



Multi-planar whole heart segmentation of 3D CT images using 2D spatial propagation CNN

Sundgaard, Josefine Vilsbøll; Juhl, Kristine Aavild; Kofoed, Klaus Fuglsang; Paulsen, Rasmus Reinhold

Published in:
Proceedings of SPIE

Link to article, DOI:
[10.1117/12.2548015](https://doi.org/10.1117/12.2548015)

Publication date:
2020

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

[Link back to DTU Orbit](#)

Citation (APA):
Sundgaard, J. V., Juhl, K. A., Kofoed, K. F., & Paulsen, R. R. (2020). Multi-planar whole heart segmentation of 3D CT images using 2D spatial propagation CNN. In *Proceedings of SPIE* (Vol. 11313). Article 113131Y SPIE - International Society for Optical Engineering. <https://doi.org/10.1117/12.2548015>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

PROCEEDINGS OF SPIE

SPIDigitalLibrary.org/conference-proceedings-of-spie

Multi-planar whole heart segmentation of 3D CT images using 2D spatial propagation CNN

Sundgaard, Josefine Vilsbøell, Juhl, Kristine Aavild, Kofoed, Klaus Fuglsang, Paulsen, Rasmus

Josefine Vilsbøell Sundgaard, Kristine Aavild Juhl, Klaus Fuglsang Kofoed, Rasmus R. Paulsen, "Multi-planar whole heart segmentation of 3D CT images using 2D spatial propagation CNN," Proc. SPIE 11313, Medical Imaging 2020: Image Processing, 113131Y (10 March 2020); doi: 10.1117/12.2548015

SPIE.

Event: SPIE Medical Imaging, 2020, Houston, Texas, United States

Multi-planar whole heart segmentation of 3D CT images using 2D spatial propagation CNN

Josefine Vilsbøll Sundgaard^a, Kristine Aavild Juhl^a, Klaus Fuglsang Kofoed^b, and Rasmus R. Paulsen^a

^aDepartment of Applied Mathematics and Computer Science, Technical University of Denmark, Kongens Lyngby, Denmark

^bDepartment of Cardiology, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark

ABSTRACT

Whole heart segmentation from cardiac CT scans is a prerequisite for many clinical applications, but manual delineation is a tedious task and subject to both intra- and inter-observer variation. Automating the segmentation process has thus become an increasingly popular task in the field of image analysis, and is generally solved by either using 3D methods, considering the image volume as a whole, or 2D methods, segmenting each slice independently. In the field of deep learning, there are significant limitations regarding 3D networks, including the need for more training examples and GPU memory. The need for GPU memory is usually solved by down sampling the input images, thus losing important information, which is not a necessary sacrifice when employing 2D networks. It would therefore be relevant to exploit the benefits of 2D networks in a configuration, where spatial information across slices is kept, as when employing 3D networks. The proposed method performs multi-class segmentation of cardiac CT scans utilizing 2D convolutional neural networks with a multi-planar approach. Furthermore, spatial propagation is included in the network structure, to ensure spatial consistency through each image volume. The approach keeps the computational assets of 2D methods while addressing 3D issues regarding spatial context. The pipeline is structured in a two-step approach, in which the first step detects the location of the heart and crops a region of interest, and the second step performs multi-class segmentation of the heart structures. The pipeline demonstrated promising results on the MICCAI 2017 Multi-Modality Whole Heart Segmentation challenge data.

Keywords: Deep learning, whole heart segmentation, cardiac CT, convolutional neural networks

1. INTRODUCTION

Cardiovascular disease (CVD) is the leading cause of death worldwide, and accounted for 31% of all deaths in 2016.¹ Early detection of CVD is crucial, and medical imaging with segmentation and registration of heart structures is a vital tool for the diagnosis and treatment. Fully automatic whole heart segmentation is challenging due to the complex geometry of the heart, unclear boundaries between heart structures and variation in orientation, shape, and size from patient to patient.² Previous attempts of whole heart segmentation have mainly focused on deformable models using statistical shape models (SSM) and atlas registration.³ Limitations of these methods include computational time for the non-rigid registration in atlas propagation, which is computationally expensive, and the extensive need for data in SSM, since examples of all variations of heart geometry is needed to create an atlas. Deformation of the model is constrained to be within the shape of the SSM, because of the strong prior in the model. With large variation in heart geometry, especially when pathological cases are involved, the need for data becomes critical.²

Since machine learning and particularly deep learning has gained ground in medical image analysis and computer vision in general, many state of the art segmentation works are based on deep learning and these have been outperforming previous methods. In the task of cardiac CT segmentation, neural networks are used to segment heart structures in a 3D image volume, which can either be performed using 3D or 2D neural networks.

Further author information:

J.V.S.: E-mail: josh@dtu.dk, Telephone: (+45) 24 45 46 96

Medical Imaging 2020: Image Processing, edited by Ivana Išgum, Bennett A. Landman,
Proc. of SPIE Vol. 11313, 113131Y · © 2020 SPIE · CCC
code: 1605-7422/20/\$21 · doi: 10.1117/12.2548015

Proc. of SPIE Vol. 11313 113131Y-1

3D neural networks treat each image volume as a training example, while 2D neural networks split the image volume into several 2D image slices used as training examples. When employing 3D neural networks the spatial context in the image volumes is captured, and the spatial consistency between slices is preserved. The limitations of 3D methods include the need for 3D volume data, which drastically reduces the number of training examples. Furthermore, 3D operations require far more GPU memory and the solution is often to down sample the data, which causes loss of information. 2D networks are lightweight when it comes to GPU memory, and require less training data since each volume contains many training examples.⁴ It is therefore relevant to investigate the ability of 2D networks in 3D image segmentation.

When analysing 3D image volumes using 2D networks, different approaches can be used. The most basic approach is to simply analyse each 2D image slice with a neural network and concatenate all of the resulting segmentations into a 3D label map, but this does not capture any spatial context across slices. A multi-planar approach with three perpendicular 2D U-nets was proposed by Mortazi et al.⁵ and Wang et al.⁶ in the MM-WHS challenge. This multi-planar approach employs three 2D networks, which perform segmentation in three perpendicular planes in the image volume. The output segmentation is created as a combination of the three resulting probability maps, resulting in a label map of the full image volume. Besides the MM-WHS challenge, several other groups have worked on the task of heart segmentation using deep learning, primarily on MRI data. Zheng et al.⁷ proposed an approach for segmentation of left and right ventricular cavity and left ventricular myocardium in MRI data. Their method was based on a 2D U-net-like network,⁸ which was propagated down through the slices of the MRI scan, from base to apex. They added a recurrent element to the network structure, using the generated segmentation of the adjacent slice as an extra input for the current slice. This way, contextual information was propagated through the data, assisting to include spatial information in the 2D segmentation.

These two approaches of combining the segmentation from 2D networks into a 3D volume have obtained good results regarding spatial context, while limiting the use of GPU memory and exploiting the large amount of 2D training data in a 3D image volume. The proposed method employs 2D networks in a novel configuration in order to exploit the computational assets of 2D networks, while addressing 3D issues regarding spatial context and avoiding the drawbacks of 3D networks. The network structure is a combination of the approaches presented in Zheng et al.⁷ and Mortazi et al.,⁵ using a multi-planar pipeline structure with a spatial propagation element in the network architecture. The pipeline is used to segment the seven largest heart structures (right and left ventricle, right and left atria, left ventricular myocardium, aorta and pulmonary arteries) in cardiac CT scans. A majority of work on heart segmentation is aimed at MR images as this is the primary modality for heart diagnostics. However, CT contrast imaging is rapidly gaining a foot-hold and therefore we have aimed our approach specifically at contrast enhanced CT scans.

2. DATA

The MICCAI MM-WHS dataset^{2,3,9,10} consists of 60 image volumes, 20 labelled CT scans for training and 40 unlabelled CT scans for testing. The images were collected in a clinical environment at Shanghai Shuguang Hospital, China with a routine cardiac CT angiography. The angiography creates a clear contrast in the images in the left ventricle, which makes it possible to distinguish between the blood cavity and myocardium in the left ventricle, while also highlighting aorta in the scan. The images cover the whole heart region from the upper abdominal to the aortic arch, with an in-plane resolution of 0.78 x 0.78 mm and a slice thickness of 1.60 mm. The training data is annotated with manual segmentation of seven of the most important substructures of the heart: right ventricle (RV), left ventricle (LV), right atria (RA), left atria (LA), left ventricular myocardium (MYO), aorta (AO) and pulmonary arteries (PA). To increase the size of the training data, data augmentation is performed with a factor of three using B-spline deformations.¹¹ This form of deformation is non-linear and affects locally, creating small alterations to the image, in order to mimic anatomic variations in the heart structures.

3. METHODS

The segmentation pipeline is a two-step approach. The first step is a region network, performing a binary segmentation of heart and background using a normal U-net.⁸ To reduce computational cost, this step is performed on a down-sampled version of the input images where each slice is resampled to dimensions 128x128. A binary segmentation is obtained in all three directions, where the pixels containing the value one represents the

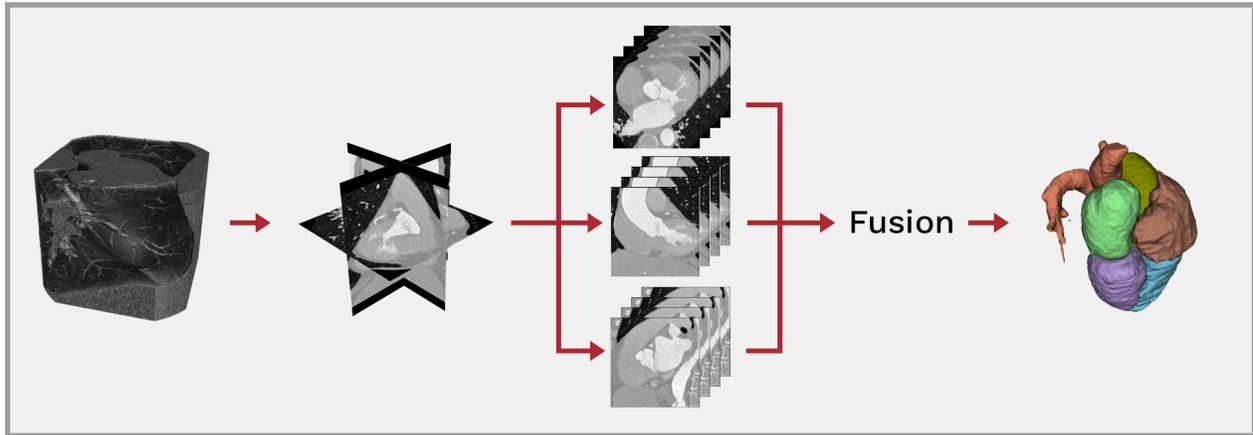


Figure 1: Overview of multi-planar segmentation

heart and the pixel with value zero represents everything that is not heart. After a fusion of the three probability maps, a final segmentation is generated and a boundary box around the heart structures is created. The original training images are then cropped based on the boundary box, generating a 2% margin around the segmented heart.

The second step takes the cropped input and performs multi-class segmentation of the seven heart structures in all three directions using the spatial propagation network. For this step, the cropped training images are downsampled to 256x256. The multiple networks are trained from the three directions (axial, sagittal and coronal views) in a 2D manner. The structure of the approach and the multi-planar setup is shown in Figure 1. The network structure is an extended version of the U-net⁸ inspired by Zheng et al.⁷ and attempts to include spatial information by introducing a recurrent element in the network structure, as seen in Figure 2. This additional contextual input propagates spatial consistency through the image volume, to ensure coherence between the individual slices.

Fusion of the three directional probability maps in each step is performed by selecting the class as the one with the highest probability across all directions and classes.

Preprocessing in the form of padding is performed to ensure equal dimensions in all three directions of the CT scans, as the original CT scans are 512x512 and varying size in the z-direction. Normalization of the images is furthermore performed, to ensure pixel values in the interval [-1,1].

In order to minimize the amount of misclassifications and errors, post processing is performed using connected component analysis. Only the largest component is assumed to be a given heart structure, and all smaller components are assigned to the nearest different heart structure or background. The Euclidian distance measure is used to determine the distance from a small component to the other classes.

3.1 Implementation details

The networks are implemented in Python using Tensorflow, and trained end-to-end on a Titan X 12 GB RAM GPU. The Adam optimizer with a learning rate of $1 \cdot 10^{-5}$ is employed for training the networks. Batch size is set to one, and due to the small batch size, a high momentum of 0.99 is employed to ensure that a large number of the previous training samples contribute to the update in the current optimization step. A categorical loss functions is used for the multi-class segmentation networks, and the loss function is minimized over a training period of 30 epochs. The training time per epoch is approximately 1600 seconds, thus the full training for each of the three networks is 13 and a half hours. The extracted 2D images are handled individually, meaning that from the 60 image volumes, 15360 training examples of dimensions 256x256 are extracted. For each epoch, all 2D training images are shuffled, to ensure a random training on each of the training images. This way, each image creates an independent change on the network, without being biased by the previous image in each epoch. Inference time is on average 2 minutes and 50 seconds, in which loading of the model takes 40 seconds and

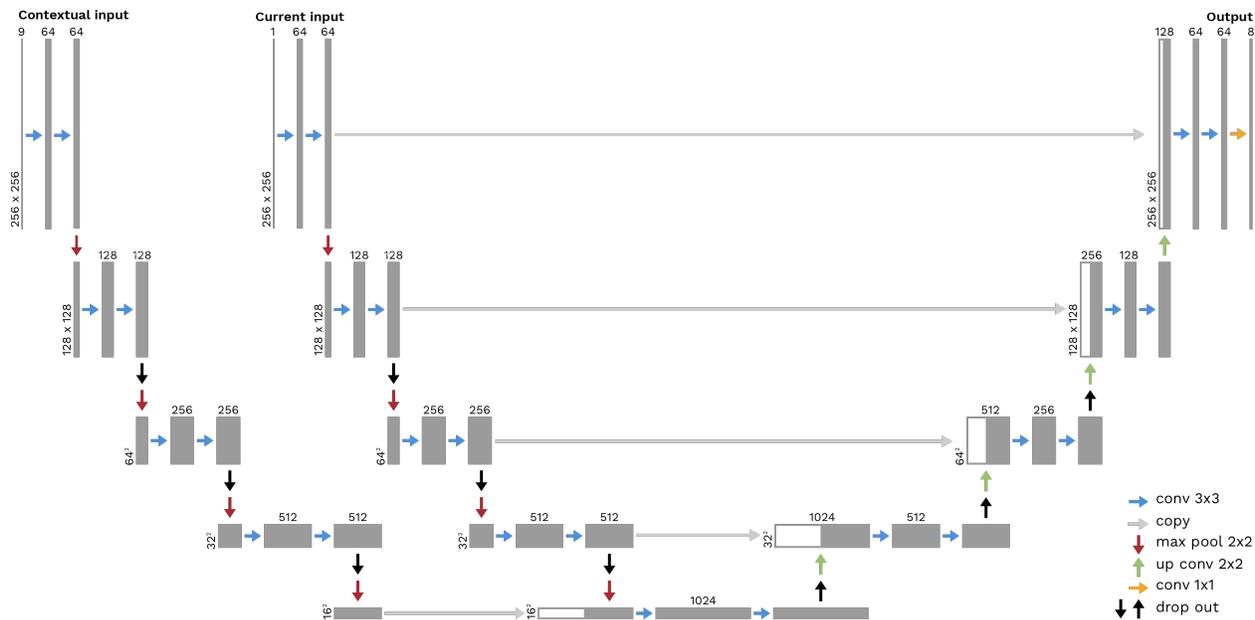


Figure 2: Segmentation network architecture with contextual input, modified from Zheng et al.⁷ The leftmost contracting path uses the previous training image and segmentation as input, while the other contracting path (to the right) uses the current training image as input. Each grey box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box, and the dimensions of each map are listed at the lower left edge. White boxes represent copied feature maps, and the arrows denote different operations

prediction of all 2D slices in one 3D volume takes 130 seconds for each network. Furthermore the post processing step takes on average 60 seconds. This adds up to 9 minutes and 30 seconds for total analysis and post processing of each 3D image volume.

4. PRELIMINARY EXPERIMENTS

Preliminary experiments were run using only the 20 annotated training examples. These were separated into 15 examples for training and 5 examples for testing, and then several methods were compared on this dataset. Table 1 shows the performance of the proposed combination method using both spatial propagation and multi-planar setup, while the performance of the separate methods are also shown together with the most simple approach of a 2D U-net of each axial slice in the image volume. The effect of the post processing of the proposed method is likewise shown in this Table. Multi-planar U-net is an approach similar to Mortazi et al.⁵ and the spatial propagation CNN used is inspired from Zheng et al.⁷ and the combined method is the novel approach of spatial propagation CNN in a multi-planar setup. These results show, that the performance is increased for all heart structures when using the combined approach, except for the left atrium, and that the post processing step increases the Dice score for all heart structures.

5. RESULTS

The proposed segmentation pipeline on the full training dataset is evaluated using the test dataset from the MM-WHS challenge consisting of 40 cardiac CT scans. Dice score for each heart structure is listed here: left ventricle 0.914 ± 0.054 , right ventricle 0.884 ± 0.062 , left atrium 0.920 ± 0.041 , right atrium 0.840 ± 0.204 , myocardium 0.832 ± 0.109 , aorta 0.940 ± 0.023 , pulmonary arteries 0.845 ± 0.112 and for whole heart segmentation 0.886 ± 0.064 . The highest Dice scores are achieved for aorta, left atrium and left ventricular blood cavity. As the cardiac CT is captured with the use of intravenous contrast and the acquisition is triggered to ensure contrast

Table 1: Average Dice scores for each heart structure and whole heart segmentation for the applied methods. Best performing method for each heart structure is marked with bold. WHS = Whole Heart Segmentation

Method	LV	RV	LA	RA	MYO	AO	PA	WHS
2D U-net	0.934	0.914	0.931	0.895	0.898	0.965	0.858	0.955
Multi-planar U-net	0.939	0.926	0.926	0.919	0.912	0.971	0.880	0.959
Spatial propagation CNN	0.934	0.917	0.937	0.900	0.892	0.970	0.878	0.956
Combined	0.942	0.928	0.931	0.919	0.912	0.972	0.891	0.960
Combined with CCA	0.943	0.930	0.933	0.922	0.912	0.978	0.899	0.962

in the left side of the heart, the boundaries are easier to detect, compared to the right side of the heart. This is clearly seen in the difference in Dice scores between the two sides of the heart.

Figure 3 shows the three best and three worst segmentations performed on the test dataset. The top rows show three smooth and nice segmentations of each heart structure, where each substructure is well separated from the others. The bottom row shows the problems and errors that occur in the segmentation. In Figure 3e and 3f, the left atrium is not found in the right location, but this class is blended together with the left ventricular myocardium instead. This is caused by the very similar Hounsfield values in these heart structures, and the fact that no anatomical or shape knowledge is introduced to these neural networks. Another typical problem area in heart segmentation is the boundary between the right ventricle and atrium, as also seen in Figure 3f, where the ventricle extends into the area of the atrium. This boundary is difficult to define even by medical experts, since there is no clear difference between these two heart structures, and this gives rise to segmentation errors.² Besides the right atrium and the ventricles, the rest of the heart structures are well defined even in the worst performing segmentations.

In Table 2 the Dice score, Jaccard index, surface-to-surface distance and Hausdorff distance is shown for this approach, the best performing approach from the challenge by Payer et al.¹² and the multi-planar approach by Mortazi et al.⁵ together with the average performance scores of the participants in the challenge employing deep learning methods for the segmentation. The results show that this approach achieves a better score on all four measurements compared to Mortazi et al.⁵ and well above the average scores of all challenge participants employing deep learning methods. Payer et al.,¹² employing a 3D U-net, outperforms this approach, which shows that more work still needs to be done, before 2D networks can capture the spatial content of a 3D image volume as well as 3D networks. When comparing to the full score board of the MM-WHS challenge, this approach ranks 2nd in regards to Hausdorff distance, 3rd in regards to surface-to-surface distance, 4th in regards to Jaccard index and achieves similar Dice scores as the 4th and 5th best performing approaches.

Table 2: Performance scores of relevant CT MM-WHS challenge participants¹³ together with the proposed method. The average performance of all participants employing deep learning methods is also shown for reference. SD = Surface-to-surface distance, HD = Hausdorff distance, MP = Multi-planar, CNN = Convolutional Neural Network

Approach	Dice score	Jaccard index	SD (mm)	HD (mm)
Payer et al. ¹²	0.908 ± 0.086	0.832 ± 0.037	1.117 ± 0.250	25.242 ± 10.813
MP spatial propagation CNN	0.886 ± 0.064	0.800 ± 0.089	1.428 ± 1.146	25.864 ± 12.965
Mortazi et al. ⁵	0.879 ± 0.079	0.792 ± 0.106	1.538 ± 1.006	28.481 ± 11.434
Average	0.875 ± 0.083	0.784 ± 0.010	1.840 ± 2.963	38.510 ± 17.890

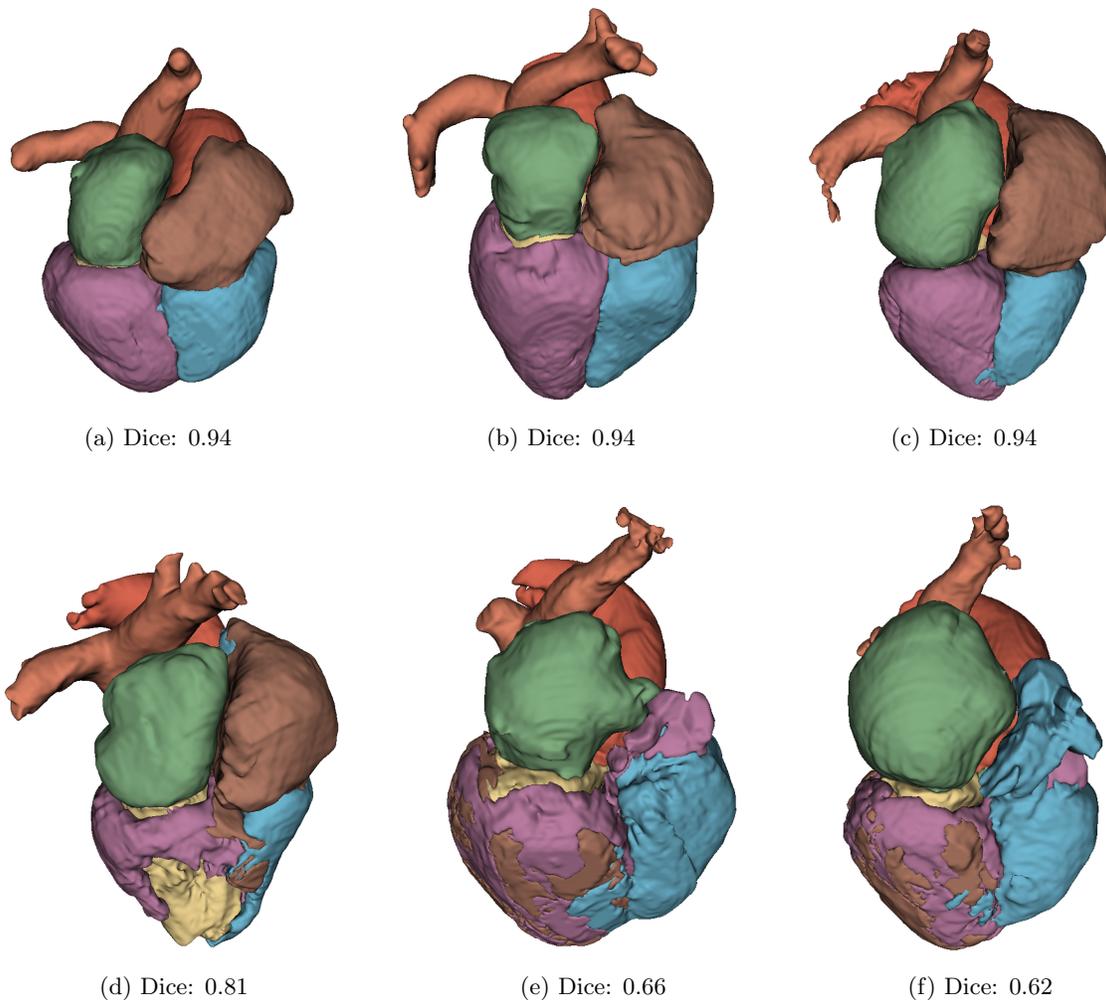


Figure 3: Top row: Test segmentations with the highest Dice scores. Bottom row: Test segmentations with the lowest Dice scores. Heart structures: Left atrium (green), right atrium (brown), left ventricular myocardium (purple), left ventricle blood cavity (yellow), right ventricle (blue), aorta (red), pulmonary arteries (orange)

6. DISCUSSION

Since the cardiac CT image data is a 3D volume, the approach for the heart segmentation could also have been a 3D convolutional neural network. One of the biggest struggles of a 3D network is the need for GPU memory for the 3D operations and the large input images, which often leads to down sampling of the training examples. Payer et al.¹² employed a 3D U-net for heart segmentation, and the input volume was down sampled to 64 x 64 x 64 for the fine-grain segmentation, while the input size for this 2D implementation was 256 x 256, and Mortazi et al.⁵ employed input images of dimensions 350 x 350 for their 2D approach. This means, that much finer details can be captured in the 2D approaches since the details are not lost in the down sampling. Baumgartner et al.⁴ found that the low through-plane resolution of MRI scans reduced the performance of 3D networks. The through-plane resolution is better in CT than in MRI scans, but it could still be comparable. In this work, it is found that the sagittal and coronal directional networks do not obtain as good results as the axial directional network, which could be due to the lower resolution. It is a benefit of the pipeline in this work, that the axial prediction is not affected by this low through-plane resolution, as would be seen in a 3D approach. Despite this issue, it is shown, that the fusion of the three directional 2D networks greatly improves the prediction in a 3D image volume compared to a single 2D network, and that the performance could be further improved by

employing the contextual input network in each direction. With this setup, the spatial information was captured in the 3D image, while still exploiting the benefits of a 2D network.

Heart segmentation involves many challenges, such as overlap of Hounsfield values between heart structures and class imbalance between object classes and background, which was sought to be solved by the two step approach employing initial region cropping. In general, the networks seem to underestimate the heart structures, and the boundaries between the substructures are the biggest challenge, which has not yet been completely solved. The problem with utilizing neural networks for this task is, that it is very difficult to determine the origin of errors, as the decision making is not comprehensible. The test predictions show that the similar Hounsfield values of right ventricle and atrium causes problems, and that the network struggles to detect the myocardium in some scans. But overall, the results indicate that the network can be used to obtain sensible heart segmentations. The heart consists of many structures besides the seven larger structures evaluated in this project. The generated segmentation is an overall model of the heart, but in and between the structures, smaller anatomical structures are present, such as papillary muscles in the ventricles. The papillary muscles constitute about 10 % of the total heart mass, and are clearly visible in the CT scans, especially in the left ventricle due to the contrast. The presence of these muscles arises an issue when predicting, since the Hounsfield values are very different from the rest of the ventricle, and the boundary between the myocardium and blood cavity is very difficult to distinguish. The same issue is present for the valves between the heart chambers, which could explain some of the boundary issues with the segmentation between atria and ventricle.

7. CONCLUSION

This work shows that 3D spatial context can be captured using 2D networks with this novel network configuration. The results show that this approach achieves promising scores on the MM-WHS challenge data. It is demonstrated how the multi-planar approach is improved, when using the spatial propagation network in all four evaluation metrics. This novel approach shows promising results in capturing the spatial information in a 3D image volume, while exploiting the benefits of 2D convolutional networks.

REFERENCES

- [1] World Health Organization, “Cardiovascular diseases (CVDs),” (2017).
- [2] Zhuang, X., “Challenges and Methodologies of Fully Automatic Whole Heart Segmentation: A Review,” *Journal of Healthcare Engineering* **4**(3), 371–407 (2013).
- [3] Zhuang, X., Bai, W., Song, J., Zhan, S., Qian, X., Shi, W., Lian, Y., and Rueckert, D., “Multiatlas whole heart segmentation of CT data using conditional entropy for atlas ranking and selection,” *Medical Physics* **42**(7), 3822–3833 (2015).
- [4] Baumgartner, C. F., Koch, L. M., Pollefeys, M., and Konukoglu, E., “An exploration of 2D and 3D deep learning techniques for cardiac MR image segmentation,” in [*MICCAI STACOM workshop*], 111–119, Springer (2017).
- [5] Mortazi, A., Burt, J., and Bagci, U., “Multi-planar deep segmentation networks for cardiac substructures from MRI and CT,” in [*MICCAI STACOM workshop*], 199–206, Springer (2017).
- [6] Wang, C. and Smedby, Ö., “Automatic whole heart segmentation using deep learning and shape context,” in [*MICCAI STACOM workshop*], 242–249, Springer (2017).
- [7] Zheng, Q., Delingette, H., Duchateau, N., and Ayache, N., “3D Consistent & Robust Segmentation of Cardiac Images by Deep Learning with Spatial Propagation,” *IEEE Transactions on Medical Imaging* **37**, 2137–2148 (2018).
- [8] Ronneberger, O., Fischer, P., and Brox, T., “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in [*MICCAI*], 234–241, Springer (2015).
- [9] Zhuang, X., Rhode, K. S., Razavi, R. S., Hawkes, D. J., and Ourselin, S., “A registration-based propagation framework for automatic whole heart segmentation of cardiac MRI,” *IEEE Transactions on Medical Imaging* (2010).
- [10] Zhuang, X. and Shen, J., “Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI,” *Medical Image Analysis* **31**, 77–87 (2016).

- [11] Rueckert, D., “Nonrigid registration using free-form deformations: Application to breast mr images,” *IEEE Transactions on Medical Imaging* (1999).
- [12] Payer, C., Štern, D., Bischof, H., and Urschler, M., “Multi-label whole heart segmentation using CNNs and anatomical label configurations,” in [*MICCAI STACOM workshop*], 190–198, Springer (2017).
- [13] Zhuang, X., Li, L., Payer, C., Štern, D., Urschler, M., Heinrich, M. P., Oster, J., Wang, C., Smedby, Ö., Bian, C., Yang, X., Heng, P. A., Mortazi, A., Bagci, U., Yang, G., Sun, C., Galisot, G., Ramel, J. Y., Brouard, T., Tong, Q., Si, W., Liao, X., Zeng, G., Shi, Z., Zheng, G., Wang, C., MacGillivray, T., Newby, D., Rhode, K., Ourselin, S., Mohiaddin, R., Keegan, J., Firmin, D., and Yang, G., “Evaluation of algorithms for Multi-Modality Whole Heart Segmentation: An open-access grand challenge,” *arXiv preprint arXiv:1902.07880* (2019).