UAV image analysis for leakage detection in district heating systems using machine learning

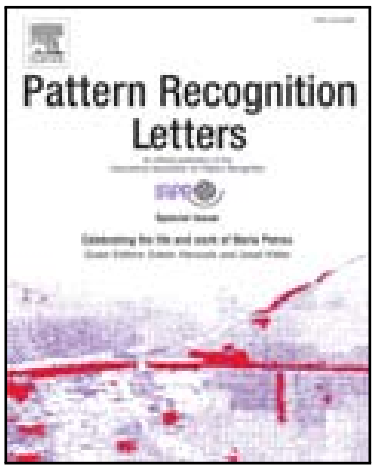
Hossain, Kabir; Villebro, Frederik; Forchhammer, Søren

Published in:
Pattern Recognition Letters

Link to article, DOI:
10.1016/j.patrec.2020.05.024

Publication date:
2020

Document Version
Peer reviewed version



UAV Image Analysis for Leakage Detection in District Heating Systems using Machine Learning

Kabir Hossain , Frederik Villebro , Søren Forchhammer

PII: S0167-8655(20)30203-8
DOI: https://doi.org/10.1016/j.patrec.2020.05.024
Reference: PATREC 7910

To appear in: *Pattern Recognition Letters*

Received date: 30 August 2019
Revised date: 9 January 2020
Accepted date: 20 May 2020

Please cite this article as: Kabir Hossain , Frederik Villebro , Søren Forchhammer , UAV Image Analysis for Leakage Detection in District Heating Systems using Machine Learning, *Pattern Recognition Letters* (2020), doi: https://doi.org/10.1016/j.patrec.2020.05.024

Highlights

- An automatic energy leakage detection for district heating systems based on UAV Infrared (IR) images is presented.
- A deep learning-based model and 8 different ML classifiers have been developed and evaluated for detection of leakages.
- A region extraction algorithm has been incorporated to extract potential leakages enabling the ML analysis.
- The proposed ML methods have been applied to ca. 13.4 million image patches and achieved an actual detection rate of 98.6%.

UAV Image Analysis for Leakage Detection in District Heating Systems using Machine Learning

Kabir Hossain[a], Frederik Villebro[a,b,] and Søren Forchhammer[a]

[a] *Department of Photonics Engineering, Technical University of Denmark*
[b]*Drone Systems, Denmark*

Keywords
CNN, SVM, RF, Adaboost, Energy Leakage detection, District Heating systems

ABSTRACT

In this paper, we propose automatic energy leakage detection in underground pipes of district heating systems based on Infrared (IR) images, captured by an Unmanned Aerial Vehicle (UAV). Hot water or steam is distributed to homes and industries through underground pipes from a central power plan. Leakages in underground pipes pose a very common problem, which can occur for many reasons, e.g. unprofessional installation and end of service life. Potentially, a leakage remains undiscovered for a very long period of time. Therefore, it is of great interest for power supply companies to monitor district heating networks to identify leakages. In this paper, the original IR images are captured in a 16 bit format by a UAV. On ground, potential leakages are extracted using a region extraction algorithm. Thereafter a Convolutional Neural Network (CNN) as well as eight conventional Machine Learning (ML) classifiers are applied on these regions to classify whether or not it is a leakage.

In total, twelve UAV sequences are captured at different cities in Denmark. Based on these, around 13.4 million samples of image patches of district heating systems are extracted. Eleven sequences are used for training and the remaining one for testing. This was performed on all splits in the leave-one-out testing. The deep learning CNN achieved an average weighted accuracy of 0.872 with a false positive and negative rate of 12.7 % and 10.4 %, respectively. This CNN model detected around 98.6 % of the true leakages. In comparison, conventional ML classifiers, i.e. Adaboost (AB), Random Forest (RF), etc. provide lower average weighted accuracy, but on the other hand they require less computational resources. We have compared our method with a state-of-art method and the result shows that the proposed method is very competitive.

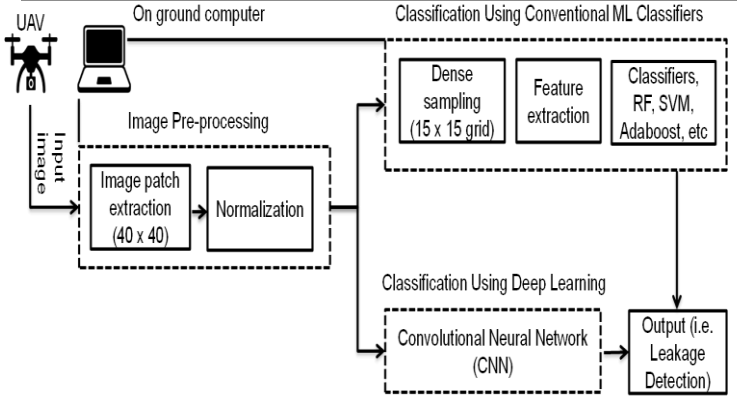


Fig. 1. Overall diagram of the proposed Model

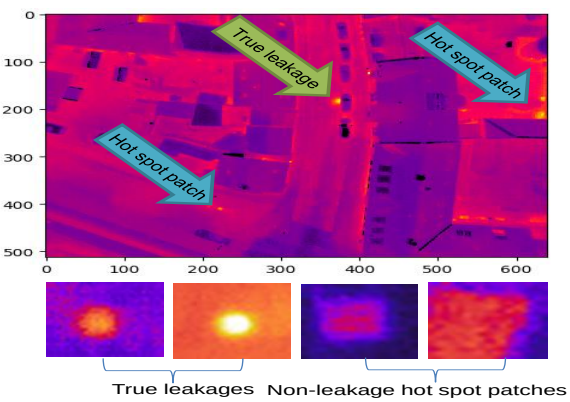


Fig. 2. 8 bit version of UAV image

Table 1. Parameters for Region Extraction

Input	Radius	Vote Kernel	Vote Fraction	Minimum Area
Values	100	5 x 5	0.5	50

1. Introduction

Central power plants produce and supply hot water to homes and industries through underground pipes in most of the northern countries. The old underground pipes start to degenerate with time (Olsson, 2001), even newer systems can leak for many reasons, as unprofessional installation, lack of maintenance, etc. Lack of efficient methods for discovering leakages can lead to unnecessary cost, creating a negative impact on the environment (Fröling, 2002) and even sometimes posing threats for humans. Therefore, it is very important for the power supply company to monitor the network regularly and find a way to discover leakages in underground pipes.

The manual way is to measure the amount of hot water coming from the central power plant and compare with the amount of hot water reached by homes and industries. This way the existence of leakages can be discovered, but that process does not give the location of leakages in underground pipes. This process involves digging up the pipes to find exact locations, which is a very expensive operation for a power supply company.

Infrared (IR) sensors have the capacity to capture heat variations, where a potential underground leakage will appear as a hot spot. Therefore, the more efficient and cheapest way is to use an IR sensor embedded on a UAV to inspect the leakages of a district heating network. In this paper, we focus on the automatic leakage detection of district heating networks by analysis of UAV images using ML. The overall system diagram of our proposed model can be seen in Fig. 1

1.1. Related Work

Over the years, researchers have worked on various methods for monitoring district heating networks. Especially detecting the location of leakages is very important (Zhou et al., 2018). Infrared sensors have been widely used over the years in a wide range of
applications, as wildlife monitoring (Ward et al., 2016), inspection of solar panels (Deitsch et al., 2018) and energy installations (O. Friman and Sjökvist, 2014; Berg et al., 2016; Berg and Ahlberg, 2014). In (Ljungberg and Rosengren, 1988; Axelsson, 1987; Bøhm and Borgstrøm, 1996), thermal cameras were already used for energy inspection. Particularly, energy leakage detection using a hand-held device was presented by Bøhm (Bøhm and Borgstrøm, 1996). The results presented in (Ljungberg and Rosengren, 1987; Axelsson, 1988; Bøhm and Borgstrøm) are outdated today due to rapid improvement of IR cameras in the past decades.

Recently, in (O. Friman and Sjökvist, 2014; Berg et al., 2016; Berg and Ahlberg, 2014), an IR camera embedded on a plane was used for energy leakage detection. In (O. Friman and Sjökvist, 2014; Berg et al., 2016; Berg and Ahlberg, 2014), orthorectification, and pipe information were used to extract regions of interest. Then they used pre-defined features, as mean, standard deviation, circularity, etc. Thereafter five conventional ML classifiers were applied for performance evaluation. Their setup is rather expensive since they were using a plane equipped with a cooled thermal camera. In contrast, we are using an UAV embedded with a cheap uncooled IR camera for energy leakage inspections. It should be noted here, a camera embedded on an UAV is often used for other applications (Wang et al., 2019). Moreover, no predefined features are used in our deep learning based CNN model. However, our conventional ML classifiers require prior feature engineering for performance evaluations.
In a more recent paper on energy leakage detection, Y. Zhong et al. (Zhong et al., 2019) utilize almost the exact same setup as us. They utilize the same IR camera core and a UAV. Their payload furthermore consists of a normal RGB camera, which is used to reduce the search area in case of no pipeline information being available. Saliency maps were used for leakage detection with favorable results on the relatively limited datasets compared to our datasets, in terms of area covered and the number of ground truth leakages. It may be remarked that the investigated leakages in (Zhong et al., 2019) are rather eye-catching which explains why saliency maps seem to be so promising. In contrast, we apply a region extraction algorithm, which acts as a coarse filter selecting potential regions containing a leakage, and perform detection using various ML classification models. The authors (Zhong et al., 2019) mention ML as a possible improvement to their system.

In a preliminary work (Hossain et al., 2019), leakage detection in district heating systems using UAV was presented. In this paper, we extend the work. We have evaluated our model for more sequences (12) with a total of 243082 images within which around 13.4 million image patches of district heating areas were identified as potential leakages using a region extraction.

A total of 1345 unique leakages were identified by DroneSystems who provided one image for each leakage marking its location with a bounding box. The identified leakages from DroneSystems were delivered to the district heating companies who confirmed that a leakage was present. This feedback validates the identified leakages and indirectly indicates that the leakage identification performance is at an acceptable

level. These annotations provided by DroneSystems are considered to be the basis for defining the ground truth in our experiments. However the initial annotation does not provide annotation for all images where a leakage is present. Thus, we extended the annotation by for each initially identified leakage also marking it in all the other images where the leakage appears. All these leakages where then marked with a bounding box manually. This forms the ground truth used in our experiments.

In our target application, the IR sensors are embedded on the UAV. The UAV captured 16 bit images. The images were later processed on a ground computer, where our region detection algorithm extracts regions of interest and cuts them into image patches. It is determined for training and testing whether a patch contains a leakage using the extended ground truth annotation. As an example, 8 bit versions of image patches both with and without a leakage are shown in Figure 2. Finally, with these true and false labelled image patches, we train our model with CNN and eight conventional ML classifiers to classify the true leakages.

1.2. Contribution

The main contribution of this work is that we present a method for automatic leakage detection in underground pipes of district heating networks using CNN and conventional ML classifiers based on UAV IR images and test it on a large dataset of actual images. We incorporate a region extraction method to extract potential leakages precisely. The proposed model is applied to the extremely large district heating dataset consisting of the 243082 UAV IR images, from which around 13.4 million samples of image patches were extracted. The machine learning is applied on the raw 16 bit format, since 16 bit provides detailed information. As a result, we provide a relatively cheap and simple solution to monitor district heating networks.

1.3. Outline

The outline of this paper as follows. In Section 2, the proposed method is described in detail. The experimental results including description of district heating datasets, and performance of the proposed method along with the discussion are presented in Section 3. Finally, Section 4 contains our conclusions.

2. Methodology

We have applied conventional ML classifiers and deep learning-based CNN for the leakage detection of district heating networks. At first, the potential leakages (image patches) are extracted from the full image using a region extraction algorithm. Then the image patches are processed before feeding them into the network or the ML classifier.

2.1. Region Extraction Algorithm

Regions potentially containing anomalies are characterized by being rather warm compared to their surroundings. Furthermore, these anomalies usually have softer gradients as the heat diffuses through the ground. The algorithm used to find these potential regions has 3 steps. The first step is to find the warmer regions within the image. Secondly, gradients of a certain magnitude within the image are found and in the last step, they are combined. Each of the steps is visualized, by the 8-bit version of the input image, in Fig. 3.

Both warmer regions and gradients are found by looking at a region around the pixel in question. Mean and standard deviations are calculated for both temperature and gradient. These values are used to determine whether the pixel is warm enough or it has the right gradient magnitude evaluated by how many standard deviations the pixel is above the mean. The

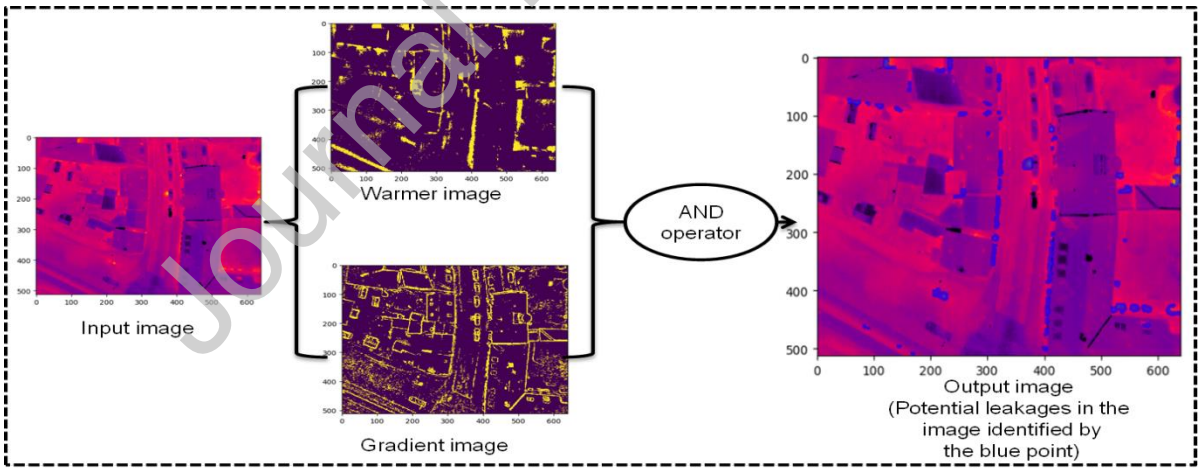


Fig. 3. Overall procedure of region extraction algorithm

warmer region is simply found by the following equation,

$$I_w(j) = \begin{cases} 1, & \mu(j) + \sigma(j) < I(j) \\ 0, & otherwise, \end{cases} \qquad (1)$$

where I is the image within the search region, j is the pixel index value. μ and σ are mean and standard deviation of the input search region, respectively.

Table 2. Parameters of Conventional ML Classifiers

Methods	Parameters
RF	Number of estimators = 200
Adaboost	Number of estimators = 200 Learning rate = 1
SVM	Penalty parameter C = 1 Kernel = 'Linear'

For the gradient image, we have used a Sobel filter with a filter size of 3 x 3 to obtain the two images, i.e. G_x and G_y which indicate changes in horizontal and vertical direction, respectively.

$$G_x = \begin{pmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{pmatrix} * I \quad (2)$$

$$G_y = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{pmatrix} * I \quad (3)$$

Based on G_x and G_y, the magnitude of the gradient, G is calculated, by

$$G = \sqrt{{G_x}^2 + {G_y}^2} \quad (4)$$

Then the mean and standard deviation of the gradient image is used to determine whether the pixels have the right gradient by the following equation,

$$I_w(j) = \begin{cases} 1, & \mu(j) + 0.5\,\sigma(j) > G(j) \\ 0, & otherwise, \end{cases} \quad (5)$$

where *j* is the index value of the gradient image, and μ and σ are mean and standard deviation, of the region surrounding pixel *j*, respectively.

Finally, these regions of warmth and sizeable gradients are combined using an AND operation followed by a majority vote within a *5 x 5* kernel and thereafter potential holes are filled and smaller regions removed. The final algorithm additionally takes four input parameters consisting of the Radius, Vote Kernel, Vote Fraction and Minimum Area. Labelled regions are returned in the form of a matrix with the same size as the input image. Radius is in pixels, Warmth bias as well as Gradient bias are standard deviations, and Vote Kernel is in pixels. Vote fraction is a threshold for the majority vote and Minimum Area is an area in pixels. The parameters used in this algorithm are given in Table 1.

2.2. Image Pre-processing

After extraction of image patches, the patches are preprocessed before applying conventional ML classifiers and CNN, respectively. At first, the patches are rescaled to 40 x 40 and then normalization (Hossain et al., 2019) is applied on each of the patches. Normalization is a very important step because it is ensures faster convergence of the CNN and makes it less sensitive to small intensity changes. The normalization is done by:

$$y = (x - \mu)/\ \sigma, \quad (6)$$

where *x* and *y* are the pixel values, σ is the standard deviation and μ is the mean value of image pixels.

2.3. Classification Using Conventional ML Classifiers

In our experiment, we have tested eight different conventional ML classifiers after training them. For conventional ML classifiers, it is required to extract features.

2.3.1 Feature Extraction

We have extracted dense SIFT and dense SURF features for training using conventional ML classifiers. We selected these features and dense sampling as in (Deitsch et al., 2018) to get a fair amount of feature points from all parts of the image. Dense sampling detects the feature point from an *n* x *n* grid, and subsequently feature descriptors are extracted from each point. In total 1152 and 576 features were extracted by dense SIFT and SURF, respectively. Then, a list of conventional ML classifiers are each trained using these features for leakage detection.

2.3.2 Conventional ML Classifiers

We have trained our model using a set of eight conventional ML classifiers, where four of them are linear and the other four are non-linear classifiers. The linear models are Logistic Regression (LR) (Puntanen, 2010), Linear Discriminant Analysis (LDA) (Duda et al., 2001), Linear Support Vector Machine (L-SVM) (Hastie et al., 2008) and Gaussian Naive Bayes (GNB) (Webb and I., 2010). The remaining four non-linear models were k-nearest neighbors (KNN) (Mucherino et al., 2009), Decision Tree (DT) (Williams, 2011), Random Forest (RF) (Breiman, 2001) and Adaboost (AB) (Bishop, 2006). Python and the scikit-learn library (Pedregosa, F. 2011) are used to implement the proposed model with the ML classifiers. The default parameters are used for most of the ML classifiers as given in (Pedregosa, F. 2011), except for three ML classifiers; RF, Adaboost and SVM for which the parameters are given in Table 2.

2.4. Classification Using Convolutional Neural Network

The 40 x 40 input image patches are directly fed into the CNN network for classification. In our implementation, multiple layers are used including three convolutional layers followed by two fully connected layers. The first two filter sizes are 64 and the last one is 128 with a kernel size of 3 x 3 inside the convolutional layers.

Each convolutional layer is followed by a leaky ReLu activation function with a rate of 0.1 and a max-pooling layer with a kernel of 2 x 2. Subsequently, dropout layers are utilized with each of the convolutional layers with a rate of 0.25 and 0.5 for the first 2 layers and the last 2 layers, respectively. A total of 128 for the first and 2 neurons (responsible for classification) are defined in each of the fully connected layers.

The loss function used in this structure is cross entropy and the optimizer is Adam (Kingma and Ba, 2015). We have used a total of 20 epochs with batch sizes of 64. The library used for CNN implementation is Keras (Chollet et al., 2015) and the language is Python.

3. Experimental Results

As shown in Table 3 (a), a total of 243082 images (16 bit format) were captured. Then ground truths are based upon

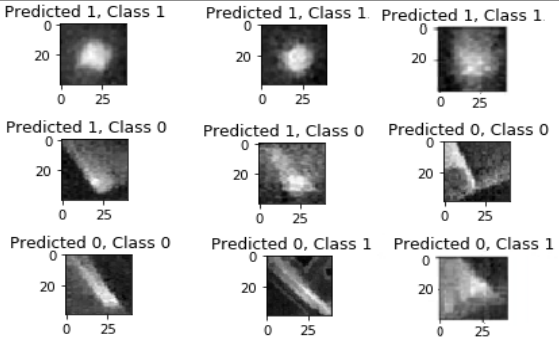


Fig. 4. Visualization of the confusion matrix in terms of image patches (1 represents true leakage and 0 means non-leakage hot spot patch)

classification initially performed by an expert within leakage detection from the IR images. As described in Sec. 1.1, all the images were visually inspected by the expert and 1345 (unique) leakages were located and each of these localized on an image using a bounding box. The provided images with a leakage from the expert were then utilized to locate the same leakages in all other images manually, again marking using a bounding box.

Out of the 243082 images, 13380450 (around 13.4 million) image patches were extracted by the region extraction algorithm (See Sec. 2.1). Out of the 13.4 million image patches, only 30290 image patches are true leakages and the remaining are false. It should be noted here that our ground truth leakages are composed of 3 categories A, B, and E corresponding to clear leakage, bad insulation/smaller leakage, and district heating systems (DSH) components, respectively. They are all treated as leakages here, but some of them are very hard to detect.

The experiments are carried out using the Convolutional Neural network (CNN) and the eight different conventional Machine Learning (ML) classifiers. In the case of CNN (See Sec. 2.4), the image patches are directly fed into the network for classification. For conventional ML classifiers, the process can be divided into two parts. Firstly, features from the images in the training dataset are extracted, using general SIFT and SURF descriptors, and then indexed into an HDF5 (The HDF Group, 1997) file format. Secondly, the network is trained using the eight ML classifiers (See Section 2.3.2).

3.1. Dataset

The sequences are obtained by night flights searching for potential leakages in the district heating network. Each of the sequences is captured at one of 12 different cities in Denmark. These UAV sequences are provided by Drone Systems.

The original UAV sequences (S1-S12) are in a 16 bit format, which are acquired with an uncooled TeAx camera. The model name is Thermal Capture 640 2.0 containing a FLIR Tau 2 640 core, which has three different versions (Industrial, Performance, and Commercial). Our sequences were captured using two versions of the camera, namely Industrial and Performance.

The original image resolution is 640 x 512 and the sensitivity up to 30mK for Industrial and 50mK for Performance. We extract the image patches with a resolution of 40 x 40 pixels.

3.2. Performance Analysis

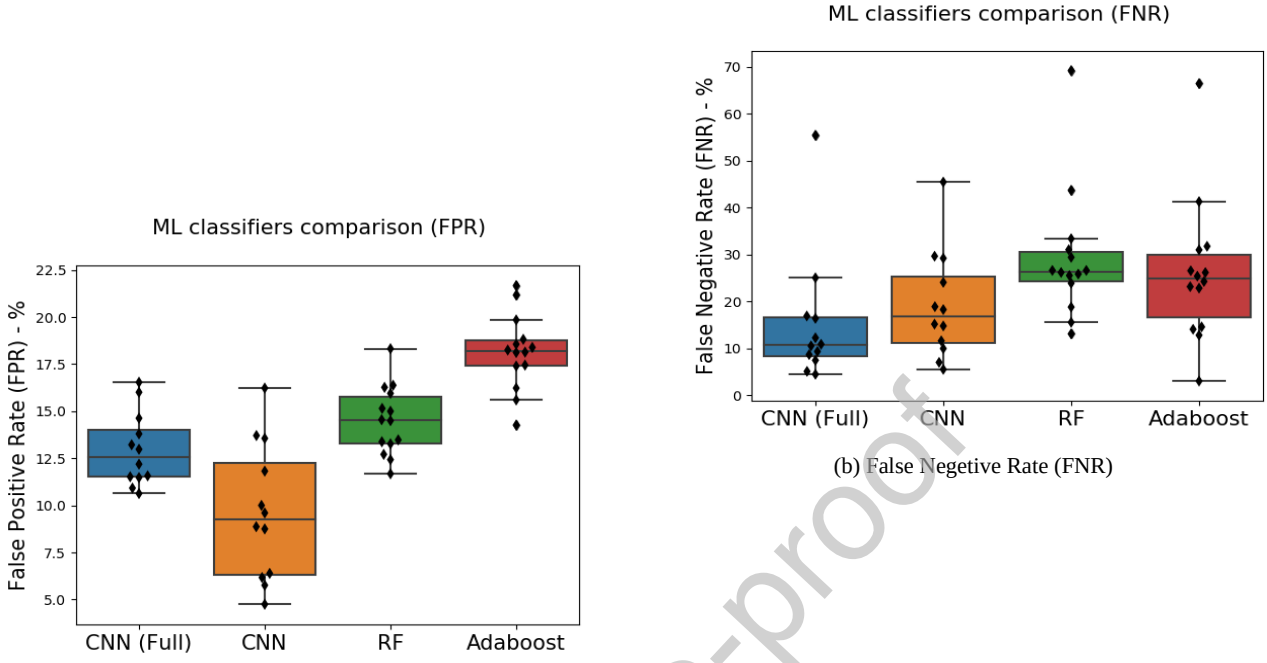
For the experiments, out of the 12 IR datasets, 11 datasets are used for training and the remaining set for testing. The experiments are carried out on all splits of datasets for evaluation in a leave-one-out setting. The results are shown in Tables 4-6. The experiments were performed using conventional ML classifiers (Sec. 2.3.2) and a deep learning based CNN (Sec. 2.4). The performance is reported by the recall, precision, False Positive Rate (FPR) (Table 5), unique ID (Table 6) and Accuracy (Table 4).

Only 30290 (out of the 13380450) image patches are true leakages. Due to this huge imbalance, the training is done on a balanced set only. For training with a balanced set, in total 60820 image patches are used for training, where 30290 of them are true and the rest of the 30530 patches are false (See Table 3 (b)). For CNN, after training with balanced datasets, the trained model is evaluated on both the balanced dataset (denoted as CNN (balanced)) and the full dataset (denoted as CNN (full)). For conventional ML classifiers, the training and testing are done only on a balanced dataset, for a first comparison.

The best results with the balanced set are found for CNN (balanced). An average weighted accuracy of 0.866 is achieved (see Table 4), which is better than all conventional ML classifiers. As CNN is the best, we also test it on the full dataset. We got an average weighted accuracy of 0.872 for CNN (full). For CNN, the corresponding False positive Rate (FPR) and False Negative Rate (FNR) are 12.7% and 10.4% for CNN (full), and 9.06% and 17.8% for CNN (balanced) sequences (See Figure 5). The corresponding recall, precision, and FPR are shown in Table 5.

Among the conventional ML classifiers, non-linear RF and AB performed best with an average weighted accuracy of 0.791 and 0.781, respectively. The results have been weighted by the total number of true patches. The detailed results, i.e. confusion matrices are given for CNN (Full), CNN (balanced), RF, and Adaboost in the Supplementary file. Some examples of image patches and the corresponding results in terms of true positive, false positive, false negative and true negative are visualized in Fig. 4. In the figure, 1 represents true leakage and 0 means non-leakage hot spot patches, e.g., the last row of Fig. 4, represents the false-negative patches.

The drone flies in one direction and then it rotates and returns to its starting position during image acquisition. In this process, there is an overlap of approximately 10-50 % from frame to frame during inspection. Therefore, the same annotated patches appeared in multiple frames.



(a) False Positive Rate (FPR)

(b) False Negetive Rate (FNR)

Fig. 5. Comparison in terms of FPR and FNR of CNN and conventional ML classifiers

Table 3. The 12 image sequences captured by UAV

(a) Overview of Datasets

Dataset / Sequences	Total # of images	Total # of patches	Percentage of false patches
S1	2150	117787	98.77%
S2	19549	834777	99.91%
S3	24334	1385508	99.80%
S4	43900	1982174	99.76%
S5	15455	2295396	99.76%
S6	13024	379048	99.82%
S7	12793	859589	99.88%
S8	28735	914790	99.88%
S9	32504	1692921	99.81%
S10	27108	535279	99.76%
S11	8676	1597745	99.64%
S12	14854	785436	99.74%
Total	243082	13380450	99.77%

(b) Number of true and false patches of the balanced dataset

Dataset / Sequences (12)	Total # of True	Total # of False	Size of balanced dataset
Total	30290	30530	60820

Table 4. Summary of performance evaluation in terms of accuracy

Dataset / Sequence (total 12 sequences)	KNN	DT	AB	RF	LR	L-SVM	LDA	GNB	CNN (balanced)
Weighted Average of 12 Sequences	0.746	0.703	0.781	0.791	0.778	0.779	0.778	0.735	**0.866**

Table 5. Performance evaluation using CNN for balanced and full datasets

Dataset / Sequence	Models	FPR (%)	Recall (%)	Precision (%)
Weighted Average of 12 Sequences	**CNN (Full)**	12.7 %	89.8 %	1.7 %
	CNN (balanced)	9.06 %	82.2 %	90.2 %

These were identified and marked as a unique ID. In our experiment, we evaluated how many annotated patches are identified in at least one of the regions by our proposed model. As shown in Table 6, our model can detect 98.6% of the true leakages evaluated by unique IDs. We refer to this as the caught rate.

3.3. Comparative Performance Evaluation

We have compared our results with our implementation of the RF method in (Berg et al., 2016) applied to our data set. As described above (See Section 1.1), they used a plane with a cooled thermal camera embedded for the leakage detection. Then a set of five ML classifiers, LDA, L-SVM, RBF-SVM, RF, and AB were applied to detect the leakages. The best result was found by the RF, which utilized 16 features. As the code was not available, we implemented our own version of the RF method in (Berg et al., 2016) utilizing 14 of the16 features and selecting the same set-up for the RF as detailed below. We refer to this as the Berg model. The original work by Berg et al. was applied to orthomosaic images and a different region extraction method than ours was used, so direct comparison is rather hard. Besides that, our version only uses 14 features as the pipeline information and housing segmentation is not available for our datasets. The 2 features not included are coverage and distance to houses. Otherwise, all features were implemented. (Flatness was computed as the standard deviation in a 5-pixel radius of the peak). As in (Berg et al., 2016), the RF consists of 120 trees with 1 randomly selected feature at each split. Features were calculated for each region of our region extraction algorithm and the RF model was trained on a balanced dataset with an equal amount of true and false samples. The model was evaluated on the full dataset on an equal footing with all other models investigated.

The comparative results are shown in Table 7. Our conventional ML classifiers with dense SIFT and SURF features provided slightly lower performance. We also evaluated the experiment with their features for balanced sets comparing with our conventional ML classifiers. Their method gave around 9.36 % improved weighted accuracy compared to our dense SIFT and dense SURF based RF. Therefore the most reasonable comparison is with the deep learning model CNN (full) as shown in Table 7.

In Table 7, our model gave higher recall and comparatively lower precision. This means our model gave comparatively higher False Positive Rate (FPR) and lower False Negative Rate (FNR), where our implementation of (Berg et al., 2016) gave lower FPR but higher FNR. It could be noted that higher FNR is much more critical compared to higher FPR in our application, since we don't want to overlook true leakages.

Accuracy is not always a good indicator when datasets are imbalanced. In our case, the datasets are imbalanced (for the full dataset), therefore, we focus on the caught rate. Our focus is to catch as many true leakages as possible. As shown in Table 7, our proposed model can detect around 98.6 %, where the Berg model (Berg et al., 2016) can detect around 95.2 % true leakage. Thus, our deep learning CNN achieved a 3.4 % higher caught rate.

3.4. Discussion

As per experimental results, CNN gave the best results for both balanced and full datasets. Among conventional classifiers, the best results are found for RF, AB, and L-SVM. The CNN (balanced) gave an improved average weighted accuracy of 9.48 %, 10.9 % and 11.2 % compared with RF, AB, and L-SVM, respectively. The comparisons are performed with CNN (balanced) because the experiments are carried out for balanced datasets for conventional ML classifiers.

For CNN (full), the precision and recall are 1.7 % and 89.8 %, respectively. This means we get higher False Positive Rate (FPR) (12.7 %) and lower False Negative Rate (FNR) (10.4 %) (See Figure 5). In our target application, the higher FNR is much more critical than a higher FPR, because we do not want to overlook a true leakage. However, a higher FPR means lower True Negative Rate (TNR) and vise-versa. In addition, we got a lower precision, which can be improved by improving the regional extraction algorithm. For comparison, we have also extracted the image patches with different parameters (Warm bias = 1.5 (Eq. 1) and Gradient Bias = 1 (Eq. 5)) to reduce the number of false patches. With these parameters, we got an improved precision rate of 5.10 %. It shows that there is a possibility to improve precision by improving the regional extraction algorithm. However, there is a risk to reduce the number of detected actual leakages in this process. Therefore, we didn't test new parameters, and further study is required.

We have compared our model with the method in (Berg et al., 2016) (See Sec. 3.3). However, a more recent paper for energy leakage detection is presented in (Zhong et al., 2019). We did not compare our model with that of (Zhong et al., 2019), since, initial testing of our implementation of the saliency map method on our data suggested poor performance, presumably caused by a lack of pipeline information and a more challenging dataset. Furthermore, the datasets in (Zhong et al., 2019) were rather limited. A total of four stitched images were used in their investigation, where three of them are from (Berg et al., 2016) (resolution of 1653 x 2870, 1150 x 3730 and 1181 x 598), which were from a plane and not from an UAV and only one UAV dataset, i.e., their own dataset (resolution of 682 x 2541). In contrast, we performed our experiment with extensive datasets (243082 full images in total). It should be noted that both states of the art methods have been evaluated on stitched images, whereas our models have been evaluated on each image separately. The reason is that, as per our analysis we have approximately 10-50% overlap from image to image during the inspection. Therefore a given leakage appears in multiple frames, which is identified as one unique ID. This gives us an advantage in terms of detecting the leakages. If we miss a leakage in one image, we have a high chance to detect the given leakage in other images.

Table 6. Evaluation of leakages by unique ID

Seq.	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	Weighted Avg
Unique ID	52	48	139	218	42	93	56	263	230	128	29	47	
Caught	52	48	135	216	41	93	50	263	229	125	29	46	
Percentage	100	100	97.1	99.1	97.6	100	89.3	100	99.6	97.7	100	97.9	98.6 %

Table 7. Performance comparison with an energy leakage detection technique

Dataset / Sequences	Accuracy	Recall (%)	FPR (%)	Precision (%)	Caught (%)
RF (Based on (Berg et al., 2016))	0.91	81.1 %	8.7 %	2.15 %	95.2 %
Proposed CNN model	0.87	89.8 %	12.7 %	1.7 %	98.6 %

In the stitched approach, there is only one chance to capture the true leakages. Furthermore, evaluation based on images may be performed immediately without stitching first. Even though the average weighted accuracy is around 0.872 %, the model successfully recognized 98.6 % of the true leakages.

It could be noted that the experimental results are highly influenced by the quality of IR sequences. As an example, the accuracy for S7 is poor compared with the rest of the sequences and the unique ID caught is only around 89.3%. The reason behind these poor results for S7 is that the weather was not so good during the capture of this sequence, and by visual inspection it is seen that the images are dark. Overall, our proposed model gave a fairly good accuracy for the energy leakage detection problem on the challenging real world datasets we processed.

4. Conclusions

We have presented a method for energy leakage detection in the underground pipes of district heating networks. The experiments are carried out for around 13.4 million image patches of district heating extracted from images captured by a UAV. Eight different ML classifiers and CNN were evaluated in this work.

For CNN, the training is done for balanced datasets and testing with both the balanced and full dataset. Generally, CNN (balanced) gave a lower false positive and false negative rate compared to conventional ML classifiers. The CNN gave FPR of 12.7 % and FNR of 10.4 % for full datasets.

During the inspection, the UAV images are highly overlapping, approximately 10-50 %, thus the same leakage appeared in many frames. Therefore each of the leakages is identified by a unique ID. As per our analysis, our CNN model detected 98.6 % of the unique IDs, which proves the usability of our proposed model.

In addition, we have also compared our method with the state-of-art method from the literature. Overall, the results show that our proposed method works well.

Declaration of Competing Interest

None

Acknowledgements

This research was funded in part by the Innovation Fund Denmark under the InnoBooster grant nr. 7041-00278B.

References

Axelsson, S.R.J., 1988. Thermal modeling for the estimation of energy losses from municipal heating networks using infrared thermography. IEEE Transactions on Geoscience and Remote Sensing 26, 686–692.

Berg, A., Ahlberg, J., 2014. Classifying district heating network leakages in aerial thermal imagery. Swedish Symposium on Image Analysis, 2014, Göteborg, Sweden.

Berg, A., Ahlberg, J., Felsberg, M., 2016. Enhanced analysis of thermographic images for monitoring of district heat pipe networks. Pattern Recognition Letters 83, pp. 215–223.

Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer-Verlag New York, Inc., Secaucus, NJ. Bøhm, B., Borgstrøm, M., 1996. A comparison of different methods for in-situ determination of heat losses from district heating pipes. Technical University of Denmark.

Breiman, L., 2001. Random Forests. Machine Learning 45, pp. 5–32.

Deitsch, S., Christlein, V., Berger, S., Buerhop-Lutz, C., Maier, A.K., Gallwitz, F., Riess, C., 2018. Automatic classification of defective photovoltaic module cells in electroluminescence images. CoRR abs/1807.02894 (on-revision).

Duda, R.O., Hart, P.E., Stork, D.G., 2001. Pattern Classification. 2nd ed. John Wiley & Sons.

Fröling, M., 2002. Environmental and thermal performance of district heating pipes. Chalmers University of Technology. Ph.D. thesis.

Hastie, T., Tibshirani, R., Friedman, J., 2008. The Elements of Statistical Learning. 2nd ed. Springer, New York, NY, USA.

Hossain, K., Villebro, F., Forchhammer, S., 2019. Leakage Detection in District Heating Systems Using UAV IR Images: Comparing Convolutional Neural Network and ML Classifiers. 27th EUSIPCO 2019 satellite workshop on Signal Processing, Computer Vision and Deep Learning for Autonomous Systems (Accepted).

Kingma, D.P., Ba, J.L., 2015. Adam: A method for stochastic optimization. ICLR, pp. 1–15.

Pedregosa, F. and Varoquaux, G. and Gramfort, A. and Michel, V. and Thirion, B. and Grisel, O. and Blondel, M. and Prettenhofer, P. and Weiss, R. and Dubourg, V. and Vanderplas, J. and Passos, A. and Cournapeau, D. and Brucher, M. and Perrot, M. and Duchesnay, E.(2011). "Scikit-learn: Machine Learning in Python". Journal of Machine Learning Research. 12: 2825–2830.

Chollet and Francois. et al., "Keras." https://keras.io (2015)

The HDF Group. Hierarchical Data Format, version 5, 1997-NNNN. https://www.hdfgroup.org/HDF5/.

Ljungberg, S.A., Rosengren, M., . Aerial thermography - a tool for detecting heat losses and defective insulation in building attics and district heating networks, in: Proc. SPIE 0780, Thermosense IX: Thermal Infrared Sensing For Diagnostics and Control, (11 May 1987).

Mucherino, Antonio, Papajorgji, J., P., Pardalos, M., P., 2009. K-Nearest Neighbor Classification. Springer New York, New York, NY. pp. 83–106. doi:10.1007/978-0-387-88615-2_4.

O. Friman, P. Follo, J.A., Sjökvist, S., 2014. Methods for large-scale monitoring of district heating systems using airborne thermography. IEEE Tran on Geosci.& Remote Sen 52, pp.5175–5182.

Olsson, M., 2001. Long-term thermal performance of polyurethaneinsulated district heating pipes. Chalmers Tekniska Hogskola, PhD thesis , i–46.

Puntanen, S., 2010. Linear regression analysis: Theory and computing by xin yan, xiao gang su. International Statistical Review 78, pp. 144–144.

Wang, X., Li, C., Yu, L., Hian, L., Deng, X., Yang, E., Ren, P., 2019. Uav first view landmark localization with active reinforcement learning. Pattern Recognition Letters, 125.

Ward, S., Hensler, J., Alsalam, B., Gonzalez, L.F., 2016. Autonomous UAVs wildlife detection using thermal imaging, predictive navigation and computer vision. IEEE Aerospace Conference, Big Sky, MT, pp. 1–8.

Webb, I., G., 2010. Naive Bayes. Springer US, Boston, MA. pp. 713–714.

Williams, G., 2011. Decision Trees. Springer New York, New York, NY. pp. 205–244. doi:10.1007/978-1-4419-9890-3_11.

Zhong, Y., Xua, Y., Wang, X., Jia, T., Xia, G., Ma, A., Zhang, L., 2019. Pipeline leakage detection for district heating systems using multisource data in mid- and high-latitude regions. ISPRS Journal of Photogrammetry and Remote Sensing 151, pp. 207–222.

Zhou, S., O'Neill, Z., O'Neill, C., 2018. A review of leakage detection methods for district heating networks. Applied Thermal Engineering 137, pp. 567–574.