

Transfer Learning 기법을 이용한 가스 누출 영역 분할 성능 비교

Performance Comparison of Gas Leak Region Segmentation Based on Transfer Learning

Marshall¹, Jang-Sik Park², Seong-Mi Park^{3*}

〈Abstract〉

Safety and security during the handling of hazardous materials is a great concern for anyone in the field. One driving point in the security field is the ability to detect the source of the danger and take action against it as quickly as possible. Via the usage of a fully convolutional network, it is possible to create the label map of an input image, indicating what object is occupying the specific area of the image. This research employs the usage of U-net, which was constructed in biomedical field segmentation to segment cells, instead of the original FCN. One of the challenges that this research faces is the availability of ground truth with precise labeling for the dataset. Testing the network after training resulted in some images where the network pronounces even better detail than the expected label map. With better detailed label map, the network might be able to produce better segmentation is something to be studied in further research.

Keywords : Gas leak detection, Image segmentation, Fully convolutional Network, Deep learning

1 경성대학교 전자공학과

2 경성대학교 전자공학과, 교수

3* 한국승강기대학교 승강기공학부, 부교수

E-mail: seongmi@kic.ac.kr

1 Main Author, Dept. of Electronic Eng., Kyung Sung University, Student

2 Author, Dept. of Electronic Eng., Kyung Sung University, Professor

3* Corresponding Author, Dept. of Lift Engineering, Korea Lift College, Associate Professor, E-mail: seongmi@klc.ac.kr

1. Introduction

The dangers of fire and other hazardous substances in any industries have always been a basic concerns for those involved in the fields. To fight against these dangers, people usually prepare soft (skill-wise e.g. safety training) and hard (equipment-wise e.g. sensors, extinguishers, compartment) measures. When an incident does occur, a quick response from any personnels will minimize the impact of damage could have been incurred due to it. For example, natural-gas processing, oil refineries, chemical plants, biochemical plants, wastewater treatment most likely involve the usage of chemical substances that are hazardous to human but vital for the processing plant. With a lot of equipments involved in large scale and number, perfectly inspecting all of them become a challenge in itself. Without a doubt, methods to help the maintenance and inspections of those equipments exist and are tested. This should be understood very well especially by those who are involved in such fields. However at the same time, those who can appreciate the existence of those safety measures should also be able to understand that those safety measures are put in layers with the exact reason that a fault can always take place even in the tightly set-up safety system.

Such small faults are the ones that are most dangerous as they can cause any compromised equipments to go unnoticed for

a long time and thus leading to accumulated wear and tear that will cause even more damage. Or even worse, the small leak that exists might only causes inconspicuous changes in the short term, before affecting the lives of the majority of the workers in said plant in the long term. Some chemical substances might not cause any apparent damage in the short term, however with accumulation in the body over time, it might damage the organs in the body beyond repair when the damages are showing their effect. As an example, in the case of natural-gas processing plant or other similar chemical plants which involves the usage or production of hydrocarbons, might not suffer from live threatening issue from small leaks [1], albeit the existence of report stating methane's potential to trigger rare respiration failure [2]. Instead the leakage of pure methane gas, as an odourless gas can go unnoticed until it reaches level which are dangerous for human (about 60% concentration in the air [1]), and even more so the potential of explosion due to the heat accumulated from the equipments. Furthermore, as methane gas is one of the greenhouse effect inducing gases, a "small" leak in any chemical plant translates to amount that might be detrimental to the environment [3], especially considering the high pressure and volume these equipments usually are.

Another research attempted to make use of preprocessed infrared camera feed to classify the existence of methane gas in the gas

emission [4]. This research instead tries to train a network to create the label map of input image, indicating the region of the gaseous object. In this manner, the trained network is expected to be able to handle gaseous-like objects, regardless of its real identity. This way the network is hoped to be able to work as a general method to recognize and locate the source of leakage more easily via the video feed of closed circuit television (CCTV) system monitoring the equipments.

Furthermore, the usage of transfer learning has been proposed as one of the way to supplement the lack of image data and label data in the study of deep learning [5, 6]. As we could not collect real-life data due to the danger potential it poses as well as the availability of such occasions to actually able to record the data, we employed transfer learning transfer learning to increase the complement our lack of training data.

2. Proposed Method

Image and video analysis in recent years are leaning toward the usage of Convolutional Neural Network (CNN) or other deep learning model and rely on the use of large amount of image data collected through the internet and Social Network Services (SNS). Deep learning techniques have improved substantially along the years, including in the field of video based surveillance.

CNN is largely used in the image processing field due to its ability to extract important information through the usage of convolutional layers. One of the most common application of CNN is to classify the object in an image. The general structure of CNN consists of layers or convolution filters, pooling function, and nonlinearity arranged as multiple sets and finally reaching the fully connected layers, which work as the core function of a neural network. The very first convolutional layer takes in an optimally resized image as its input. The convolution layers calculate the relationship of a certain pixel at a position with its surrounding neighbor pixels by using small squares of input data. These values then will be passed to the activation function, which will introduce non-linearity to the network and help in deciding the activation of the neuron. The pooling layers are there to reduce the number of parameters, and hence the computational cost, as well as allowing the reduction of spatial size of the data while the still retaining the most important information. And at the end of the network lies the fully-connected layers which hold the feature vectors obtained from composite and aggregated local features of the previous layers. These feature vectors will be then fed into a last activation function which will classify the output as one of the predetermined class labels.

A fully convolutional network (FCN), instead, replaces the fully connected layers

with up-sampling layers, allowing the extracted data to be reconstructed as the label map of the input image. In this manner, an FCN is able to take input image of arbitrary size and output a similar sized image with class labels assigned to them [7].

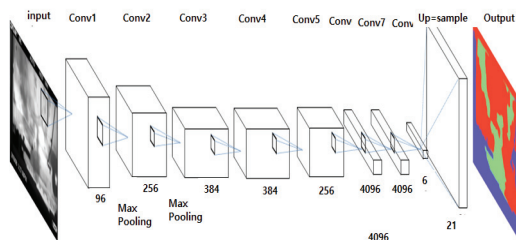


Fig. 1 A simplified illustration of FCN

After going through many convolutional layers, the outputted size is small compared to the original data. Then up-sampling these local features by 32 times (based on the reduced dimension) is done to create a label map with size similar or same to the original image. However, the resulting label map look coarse, since going through many convolutional and pooling layers means the resulting local features contain important features while having lost their spatial location information. To counter this issue, by applying a skip architecture, the network can combine the deep, coarse, semantic information and the shallow, fine, appearance information to enhance the final result, as described in figure 2.

Previously we employed FCN-AlexNet as our network and ended with limited result. In this study, we choose U-net as our fully

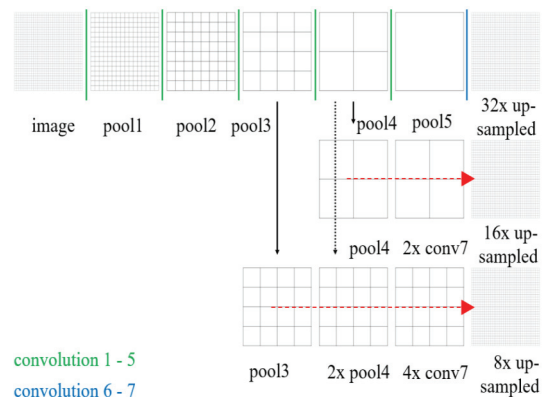


Fig. 2 Illustration of skip architecture used in FCN

convolutional network [8]. U-net is a more developed FCN originally intended for biomedical image segmentation.

U-Net is a model that generates an image of the same size as the input image through up-sampling based on the features extracted through the convolutional layer. At this time, in the upsampling process, the process of increasing the accuracy through the process of reusing the value of the convolution layer is performed. Figure 4 shows a model with an input size of 256x256x3 in the U-Net structure. The total number of learning factors is 31,031,685 [11].

The important network features of U-net are the usage of concatenation operator instead of summation operator at its skip connections and the symmetrical shape of the network as illustrated in figure 3.

Due to its symmetrical shape, U-net contains a large number of feature maps in the up-sampling path which enables information transfer. By combining the spatial information from the down-sampling step and

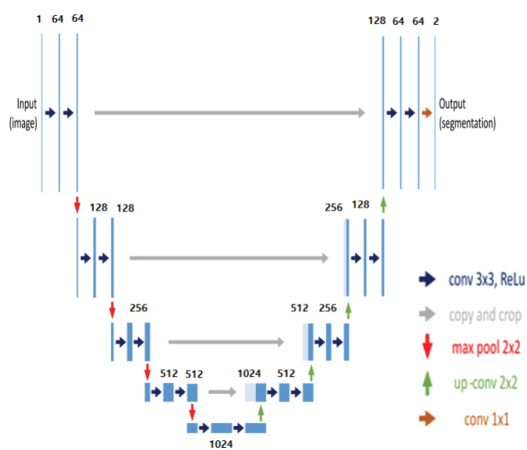


Fig. 3 U-net architecture, notice its shape is where its name come from

the contextual information from the up-sampling step, it can obtain the input image's general information which is important for image segmentation. On the other hand, a typical FCN only has number of classes feature maps in the up-sampling step.

This study also employs the usage of different architecture for the down-sampling path. Residual network(ResNet) is employed as the backbone of the modified network. ResNet's special feature is the usage of shortcut connection which provides identity mapping and allows the network to go deeper and optimizes with less degradation [9].

Transfer learning is a learning methodology used in machine learning. It is a concept that encompasses techniques that utilize knowledge of other relevant fields or tasks when learning a model to perform a specific task. For example, if there is only an input value and no label in the training data for

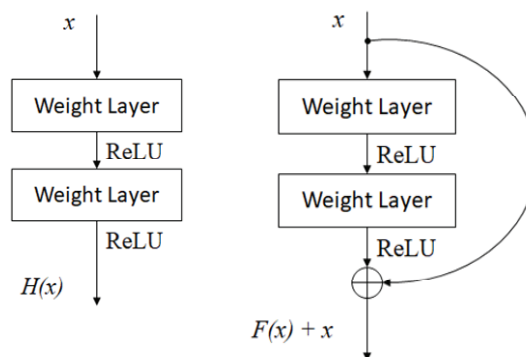


Fig. 4 Illustration of standard learning and residual learning block

the target task, the transfer learning is similar to constructing a predictive model using data from other domains.

Transfer learning is one of machine learning techniques that allows learning process to use different types of data, even when the form or marginal distribution of the training data and testing data is different [10]. It can be classified according to the domain and task which each data belongs to. A domain means the feature space and marginal distributions of a data. Transfer learning when the task for training and testing is the same with differing domain is called as domain adaptation.

To train any deep learning network properly, commonly a large size of training dataset is required. This is true for our gas leakage segmentation model as well. To obtain gas leakage images with features visible enough to be captured, sophisticated thermal camera is required. Since images need to be taken from the gas leakage site, the previously described potential dangers

and hazards become a real concern. In this paper, we propose a method to train deep learning models by using smoke images which share similar image characteristics to gas leak images.

3. Dataset

The dataset used for this study is made up from 3 different datasets: the gas leak dataset which contain gray scale images of taken from gas leakage videos 3 with classes of “gas leak”, “machine”, and “background”; the smoke dataset, which contains images of man-made smoke; and the Kaggle smoke detection dataset, which carries various images with binary mapping. Figure 5 shows some samples of images inside the dataset.

We obtained 839 images for training, 240 images for validation, and 120 images for testing; in total the dataset contains 1199 images.

4. Experimental Results

The network was trained on computer with Intel® Xeon E5-2650 CPU, 64GB RAM, and Geforce Quadro M5000 GPU on Ubuntu 16.04 operating system. As well as employing CUDA 10.0.130, cuDNN 7.5.0, Keras 2.2.0, tensorflow 2.0 and python 3.

This study makes use of U-net with

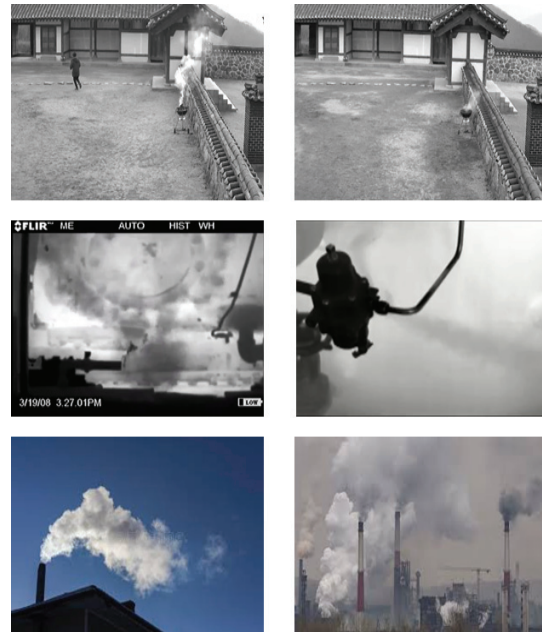


Fig. 5 Sample images from different dataset, for each row: Smoke dataset, Gas Leak dataset, Kaggle Smoke Detection dataset

backbone of ResNet-18. The network training makes use of Adam optimizer with initial learning rate of 1×10^{-4} and the network was trained for 20 epochs. The network is trained on the ‘gas-smoke’ and ‘machine’ class only, while the rest of the classes are considered as part of the ‘background’ class.

Data augmentation is also included during the training process to make a more robust data for training. Some of the data augmentation technique used includes horizontal flip, shifting, scaling, rotating, random crop, additive Gaussian noise, intensity level changes, sharpening and blurring, etc. A sample of the augmented dataset used for training is shown in figure 6 below.



Fig. 6 Augmented dataset used for training and its corresponding class map

This study observes the changes in performance when different activation functions are used in the network. The activation functions observed in this experiment are sigmoid, softmax, and linear. The performance here is calculated by taking the average intersection over union (IoU) score of 60 testing images. Table 1 describes the performance difference between the tested activation functions.

Table 1. Testing scores with different activation function

Parameters	Linear	Sigmoid	Softmax
Loss	0.41345	0.069833	0.1253
Mean IoU score	0.34182	0.80855	0.84532
Mean F1 score	0.4696	0.8908	0.91119

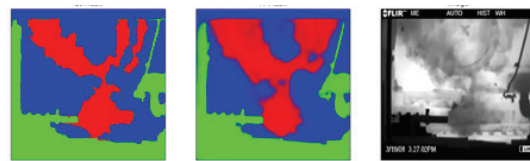
From the table above, we can see the network with linear activation function performs especially poorly with a mean score IoU of 0.34182. This is most likely due to the incompatibility of linear function in handling multi-class problems. The network with sigmoid activation function and softmax activation function performs decently with mean IoU score of 0.80855 and 0.84532 respectively, while it's worth to note that the

network with sigmoid activation function actually has less loss value compared to the latter.

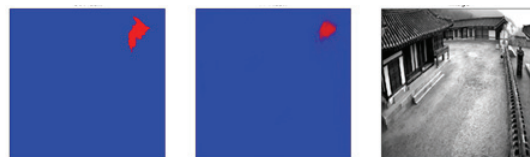
Figure 7 shows some of the resulting segmented images from the network with softmax activation function. From figure 7(a), (b), (c), (e), we can see that the network manages to segment “gas-smoke” and “machine” classes rather well, even the images which come from the Smoke dataset and Kaggle Smoke Detection dataset. However, in test image (d) and (f), we can see some mis-segmentation by the network, where on (d) the network simply failed to recognize a large portion of the smoke while still indicating some patches of segmented smoke, and on (e) the network might have assumed the cloud/fog/sky to appear similar to the “gas-smoke” class.



(a)



(b)



(c)

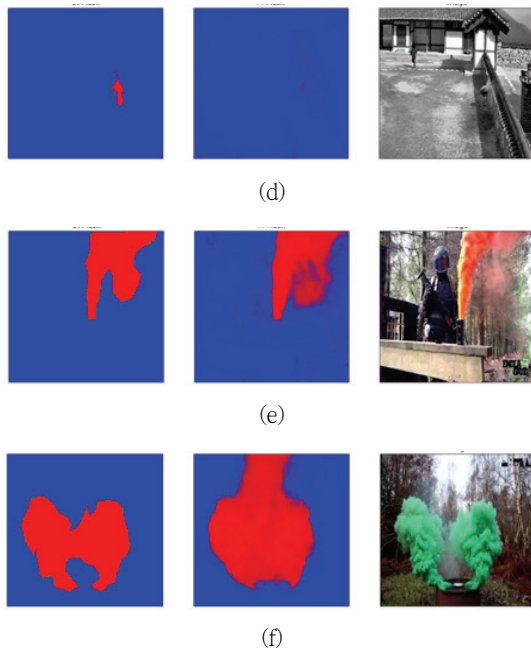


Fig. 7 The segmentation result using the testing dataset. The first column is the class map ground truth of the input image shown in the last column. The middle column shows the resulting image segmentation by the trained network.

4. Conclusion

In this study, we employ the fully convolutional network, U-net, with differing final activation function to perform semantic segmentation and make use of transfer learning to complement the lack of raining data for our network training. Due to U-net's structure allowing it to transfer more spatial information to its upward path, we obtain a much better semantic segmentation result compared to traditional FCN result. Training

on 1079 images of combined gas leakage and smoke images, resulted in network with softmax activation function with the highest mean IoU score of 0.84532 from the first 60 test images.

Acknowledgements

This research(or exhibition) was supported by Kyungsoong University Research Grants in 2019.

References

- [1] Duncan, I. J., "Does methane pose significant health and public safety hazards?—A review," *Environmental Geosciences*, vol. 22, no. 3, pp. 85-96, (2015).
- [2] Jo J. Y., Kwon Y. S., Lee J. W., Park J. S., Rho B. H., and Choi W. I., "Acute respiratory distress due to methane inhalation," *Tuberculosis and Respiratory Diseases*, Seoul, vol. 74, no. 3, pp. 120-123, (2013).
- [3] Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex V., and Midgley, P.M. (eds.), "Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change," Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, IPCC, pp. 1535, (2013).
- [4] Wang, J., Tchappmi, L. P., Ravikumar, A. P., McGuire, M., Bell, C. S., Zimmerle, D., Savarese S., and Brandt, A. R., "Machine vision for natural gas methane emissions detection

- using an infrared camera,” *Applied Energy*, vol.257, pp. 113998, (2020).
- [5] Thrun, S., Pratt, L., “Learning to learn,” Springer Science & Business Media, (2012).
- [6] Ganin, Y., et al., “Domain-adversarial training of neural networks,” *The Journal of Machine Learning Research* 17.1, pp. 2096-2030, (2015).
- [7] Long, J., Shelhamer, E., and Darrell, T., “Fully convolutional networks for semantic segmentation,” *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431-3440, (2015).
- [8] Ronneberger, O., Fischer, P., and Brox, T., “U-Net: Convolutional networks for biomedical image segmentation,” *International Conference on Medical image computing and computer-assisted intervention*, pp. 234-241, (2015).
- [9] He, K., Zhang, X., Ren, S., Sun, J., “Deep Residual Learning for Image Recognition,” *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 770-778, (2016).
- [10] Pan, S. J., and Yang, Q., “A survey on transfer learning,” *IEEE Transactions on Knowledge and Data Engineering*, 22.10, pp. 1345-1359, (2009).
- [11] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” *International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 234-241, (2015).

(Manuscript received April 13, 2020;

revised May 14, 2020; accepted May 21, 2020)