Video Based Smoke and Flame Detection Using Convolutional Neural Network

1st GeumYoung Son
Department of Electrical and Electronic Engineering
Kyungsung University
Busan, Republic of Korea
zhflgk135@ks.ac.kr

2nd Jang-Sik Park
Department of Electrical and Electronic Engineering
Kyungsung University
Busan, Republic of Korea
jsipark@ks.ac.kr

3rd Byung-Woo Yoon
Department of Electrical and Electronic Engineering
Kyungsung University
Busan, Republic of Korea
bwyoon@ks.ac.kr

4th Jong-Gwan Song
Department of Electrical and Electronic Engineering
Kyungsung University
Busan, Republic of Korea
jsong@ks.ac.kr

Abstract—**This paper proposes a method to detect fire using deep learning. Deep learning networks used in the proposal were AlexNet, GoogLeNet and VGG-16 in three ways. Image input from a closed-circuit television (CCTV) camera is classified into three different states (normal, smoke and flame) and then the network is trained to recognize each corresponding state. The datasets used consist of images coming from three areas which are housing, high-rise building, and mountain where fire incidents occurred. A quarter of all datasets were used as test images, half of them for validation, and the rest were used for testing purposes. Experimental results showed that all three network models were classifying fire detection at over ninety percent accuracy.**

Keywords—AlexNet, GoogLeNet, VGG-16l, smoke, flame, fire detection

I. INTRODUCTION

The fire incidents are currently on the rise again after its recent decline within the past 5 years. While the number of fire occurrences has decreased by 2,943 cases between 2011 and 2013, it has increased by an average of 4.2% annually since then. With no improvement to the casualties produced by fire incidents, property damage went up to ₩400 billion since 2013. In order to reduce the damage caused by fire, it is necessary to develop a fire monitoring system [1].

24-hour monitoring of commercial and residential areas is an effective way to reduce personal and property losses from fire. However, when the surveillance monitors the images of various cameras at the same time, difficulty to recognize fire at its early stage or late detection might become an issue. Therefore, it is necessary to study fire detection system that works on sensors and one that detects on surveillance camera.

The detection method using sensors constitutes of fire monitoring system which reacts to changes in a set temperature or a rapid temperature change, a smoke detector using an ionization type, or photoelectric type sensor. We have implemented a wireless multi-sensor fire detection system with carbon monoxide sensor and optoelectronic smoke sensor, fire detection and location alerting [2] and sensor combined algorithm to improve fire detection performance in the presence of obstacles [3], Distributed Temperature Sensing (DTS) which uses fiber optic to measure delicate temperature changes and to locate the heat origin [4], and research on high-sensitivity gas sensors based on metal oxides for realizing a high-performance fire detection system [5].

The fire detection method using CCTV camera uses real-time image analysis and deep learning to perform its functions. In recent years, there have been a lot of studies on the method of detecting smoke and flame in image using Convolution Neural Network (CNN) [6], using ViBe algorithm to extract and detecting fire and smoke regions [7], using the YOLO model to detect flame and sparks [8], and the disaster management system for detecting early fire and automatic response in indoor environment using CNN [9].

Sensor based fire detection has a high probability of malfunctioning due to the lack of intuitive information, and followed by the issue that it cannot deduce the initial fire position, smoke propagation direction, and scale of fire.

In the case of detecting fire by image, it can be intuitively seen to reduce the incidence of malfunction and it is also possible to use existing sensors to supply the other information that are hard to be obtained through image only. Another method is by using videos obtained from the existing CCTV cameras as the input for the fire detection system to reduce the expense for such system [10].

In this paper, a video analysis system using AlexNet, GoogLeNet and VGG-16 network detects three instances: normal, smoke and flames. A short introduction of each network will be described in section II Network Model, followed by section III describing how the datasets containing real fire incidents are used and how the experiment was carried out. And finally, section IV will have detailed result and observation from each network compared to one another with more suggestions that can be done to improve the video analysis based fire detection system.

II. DEEP LEARNING MODEL

The network models used for fire detection are taken from ImageNet Large Scale Visual Recognition Challenge (ILSVRC) championship: AlexNet which won the 2012 competition, GoogLeNet which won first place in the 2014 competition, and a VGG-16, the first runner up of 2014 ILSVRC. The three of them will be described in this section.

A. AlexNet

At ILSVRC championship 2012, AlexNet won the competition network model by bringing down the top five errors, which averaged at 26%, at other model networks to 15.3%. SuperVision Group, consisting of Alex Krizhevsky, Geoffrey Hinton and Ilya Sutskever, first presented AlexNet at ILSVRC 2012. As shown in Fig. 1, there is only one input image, but two GPUs are used. The first convolution layer takes 4x4 pixels from an input image with the size of 224x224x3 and filter it through 96 kernels with the size of 11x11x3 since it is the distance from the center of the acceptance field and the neighboring neurons in the kernel map. The second convolution layer receives the normalized pooling output of the first convolution layer as input and filters it into 256 kernels of size 5x5x48. The third convolution layer links 384 kernels of size 3x3x256 of the normalized pooling from the second convolution layer. The fourth convolution layer has 384 kernels of size 3x3x192 and the fifth convolution layer contains 256 kernels with the size of 3x3x192. Each fully connected layer carries 4,096 neurons. Fig. 1 illustrates the role between GPUs in the architecture. The input to the network is 150,528 dimensions, and the number of neurons in the other layer of AlexNet network is 253,440-186,624-64,896-64,896-43,264-4096-4096-1000, and the number of neurons decreases as the layer passes [11].

B. GoogLeNet

GoogLeNet won the ILSVRC 2014 competition with error rate less than half of AlexNet's 15.3% which won the ILSVRC in 2012. The top five errors in GoogLeNet are 6.67%, which is very close to human performance. CNN's structure has begun to change since GoogLeNet appeared in 2014. While previous networks were made up of 10 layers, GoogLeNet with 22 layers had shown improvement in the performance of CNN directly. GoogleNet uses 4 million parameters, which is about 12 times less compared to AlexNet (60 million). GoogLeNet performs convolution operations of 1x1, 3x3, and 5x5 with the same input image for better feature extraction. As shown in Fig. 2, while we perform 3x3 max-pooling, after reducing the channel by 1x1 convolution they are re-expanded by passing them through 3x3 and 5x5 convolution. Repeating this process 9 times will greatly reduce the amount of computation required at each next step. The 1x1 convolution appended after the max pooling layer is to make its result have the same number of
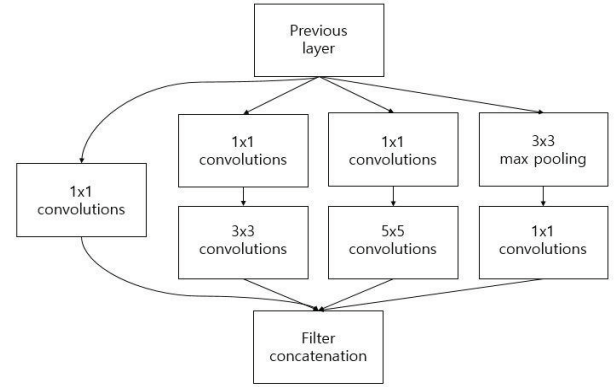

Fig. 2 1×1 filters in the reduction layer used before the 3×3 and 5×5 convolutions[12].

channels as its previous step [12].

C. VGG-16

The VGG model, as shown in Fig. 3, is much simply structured when compared to the complex GoogLeNet structure, and when the same image was run through the three models, the error produced by VGG was smaller than AlexNet. Since convolution on filter with even-numbered size cannot be done easily hence only odd-numbered filters are used. Excluding 1x1 filter which cannot perform clustering calculation, the most optimal filter size is 3x3. Having more number of the optimal filter will increase the performance. Hence a stride of 1 is always used whenever filter size of 3x3 appears on a convolution layer. Generally, VGG performs parallel processing on the starting filters; it starts with 64 filters and doubles at each following step. The max-pooling uses filter with the size of 2x2 and strides of 2 which halves the resulting feature map. Dropout applies only to the Fully-Connected (FC) layer with a rate of 0.5. Starts the initial learning rate at 0.1 and if the error does not decreases, slightly increases the learning rate gradually. Among the top-5 test results, the VGG-16 model error percentage is at 7.0% , 0.9% lower than the single model of GoogLeNet (7.9%), and the difference against the ensemble model of GoogLeNet is not even lower than 0.1%[13].

III. FIRE DETECTION METHOD

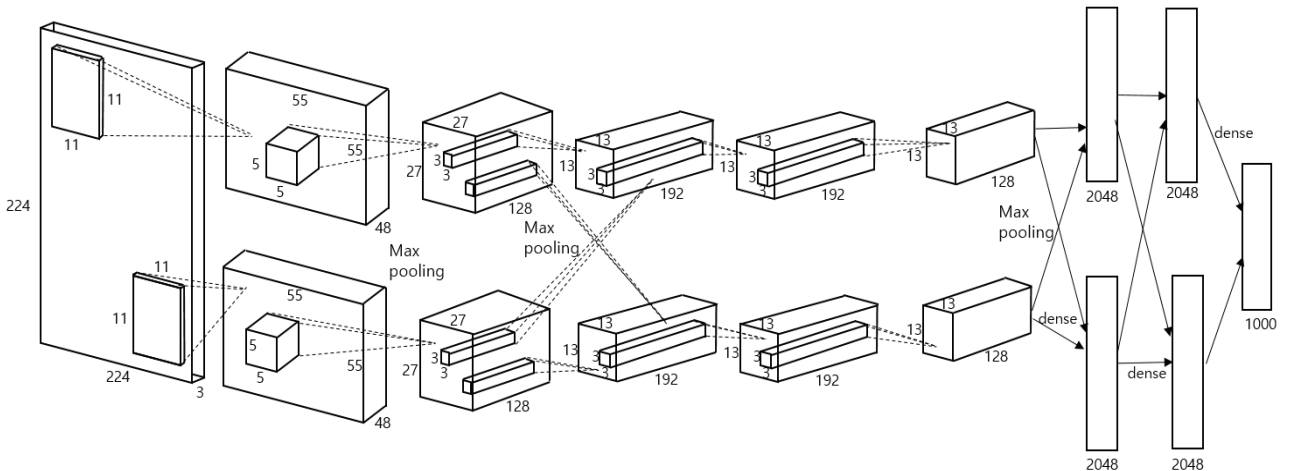We propose a method to detect smoke or flame by taking

Fig. 1 AlexNet model which works on 2 GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom[11].
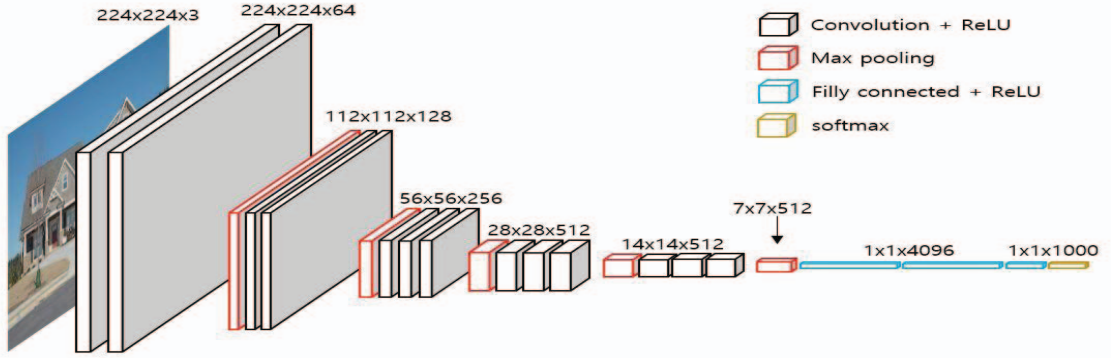
Fig. 3 VGG-16 Architecture model. The innovation comes from using small convolutional filters in a deeper neural network.

RGB colored video footages from CCTV camera and running them through deep learning networks (AlexNet, GoogLeNet and VGG-16) to detect instances of "normal" state, "smoke" state, and "flame" state. The fire dataset which consists of real fire incident images that occurred on building, mountain, et cetera, is used to train the network and then images taken at certain interval from a single channel CCTV camera is used as the input for the fire detection network to determine whether a fire occurs or not and alerts the monitoring personnel or related institutions when it is judged as a fire incident. Optimization on single channel CCTV can lead to application on multichannel CCTV fire detection system. Multichannel mode also stores the colored images from the CCTV cameras or videos every certain interval with the images' name following the cameras' ID and their location, and then using the deep learning network of fire detection system distinguishes the state of each image whether it contains "normal" event or "fire" event ("flame" or "smoke") and alerts the user or the related agencies to help them cope with the situation promptly. Further studies to

incorporate video analysis to supplement the malfunctions of sensor based fire detection system assuring a faster response time to cope with any fire incidence will be continued.

IV. SIMULATION RESULTS AND CONSIDERATIONS

In order to check the proposed deep learning based fire detection, we ran tests using the smoke and fire datasets with AlexNet, GoogLeNet, and VGG-16 respectively. Table 1 shows the specifications of the computer used in the tests. The datasets of AlexNet, GoogLeNet and VGG-16 were created by resizing them to 256x256 images, and the learning classifications were classified as "normal", "smoke", and "flame". Fig. 4 shows the image of each classification with some data of used dataset.

The experimental dataset is classified into 581 "normal" images, 343 "smoke" images and 609 "flame" images, and the total number of images is 1,533. The number of images used in the training process is 768 pieces, 50% of the total images, and the validation images and the test images use 383 pieces, which is 25% of the total image, respectively. The learning rate was set to 0.001, with the batch size of 50, and the training was done for 50 epochs. AlexNet's accuracy was 94.24%, close to 95.00%. As shown in Table 3, the "smoke" detection rate is 89.41% and the detection rate is lower than "normal" and "flame". The result of GoogLeNet's experiment yields "normal" and "flame" detection rate above 95.00%. "Smoke" detection was 95.29%, 5.88% higher than AlexNet's. The accuracy of GoogLeNet was 95.55%. The VGG-16 detected "smoke" at 100.00%, showing better performance than the other two networks, with an accuracy of 98.95%, making it the most accurate among the three networks. Table 3 shows the test results for each network

TABLE I. Experiment environment

Category	Version and Specifications
OS	Ubuntu16.04LTS
CPU	Intel® xeon(R) E5-2650
GPU	NVIDIA qudro p5000 16GB
Memory	Samsung DDR4 64GB

TABLE II. Number of dataset images

Dataset	Normal	Smoke	Flame	Total
Train	291	171	306	768
Validation	145	86	152	383
Test	145	86	152	382
Total	581	343	609	1,533

TABLE III. Test results of AlexNet, GoogLeNet and VGG-16

Network	class	Normal	Smoke	Flame	Accuracy
AlexNet	Normal	141	4	0	97.24%
	Smoke	3	76	6	89.41%
	Flame	3	6	143	94.08%
GoogLe-Net	Normal	142	3	0	97.93%
	Smoke	2	81	2	95.29%
	Flame	2	8	142	93.42%
VGG-16	Normal	145	0	0	100.00%
	Smoke	0	85	0	100.00%
	Flame	0	4	148	97.37%

Fig. 4 Three class ("Building", "Mountain" and "House") examples of our fire datasets ("normal", "smoke" and "flame")

V. CONCLUSIONS

In this paper, we propose and test fire detection method using AlexNet, GoogLenet and VGG-16. Using the 768 fire image dataset, we trained the network to recognize the "fire" event ("smoke" and "flame") and using its result, AlexNet showed more than 94.00% accuracy, GoogLeNet had more than 95.00% accuracy, and VGG-16 had more than 98.00% accuracy. Images that are erroneously identified commonly contain features similar to the classified images such as sunset, fog, mist, and so on. In order to reduce the false positives, it is necessary to collect various fire images and similar images in the next research to produce more accurate detection rate, and performing test and supplementing various detection algorithms. This study is intended to help fire prevention using CCTV.

REFERENCES

[1] National Security Agency, Announcement No. 2016-344," Fire Safety Policy Implementation Plan for 2017 in accordance with the 1st Basic Fire Safety Policy Basic Plan (2017~ 2021)". 2017.

[2] H. Hu, G. Wang, and Q. Zhang, "Design wireless multi-sensor fire detection and alarm system based on ARM," 9th International Conference on Electronic Measurement & Instruments, Aug. 2009, pp. 285-288

[3] A. Solórzano, J. Fonollosa, L. Fernández, J. Eichmann, and S. Marco, "Fire detection using a gas sensor array with sensor fusion algorithms," ISOCS/IEEE International Symposium on Olfaction and Electronic Nose (ISOEN), May. 2007, pp. 1-3

[4] H. Hoff, "Using Distributed Fibre Optic Sensors for Detecting Fires and Hot Rollers on Conveyor Belts," 2017 2nd International Conference for Fibre-optic and Photonic Sensors for Industrial and Safety Applications (OFSIS), July 2017, pp. 70-76

[5] K.-J. Lee, Y.-S. Lee, Y.-S. Shim, Y.-G. Song, S.-D. Han, and C.-Y. Kang, "Highly Sensitive Sensors Based on Metal-Oxide Nanocolumns for Fire Detection," SENSORS; Feb. 2017, 17 2, 11p.

[6] S. Frizzi, R. Kaabi, M. Bouchouicha, J.-M.Ginoux, Eric Moreau, and Farhat Fnaiech, "Convolutional neural network for video fire and smoke detection," IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society, Dec. 2016, pp. 877-882

[7] X. Wu, X. Lu, and H. Leung, "An adaptive threshold deep learning method for fire and smoke detection," 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Oct. 2017, pp. 1954-1959

[8] D. Shen, X. Chen, M. Nguyen, and W. Q. Yan, "Flame detection using deep learning," 2018 4th International Conference on Control, Automation and Robotics (ICCAR), Apr. 2018

[9] Y.-T. Do, "Visual Sensing of Fires Using Color and Dynamic Features," J. of Sensor Science and Technology, Vol. 21, No. 3, 2012, pp. 211-216

[10] M. Khan, A. Jamil, and B.-S. Wook, "Early fire detection using convolutional neural networks during surveillance for effective disaster management," NEUROCOMPUTING; May. 2 2018, 288 p30-p42, 13p.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Advances in Neural Information Processing Systems 25 (NIPS 2012), Dec. 2012, pp. 1097-1105

[12] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," Computer Vision and Pattern Recognition Wed, 17 Sep. 2014, pp. 1-9

[13] K. Simonyan, A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," ICLR 2015, arXiv:1409.1556v6 [cs.CV] , 10 Apr 2015