Towards Efficient Fire Detection in IoT Environment: A Modified Attention Network and Large-Scale Dataset

Naqqash Dilshad, *Student Member, IEEE*, Samee Ullah Khan, *Student Member, IEEE*, Norah Saleh Alghamdi, Tarik Taleb, *Senior Member, IEEE*, and JaeSeung Song, *Senior Member, IEEE*

***Abstract*—Advancements in deep learning and the Internet of Things (IoT) enable early fire detection through vision-based systems, reducing ecological, social, and economic damage. These systems necessitate lightweight, cost-effective convolutional neural networks (CNNs) for real-time operation. Effective deployment on AI-assisted edge devices is crucial for optimal performance. To mitigate this problem, we present the optimized fire attention network (*OFAN*) for effective and efficient fire detection. In the re-engineered attention block, we swapped the convolution layers by dilated variants and integrated additional dense layers to capture global context and refine more weight optimization. We calibrate the *OFAN* for real-time processing using a lightweight and efficient feature extractor backbone model. Additionally, a challenging fire dataset is a critical contribution that contains extremely diverse, blazing, and non-fire, captured in lighting and foggy environments. It advances traditional fire detection samples by considering low-light and foggy conditions. A comprehensive experiment is conducted over three widely used fire detection datasets, and our proposed *OFAN* outperforms state-of-the-art. The proposed *OFAN* achieved 96.23%, 96.54% and 94.63% accuracies over BoWFire, FD and the newly proposed *DiverseFire* dataset, respectively. Our research sets a standard for fire detection over edge devices, offering improved accuracy and better frames per second (FPS) performance.**

***Index Terms*—Attention Mechanism, Convolutional Neural Network, Deep Learning, Disaster Management, Fire Classification, Fire Detection, Fire Localization, IoT, Surveillance System.**

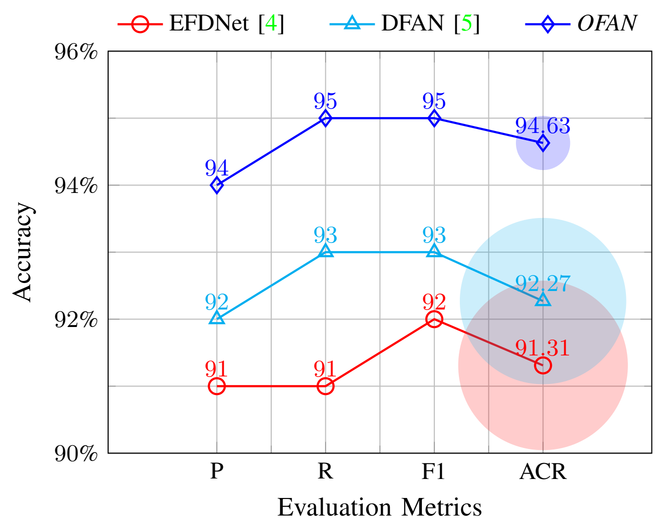


Fig. 1. A comparative analysis between the proposed *OFAN* and the current studies in terms of Precision (P), Recall (R), F1-Score (F1), and Accuracy (ACR). The variant circle size in the graph represents the number of parameters. The larger circles indicate a higher number of parameters. The proposed *OFAN* has fewer trainable parameters and outperforms as compared to state-of-the-art models in terms of various evaluation metrics, demonstrating its effectiveness and efficiency.

I. Introduction

FIRE is a disaster deemed highly dangerous due to its rapid propagation and devastating impact. Promptly controlling a fire is exceptionally challenging, especially in areas abundant with highly combustible materials, such as forest woodlands and residential zones. Fires can ignite due to human activity, equipment failure, elevated temperatures, climate variations, among other factors. An uncontrolled fire can inflict substantial damage on an area's economy, ecology, and environment. Forest and bushfires represent the most perilous types of fires, as their intensity can rapidly escalate, leading to significant environmental devastation. For example, the bushfires in Australia from January to March 2020 burned close to 19 million hectares, took the lives of 33 people, killed around 1500 million animals and destroyed more than 3,000 shelters [1] [2] [3].

A similarly devastating fire in California led to numerous fatalities and extensive property damage [6]. Furthermore, fires in buildings and automobiles are a threat to human safety and property. A report from early 2018 reveals that between 1993 and 2016, approximately 4.5 million fires cases were outlined

Manuscript received XXXX; Accepted: XXXX, Published: XXXX.

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (2020-0-00959, Fast Intelligence Analysis HW/SW Engine Exploiting IoT Platform for Boosting On-device AI in 5G Environment) and also conducted in ICTFICIAL OY and is partially supported by the European Union's Horizon Europe program for Research and Innovation through the aerOS project under Grant No. 101069732, and by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R40), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Corresponding authors: JaeSeung Song and Tarik Taleb.

Naqqash Dilshad and JaeSeung Song are with the Department of Convergence Engineering for Intelligent Drone, Sejong University, Seoul 05006, South Korea (Email: naqqash@ieee.org and jssong@sejong.ac.kr), Samee Ullah Khan is with Sejong University, Seoul 05006, South Korea (Email: samee@ieee.org), Norah Saleh Alghamdi is with the Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia (Email: nosalghamdi@pnu.edu.sa), Tarik Taleb is with University of Oulu, Oulu 90570, Finland and the Department of Convergence Engineering for Intelligent Drone, Sejong University, Seoul 05006, South Korea (Email: talebtarik@gmail.com).

across more than 50 countries, resulting in an estimated 62,000 fatalities [7]. In China, between 2009 and 2015, the average annual count of vehicle fires was 20,000, leading to an annual financial burden of roughly 370 million chinese yuan [8]. In the United States, data from 2019 showed that there were approximately 189,500 highway vehicle fires, leading to 550 deaths [9]. By leveraging smart techniques that utilize various forms of sensory data, early fire detection can substantially mitigate loss of life and property damage in most scenarios.

Over the past few years, various scalar sensor-based systems for fire detection have been investigated by numerous researchers. A variety of smart devices and services are being tested for fire detection. These cost-effective systems utilize sensors such as flame, particle, temperature, and smoke detectors [10]. While these devices may be suitable for indoor environments, they require human interaction to activate, as well as proximity to the fire. Additionally, assorted components of equipment are necessary for timely delivery of information about the fire, such as its degree of burning, size, and location. Conversely, various authors [11] [12] have probed vision sensor-based fire detection systems, which offer numerous benefits when compared with scalar sensor-based systems, including no need for human intervention, an immediate response, wide coverage, and environmental robustness.

Vision sensor-based fire detection techniques are mainly classified into two categories: machine learning (ML) and DL-based methods. ML-based approaches for fire detection concentrate on the shape, motion, color, and texture features of an input frame [13] [10]. These techniques heavily rely on the handcrafted features. However, selecting the most prominent features is a demanding task, as various materials have different flame colors, light and air flow have different affects on fire shape. Hence, balancing the accuracy, loss, and false-positive rate (FPR) and false-negative rate (FNR) metrics is an unresolved challenge through these techniques.

Nowadays, DL approaches have gained popularity in various computer vision (CV) domains and show tremendous performance in fire detection [10]. DL methods have shown better detection accuracy and significantly lower FAR when compared to ML-based techniques. However, DL has limited capabilities to classify and locate complex fire scenes, including fire-colored lighting, fire-like objects, and sunlight that appears like fire. The existing literature [4]–[6] reveals that there is a scarcity of fire image samples for the purpose of training and testing. The publicly available datasets lack diversity, which are mostly small in size and are insufficient to produce reliable and efficient models. As a result, in this study, we accumulated multiple benchmark datasets containing fire and non-fire samples. These datasets were merged to create composite samples thoroughly utilized in this work. Furthermore, the inadequate performance of existing deep models in terms of both efficiency and accuracy poses a challenge to the practical application of the systems.

As a solution to these challenges, this article introduces *OFAN* that utilizes a lightweight CNN model followed by a modified fire attention module formulated through optimized channel attention (OCA) and optimized spatial attention (OSA) mechanisms. As a result of the attention module, deep models are able to detect the presence of fire, fire-like objects, and colors. The proposed method outperforms the state-of-the-art in terms of accuracy and having fewer trainable parameters, as presented in Fig. 1. To choose the optimal collection for the proposed model numerous backbone models are deeply checked with different attention combinations. To this end, a comprehensive empirical study effectively analyzes the performance of the individual model setup. After rigorous analysis, we found that MobileNetV3Small showed the best scene classification score due to usage of variant kernel size, hard-swish activation function, lightweight squeeze-and-excite attention, and our modified attention module. In the modified attention module the OCA module exploits the inter-channel relationships between features in order to refine backbone features from pre-trained CNNs, with the tensor maps of each channel being viewed as single detector. By leveraging the inter-spatial relationships of pixels, the OSA module identifies informative features, thereby complementing the OCA module's activities. In terms of diversity of data, we collected highly diverse samples from nine different public benchmark datasets. The entire contributions of our work are summarized as follows:

1) We present a novel approach called the *OFAN* that incorporates OCA and OSA modules. By integrating these modules with various backbone models, our research achieves state-of-the-art classification results, expanding the potential application domains. Our proposed method yields an inference model of an equitable size, roughly 12 megabytes (MB), making it highly suitable for deployment on edge devices equipped with embedded vision capabilities. Through rigorous experimentation, we have demonstrated that the proposed method outperforms state-of-the-art approaches in terms of accuracy and efficiency. As a result, we consider our method to be a strong contestant for integration into disaster management systems, where timely and accurate fire detection is of paramount significance.
2) To overcome the scarcity of fire images and facilitate researchers in vision-based fire detection, we combined fire images from various well-known public datasets, resulting in a diverse collection of 47,000 images. This dataset includes challenging fire scenarios such as burning clouds, intense light sources, and red objects. Although benchmark datasets exist for specific environments, uncertain surveillance environments are currently not represented in any dataset. To fill this gap, we have curated a dataset that incorporates synthetic fire images with fog and low light conditions. By including these challenging scenarios, our dataset aims to support the development of robust fire detection algorithms specifically tailored to uncertain environments. Our dataset will be publicly available at (https://github.com/naqqashdilshad/OFAN) to encourage collaborative research in the field.
3) In order to evaluate how well and diverse our proposed *DiverseFire* dataset performed, we re-implemented the state-of-the-art methods in order to determine its performance. We identified the most suitable approaches

for each challenge posed by the *DiverseFire* dataset through a benchmarking study, that enabled us to determine which ones were most suitable for addressing the challenges. Furthermore, the study provides a starting point for further advances in fire detection, as it provides a deep understanding of the performance landscape of this field in addition to providing a baseline for assessing future progress.

The subsequent sections of this paper are structured as follows: Section II presents a comprehensive literature review, outlining the ML-based, and DL-based methods with concise descriptions. Section III provides a brief overview of the proposed method. In Section IV, we discuss the dataset, performance evaluation, parameter settings, quantitative along with qualitative results obtained through our research. Finally, in Section V, we conclude the article, highlighting the limitations of our study and suggesting potential areas for future research.

II. Related Work

The field of fire detection has attracted a significant amount of research interest, with a particular focus on developing CV-based approaches for the early and accurate identification of fire. These methods can be broadly classified into two categories: ML-based and DL-based fire detection methods.

A. ML-based Methods

The primary objective of these methods is to leverage shape, motion, color and texture features present in an input image for effective fire detection. In baseline methods, color features are extracted using color spaces such as red green blue (RGB) and luminance, chroma blue, chroma red (YCbCr) [14] [15]. There has also been research that combines statistical color features, super-pixel texture discrimination, and fuzzy logic into an integrated form to detect fires. In addition, the analysis of objects in motion [13], employing optical flow features [16], has been widely used to detect fires. As a result, there is a vast array of research going on in the fire detection domain, with CV-based methods playing a central role in facilitating the accurate and timely detection of fires.

However, ML-based methods are often susceptible to environmental factors such as fire-like objects that are in motion and objects with orange or red tint, which contribute to higher FPR. Moreover, brightness fidelity cannot adequately depict the overall impression of fire, and optical flow-based techniques are computationally expensive. To address these issues, researchers have utilized trainable classifiers to reduce subjective apprehension effects in the classification phase. For instance, [17] propose the use of uni-modal covariance features from spatial-temporal blocks, Gaussian, and a support vector machine (SVM) to extract and classify fire-colored regions that are constantly in motion. Due to the presence of moving objects that look like fires, varying lighting conditions, shadows, and low accuracy, ML-based approaches may be challenging and time-consuming to use for early fire scene classification and alarm generation. This led researchers to investigate various end-to-end DL-based techniques to enhance fire classification.

B. DL-based Methods

Researchers have investigated the use of CNNs for fire detection in order to address the limitations of ML-based methods. With CNNs, fire detection is more convenient and reliable because features are extracted and classified automatically. As described in this study [18], the best fire detection performance was achieved by GoogLeNet, over VGG13 and AlexNet. Authors in [19] employed VGG16 and ResNet50 models for fire classification, where ResNetFire outperformed VGG16 in terms of accuracy, although the dataset used for experimentation was small. LeNet-5 were utilized for fire classification in [20], and this model achieved superior results than ML-based techniques, but the large footprint and computational complexity of LeNet-5 method make it inappropriate for resource-constrained devices (RCDs).

Furthermore, researchers have attempted to combine ML with CNNs to enhance fire classification. Authors in [21] utilized the combination of SVM and CNN to detect fires by employing Haar-like features and an AdaBoost cascade classifier for region of interest (RoI) extraction. Furthermore, for fire classification, they used a two-tier SVM based on a CNN architecture with four layers. To detect smoke and fire, the authors fused a CNN with motion detection and irregularities [22]. Furthermore, saliency detection was employed to extract RoI in [23]. Authors in [24] introduced temporal and spatial features for fire classification. Additionally, authors in diverse fields have integrated attention-based techniques with CNN architectures to enhance the efficiency of their methods [25]. These techniques show convincing performances due to the selection of most prominent features before classification, which prompted numerous investigators to utilize them for the precise localization and classification of fire [4] [6]. Nonetheless, in these methods, the OCA module was exclusively employed for fire scene localization, but it lacks the necessary level of detail for precise fire localization. The literature indicates that there is scope for a superior performing model for fire classification and localization. Moreover, the scarcity of diversity in the datasets poses a challenge for researchers who seek to evaluate the robustness of their proposed models.

In summary, the existing literature has the following limitations:

- **High False Alarm Rates:** Existing deep learning models for fire detection often suffer from high false alarm rates, meaning they frequently mistake non-fire events for fires. This inaccuracy can lead to unnecessary panic and resource deployment.
- **Slow Inference Speeds:** Conventional fire detection models often grapple with slow inference speeds. This inefficiency poses significant challenges to timely fire detection and subsequent emergency responses, thus hindering their application in real-time scenarios that demand swift action.
- **Insufficient Attention to Pertinent Channels:** Existing models might not effectively accentuate the most relevant channels and capture spatial details with considerable depth, leading to sub-optimal performance. Furthermore, the current systems might not be able to discern the depth

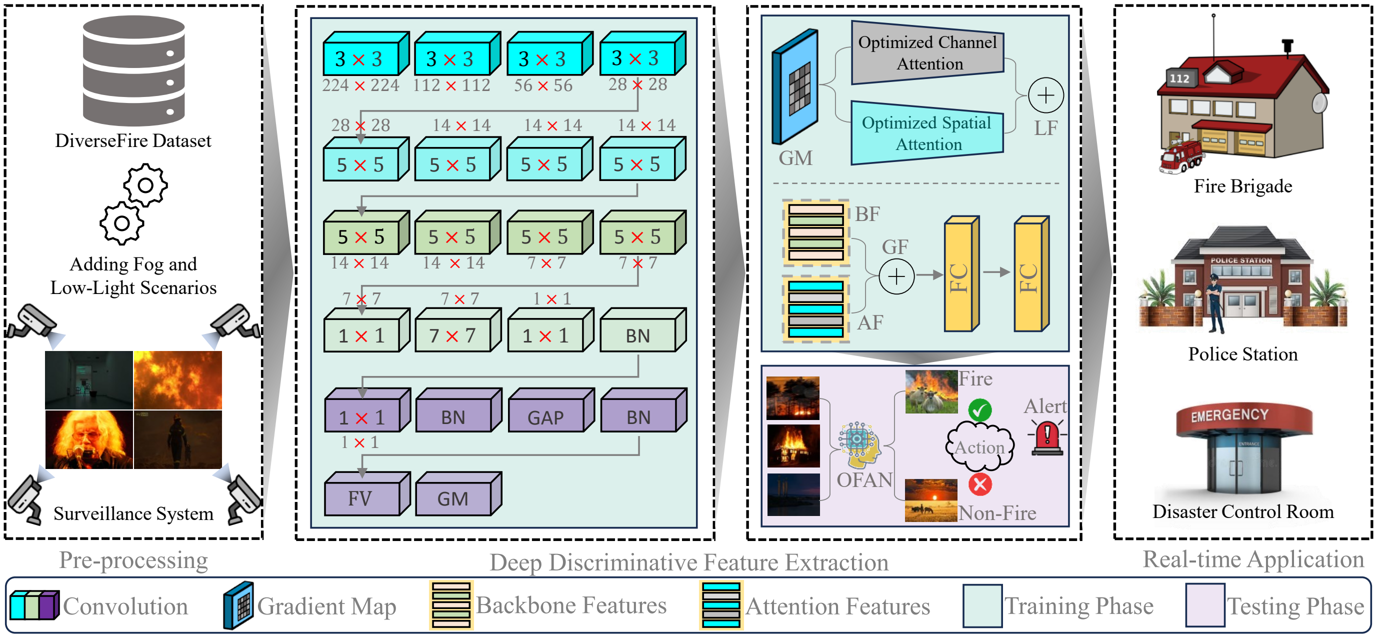


Fig. 2. The proposed *OFAN* framework: Enhancing fire detection with deep discriminative feature extraction, optimized attention mechanisms, and synthesized *DiverseFire* benchmark.

and scope of fires, which can limit their effectiveness in detecting and assessing the size of the fire.

- **Inability to Adapt to New Fire Patterns:** Given the rapidly evolving nature of fire incidents, it can be challenging for existing models not trained on newer fire patterns to detect and classify them accurately.
- **Complex Architecture:** Many existing models employ complex backbones for feature extraction, which can be computationally heavy and unsuitable for real-time usage, especially on edge devices.

As a result, in this study, we introduce a cost-effective model, which is an effective technique for fire scene classification and localization. In addition, it was trained on an extremely diverse set of data.

III. The Proposed Methodology

In this section, the technical details of each component used in the proposed model are provided. The key contributions in this study are as follows: (a) a transfer learning approach is used to compress the parameters of the lightweight pre-trained MobileNetV3, also tuned it for fire scene classification; (b) we optimized the current attention modules for enhancing their learning capability and obtain more representative features; (c) Additionally, we developed a large-scale diverse fire dataset which covers various challenging scenarios. The summarized description of the proposed work is given in the subsequent paragraph.

Overview: We introduced an attention module composed of both OCA and OSA mechanisms in the proposed *OFAN* model. As a result of this module, we are able to obtain more accurate spatial details and more informative channels when determining fire classification. The main architecture of the *OFAN* model is composed of three key steps, including pre-processing training followed by the testing phase, the graphical representation of the entire framework is illustrated in Fig. 2. In the pre-processing step the various benchmark datasets are integrated to create a composite dataset, to make it more complex, we added synthetic foggy and low-light uncertainties. While the training phase involves an efficient backbone followed by a modified attention module to extract more discriminative, representative, and optimized features. Lastly, the proposed *OFAN* is tested over complex data to investigate its compatibility for real-time processing. The step-by-step process is outlined in Algorithm 1. It is particularly designed for classifying various fires into their respective class. To this end, the fire dataset is initially loaded followed by pre-trained CNN model weights. Next, the entire procedure is divided into training and testing. For the training, set all the hyper-parameters and extract features from the backbone model. Later, significant information is forwarded to the attention module. This module mainly comprises OCA and OSA that obtain more meaningful and precise features. Later, the attention-obtained features are fused to make it a more refined descriptor. Next, these are incorporated with backbone features to make a resultant descriptor. Eventually, these feature vectors are passed to the two fully connected, followed by the activation function softmax for classification. When the training phase is completed, the proposed model is saved. In the testing phase, the trained model is loaded and evaluated on various fire samples regarding accuracy, precision, and F1 score. We elaborate on it further in the subsequent sections.

A. Pre-Processing

Pre-processing is an essential step in data analysis, particularly in ML and DL, where data cleaning, transformation,

Algorithm 1: Pseudocode of the proposed *OFAN* for fire detection.

Input: Load Fire Image Dataset
Output: Classifying Fire and Non-Fire with Label and Confidence
Pre-Initialization:
Load Dataset $\rightarrow (D)\Sigma_{i=1}^{n}()$;
Load Pre-trained CNN Weights $\rightarrow M$;
Model Training Procedure:
Proposed Model $\rightarrow P$;
Set The Model Hyper-parameters $\rightarrow \Psi$;
Steps:
while $D \neq 0$ **do**
  for $i \rightarrow I_i \rightarrow (I_1, I_2, I_3, ...I_n)$ **do**
    Backbone features $\rightarrow (\beta) \rightarrow (I_i) \rightarrow (GF)$;
    Modified Attention Block → Æ;
    Opt. Channel Attention → OCA → (GF) $\rightarrow (F_1)$;
    Opt. Spatial Attention → OSA → (GF) $\rightarrow (F_2)$;
    Attention Fusion → (AF) $\rightarrow (F_1) \oplus (F_2)$;
    Resultant Fusion → RF → $(GF) \oplus (AF) \rightarrow M(F_3)$;
    $M(F_3) \rightarrow FC1 \rightarrow FC2 \rightarrow SoftMax$;
Save (model) $\rightarrow \Omega$
Model Testing Procedure:
Load Trained Model $\rightarrow \Omega$;
Load Testing Data $\rightarrow \tau$;
for $j \rightarrow I_j \rightarrow (I_1, I_2, I_3, ...I_n)$ **do**
  Extract Features $\rightarrow (\Gamma)$;
  Evaluate Performance $\rightarrow \alpha$;
  Accuracy → (A);

and filtering make data suitable for analysis and training. In fire datasets, pre-processing ensures that the data is appropriate for the intended scenario as fire datasets are collected in different environments with varying conditions that can affect data quality. The surveillance environment changes over time, leading to fog and darkness in outdoor surveillance scenarios, significantly impacting the model's ability to detect fire, resulting in high false alarm rates. In order to address these challenges and improve the model robustness in such an uncertain environment, the dataset was pre-processed. The proposed dataset was pre-processed by adding fog and low-light conditions through image processing techniques into all images to ensure the model's generalization ability in complex and critical scenarios, reducing the risk of high false alarm rates, and improving the accuracy and detection of the fire in different environments and scenarios. The fog effect was simulated by setting a coefficient of -1, while a coefficient of 0.7 was used for low-light conditions. These specific values were determined through experimentation to strike a balance between realism and model performance. By introducing these challenging scenarios during pre-processing, the model becomes more adept at detecting fires under adverse conditions, thus strengthening its generalization capability and ensuring reliable performance in real-world scenarios.

B. Motivation for using MobileNetV3Small for Fire Detection

In scenarios where resources are limited and timely decision-making is crucial, such as disaster management,

TABLE I
STATISTICS COMPARISON OF ALL LIGHTWEIGHT VARIANTS OF MOBILENET

Parameters	MobileNetV1_1.0	MobileNetV2_1.0	MobileNetV3Small_1.0
MACs (millions)	569	300	66
Parameter (millions)	4.24	3.47	2.9
Top-1 Accuracy (%)	70.9	71.8	68.1
Top-5 Accuracy (%)	89.9	91.0	87.4

the selection of an appropriate model is of utmost importance. Among various pre-trained CNN architectures, MobileNetV3Small is preferred due to its compatibility with hardware architectures which has limited processing capabilities such as field programmable gate arrays (FPGAs), smart sensors, and Raspberry Pi, and its suitability for IoT environments. Compared to MobileNetV1 and MobileNetV2, MobileNetV3Small offers better accuracy with fewer computations and learned parameters, making it a more compact and efficient choice. MobileNetV3Small employs depthwise separable convolutional layers instead of traditional convolution. The kernels processing occurs at the same time across the channels while the traditional convolution employs filter over individual channels that is computationally complex procedure. Table I provides a comparison of statistics for all three variants.

The expression "1.0" denotes the version of MobileNetV3Small, and MAC stands for Multiply-Accumulate operations, which measure the number of computations required to make an inference on a single image having input size of 224×224×3. According to this measure, V3Small is almost nine times as fast as V1 and five times as V2. Mobile devices are also much slower at computing than they are at accessing memory, so V3Small has fewer parameters than both V1 and V2. Additionally, V3Small has slightly less accuracy on the ImageNet dataset but outperforms both V1 and V2 in terms on inference speed. These features validate our selection of MobileNetV3Small in our proposed *OFAN* model.

C. Deep Discriminative Feature Extraction

CNNs have shown remarkable abilities in automatically selecting the most prominent features from the input image, making them ideal for different CV applications. However, selecting a domain-specific CNN architecture that provides precise predictions and a balanced computational complexity for real-world applications is a challenging task. Researchers have conducted significant contributions in vision-based fire detection utilizing pre-trained models to classify and localize fire scenes based on their target datasets [6] [26]. Fine-tuning pre-trained models enables the neural network to acquire domain-specific visual features by making effective use of their trainable parameters. It is an effective initialization technique for vision-based classification tasks since pre-trained networks feature an extensive and robust feature extraction pipeline. Utilizing feature extraction methods that have already proven successful in many CV domains, we have employed multiple lightweight backbone feature extractors, such as Xception, EfficientNetB0, MobileNetV1, MobileNetV2, and

MobileNetV3Small [27] to decide the prominent feature selection mechanism for fire classification in extremely uncertain and challenging scenarios. The main reason of selecting these models is that they are efficient and having higher Top-1 accuracy among others [27].

It is widely accepted in the research community that CNNs have the capability to acquire informative and distinctive features from raw input data. However, finding the optimal configuration for a CNN requires careful consideration of evaluation metrics, data quantity, quality, and the specific problem being tackled. We evaluated various CNN architectures and parameter settings for fire detection in both certain and uncertain environments. Subsequent to substantial experimentation with attention mechanism, we found that the MobileNetV3Small outperformed other pre-trained models such as Xception, EfficientNetB0, MobileNetV1, and MobileNetV2.

We used a fine-tuned version of the MobileNetV3Small architecture that is customized for fire detection in uncertain surveillance environments. The baseline MobileNetV3Small was initially trained on the ImageNet dataset, which contains 1,000 object classes for classification. However, since MobileNetV3Small can learn richer features than various other CNN models, we reused its acquired weights to achieve precise fire detection. To achieve this, we added global average pooling layer (GAP) that calculates the average of each feature map across its spatial dimensions, condensing the information into a one-dimensional vector, providing a global summary of the input feature maps to the last fully connected layer, allowing for classification into fire and non-fire categories. We also added an optimized attention modules to the main building block of the MobileNetV3Small to further enhance its performance for fire detection. The proposed architecture comprises of 16 layers as depicted in Fig. 2, the model pipeline starts with a (3×3) convolution layer with H-swish activation function which takes the input in (224×224) dimension before it passes the given input to the bottleneck block. The bottleneck layer contains three (3×3) layers which utilizes the ReLU6 activation function, followed by eight consecutive (5×5) bottleneck layers which also employs H-swish function. After that (1×1) convolution layer, a (7×7) pooling layer and another (1×1) convolution layer coupled with batch normalization (BN) summons up the backbone architecture. In addition, the obtained deep discriminative features from the backbone is termed as gradient map (GM), GM is forwarded to the optimized attention block where local fusion (LF) of features occurs. The attention features (AF) from attention block and backbone features (BF) are concatenated as global features (GF) are then forwarded to the dense layer. According to Fig. 2 an input image is processed through the proposed architecture in order to obtain an inference on it, which provides discriminative features that are further processed to the attention block.

D. Optimized Attention Mechanism

Several CNN-based models followed by attention mechanisms in various domains provide exemplary performance for video stream due to the high frame resemblance between

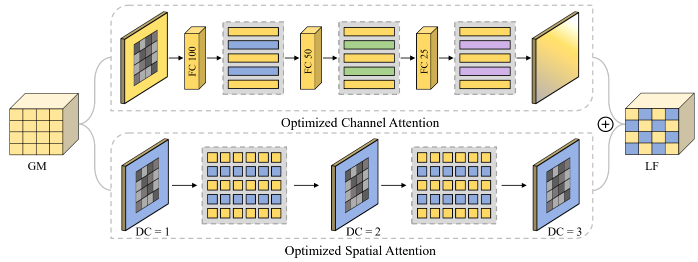


Fig. 3. Block diagram of a modified attention block, where fully connected (FC) and dilated convolution (DC) with variable dilation rate are employed to extract more prominent features from the given gradient map (GM). Furthermore, parallel attention feature maps are integrated as local fusion (LF).

frames [28]. In these studies, individual channel attention (CA) and spatial attention (SA) modules were used and limited performance was obtained on image data due to the diversity of the data and module selection. For image-based fire detection, some studies have incorporated only a channel module into CNN architectures [4] [6]. For simple scenarios, integrated channel information with a DL model is highly effective; however, in complex and uncertain scenes present in the *DiverseFire* dataset, an integrated CA module has limited performance. To improve fire classification and fire localization, we integrate OCA and OSA as part of the *OFAN*, which represents the fire attention component. The integrated (OCA + OSA) attention scheme is used in the fire attention module to extract and localize the most important regions.

Taking inspiration from the approach contained in [29], an optimized attention mechanism block incorporating the OCA and OSA are employed for detecting fire in which each block has a diverse function based on its architecture. An illustration of the optimized attention block can be found in the Fig. 3. A primary principle of its design is that the foreground and important features are the only factors that are focused on rather than unimportant background or redundant information. We therefore follow their policy in this study and make adjustments attention module for the purpose of detecting fire. There are two attention blocks in this module, which are called OCA and OSA. Each fully connected layer in the OCA infers high-level features that are considered the semantic information specific to an object/class and differ in the way they encode these features. Our objective was to enhance the representation of the features in particular semantics of the channel maps to ensure that mutual information is maintained between them. The feature maps are analyzed by noting their associations.

Content-related information is extracted from a wide range of data in order to represent discriminative data features relevant to understanding a scene. Due to its ability to gather homogeneous contextual information from a wide range of sources, OSA is sometimes called the rich relationship module. Moreover, this technique improves the model's feature representation ability by encoding a great deal of contextual information into native features. We modified the existing attention module where each module focuses on its intended purpose. The OCA combines the input data using global average pooling (GAP) and then uses three dense layers to obtain channel information. In the same way, the OSA uses dilated convolutional layers (DCL) to focus on the most important parts. The overall attention module process is presented using mathematical formulas, while the detailed architecture is given in Fig. 3. For the given input activation map V $\in$ $R^{i \times j \times k}$ specifies a detailed attention map based on a 3D map. The final feature map V′ is evaluated as:

$$V\prime = V + V \times M(V), \tag{1}$$

Two attention mechanisms are involved in the proposed method in a unified manner, where the skip learning mechanism followed by attention modules assist the gradient flow. In Equation (1), we perform element-wise multiplication. To calculate the final attention map $M(V)$, we first compute the OCA $Mc(V) \in R^{X}$ and the OSA $Ms(V) \in R^{Y \times Z}$ using the given Equation (2).

$$M(V) = \delta\big(Mc(V) + Ms(V)\big), \tag{2}$$

The activation function used in this framework is represented by the sigmoid function, which is denoted as σ. To accommodate these features, we resize the attention output before integrating them i.e., $R^{\,i \times j \times k}$. Each channel comprises more representative features that exploit the relationship between the channels. To produce a channel vector Vc $\in$ R^{X}, a feature map is compiled against each frequency using a pooling operation. Multilayer perceptron (MLP) evaluates channel information from the resulting vector, which encodes global information. The SA output is adjusted and re-scaled using BN (B_{Norm}), despite using the MLP layer. Eventually, the OCA is computed by using the Equation (3).

$$\begin{aligned} Mc(V) &= B_{Norm}\Big(Vc\big(AP(V)\big)\Big) \\ &= B_{Norm}\big(r_1(r_0 P(V) + k_0 + k_1\big), \end{aligned} \tag{3}$$

The terms "k_0" and "k_1" represent single and the whole channel details, correspondingly. On the other hand, "r_0" and "r_1" are utilized to compute the OCA over the channel by applying a certain scaling down strategy.

$$Ms(V) = B_{Norm}\left(\int_1^{1\times1}\Big(\int_2^{3\times3}\big(\int_3^{1\times1}(V)\big)\Big)\right). \tag{4}$$

In OSA, salient locations are identified within an image using spatial feature maps. Focusing on salient locations requires a filter with a large receptive field. In this study, DCLs with dilation rate ($d = 1, 2, 3$) were utilized to enlarge the receptive field and reduce the total sum of trainable parameters. This work utilizes a (1×1), (3×3) and (1×1) DCLs to accumulate the most prominent contextual features with a scale down strategy same as OCA. Lastly, a (1×1) convolution with a B_{Norm} is utilized to improve and resize the feature map, which is computed through Equation (4).

In summary the mathematical modeling of the proposed scheme work is provided as follow, Equation 5, 6, 7, 8, 9, 10, and 11 represents depthwise convolution, global average pooling, BN, SA, CA, feature fusion and softmax respectively.

$$\hat{D}_{x,y,z} = \sum_{v,w} \hat{X}_{v,w,z} \cdot F_{x+v-1,y+w-1,z} \tag{5}$$

where $\hat{X}$ is the depthwise convolutional kernel with size $D_X \times D_X \times Z$, where the filter z_{th} in $\hat{X}$ is applied to z_{th} channel in F to yield the z_{th} channel of the filtered output feature map $\hat{D}$.

$$\text{GAP}(X)_c = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W} X_{i,j,c} \tag{6}$$

$\text{GAP}(X)_c$ represents the average value for the c_{th} channel computed by taking the average of all values in that channel across the entire spatial dimensions of the tensor.

$$\hat{x} = \frac{x-\mu}{\sqrt{\sigma^2+\epsilon}} \cdot \gamma + \beta \tag{7}$$

$\hat{x}$ depicts how BN normalizes the input x by subtracting the mean, dividing by the standard deviation, and then scaling and shifting the result using learned parameters γ and β.

$$S\left(F\right) = Softmax\left(W_8 * F + B_8\right) \tag{8}$$

Here, $*$ denotes the convolution operation, W_8 are the weights, and B_8 are the biases for the SA mechanism.

$$C\left(F\right) = \sigma\left(W_c F + b_c\right) \tag{9}$$

W_c and B_c are the weights and biases for the CA mechanism, and σ denotes the sigmoid activation function.

$$F_{\text{fused}} = F_{backbone} \cdot S\left(F_{backbone}\right) \cdot C\left(F_{backbone}\right) \tag{10}$$

Here, we incorporate the backbone feature map with the attention mechanisms. Assuming $F_{backbone}$ is the backbone feature map, we apply both SA and CA.

$$\text{Softmax}\left(z_i\right) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{11}$$

The Softmax output is the probability distribution over the classes indicating the presence or absence of fire in the input scene.

IV. Experimental Results

This section, provide a concise description of the experimental setup, the included datasets, training, and evaluation metrics. Following that, we present both quantitative and qualitative analysis of *OFAN* in comparison with state-of-the-art methods. Finally, we conduct an ablation study to assess the performance of the *OFAN* model.

A. System Configuration and Implementation Details

All experiments were performed using Python 3.8, Keras DL framework with TensorFlow backend implementation, on an Intel Core i9 12900K clocked at 5.0 gigahertz (GHz) with an NVIDIA GeForce RTX 3090 GPU renowned for its exceptional DL acceleration capabilities, is equipped with a staggering 24GB of onboard memory. With its high floating-point arithmetic performance, this powerhouse GPU can reach 35.58 tera-floating-point operations (TFLOPs), ensuring high-performance computing for demanding AI tasks. The *OFAN* model, along with the ablation models, were trained for 30 epochs with the default input size of (224×224×3) of the proposed model. The batch size was set to 32, and Adam optimizer was employed, using a learning rate of 1e-4 and a momentum of 0.9. These hyper-parameters were selected based on comprehensive experiments.

B. Datasets and Evaluation Metrics

In real-world fire detection scenarios, the manifestations of fire can be incredibly varied, influenced by diverse factors such as environmental conditions, fire types, fire size and intensity, and lighting conditions. Existing fire detection datasets such as BoWFire [16], Yar [30], Sharma [31], DeepQuestAI [32], Saied [33], Carlo [34], Foggia [35], SV-Fire [11], DeepFire [26] and FD [4] each capture different aspects and characteristics of fire scenarios. However, no single dataset encompasses all the potential variations and complexities that might be encountered in actual fire incidents.

Thus, in order to ensure a robust and versatile fire detection model, it is crucial to train on a composite dataset that reflects a broad spectrum of fire conditions. This justifies our decision to combine the publicly available benchmark datasets for this study. This composite dataset not only includes a wider range of fire scenarios, but also covers diverse environmental conditions and fire characteristics. This ensures our model is trained on and validated against a more comprehensive and representative sample of potential real-world fire situations, ultimately enhancing its generalizability and reliability in practical applications. For a fair evaluation of the proposed *OFAN*, we employed the following benchmark datasets.

1) **BoWFire dataset** [16] is highly imbalanced, diverse and compact, with only two binary classes, namely fire and non-fire. The fire class comprises 119 samples, whereas the non-fire class contains 107 samples.
2) **FD dataset** [4] It is a combined dataset, comprised of Foggia's [35] and the BoWFire datasets [16], with additional images collected from the Internet, in order to upgrade the dataset and add new instances about fire and non-fire scenarios. Finally, a dataset containing 50,000 images was generated for fire detection, each class containing 25,000 images.
3) ***DiverseFire* dataset** comprises of a collection of samples that have been classified into two distinct categories, one containing fire and the other without fire. The dataset is mainly a composite dataset that comprises of BoWFire [16], Yar [30], Sharma [31], DeepQuestAI [32], Saied [33], Carlo [34], Foggia [35], SV-Fire [11], DeepFire [26]. The lack of benchmark datasets representing uncertain surveillance environments poses a significant challenge for researchers in developing effective and adaptable systems. The dataset has a diverse range of images captured from different angles and backgrounds to ensure that the model is effectively trained to distinguish between the two classes. A total of 24,144 in fire and 22,980 in non-fire category were initially compiled for this dataset. Additionally, foggy and low-light conditions were introduced in the dataset to enrich the complexity of fire detection scenarios for DL models. This further augmented the total number of images to 47,124, thereby making the dataset a challenging one to work with. A comprehensive detail of all the included datasets are listed in Table II, while the exemplar samples from the newly proposed *DiverseFire* dataset are illustrated in Fig. 4. The main reason behind the name of *DiverseFire* dataset are as follow: (a) In real-world the scenes are complex and highly diverse, which means that type, size, intensity, and illumination conditions in fire is very challenging; (b) Existing fire datasets mainly cover different aspects rather then focuses on complexity scenarios; (c) Obtaining best performance over a wide spectrum of fire conditions indicating the robustness and generalizability of the model.

In our evaluation, we employed the same set of parameters

TABLE II
OVERALL STATISTICS OF THE INCLUDED DATASETS AND THE NEWLY PROPOSED DATASET HIGHLIGHTED IN **BLUE** TEXT.

#	Dataset	Fire	Non-Fire	Total Images
1	BoWFire [16]	119	117	226
2	Yar [30]	1000	1000	2000
3	Sharma [31]	110	541	651
4	DeepQuestAI [32]	1000	1000	2000
5	Saied [33]	755	244	999
6	Carlo [34]	120	120	240
7	Foggia [35]	7018	7018	14036
8	SV-Fire [11]	1000	500	1500
9	DeepFire [26]	950	950	1900
10	Fog Scenario	12072	11490	23562
	Low-light Scenario	12072	11490	23562
	DiverseFire	**24144**	**22980**	**47124**

Fig. 4. Various instances of Actual fire and fire like scenes from the newly proposed *DiverseFire* dataset, the first and third column depicts fire while the second and the last row shows non-fire samples.

TABLE III
QUANTITATIVE ANALYSIS OF THE PROPOSED *OFAN* WITH STATE-OF-THE-ART ON THE INCLUDED BENCHMARK DATASETS. THE **BLUE** AND **RED** TEXT REPRESENT THE BEST AND SECOND-BEST PERFORMANCES, RESPECTIVELY.

Methods			BoWFire [16]				FD [4]				*DiverseFire*			
			P	R	F1	ACR	P	R	F1	ACR	P	R	F1	ACR
ML	Traditional Models	FD-GCM [15]	0.55	0.54	0.54	-	0.63	0.90	0.74	0.69	-	-	-	-
		FFD-ANN [36]	-	-	-	-	0.71	0.73	0.72	0.71	-	-	-	-
		FPC [13]	-	-	-	-	0.52	0.99	0.68	0.53	-	-	-	-
		EFD-IP [14]	0.75	0.15	0.25	-	-	-	-	-	-	-	-	-
		BoWFire [16]	0.51	0.65	0.67	-	-	-	-	-	-	-	-	-
DL	Large Models	ResNetFire [19]	-	-	-	92.50	0.84	0.97	0.90	90.07	0.86	0.87	0.87	86.74
		VGGFire [19]	-	-	-	91.32	-	-	-	-	0.86	0.84	0.84	84.67
		DeepFire [26]	-	-	-	-	-	-	-	-	0.86	0.85	0.85	85.33
		ViT-B/32 [12]	-	-	-	-	-	-	-	-	0.89	0.90	0.90	89.70
	Lightweight Models	LW-CNN [30]	0.86	0.78	0.77	79.00	0.82	0.81	0.81	81.00	0.78	0.77	0.77	77.16
		E-FireNet [11]	0.82	0.80	0.80	80.66	0.84	0.83	0.84	84.00	0.79	0.78	0.80	79.00
		EMNFire [37]	0.90	0.93	0.92	92.04	0.88	0.98	0.93	92.80	0.88	0.87	0.87	87.50
	Attention-based Models	EFDNet [4]	0.81	0.83	0.81	83.33	0.93	0.97	0.95	95.30	0.91	0.91	0.92	91.31
		DFAN [5]	0.95	0.94	0.95	95.00	0.96	0.97	0.96	96.17	0.92	0.93	0.93	92.27
		OFAN	0.96	0.95	0.96	96.23	0.97	0.96	0.97	96.54	0.94	0.95	0.95	94.63

that have been utilized to assess various state-of-the-art fire detection techniques, as referenced in [19] [4] [30] [11]. These parameters include ACR, P, F1, R, FPR, and FNR. Furthermore, mathematical formulations of these parameters are provided in the preceding sources.

C. Performance Evaluation of OFAN

In this section, we will briefly compare and contrast the quantitative and qualitative performances of the *OFAN* and state-of-the-art methods as applied to the analysis of data.

1) Quantitative Analysis: We performed a quantitative analysis of the performances of *OFAN* and state-of-the-art CNN based methods. To demonstrate the suitability of the *OFAN* for the classification and localization of fire scenes, we used three benchmark datasets and our proposed *DiverseFire* dataset. Upon learning that several baseline CNN models could be used as baselines for the evaluation of the proposed method, we conducted an ablation study. There is a quantitative analysis of the existing datasets presented in Table III, and an ablation study of *OFAN* using the proposed dataset is presented in Table IV. According to the results of these detailed experiments, the *OFAN* outperformed the state-of-the-art methods on both the existing dataset and the proposed dataset when it comes to ACR, P, R, and F1.

a) Comparison with ML-based Methods: We compare the performance of the *OFAN* method to the ML-based method using the benchmark datasets BoWFire, FD and the newly proposed *DiverseFire* datasets, which are considered in this study, using the evaluation parameters of [4] [37]. It can be seen from Table III that the *OFAN* has outperformed the current ML-based methods on the datasets referred to in the above. The best methods to determine BowFire's evaluation in terms of the listed metrics are the proposed *OFAN* and DFAN [5]. The proposed *OFAN* outperformed [14] [15], and [16] methods. In addition to that, the proposed model improved the ACR from 95.00% to 96.23% when compared with previous studies. There was a 3.90% greater R value in [13] on the FD dataset. In comparison to ML methods, *OFAN* achieved the best performance in terms of ACR, P, and F1, which illustrates that the proposed *OFAN* is robust and adaptable to various environments because it can autonomously discern intricate patterns and representations from the data, reducing the need for manual feature engineering, as shown in Table III.

b) Comparison with DL-based Methods: Considering the complex objectives associated with fire scene classification, it is crucial for fire disaster management systems to demonstrate robustness in handling various scenarios, including objects that are burning and objects that look like they are burning. Furthermore, certain fire scenarios may suffer from occlusion due to thick fog or being situated at a considerable distance from the camera or low-light hazy environment. Consequently, to assess the effectiveness of fire scene classification and localization, we conducted an evaluation of the *OFAN* on two widely used benchmark datasets and a newly comprised *DiverseFire* as presented in Table III. The re-implemented models are categorized into three sub-categories, that is, large, lightweight and attention-based models. However, our evaluation results demonstrate that *OFAN* achieved the superior performance i.e., 96.23% on the BoWFire [16] dataset, followed closely by DFAN [5] with 95.00%, while the worst accuracy of 79.00% was obtained by LW-CNN [30]. While analyzing the results indicate that our proposed technique exhibits greater robustness compared to state-of-the-art approaches, including EFDNet [4], DFAN [5] and EMNFire [37]. The proposed *OFAN* shows convincing performance in terms of ACR and efficiency. To verify this claim, comprehensive qualitative and quantitative results are conducted followed by time complexity to ensure its smooth execution over the edge devices. All the empirical results are listed in Table III, IV, V, VI and VII.

Within the domain of fire detection, FD [4] dataset holds a prominent position as a currently large scale publicly available fire benchmark. This dataset is ranked among the most challenging for fire detection. By employing this particular dataset, as both fire and non-fire classes share a similar

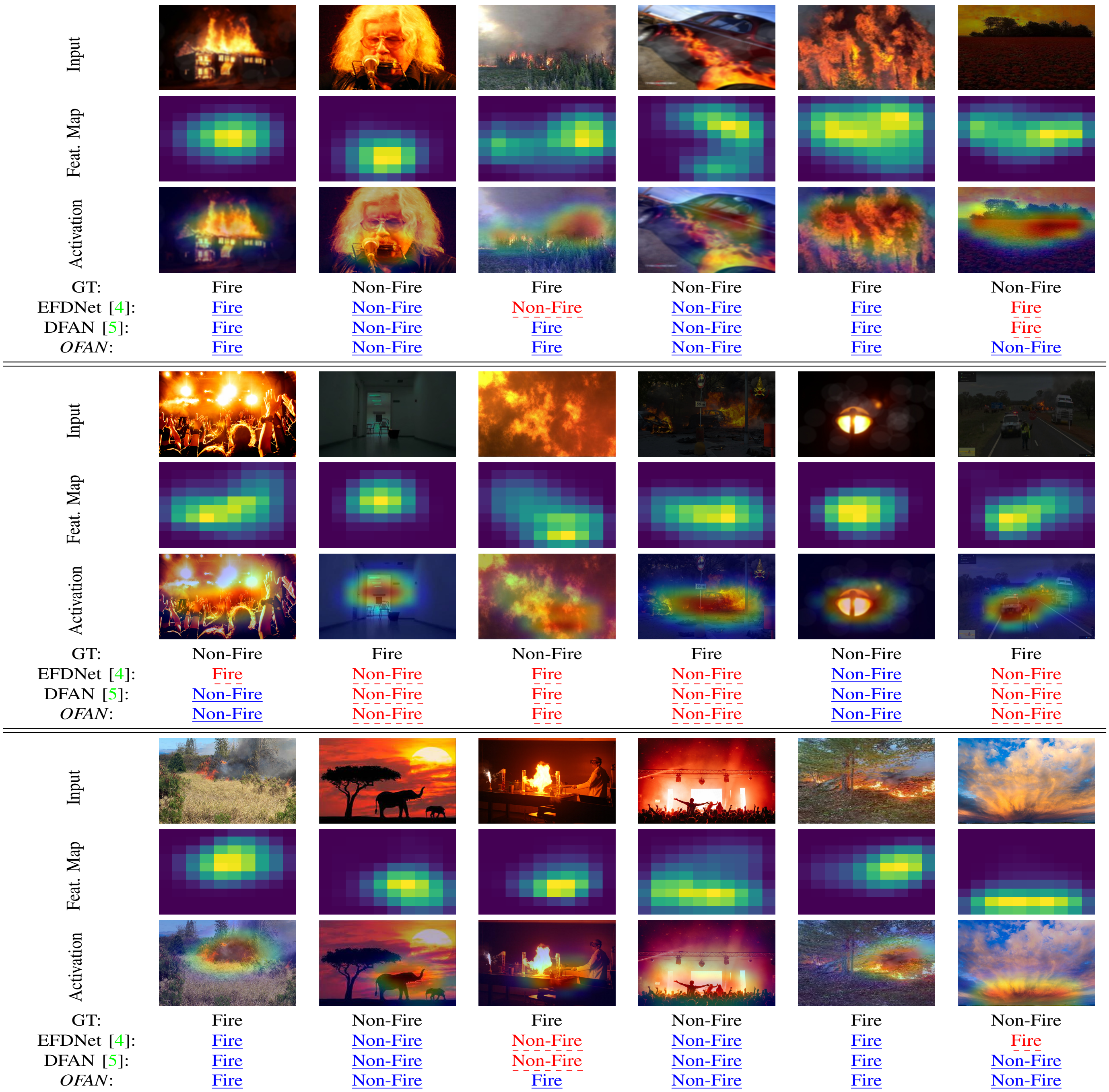


Fig. 5. The objective of fire detection is to determine the presence of a fire. Certain unique scenarios pose challenges to fire detection. We conducted a visual comparison of our *OFAN* on the newly proposed *DiverseFire* dataset. Correct classification results are highlighted in **BLUE** text, while incorrect classification results are indicated by **RED** text.

visual appearance, EMNFire [37] attains the highest R value of 0.98. Nevertheless, in terms of P, F1 score, and ACR values, the *OFAN* outperforms state-of-the-art DL methods with accuracy of 96.54%. The F1 serves as a comprehensive metric that levels the considerations of both P and R. Amid the analyzed methods, DFAN [5] demonstrates the next best performance, at the same time the *OFAN* exhibits the best performance. Overall, the proposed model excels at classification of challenging fire scenes, as listed in the quantitative analysis Table III. Furthermore, for the *DiverseFire* dataset the proposed *OFAN* attains the highest P, R, F1 and ACR of 0.94, 0.95, 0.95, and 94.63%, respectively. Followed closely by DFAN [5] which achieved 0.92, 0.93, 0.93, and 92.27%. Although it seems that the difference is small, the proposed *OFAN* has 21 times fewer trainable parameters, and has 6 times smaller footprint when compared with the second-best performing method i.e., DFAN [5]. In summary the proposed *OFAN* achieved 96.23%, 96.54% and 94.63% accuracy on BoWFire, FD and the newly proposed *DiverseFire* dataset.

Fig. 6. Qualitative analysis of *OFAN* for fire detection on distant objects in outdoor environment. Correct classification results are highlighted in **BLUE** text.

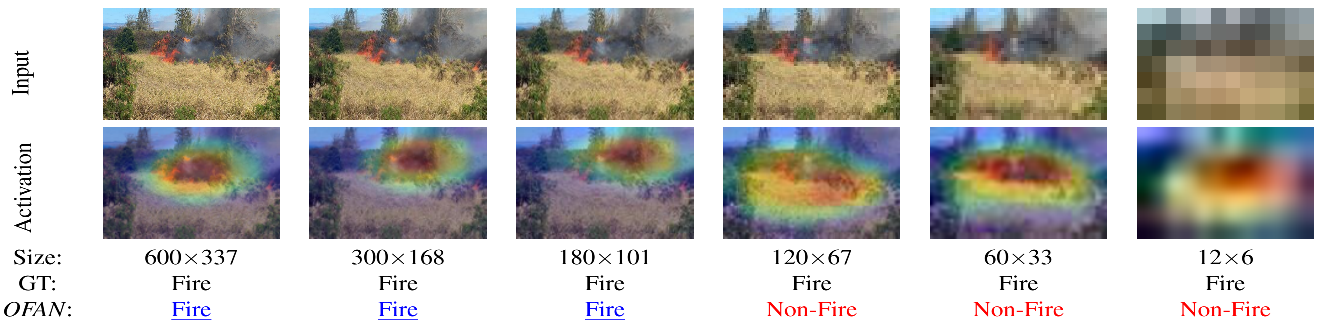


Fig. 7. Qualitative analysis of *OFAN* for fire detection relevant to image size and resolution. Correct classification results are highlighted in **BLUE** text, while incorrect classification results are indicated by **RED** text.

2) Qualitative Analysis: To assess the qualitative performance of EFDNet [4], DFAN [5] and the proposed *OFAN* based on localization and class activation, we conducted an analysis. Fig. 5 presents the results, demonstrating the robustness of the *OFAN* in detecting fire regions within challenging scenes in comparison with other state-of-the-art. Additionally, for every single test sample, we added feature map based on the backbone and activation maps of the *OFAN*, highlighting the most salient parts of the input image that captured the model's attention. Fig. 5 showcases the visual outcomes of the *OFAN* for the most puzzled samples obtained from the proposed *DiverseFire* dataset. The first, sixth and eleventh row for each set represents the input images from the newly proposed and challenging *DiverseFire* dataset. The second and third row depicts the backbone feature maps and activation maps for each class. While the last row represents the predicted label via EFDNet [4], DFAN [5] and the proposed *OFAN*. The first set of input images were correctly classified and localized regions are highlighted with gradient-weighted class activation mapping (Grad-CAM) heat maps via *OFAN*, but EFDNet [4] misclassified the third sample as the fire is quite fire and also occluded by plants. For the last sample, EFDNet [4] and DFAN [5] both inferred inaccurate class activation due to the complex scenario of the sample.

However, for some complex samples shown in the second set of input images, misclassifications occur due to the presence of confusing patterns among the fire and other scenes, making it difficult to distinguish accurately. For the first sample EFDNet [4] confused it with Fire due to high orange tint and fire like illumination in the background. In the second set of input images of Fig. 5, it is apparent that the fire scene was confused with a non-fire scenario by all models, as evident from the feature map and activation the proposed *OFAN* is focusing at the door alley. The third sample which represents a non-fire scene that is, fire like clouds which is really challenging for EFDNet [4], DFAN [5] and the proposed *OFAN* to distinguish it properly is also misclassified as fire due to visual similarity. Furthermore, the fourth sample which is a fire scene is also incorrectly classified because there is a road sign pole in front of the object of interest and secondly this fire scene is has low-light conditions which makes it extremely challenging as a fire scene. The last sample was also incorrectly classified as non-fire as the model is focusing on an ambulance which is in foreground while the vehicle on fire is in the background and the lighting conditions are poor which pose a significant challenge for the included EFDNet [4], DFAN [5] and the proposed *OFAN*.

Although the third set of input images were correctly classified by the proposed *OFAN*, the fire samples presented in this row have low volume and spread of fire, depicting early-stages of fire. In the first and fifth image the fire takes place in bushes, if not detected at early-stages the consequences can be devastating for both human and animals, while the third fire image depicts fire in a laboratory if not distinguished in time, can be hazardous to human lives due to the presence of highly flammable chemicals. The fire in this sample looks like a glass lamp that is why both EFDNet [4] and DFAN [5] depicts incorrect label. Also, the last image which is a cloudy image but tricky for the EFDNet [4] model due to the orange color and flairs. These observations illustrate both the capabilities and limitations of the *OFAN* in effectively identifying and classifying fire regions within complex scenes. The analysis presented in Fig. 5 provides compelling evidence of the *OFAN* exceptional capabilities in accurately detecting fire regions even in challenging conditions. The visual representation clearly illustrates the model's proficiency in this regard. However, upon closer examination, certain images within Fig. 5 exhibit instances where misclassification and inaccurate localization occur; this is due in part to the similar visual characteristics in both classes.

Furthermore, qualitative analysis of the proposed *OFAN* was conducted related to distant objects in Fig. 6. The included samples contain fire at a long range of approximately 30-meter or more. *OFAN* localized the fire for the included samples correctly. In another qualitative analysis related to image resolution is depicted in Fig. 7. The figure showcases one single early-stage fire image, the proposed *OFAN* was tested with different levels of image size and resolution. The proposed *OFAN* correctly labelled the image until the resolution went so low that the channel and spatial features of the input image are no longer like the original.

D. Ablation Study

We conducted ablation studies to identify the optimal configuration for the newly proposed *OFAN*. These studies involved exploring different combinations of attention modules and evaluating the efficacy of the proposed *OFAN* attention method across various composition. The results of these experiments are provided in Table IV and are discussed in the following subsections.

a) Impact of the Optimized Attention Module: To enhance the accuracy of fire scene classification and localization on the newly proposed *DiverseFire* dataset, we incorporated the modified attention modules that is (OCA + OSA) into the several lightweight baseline methods namely Xception, EfficientNetB0, MobileNetV1, MobileNetV2, and MobileNetV3Small. The results, as presented in Table IV, indicates that the baseline CNN models with the optimized attention modules outperformed the methods relying solely on OSA and OCA deep features. This observation can be attributed to the inherent challenges associated with fire classification, which surpasses the complexity of simple ImageNet classification tasks. The integration of optimized attention modules effectively improves the extraction of distinctive features from the input sample, thereby contributing to the improved accuracy of fire classification. Among the tested combinations with baseline methods, the integration of OSA and OCA features with a softmax classifier yielded the best performance due to its superior feature extraction capabilities. Additionally, a closer examination of Table IV reveals that the *OFAN* achieves the highest overall results, whereas OCA and OSA perform relatively poorly when coupled separately with the baseline feature extractor. In summary, the incorporation of both attention modules, in conjunction with baseline features, yields the most favorable outcomes in terms of fire classification and localization on the newly proposed *DiverseFire* dataset.

TABLE IV
PERFORMANCE COMPARISON AMONG VARIOUS STRUCTURES ON *DiverseFire*. OCA DENOTES OPTIMIZED CHANNEL ATTENTION, OSA REPRESENTS OPTIMIZED SPATIAL ATTENTION MODULES. THE **BLUE** TEXT IMITATES THE BEST PERFORMANCE, WHILE THE **RED** TEXT IMITATES THE SECOND-BEST PERFORMANCE.

Baseline Methods	OCA	OSA	P	R	F1	ACR
Xception	-	✓	0.81	0.83	0.81	81.54
	✓	-	0.82	0.82	0.83	82.27
	✓	✓	0.84	0.86	0.84	84.67
EfficientNetB0	-	✓	0.88	0.86	0.85	86.39
	✓	-	0.90	0.87	0.89	88.64
	✓	✓	0.91	0.92	0.92	91.79
MobileNet	-	✓	0.86	0.80	0.83	82.50
	✓	-	0.85	0.83	0.84	84.16
	✓	✓	0.87	0.83	0.85	85.00
MobileNetV2	-	✓	0.90	0.89	0.89	89.33
	✓	-	0.92	0.91	0.91	91.41
	✓	✓	0.93	0.92	0.92	92.50
MobileNetV3Small	-	✓	0.91	0.90	0.90	90.43
	✓	-	0.93	0.92	0.93	92.66
	✓	✓	0.94	0.95	0.95	94.63

TABLE V
FIVE-FOLD CROSS VALIDATION ACCURACIES OF *OFAN* FOR THE INCLUDED DATASETS. THE AVERAGE TEST ACCURACIES ARE LISTED IN **BLUE** TEXT.

Fold	Dataset		
	BoWFire	FD	***DiverseFire***
1	0.9116	0.9369	0.9204
2	0.9243	0.9341	0.9137
3	0.9067	0.9288	0.9210
4	0.9232	0.9313	0.9121
5	0.9174	0.9283	0.9187
Average Test Accuracy	0.9166	0.9318	0.9171

We conducted experiments focusing on the *OFAN* model with different settings: A) Integration of OCA with MobileNetV3Small, B) Integration of OSA with MobileNetV3Small, and C) Integration of both OCA and OSA with the MobileNetV3Small, referred to as *OFAN*. We exclusively evaluated the performance of the optimized attention modules due to the limited robustness observed with and without the module. The experimental evaluations, as presented in Table IV, reveal that the individual integration of OSA with the baseline performs relatively poor as compared to OCA. In addition, the integration of OCA yielded better results compared to the OSA in terms of ACR from 90.43% to 92.66%. Notably, when both modified attention modules are integrated with the baseline method exhibited the highest performance among the evaluated combinations. *OFAN* surpassed MobileNetV3Small + OCA by approximately 2.00% in terms of ACR, as depicted in Table IV. Although, the decisive effect of optimized attention module in *OFAN* yields the best results, when coupled with MobileNetV2 and EfficientNetB0 yields second-best P and R values. Overall, the *OFAN* model achieved the best performance among all the possible combinations, showcasing the efficiency and effectiveness of the proposed approach.

b) K-Fold Cross Validation Evaluation: To further evaluate the strength of the proposed *OFAN* and the included

TABLE VI
TEN-FOLD CROSS VALIDATION ACCURACIES OF *OFAN* FOR THE INCLUDED DATASETS. THE AVERAGE TEST ACCURACIES ARE LISTED IN **BLUE** TEXT.

Fold	Dataset		
	BoWFire	FD	***DiverseFire***
1	0.8907	0.9171	0.8940
2	0.8944	0.9139	0.8971
3	0.8884	0.9012	0.8966
4	0.8826	0.9151	0.9032
5	0.8802	0.9175	0.8980
6	0.8961	0.9188	0.9046
7	0.8713	0.9093	0.8880
8	0.8836	0.9142	0.8977
9	0.8851	0.9074	0.8958
10	0.8768	0.9162	0.8897
Average Test Accuracy	0.8849	0.9130	0.8964

TABLE VII
THE PARAMETERS, MODEL SIZE, MFLOPS, AND TIME COMPLEXITY OF THE PROPOSED *OFAN* AND OTHER STATE-OF-THE-ART. THE **BLUE** TEXT INDICATES THE BEST PERFORMANCE AND THE **RED** TEXT INDICATES THE SECOND-BEST PERFORMANCE.

Method	Parameters (M) ↓	Size (MB) ↓	MFLOPs ↓	Frames Per Second ↑	
				RPi 4B+	CPU
ResNetFire [19]	25.60	98.00	3800	1.07	11.70
VGGFire [19]	138.4	528.00	15500	0.60	10.02
ViT-B/32 [12]	86.00	345.63	8560	-	21.94
DeepFire [26]	143.7	549.00	19670	0.45	8.66
LW-CNN [30]	3.31	25.3	9.6	16.02	27.93
E-FireNet [11]	7.63	76.3	7649	3.71	18.49
EMNFire [37]	3.47	13.23	300	5.0	21.25
EFDNet [4]	23.90	4.80	1130	-	4.43
DFAN [5]	22.47	83.63	141.25	0.83	12.98
OFAN	1.02	12.2	19.42	8.37	25.50

datasets, we conducted a thorough assessment by subjecting it to both 5-fold and 10-fold cross-validation across. The outcomes of the cross-validation analysis reveal that *OFAN* maintains a robust performance across all folds, despite a marginal decline in average test accuracy when dealing with smaller training samples within each fold compared to the entire dataset. This sustained level of performance underscores the resilience and dependability of the *OFAN* model. For a more detailed breakdown of the cross-validation results, please refer to Table V and VI, which provide a comprehensive summary of the 5-fold and 10-fold cross-validation accuracies for each dataset, along with the average test accuracy computed across the 5 and 10-folds. These findings further substantiate the efficacy of *OFAN* model in effectively handling complex datasets, consistently delivering robust results that hold practical significance in real-world fire detection.

E. Time Complexity Analysis

In order to evaluate the proposed model, we compared it with several state-of-the-art methods in terms of parameters, model size, mega-floating-point operations (MFLOPs), and inference time. These factors play a pivotal role in determining the inference speed of DL methods. For the comparison, we selected nine state-of-the-art models namely ResNetFire [19], VGGFire [19], ViT-B/32 [12], DeepFire [26], LW-CNN [30], E-FireNet [11], EMNFire [37], EFDNet [4], and DFAN [5]. We analyzed the computational complexity by examining the MFLOPs and the model size on the disk in MB for each model as listed in Table VII. For evaluation purposes, we ran experiments with two settings: 1) Intel Core i9-12900K clocked at 5.0 GHz coupled with 64GB Quad Channel Configuration RAM clocked at 2,400 megahertz (MHz), and 2) Raspberry Pi model 4 (B+) system-on-chip (SoC) is powered by a 64-bit quad-core Cortex-A72 processor with a clock speed of 1.5 GHz and is accompanied by a generous four GB of onboard main memory. In the large model category DeepFire [26] exhibits higher MFLOPs and a larger model size followed by VGGFire [19], ResNetFire [19], and ViT-B/32 [12]. Similarly, for lightweight models LW-CNN [30] has the lowest number of MFLOPs that is, 9.6 but EMNFire [37] has a smaller size of 13.23 MB. Additionally, in attention-based model EFDNet [4] and DFAN [5] has 1130 and 141.25 MFLOPs respectively, while the proposed *OFAN* method has 19.42 MFLOPs beating the state-of-the-art by a big margin in terms of model complexity, and number of parameters. During the experiments, we observed that the performance limitations, computational complexity, and slower FPS values of EFDNet [4] and DFAN [5] make them less suitable for real-world implementation. Although EFDNet [4] and LW-CNN [30] have smaller model sizes, their ACR values are lower than those of *OFAN*. While in terms of inference speed LW-CNN [30] achieved higher values due to its low number of MFLOPs. Additionally, the proposed *OFAN* inference was hypothetically tested for a high-speed vehicle over a 30-meter distance, the model extracted approximately 4.65 frames. A single frame proved sufficient for detecting fire on a high-speed vehicle. The analytical study is listed in Table IX.

TABLE VIII
THE PARAMETERS, MODEL SIZE, MFLOPS, AND TIME COMPLEXITY OF THE MODELS HAVING MORE THAN **90%** ACCURACY AGAINST THE PROPOSED *OFAN*. THE **BLUE** TEXT DENOTES THE BEST PERFORMANCE AND THE **RED** TEXT DENOTES THE SECOND-BEST PERFORMANCE.

Method	Parameters (M) ↓	Size (MB) ↓	MFLOPs ↓	Frames Per Second ↑	
				RPi 4B+	CPU
EMNFire [37]	3.47	13.23	300	5.0	21.25
EFDNet [4]	23.90	4.80	1130	-	4.43
DFAN [5]	22.47	83.63	141.25	0.83	12.98
OFAN	1.02	12.2	19.42	8.37	25.50

TABLE IX
ANALYSIS OF *OFAN* FOR A HIGH SPEED VEHICLE OVER A 30-METER DISTANCE AND THE EXTRACTED FRAMES VIA RPI 4B+. THE RESULTS FOR HIGH SPEED IS HIGHLIGHTED IN **BLUE** TEXT.

Speed (Kph)	Distance Covered (Seconds)	Extracted Frames
25	4.00	33.48
50	2.16	18.07
75	1.07	8.95
100	0.83	6.94
125	0.60	5.02
150	0.55	4.65

In summary, the proposed *OFAN* outperforms other state-of-the-art studies in terms of parameters, ACR and performs the second-best in terms of size on disk, MFLOPs and FPS when compared with all deep models, on the other hand our method comes on top when compared with models having more than 90% accuracy details of which are listed in Table VIII.

Due to the usage of depthwise separable convolutions, novel activation functions, and architectural enhancements, resulting in a compact model that consumes less memory, offers faster inference times. Additionally, the optimized attention module which mainly focuses on the most important and pertinent features of an input image, the proposed *OFAN* approach achieves higher performance.

V. Conclusion

In the CV, the utilization of CNNs has significantly improved the efficacy of fire detection models. Although, existing CNN-based fire detection approaches exhibit certain disadvantages. They tend to miss-classify fire scenes in challenging and uncertain environments, and their large model sizes and high time complexities make them unsuitable for deployment on RCDs. In order to deal and offset these challenges, we propose a fire detection framework that employs a MobileNetV3Small model coupled with optimized CA and SA mechanisms. Compared with existing CNN-based fire detection methods, our method offers a better balance of accuracy, model size, and inference speed. Aside from its relatively small size, the model is highly suitable for industrial applications of vision-based fire detection techniques. In future, we aim to use object detection techniques and semantic segmentation techniques, which will enables us to accurately identify the fire area in the input samples.

References

[1] A. I. Filkov, T. Ngo, S. Matthews, S. Telfer, and T. 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*, vol. 1, no. 1, pp. 44–56, 2020.

[2] M. R. Nosouhi, K. Sood, N. Kumar, T. Wevill, and C. Thapa, "Bushfire risk detection using internet of things: An application scenario," *IEEE Internet of Things Journal*, vol. 9, no. 7, pp. 5266–5274, 2022.

[3] O. M. Bushnaq, A. Chaaban, and T. Y. Al-Naffouri, "The role of uav-iot networks in future wildfire detection," *IEEE Internet of Things Journal*, vol. 8, no. 23, pp. 16 984–16 999, 2021.

[4] 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*, vol. 29, pp. 8467–8475, 2020.

[5] 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*, vol. 31, pp. 6331–6343, 2022.

[6] S. Majid, F. Alenezi, S. Masood, M. Ahmad, E. S. Gündüz, and K. Polat, "Attention based cnn model for fire detection and localization in real-world images," *Expert Systems with Applications*, vol. 189, p. 116114, 2022.

[7] F. Bu and M. S. Gharajeh, "Intelligent and vision-based fire detection systems: A survey," *Image and vision computing*, vol. 91, p. 103803, 2019.

[8] D. Zhang, L. Xiao, Y. Wang, and G. Huang, "Study on vehicle fire safety: Statistic, investigation methods and experimental analysis," *Safety science*, vol. 117, pp. 194–204, 2019.

[9] N. Csápaiová, "The impact of vehicle fires on road safety," *Transportation research procedia*, vol. 55, pp. 1704–1711, 2021.

[10] A. Gaur, A. Singh, A. Kumar, A. Kumar, and K. Kapoor, "Video flame and smoke based fire detection algorithms: A literature review," *Fire technology*, vol. 56, pp. 1943–1980, 2020.

[11] N. Dilshad, T. Khan, and J. Song, "Efficient deep learning framework for fire detection in complex surveillance environment," *Computer Systems Science and Engineering*, vol. 46, no. 1, pp. 749–764, 2021.

[12] M. Shahid and K.-l. Hua, "Fire detection using transformer network," in *Proceedings of the 2021 International Conference on Multimedia Retrieval*, 2021, pp. 627–630.

[13] T. Celik, H. Ozkaramanli, and H. Demirel, "Fire pixel classification using fuzzy logic and statistical color model," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, vol. 1. IEEE, 2007, pp. I–1205.

[14] T.-H. Chen, P.-H. Wu, and Y.-C. Chiou, "An early fire-detection method based on image processing," in *2004 International Conference on Image Processing, 2004. ICIP'04.*, vol. 3. IEEE, 2004, pp. 1707–1710.

[15] T. Celik and H. Demirel, "Fire detection in video sequences using a generic color model," *Fire safety journal*, vol. 44, no. 2, pp. 147–158, 2009.

[16] D. Y. Chino, L. P. Avalhais, J. F. Rodrigues, and A. J. Traina, "Bowfire: detection of fire in still images by integrating pixel color and texture analysis," in *2015 28th SIBGRAPI conference on graphics, patterns and images*. IEEE, 2015, pp. 95–102.

[17] Y. H. Habiboğlu, O. Günay, and A. E. Çetin, "Covariance matrix-based fire and flame detection method in video," *Machine Vision and Applications*, vol. 23, pp. 1103–1113, 2012.

[18] W. Lee, S. Kim, Y.-T. Lee, H.-W. Lee, and M. Choi, "Deep neural networks for wild fire detection with unmanned aerial vehicle," in *2017 IEEE international conference on consumer electronics (ICCE)*. IEEE, 2017, pp. 252–253.

[19] J. Sharma, O.-C. Granmo, M. Goodwin, and J. T. Fidje, "Deep convolutional neural networks for fire detection in images," in *Engineering Applications of Neural Networks: 18th International Conference, EANN 2017, Athens, Greece, August 25–27, 2017, Proceedings*. Springer, 2017, pp. 183–193.

[20] W. Mao, W. Wang, Z. Dou, and Y. Li, "Fire recognition based on multi-channel convolutional neural network," *Fire technology*, vol. 54, pp. 531–554, 2018.

[21] Z. Wang, Z. Wang, H. Zhang, and X. Guo, "A novel fire detection approach based on cnn-svm using tensorflow," in *Intelligent Computing Methodologies: 13th International Conference, ICIC 2017, Liverpool, UK, August 7-10, 2017, Proceedings, Part III 13*. Springer, 2017, pp. 682–693.

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

[23] L. Shi, F. Long, C. Lin, and Y. Zhao, "Video-based fire detection with saliency detection and convolutional neural networks," in *Advances in Neural Networks-ISNN 2017: 14th International Symposium, ISNN 2017, Sapporo, Hakodate, and Muroran, Hokkaido, Japan, June 21–26, 2017, Proceedings, Part II 14*. Springer, 2017, pp. 299–309.

[24] C. Hu, P. Tang, W. Jin, Z. He, and W. Li, "Real-time fire detection based on deep convolutional long-recurrent networks and optical flow method," in *2018 37th Chinese Control Conference (CCC)*. IEEE, 2018, pp. 9061–9066.

[25] X. Shu, L. Zhang, G.-J. Qi, W. Liu, and J. Tang, "Spatiotemporal co-attention recurrent neural networks for human-skeleton motion prediction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 6, pp. 3300–3315, 2021.

[26] A. Khan, B. Hassan, S. Khan, R. Ahmed, and A. Abuassba, "Deepfire: A novel dataset and deep transfer learning benchmark for forest fire detection," *Mobile Information Systems*, vol. 2022, 2022.

[27] Keras, "Keras Applications," https://keras.io/api/applications/, 2020, (Online; accessed March 22, 2023).

[28] J. Tang, X. Shu, R. Yan, and L. Zhang, "Coherence constrained graph lstm for group activity recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 2, pp. 636–647, 2019.

[29] J. Park, S. Woo, J.-Y. Lee, and I. S. Kweon, "Bam: Bottleneck attention module," *arXiv preprint arXiv:1807.06514*, 2018.

[30] H. Yar, T. Hussain, Z. A. Khan, D. Koundal, M. Y. Lee, S. W. Baik *et al.*, "Vision sensor-based real-time fire detection in resource-constrained iot environments," *Computational intelligence and neuroscience*, vol. 2021, 2021.

[31] A. Sharma, "Fire Detection Dataset," https://www.kaggle.com/general/181150, 2020, (Online; accessed March 22, 2023).

[32] DeepQuestAI, "Fire Detection Dataset," https://github.com/DeepQuestAI/Fire-Smoke-Dataset, 2020, (Online; accessed March 22, 2023).

[33] A. Saied, "Fire Dataset," https://www.kaggle.com/datasets/phylake1337/fire-dataset?select=fire_dataset, 2020, (Online; accessed March 22, 2023).

[34] Carlo, "Fire and smoke.zip," https://www.kaggle.com/datasets/carlo946/fire-and-smokezip, 2020, (Online; accessed March 22, 2023).

[35] 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*, vol. 25, no. 9, pp. 1545–1556, 2015.

[36] D. Zhang, S. Han, J. Zhao, Z. Zhang, C. Qu, Y. Ke, and X. Chen, "Image based forest fire detection using dynamic characteristics with artificial neural networks," in *2009 International Joint Conference on Artificial Intelligence*. IEEE, 2009, pp. 290–293.

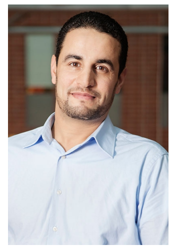
[37] K. Muhammad, S. Khan, M. Elhoseny, S. H. Ahmed, and S. W. Baik, "Efficient fire detection for uncertain surveillance environment," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 3113–3122, 2019.

Naqqash Dilshad (Student Member, IEEE) received his Bachelor's degree in Computer Sciences from Abdul Wali Khan University, Mardan, Pakistan in 2014. He acknowledged his Master's degree in Computer Sciences from COMSATS University, Islamabad, Pakistan in 2018. Currently, he is enrolled in a Ph.D. program at Sejong University, Seoul, South Korea, and serving as a Graduate Research Assistant at Software Engineering and Security Lab (SESLab). His major research domains are Video Analytics, Image Processing, Pattern Recognition, Object Detection, Scene Understanding, Deep Learning for Multimedia Data understanding, IoT, IIoT, and Resource-Constrained Programming. He has published several articles in peer-reviewed journals and conferences in reputed venues including IEEE INTERNET OF THINGS JOURNAL, Elsevier JOURNAL OF NETWORKS and COMPUTER APPLICATIONS, ELSEVIER DIGITAL COMMUNICATIONS AND NETWORKS, TSP COMPUTER SYSTEMS SCIENCE & AND ENGINEERING, MDPI SENSORS, MDPI SMART CITIES, MDPI LIFE, IEEE INTERNATIONAL CONFERENCE ON INFORMATION AND COMMUNICATION TECHNOLOGY CONVERGENCE (ICTC), IEEE INTERNATIONAL CONFERENCE ON SMART INTERNET OF THINGS (SmartIoT) and ASIA PACIFIC INTERNATIONAL CONFERENCE ON INFORMATION SCIENCE AND TECHNOLOGY (APIC-IST).

Mr. Dilshad actively participates in the peer review process, contributing as a Reviewer for various reputable journals, such as IEEE INTERNET OF THINGS JOURNAL, IEEE COMMUNICATIONS LETTERS, SPRINGER INTERNATIONAL JOURNAL OF COMPUTER VISION, MDPI Sensors, MDPI Fire, MDPI Drones, MDPI Sustainability, TSP COMPUTER SYSTEMS SCIENCE & ENGINEERING, TSP COMPUTERS, MATERIALS & CONTINUA and TSP INTELLIGENT AUTOMATION & SOFT COMPUTING).

Samee Ullah Khan (Student Member, IEEE) received the Ph.D. degree in software convergence from Sejong University, Seoul, South Korea, in 2023. He is currently a Senior Research Fellow with the Intelligent Media Laboratory, Sejong University. His main research interests are multimedia data analysis, including person re-identification, action and activity recognition, video summarization, and time series data analysis for power generation and consumption prediction/forecasting. He has published several articles in peer-reviewed journals in reputed venues, including, IEEE INDUSTRIAL INFORMATICS, IEEE INTERNET OF THINGS JOURNAL, IEEE MULTIMEDIA, IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, ENERGY AND BUILDINGS, ALEXANDRIA ENGINEERING JOURNAL, RENEWABLE AND SUSTAINABLE ENERGY REVIEWS, INTERNATIONAL JOURNAL OF INTELLIGENT SYSTEMS, and SPRINGER MULTIMEDIA TOOLS AND APPLICATIONS. He provides professional review services in various reputed journals, including IEEE INTERNET OF THINGS JOURNAL, IEEE OPEN JOURNAL OF ENGINEERING IN MEDICINE AND BIOLOGY, IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING, IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS, ASCJ.

Norah Saleh Alghamdi is an Associate Professor in the Department of Computer Sciences at Princess Nourah bint Abdulrahman University (PNU) in Riyadh, Saudi Arabia. Serving as Vice-Dean of Quality Assurance since 2019, she currently directs business and projects management in her college. She holds a Bachelor's in Computer Science from Taif University and a Master's/Ph.D. from La Trobe University, Australia. Her research spans data mining, machine learning, text analytics, image classification, bio-engineering, and deep learning. Actively engaged in international conferences, she has authored/coauthored numerous articles in reputable journals. She is also a valued member of reviewer committees for IEEE, MDPI, Emerald, and Elsevier journals. Her multifaceted contributions highlight her dedication to advancing computer science education and research.



Tarik Taleb (Senior Member, IEEE) received the B.E. degree (with Distinction) in information engineering and the M.Sc. and Ph.D. degrees in information sciences from Tohoku University, Sendai, Japan, in 2001, 2003, and 2005, respectively.

He is currently a Professor with the Center of Wireless Communications, The University of Oulu, Oulu, Finland, where he is the Founder and the Director of the MOSA!C Lab. He was a Professor with the School of Electrical Engineering, Aalto University, Espoo, Finland, from October 2014 and December 2021. Prior to that, he was working as a Senior Researcher and a 3GPP Standards Expert with NEC Europe Ltd., Heidelberg, Germany. Before joining NEC and till March 2009, he worked as an Assistant Professor with the Graduate School of Information Sciences, Tohoku University in a lab fully funded by KDDI. He worked as a Research Fellow with the Intelligent Cosmos Research Institute, Sendai, from October 2005 to March 2006. His research interests lie in the field o telco cloud, network softwarization and network slicing, AI-based SW-defined security, immersive communications, mobile multimedia streaming, and next generation mobile networking.

JaeSeung Song (Senior Member, IEEE) is a professor in the Department of Computer & Information Security Sejong University. He holds the position of Technical Plenary Vice Chair of the oneM2M global IoT standards initiative. Prior to his current position, he worked for NEC Europe Ltd. and LG Electronics in various positions. He received a Ph.D. at Imperial College London in the Department of Computing, United Kingdom. He holds B.S. and M.S. degrees in computer science from Sogang University. His research interests span the areas of beyond 5G and 6G, AI/ML enabled network systems, software engineering, networked systems and security, with focus on the design and engineering of reliable and intelligent IoT/M2M platforms. He also holds leadership roles in several journals and conferences such as an associate editor of IEEE INTERNET OF THINGS JOURNAL, an IoT series editor of IEEE COMMUNICATIONS STANDARDS MAGAZINE and an IoT track chair of IEEE CONFERENCE ON STANDARDS FOR COMMUNICATIONS AND NETWORKING.