization precision and sensitivity. Standard ULM methods ordinarily locate the centroids of isolated MBs, therefore, overlapping point spread functions (PSFs) need to be avoided. Diluted low-concentrations of MBs are commonly employed to minimize the overlapping PSFs for accurate localization. Even so, some overlapping PSFs still appear since MBs cannot easily be controlled after injection. The high-resolution of ULM is related to precise MB localization, so the overlapping MB PSFs are often rejected. However, low-concentrations of MBs and overlapping PSF rejection limit the number of detectable MBs in an image frame, and eventually require a long data acquisition time. To cope with this limitation, there have been efforts to achieve SRUS at high-concentrations of MBs [10]–[12].

Recently, several deep learning-based methods have been proposed to localize MBs across high-concentration regimes with overlapping PSFs [13]–[16]. Here we analyze two models and assess their capability in terms of localization accuracy and computational complexity. One approach is a data-driven encoder-decoder convolutional neural network (deep-ULM) [13], and the other is a model-based deep unfolded network that embeds a sparsity prior (deep unfolded ULM) [14]. These algorithms were compared along with the centroid detection method as baseline under challenging simulation scenarios.

II. Method

A. Data Generation

Ultrasound data were simulated in Field II pro [17]–[19] for training and evaluating deep learning models. The simulated data were chosen over measured data for training because it

is difficult to obtain ground-truth (i.e., MB positions) from the measured data. Radiofrequency (RF) channel data were simulated using a transducer modeled following the Verasonics L11-4v and a single cycle 6.9 MHz sinusoidal pulse. For one image frame, eleven plane waves with different angles were transmitted after placing ultrasound scatterers randomly in the region of interest. The RF channel data were then delay-and-sum beamformed with a dynamic apodization on a $\lambda/4$ grid, and the beamformed images were subsequently coherently compounded. The simulation parameters are presented in Table I. For the training set, 256 image frames were generated.

TABLE I
FIELD II SIMULATION PARAMETERS

	Parameter	Value
Transducer	Transmit frequency	6.9 MHz
	Pitch	30 mm
	Element height	5 mm
	Element width	27 mm
	Number of elements	128
Imaging	Wave type	Plane
	Steering angles	$2 \cdot i^{\circ}$, $i \in \{-5, \ldots, 5\}$
	F#	0.5
	# of elements in TX	128
	Apodization in TX	Hann window
	Apodization in RX	Hann window
Environment	Speed of sound	1480 m/s
	Field II sampling frequency	180 MHz

B. Deep learning-based Localization

Deep learning methods were designed to estimate MB positions from beamformed RF data. The MB positions (i.e., output) that were used to train networks were quantized and represented in a $\lambda/16$ image grid. The values of pixels containing MBs were set to one, and the others were zero. The higher-resolution grid was used than the beamformed images (i.e., input) to increase localization precision of estimated MBs.

The deep neural networks were trained by minimizing the difference between true MB positions and estimated MB positions using the ADAM [20] optimizer. The difference was captured by a loss function,

$$\mathcal{L}\left(\mathbf{x}, \mathbf{y}; \theta, \sigma\right) = \frac{1}{N} \sum_{i=1}^{N} \| G\left(\mathbf{y}_i; \sigma\right) - f\left(\mathbf{x}_i; \theta\right) \|_F^2, \quad (1)$$

where $\mathbf{x}_i$ and $\mathbf{y}_i$ are the i-th ultrasound image and MB positions, N is the number of samples, G is the 2-D Gaussian filtering with a standard deviation of σ, $f\left(\cdot; \theta\right)$ is the neural network function with learning parameters θ, and $\|\cdot\|_F$ is the Frobenius norm. Smoothing was applied to the true MB positions to provide larger gradients to ensure training stability.

1) Deep-ULM: Deep-ULM uses an encoder-decoder convolutional neural network (CNN), which is widely used for computer vision and image processing problems such as segmentation [21], [22] and image generation [23]. It mainly

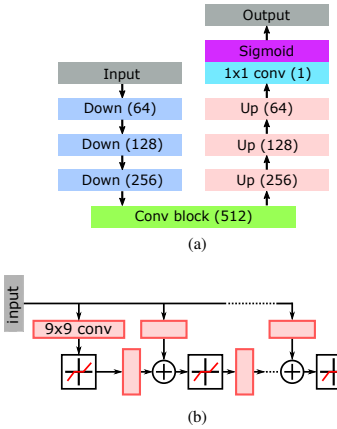


Fig. 1. Deep neural networks for MB localization. (a) Deep-ULM: encoder-decoder convolutional neural network and (b) Deep unfolded ULM: model-based neural network.

consists of *down*, *conv*, and *up* blocks, as shown in Fig. 1a. In the encoding path, the *down* blocks extract features using a series of convolution layers while downsampling the features from the previous layer by a factor of 2. In the decoding path, the MB positions are reconstructed based on the extracted features in the encoding path. To obtain the MB positions in the higher-resolution grid, the first *up* block upsamples the features by a factor of 2 and the other *up* blocks perform upsampling by a factor of 4. A detailed description of *down*, *conv*, and *up* blocks can be found in [15], [16].

The encoder-decoder CNN is a fully data-driven method and requires millions of learning parameters, which has a high chance of overfitting to the training data distributions. Therefore, considering the training data were simulated, deep-ULM may work well on the data simulated in the same way but not on data simulated differently or measured data.

2) Deep unfolded ULM: Deep unfolded ULM has been proposed to overcome the limitations of generalizability of deep-ULM [14], [24]. It solves ULM as a sparse coding problem, which can be formalized as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \quad (2)$$

where $\mathbf{y}$ is the low-resolution MB ultrasound image, $\mathbf{A}$ represents the PSF, $\mathbf{x}$ is the MB positions on the high-resolution grid, and $\mathbf{n}$ is noise.

It can be assumed that $\mathbf{x}$ is sparse because the MB positions are represented in a higher-resolution grid. The optimal $\mathbf{x}$ can then be estimated by solving an optimization problem with a sparsity prior, i.e., the ℓ_1-penalty:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1, \quad (3)$$

where λ is the regularization coefficient. The problem (3) can be solved using the proximal gradient method. However, such

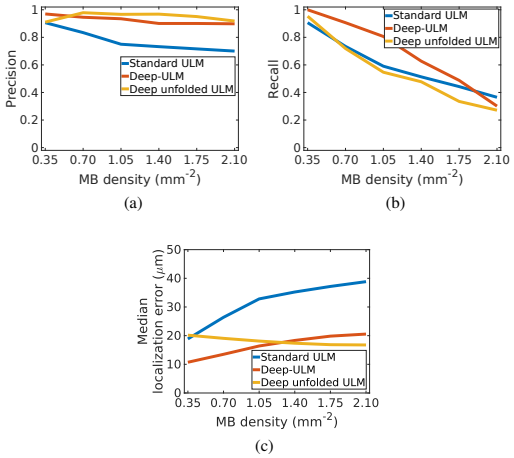


Fig. 2. Comparison of the methods on test sets simulated by placing scatterers randomly at different MB densities where (a) is precision, (b) is recall, and (c) is the median of localization error.

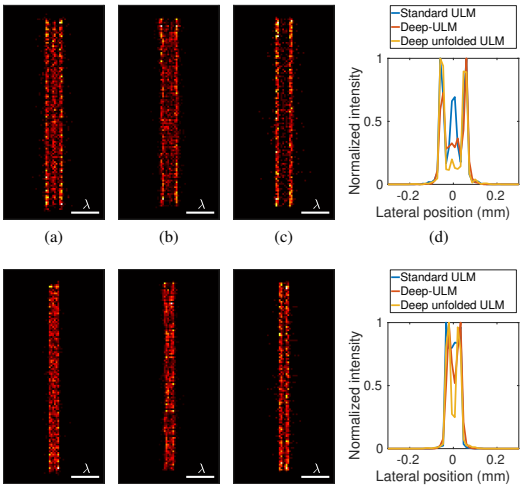


Fig. 3. Comparison of the methods on the simulation of a pair of parallel tubes. (a) - (d) are the results of tubes separated by $\lambda/2$ and (e) - (h) are the results of tubes separated by $\lambda/4$, where (a), (e) are stand ULM, (b), (f) are deep-ULM, (c), (g) are deep unfolded ULM, and (d), (h) are the intensity profile of each method along the lateral direction.

iterative methods may take a long time to converge and their performance highly depends on the hyper-parameters such as the regularization coefficient, the PSF model, and the step size at each iteration, so that empirical tuning is necessary.

Deep unfolded ULM solves the optimization problem using Learned ISTA (LISTA) [25]. LISTA is constructed by unfolding the iteration part as a K-layer neural network, as shown in Fig. 1b. In this work, a 10-layer network was used. LISTA is fast and tuning-free since the iteration is not required and the hyper-parameters, which need to be tuned in the proximal gradient scheme, are embedded in the model, as learning parameters. That allows more robust MB localization by learning more diverse PSF models, unlike the proximal gradient methods which require a specific PSF model [26]. Deep unfolded ULM does not include upsampling in the model, so the input data were upsampled by a factor of 4 before being applied to the network.

III. Results

The trained deep learning models were compared with standard ULM (centroid detection) on two different simulated test sets. One test set comprised independent frames simulated in the same way as the training data at different MB densities. The other test set was composed of consecutive frames simulated using a pair of closely spaced parallel tubes in which scatterers flowed in the opposite directions to each other.

A. Randomly Placed Scatterers

The capability of the models at various MB densities was investigated using a randomly placed scatterer test set. Three evaluation metrics were used: precision, recall, and the median of localization error, defined as

$$\text{precision} = \frac{TP}{TP+FP}, \quad \text{recall} = \frac{TP}{TP+FN}, \tag{4}$$

where TP is the number of true positive (true detection), FP is the number of false positive (false detection), and FN is the number of false negative (missed target).

The results are shown in Fig. 2. For standard ULM, all three metrics got worse as the density increases. At higher densities, a larger number of overlapping PSFs appeared, so that MB localization became more challenging. On the other hand, deep-ULM was not degraded as much as standard ULM at the high densities because deep learning models can deal with a certain degree of overlapping PSFs in MB localization. Deep unfolded ULM achieved comparable precision and localization uncertainty to deep-ULM, but the recall was not as good as deep-ULM. This shows that deep-ULM can achieve better performance on the data set that have the same distribution as the training set, i.e., randomly placed scatterer data, by exploiting a larger number of learning parameters.

B. Parallel Tubes

For more realistic experiments, 1024 consecutive frames were simulated using a pair of parallel tubes separated by $\lambda/2$ and $\lambda/4$. The resulting super-resolution images of each method and their MB intensity profile along the lateral direction are shown in Fig. 3.

The limitation of standard ULM at a high MB density is clearly shown. In the middle of the tubes where no MBs were supposed to be detected, a larger number of false detection

TABLE II
SUMMARY OF DEEP-ULM AND DEEP UNFOLDED ULM

	Deep-ULM	Deep unfolded ULM
Scheme	Fully data-driven	Model-based data-driven
# of learning parameters	5 998 785	1735
Floating point operations (FLOPs)	788 259 839	3462
Generalizability to out of data distributions	Not good	Good

appeared and high MB intensity along the lateral direction was achieved. Both deep learning models worked better than standard ULM and deep unfolded ULM resulted in better-resolved images with much fewer parameters. This shows that deep unfolded ULM achieves better generalization to various data distributions that are different from the training data, consistent with [24], [26].

IV. DISCUSSION

A summary of deep-ULM (a fully data-driven method) and deep unfolded ULM (a model-based data-driven method) are shown in Table II. Deep unfolded ULM, required much fewer parameters and operations while achieving comparable results to deep-ULM. The model-based approach allowed not only to reduce the number of learning parameters and operations, but also to achieve better generalizability to out of training data distributions. Deep unfolded ULM showed better performance on the test set of parallel tubes, which had scatterers located inside the tubes contrary to the training data which had randomly placed scattereres. Under the better generalizability, deep unfolded ULM will possibly be able to achieve more robust MB localization than deep-ULM on measured data.

ACKNOWLEDGMENT

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan V Volta GPU used for this research.

REFERENCES

[1] O. Couture, B. Besson, G. Montaldo, M. Fink, and M. Tanter, "Microbubble ultrasound super-localization imaging (MUSLI)," in *Proc. IEEE Ultrason. Symp.*, 2011, pp. 1285–1287.

[2] O. M. Viessmann, R. J. Eckersley, K. Christensen-Jeffries, M. X. Tang, and C. Dunsby, "Acoustic super-resolution with ultrasound and microbubbles," *Phys. Med. Biol.*, vol. 58, pp. 6447–6458, 2013.

[3] M. A. O'Reilly and K. Hynynen, "A super-resolution ultrasound method for brain vascular mapping," *Med. Phys.*, vol. 40, no. 11, pp. 110 701–7, 2013.

[4] C. Errico, J. Pierre, S. Pezet, Y. Desailly, Z. Lenkei, O. Couture, and M. Tanter, "Ultrafast ultrasound localization microscopy for deep super-resolution vascular imaging," *Nature*, vol. 527, pp. 499–502, November 2015.

[5] K. Christensen-Jeffries, R. J. Browning, M. Tang, C. Dunsby, and R. J. Eckersley, "In vivo acoustic super-resolution and super-resolved velocity mapping using microbubbles," *IEEE Trans. Med. Imag.*, vol. 34, no. 2, pp. 433–440, February 2015.

[6] K. Christensen-Jeffries, O. Couture, P. A. Dayton, Y. C. Eldar, K. Hynynen, F. Kiessling, M. O'Reilly, G. F. Pinton, G. Schmitz, M. Tang *et al.*, "Super-resolution ultrasound imaging," *Ultrasound Med. Biol.*, vol. 46, no. 4, pp. 865–891, 2020.

[7] F. Lin, S. E. Shelton, D. Espindola, J. D. Rojas, G. Pinton, and P. A. Dayton, "3-D ultrasound localization microscopy for identifying microvascular morphology features of tumor angiogenesis at a resolution beyond the diffraction limit of conventional ultrasound," *Theranostics*, vol. 7, no. 1, pp. 196–204, 2017.

[8] S. B. Andersen, C. A. V. Hoyos, I. Taghavi, F. Gran, K. L. Hansen, C. M. Sørensen, and J. A. J. M. B. Nielsen, "Super-resolution ultrasound imaging of rat kidneys before and after ischemia-reperfusion," in *Proc. IEEE Ultrason. Symp.*, 2019, pp. 1–4.

[9] D. Ghosh, J. Peng, K. Brown, S. Sirsi, C. Mineo, P. W. Shaul, and K. Hoyt, "Super-resolution ultrasound imaging of skeletal muscle microvascular dysfunction in an animal model of type 2 diabetes," *J. Ultrasound Med.*, vol. 38, no. 10, pp. 2589–2599, 2019.

[10] A. Bar-Zion, C. Tremblay-Darveau, O. Solomon, D. Adam, and Y. C. Eldar, "Fast vascular ultrasound imaging with enhanced spatial resolution and background rejection," *IEEE Trans. Med. Imag.*, vol. 36, no. 1, pp. 169–180, 2016.

[11] A. Bar-Zion, O. Solomon, C. Tremblay-Darveau, D. Adam, and Y. C. Eldar, "Sushi: Sparsity-based ultrasound super-resolution hemodynamic imaging," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 65, no. 12, pp. 2365–2380, 2018.

[12] O. Solomon, R. J. G. van Sloun, H. Wijkstra, M. Mischi, and Y. C. Eldar, "Exploiting flow dynamics for super-resolution in contrast-enhanced ultrasound," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 60, no. 10, pp. 1573–1586, 2019.

[13] R. J. G. van Sloun, O. Solomon, M. Bruce, Z. Z. Khaing, H. Wijkstra, Y. C. Eldar, and M. Mischi, "Super-resolution ultrasound localization microscopy through deep learning," arXiv:1804.07661v2 [eess.SP], 2018.

[14] R. J. G. van Sloun, R. Cohen, and Y. C. Eldar, "Deep learning in ultrasound imaging," *IEEE Proc.*, vol. 108, no. 1, pp. 11–29, 2020.

[15] J. Youn, M. L. Ommen, M. B. Stuart, E. V. Thomsen, N. B. Larsen, and J. A. Jensen, "Ultrasound multiple point target detection and localization using deep learning," in *Proc. IEEE Ultrason. Symp.*, 2019.

[16] J. Youn, M. L. Ommen, M. B. Stuart, E. V. Thomsen, N. B. Larsen, and J. A. Jensen, "Detection and localization of ultrasound scatterers using convolutional neural networks," *IEEE Trans. Med. Imag.*, 2020.

[17] J. A. Jensen and N. B. Svendsen, "Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 39, no. 2, pp. 262–267, 1992.

[18] J. A. Jensen, "Field: A program for simulating ultrasound systems," *Med. Biol. Eng. Comp.*, vol. 10th Nordic-Baltic Conference on Biomedical Imaging, Vol. 4, Supplement 1, Part 1, pp. 351–353, 1996.

[19] ——, "A multi-threaded version of Field II," in *Proc. IEEE Ultrason. Symp.* IEEE, 2014, pp. 2229–2232.

[20] D. Kingma and L. Ba, "Adam: A method for stochastic optimization," arXiv:1412.6980 [cs.LG], 2015.

[21] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.

[22] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *V. Badrinarayanan and A. Kendall and R. Cipolla*, vol. 39, no. 12, pp. 2481–2495, 2017.

[23] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," arXiv:1611.07004v3 [cs.CV], 2016.

[24] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," arXiv:1912.10557v3 [eess.IV], 2019.

[25] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Int. Conf. Machine Learning*, 2010, pp. 399–406.

[26] G. Dardikman-Yoffe and Y. C. Eldar, "Learned sparcom: Unfolded deep super-resolution microscopy," arXiv:2004.09270v2 [eess.IV], 2020.

Paper 5

Model-based Deep Learning on Ultrasound Channel Data for Fast Ultrasound Localization Microscopy

Jihwan Youn, Ben Luijten, Matthias Bo Stuart, Yonina C. Eldar, Ruud J. G. van Sloun, Jørgen Arendt Jensen

Name of journal in:
In preparation

Document Version:
In preparation

MODEL-BASED DEEP LEARNING ON ULTRASOUND CHANNEL DATA FOR FAST ULTRASOUND LOCALIZATION MICROSCOPY

Jihwan Youn[⋆]*, Ben Luijten*[†]*, Matthias Bo Stuart*[⋆]*, Yonina C. Eldar*[§]
Ruud J. G. van Sloun[†‡]*, Jørgen Arendt Jensen*[⋆]

[⋆] Dept. of Health Technology, Technical University of Denmark, Lyngby, Denmark
[†] Dept. of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands
[‡] Philips Research, Eindhoven, The Netherlands
[§] Dept. of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel

ABSTRACT

Ultrasound localization microscopy (ULM) can break the diffraction limit of ultrasound. However, a long data acquisition time is often required due to the use of low-concentrations of microbubbles (MBs) for high localization precision. Lately, deep learning-based methods that can localize high-concentrations of microbubbles (MBs) robustly have been proposed to overcome this constraint. In particular, deep unfolded ULM has shown promising results with few parameters by using a sparsity prior. In this work, deep unfolded ULM is further extended to perform beamforming as well as MB localization. The proposed network learns data-dependent apodization weights that are optimal for deep unfolded ULM to locate MBs. The beamformed images by the network were sharper than delay-and-sum beamformed images. In a test set simulated at an MB density of $3.84\,\mathrm{mm}^{-1}$, the proposed network reconstructed $87\,\%$ of MBs while achieving comparable localization accuracy to deep unfolded ULM, when centroid detection and deep unfolded ULM reconstructed $42\,\%$ and $67\,\%$ of MBs, respectively.

***Index Terms*—** adaptive beamformer, deep unfolded network, model-based neural network, ultrasound localization microscopy

1. INTRODUCTION

Super-resolution ultrasound imaging aims to separate targets that are spaced closer than the diffraction limit of ultrasound. One approach is ultrasound localization microscopy (ULM) that localizes individual microbubbles (MBs) which are injected into the bloodstream as contrast agents, and superimposes their centroids over time into one image frame. The capability and potential of ULM for vascular imaging have been extensively studied [1–6]. Clinically, ULM images are expected to be used for the diagnosis of early-stage cancer [7], ischemic kidney disease [8], and diabetes [9], as well as functional ultrasound [10].

The resolution of ULM is essentially determined by the sensitivity and localization precision of estimated MBs. Therefore, diluted low-concentrations of MBs are often used to minimize MB localization uncertainty by avoiding overlapping point spread functions (PSFs). However, a long data acquisition time is required to reconstruct the target structure since the number of detectable MBs becomes limited. Super-resolution imaging across high-concentrations of MBs using sparse recovery has been proposed [11–13], and ULM at high-concentrations using deep learning has been investigated [14–18] to shorten the data acquisition time.

Among deep learning-based methods, deep unfolded ULM [16] has shown comparable performance to deep-ULM [14] with far fewer parameters thanks to its model-based approach, which also led to better generalizability [17]. In this work, we propose a deep learning model by incorporating Adaptive Beamforming by deep LEarning (ABLE) [19] into deep unfolded ULM. ABLE can perform fast content-adaptive apodization calculation, which results in high-quality ultrasound images. By placing ABLE before deep unfolded ULM, and training the whole network in an end-to-end fashion, apodization weights optimized for MB localization by deep unfolded ULM can be learned. The localization performance of the proposed method was assessed and compared with deep unfolded ULM under various simulation scenarios. The proposed network reconstructed more MBs than deep unfolded ULM while keeping localization accuracy.

2. METHOD

2.1. Ultrasound Data Generation

To train deep learning models, ultrasound data with ground truth (i.e, MB positions) are necessary. It is extremely difficult to acquire true MB positions in measured data, especially when there are overlapping PSFs. Therefore, training and test data were simulated using Field II pro [20–22]. The parame-

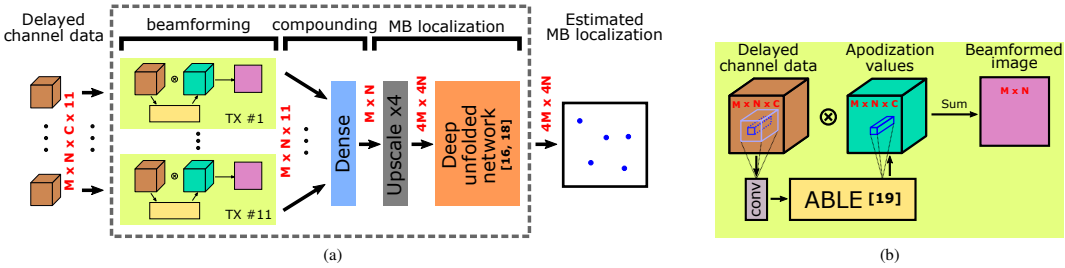


Fig. 1. A schematic overview of the proposed network. (a) shows the whole pipeline and (b) shows the beamforming process for one transmit event. The proposed network takes delayed RF channel data as input and performs beamforming. Here, optimal apodization weights for the downstream task (i.e., MB localization) are learned by ABLE [19]. After that, beamformed signals from each transmit are compounded using a dense layer, and MBs are localized using deep unfolded ULM [16, 18] in the image beamformed and compounded by the network. The red text represents data size.

ters for the simulation are presented in Table 1. For one image frame, ultrasound scatterers were randomly placed in the region of interest and eleven plane waves steered at different angles were transmitted. Simulated RF channel data were then delayed (time-of-flight corrected) on a $\lambda/4$ grid but not summed. The resulting data size for one image frame (eleven transmit events) was $\mathrm{M} \times \mathrm{N} \times \mathrm{C} \times 11$, where M and N are the numbers of image points in the axial and lateral directions and C is the number of transducer elements. A total of 768 frames were generated for training.

Table 1. Field II simulation parameters

	Parameter	Value
Transducer	TX frequency	6.9 MHz
	# of TX pulse cycle	1
	Pitch	0.30 mm
	Element height	5.00 mm
	Element width	0.27 mm
	Number of elements	128
Imaging	Wave type	Plane
	Steering angles	$2 \cdot i°, i \in \{-5, \ldots, 5\}$
	# of elements in TX	128
	Apodization in TX	Hann window
	Apodization in RX	Hann window
Environment	Speed of sound	1480 m/s
	Field II sampling frequency	180 MHz

2.2. Network Architecture

The proposed neural network was constructed by combining ABLE [19] and deep unfolded ULM [16]. An overview of the framework is illustrated in Fig. 1a. The network takes delayed RF channel data as input and performs beamforming that is optimized for the downstream task (i.e., MB localization). Specifically, to that end, the beamforming part of the network actively adapts apodization weights. After that, signals from each transmit event are compounded. Lastly, MBs are localized in the image that is beamformed and compounded by the network.

The beamforming for one transmit event is illustrated in Fig. 1b, where ABLE calculates content-adaptive apodization weights for MB localization. The detailed network structure of ABLE can be found in [19]. Here, a 5×5 convolution layer was additionally used before ABLE to consider neighboring pixels for apodization by offering a larger receptive field. A distinct ABLE network was defined for each transmit event (i.e., eleven ABLE networks). The beamformed signals from each transmit are compounded through a dense layer, which effectively learns a weighted summation. A compounded ultrasound image can be obtained after this dense layer, whose size is $\mathrm{M} \times \mathrm{N}$.

MB positions were represented in a $\lambda/16$ grid for more precise localization, however, the beamforming was performed in a $\lambda/4$ grid. Therefore, the beamformed images were upscaled by a factor of 4 using the nearest neighbor interpolation, and deep unfolded ULM localized MBs on the upscaled images. Deep unfolded ULM is a model-based neural network [23] that solves ULM as a sparse coding problem [24], which can be expressed as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{y}$ is the MB ultrasound image, $\mathbf{A}$ is its shifted versions of the PSF, $\mathbf{x}$ is the MB positions, and $\mathbf{n}$ is noise. The MB positions $\mathbf{x}$ can be estimated by solving the following optimization problem with the ℓ_1-penalty:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1, \tag{2}$$

where λ is the regularization coefficient.

Proximal gradient methods can be used to solve the problem (2). But, such methods may require many iterations and

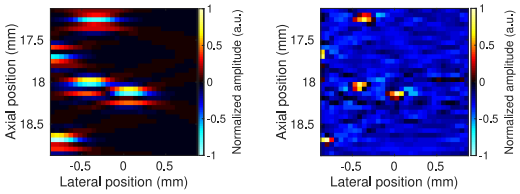


Fig. 2. A comparison of beamformed and compounded RF images by (a) delay-and-sum with a dynamic apodization where F# is 0.5 and (b) ABLE trained jointly with deep unfolded ULM.

are highly sensitive to hyper-parameters (e.g., the step size, the regularization coefficient, and the PSF model). Deep unfolded ULM instead uses Learned ISTA (LISTA) [25] to solve the optimization problem, which is a K-layer neural network built by unfolding the iteration part. The hyper-parameters in proximal gradient methods are embedded in the network so that they can be learned from data during training. Therefore, deep unfolded ULM can achieve more robust MB localization by learning diverse PSF models and better generalization with few parameters thanks to its model-based approach [26]. Further details on deep unfolded ULM can be found in [16, 18]. In this work, a 10-layer network was used, which consisted of 9×9 convolution layers.

The training was performed by minimizing the following loss function with the ADAM [27] optimizer,

$$\mathcal{L}\left(\mathbf{x}, \mathbf{y}; \theta, \sigma\right) = \frac{1}{n} \sum_{i=1}^{n} \| G\left(\mathbf{y}_i; \sigma\right) - f\left(\mathbf{x}_i; \theta\right) \|_F^2, \quad (3)$$

where $\mathbf{x}_i$ and $\mathbf{y}_i$ are the i-th delayed RF channel data and MB positions, n is the number of samples, G is the 2-D Gaussian filtering with a standard deviation of σ, $f\left(\cdot; \theta\right)$ is the neural network function with learning parameters θ, and $\|\cdot\|_F$ is the Frobenius norm. Here, smoothing was applied to gain training stability by providing larger gradients. The standard deviation was chosen to be 1 pixel through cross-validation.

3. RESULTS

Conceptually, the proposed network is composed of beamforming and MB localization, although a mapping from delayed RF data to MB positions is learned in an end-to-end fashion. An image, beamformed by the proposed network, can be obtained by taking the intermediate layer output. One example is shown in Fig. 2 with a delay-and-sum beamformed image with a dynamic apodization where F# is 0.5. The network beamformed image resulted in sharper peaks at MB positions while producing noise in the other region, which can easily be handled by deep unfolded ULM.

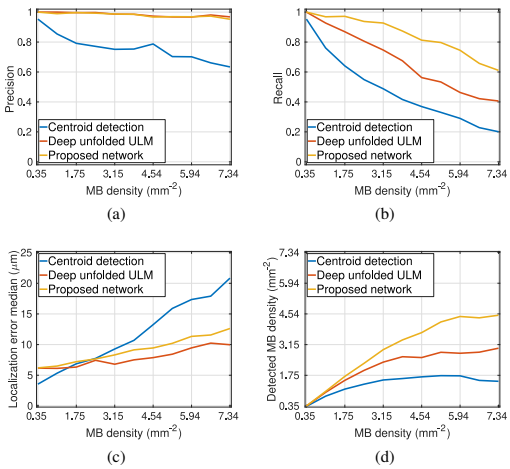


Fig. 3. Comparison of the methods on test sets simulated by placing scatterers randomly at different MB densities where (a) is precision, (b) is recall, and (c) is the median of localization error.

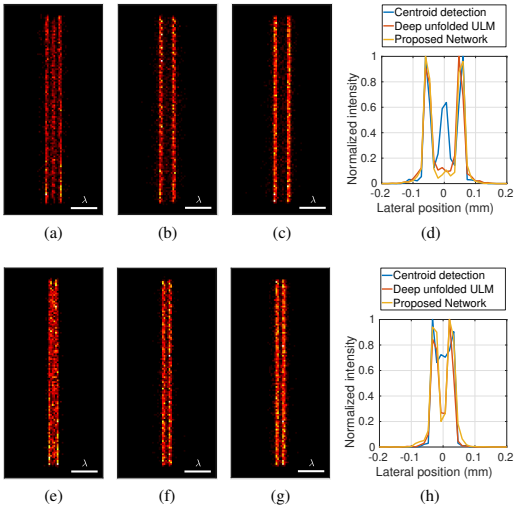


Fig. 4. Comparison of the methods on the simulation of a pair of parallel tubes. (a) - (d) are the results of tubes separated by $\lambda/2$ and (e) - (h) are the results of tubes separated by $\lambda/4$, where (a), (e) are ULM images by centroid detection, (b), (f) are ULM images by deep unfolded ULM, (c), (g) are ULM images by the proposed method, and (d), (h) are the intensity profiles along the lateral direction.

Table 2. Parallel tube simulation results

Method	$\lambda/2$		$\lambda/4$	
	precision	recall	precision	recall
Centroid detection	0.60	0.29	0.68	0.30
Deep unfolded ULM	0.72	0.53	0.83	0.52
Proposed network	0.80	0.71	0.80	0.66

Precision, recall, localization error median, and detected MB density were calculated on test data at different MB densities, which is defined as

$$\text{precision} = \frac{TP}{TP+FP}, \ \text{recall} = \frac{TP}{TP+FN}, \\ \text{Detected MB density} = \text{recall} \times \text{MB density}, \tag{4}$$

where TP is the number of true positive (true detection), FP is the number of false positive (false detection), and FN is the number of false negative (missed target). The localization performance degraded as the MB density increased for all methods, as shown in Fig. 3 since more overlapping PSFs started to appear. The performance drop of centroid detection was more drastic because centroid detection cannot handle overlapping PSFs. Precision and localization error were comparable for deep unfolded ULM and the proposed method. However, the sharper peaks allowed for the proposed network to reconstruct larger numbers of MBs, especially at high MB densities than deep unfolded ULM without sacrificing precision and localization error.

More realistic experiments were performed by simulating 1024 consecutive frames using scatterers flowing in two pairs of parallel tubes separated by $\lambda/2$ and $\lambda/4$. The ULM images and their MB intensity profiles along the lateral direction are shown in Fig. 4, and precision and recall are presented in Table 2. The limitation of centroid detection at high MB densities was explicitly shown from not only low precision and recall but also high MB intensities in the middle of the tubes, where MBs should not be detected. The model-based approach allowed deep unfolded ULM and the proposed network to be well generalized to the parallel tube simulation data that came from a different data distribution than the training data (i.e., randomly placed scatterer data) [18,23,26]. The proposed network again showed better recall than deep unfolded ULM.

4. CONCLUSION

A model-based neural network that can localize high-concentrations of MBs for ULM is proposed. The network is constructed by incorporating ABLE into deep unfolded ULM. By doing so, adaptive apodization weights that are optimal for deep unfolded ULM to locate MBs can be learned from data. The neural network beamforming resulted in a sharper image than delay-and-sum beamforming, as shown in Fig. 2. The proposed network detected more MBs than deep unfolded ULM while keeping similar localization accuracy by locating MBs in the sharper image. The proposed network can possibly be used to reduce the data acquisition time of ULM by localizing the more MBs precisely using high-concentrations of MBs.

5. COMPLIANCE WITH ETHICAL STANDARDS

This is a numerical simulation study for which no ethical approval was required.

6. ACKNOWLEDGMENT

This work is partially supported by the Fondation Idella. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan V Volta GPU used for this research.

7. REFERENCES

[1] O. Couture, B. Besson, G. Montaldo, M. Fink, and M. Tanter, "Microbubble ultrasound super-localization imaging (MUSLI)," in *Proc. IEEE Ultrason. Symp.*, 2011, pp. 1285–1287.

[2] O. M. Viessmann, R. J. Eckersley, K. Christensen-Jeffries, M. X. Tang, and C. Dunsby, "Acoustic super-resolution with ultrasound and microbubbles," *Phys. Med. Biol.*, vol. 58, pp. 6447–6458, 2013.

[3] M. A. O'Reilly and K. Hynynen, "A super-resolution ultrasound method for brain vascular mapping," *Med. Phys.*, vol. 40, no. 11, pp. 110701–7, 2013.

[4] C. Errico, J. Pierre, S. Pezet, Y. Desailly, Z. Lenkei, O. Couture, and M. Tanter, "Ultrafast ultrasound localization microscopy for deep super-resolution vascular imaging," *Nature*, vol. 527, pp. 499–502, November 2015.

[5] K. Christensen-Jeffries, R. J. Browning, M. Tang, C. Dunsby, and R. J. Eckersley, "In vivo acoustic super-resolution and super-resolved velocity mapping using microbubbles," *IEEE Trans. Med. Imag.*, vol. 34, no. 2, pp. 433–440, February 2015.

[6] K. Christensen-Jeffries, O. Couture, P. A. Dayton, Y. C. Eldar, K. Hynynen, F. Kiessling, M. O'Reilly, G. F. Pinton, G. Schmitz, M. Tang, et al., "Super-resolution ultrasound imaging," *Ultrasound Med. Biol.*, vol. 46, no. 4, pp. 865–891, 2020.

[7] F. Lin, S. E. Shelton, D. Espindola, J. D. Rojas, G. Pinton, and P. A. Dayton, "3-D ultrasound localization microscopy for identifying microvascular morphology features of tumor angiogenesis at a resolution beyond the diffraction limit of conventional ultrasound," *Theranostics*, vol. 7, no. 1, pp. 196–204, 2017.

[8] S. B. Andersen, C. A. V. Hoyos, I. Taghavi, F. Gran, K. L. Hansen, C. M. Sørensen, and J. A. Jensen M. B. Nielsen, "Super-resolution ultrasound imaging of rat kidneys before and after ischemia-reperfusion," in *Proc. IEEE Ultrason. Symp.*, 2019, pp. 1–4.

[9] D. Ghosh, J. Peng, K. Brown, S. Sirsi, C. Mineo, P. W. Shaul, and K. Hoyt, "Super-resolution ultrasound imaging of skeletal muscle microvascular dysfunction in an animal model of type 2 diabetes," *J. Ultrasound Med.*, vol. 38, no. 10, pp. 2589–2599, 2019.

[10] T. Deffieux, C. Demene, M. Pernot, and M. Tanter, "Functional ultrasound neuroimaging: a review of the preclinical and clinical state of the art," *Curr. Opin. Neurol.*, vol. 50, pp. 128–135, 2018.

[11] A. Bar-Zion, C. Tremblay-Darveau, O. Solomon, D. Adam, and Y. C. Eldar, "Fast vascular ultrasound imaging with enhanced spatial resolution and background rejection," *IEEE Trans. Med. Imag.*, vol. 36, no. 1, pp. 169–180, 2016.

[12] A. Bar-Zion, O. Solomon, C. Tremblay-Darveau, D. Adam, and Y. C. Eldar, "SUSHI: Sparsity-based ultrasound super-resolution hemodynamic imaging," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 65, no. 12, pp. 2365–2380, 2018.

[13] O. Solomon, R. J. G. van Sloun, H. Wijkstra, M. Mischi, and Y. C. Eldar, "Exploiting flow dynamics for super-resolution in contrast-enhanced ultrasound," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 60, no. 10, pp. 1573–1586, 2019.

[14] R. J. G. van Sloun, O. Solomon, M. Bruce, Z. Z. Khaing, H. Wijkstra, Y. C. Eldar, and M. Mischi, "Super-resolution ultrasound localization microscopy through deep learning," `arXiv:1804.07661v2 [eess.SP]`, 2018.

[15] J. Youn, M. L. Ommen, M. B. Stuart, E. V. Thomsen, N. B. Larsen, and J. A. Jensen, "Ultrasound multiple point target detection and localization using deep learning," in *Proc. IEEE Ultrason. Symp.*, 2019.

[16] R. J. G. van Sloun, R. Cohen, and Y. C. Eldar, "Deep learning in ultrasound imaging," *IEEE Proc.*, vol. 108, no. 1, pp. 11–29, 2020.

[17] J. Youn, M. L. Ommen, M. B. Stuart, E. V. Thomsen, N. B. Larsen, and J. A. Jensen, "Detection and localization of ultrasound scatterers using convolutional neural networks," *IEEE Trans. Med. Imag.*, 2020.

[18] J. Youn, B. Luijten, M. B. Stuart, Y. C. Eldar, R. J. G. van Sloun, and J. A. Jensen, "Deep learning models for fast ultrasound localization microscopy," in *Proc. IEEE Ultrason. Symp.*, 2020, pp. 1–4.

[19] B. Luijten, R. Cohen, F. J. De Bruijn, H. A. W. Schmeitz, M. Mischi, Y. C. Eldar, and R. J. G. Van Sloun, "Adaptive ultrasound beamforming using deep learning," *IEEE Trans. Med. Imag.*, 2020.

[20] J. A. Jensen and N. B. Svendsen, "Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 39, no. 2, pp. 262–267, 1992.

[21] J. A. Jensen, "Field: A program for simulating ultrasound systems," *Med. Biol. Eng. Comp.*, vol. 10th Nordic-Baltic Conference on Biomedical Imaging, Vol. 4, Supplement 1, Part 1, pp. 351–353, 1996.

[22] J. A. Jensen, "A multi-threaded version of Field II," in *Proc. IEEE Ultrason. Symp.* 2014, pp. 2229–2232, IEEE.

[23] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," `arXiv:1912.10557v3 [eess.IV]`, 2020.

[24] Y. C. Eldar, *Sampling Theory: Beyond Bandlimited Systems*, Cambridge University Press, 2015.

[25] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Int. Conf. Machine Learning*, 2010, pp. 399–406.

[26] G. Dardikman-Yoffe and Y. C. Eldar, "Learned SPARCOM: Unfolded deep super-resolution microscopy," *Opt. Express*, vol. 28, no. 19, pp. 27736–27763, 2020.

[27] D.P. Kingma and L.J. Ba, "ADAM: A method for stochastic optimization," `arXiv:1412.6980 [cs.LG]`, 2015.