

Article

BoucaNet: A CNN-Transformer for Smoke Recognition on Remote Sensing Satellite Images

Rafik Ghali  and Moulay A. Akhloufi * 

Perception, Robotics and Intelligent Machines (PRIME), Department of Computer Science, Université de Moncton, Moncton, NB E1A 3E9, Canada; rafik.ghali@umoncton.ca

* Correspondence: moulay.akhloufi@umoncton.ca

Abstract: Fire accidents cause alarming damage. They result in the loss of human lives, damage to property, and significant financial losses. Early fire ignition detection systems, particularly smoke detection systems, play a crucial role in enabling effective firefighting efforts. In this paper, a novel DL (Deep Learning) method, namely BoucaNet, is introduced for recognizing smoke on satellite images while addressing the associated challenging limitations. BoucaNet combines the strengths of the deep CNN EfficientNet v2 and the vision transformer EfficientFormer v2 for identifying smoke, cloud, haze, dust, land, and seaside classes. Extensive results demonstrate that BoucaNet achieved high performance, with an accuracy of 93.67%, an F1-score of 93.64%, and an inference time of 0.16 seconds compared with baseline methods. BoucaNet also showed a robust ability to overcome challenges, including complex backgrounds; detecting small smoke zones; handling varying smoke features such as size, shape, and color; and handling visual similarities between smoke, clouds, dust, and haze.

Keywords: smoke recognition; BoucaNet; satellite images; deep learning



Citation: Ghali, R.; Akhloufi, M.A. BoucaNet: A CNN-Transformer for Smoke Recognition on Remote Sensing Satellite Images. *Fire* **2023**, *6*, 455. <https://doi.org/10.3390/fire6120455>

Academic Editor: Grant Williamson

Received: 30 October 2023

Revised: 24 November 2023

Accepted: 28 November 2023

Published: 29 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Fires cause severe damage to economies, properties, ecosystems, and human lives. They destroy properties, homes, and resources, leading to considerable financial losses, and contribute to ecological imbalances. For example, since 1990, wildfires have destroyed an average of 2.5 million hectares per year in Canada [1]. In addition, over the past decade, the cost of firefighting in Canada ranged between \$800 million and \$1.5 billion a year [1]. Since January 2023, 260,000 hectares have already burned in the European Union [2]. Researchers have focused on developing fire ignition and early detection systems to reduce this alarming statistic and improve firefighting capabilities [3,4]. Both smoke and fire detection systems are used to provide comprehensive early warning and fire protection. Fire detection systems are used to detect the presence of flames, while smoke detection systems are adopted to identify the first signs of smoke, even before flames are visible.

Recently, smoke recognition methods made significant progress by exploiting visible features captured by vision sensors [5]. Additionally, classical machine learning methods, such as dynamic texture and optical flow, were employed to manually extract smoke features from images or videos. These extracted features were then used to identify the presence of smoke using various classifiers, such as SVM (Support Vector Machine), Random Forest, and AdaBoost. These approaches showed interesting efficiency, but were related to false alarms and the identification of relevant features that accurately represented the smoke recognition problem [5].

Deep learning models were successfully employed in many fields and industries [6,7]. More specifically, they were used for fire ignition detection due to their ability to learn to automatically extract smoke features from large amounts of data. They provide diverse and informative feature maps, which are often better than manually generated features in terms of performance and robustness [8,9]. More recently, satellite remote sensing images

were adopted for this task, representing a great opportunity thanks to the advantages of satellite remote sensing, including timeliness and large coverage areas [10,11].

High false-alarm rates are still present due to background complexity; the variability of smoke regarding its size, intensity, and shape; and the presence of smoke-like objects, such as haze, dust, and clouds. These objects often have very similar textures, colors, shapes, and spectral features to smoke, leading to false results in detecting smoke. Therefore, this paper presents a novel ensemble learning method, namely BoucaNet, for recognizing smoke on remote sensing satellite images, addressing these challenging limitations. BoucaNet employed a vision transformer, EfficientFormer v2 [12], and a deep CNN (Convolutional Neural Network), EfficientNet v2 [13], to extract smoke features from satellite images. It was trained and evaluated using a satellite dataset, USTC_SmokeRS [14], which comprises six classes (smoke, cloud, haze, dust, seaside, and land). This paper presents three main contributions:

1. A novel DL method, BoucaNet, is introduced to detect the presence of smoke in satellite images, thereby improving the performance of DL-based smoke classification methods.
2. BoucaNet demonstrated a robust ability to handle challenging situations such as background complexity and dynamism; detecting small smoke areas; varying characteristics of smoke regarding its air concentration, flow pattern, intensity, shape, and color; and handling its visual similarity to haze, dust, and clouds. This ability reduces false alarms, making BoucaNet a reliable solution for smoke remote sensing applications with high accuracy.
3. An optimized architecture is proposed in this study, achieving fast inference time, which is an important aspect in developing an early smoke-detection system.

The remainder of this paper is structured as follows: Section 2 presents state-of-the-art methods for smoke recognition using DL approaches. Section 3 introduces the proposed method, BoucaNet, and provides details about the satellite dataset, USTC_SmokeRS. Section 4 reports and discusses the experimental results of BoucaNet. Section 5 concludes the paper.

2. Related Works

Over the years, numerous DL methods were developed to improve the performance of smoke classification in different fields of application, as presented in Table 1. Among them, Tao et al. [15] suggested a simple CNN to recognize smoke in ground images, addressing challenging limitations such as varying smoke colors, shapes, and textures. The proposed CNN is a modified AlexNet [16] by changing the order of the max pooling layers and normalization layers, which follow the first and second convolutional layers. The modified AlexNet was trained and evaluated using the Yuan dataset (5695 smoke images and 18,522 non-smoke images) [17], resulting an accuracy of 96.88%. Yin et al. [18] proposed a new deep normalization CNN, namely DNCNN, to improve smoke detection performance. DNCNN incorporates batch normalization into convolutional layers to deal with overfitting and gradient dispersion. Data augmentation techniques (vertical flipping, rotation, and horizontal flipping) were also used to address the challenges of imbalanced data between smoke and non-smoke images (5695 smoke images and 18,522 non-smoke images [17]). Test results showed that DNCNN achieved an impressive performance with an accuracy of 98.08%, surpassing popular CNNs such as AlexNet, ZF-Net [19], and VGG-16 [20]. Khan et al. [21] studied three CNN models (AlexNet, VGG-16, and GoogleNet [22]) to identify smoke in a normal and foggy IoT environment. Experimental tests were performed using a very large dataset, comprising 18,532 smoke images, 17,474 non-smoke images, 17,474 non-smoke images with fog, and 18,532 smoke images with fog. VGG-16 obtained the higher performance with an accuracy of 97.72% compared with AlexNet, GoogleNet, and published fire models, demonstrating its ability to detect smoke in a foggy environment.

Peng and Wand [23] proposed a video smoke detection method to recognize smoke in complex environments. First, a GMM (Gaussian Mixture Model) [24] was employed as an image processing method to extract the suspected smoke areas from images collected from

surveillance cameras. Then, the SqueezeNet model [25] was adopted to detect the presence of smoke. Using a large dataset (25,000 smoke images and 25,000 non-smoke images), this proposed method showed a high performance with an accuracy of 97.12% and a high prediction time compared with existing wildfire models such as AlexNet, ShuffleNet [26], Xception [27], and MobileNet [28]. Gu et al. [29] developed a DCNN (Deep Dual-Channel Neural Network) as a smoke recognition method. The DCNN is composed of two deep subnet channels, SBNN (Selective-based Batch Normalization Network) and SCNN (Skip Connection-based Neural Network). SBNN comprises six convolutional layers, four normalization layers, three max pooling layers, and three fully connected layers. SCNN includes eleven convolutional layers, seven normalization layers, three max pooling layers, and one global average pooling layer. DCNN was trained on large public learning data [17], comprising 5695 smoke images and 18,522 non-smoke images, and data augmentation techniques (rotation of 90, 180, and 270 degrees). It achieved an accuracy of 99.5%, higher than hand-crafted methods and state-of-the-art DL methods such as DNCNN [18], AlexNet, VGG, GoogLeNet, Xception, ResNet, etc.

Zhang et al. [30] presented a DL method, called DC-CNN (Dual-Channel Convolutional Neural Network), for detecting smoke. DC-CNN is composed of two channels. The first channel employs a pretrained AlexNet in extracting smoke features. The second channel is a simple CNN architecture, consisting of four convolutional layers, a pooling layer, and two fully connected layers for generating more advanced characteristics. Extensive studies were conducted using learning data, including 9794 smoke and 9794 non-smoke images, to handle the challenges related to smoke features, such as transparency properties, homogeneity, and visual similarity to clouds, steam, haze, and fog. DC-CNN obtained the highest accuracy of 99.33% compared with baseline DL models such as LeNet, AlexNet, VGG-16, and DNCNN [18]. Jia et al. [31] designed a new method for detecting smoke in videos. Firstly, GMM-based domain knowledge of smoke was adopted to segment the suspected areas of smoke. Then, three pretrained deep learning models (AlexNet, Inception v3, and ResNet50 [32]) were used to recognize smoke. ResNet50 with GMM performed best, with an F1-score of 99.32% compared with the other models using 138 smoke videos as testing data. He et al. [33] proposed a DL method for smoke detection in a foggy environment. This method combines the VGG-16 method as a backbone to extract smoke features and an attention method, which consists of channel attention and spatial attention to improve the detection of small smoke areas. It was also trained and evaluated using 33,666 images (8342 smoke images, 8522 smoke with fog images, 8401 non-smoke images, and 8401 non-smoke with fog images). It achieved an F1-score of 99.97%, outperforming the AlexNet, VGG-16, and SqueezeNet methods.

Zhang et al. [34] developed an end-to-end CNN method to identify smoke. Two CNNs (spatial stream and temporal stream), each comprising five convolutional layers, three max pooling layers, and an attention module to suppress noise, and which extract salient features from temporal and spatial feature maps and improve detection performance, were adopted to extract the spatial and temporal features of smoke. This method achieved an accuracy of 96.8%, better than state-of-the-art methods using 116 fire videos and 89 non-fire videos. Cheng et al. [35] presented a deep convolutional network, namely PACNN, to improve the robustness of smoke recognition tasks. PACNN is a deep CNN with a PAAModule (Pixel Aware Attention Module), which integrates into the residual structure via element-wise addition and skip connection on two feature maps. Testing results showed that PACNN reached a high accuracy of 98.91% compared with popular CNNs (AlexNet, Inception v4, ResNet34, SEResNet34, DenseNet-121, and DNCNN) and vision transformers (ViT, Swin-T, and DeiT-Ti) using the Yuan dataset.

Tao and Duan [36] introduced a video smoke recognition method, AFSNet, to address slow-moving smoke challenges. AFSNet is composed of three main modules: AFSM (Adaptive Frame Selection Module) for extracting multi-scale spatial and spatiotemporal features; FEM (Feature Extraction Module) for incorporating a context attention module, an enhanced dilated convolution module, and a spatiotemporal feature attention module to minimize the loss of detailed information; and RM (Recognition Module) for detecting

smoke presence. AFSNet was trained on two large datasets, SRSet (14,100 smoke images and 15,380 non-smoke images) and RISE (12,567 videos). It achieved impressive F1-scores of 96.57% and 91.00% using the SRSet and RISE datasets, respectively, surpassing classical machine learning methods and existing deep learning models. Cheng et al. [37] proposed a novel vision transformer, called CViTNet (Convolution-enhanced Vision Transformer Network), for identifying smoke. CViTNet consists of three stages (s1, s2, and s3). The first stage, s1, comprises a convolutional stem and a ViT transformer encoder. Each of the s1 and s2 stages includes a ViT transformer encoder [38] and a convolutional token embedding, which was proposed to improve the multiscale feature representation of tokenization. Using the Yuan dataset, CViTNet achieved a high accuracy of 99.20% compared with existing CNNs (AlexNet, ResNet, SEResNet, DenseNet, DNCNN, etc.) and vision transformer methods (ViT-B, DeiT-S, conViT-Ti, Swin-T, etc.) [37].

In the study conducted by Mohammed [39], a pretrained InceptionResNet v2 model [40] was employed for the detection of forest smoke and fires. Mohammed utilized a dataset comprising aerial and ground images (1102 fire images and 1102 smoke images). Data augmentation methods, including scaling and horizontal/vertical flipping, were applied during the training phase. Testing results showed that InceptionResNet v2 achieved an impressive accuracy of 99.09%. Chen et al. [41] studied the effectiveness of five DL methods (LeNet5, VGG-16, ResNet18, MobileNet v2 [42], and Xception) for wildland smoke/fire recognition on aerial images. These models were trained using a large dataset comprising a total of 53,451 images, which were divided into three categories: 25,434 fire/smoke images, 14,317 fire/no-smoke images, and 13,700 no-fire/no-smoke images. VGG-16 obtained an accuracy of 99.91%, surpassing MobileNet v2, ResNet18, LeNet5, Xception, and a traditional machine learning method (Logistic Regression) by 0.56%, 1.52%, 4.58%, 5.35%, and 9.54%.

Dilshad et al. [43] proposed a fire detection model, E-FireNet, to recognize fires in a surveillance environment. E-FireNet is a modified VGG-16 by deleting block 5 and adjusting the convolutional layers of block 4. The experimental setup was performed using data augmentation techniques (horizontal flipping, rotation, and scaling). E-FireNet achieved an accuracy 98% better than that of the pretrained MobileNet v1, VGG-19, EfficientNet-B0, VGG-16, and NASNetMobile v1 models using the SV-Fire dataset (1500 images) [43]. Yar et al. [44] developed a modified YOLO v5 method for detecting and locating fires in smart cities. A total of 1957 images, comprising indoor fires (118 images), building fires (723 images), and vehicle fires (1116 images), were used to train and evaluate the proposed model, achieving an F1-score of 84%.

Priya and Vani [45] introduced a CNN based on Inception v3 architecture [46] for the recognition of forest smoke/fires using satellite images. Their study utilized a dataset consisting of 534 satellite images, with 239 fire images and 295 no-fire images, for both training and testing purposes. Their proposed method achieved an accuracy of 98%. Ba et al. [14] also proposed a DL method, namely SmokeNet, to address the challenge of recognizing smoke on satellite data, including varying smoke features such as colors, shapes, and spectral overlaps. SmokeNet is a CNN model with channel-wise and spatial attention. A novel satellite dataset, namely USTC_SmokeRS, comprising 6225 satellite images divided into six classes (smoke, cloud, haze, dust, seaside, and land), was used in the training and testing phases. SmokeNet showed high performance with an accuracy of 92.75%.

As described in Table 1, deep learning methods performed better in recognizing smoke. However, several challenging limitations persist, including the complexity and dynamics of the background; the visual similarity between smoke, clouds, dust, and haze; the varying characteristics of smoke regarding its air concentration, flow pattern, and color; and detecting small smoke zones.

Table 1. Deep learning models for smoke recognition.

Ref.	Methodology	Object Detected	Dataset	Image Type	Results (%)
[41]	VGG-16	Smoke/Flame	FLAME2: 53,451 images	Aerial	Accuracy = 99.91
[39]	InceptionResNet v2	Smoke/Flame	Private: 1102 fire images and 1102 smoke images	Aerial Ground	Accuracy = 99.90
[15]	Modified AlexNet	Smoke	Yuan dataset: 5695 smoke images and 18,522 non-smoke images	Ground	Accuracy = 96.88
[18]	DNCNN	Smoke	Yuan dataset: 5695 smoke images and 18,522 non-smoke images	Ground	Accuracy = 98.08
[21]	VGG-16	Smoke	Private: 18,532 smoke images, 17,474 non-smoke images, 17,474 non-smoke images with fog, and 18,532 smoke images with fog	Ground	Accuracy = 97.72
[23]	GMM and SqueezeNet	Smoke	Private: 25,000 smoke images and 25,000 non-smoke images	Ground	Accuracy = 97.12
[29]	DCNN	Smoke	Yuan dataset: 5695 smoke images and 18,522 non-smoke images	Ground	Accuracy = 99.50
[30]	DC-CNN	Smoke	Private: 9794 smoke and 9794 non-smoke images	Ground	Accuracy = 99.33
[31]	GMM and ResNet50	Smoke	VisiFire: 138 smoke video and PascalVoc2012: 17,708 images	Ground	F1-score = 99.32
[33]	VGG-16 and attention module	Smoke	Private: 33,666 images (560 videos): 8342 smoke images, 8522 smoke with fog images, 8401 non-smoke images, and 8401 non-smoke with fog images	Ground	F1-score = 99.97
[34]	CNN with attention	Smoke	Private: 116 fire videos and 89 non-fire videos	Ground	Accuracy = 96.80
[35]	PACNN	Smoke	Yuan dataset: 5695 smoke images and 18,522 non-smoke images	Ground	Accuracy = 98.91
[36]	AFSNet	Smoke	RSet: 29,480 images (14,100 smoke images and 15,380 non-smoke images)	Ground	F1-score = 96.57
[37]	CViTNet	Smoke	RISE: 12,567 videos	Ground	F1-score = 91.00
[43]	E-FireNet	Flame	Yuan dataset: 5695 smoke images and 18,522 non-smoke images	Ground	Accuracy = 99.20
[44]	Modified YOLO v5	Flame	SV-Fire dataset: 1500 images	Ground	Accuracy = 98.00
[45]	CNN based on Inception v3	Smoke/Flame	Private: 723 building fire images, 118 indoor electric fire images, and 1116 vehicle fire images	Ground	F1-score = 84.00
[14]	SmokeNet	Smoke	Private: 534 images (239 fire images and 295 no-fire images)	Satellite	Accuracy = 98.00
			USTC_SmokeRS: 6225 satellite images	Satellite	Accuracy = 92.75

3. Materials and Methods

In this section, the proposed DL method, BoucaNet, designed for the recognition of smoke using satellite images, is introduced. Subsequently, an overview of the dataset employed to train and test the BoucaNet model is provided. Finally, the evaluation metrics (F1-score, accuracy, and inference time) used in this paper are presented.

3.1. Proposed Method for Smoke Classification

In this paper, a new ensemble learning approach, namely BoucaNet, is introduced for recognizing smoke in satellite images and for addressing challenging limitations, including background complexity and dynamics due to the presence of dynamically changing backgrounds in input satellite images; visual similarities of smoke with clouds, dust, and haze; and varying features of smoke regarding its shape, form, color, flow pattern, and texture. BoucaNet combines the deep CNN EfficientNet v2 (EfficientNetV2M) [13] and the vision transformer EfficientFormer v2 (EfficientFormerV2L) [12]. EfficientNet v2 [13] is a new family of CNN. It is proposed to address the training limitation of EfficientNet models [47], showing a better parameter efficiency and faster learning speed compared with these models. It adopts an improved progressive learning method, which adaptively adjusts regularization techniques such as data augmentation techniques and dropout methods along with input image size. EfficientNet v2 achieves a high performance with top-1 accuracy of 87.3% using ImageNet21K dataset [48], surpassing the popular vision transformers (ViT, DeiT, and T2T-ViT) and existing CNNs (EfficientNet, RegNetY, ResNetSt, NFNet, BotNet, etc.) [13]. EfficientFormer v2 was developed by Li et al. [12] to improve the size and latency of vision transformers while maintaining high performance. This model is

an updated version of the EfficientFormer model, integrating a fine-grained joint search method, which optimizes the speed and size of the model, simultaneously. Using the ImageNet-1K dataset [49] as the learning data, it achieves an impressive top-1 accuracy of 83.5% and a low latency of 0.9 ms on iPhone 12 (iOS 16), outperforming existing competitive CNN methods (MobileNet v2, EfficientNet, ResNet, etc.) and vision transformer models (Mobile ViT, EdgeViT, LeViT, DeiT, T2T-ViT, Swin-Tiny, CSwin, etc.) [12].

To employ EfficientNet v2 and EfficientFormer v2 models in the specific task of smoke recognition, their classification layers (last layers), originally developed for different classification tasks, are removed. As depicted in Figure 1, the preprocessing steps start with resizing the input satellite images to 224×224 pixels. Next, four data augmentation techniques, including rotation, shearing, shifting, and zooming, are utilized to diversify learning data, improve the potential of BoucaNet to generalize different real-world scenarios, and avoid overfitting. Then, the input satellite images and the generated images are simultaneously fed into the EfficientNet v2 and EfficientFormer v2 models to extract complex contextual features, comprising both smoke plume patterns and background contextual information, and provide a comprehensive representation of various smoke scenarios. After concatenating the two feature maps generated by the EfficientNet v2 and EfficientFormer v2 models, the Gaussian dropout regularization technique with a rate of 0.3 is employed. This method adds random noise from a Gaussian distribution to the input satellite data, improving BoucaNet's generalization ability and avoiding overfitting. Finally, a Softmax function generates a probability score ranging from 0 to 1, determining the appropriate class, such as smoke, cloud, haze, dust, seaside, or land, for the input satellite images.

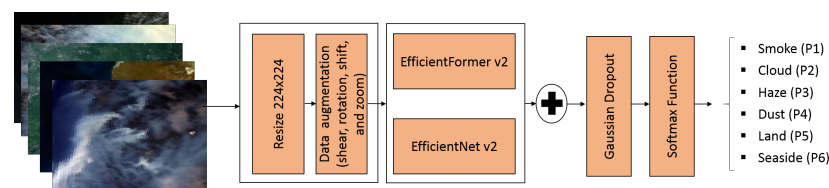


Figure 1. The proposed architecture of BoucaNet. P1, P2, P3, P4, P5, and P6 correspond to the predicted probabilities of the input image belonging to the smoke, cloud, haze, dust, land, or seaside class.

3.2. Datasets

Many large fire datasets are made available to help researchers in benchmarking and comparing DL techniques dealing with the same problem. However, this is not the case for smoke recognition problems, especially when using satellite data, thus making the evaluation of these DL methods a little challenging.

To train and test the proposed smoke recognition method, BoucaNet, the available satellite data, USTC_SmokeRS [14], is utilized. This dataset is collected using MODIS (Moderate Resolution Imaging Spectroradiometer) and represents numerous smoke scenes through satellite remote sensing. It is selected from a remote sensing platform in Hefei, China, and the Level-1 and Atmosphere Archive & Distribution System (LAADS) Distributed Active Archive Center (DAAC) situated at the Goddard Space Flight Center in Greenbelt, Maryland, USA. The USTC_SmokeRS dataset comprises a total of 6225 satellite images with dimensions of 256×256 pixels and a spatial resolution of 1 km. It comprises six classes:

- Smoke (1016 satellite images) as the target class for wildfire detection.
- Dust (1009 satellite images) and haze (1002 satellite images) as negative classes to smoke, which share similar features (texture and spectral) with smoke.
- Cloud (1164 satellite images) as the most common class in satellite images, with similar color, shape, and spectral characteristics to smoke.
- Land (1027 satellite images) and seaside (1007 satellite images) as background classes for fire smoke scenes.

Figures 2 and 3 depict a USTC_SmokeRS dataset example.

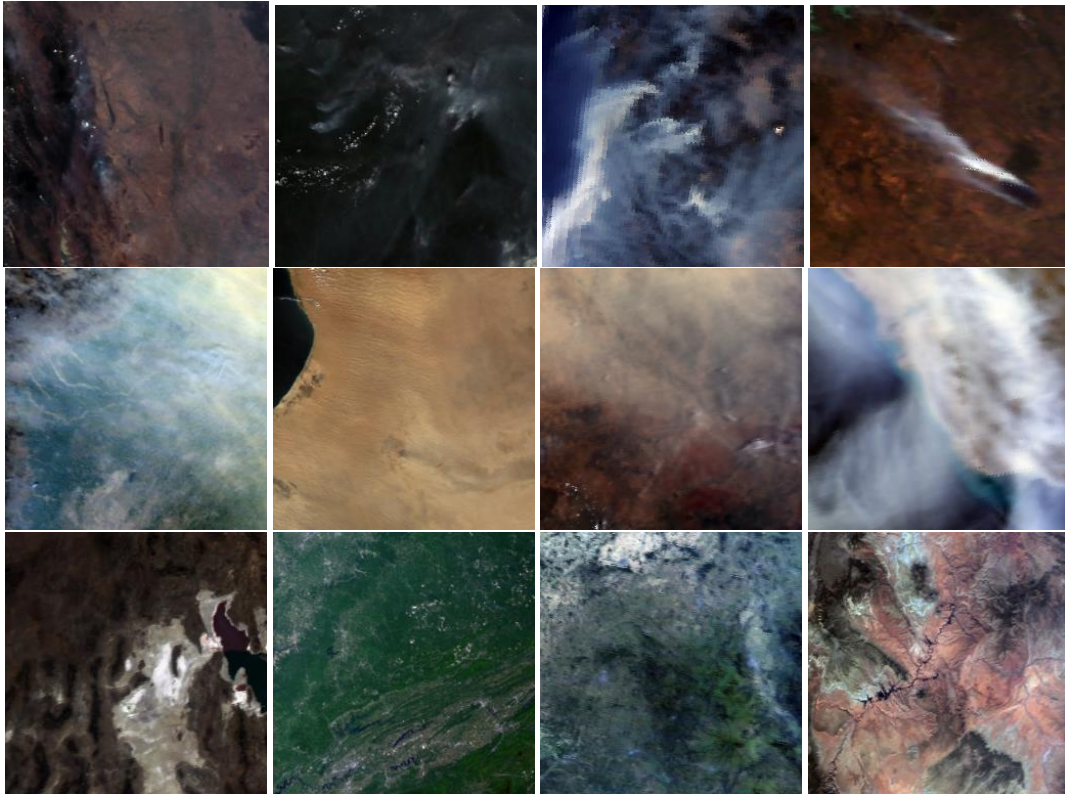


Figure 2. USTC_SmokeRS dataset example from top to bottom: smoke images, dust images, and land images.

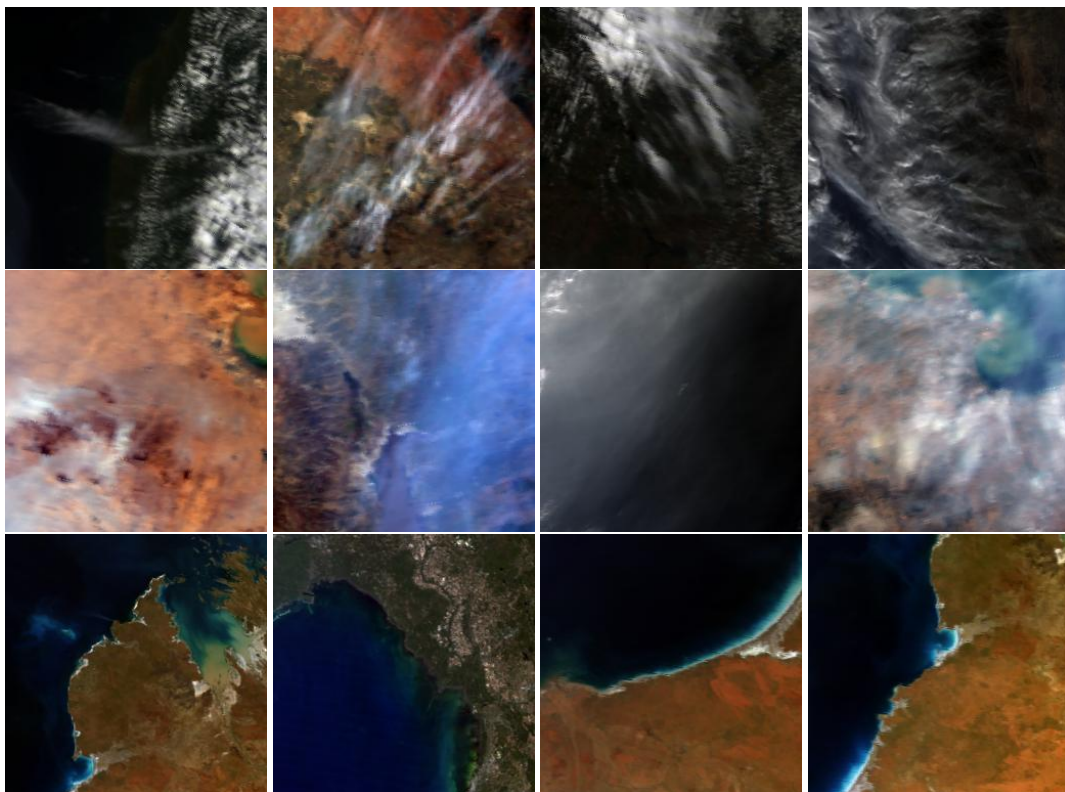


Figure 3. USTC_SmokeRS dataset example from top to bottom: cloud images, haze images, and seaside images.

3.3. Evaluation Metrics

In this work, three metrics (accuracy, F1-score, and inference time) are used to evaluate the proposed ensemble learning approach, BoucaNet. The accuracy and F1-score metrics are determined using the true positive rate (TP), false positive rate (FP), true negative rate (TN), and false negative rate (FN).

- Accuracy is the proportion of accurate predictions relative to the total number of predictions, as shown in Equation (1).

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

- F1-score integrates precision and recall metrics to calculate the performance of the proposed model, as presented in Equation (2).

$$F1 - score = \frac{2 * TP}{2 * TP + FN + FP} \quad (2)$$

- The inference time is the average time taken by BoucaNet to identify and recognize the presence of smoke in an input satellite image during the test step.

4. Results and Discussion

The proposed DL model, BoucaNet, was developed using Python and TensorFlow version 2.11 [50]. For training and testing this model, a machine equipped with an NVIDIA GeForce RTX 2080Ti GPU, an Intel(R) Xeon(R) CPU (E5-2620 v4), and 64GB of RAM was utilized.

BoucaNet was trained using the USTC_SmokeRS satellite dataset. This dataset allowed BoucaNet to learn on various classes and scenarios, thereby enabling it to learn and recognize various aspects of smoke in satellite images. It comprises a total of 6225 satellite images, divided into six distinct classes. These images were split into three sets as shown in Table 2:

- Training set: a total of 4181 images were used, including 782, 678, 673, 690, 676, and 682 satellite images for the cloud, dust, haze, land, seaside, and smoke classes, respectively.
- Validation set: a total of 796 images were utilized, including 149 images for cloud, 129 images for dust, 128 images for haze, 131 images for land, 129 images for seaside, and 130 images for smoke.
- Testing set: a total of 1248 images were selected for evaluation. This test set is composed of 233 images for cloud, 202 images for dust, 201 images haze, 206 images for land, 202 images for seaside, and 204 images for smoke.

Table 2. Dataset subsets.

Data	Cloud	Dust	Haze	Land	Seaside	Smoke	Total
Training set	782	678	673	690	676	682	4181
Validation set	149	129	128	131	129	130	796
Testing set	233	202	201	206	202	204	1248

During the training process, various hyperparameters were selected to optimize the learning of BoucaNet, including a learning rate of 0.001, the Adam optimizer, a total of 150 training epochs, and a batch size of eight. Additionally, the categorical cross-entropy loss function (see Equation (3)) was employed.

$$Cross - entropy = - \sum_{i=1}^A z_i \log(p) \quad (3)$$

where z is the binary indicator, A is the number of classes (six classes, including smoke, cloud, haze, dust, land, and seaside), and p is the predicted probability.

The experimental setup utilized input satellite images with a size of 224×224 pixels. To improve BoucaNet's performance and avoid overfitting, four data augmentation techniques, such as shear, rotation, shift, and zoom, were employed, enabling BoucaNet to handle a wide range of real-world scenarios. Additionally, the GPU was used to facilitate model training and calculate the inference time.

The evaluation of BoucaNet includes several key aspects. Firstly, its performance was analyzed in terms of F1-score, accuracy, and inference time with the method, namely CT-Fire, which combines EfficientFormer v2 [14] and RegNetY [51] models as the backbone, RegNetY-16GF [51], the vision transformer EfficientFormer v2 [12], and SmokeNet [14] as the state-of-the-art smoke detection method. Next, the obtained F1-scores of these models for each class, namely smoke, cloud, dust, haze, land, and seaside, were presented. Then, the resulting confusion matrix generated by BoucaNet was illustrated and discussed. Finally, visual results of the input images predicted by these models were presented.

Testing results (loss, F1-score, accuracy, and inference time) of the proposed BoucaNet, CT-Fire, RegNetY-16GF, and EfficientFormer v2 are reported in Table 3. RegNetY-16GF and EfficientFormer v2 were selected due to their excellent performance in classifying objects. CT-Fire is an ensemble learning method, which combines EfficientFormer v2 and RegNetY-16GF to extract features. Then, the Gaussian drop regularization method and the softmax function were used to recognize the presence of smoke. BoucaNet showed a high performance during testing, achieving a loss of 0.2184, an accuracy of 93.67%, and an F1-score of 93.64%. This performance was obtained thanks to the diversity of feature maps extracted by EfficientNet v2 and EfficientFormer v2 models, including details, complexity, and local and global feature (colors, shapes, textures, etc.) for the smoke, cloud, haze, seaside, land, and dust classes, thus enabling BoucaNet to distinguish between smoke and complex backgrounds and identify small areas of smoke. In terms of F1-score, BoucaNet outperformed CT-Fire, RegNetY-16GF, and EfficientFormer by 2.75%, 1.38%, and 1.50%, respectively. This proposed model also performed better than the state-of-the-art method SmokeNet, which achieved an accuracy of 92.75% using the USTC_SmokeRS dataset [14]. It demonstrated its potential to address and overcome challenging limitations related to recognizing smoke in satellite images. These challenges include complex backgrounds, comprising various land covers and geographical features, which can make it difficult to accurately identify smoke in input satellite images. Additionally, BoucaNet handled the varying and dynamic nature of smoke in terms of its shape, color, intensity, and flow pattern features, as well as the visual similarities of smoke, including color, shape, and spectral characteristics, which are often shared with clouds, dust, and haze. On the other hand, BoucaNet achieved an efficient processing speed with an inference time of 0.16 seconds, slightly surpassing the inference times of EfficientFormer v2, CT-Fire, and RegNetY-16GF. This inference time showed BoucaNet's suitability for real-time processing of satellite images for smoke recognition while maintaining high performance.

Table 3. Comparative analysis of BoucaNet and other models on USTC_SmokeRS dataset.

Models	Loss	Accuracy (%)	F1-Score (%)	Inference Time (s)
CT-Fire	0.2611	90.95	90.89	0.10
RegNetY-16GF	0.2668	92.31	92.26	0.04
EfficientFormer v2	0.2643	92.23	92.14	0.07
SmokeNet [14]	–	92.75	–	–
BoucaNet	0.2184	93.67	93.64	0.16

Table 4 illustrates the comparative analysis of BoucaNet, RegNetY-16GF, CT-Fire, and EfficientFormer v2 for recognizing smoke, cloud, haze, dust, land, and seaside classes. BoucaNet achieved superior results with an F1-score of 95.58%, 91.00%, 90.82%, 95.01%, 98.76%, and 90.36% for recognizing cloud, dust, haze, land, seaside, and smoke classes, respectively, compared with CT-Fire, RegNetY-16GF, and EfficientFormer v2. It demonstrated its ability

to accurately differentiate between cloud, smoke, haze, dust, land, and seaside features, thereby proving its capability to overcome challenges related to background complexity and visual similarities, including color, shape, and spectral characteristics, between smoke and other classes (cloud, dust, and haze).

Table 4. Comparative analysis of BoucaNet and other DL methods for smoke, cloud, haze, dust, land, and seaside classes.

Models	Cloud	Dust	F1-Score (%)			
			Haze	Land	Seaside	Smoke
CT-Fire	94.14	86.49	86.89	91.86	96.53	88.94
RegNetY- 16GF	95.34	88.61	87.32	95.47	97.80	88.84
EfficientFormer v2	94.58	88.56	87.03	94.76	97.80	89.66
BoucaNet	95.58	91.00	90.82	95.01	98.76	90.36

Figure 4 depicts a confusion matrix of BoucaNet for the six classes (smoke, dust, cloud, haze, seaside, and land) using the testing set. The results obtained provide a comprehensive view of BoucaNet's performance in recognizing these classes and overcoming challenges. BoucaNet performed well in distinguishing between features of the smoke (178 instances), cloud (227 instances), dust (187 instances), haze (178 instances), land (200 instances), and seaside (199 instances) classes. These results demonstrate the robustness of BoucaNet in identifying smoke in varying environmental conditions and complex backgrounds, despite the overlap in visual features between smoke, clouds, dust, and haze. However, it misclassified a small number of smoke instances as clouds (eight instances), dust (six instances), haze (five instances), land (six instances), and seaside (one instance). These misclassifications can be attributed to the complex nature of smoke, which shares visual characteristics (color, shape, spectral texture, etc.) with other classes.

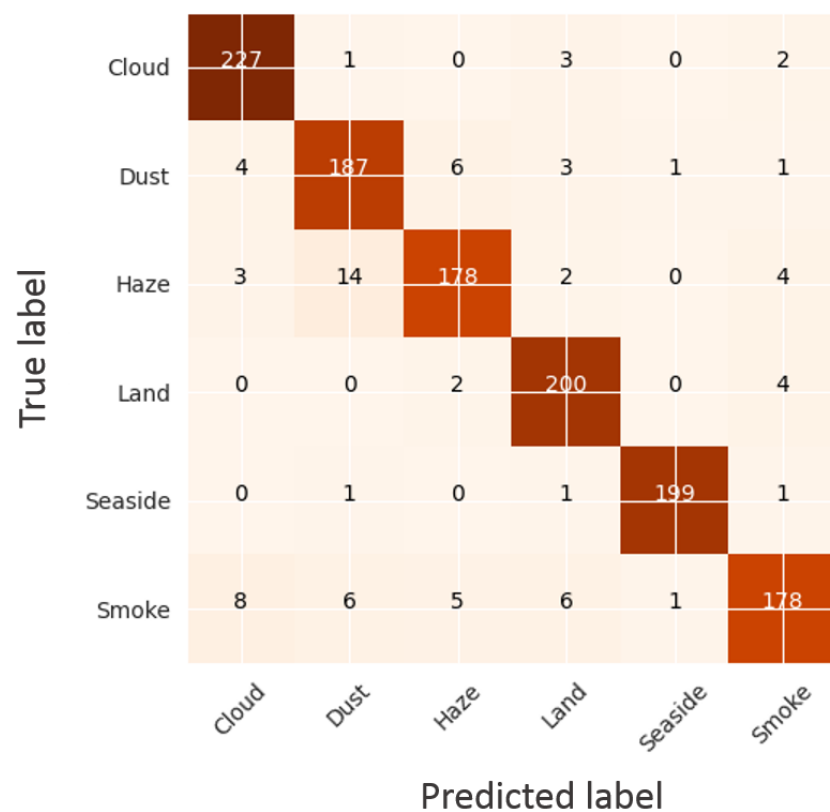


Figure 4. Confusion matrix of BoucaNet on USTC_SmokeRS data test set.

Similar to its quantitative performance, BoucaNet performed well in predicting and identifying the presence of smoke, clouds, dust, haze, land, and seaside in input satellite images with high confidence scores (see Figures 5–10). For instance, it correctly predicted a smoke image as smoke with a confidence score of 0.99 (as shown in Figure 5c), a cloud instance as cloud with a confidence score of 0.98 (as depicted in Figure 6a), a dust instance as dust with a confidence score of 0.88 (see Figure 7c), and a haze instance as haze with a confidence score of 0.99 (as shown in Figure 8c). Additionally, CT-Fire made incorrect predictions, such as classifying clouds as dust with a confidence score of 0.99 (see Figure 6b) and haze as land with a confidence score of 0.93 (as shown in Figure 8b). RegNetY-16GF also misclassified haze as land with a confidence score of 0.63 (see Figure 8b). EfficientFormer also performed poorly in detecting land as haze with a confidence score of 0.94 (as depicted in Figure 9b).

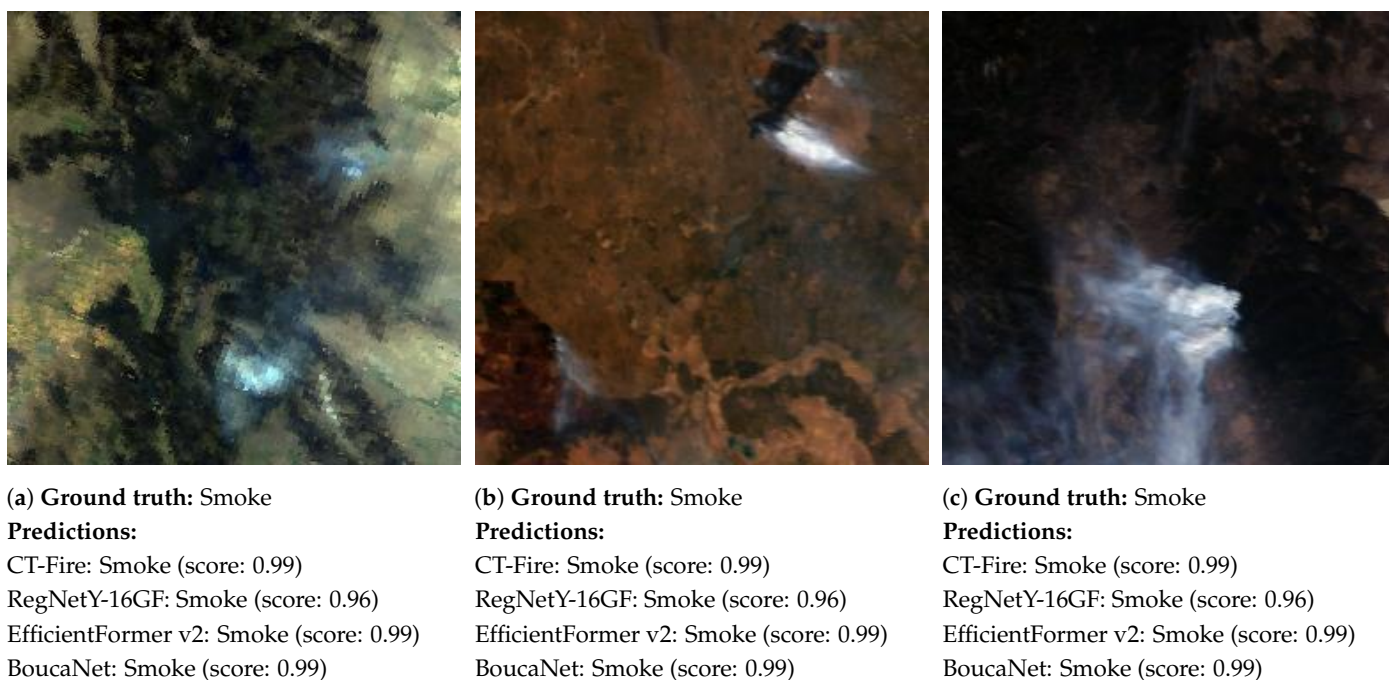
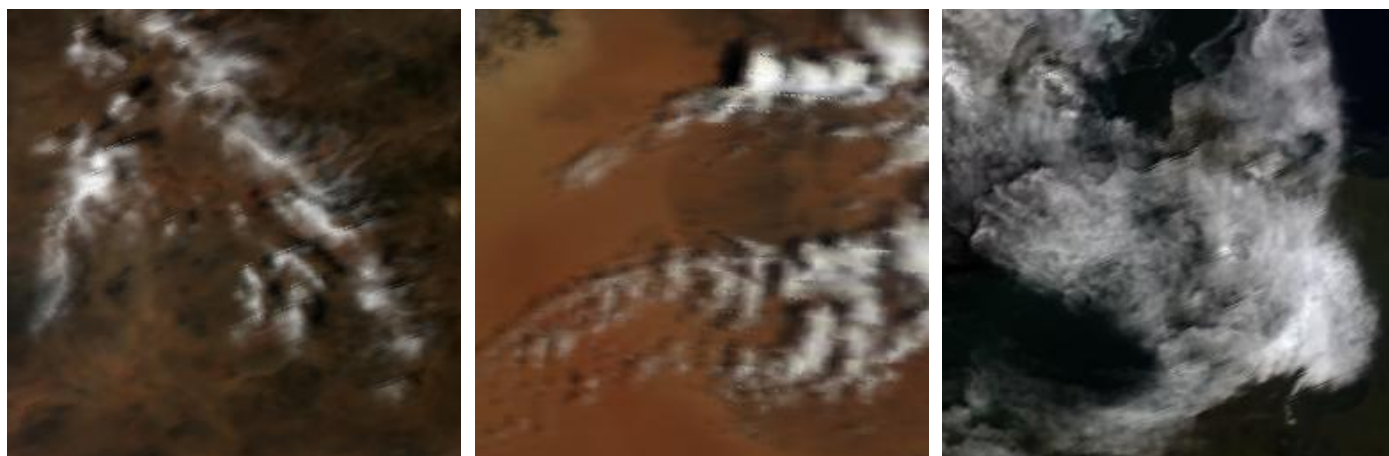


Figure 5. Smoke classification results of the proposed models.

In conclusion, BoucaNet performed well in recognizing smoke in satellite images compared with baseline models (EfficientFormer v2, RegNetY-16GF, CT-Fire, and SmokeNet). Notably, it demonstrated its potential to address challenging limitations, including complex backgrounds; the dynamic nature of smoke in terms of its shape, intensity, and color; detecting small areas of smoke; and distinguishing visual similarities in terms of color, shape, and spectral characteristics between smoke and other elements, including clouds, dust, and haze. Additionally, BoucaNet achieved an interesting inference time.

**(a) Ground truth:** Cloud**Predictions:**

CT-Fire: Cloud (score: 0.97)

RegNetY-16GF: Cloud (score: 0.98)

EfficientFormer v2: Cloud (score: 0.97)

BoucaNet: Cloud (score: 0.98)

(b) Ground truth: Cloud**Predictions:**

CT-Fire: Dust (score: 0.99)

RegNetY-16GF: Cloud (score: 0.93)

EfficientFormer v2: Cloud (score: 0.95)

BoucaNet: Cloud (score: 0.96)

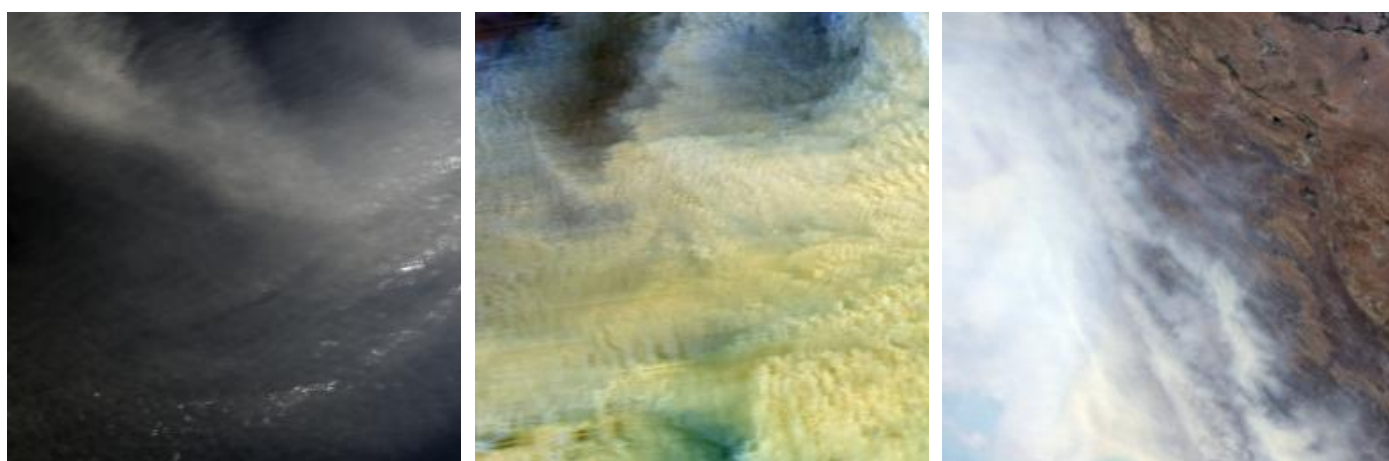
(c) Ground truth: Cloud**Predictions:**

CT-Fire: Cloud (score: 0.99)

RegNetY-16GF: Cloud (score: 0.98)

EfficientFormer v2: Cloud (score: 0.99)

BoucaNet: Cloud (score: 0.99)

Figure 6. Cloud classification results of the proposed models.**(a) Ground truth:** Dust**Predictions:**

CT-Fire: Dust (score: 0.96)

RegNetY-16GF: Dust (score: 0.90)

EfficientFormer v2: Dust (score: 0.88)

BoucaNet: Dust (score: 0.87)

(b) Ground truth: Dust**Predictions:**

CT-Fire: Dust (score: 0.81)

RegNetY-16GF: Dust (score: 0.93)

EfficientFormer v2: Dust (score: 0.99)

BoucaNet: Dust (score: 0.99)

(c) Ground truth: Dust**Predictions:**

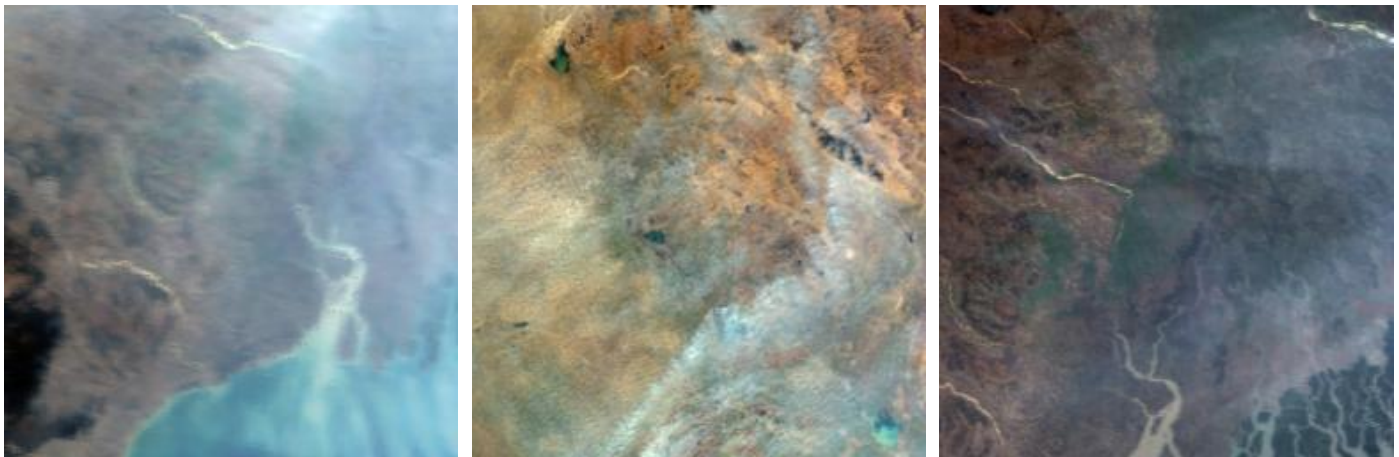
CT-Fire: Dust (score: 0.92)

RegNetY-16GF: Dust (score: 0.82)

EfficientFormer v2: Dust (score: 0.84)

BoucaNet: Dust (score: 0.88)

Figure 7. Dust classification results of the proposed models.

**(a) Ground truth: Haze****Predictions:**

CT-Fire: Haze (score: 0.99)

RegNetY-16GF: Haze (score: 0.92)

EfficientFormer v2: Haze (score: 0.52)

BoucaNet: Haze (score: 0.99)

(b) Ground truth: Haze**Predictions:**

CT-Fire: Land (score: 0.93)

RegNetY-16GF: Land (score: 0.63)

EfficientFormer v2: Haze (score: 0.50)

BoucaNet: Haze (score: 0.88)

(c) Ground truth: Haze**Predictions:**

CT-Fire: Haze (score: 0.99)

RegNetY-16GF: Haze (score: 0.94)

EfficientFormer v2: Haze (score: 0.97)

BoucaNet: Haze (score: 0.99)

Figure 8. Haze classification results of the proposed models.**(a) Ground truth: Land****Predictions:**

CT-Fire: Land (score: 0.99)

RegNetY-16GF: Land (score: 0.98)

EfficientFormer v2: Land (score: 0.97)

BoucaNet: Land (score: 0.99)

(b) Ground truth: Land**Predictions:**

CT-Fire: Land (score: 0.86)

RegNetY-16GF: Land (score: 0.95)

EfficientFormer v2: Haze (score: 0.94)

BoucaNet: Land (score: 0.55)

(c) Ground truth: Land**Predictions:**

CT-Fire: Land (score: 0.99)

RegNetY-16GF: Land (score: 0.97)

EfficientFormer v2: Land (score: 0.99)

BoucaNet: Land (score: 0.99)

Figure 9. Land classification results of the proposed models.

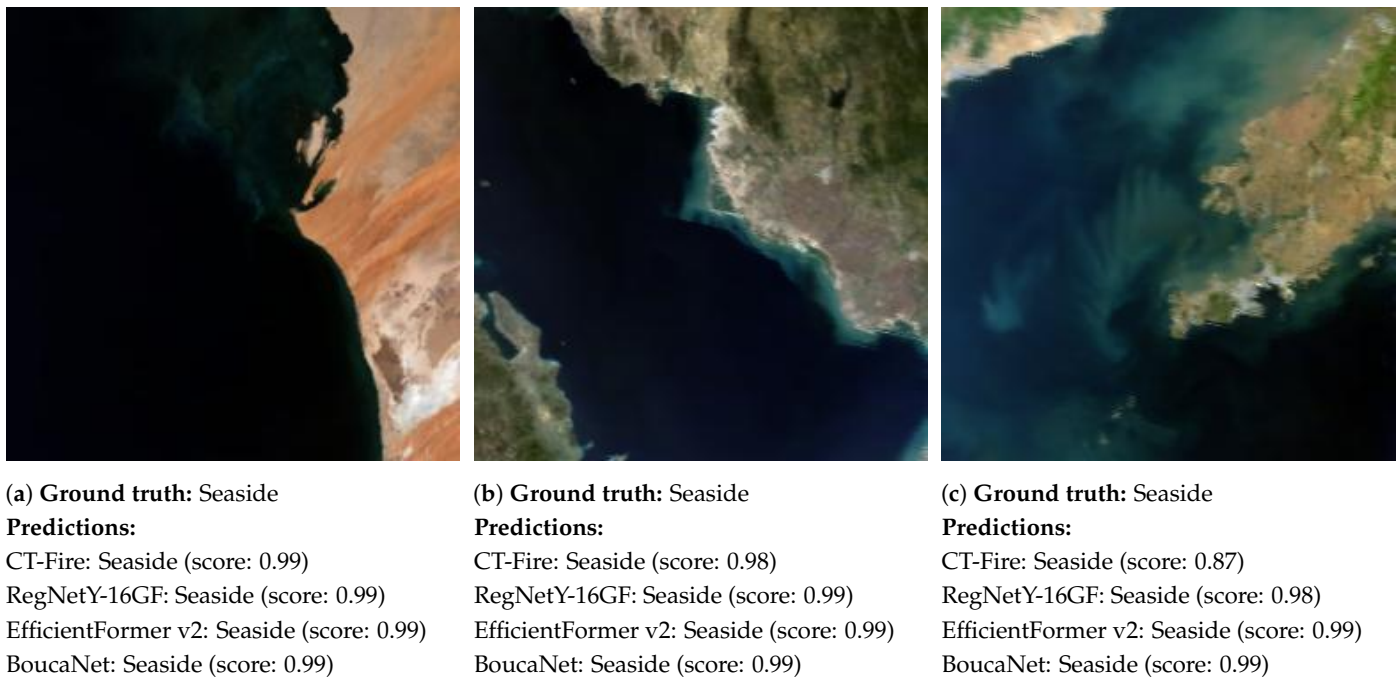


Figure 10. Seaside classification results of the proposed models.

5. Conclusions

In this paper, a novel ensemble learning method, namely BoucaNet, was presented for recognizing smoke in satellite images while addressing the associated challenges. BoucaNet combines the strengths of EfficientNet v2 and EfficientFormer v2 to extract rich and diverse feature maps for smoke, cloud, haze, dust, land, and seaside classes. It demonstrated a high performance, with an accuracy of 93.67% and an F1-score of 93.64%, using the USTC_SmokeRS dataset, which consists of 6225 satellite images. Furthermore, BoucaNet outperformed existing deep learning models for object classification, specifically EfficientFormer v2 and RegNetY-16GF, as well as state-of-the-art methods, including SmokeNet. It also showed an interesting processing speed, with an inference time of 0.16 s. Additionally, BoucaNet demonstrated its potential as a robust solution to the challenges of recognizing smoke in satellite images, including complex backgrounds; the dynamic nature of smoke, which can present variations in shape, intensity, and color; detecting small areas of smoke; and visual similarities between smoke and other elements, such as clouds, dust, and haze.

As future work, the evaluation of BoucaNet is planned for detecting smoke and fires using large scale satellite and/or aerial images in both forest and urban environments.

Author Contributions: Conceptualization, M.A.A. and R.G.; methodology, R.G. and M.A.A.; validation, R.G. and M.A.A.; formal analysis, R.G. and M.A.A.; writing—original draft preparation, R.G.; writing—review and editing, M.A.A.; funding acquisition, M.A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was enabled in part by support provided by the Natural Sciences and Engineering Research Council of Canada (NSERC), funding reference number RGPIN-2018-06233.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: This work uses a publicly available dataset; see reference [14] for data availability. More details about this dataset are available in Section 3.2.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

DL	Deep Learning.
SVM	Support Vector Machine.
CNN	Convolutional Neural Network.
GMM	Gaussian Mixture Model.
RM	Recognition Module.
AFSM	Adaptative Frame Selection Module.
FEM	Feature Extraction Module.
SBNN	Selective-based Batch Normalization Network.
SCNN	Skip Connection-based Neural Network.
CViTNet	Convolution-enhanced Vision Transformer Network.
DC-CNN	Dual-Channel Convolutional Neural Network.
PAAModule	Pixel Aware Attention Module.
MODIS	Moderate Resolution Imaging Spectroradiometer.

References

1. Government of Canada. Forest Fires. 2023. Available online: <https://natural-resources.canada.ca/our-natural-resources/forests/wildland-fires-insects-disturbances/forest-fires/13143> (accessed 30 September 2023).
2. European Commission. Wildfires in the Mediterranean. 2023. Available online: https://joint-research-centre.ec.europa.eu/jrc-news-and-updates/wildfires-mediterranean-monitoring-impact-helping-response-2023-07-28_en (accessed on 30 September 2023).
3. Ghali, R.; Akhloufi, M.A. Wildfires Detection and Segmentation Using Deep CNNs and Vision Transformers. In Proceedings of the Pattern Recognition, Computer Vision, and Image Processing, ICPR 2022 International Workshops and Challenges, Montreal, QC, Canada, 21–25 August 2022; pp. 222–232.
4. Ghali, R.; Akhloufi, M.A. Deep Learning Approaches for Wildland Fires Remote Sensing: Classification, Detection, and Segmentation. *Remote Sens.* **2023**, *15*, 1821. [\[CrossRef\]](#)
5. Chaturvedi, S.; Khanna, P.; Ojha, A. A Survey on Vision-based Outdoor Smoke Detection Techniques for Environmental Safety. *ISPRS J. Photogramm. Remote Sens.* **2022**, *185*, 158–187. [\[CrossRef\]](#)
6. Madhavi, K.R.; Kora, P.; Reddy, L.V.; Avanija, J.; Soujanya, K.; Telagarapu, P. Cardiac Arrhythmia Detection Using Dual-tree Wavelet Transform and Convolutional Neural Network. *Soft Comput.* **2022**, *26*, 3561–3571. [\[CrossRef\]](#)
7. Skandha, S.; Saba, L.; Gupta, S.K.; Kumar, V.K.; Johri, A.M.; Khanna, N.N.; Mavrogeni, S.; Laird, J.R.; Pareek, G.; Sfrikakis, P.P.; et al. Magnetic Resonance Based Wilson’s Disease Tissue Characterization in an Artificial Intelligence Framework Using Transfer Learning. In *Multimodality Imaging, Volume 1*; IOP Publishing: Bristol, UK, 2022; pp. 5–22. [\[CrossRef\]](#)
8. Ghali, R.; Akhloufi, M.A.; Jmal, M.; Mseddi, W.S.; Attia, R. Forest Fires Segmentation using Deep Convolutional Neural Networks. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC), Melbourne, QSD, Australia, 17–20 October 2021; pp. 2109–2114.
9. Ghali, R.; Akhloufi, M.A.; Souidene Mseddi, W.; Jmal, M. Wildfire Segmentation Using Deep-RegSeg Semantic Segmentation Architecture. In Proceedings of the 19th International Conference on Content-Based Multimedia Indexing, Graz, Austria, 14–16 September 2022; pp. 149–154.
10. Ghali, R.; Akhloufi, M.A. Deep Learning Approaches for Wildland Fires Using Satellite Remote Sensing Data: Detection, Mapping, and Prediction. *Fire* **2023**, *6*, 192. [\[CrossRef\]](#)
11. Xie, Z.; Song, W.; Ba, R.; Li, X.; Xia, L. A Spatiotemporal Contextual Model for Forest Fire Detection Using Himawari-8 Satellite Data. *Remote Sens.* **2018**, *10*, 1992. [\[CrossRef\]](#)
12. Li, Y.; Hu, J.; Wen, Y.; Evangelidis, G.; Salahi, K.; Wang, Y.; Tulyakov, S.; Ren, J. Rethinking Vision Transformers for MobileNet Size and Speed. *arXiv* **2022**, arXiv:2212.08059.
13. Tan, M.; Le, Q. EfficientNetV2: Smaller Models and Faster Training. In Proceedings of the 38th International Conference on Machine Learning, Virtual event, 18–24 July 2021; pp. 10096–10106.
14. Ba, R.; Chen, C.; Yuan, J.; Song, W.; Lo, S. SmokeNet: Satellite Smoke Scene Detection Using Convolutional Neural Network with Spatial and Channel-Wise Attention. *Remote Sens.* **2019**, *11*, 1702. [\[CrossRef\]](#)
15. Tao, C.; Zhang, J.; Wang, P. Smoke Detection Based on Deep Convolutional Neural Networks. In Proceedings of the International Conference on Industrial Informatics—Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII), Wuhan, China, 3–4 December 2016; pp. 150–153.
16. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
17. Yuan, F. Video-based Smoke Detection with Histogram Sequence of LBP and LBPV Pyramids. *Fire Saf. J.* **2011**, *46*, 132–139. [\[CrossRef\]](#)

18. Yin, Z.; Wan, B.; Yuan, F.; Xia, X.; Shi, J. A Deep Normalization and Convolutional Neural Network for Image Smoke Detection. *IEEE Access* **2017**, *5*, 18429–18438. [\[CrossRef\]](#)
19. Zeiler, M.D.; Fergus, R. Visualizing and Understanding Convolutional Networks. In Proceedings of the Computer Vision—ECCV 2014, Zurich, Switzerland, 6–12 September 2014; pp. 818–833.
20. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-scale Image Recognition. In Proceedings of the 3rd International Conference on Learning Representations (ICLR), Banff, AB, Canada, 14–16 April 2014; pp. 1–14.
21. Khan, S.; Muhammad, K.; Mumtaz, S.; Baik, S.W.; de Albuquerque, V.H.C. Energy-Efficient Deep CNN for Smoke Detection in Foggy IoT Environment. *IEEE Internet Things J.* **2019**, *6*, 9237–9245. [\[CrossRef\]](#)
22. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper With Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9.
23. Peng, Y.; Wang, Y. Real-time Forest Smoke Detection using Hand-designed Features and Deep Learning. *Comput. Electron. Agric.* **2019**, *167*, 105029. [\[CrossRef\]](#)
24. Manchanda, S.; Sharma, S. Analysis of Computer Vision Based Techniques for Motion Detection. In Proceedings of the 6th International Conference—Cloud System and Big Data Engineering (Confluence), Noida, India, 14–15 January 2016; pp. 445–450.
25. Iandola, F.N.; Moskewicz, M.W.; Ashraf, K.; Han, S.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <1 MB model size. *arXiv* **2016**, arXiv:1602.07360.
26. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 6848–6856.
27. Chollet, F. Xception: Deep Learning With Depthwise Separable Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
28. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:abs/1704.04861.
29. Gu, K.; Xia, Z.; Qiao, J.; Lin, W. Deep Dual-Channel Neural Network for Image-Based Smoke Detection. *IEEE Trans. Multimed.* **2020**, *22*, 311–323. [\[CrossRef\]](#)
30. Zhang, F.; Qin, W.; Liu, Y.; Xiao, Z.; Liu, J.; Wang, Q.; Liu, K. A Dual-Channel Convolution Neural Network for Image Smoke Detection. *Multimed. Tools Appl.* **2020**, *79*, 34587–34603. [\[CrossRef\]](#)
31. Jia, Y.; Chen, W.; Yang, M.; Wang, L.; Liu, D.; Zhang, Q. Video Smoke Detection with Domain Knowledge and Transfer Learning from Deep Convolutional Neural Networks. *Optik* **2021**, *240*, 166947. [\[CrossRef\]](#)
32. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
33. He, L.; Gong, X.; Zhang, S.; Wang, L.; Li, F. Efficient Attention based Deep Fusion CNN for Smoke Detection in Fog Environment. *Neurocomputing* **2021**, *434*, 224–238. [\[CrossRef\]](#)
34. Zhang, Z.; Jin, Q.; Wang, L.; Liu, Z. Video-based Fire Smoke Detection Using Temporal-spatial Saliency Features. *Procedia Comput. Sci.* **2022**, *198*, 493–498. [\[CrossRef\]](#)
35. Cheng, G.; Chen, X.; Gong, J. Deep Convolutional Network with Pixel-aware Attention for Smoke Recognition. *Fire Technol.* **2022**, *58*, 1839–1862. [\[CrossRef\]](#)
36. Tao, H.; Duan, Q. An Adaptive Frame Selection Network with Enhanced Dilated Convolution for Video Smoke Recognition. *Expert Syst. Appl.* **2023**, *215*, 119371. [\[CrossRef\]](#)
37. Cheng, G.; Zhou, Y.; Gao, S.; Li, Y.; Yu, H. Convolution-Enhanced Vision Transformer Network for Smoke Recognition. *Fire Technol.* **2023**, *59*, 925–948. [\[CrossRef\]](#)
38. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2020**, arXiv:abs/2010.11929.
39. Mohammed, R. A Real-time Forest Fire and Smoke Detection System using Deep Learning. *Int. J. Nonlinear Anal. Appl.* **2022**, *13*, 2053–2063. [\[CrossRef\]](#)
40. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI-17), San Francisco, CA, USA, 4–9 February 2017; pp. 4278–4284.
41. Chen, X.; Hopkins, B.; Wang, H.; O'Neill, L.; Afghah, F.; Razi, A.; Fulé, P.; Coen, J.; Rowell, E.; Watts, A. Wildland Fire Detection and Monitoring Using a Drone-Collected RGB/IR Image Dataset. *IEEE Access* **2022**, *10*, 121301–121317. [\[CrossRef\]](#)
42. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
43. Dilshad, N.; Khan, T.; Song, J. Efficient Deep Learning Framework for Fire Detection in Complex Surveillance Environment. *Comput. Syst. Sci. Eng.* **2023**, *46*, 749–764. [\[CrossRef\]](#)
44. Yar, H.; Khan, Z.A.; Ullah, F.U.M.; Ullah, W.; Baik, S.W. A modified YOLOv5 architecture for efficient fire detection in smart cities. *Expert Syst. Appl.* **2023**, *231*, 120465. [\[CrossRef\]](#)

45. Priya, R.S.; Vani, K. Deep Learning Based Forest Fire Classification and Detection in Satellite Images. In Proceedings of the 11th International Conference on Advanced Computing (ICoAC), Chennai, India, 18–20 December 2019; pp. 61–65.
46. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 2818–2826.
47. Tan, M.; Le, Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
48. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. Imagenet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [[CrossRef](#)]
49. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. ImageNet: A Large-scale Hierarchical Image Database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
50. Dillon, J.V.; Langmore, I.; Tran, D.; Brevdo, E.; Vasudevan, S.; Moore, D.; Patton, B.; Alemi, A.; Hoffman, M.D.; Saurous, R.A. TensorFlow Distributions. *arXiv* **2017**, arXiv:1711.10604.
51. Radosavovic, I.; Kosaraju, R.P.; Girshick, R.; He, K.; Dollar, P. Designing Network Design Spaces. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, DC, USA, 14–19 June 2020; pp. 10428–10436.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.