# A Survey of Vision-based Fire Detection using Convolutional Neural Networks

Gong-suo Chen
*International College of Digital Innovation*
*Chiang Mai University*
Chaing Mai, Thailand
*School of Information and Engineering*
*Sichuan Tourism University*
Chengdu, China
gongsuo_chen@cmu.ac.th

Tirapot Chandarasupsang*
*International College of Digital Innovation*
*Chiang Mai University*
Chaing Mai, Thailand
*Corresponding author
tirapot@gmail.com

Xiao-dong Luo
*School of Information and Engineering*
*Sichuan Tourism University*
Chengdu, China
luoxd@sctu.edu.cn

Annop Tananchana
*International College of Digital Innovation*
*Chiang Mai University*
Chaing Mai, Thailand
Annop.t@cmuic.net

Lei Mu
*School of Information and Engineering*
*Chengdu University*
Chengdu, China
mulei@cdu.edu.cn

2024 4th International Conference on Big Data Engineering and Education (BDEE) | 979-8-3503-8826-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/BDEE63226.2024.00015

***Abstract*—Fire can harm humans and animals and destroy the natural environment. Early and accurate fire detection is crucial. However, detecting fire is very challenging due to its varying color, size, shape, texture, and the uncertainty of different environments. In this paper, we review literature on datasets, attention mechanisms, and networks, which are the most important components of a fire detection model. We first present key notations to describe the network structure and its key components. Then, we introduce several attention mechanisms to reduce the false alarm rate. After that, we present numerous state-of-the-art neural networks for fire, flame, and smoke detection. Finally, we discuss several challenges and opportunities at the end of this paper.**

***Index Terms*—fire detection, fire classification, fire recognition, fire detection datasets, convolutional neural networks, attention mechanism**

## I. Introduction

Fire is one of the most terrible and threatening disasters worldwide due to its destructive nature. Without any control, fire can quickly spread and consume everything it encounters. Fire can cause severe damage to the natural environment, animals, human lives, property, and more. Over the past few years, several countries have suffered from fires. In the United States, approximately 189,500 highway vehicle fires occurred, causing 550 deaths in 2019 [1]. From January to March 2020, 33 people and 1.5 billion animals were killed by forest and bush fires in Australia, with 3,000 homes and 19 million hectares burned [2]. Similar incidents have occurred in California [3]. From 2009 to 2015, there were about 20,000 vehicle fires annually in China, causing nearly 370,000 yuan in losses [4]. From 1993 to 2016, 4.5 million building fires were reported in 57 countries, resulting in 62,000 deaths [5]. To avoid large-scale disasters caused by fires, accurate and timely detection of various types of fire is essential.

Recently, several researchers have explored sensor-based methods for early fire detection, such as those detecting smoke, particles, and temperature. These sensors are extremely easy to install and deploy, and they are also very affordable. However, they need to come into direct contact with smoke or fire, leading to long response times and missing the optimal time to suppress the fire. To overcome these limitations, vision-based algorithms have attracted significant attention from researchers.

In the past few years, several meaningful surveys have reviewed vision-based fire detection technologies to help researchers gain a deeper understanding of fire [6, 7, 8, 9]. However, these surveys mostly focus on traditional image processing and some convolutional neural networks, but they overlook attention mechanisms, which are widely used to improve the ability to distinguish fire from fire-like objects and reduce the false alarm rate.

In this paper, we summarize five of the most representative datasets, four attention mechanisms, and seven state-of-the-art fire detection convolutional neural networks, and we use unified formal methods to describe them.

The rest of this paper is organized as follows: Section II introduces key notations that describe networks and key components. Section III presents some commonly used datasets. Section IV introduces several attention mechanisms used to enhance fire extraction features. Section V presents representative networks. Some challenges and suggestions are discussed in Section VI. Finally, Section VII provides the conclusion.

979-8-3503-8826-8/24/$31.00 ©2024 IEEE
DOI 10.1109/BDEE63226.2024.00015

## II. NOTATIONS

In this section, we introduce some key notions that used to describe the network and its key components.

TABLE I: Details of notations.

| Name | Remark |
|---|---|
| $\mathbf{X}$ | The input of the module. |
| $\mathbf{Y}$ | The output of the module. |
| $\sigma$ | Sigmoid activation function. |
| $\delta$ | ReLU activation function. |
| $FC_n$ | Fully connected layers with $n$ neurons. |
| $f^{k\times k}$ | $k \times k$ convolution operation |
| $g_m^n$ | Module $m$ repeat $n$ times, i.e. $g_m(\dots g_m(\mathbf{X}))$. |
| $\varphi$ | attention module |
| $\phi$ | global average pooling |
| $\psi$ | global max pooling |
| $\varphi_c$ | Channel attention mapping. |
| $\varphi_s$ | Spatial attention mapping. |

## III. FIRE CLASSIFICATION DATASETS

In this section, we introduce some commonly used fire detection datasets and discuss the pros and cons of these datasets.

### A. BowFire

The BoWFire dataset is one of the most commonly used datasets to validate the effectiveness and generalization of fire detection models, as shown in studies such as [10], [11], [12], and [13]. The main contribution of this dataset is its extreme diversity and challenging nature, consisting of many fire-like objects, such as red cars, sunsets, and glare lights. The dataset is comparatively small, containing 226 images in total: 119 fire images and 107 normal images.

### B. Foggia's dataset

Foggia's dataset [14] consists of 31 videos, of which 14 are captured in fire environments and 17 are normal. This dataset is very challenging for traditional color-based methods due to scenes containing red objects and for motion-based methods due to mountains containing smoke, clouds, or fog. However, it is not as difficult for deep learning-based approaches. Additionally, small fires or fire images captured from long distances are also comparatively challenging for existing fire detection methods.

### C. FD

FD, proposed by [15], is currently the largest fire dataset, consisting of 50,000 images, with 25,000 fire images and 25,000 normal images. It is based on two benchmark datasets, [14] and [16], and the remaining images are collected from the internet, containing scenes such as burning clouds, sunsets, glare lights, and red objects. The main purpose of FD is to effectively promote vision-based fire detection research, similar to the role of ImageNet [17]. Compared to BoWFire and Foggia's dataset, FD includes a larger number of fire categories, such as boat fires, car fires, building fires, and forest fires, enhancing the model's ability to handle real-time complex fire environments.

### D. Yar's dataset

Although Foggia's dataset is extremely large and includes images captured from both indoor and outdoor environments, most of the images are very similar to each other due to an unreasonable frame extraction strategy. To tackle this problem, [18] developed a new small and diverse fire dataset consisting of 2,000 images, with 1,000 fire images and 1,000 non-fire images. To further improve diversity, the dataset includes images from different countries, such as Italy and India, and from various sources, such as YouTube, Facebook, and DW News. Additionally, different videos contribute different images based on image similarity, ensuring a broader range of scenarios.

### E. DFAN fire dataset

Existing fire datasets, such as those proposed by [14], [18], and [15], often lack diversity and only include fire and normal classification statuses, which limits the ability of models trained on these datasets to recognize fires in real-time complex environments. To address this issue, [19] initially collected videos from YouTube, Facebook, and disaster management agency records. They extracted frames using a 40-60-frame skipping strategy to create a small, imbalanced, and extremely diverse fire classification dataset, namely DFAN, which features 12 different categories of fire, as depicted in Table II.

TABLE II: Description of DFAN dataset which is small, imbalanced but diverse.

| Category of fire | Number |
|---|---|
| Boat fire | 338 |
| Building fire | 305 |
| Bus fire | 400 |
| Car fire | 579 |
| Cargo fire | 207 |
| Electric pole fire | 300 |
| Forest fire | 480 |
| Normal | 97 |
| Pick-up fire | 257 |
| SUV fire | 240 |
| Train fire | 300 |
| Van fire | 300 |
| Total | 3804 |

## IV. FIRE DETECTION ATTENTION MECHANISMS

In this section, we present and formulate uniformly some commonly used attention mechanisms, as depicted in Table III, that applied in the fire detection convolutional neural networks.

### A. SENet Channel Attention Module

Let $\mathbf{X} \in \mathbb{R}^{C\times H\times W}$ be the input feature maps, $\mathbf{Y} \in \mathbb{R}^{C\times H\times W}$ be the output features that through the channel attention module, $x_c(i,j)$ be the $c$-th input features which location at position $(i,j)$, where $i \in [1,H], j \in [1,W]$, $\mathbf{X} = [\mathbf{x_1}, \dots, \mathbf{x_C}], \mathbf{Y} = [\mathbf{y_1}, \dots, \mathbf{y_C}]$. Global average pooling

TABLE III: Development of attention mechanism.

| Proposed Year | Attention |
|---|---|
| 2014 | NLP [20] |
| 2018 | SENet [21] |
| 2018 | CBAM [22] |
| 2019 | CCNet [23] |
| 2020 | ECANet [24] |
| 2020 | CTAM [25] |
| 2020 | MSCAM [26] |
| 2021 | FcaNet [27] |

was first applied to input features to acquired squeezed feature map vectors $\mathbf{z} = [z_1, \ldots, z_C]$, and $z_c$ can be calculated by:

$$z_c = \phi(u_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x_c(i, j) \tag{1}$$

where $\phi$ represents the global average pooling operation, $c \in [1, C]$.

Then, the vector $\mathbf{z}$ is feed into two stacked fully connected layers,

$$\mathbf{s} = \sigma(\mathbf{W_2} \cdot \delta(\mathbf{W_1} \cdot \mathbf{z})) \tag{2}$$

where $\sigma$ is sigmoid function, $\delta$ is ReLU activation function, $\mathbf{W_1}$ and $\mathbf{W_2}$ are weighted parameters of the two fully connected layers.

Finally, the learned channel weights $\mathbf{s}$ of the input feature maps can be multiplied by the input features to get final weighted maps.

$$\mathbf{Y} = \mathbf{s} \otimes \mathbf{X} \tag{3}$$

where $\otimes$ denotes element-wise multiplication.

[28] used a channel attention module in the last 8 convolutional layers of VGG-19 to detect fire level ratings, achieving approximately a 7% relative improvement over the baseline. [15] applied dense layers to gradually increase the number of feature map channels and used a transition layer to reduce it, avoiding channel explosion. They used a channel attention module before the transition layer to refine all the channels, resulting in approximately a 2% improvement over the baseline.

Due to the significant impact of small outputs from activation layers on feature fusion results, [29] introduced L2-normalization before the global average pooling of SENet to regularize feature maps. They applied this improved SENet module in the two dense blocks of an improved VGG-16, achieving slightly better results than the baseline.

### B. Channel and Spatial Fusion Attention

[30] applied SENet channel attention and two convolutaional layers of spatial attention in the end of the feature extraction module to detect fire.

### C. Convolutional Block Attention Module

The convolutional block attention module (CBAM) is an efficient and effective attention module for feed forward neural networks, channel and spatial attention are simultaneously integrated in one module. Let $\mathbf{X} \in \mathbb{R}^{C \times H \times W}$ be the input feature maps, $\mathbf{Y} \in \mathbb{R}^{C \times H \times W}$ be the output, $\mathbf{U} \in \mathbb{R}^{C \times H \times W}$ be the output of channel attention module, $\varphi_c : \mathbf{X} \longrightarrow \mathbf{U}$ represents channel attention transformation, $\varphi_s : \mathbf{U} \longrightarrow \mathbf{Y}$ denotes spatial attention mapping. The CBAM can be calculated by:

$$\begin{aligned} \mathbf{U} &= \varphi_c(\mathbf{X}) \otimes \mathbf{X} \\ \mathbf{Y} &= \varphi_s(\mathbf{U}) \otimes \mathbf{U} \end{aligned} \tag{4}$$

Channel attention $\varphi_c$ can be obtained by:

$$\begin{aligned} \varphi_c(\mathbf{X}) &= \sigma(MLP(\phi(\mathbf{X})) + MLP(\psi(\mathbf{X}))) \\ &= \sigma(\mathbf{W_2} \cdot \delta(\mathbf{W_1} \cdot \mathbf{X}_{avg}) + \mathbf{W_2} \cdot \delta(\mathbf{W_1} \cdot \mathbf{X}_{max})) \end{aligned} \tag{5}$$

where $\mathbf{W_1} \in \mathbb{R}^{C/r \times C}, \mathbf{W_2} \in \mathbb{R}^{C \times C/r}$, $\phi$ denotes global average pooling, $\psi$ represents global max pooling.

Spatial attention $\varphi_s$ is computed as:

$$\varphi(\mathbf{U}) = \sigma(f^{7\times7}(\odot([\phi(\mathbf{U}); \psi(\mathbf{U}])))) \tag{6}$$

where $\odot$ represents concatenation operation and $f^{7\times7}$ denotes a convolutional with $7 \times 7$ kernel size.

[31] applied CBAM in the first and forth convolution block to improve the discrimination feature representation and acquired the usefull global information.

### D. Dual Fire Attention Module

Let $\mathbf{X} \in \mathbb{R}^{C \times H \times W}$ be the input features, $\mathbf{Y} \in \mathbb{R}^{C}$ be the output.

$$\mathbf{Y} = \odot([\phi(\mathbf{X}); \varphi_c(\mathbf{X}); \varphi_s(\mathbf{X})]) \tag{7}$$

where channel attention can be computed by:

$$\varphi_c(\mathbf{X}) = \mathbf{W_2} \cdot \delta(\mathbf{W_1} \cdot \phi(\mathbf{X})) \tag{8}$$

and spatial attention can be obtained by:

$$\varphi_s(\mathbf{X}) = \phi(f^{1\times1}(\delta(f^{3\times3}(\delta(f^{1\times1}(\mathbf{X})))))) \tag{9}$$

## V. Fire Detection Convolutional Neurual Networks

In this section, we will first introduce some representative convolutional neural networks for fire detection. We will then discuss the pros and cons of these different networks. Finally, we will provide further suggestions regarding network design.

### A. ResNetFire

Compared with traditional image processing technologies, CNN-based fire detection methods achieve sufficient accuracy. However, these networks are often too shallow, such as LeNet-5, and the datasets used to train them are very balanced. This is inconsistent with the real world, as fires are rare events. To address this problem, ResNetFire, which is based on ResNet[32], was proposed by [33]. It can be computed as follows:

$$\mathbf{Y} = FC_2(FC_{4096}(g_{resnet50}(\mathbf{X}))) \tag{10}$$

where $g_{resnet50}$ represents ResNet50 network without classification layers.

Moreover, ResNetFire indicates that adding one more fully connected layer after ResNet50 could increase accuracy, although it brings more computational overhead.

The drawback of ResNetFire is that the dataset is very small, with only 651 images, and its performance is only compared with VGG16, leaving many other networks unexplored.

### B. ANetFire

ResNetFire achieved better accuracy than shallow networks; however, it was only tested on a small, imbalanced dataset. Foggia's dataset, consisting of 31 videos captured in both indoor and outdoor environments, is one of the most popular datasets used in fire detection. ANetFire, proposed by [11], can be computed by:

$$\mathbf{Y} = FC_2(g_{alexnet}(\mathbf{X})) \tag{11}$$

where $g_{alexnet}$ indicates AlexNet network without classification layers.

However, the performance of ANetFire was only compared with traditional machine learning approaches using color, motion, or shape features.

### C. GNetFire

Although convolutional neural network-based fire detection methods achieve better accuracy than traditional image processing methods, CNN-based models are difficult to deploy on resource-constrained devices such as surveillance networks. To tackle this issue, GNetFire was proposed by [12] and it can be computed by:

$$\mathbf{Y} = FC_2(g_{googlenet}(\mathbf{X})) \tag{12}$$

where $g_{googlenet}$ represents GoogLeNet without classification layer.

### D. CNNFire

CNNFire [13] explored how to make a good tradeoff between accuracy and efficiency, and it could be computed by:

$$\mathbf{Y} = FC_2(g_{squeezenet}(\mathbf{X})) \tag{13}$$

where $g_{squeezenet}$ denotes SqueezeNet without classification layer.

### E. EMNFire

Existing methods may fail in uncertain IoT environments with fog, snow, and smoke. To address this drawback, EMNFire was proposed by [34] and it can be computed by:

$$\mathbf{Y} = FC_2(g_{mobilenetv2}(\mathbf{X})) \tag{14}$$

where $g_{mobilenetv2}$ is MobileNetV2 without classification layer.

### F. EFDNet

In order to make a good tradeoff between model accuracy, size and inference speed, [15] proposed EFDNet and it can be formulated as follows:

$$\mathbf{Y} = g_{class}(g^4_{ids\_ca}(g^3_{mfe}(g_{head}(\mathbf{X})))) \tag{15}$$

where $g_{head}$ represents the head layer, $g_{mfe}$ the multifeature extraction module, $g_{ids\_ca}$ the implicit deep supervision and channel attention module, $g_{class}$ the classification layer.

Then the head layer can be computed by:

$$g_{head}(\mathbf{X}) = f^{3\times3}(f^{3\times3}(\mathbf{X})) \tag{16}$$

In order to simplify the equation symbol, batch normalization and relu activation are default followed by every $f^{3\times3}$ convolution.

The multifeature extraction module is actually an inception module and it can be acquired by:

$$\begin{aligned} g_{mfe}(\mathbf{X}) = \odot[&f^{1\times1}(\mathbf{X}); f^{3\times3}(f^{1\times1}(\mathbf{X})); \\ &f^{5\times5}(f^{1\times1}(\mathbf{X})); \\ &f^{1\times1}(f^{pooling}(\mathbf{X}))] \end{aligned} \tag{17}$$

The $g_{ids\_ca}$ module can be computed by:

$$g_{ids\_ca}(\mathbf{X}) = g_{tran}(g_{ca}(g_{ids}(\mathbf{X}))) \tag{18}$$

where $g_{ca}$ can be computed by Equation 3.

$$g_{tran}(\mathbf{X}) = \phi(f^{1\times1}(\mathbf{X})) \tag{19}$$

where $f^{1\times1}$ and $\phi$ are used to reduce the dimension and the spatial of the feature map.

Let $\mathbf{X_0} = \mathbf{X}$ be the input of the $g_{ids}$ module, which has $l$ basic layers, $\mathbf{Y} = \mathbf{X_l}$ be the output and every basic layer could be computed by Equation 21. Then $\mathbf{Y}$ can be computed by:

$$\begin{aligned} \mathbf{X_1} &= \odot[\mathbf{X_0}; g'_1(\mathbf{X_0})] \\ \mathbf{X_2} &= \odot[\mathbf{X_1}; g'_2(\mathbf{X_1})] \\ &= \odot[\mathbf{X_0}; g'_1(\mathbf{X_0}); g'_2(\mathbf{X_1})] \\ &\dots \\ \mathbf{X_l} &= \odot[\mathbf{X_0}; g'_1(\mathbf{X_0}); \dots, g'_l(\mathbf{X_{l-1}})] \end{aligned} \tag{20}$$

where $g'$ can be computed by:

$$g'(\mathbf{X}) = f^{3\times3}(f^{1\times1}(\mathbf{X})) \tag{21}$$

### G. DFAN

Existing networks are typically used for fire and normal classification, which limits their ability to recognize different fire categories. To overcome this issue, DFAN was proposed by [19], which uses InceptionV3 [35] as the backbone network to extract features and develops a new dual attention module that fuses channel and modified spatial attention. It can be formulated as follows:

$$\mathbf{Y} = \varphi_{dual}(g_{inception\_v3}(\mathbf{X})) \tag{22}$$

where $\varphi_{dual}$ is a dual attention module and can be computed by Equation 7 and $g_{inception\_v3}$ is for InceptionV3 network without classification layers.

## VI. Challenges and Suggestions

Although significant progress has been made in fire detection using convolutional neural networks, there are still many research areas to explore in the future. Specifically, partial convolution [36], octave convolution [37], and other newly developed convolutional operators proposed by CVPR, ICCV, and AAAI in recent years have not yet been evaluated in the domain of fire detection. Exploring the application of these new convolution operators in fire detection could be a promising research direction.

Making a good tradeoff between accuracy, model size, and inference speed is still an important research problem. Although modern model compression methods have been developed, few have been applied to fire detection models.

Additionally, convolutional neural networks designed by humans can be very cumbersome. Recently, many heuristic approaches for automatically discovering networks for specific tasks have developed rapidly. However, only a few methods have been explored for fire detection. It remains a significant challenge to find efficient and effective networks for fire detection, particularly those that include crucial components to improve fire discrimination ability.

Finally, fire detection benchmark datasets are still insufficient, especially for specific domain applications such as tourist spot fires, kitchen fires, and other representative types of fire. Moreover, not all fires are harmful to human beings; some fires, like gas fires used for cooking, are necessary for daily life. Therefore, developing models that can distinguish between beneficial and harmful fires is also important for a deeper understanding and practical application.

## VII. Conclusion

In this paper, we reviewed important literature on fire detection using vision-based convolutional neural networks. We summarized key aspects including fire detection datasets, attention mechanisms, and various convolutional neural network architectures. Finally, we concluded by highlighting challenges and offering reasonable suggestions for future research directions.

## Acknowledgment

This work was supported in part by the Sichuan Science and Technology Program of China under grant 2023YFG0130, 2023YFG0099, 2023ZYD0148, 2023YFS0467, 2023YFS0431 and in part by the ABa Achievements Transformation Program under grant S24CGZH0004, R22CGZH0007, and in part by the Scientific Research Project of Sichuan Tourism University under grant 2023SCTUZK98.

## References

[1] Nikoleta Csápaiová. The impact of vehicle fires on road safety. *Transportation research procedia*, 55:1704–1711, 2021.

[2] Alexander I Filkov, Tuan Ngo, Stuart Matthews, Simeon Telfer, and Trent D Penman. Impact of australia's catastrophic 2019/20 bushfire season on communities and environment. retrospective analysis and current trends. *Journal of Safety Science and Resilience*, 1(1):44–56, 2020.

[3] Saima Majid, Fayadh Alenezi, Sarfaraz Masood, Musheer Ahmad, Emine Selda Gündüz, and Kemal Polat. Attention based cnn model for fire detection and localization in real-world images. *Expert Systems with Applications*, 189:116114, 2022.

[4] DL Zhang, LY Xiao, Y Wang, and GZ Huang. Study on vehicle fire safety: Statistic, investigation methods and experimental analysis. *Safety science*, 117:194–204, 2019.

[5] RMDIM Rathnayake, P Sridarran, and MDTE Abeynayake. Fire risk of apparel manufacturing buildings in sri lanka. *Journal of Facilities Management*, 20(1):59–78, 2022.

[6] Princy Matlani and Manish Shrivastava. A survey on video smoke detection. In *Information and Communication Technology for Sustainable Development: Proceedings of ICT4SD 2016, Volume 1*, pages 211–222. Springer, 2018.

[7] Fengju Bu and Mohammad Samadi Gharajeh. Intelligent and vision-based fire detection systems: A survey. *Image and vision computing*, 91:103803, 2019.

[8] Anshul Gaur, Abhishek Singh, Anuj Kumar, Ashok Kumar, and Kamal Kapoor. Video flame and smoke based fire detection algorithms: A literature review. *Fire technology*, 56:1943–1980, 2020.

[9] S Geetha, CS Abhishek, and CS Akshayanat. Machine vision based fire detection techniques: A survey. *Fire technology*, 57:591–623, 2021.

[10] K. Muhammad, J. Ahmad, and S. W. Baik. Early fire detection using convolutional neural networks during surveillance for effective disaster management. *Neurocomputing*, 288:30–42, 2018.

[11] K. Muhammad, J. Ahmad, and S. W. Baik. Early fire detection using convolutional neural networks during surveillance for effective disaster management. *Neurocomputing*, 288:30–42, 2018.

[12] K. Muhammad, J. Ahmad, I. Mehmood, S. Rho, and S. W. Baik. Convolutional neural networks based fire detection in surveillance videos. *IEEE Access*, 6:18174–18183, 2018.

[13] K. Muhammad, J. Ahmad, Z. Lv, P. Bellavista, P. Yang, and S. W. Baik. Efficient deep cnn-based fire detection and localization in video surveillance applications. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 49(7):1419–1434, 2019.

[14] P. Foggia, A. Saggese, and M. Vento. Real-time fire detection for video-surveillance applications using a combination of experts based on color, shape, and motion. *IEEE Transactions on Circuits and Systems for Video Technology*, 25(9):1545–1556, 2015.

[15] S. Li, Q. Yan, and P. Liu. An efficient fire detection method based on multiscale feature extraction, implicit deep supervision and channel attention mechanism. *IEEE Transactions on Image Processing*, 29:8467–8475, 2020.

[16] D. Y. T. Chino, L. P. S. Avalhais, J. F. Rodrigues, and A. J. M. Traina. Bowfire: Detection of fire in still images by integrating pixel color and texture analysis. In *Brazilian Symposium of Computer Graphic and Image Processing*, volume 2015-October, pages 95–102. IEEE Computer Society, 2015.

[17] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. IEEE, 2009.

[18] H. Yar, T. Hussain, Z. A. Khan, D. Koundal, M. Y. Lee, and S. W. Baik. Vision sensor-based real-time fire detection in resource-constrained iot environments. *Computational Intelligence and Neuroscience*, 2021, 2021.

[19] H. Yar, T. Hussain, M. Agarwal, Z. A. Khan, S. K. Gupta, and S. W. Baik. Optimized dual fire attention network and medium-scale fire classification benchmark. *IEEE Transactions on Image Processing*, 31:6331–6343, 2022.

[20] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. In *Advances in Neural Information Processing Systems*, pages 1724–1732. 2014.

[21] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu. Squeeze-and-excitation networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(8):2011–2023, 2020.

[22] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018.

[23] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu. Ccnet: Criss-cross attention for semantic segmentation. In *17th IEEE/CVF International Conference on Computer Vision, ICCV 2019*, volume 2019-October, pages 603–612. Institute of Electrical and Electronics Engineers Inc., 2019.

[24] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu. Eca-net: Efficient channel attention for deep convolutional neural networks. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020*, pages 11531–11539. IEEE Computer Society, 2020.

[25] D. Misra, T. Nalamada, A. U. Arasanipalai, and Q. B. Hou. Rotate to attend: Convolutional triplet attention module. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3139–3148, 2020.

[26] Y. Dai, F. Gieseke, S. Oehmcke, Y. Wu, and K. Barnard. Attentional feature fusion. *Attentional Feature Fusion*, pages 3560–3569, 2021.

[27] Z. Qin, P. Zhang, F. Wu, and X. Li. Fcanet: Frequency channel attention networks. In *18th IEEE/CVF International Conference on Computer Vision, ICCV 2021*, pages 763–772. Institute of Electrical and Electronics Engineers Inc., 2021.

[28] Y. Wu, Y. He, P. Shivakumara, Z. Li, H. Guo, and T. Lu. Channel-wise attention model-based fire and rating level detection in video. *CAAI Transactions on Intelligence Technology*, 4(2):117–121, 2019.

[29] S. Jin, T. Zhao, and X. An. Joint network smoke recognition based on channel attention mechanism. In *Journal of Physics: Conference Series*, volume 1748. IOP Publishing Ltd, 2021.

[30] Z. Deng, S. Hu, S. Yin, Y. Wang, A. Basu, and I. Cheng. Multi-step implicit adams predictor-corrector network for fire detection. *IET Image Processing*, 16(9):2338–2350, 2022.

[31] T. Li, H. Zhu, C. Hu, and J. Zhang. An attention-based prototypical network for forest fire smoke few-shot detection. *Journal of Forestry Research*, 33(5):1493–1504, 2022.

[32] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[33] J. Sharma, O. C. Granmo, M. Goodwin, and J. T. Fidje. Deep convolutional neural networks for fire detection in images. In L. Iliadis, A. Likas, C. Jayne, and G. Boracchi, editors, *Communications in Computer and Information Science*, volume 744, pages 183–193. Springer Verlag, 2017.

[34] K. Muhammad, S. Khan, M. Elhoseny, S. Hassan Ahmed, and S. Wook Baik. Efficient fire detection for uncertain surveillance environment. *IEEE Transactions on Industrial Informatics*, 15(5):3113–3122, 2019.

[35] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[36] Jierun Chen, Shiu-hong Kao, Hao He, Weipeng Zhuo, Song Wen, Chul-Ho Lee, and S-H Gary Chan. Run, don't walk: Chasing higher flops for faster neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12021–12031, 2023.

[37] Yunpeng Chen, Haoqi Fan, Bing Xu, Zhicheng Yan, Yannis Kalantidis, Marcus Rohrbach, Shuicheng Yan, and Jiashi Feng. Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3435–3444, 2019.