



Multi-Scale Prediction For Fire Detection Using Convolutional Neural Network

Myeongho Jeon, Department of Computational Science and Technology, Seoul National University, Seoul, Republic of Korea

Han-Soo Choi, Research Institute of Mathematics, Seoul National University, Seoul, Republic of Korea

Junho Lee, Department of Computational Science and Technology, Seoul National University, Seoul, Republic of Korea

Myungjoo Kang, Department of Mathematical Sciences, Seoul National University, Seoul, Republic of Korea*

Received: 5 January 2021/**Accepted:** 22 April 2021/**Published online:** 8 May 2021

Abstract. The automation of fire detection systems can reduce the loss of life and property by allowing a fast and accurate response to fire accidents. Although visual techniques have some advantages over sensor-based methods, conventional image processing-based methods frequently cause false alarms. Recent studies on convolutional neural networks have overcome these limitations and exhibited an outstanding performance in fire detection tasks. Nevertheless, previous studies have only used single-scale feature maps for fire image classification, which are insufficiently robust to fires of various sizes in the images. To address this issue, we propose a multi-scale prediction framework that exploits the feature maps of all the scales obtained by the deeply stacked convolutional layers. To utilize the feature maps of various scales in the final prediction, this paper proposes a feature-squeeze block. The feature-squeeze block squeezes the feature maps spatially and channel-wise to effectively use the information from the multi-scale prediction. Extensive evaluations demonstrate that the proposed method outperforms the state-of-the-art convolutional neural networks-based methods. As a result of the experiment, the proposed method shows 97.89% for F1-score and 0.0227 for false positive rate in the average of evaluations for multiple.

Keywords: Fire detection, Multi-scale prediction, Convolutional neural network, Deep learning

1. Introduction

Fire accidents cause numerous casualties and considerable property loss. To respond quickly to fire accidents, various fire detection technologies have been suggested. Fire detection is largely divided into fire sensor-based (smoke, flame, heat, etc.) detection and image processing-based detection with camera sensors.

*Correspondence should be addressed to: ; Myungjoo Kang, E-mail: mkang@snu.ac.kr
Myeongho Jeon and Han-Soo Choi have contributed equally to this work.



Previous studies on fire detection technology have mainly focused on fire sensors, but sensor-based detection methods have the limitation that the performance of the system can be considerably reduced due to the effects of wind, ultraviolet light, and ambient temperature. In addition, there is a delay until the elements created by the fire (smoke, heat, etc.) have reached the sensor.

On the other hand, image processing-based fire detection using a camera has the following strengths: (1) The cost of infrastructure construction can be reduced by using cameras such as closed-circuit televisions (CCTVs) that are already installed for other surveillance purposes. (2) Since the image is captured in real-time by the camera, flames can be detected quickly compared to detection in response to particles.

Conventional image processing-based methods capture a fire's features such as color, shape, flickering, frequency, dynamic textures. The RGB, YUV, YCbCr, and CIE Lab color spaces have been used to detect fire [1–4]. Motion information has been used in addition to color information. Methods such as background subtraction and optical flow analysis have been used to distinguish fire from other objects, which increased the utilization of the movement of the flame for detecting the fire [5, 6]. Probability-based machine learning methods that use various features of fire have also been proposed [7–9]. Despite these studies on image processing-based methods, sensor-based methods have been widely used in the field. Because the conventional image processing-based methods were not robust enough to be utilized in the field.

Since deep learning-based methods, especially convolutional neural network (CNN)-based methods, have exhibited an outstanding performance in visual tasks, several studies have recently used neural networks for fire detection. In the whole process of fire detection system, the frames obtained from the CCTV video are used as input to the CNN-based method and the prediction result is transmitted to the alarm system. Most existing methods [10–19] use a sequential CNN model, consisting of convolutional, down-sampling, and fully connected layers. The deeply stacked layers effectively extract various features of the input images, which achieve high performance in fire detection while reducing false alarms. The existing CNN models, i.e. GoogLeNet [20] and VGGNet [21], have been applied in fire detection tasks by using a fine-tuning technique [22, 23]. For real-time fire detection, Muhammad et al. [24] have applied MobileNet [28], which has a similar performance to other CNN models but considerably reduces the computational cost.

Although some methods using multi-scale [25, 26] were presented for object detection task, all CNN-based methods for fire image classification exploited a single-scale prediction, which uses only the last feature maps of the network in the prediction. However, given varying sizes of fires in images, multi-scale feature maps can aid the robustness of the Prediction for model. As the input passes through the network, the scale of the feature map decreases, and its receptive field expands. The larger receptive field of the feature map means that one pixel of the feature map represents a larger range of the feature in the input image (see Fig. 1). The large receptive field of the last feature maps is suitable for detecting large fires. However, it is difficult to activate a small area in the image because the concentration of the convolution filter is distributed. To reliably activate the

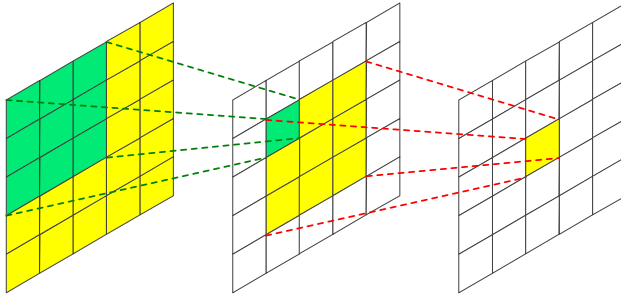


Figure 1. Receptive field.

region where a small fire exists, it is effective to use a small receptive field. The size of a fire in this study refers to what proportion of the image the fire occupies within the image rather than the fire's actual size (see Fig. 6).

Inspired by Ren et al. [27], we propose a multi-scale prediction that exploits the feature maps of all the scales. To the best of our knowledge, this paper apply multi-scale feature map prediction on fire image classification task first. The proposed multi-scale prediction does not simply use the feature maps of all the scales for prediction, but it is a weighted voting algorithm that predicts fire through feature-squeeze (FS) blocks and softmax functions at each scale, and then synthesizes those predictions to make the final prediction. An FS block is a block that squeezes the activated region of the feature maps and allows the network to make a prediction for each scale based on the squeezed information. In Section 4, we demonstrate, with extensive evaluation, that the proposed method effectively predicts fire. The proposed model can process images at 364 frames per second (FPS) in an Intel Core i7-9700K CPU and NVIDIA Titan RTX GPU environment, so it can be used as a real-time surveillance model in the field. In summary, the contributions of this work are as follows:

1. We propose a multi-scale prediction model to capture flames of various sizes in the input images. The multi-scale prediction significantly reduces the false positive rate (FPR) by responding robustly to various resolutions resulting from the distance between the camera and the fire.
2. We propose an FS block to utilize the feature maps of various scales for final prediction. The FS block squeezes the feature maps spatially and channel-wise to effectively use their information.
3. We perform experiments on various data sets and demonstrate that the proposed method outperforms existing state-of-the-art methods. In particular, we use an unseen data set for training to evaluate the generalization of our proposed method.

The remainder of this paper is organized as follows: In Sect. 2, we briefly comment on recent studies on image-based fire detection methods. In Sect. 3, we describe the proposed method for fire detection. In Sect. 4, we demonstrate the validity of the proposed method through experiments. Finally, Sect. 5 presents the

conclusions.

2. Related Work

To detect fire, various visual methods have been proposed. All can be categorized into two parts: conventional image processing-based methods and CNN-based methods. We will briefly review these methods below.

2.1. Conventional Image Processing-Based Fire Detection Methods

Chen et al. [1] extracted fire-pixels by deducing the intensity and saturation of the R component in the RGB channel. Töreyn et al. [29] used spatial and temporal wavelet transforms to detect quasi-periodic behavior and color variations in flames. Marbach et al. [2] used the YUV color model to extract the candidate flame region. Celik et al. [5] proposed a background subtraction method to segment the fire candidate pixels and a generic statistical model for refined fire-pixel classification. Zhang et al. [30] combined wavelet transforms and Fourier transforms to detect forest fire. Celik et al. [3] used the YCbCr color space to separate luminance from chrominance. Ko et al. [7] applied a support vector machine after removing non-fire pixels using a luminance map. Celik [4] used the CIE Lab color space for fire pixel detection. Günay et al. [31] used Markov models to differentiate flame movement from other object movements with spatial wavelet transforms for color variation in a fire. Borges et al. [32] analyzed frame-to-frame changes in specific low-level features to estimate potential fire regions. Qiu et al. [33] first detected the coarse and superfluous edges in a fire image and then identified the edges of the fire and removed the irrelevant artifacts. Rinsurongkawong et al. [6] used the Lucas-Kanade optical flow algorithm to detect fire in real time in a video stream. The method detected the exact fire location and spread direction. In Wang et al. [34], the feature vectors extracted based on color and the Wald-Wolfowitz randomness test is applied to the motion probabilities to obtain the prior flame probability. Subsequently, the convolution operation is used to update the prior probability into a posterior probability. Foggia et al. [35] used complementary information from a multi-expert system and bag-of-words approach to represent motion. Zhou et al. [8] proposed a shape features selection preference based on AdaBoost for fire detection. Kong et al. [9] determined the candidate fire region through the color component ratio and the motion cue of the flame. Logistic regression was used to identify the genuine flame.

2.2. CNN-Based Fire Detection Methods

To address the frequent false alarms of conventional image processing-based fire detection methods, attempts have been made to apply CNN to fire detection tasks. Zhang et al. [15] used both the full image and the fine-grained image to train the CNN model, demonstrating that a deep learning-based method can be used in the field of fire detection. Frizzi et al. [13] applied a CNN model to detect fire and smoke. The CNN model performed a feature extraction and classification

within the same architecture. Maksymiv et al. [17] used a combination of Ada-boost and local binary patterns to obtain the region of interest, which is used as input for the CNN model. Sharma et al. [16] proposed a data set assembled to replicate real-world scenarios. The data set was utilized to validate VGG16 [21] and ResNet50 [36]. Muhammad et al. [22] applied GoogLeNet, proposed by [20]. They used a transfer learning technique to balance efficiency and accuracy. Mao et al. [10] proposed a multi-channel CNN. Muhammad et al. [23] fine-tuned Alex-Net, proposed by [37], to detect fire in varying indoor and outdoor environments. They also proposed an adaptive prioritization mechanism to adaptively switch the status of the camera nodes. Dunnings et al. [14] proposed Inception V1-OnFire, which is a modified version of the GoogLeNet model to detect fire more efficiently. Jadon et al. [11] designed a CNN model to deal with the issues of both performance and model size. Saeed et al. [12] proposed three different models using both sensor data and images for fire detection. They are hybrid models using Adaboost and MLP networks, Adaboost-LBP, and CNNs. Muhammad et al. [24] proposed a CNN-based method, which requires sufficient light to be utilized in surveillance networks. Zhang et al. [38] proposed a CNN architecture consisting of three convolutional layers and three fully connected layers. Multi-scale object detection models are presented. Li et al. [25] used recently presented object detection methods, Faster-RCNN, R-FCN, SSD, and YOLO v3, to detect fire. Wu et al. [26] proposed a tiny version of YOLO, which is practical for forest safety. Yang et al. [18] presented cost-effective CNN-based method to be implemented in lightweight devices. Li et al. [19] proposed multi-scale feature extraction, implicit deep supervision, and channel attention mechanism to balance between accuracy, model size, and speed.

3. Proposed Approach

The proposed method extracts the features of a given image through deep-stacked layers. The densely connected residual (DCR) block and the convolution operation gradually reduce the scale of the feature maps, which, in turn, become a 2×1 vector (see Sect. 3.2 for the DCR block). The feature maps of different scales obtained while passing through the entire network are all used in a multi-scale prediction (see Fig. 2 and Table. 1).

3.1. Multi-Scale Prediction

Since detecting a fire only with a final-scale feature map may cause false alarms or fail to detect small fires, we propose a multi-scale prediction framework for fire detection. The last feature maps of each scale are used to predict fires in the binaries via the FS block and softmax function. These obtained predictions are used to determine the final prediction through a weighted voting algorithm. The operation can be expressed as follows:

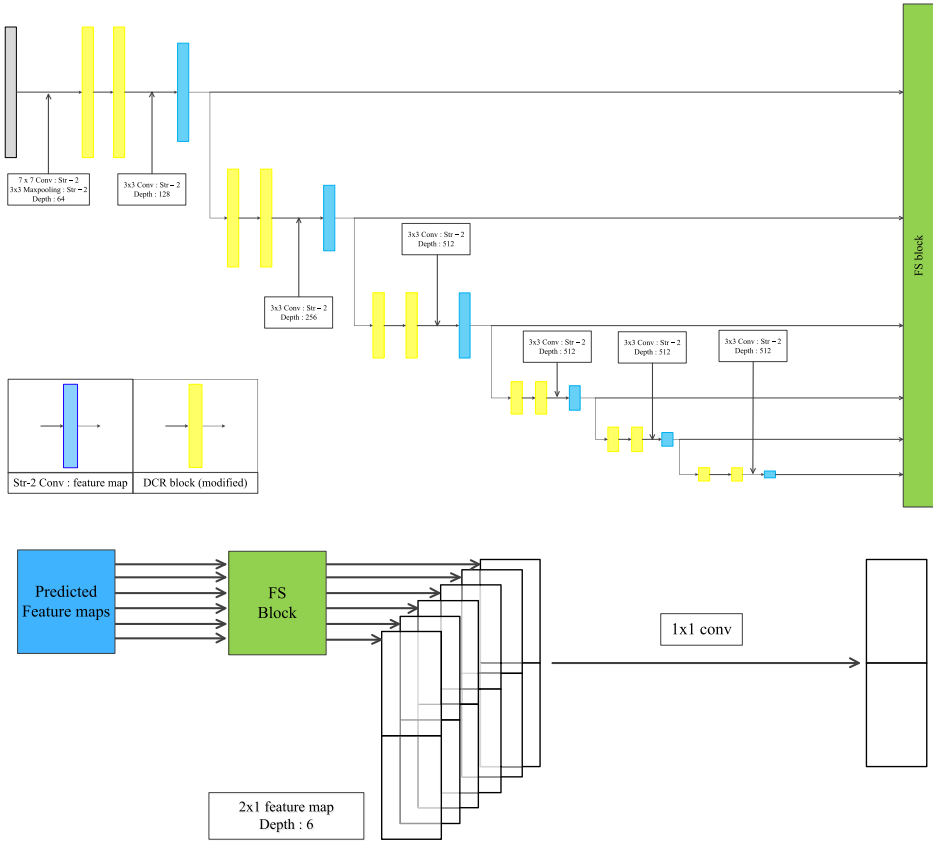


Figure 2. Proposed architecture.

$$f_{\text{softmax}}(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)} \quad (1)$$

$$y_1 = f_{fs}(X_1), y_2 = f_{fs}(X_2), \dots, y_n = f_{fs}(X_n), \quad (2)$$

$$y_{\text{prediction}} = f_{\text{softmax}}(f_w([y_1, y_2, \dots, y_n])), \quad (3)$$

where f_{softmax} denotes the softmax function, f_{fs} denotes the FS block (see Sect. 3.3), f_w denotes the weight voting algorithm and $y_{\text{prediction}}$ denotes 2×1 prediction vector.

The reasons for using a multi-scale prediction framework are twofold: (1) Using receptive fields of different sizes is suitable for extracting the features of fires of

Table 1
Summary of Proposed Architecture

Block type	Layers	Size of feature maps
Inputs		$224 \times 224 \times 3$
Entrance	7×7 conv(str-2) Maxpooling(str-2)	$56 \times 56 \times 64$
Downscaling1	DCR block $\times 2$ + conv(str-2)	$28 \times 28 \times 128$
Downscaling2	DCR block $\times 2$ + conv(str-2)	$14 \times 14 \times 256$
downscaling3	DCR block $\times 2$ + conv(str-2)	$7 \times 7 \times 512$
downscaling4	DCR block $\times 2$ + conv(str-2)	$4 \times 4 \times 512$
Downscaling5	DCR block $\times 2$ + conv(str-2)	$2 \times 2 \times 512$
Downscaling6	DCR block $\times 2$ + conv(str-2)	$1 \times 1 \times 512$
Prediction	CMFC block + concat	$2 \times 1 \times 6$
Output	1×1 conv + softmax	2×1

various sizes in an image. (2) The robustness of the model increases because it comprehensively detects fires through feature maps of different sizes for fires.

The size of the receptive field covered by one pixel of the feature map differs according to the depth of the stacked convolutional layers, which makes the size of the area to be focused on in the image different (see Fig. 1 in Sect. 1). Therefore, small fires can be effectively detected using small receptive fields, and large receptive fields are required for large fires to be covered.

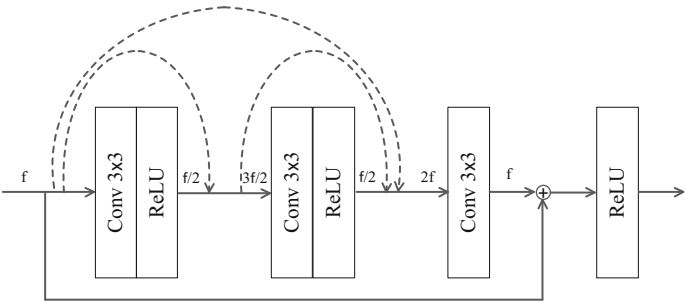


Figure 3. Densely connected residual block.

3.2. Densely Connected Residual Block

In this study, a DCR block, proposed by Park et al. [39], is used to extract the features of fire. The DCR block receives the feature maps as input and outputs feature maps of the same scale and the same depth. The DCR block consists of three convolution layers and three activation layers with a dense connection and a residual connection (see Fig. 3). These connections provide a sufficient flow of information to extract various features of the fire images. The gradient vanishing phenomenon that occurs during backpropagation is prevented by the connection between several different layers. This makes the training process of the entire network efficient and effective. We use ReLU as an activation function instead of PReLU, which was applied in the existing DCR block. We modify the position of ReLU from before-addition to after-addition because, as shown in [40], ReLU before addition makes the forward propagated signal increase monotonically and worsens the performance.

3.3. Feature Squeeze Block

To detect the fire from the feature maps of each scale, we propose an FS block. The FS block receives the feature maps as an input and outputs the 2×1 vector for binary prediction regardless of the feature maps' scale. The softmax function is used as the activation function of the 2×1 vector to calculate the confidence of fire or non-fire. The feature maps are squeezed to fire and no-fire confidences in two steps: (1) Multiple feature maps are channel-wise squeezed to two feature maps by a convolution operation. (2) Two feature maps are spatially squeezed to the softmax vector through a max-pooling operation and the softmax function (see Fig. 4). These two steps can be expressed as follows:

$$F = f_{\text{conv}}(X) \tag{4}$$

$$y_i = f_{\text{softmax}}(f_{fc}(f_{\text{max}}(F))), \tag{5}$$

where f_{conv} denotes the 1×1 convolution operation, f_{max} denotes the max-pooling operation, f_{softmax} denotes the softmax function, and f_{fc} denotes the fully connected layer (see Eq. 1).

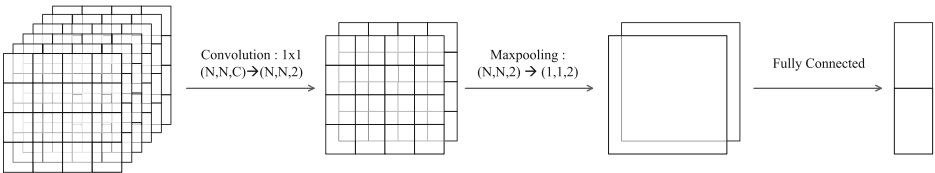


Figure 4. Feature-squeeze block.

4. Experimental Result

4.1. Data Set

In this study, the RESCURE [41], BoW Fire [42], and Sharma [16] data sets are used for training and evaluation, and the Still [43] data set is used to verify the validity of the proposed method. The RESCURE data set consists of videos depicting fires at construction sites. We extract one image every 20 frames of the video clips. This is because extracting all the frames of video results in too many similar images, which causes overfitting when training the CNN model. The BoW and Sharma data sets cover various fires in different situations. The Still data set is a data set for fire segmentation and classification. We use only fire and no-fire among the images labeled fire, smoke, and no-fire. The Still data set reflects reality well in that it contains no-fire images that can be misunderstood as fire. We randomly divide each data set into a train set and validation set at a ratio of 8:2 (see Table 2). All the Still data set images are used only for testing to evaluate the proposed method more rigorously.

To train the model to accurately predict fire in diverse environments, we apply two data augmentation techniques: horizontal flipping and a brightness adjustment (see Fig.5). Since other symmetry and rotation augmentation techniques deteriorate the features of the fire, only horizontal flipping is performed. The brightness of the images is adjusted using gamma correction. The gamma correction can be expressed as follows:

$$T_{\gamma}(x) = (\frac{x}{255})^{\frac{1}{\gamma}} \times 255,$$
 (6)

where x denotes RGB pixels and γ denotes the degree of brightness. We set γ to 0.5 and 1.2.

4.2. Results

We extensively evaluate the model by using several metrics (see Eqs. 7, 8, 9, 10 and 11). The models in Tables 3, 4, 5, 6, 7 and 8 are newly trained according to the learning conditions mentioned in each paper by adding the augmentation technique in Sect. 4.1. All the methods are trained on a combined set of RESCURE,

Table 2
Experimental Data Set. Number of Images for Fire, No-Fire, Train Set, and Evaluation Set

	Fire	No-fire	Train	Evaluation
RESCURE [41]	1007	406	1130	283
BoW Fire [42]	119	107	180	46
Sharma [16]	110	541	520	131
Still [43]	1775	4611	—	6386



Figure 5. Data augmentation example: horizontal flipping and bright adjustment.

Table 3
Accuracy Result

	Entire(All)	RESCURE	BoW	Sharma
Muhammad et al. [22]	0.962205	0.957597	0.934066	0.977011
Muhammad et al. [23]	0.883465	0.929329	0.802198	0.862069
Jadon et al. [11]	0.899213	0.964664	0.780220	0.842912
Muhammad et al. [24]	0.927591	0.968198	0.835165	0.927203
Yang et al. [18]	0.922835	0.943463	0.879121	0.915709
Li et al. [19]	0.941732	0.954064	0.879121	0.950192
Proposed Method	0.979528	0.982332	0.934065	0.992337

Bold values are the best one

Table 4
Precision Result

	Entire(All)	RESCURE	BoW	Sharma
Muhammad et al. [22]	0.949206	0.951691	0.947368	0.941176
Muhammad et al. [23]	0.887755	0.954317	0.854167	0.653061
Jadon et al. [11]	0.894040	0.979695	0.833333	0.604167
Muhammad et al. [24]	0.940625	1.000000	0.743590	0.952607
Yang et al. [18]	0.941368	0.984965	0.859649	0.862745
Li et al. [19]	0.954397	0.969849	0.894737	0.960784
Proposed Method	0.977273	0.980198	0.947368	1.00000

Bold values are the best one

BoW, Sharma data sets and evaluated on each of them to confirm whether the model is trained with a bias to a certain data set. Additional experiment on Still data set is dedicated to evaluate whether it is robust to the data set not used for training. Compared with state-of-the-art CNN-based fire detection methods for the RESCURE, BoW, and Sharma data sets, the proposed model exhibits outstanding performance evenly across multiple evaluation metrics (see Tabsles 3, 4, 5, 6 and 7). Since evaluation metrics conflict with one another, a model focused only on a specific evaluation metric is not stable. The proposed method outper-

Table 5
Recall Result

	Entire(All)	RESCURE	BoW	Sharma
Muhammad et al. [22]	0.949206	0.989950	0.947368	0.941176
Muhammad et al. [23]	0.864238	0.944724	0.788462	0.627451
Jadon et al. [11]	0.894040	0.969849	0.769231	0.568627
Muhammad et al. [24]	0.917683	0.892857	0.852941	0.957143
Yang et al. [18]	0.954397	0.969849	0.894737	0.960784
Li et al. [19]	0.927215	0.965000	0.910714	0.816666
Proposed Method	0.980456	0.994974	0.947368	0.960784

Bold values are the best one

Table 6
F1-Score Result

	Entire(All)	RESCURE	BoW	Sharma
Muhammad et al. [22]	0.949206	0.970443	0.947368	0.941176
Muhammad et al. [23]	0.875839	0.949495	0.820001	0.640000
Jadon et al. [11]	0.894040	0.974747	0.800000	0.585859
Muhammad et al. [24]	0.929012	0.943396	0.794521	0.954869
Yang et al. [18]	0.921850	0.960784	0.899083	0.800000
Li et al. [19]	0.940610	0.967419	0.902655	0.882883
Proposed Method	0.978862	0.987531	0.947368	0.980000

Bold values are the best one

Table 7
False Positive Rate Result

	Entire(All)	RESCURE	BoW	Sharma
Muhammad et al. [22]	2.519685	3.533569	3.296703	1.149425
Muhammad et al. [23]	5.196850	3.180212	7.692308	6.513410
Jadon et al. [11]	5.039370	1.413428	8.791209	5.263158
Muhammad et al. [24]	2.992126	0.000000	10.989011	3.831417
Yang et al. [18]	5.714286	4.054054	20.512821	3.465347
Li et al. [19]	4.388715	7.228916	17.142857	0.995025
Proposed Method	0.022727	0.019802	0.052632	0.000000

Bold values are the best one

forms the state-of-the-art methods even for the Still data set, which is not used for training (see Table 8). This means that the proposed method can detect fire in various environments. FPS of the proposed method is 364.8511 on a computer system equipped with Intel Core i7-9700K CPU and NVIDIA Titan RTX GPU.

Table 8
Evaluation Result on Still Image Data Set

	Accuracy	Precision	Recall	F1-score	FPR
Muhammad et al. [22]	0.873920	0.767836	0.760935	0.764370	8.382450
Muhammad et al. [23]	0.827680	0.657951	0.738766	0.696020	13.992580
Jadon et al. [11]	0.792800	0.601630	0.663271	0.630948	16.000870
Muhammad et al. [24]	0.764960	0.541493	0.781905	0.639862	24.121300
Yang et al. [18]	0.781600	0.742960	0.569853	0.644993	10.530191
Li et al. [19]	0.868160	0.820851	0.722955	0.768799	6.865672
Proposed Method	0.887680	0.822548	0.738766	0.778409	5.806592

Bold values are the best one

$$\text{Accuracy} = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \quad (7)$$

$$\text{Precision} = \frac{T_P}{T_P + F_P} \quad (8)$$

$$\text{Recall} = \frac{T_P}{T_P + T_N} \quad (9)$$

$$F_1 \text{ score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

$$\text{False Positive Rate (FPR)} = \frac{F_P}{F_P + T_N} \quad (11)$$

4.3. Analysis

Although the proposed method is generally good in performance, it is not the best for all data sets and evaluation metrics (see Table 3, 4, 5, 6, 7 and 8). These results can be explained through (1) quantitative comparisons between data sets and (2) between evaluation metrics. Comparing the performance of the trained model for each evaluation set indicates whether the training is biased towards a specific data set. Additional experiments on Still data set, which is unseen during training, show the generalization ability of method for fire images from multiple data sets. In particular, because precision and recall are mutually exclusive, F1-score, the harmonic mean of these two indicators, represents them. Recall, precision and FPR can be referenced in terms of application, but accuracy and F1-score are suitable for objective model performance evaluation. From these points

Table 9
The Ablation Study for Single-Scale and Multi-Scale Prediction

	Accuracy	Precision	Recall	F1-score	FPR
Single-scale	0.870720	0.765905	0.742960	0.754258	8.279903
Multi-part	0.886080	0.817940	0.737567	0.775677	5.981227
Multi-all (proposed)	0.887680	0.822548	0.738766	0.778409	5.806592

Bold values are the best one
In Multi-Part, We Used The Feature Maps of 28×28 , 7×7 , and 2×2 Sizes

of view, the overall performance on the evaluation set that combines the three data sets show the proposed method outperforms the state-of-the-art models. Furthermore, the results for Still data set demonstrate proposed method has good generalization ability.

Additional experiments were conducted to check the contribution of the multi-scale prediction. An ablation study for the multi-scale prediction showed that using feature maps of multiple scales contributes to fire detection performance (see Table 9).

We examined the softmax vectors from different feature maps. To quantitatively represent the size of the fire, we classified the size of the fire into three types according to the number of grids occupied by the fire in an image (see Fig. 6). It was confirmed that the feature maps of different scales complementarily contributed to detecting fires of various sizes through the weighted voting algorithm (see Fig. 7).

In addition, we investigated the sensitivity of the proposed method to fire itself. Referring to the experiments in [22], the prediction of the proposed method was investigated after manipulating the fire region in the image. Three fire-image manipulations were performed: (1) copying and pasting another part of the image onto the fire area, (2) completely covering the red and white box, and 3) covering some part of the fire. (see Fig. 8). The results show that the proposed method detects fire based solely on the inherent characteristics of the flames, which makes the prediction robust. In particular, predicting fire from images in fourth column



Figure 6. Images of three fire sizes. left: large-size, middle: middle-size, right: small size.

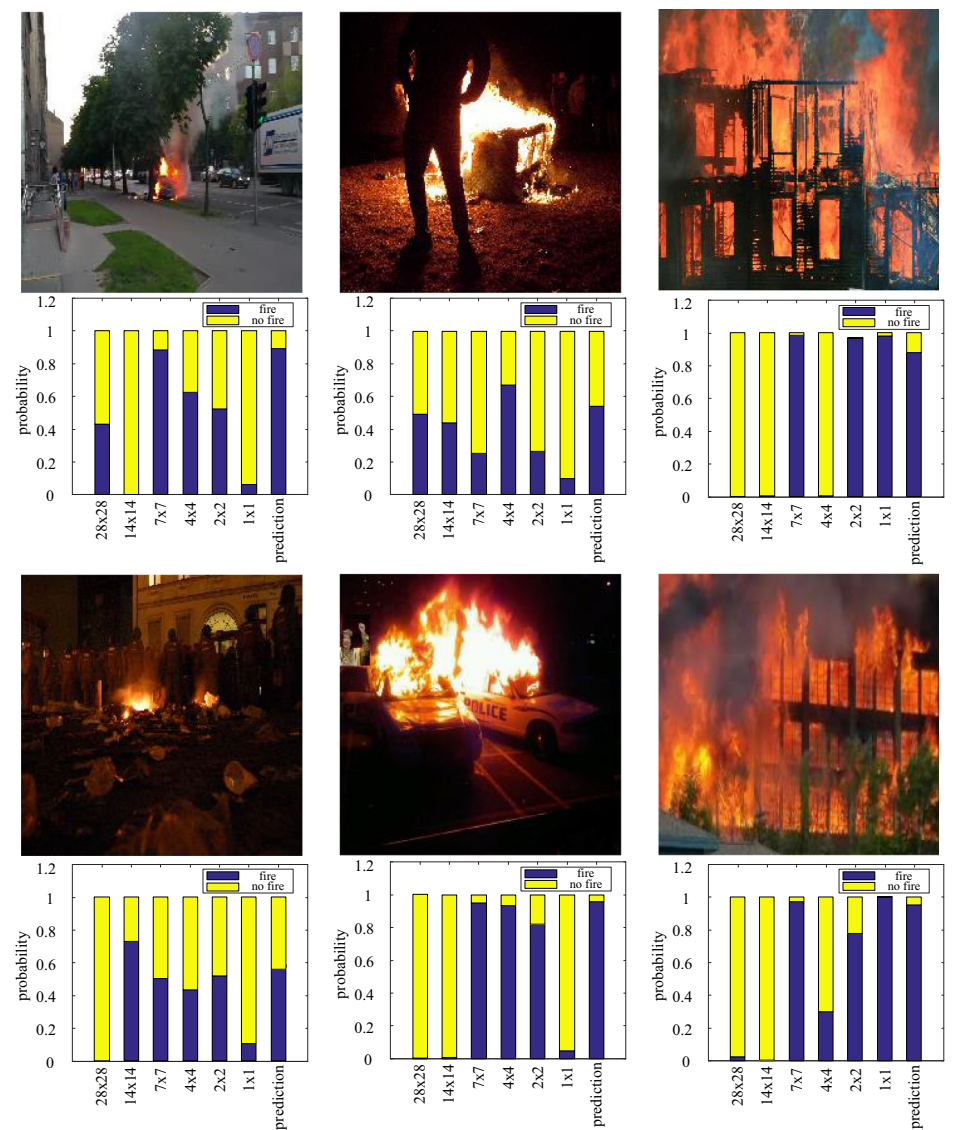


Figure 7. The contribution of each feature map for final prediction. The values displayed in the graph of each image denote the result values of FS block for multi-scale feature maps. fire denotes the confidence score of fire and no fire denotes that of no fire. left: small size, middle: middle size, right: large size.

of Figure 8 means that fire can be accurately recognized, even when it is partially hidden from the camera screen.

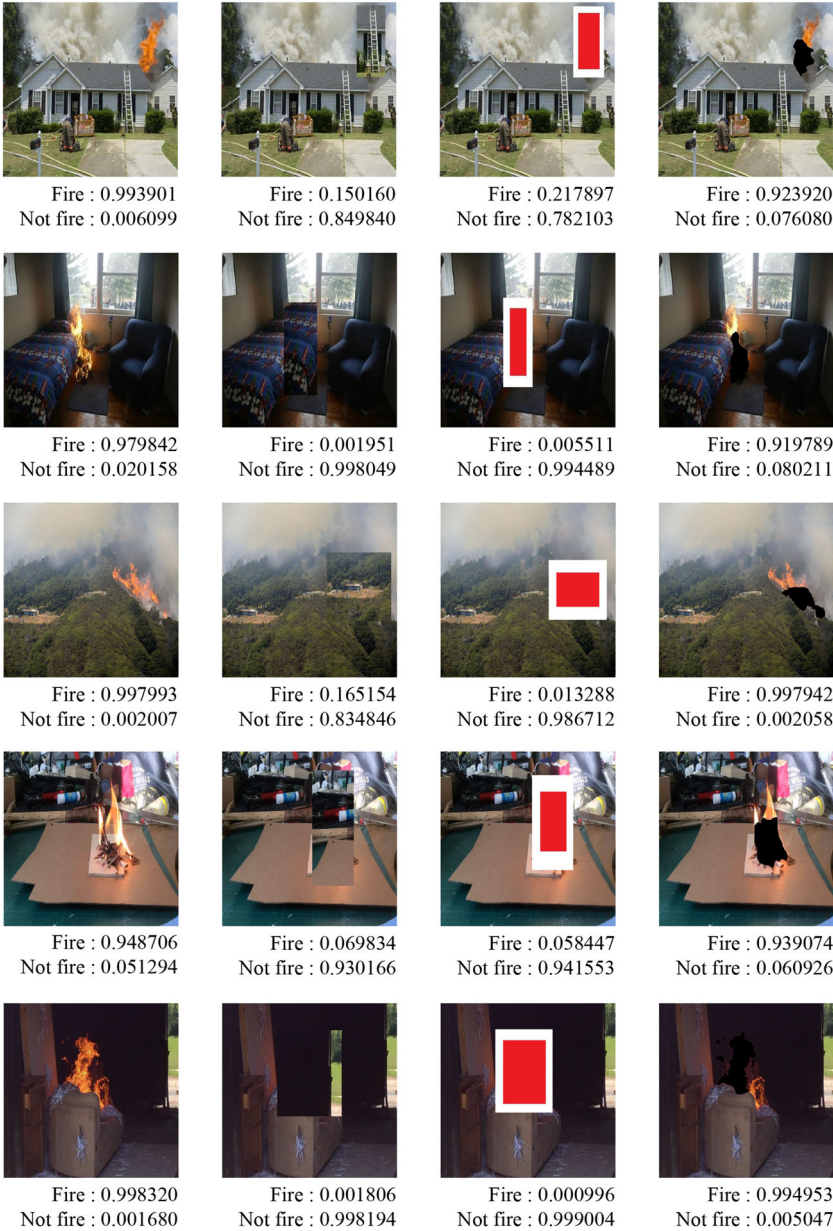


Figure 8. Blinded image test result.

5. Conclusion

In this work, we proposed a framework that strengthens the robustness of the existing CNN-based fire image classification model, focusing on the varying sizes of fires in images. To improve robustness, we exploited multi-scale feature maps obtained sequentially from deeply stacked CNN. The proposed FS block was used to squeeze feature information from multi-scale feature maps. Experiments on four fire image data sets for several evaluation metrics demonstrated that the proposed method performs well for images with various fire features. F1-score and FPR on three validation data sets are 97.89% and 0.0227, surpassing by 2.97% and decreasing by 2.4969 respectively compared to the second-best model. In particular, the proposed method outperformed several state-of-the-art models on an unseen data set, which showed multi-scale prediction has good generalization ability. Further experiments confirmed the correlation between the confidence score of the multi-scale feature maps and the size of the fire. This demonstrated that multi-scale prediction contributes to the robustness of the proposed method by referring to the multi-scale features complementarily.

This work contributes to the field by demonstrating that the multi-scale prediction can effectively aid the robustness and generalization of fire detection. We hope that this framework will add value to the future research.

Acknowledgements

Myungjoo Kang was supported by the National Research Foundation(NRF) grant of Korea (2015R1A5A1009350, 2021R1A2C3010887) and the ICT R&D program of MSIT/IITP(No. 1711117093).

Funding

Myungjoo Kang was supported by the NRF grant [2015R1A5A1009350] and the ICT RD program of MSIT/IITP[1711117093].

Availability of data and material

We implement our experiments on open data sets as follows. 1. RESCURE data set available in Project FP7-ICT-2013-EU-Brazil. RESCURE- Reliable and Smart Crowdsourcing Solution for Emergency and Crisis Management". In: 2013. 2. BoW Fire data set available in <https://bitbucket.org/gbdi/bowfire-dataset> 3. Sharma data set available in <https://github.com/cair/Fire-Detection-Image-Dataset> 4. Still data set available in http://www.fit.vutbr.cz/research/view_public.php?id=12124 Code availability Unfortunately, the code is not available because the authors of this paper use it for Project and Research and Development.

Code availability

Unfortunately, the code is not available because the authors of this paper use it for Project and Research and Development.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Please consult additional attached file.

References

1. Chen TH, Wu PH, Chiou YC (2004) An early fire-detection method based on image processing. In: 2004 International conference on image processing, 2004 (ICIP'04), vol. 3, pp. 1707–1710. IEEE
2. Marbach G, Loeffe M, Brupbacher T (2006) An image processing technique for fire detection in video images. *Fire Safety J* 41(4):285–289
3. Celik T, Demirel H (2009) Fire detection in video sequences using a generic color model. *Fire Safety J* 44(2):147–158
4. Celik T (2010) Fast and efficient method for fire detection using image processing. *ETRI J* 32(6):881–890
5. Celik T, Demirel H, Ozkaramanli H, Uyguroglu M (2007) Fire detection using statistical color model in video sequences. *J Vis Commun Image Represent* 18(2):176–185
6. Rinsurongkawong S, Ekpanyapong M, Dailey MN (2012) Fire detection for early fire alarm based on optical flow video processing. In: 2012 9th International conference on electrical engineering/electronics, computer, telecommunications and information technology, pp. 1–4. IEEE
7. Ko BC, Cheong KH, Nam JY (2009) Fire detection based on vision sensor and support vector machines. *Fire Safety J* 44(3):322–329
8. Zhou Q, Yang X, Bu L (2015) Analysis of shape features of flame and interference image in video fire detection. In 2015 Chinese Automation Congress (CAC), pp. 633–637. IEEE
9. Kong SG, Jin D, Li S, Kim H (2016) Fast fire flame detection in surveillance video using logistic regression and temporal smoothing. *Fire Safety J* 79:37–43
10. Mao W, Wang W, Dou Z, Li Y (2018) Fire recognition based on multi-channel convolutional neural network. *Fire Technol* 54(2):531–554
11. Jadon A, Omama M, Varshney A, Ansari MS, Sharma R (2019) Firenet: A specialized lightweight fire & smoke detection model for real-time IoT applications. *arXiv preprint arXiv:1905.11922*
12. Saeed F, Paul A, gaikumar P, Nayyar A (2019) Convolutional neural network based early fire detection. *Multimed Tools Appl* 79:9083–9099
13. Frizzi S, Kaabi R, Bouchouicha M, Ginoux JM, Moreau E, Fnaiech F (2016) Convolutional neural network for video fire and smoke detection. In: IECON 2016-42nd annual conference of the IEEE industrial electronics society, pp. 877–882. IEEE
14. Dunning AJ, Breckon TP (2018) Experimentally defined convolutional neural network architecture variants for non-temporal real-time fire detection. In: 2018 25th IEEE International conference on image processing (ICIP), pp. 1558–1562. IEEE

15. Zhang Q, Xu J, Xu L, Guo H (2016) Deep convolutional neural networks for forest fire detection. In: 2016 International forum on management, education and information technology application. Atlantis Press
16. Sharma J, Granmo OC, Goodwin M, Fidje JT (2017) Deep convolutional neural networks for fire detection in images. *Int Conf Eng Appl Neural Netw Springer, Cham*, pp 183–193
17. Maksymiv O, Rak T, Peleshko D (2017) Real-time fire detection method combining AdaBoost, LBP and convolutional neural network in video sequence. In: 2017 14th international conference the experience of designing and application of CAD Systems in microelectronics (CADSM), pp. 351–353, IEEE
18. Yang H, Jang H, Kim T, Lee B (2019) Non-temporal lightweight fire detection network for intelligent surveillance systems. *IEEE Access* 7:169257–169266
19. Li S, Yan Q, Liu P (2020) An efficient fire detection method based on multiscale feature extraction, implicit deep supervision and channel attention mechanism. *IEEE Trans Image Process* 29:8467–8475
20. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In *Proceedings of the IEEE Conference on computer vision and pattern recognition*, pp. 1–9
21. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. *arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)*
22. Muhammad K, Ahmad J, Mehmood I, Rho S, Baik SW (2018) Convolutional neural networks based fire detection in surveillance videos. *IEEE Access* 6:18174–18183
23. Muhammad K, Ahmad J, Baik SW (2018) Early fire detection using convolutional neural networks during surveillance for effective disaster management. *Neurocomputing* 288:30–42
24. Muhammad K, Khan S, Elhoseny M, Ahmed SH, Baik SW (2019) Efficient fire detection for uncertain surveillance environment. *IEEE Trans Ind Informat* 15(5):3113–3122
25. Li P, Zhao W (2020) Image fire detection algorithms based on convolutional neural networks. *Case Studies Thermal Eng* 19:100625
26. Wu S, Zhang L (2018) Using popular object detection methods for real time forest fire detection. In 2018 11th International symposium on computational intelligence and design (ISCID), vol. 1, pp. 280–284, IEEE
27. Ren S, He K, Girshick R, Sun J (2015) Faster r-cnn: Towards real-time object detection with region proposal networks. *arXiv preprint [arXiv:1506.01497](https://arxiv.org/abs/1506.01497)*
28. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H (2017) Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint [arXiv:1704.04861](https://arxiv.org/abs/1704.04861)*
29. Töreyn BU, Dedeoğlu Y, Güdükbay U, Çetin AE (2006) Computer vision based method for real-time fire and flame detection. *Pattern Recogn Lett* 27(1):49–58
30. Zhang Z, Zhao J, Zhang D, Qu C, Ke Y, Cai B (2008) Contour based forest fire detection using FFT and wavelet. In 2008 International conference on computer science and software engineering, vol. 1, pp. 760–763, IEEE
31. Günay O, Taşdemir K, Töreyn BU, Çetin AE (2010) Fire detection in video using LMS based active learning. *Fire Technol* 46(3):551–577
32. Borges PVK, Izquierdo E (2010) A probabilistic approach for vision-based fire detection in videos. *IEEE Trans Circuits Syst Video Technol* 20(5):721–731
33. Qiu T, Yan Y, Lu G (2011) An autoadaptive edge-detection algorithm for flame and fire image processing. *IEEE Trans Instrum Measure* 61(5):1486–1493
34. Wang DC, Cui X, Park E, Jin C, Kim H (2013) Adaptive flame detection using randomness testing and robust features. *Fire Safety J* 55:116–125

35. Foggia P, Saggese A, Vento M (2015) Real-time fire detection for video-surveillance applications using a combination of experts based on color, shape, and motion. *IEEE Trans Circuits Syst Video Technol* 25(9):1545–1556
36. He K, Zhang X, Ren S, & Sun J (2016) Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778
37. Krizhevsky A, Sutskever I, & Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pp. 1097–1105
38. Zhang G, Wang M, Liu K (2019) Forest fire susceptibility modeling using a convolutional neural network for Yunnan province of China. *Int J Disaster Risk Sci* 10(3):386–403
39. Park B, Yu S, Jeong J (2019) Densely connected hierarchical network for image denoising. In *Proceedings of the IEEE Conference on computer vision and pattern recognition workshops*, pp. 0–0
40. He K, Zhang X, Ren S, Sun J (2016) Identity mappings in deep residual networks. In *European conference on computer vision*. Springer, Cham, pp. 630–645
41. Project FP7-ICT-2013-EU-Brazil (2013) RESCURE-Reliable and Smart Crowdsourcing Solution for Emergency and Crisis Management
42. Daniel YT Chino et al. (2015) Bowfire: detection of fire in still images by integrating pixel color and texture analysis. In: 2015 28th SIBGRAPI conference on graphics, patterns and images, pp. 95–102, IEEE
43. Mlich, J, Koplík, K, Hradiš M, Zemčík P (2020) Fire segmentation in still images. In: *International conference on advanced concepts for intelligent vision systems*. Springer, Cham, pp. 27–37