

Article

Wildfire Detection via a Dual-Channel CNN with Multi-Level Feature Fusion

Zhiwei Zhang ¹, Yingqing Guo ^{1,*}, Gang Chen ¹ and Zhaodong Xu ²

¹ College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing 210037, China; zhiwei_zhang@njfu.edu.cn (Z.Z.); chengang8059@njfu.edu.cn (G.C.)

² China-Pakistan Belt and Road Joint Laboratory on Smart Disaster Prevention of Major Infrastructures, Southeast University, Nanjing 210096, China; zhdxu@163.com

* Correspondence: gyingqing@126.com; Tel.: +86-13951835602

Abstract: Forest fires have devastating impacts on ecology, the economy, and human life. Therefore, the timely detection and extinguishing of fires are crucial to minimizing the losses caused by these disasters. A novel dual-channel CNN for forest fires is proposed in this paper based on multiple feature enhancement techniques. First, the features' semantic information and richness are enhanced by repeatedly fusing deep and shallow features extracted from the basic network model and integrating the results of multiple types of pooling layers. Second, an attention mechanism, the convolutional block attention module, is used to focus on the key details of the fused features, making the network more efficient. Finally, two improved single-channel networks are merged to obtain a better-performing dual-channel network. In addition, transfer learning is used to address overfitting and reduce time costs. The experimental results show that the accuracy of the proposed model for fire recognition is 98.90%, with a better performance. The findings from this study can be applied to the early detection of forest fires, assisting forest ecosystem managers in developing timely and scientifically informed defense strategies to minimize the damage caused by fires.

Keywords: wildfire hazard; wildfire detection; dual-channel CNN; multi-level feature fusion

1. Introduction



Citation: Zhang, Z.; Guo, Y.; Chen, G.; Xu, Z. Wildfire Detection via a Dual-Channel CNN with Multi-Level Feature Fusion. *Forests* **2023**, *14*, 1499. <https://doi.org/10.3390/f14071499>

Received: 17 April 2023

Revised: 6 July 2023

Accepted: 20 July 2023

Published: 21 July 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/>).

Forests are often referred to as the “lungs of the Earth” due to their two important values. One is their visible economic value, and the other is their intangible ability to regulate the climate and maintain ecological balance. Forest fires have occurred frequently in recent years due to extreme events such as lightning, volcanic eruptions, and human activities [1–4]. For example, in 2015, forest fires in Indonesia burned over 2.6 million hectares of land, resulting in more than USD 16 billion in economic losses, and posing significant threats to the local biodiversity and endangered species [5,6]. Since September 2019, the prolonged Australian bushfires have burned over 10 million hectares of land, causing at least 33 deaths and displacing billions of animals [7,8]. Based on the analysis conducted by Lukic et al. [9] in 2012, the frequency of fires in the Tara Mountains exceeded the average level of fire occurrence in Serbia, primarily due to the strong influence of climatic conditions [10]. The occurrence of forest fires and the extent of the damage they inflict are also influenced by factors connected to the forest itself [11–13], such as the species of trees present [14], the combustibility of the forest fuels, and the water content [15,16]. Therefore, adopting proactive and effective forest fire monitoring methods is crucial for ecological, social, and economic reasons [17–19]. The use of image processing and deep learning methods has become the mainstream research direction in this field [20].

Traditional image-based fire detection methods employ feature extraction algorithms based on prior knowledge to analyze flame or smoke characteristics, and machine learning algorithms are then used to determine if a fire has occurred. Among these methods, flame or smoke features are analyzed under various color spaces [21] such as RGB [22], YCbCr [23],

and YUV [24], or multiple features related to flame or smoke are fused (e.g., motion information [25], shape [26], and area [27]) to construct expert systems for fire detection. In 2009, Rudz et al. [28] accurately identified fire characteristics and achieved better fire detection through the use of fire color space and segmented fire images. In 2019, Matlani et al. [29] improved the accuracy of detection by integrating features such as Haar, SIFT, and others. Some researchers construct fire models by evolving dynamic textures with multiple spatiotemporal features [30] or by detecting color changes in fire motion regions [31] and irregular boundaries [32], among others. In 2013, Wang et al. [33] developed prior flame probability by combining flame color probability and the results of the Wald–Wolfowitz randomness test. This method was shown to have good robustness and the ability to adapt to different environments, making it a reliable and practical approach to flame detection.

Forest fires can be roughly divided into the following five stages: the ignition stage, the propagation stage, the peak burning stage, the stage of fire suppression and weakening, and the stage of fire termination [34]. Each stage has distinct fire characteristics [35]. For example, during the ignition stage [36], the fire may only produce faint smoke, while in the propagation stage, the fire can spread rapidly. At the peak burning stage, flames and smoke may become very thick, and the fire can reach its maximum intensity. Therefore, fire monitoring algorithms that are based solely on empirical knowledge may only perform well in specific scenarios and may have limited generalization to other situations [37].

In contrast, convolutional neural networks (CNNs) automatically extract features from provided data [38,39], thus avoiding the limitations of manually selected features [40,41]. Furthermore, an excellent CNN can also achieve good results in other application domains. In 2019, Kim et al. [42] and Lee et al. [43] both utilized Faster-RNN to identify fire regions. The former directly recognized the fire and smoke characteristics, while the latter used a combination of local and global frame features as the judgment basis. In 2017, Wang et al. [44] replaced the CNN’s fully connected layer with SVM to obtain better detection results. In 2016, Frizzi et al. [45] significantly reduced the time cost by designing a feature map instead of the original frame. In 2020, Liu et al. [46] proposed a fire recognition model based on a two-level classifier. They used a first-level classifier composed of HOG and Adaboost for preliminary recognition and a second-level classifier composed of CNN and SVM for further fire recognition with a higher accuracy. In 2022, Guo et al. [47] designed a Lasso_SVM layer to replace the fully connected layer in the original model and improved the model detection accuracy through segmented training. In the same year, Qian et al. [48] proposed a weighted fusion algorithm for forest fire source recognition based on two weakly supervised models, Yolov5 and EfficientDet. In 2020, Xie et al. [49] designed a network called Controllable Smoke Image Generation Neural Networks-V2 (CSGNet-v2) to generate realistic smoke images based on smoke characteristics, which can be used for smoke detection in forest fires. In 2021, Ding et al. [50] proposed a dual-stream convolutional neural network based on attention mechanisms, which pays more attention to the spatiotemporal characteristics of smoke and enhances the ability to segment and recognize small smoke particles.

In addition to simply improving the CNN structure, some researchers improved their algorithms by complementing the advantages of traditional image processing methods with those of CNNs. For example, in 2022, Yang et al. [51] optimized the method for identifying early spring green tea by using semi-supervised learning and image processing. Wu et al. [52] proposed an adaptive deep-learning flame and smoke classifier based on traditional feature extraction algorithms in the Caffe framework in 2017. In 2019, Wang et al. [53] combined the advantages of traditional image processing and CNNs to propose an adaptive pooling convolutional neural network that effectively extracts features by pre-learning the flame segmentation area features and avoiding the blind nature of traditional feature extraction, thus improving the effectiveness of CNN learning. In 2023, Zheng et al. [54] effectively extracted target features by using cross-attention blocks to capture differences in global information and local color and texture feature information.

The accuracy and generalizability of CNN-based methods are superior to the suggested fire detection algorithms that are based on a priori knowledge. A novel two-channel CNN model is put forth in this article, and the basic network is optimized using two feature fusion techniques and the addition of an attention mechanism. A more potent dual-channel network is created by fusing two single-channel networks that have varying input sizes but similar structural characteristics. Migration learning provides a solution to the issue of the model training's simple overfitting and lowers the model training's time cost. The research methodology outlined in this paper can be applied to the task of early detection and identification of forest fires. Its findings will assist forest ecosystem managers in gaining timely awareness of forest fires and implementing scientific fire prevention measures to mitigate the losses caused by such fires. This is crucial for the protection of forest resources.

2. Materials and Methods

2.1. Construction of the Experimental Dataset

The dataset plays a crucial role in deep learning research and is one of the key factors in achieving exceptional model performance. In this study, a dataset of 14,000 images was utilized, which was obtained by members of the research team through the Google search engine. The images can be categorized into the following two groups: "non-fire" and "fire" [47]. The "fire" category encompasses images of flames, white smoke, black smoke, dense smoke, and thin smoke generated by fires in different settings, such as forests, grasslands, fields, and urban areas. The dataset creation method was inspired by references [17,55,56], which involved including various angles of regular forest images in the "non-fire" category, as well as incorporating more complex and disruptive images, such as clouds, sunlight, and pyrocumulus clouds, under different weather conditions. By introducing these disruptive factors, the objective was to enhance the model's ability to recognize smoke and flames and improve its robustness. Following an approximate 7:3 ratio, the images from the different categories in the dataset were divided into training and testing sets, with 10,000 images in the training set and the remaining images used for testing. Figure 1 displays some sample images that can facilitate a better understanding of the dataset's composition and characteristics.

2.2. Essential Basic Knowledge

Before building the network model in this article, it is necessary to introduce some fundamental concepts, including feature fusion, transfer learning, and attention mechanism.

2.2.1. Feature Fusion

Feature fusion is a widely used technique applied in fields such as computer vision and natural language processing. It involves combining feature information from diverse sources through concatenation, merging, stacking, and cascading operations. This technique effectively tackles issues such as data sparsity and noise interference, and enhances feature expression and generalization ability [57–61].

This article presents two feature fusion methods. The first method is inspired by the SPP idea, which replaces a single pooling layer in the base network with a combination of a max pooling layer and an average pooling layer that acts simultaneously and fuses the results of the two pools. The second approach fuses the shallow and deep features of the network and is influenced by the idea of residual networks. These methods enhance the richness of feature information, enabling the model to learn data features from multiple perspectives and provide a more comprehensive and accurate description of the data's essence. Based on the inspiration from SPP and residual structures, two network architectures for feature fusion methods were designed in this study, as shown in Figure 2.

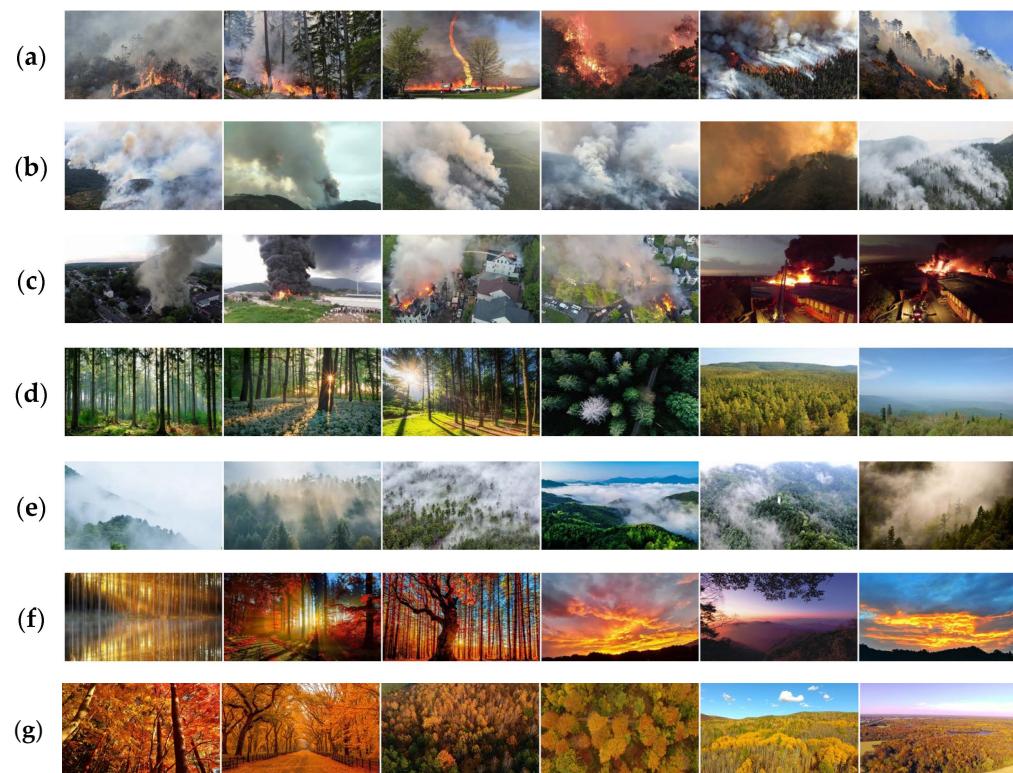


Figure 1. Example images from the dataset, including images of fire and smoke in forests and cities, typical forest images, and interference images. (a–c) The fire images mainly come from various scenes such as forests, wilderness areas, and urban areas, depicting flames, white smoke, black smoke, dense smoke, and other associated phenomena during fire incidents. (d) Non-fire undisturbed forest images were captured from various angles, including shots from inside the forest and aerial views from above the forest. (e–g) Non-fire images are accompanied by various disruptive elements, including clouds, sunlight, tree leaves with colors similar to fire, and pyrocumulus clouds.

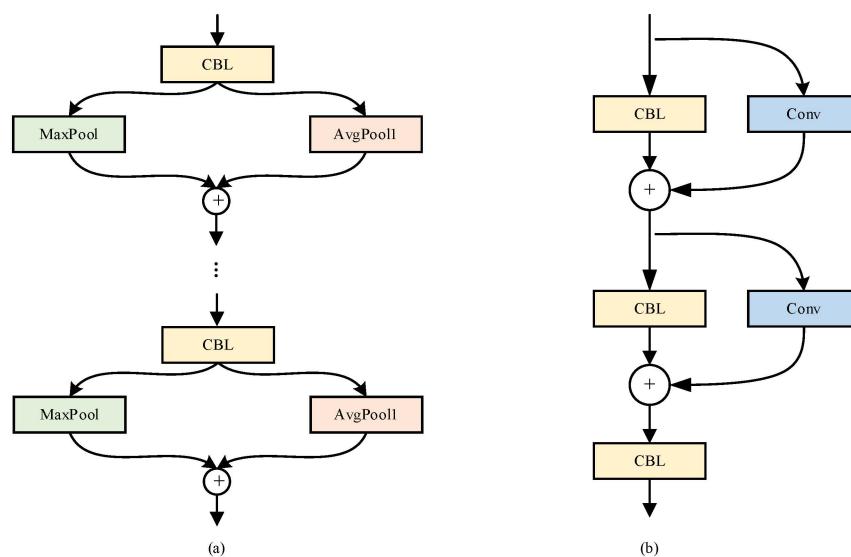


Figure 2. Network architectures for two feature fusion methods. (a) The first feature fusion network structure, inspired by the SPP structure. (b) The second feature fusion network structure, inspired by the residual structure.

2.2.2. Transfer Learning

Training a deep learning model for a specific task is a complex and expensive process that involves various challenges. Firstly, creating a dataset requires a lot of effort and resources to collect and label the data accurately. Secondly, training a network from scratch is time consuming and results in the network having a slow convergence speed. More powerful hardware is required to mitigate this issue, which can be costly.

Transfer learning was proposed to address the problem of limited annotated training data. The schematic diagram of transfer learning is shown in Figure 3. Transfer learning involves transferring the model parameters that were trained on a larger, more general dataset to a new target task network [62]. By leveraging the knowledge learned from previous tasks, the model can achieve higher accuracy and faster convergence with less task-specific annotated data [63–65].

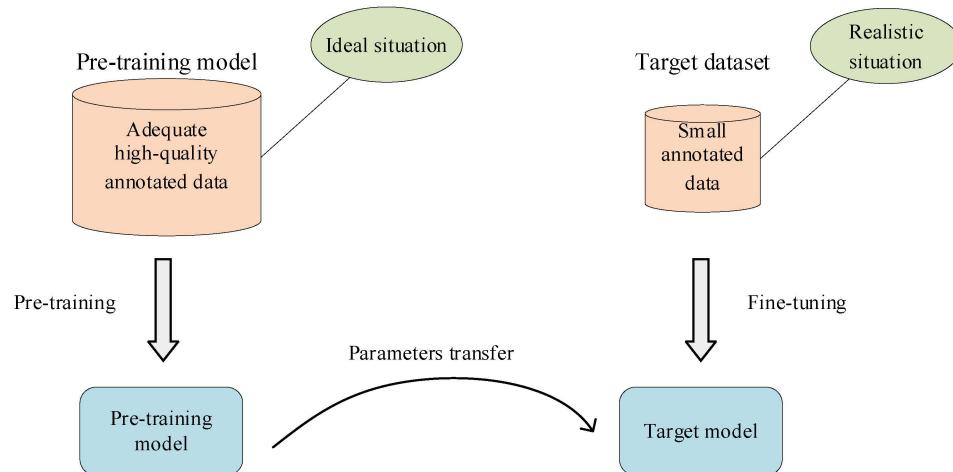


Figure 3. Schematic diagram of transfer learning.

The parameter migration of the single-channel network model structure in this paper is shown in Figure 4. The parameters of the C1–C2 convolutional layers and the parameters of the three fully connected layers of the pre-trained Alexnet model on the Imagenet dataset are migrated to the modified single-channel model and then trained using the fire dataset in this paper.

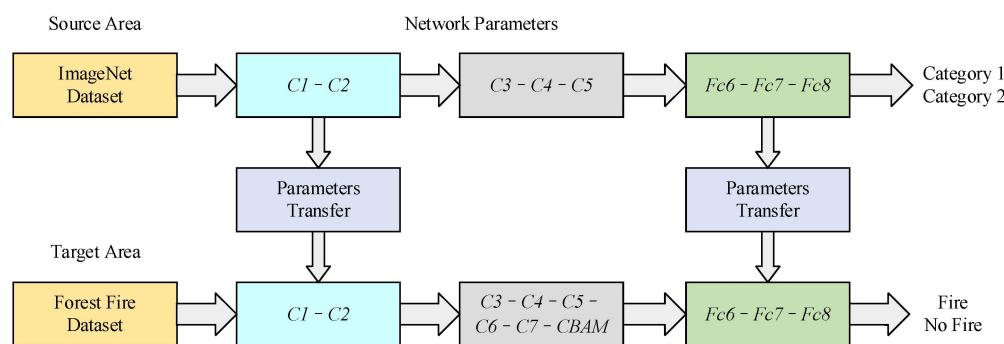


Figure 4. Schematic diagram of migration learning for single-channel network.

2.2.3. Attention Mechanism

The neural attention mechanism is a technique that enables neural networks to focus on input features and applies to data of various shapes. It helps the network to locate the relevant information from complex backgrounds and to suppress irrelevant information, thereby improving the network's performance and simplifying its structure.

Convolutional block attention module (CBAM) [66] is a lightweight attention mechanism network that can be seamlessly integrated into CNNs to filter out key information from a large amount of irrelevant background with limited resources and negligible overhead. The central idea of this module is to attend to the “what” along the channel axis and the “where” along the spatial axis, and to enhance the meaningful features along both dimensions by sequentially applying channel and spatial attention modules. By doing so, the module can capture the most relevant information in both the channel and spatial dimensions, resulting in improved feature representations [67,68].

The detailed module structure is shown in Figure 5. The parameter settings for the CBAM module are shown in Table 1.

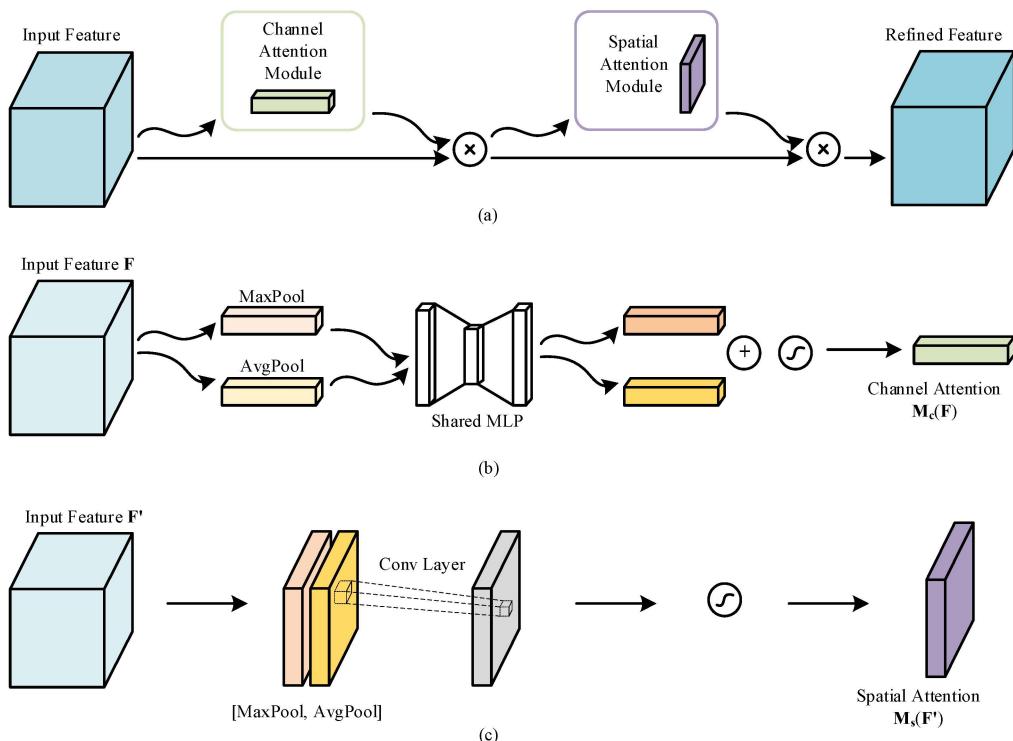


Figure 5. (a) Structure of the CBAM, (b) structure of the channel attention module, (c) structure of the spatial attention module.

Table 1. The parameter settings for the CBAM module.

Module	Type	Input Size	Kernel Size	Kernel Number	Output Size	Stride	Padding
Channel	MaxPool	$H \times W \times C$	$H \times W$	None	$1 \times 1 \times C$	1	None
	AvgPool	$H \times W \times C$	$H \times W$	None	$1 \times 1 \times C$	1	None
	Fc1	C	None	None	$C/16$	None	None
	Fc2	$C/16$	None	None	C	None	None
Spatial	Conv1	$H \times W \times 2$	7×7	1	$H \times W \times 1$	1	3
	Pool1	$56 \times 56 \times 64$	3×3	None	$27 \times 27 \times 64$	2	0

The calculation formulas for the channel attention module and spatial attention module are shown in Equations (1) and (2), respectively.

$$M_c(F) = \sigma(MLP(\text{MaxPool}(F)) + MLP(\text{AvgPool}(F))) \quad (1)$$

$$M_s(F') = \sigma\left(f^{7 \times 7}([AvgPool(F'); MaxPool(F')])\right) \quad (2)$$

Assuming that the size of the inputs F and F' for both the channel attention module and spatial attention module is $R^{C \times H \times W}$, the corresponding output sizes for M_c and M_s are

$R^{C \times 1 \times 1}$ and $R^{1 \times H \times W}$, respectively. The notation $f^{7 \times 7}$ represents a convolution operation with a 7×7 kernel size, where σ denotes the sigmoid function.

2.3. Establishment of an Improved Single-Channel Model

The classical Alexnet network [69] is chosen as the base network, which consists of 5 convolutional layers, 3 pooling layers, and 3 fully connected layers.

The improvement of the single-channel model is divided into two main aspects. On one hand, there is feature fusion. One approach is to perform global average pooling and global maximum pooling operations on the extracted features, and then combine the results to improve feature characterization. In this case, the two pooling layers within the same layer share the convolution kernel size, step size, padding, and other parameters. The second fusion uses convolutional layers to combine shallow and deep features. Conv6 and Conv7, a 1×1 convolutional layer, are added in parallel at the location of the original network Conv3 and Conv4, respectively, for the fusion of shallow and deep features. The Conv6 layer adjusts the output of the Pool2 layer to $13 \times 13 \times 384$ and then fuses it with the output of the Conv3 layer. Conv7 processes the first feature fusion result and then fuses it with the result of the Conv4 layer.

The structure diagram of the improved single-channel model is shown in Figure 6.

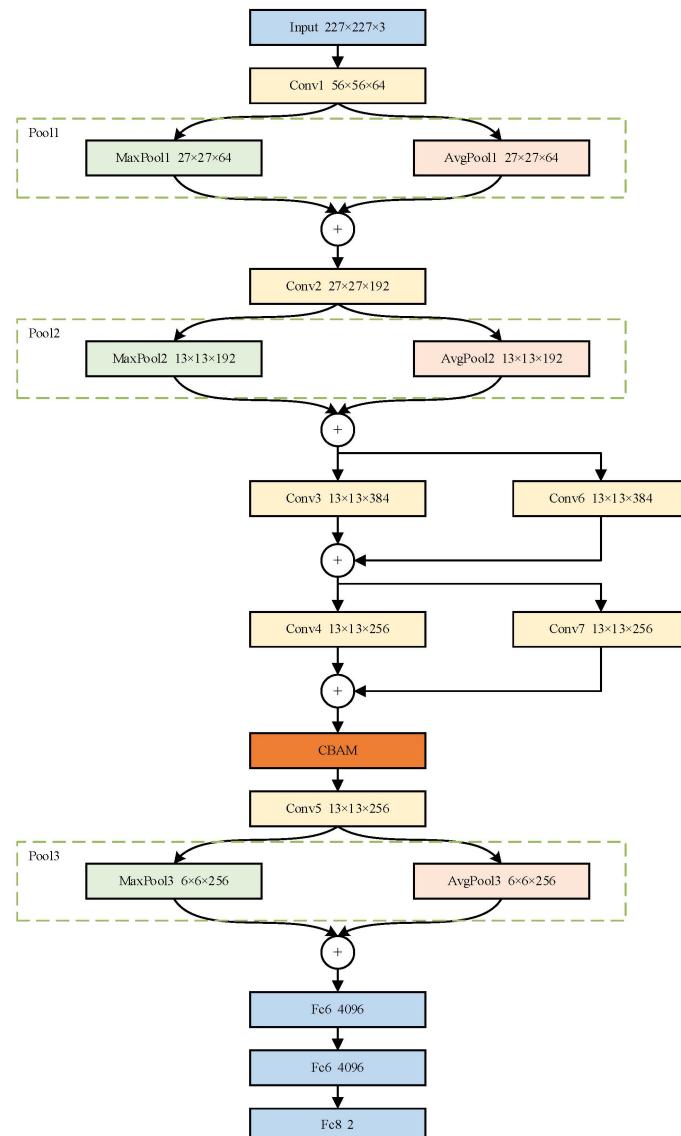


Figure 6. Structure of the improved single-channel network.

2.4. Establishment of a Novel Dual-Channel Network

In this part, a two-channel CNN model that can cover different-sized fire scenes is designed. First, another network with structurally similar inputs of $336 \times 336 \times 3$ is designed based on the improved single-channel model. Then, the feature extraction results of the two networks are fused to obtain a novel two-channel network, whose structure is shown in Figure 7.

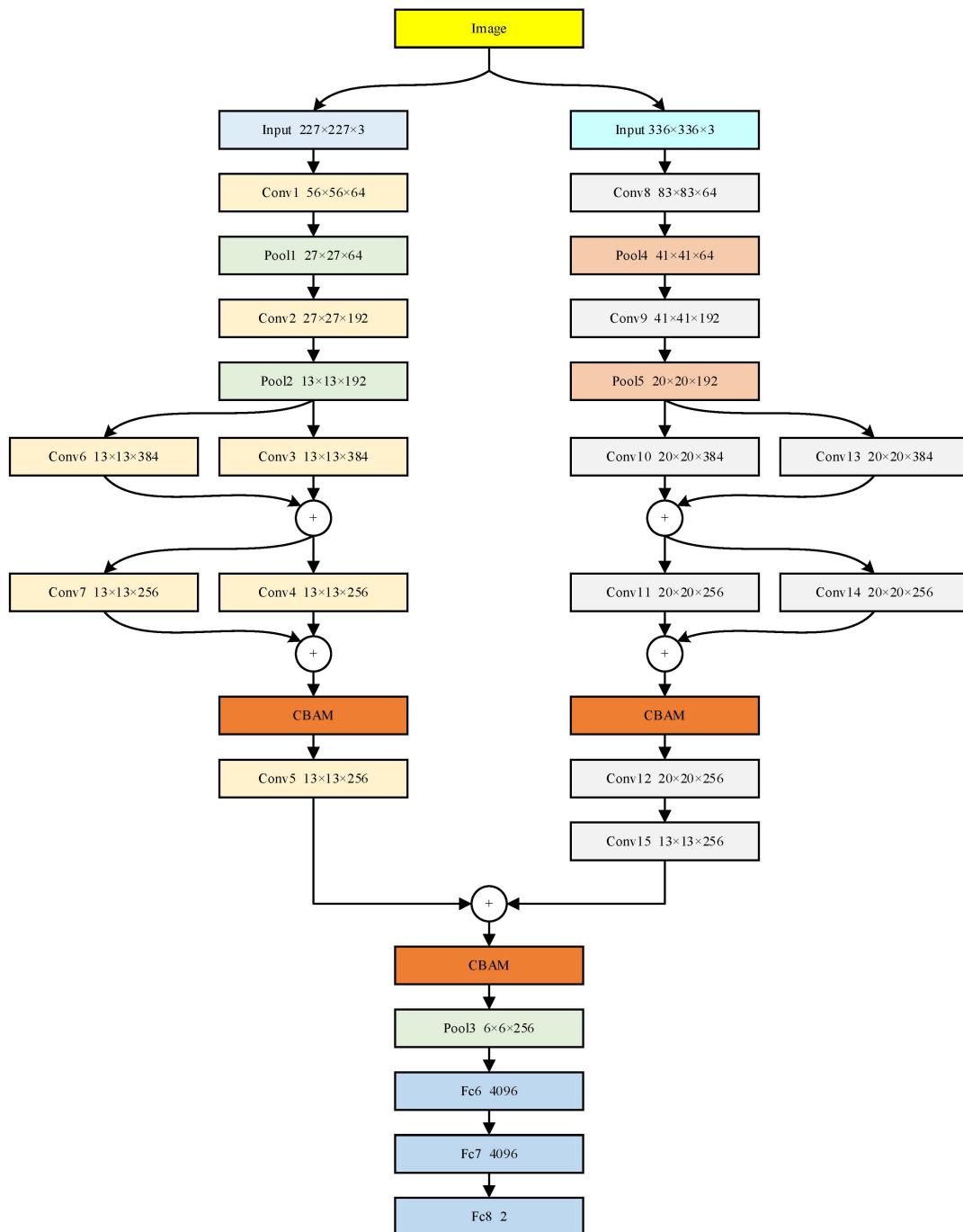


Figure 7. Structure of the novel dual-channel network.

To facilitate the distinction, the two single-channel networks are renamed; the channel network where the input is $227 \times 227 \times 3$ is called the first channel, and the one with the input of $336 \times 336 \times 3$ is called the second channel. The parameters of the novel dual-channel network are shown in Table 2.

Table 2. Parameters of the novel dual-channel network.

Type	Input Size	Kernel Size	Kernel Number	Output Size	Stride	Padding
Conv1	227 × 227 × 3	11 × 11	64	56 × 56 × 64	4	2
Pool1	56 × 56 × 64	3 × 3	none	27 × 27 × 64	2	0
Conv2	27 × 27 × 64	5 × 5	192	27 × 27 × 192	1	2
Pool2	27 × 27 × 192	3 × 3	none	13 × 13 × 192	2	0
Conv3	13 × 13 × 192	3 × 3	384	13 × 13 × 384	1	1
Conv4	13 × 13 × 384	3 × 3	256	13 × 13 × 256	1	1
Conv5	13 × 13 × 256	3 × 3	256	13 × 13 × 256	1	1
Conv6	13 × 13 × 192	1 × 1	384	13 × 13 × 384	1	0
Conv7	13 × 13 × 384	1 × 1	256	13 × 13 × 256	1	0
Conv8	336 × 336 × 3	11 × 11	64	83 × 83 × 64	4	2
Pool4	83 × 83 × 64	3 × 3	none	41 × 41 × 64	2	0
Conv9	41 × 41 × 64	5 × 5	192	41 × 41 × 192	1	2
Pool5	41 × 41 × 192	3 × 3	none	20 × 20 × 192	2	0
Conv10	20 × 20 × 192	3 × 3	384	20 × 20 × 384	1	1
Conv11	20 × 20 × 384	3 × 3	256	20 × 20 × 256	1	1
Conv12	20 × 20 × 256	3 × 3	256	20 × 20 × 256	1	1
Conv13	20 × 20 × 192	1 × 1	384	20 × 20 × 384	1	0
Conv14	20 × 20 × 384	1 × 1	256	20 × 20 × 256	1	0
Conv15	20 × 20 × 256	1 × 1	256	13 × 13 × 256	2	3
Pool3	13 × 13 × 256	3 × 3	none	6 × 6 × 256	2	0
Fc6	9216	none	none	4096	none	none
Fc7	4096	none	none	4096	none	none
Fc8	4096	none	none	2	none	none

It should be noted that the output of the second channel at the Conv12 layer is $20 \times 20 \times 256$, which is different from the output size of the Conv12 layer of the first channel. Therefore, the two need to be resized before fusion. In this paper, we chose to build a 1×1 convolutional unit (Conv15), and the output of the Conv12 layer was adjusted to $13 \times 13 \times 256$. After the fusion of the features of the two channels, it is still necessary to process them using CBAM and then continue with operations such as pooling layers (Pool3) and fully connected layers (Fc6–Fc7–Fc8).

3. Results

The experiment was carried out in the Pytorch framework of Windows 10, using an Intel® Core™ i7-12700k CPU (Santa Clara, CA, USA) running at a standard frequency of 5.00 GHz, 128 GB of RAM, and a GPU with an NVIDIA RTX 3090 (Santa Clara, CA, USA) acting as the hardware gas pedal for model training.

3.1. Simulation Analysis of the Improved Single-Channel Model

The image size of the forest fire dataset photos is scaled to $227 \times 227 \times 3$ by scaling the transformation to meet the Alexnet network's input image size requirement. After 60 rounds of training using the resized pictures, the model's test set accuracy reaches 95.30%.

The improved single-channel model achieves superior results in the task of recognizing forest fires because the fusion of deep and shallow features enhances the feature representation. Figure 8 displays the accuracy curves of the enhanced Alexnet model before and after the enhancement of the test set. When compared to the accuracy before modification, the improved Alexnet forest fire recognition accuracy is 1.50% higher at 96.8%, and the recognition error rate drops by 31.9%.

The accuracy curves of the training set and test set of the improved single-channel model are shown in Figure 9. With a training set accuracy of 99.59% and a test set accuracy of 96.80%, the training set accuracy is higher, and the model exhibits significant overfitting.

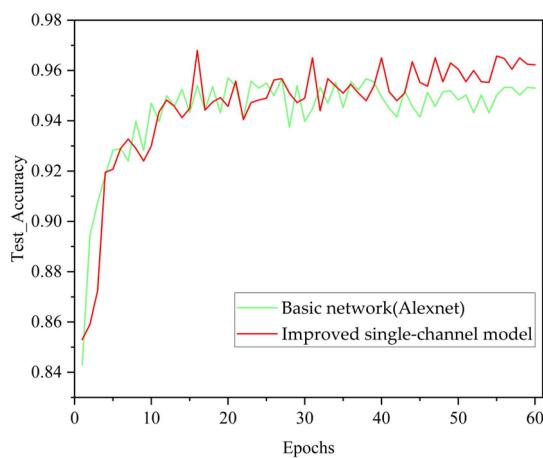


Figure 8. The test set accuracy curve of the basic model before and after improvement.

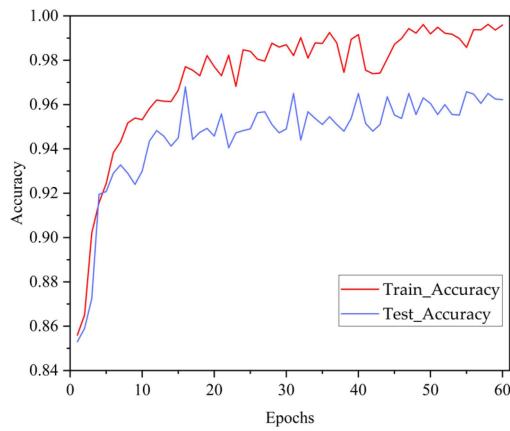


Figure 9. Accuracy curves of the training and test sets of the improved single-channel model.

It was demonstrated that transfer learning can address the issue of model overfitting or insufficient datasets [70–72]. This approach is also used in this study to address the ease of model overfitting and reduce the time cost. The accuracy curves of the test set of the model before and after using migration learning are shown in Figure 10, and it is clear that migration learning improved the test accuracy of the model. The accuracy of the model test set before using migration learning is 96.80%, and after using it, the accuracy is 98.45%, with a 1.65% improvement in accuracy and a 51.56% decrease in the recognition error rate.

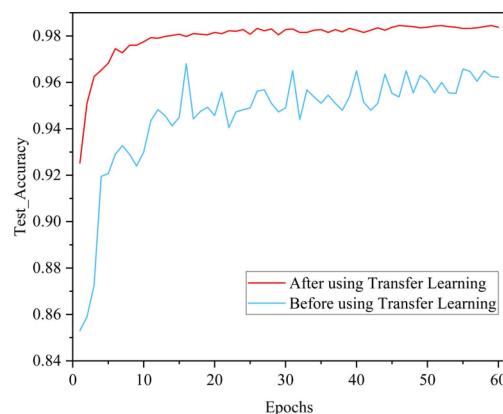


Figure 10. Accuracy curves of the test set before and after using migration learning.

3.2. Simulation Analysis of the New Dual-Channel Network

The second channel CNN with the input size of $336 \times 336 \times 3$ is designed concerning the improved single-channel network. It is trained separately, and the accuracy of the model is 98.15% after 60 rounds of training, which is higher than the accuracy of the first channel. This indicates that designing models with different input sizes is effective in improving accuracy.

The test set accuracy of the two-channel network obtained from the fusion of the two single-channel networks is 98.90%, which is an improvement of 0.45% and 0.75% in accuracy relative to the first and second channel models, respectively. The accuracy curves of the two-channel network as well as the two separate networks on the test set are shown in Figure 11.

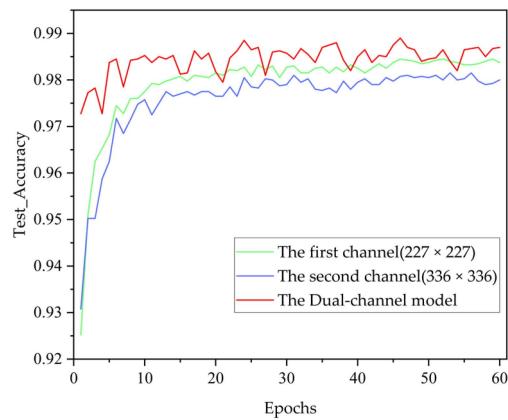


Figure 11. Accuracy curves of test sets for different models.

The performance of the novel two-channel model proposed in this paper is analyzed with the help of a confusion matrix. The confusion matrix is shown in Figure 12, and there are only 44 recognition errors among 4000 test samples. The model has a strong performance with an accurate prediction rate of 98.9%, a precision of 99.24%, a recall rate of 98.55%, and a specificity of 98.56%. An inspection of the misidentified images shows that the model misidentifies images of heavy clouds, images of fire clouds, and images of forests with large fire-like colors.

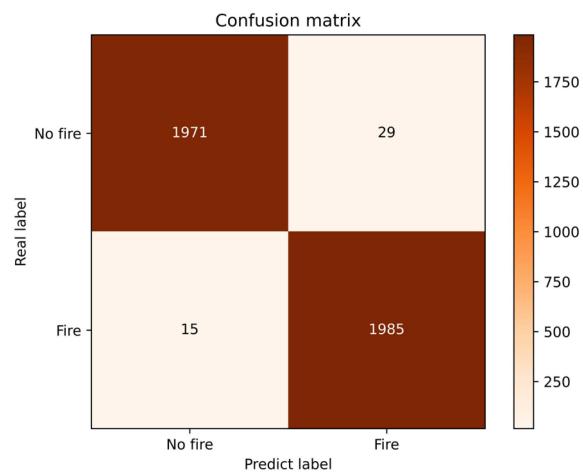


Figure 12. The confusion matrix of the dual-channel CNN.

Additionally, this study contrasts VGG16 [73] and Resnet50 [74], which are two different deep-learning methods. The test results are provided in Table 3, and the model in this research has the greatest accuracy of 98.90% among these methods.

Table 3. Performance comparison between different models.

The Relevant Literature	Models	Model Accuracy (%)
This paper [73] [74]	Novel dual-channel CNN	98.90
	VGG16	95.88
	Resnet50	97.78

4. Discussion

As shown in Table 4, the performance of the two-channel network model previously proposed in this paper and the subject group [47] is compared. Obviously, the new two-channel network outperforms the previous network in terms of accuracy, precision, and recall. It is worth mentioning that the recognition error rate of the new two-channel network is reduced by 28.89% compared with the previously proposed two-channel network, in which the probability of misclassifying a fire as a non-fire is reduced by 48.28%.

Table 4. Performance comparison between dual-channel CNNs.

The Literature	TP	FN	FP	TN	Accuracy	Precision	Recall
This paper [47]	1971 1968	29 32	15 29	1985 1971	98.90 98.48	99.24 98.55	98.55 98.40

The model described in this research is also compared to other models in the field. The RMSN proposed in [75] is a fire detection model based on RNN architecture. It combines temporal and spatial network features to achieve enhanced accuracy in fire detection. However, the use of RNNs in this model results in slower detection speeds and increased hardware requirements.

Furthermore, the dataset itself is an important limiting factor. The dataset utilized in [75] focuses solely on fire smoke images in different scenarios, disregarding the consideration of other potential interferences such as clouds, fog, or fire-like objects and their impact on fire detection. This limitation is not only evident in [75] alone, but also in [76,77], and in the other related literature. These datasets predominantly consider the influence of specific factors while neglecting the comprehensive consideration of multiple factors. However, the dataset in this study is more comprehensive. In addition to the typical forest fire data, it also includes a wide range of interfering factors such as clouds, sunlight, fire clouds, and forest areas with fire-like colors. It is foreseeable that in future research on deep-learning-based or machine-learning-based forest fire identification methods, there will be an increased focus on improving the dataset in addition to enhancing the network structure.

5. Conclusions

Forest fires have the potential to negatively impact forest ecosystems, cause economic losses, and even pose a threat to human lives. Therefore, the development of an effective fire detection method is crucial. In this paper, a novel two-channel network forest fire identification method is proposed that achieves higher accuracy in fire detection. Two different feature fusion approaches are utilized to combine features at various stages of the underlying network, thereby enhancing the characterization capability of the features. Subsequently, the enhanced features are streamlined, key information is extracted, and feature redundancy is reduced through the use of an attention mechanism. This process improves the effectiveness of the extracted features. And the dual-channel model exhibits superior performance as it is capable of accommodating varying input sizes. Furthermore, transfer learning plays a pivotal role in mitigating model overfitting and minimizing training time costs. The experimental results show that this new two-channel network significantly outperforms the single-channel network in fire recognition with an accuracy of 98.90%.

However, there are still some issues in particular circumstances. The thick fog in the forest might closely match the thick smoke characteristics produced during the peak stages of a fire when the weather is exceptionally foggy. Reddish-hued forests are already prone to erroneous interpretation; sunlight interference makes this risk even worse. Given that both fog and sunlight frequently occur in the actual world, future research on image-based forest fire monitoring should pay particular attention to these two sources of interference. To further improve the model, we intend to increase the quality and quantity of wildfire photographs, with a special focus on examining fog properties, sunshine during various seasons, and leaf traits that resemble fire colors. In order to increase the model's accuracy and resilience, we will also add more difficult images that can result in identification failures to the training dataset.

Author Contributions: Conceptualization, Y.G. and Z.Z.; methodology, Z.Z.; software, Z.Z.; validation, G.C. and Y.G.; formal analysis, Z.Z.; investigation, G.C.; resources, Z.X.; data curation, Z.Z. and G.C.; writing—original draft preparation, Z.Z.; writing—review and editing, Y.G.; visualization, G.C.; supervision, Y.G.; project administration, Z.X.; funding acquisition, Y.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Program on the Key R&D Project of China, grant number 2020YFB2103500, and the Major Project of Fundamental Research on Frontier Leading Technology of Jiangsu Province, grant number BK20222006.

Data Availability Statement: Not applicable.

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

References

1. Li, W.; Xu, Q.; Yi, J.; Liu, J. Predictive Model of Spatial Scale of Forest Fire Driving Factors: A Case Study of Yunnan Province, China. *Sci. Rep.* **2022**, *12*, 19029. [[CrossRef](#)] [[PubMed](#)]
2. Sachdeva, S.; Bhatia, T.; Verma, A.K. GIS-Based Evolutionary Optimized Gradient Boosted Decision Trees for Forest Fire Susceptibility Mapping. *Nat. Hazards* **2018**, *92*, 1399–1418. [[CrossRef](#)]
3. Boer, M.M.; Resco De Dios, V.; Bradstock, R.A. Unprecedented Burn Area of Australian Mega Forest Fires. *Nat. Clim. Chang.* **2020**, *10*, 171–172. [[CrossRef](#)]
4. Rogelj, J.; Meinshausen, M.; Knutti, R. Global Warming under Old and New Scenarios Using IPCC Climate Sensitivity Range Estimates. *Nat. Clim. Chang.* **2012**, *2*, 248–253. [[CrossRef](#)]
5. Edwards, R.B.; Naylor, R.L.; Higgins, M.M.; Falcon, W.P. Causes of Indonesia's Forest Fires. *World Dev.* **2020**, *127*, 104717. [[CrossRef](#)]
6. Purnomo, H.; Shantiko, B.; Sitorus, S.; Gunawan, H.; Achdiawan, R.; Kartodihardjo, H.; Dewayani, A.A. Fire Economy and Actor Network of Forest and Land Fires in Indonesia. *For. Policy Econ.* **2017**, *78*, 21–31. [[CrossRef](#)]
7. Abram, N.J.; Henley, B.J.; Sen Gupta, A.; Lippmann, T.J.R.; Clarke, H.; Dowdy, A.J.; Sharples, J.J.; Nolan, R.H.; Zhang, T.; Wooster, M.J.; et al. Connections of Climate Change and Variability to Large and Extreme Forest Fires in Southeast Australia. *Commun. Earth Environ.* **2021**, *2*, 8. [[CrossRef](#)]
8. Collins, L.; Bradstock, R.A.; Clarke, H.; Clarke, M.F.; Nolan, R.H.; Penman, T.D. The 2019/2020 Mega-Fires Exposed Australian Ecosystems to an Unprecedented Extent of High-Severity Fire. *Environ. Res. Lett.* **2021**, *16*, 044029. [[CrossRef](#)]
9. Lukić, T.; Marić, P.; Hrnjak, I.; Gavrilov, M.B.; Mladjan, D.; Zorn, M.; Komac, B.; Milošević, Z.; Marković, S.B.; Sakulski, D.; et al. Forest Fire Analysis and Classification Based on a Serbian Case Study. *Acta Geogr. Slov.* **2017**, *57*, 51–63. [[CrossRef](#)]
10. Novković, I.; Goran, B.; Markovic, G.; Lukic, D.; Dragicevic, S.; Milosevic, M.; Djurdjic, S.; Samardžić, I.; Lezaic, T.; Tadic, M. GIS-Based Forest Fire Susceptibility Zonation with IoT Sensor Network Support, Case Study-Nature Park Golija, Serbia. *Sensors* **2021**, *21*, 6520. [[CrossRef](#)]
11. Moritz, M.A.; Batllori, E.; Bradstock, R.A.; Gill, A.M.; Handmer, J.; Hessburg, P.F.; Leonard, J.; McCaffrey, S.; Odion, D.C.; Schoennagel, T.; et al. Learning to Coexist with Wildfire. *Nature* **2014**, *515*, 58–66. [[CrossRef](#)]
12. Tian, X.; Zhao, F.; Shu, L.; Wang, M. Distribution Characteristics and the Influence Factors of Forest Fires in China. *For. Ecol. Manag.* **2013**, *310*, 460–467. [[CrossRef](#)]
13. Page, S.; Siegert, F.; Rieley, J.; Boehm, H.-D.; Jaya, A.; Limin, S. The Amount of Carbon Released from Peat and Forest Fires in Indonesia During 1997. *Nature* **2002**, *420*, 61–65. [[CrossRef](#)]
14. Odion, D.C.; Hanson, C.T.; Arsenault, A.; Baker, W.L.; DellaSala, D.A.; Hutto, R.L.; Klenner, W.; Moritz, M.A.; Sherriff, R.L.; Veblen, T.T.; et al. Examining Historical and Current Mixed-Severity Fire Regimes in Ponderosa Pine and Mixed-Conifer Forests of Western North America. *PLoS ONE* **2014**, *9*, e87852. [[CrossRef](#)] [[PubMed](#)]

15. Rosavec, R.; Barčić, D.; Španjol, Ž.; Oršanić, M.; Dubravac, T.; Antonović, A. Flammability and Combustibility of Two Mediterranean Species in Relation to Forest Fires in Croatia. *Forests* **2022**, *13*, 1266. [[CrossRef](#)]
16. Tošić, I.; Mladjan, D.; Gavrilov, M.; Zivanovic, S.; Radaković, M.; Putniković, S.; Petrović, P.; Krstic-Mistridzelovic, I.; Markovic, S. Potential Influence of Meteorological Variables on Forest Fire Risk in Serbia during the Period 2000–2017. *Open Geosci.* **2019**, *11*, 414–425. [[CrossRef](#)]
17. Li, A.; Zhao, Y.; Zheng, Z. Novel Recursive BiFPN Combining with Swin Transformer for Wildland Fire Smoke Detection. *Forests* **2022**, *13*, 2032. [[CrossRef](#)]
18. Sivrikaya, F.; Küçük, Ö. Modeling Forest Fire Risk Based on GIS-Based Analytical Hierarchy Process and Statistical Analysis in Mediterranean Region. *Ecol. Inform.* **2022**, *68*, 101537. [[CrossRef](#)]
19. Ciprián-Sánchez, J.F.; Ochoa-Ruiz, G.; Gonzalez-Mendoza, M.; Rossi, L. FIRe-GAN: A Novel Deep Learning-Based Infrared-Visible Fusion Method for Wildfire Imagery. *Neural Comput. Appl.* **2021**, *1*–13. [[CrossRef](#)]
20. Wang, S.; Zhao, J.; Ta, N.; Zhao, X.; Xiao, M.; Wei, H. A Real-Time Deep Learning Forest Fire Monitoring Algorithm Based on an Improved Pruned + KD Model. *J. Real-Time Image Proc.* **2021**, *18*, 2319–2329. [[CrossRef](#)]
21. Lou, L.; Chen, F.; Cheng, P.; Huang, Y. Smoke Root Detection from Video Sequences Based on Multi-Feature Fusion. *J. For. Res.* **2022**, *33*, 1841–1856. [[CrossRef](#)]
22. Chen, T.-H.; Wu, P.-H.; Chiou, Y.-C. An Early Fire-Detection Method Based on Image Processing. In Proceedings of the 2004 International Conference on Image Processing, 2004, ICIP '04, Singapore, 24–27 October 2004; Volume 3, pp. 1707–1710.
23. Çelik, T.; Demirel, H. Fire Detection in Video Sequences Using a Generic Color Model. *Fire Saf. J.* **2009**, *44*, 147–158. [[CrossRef](#)]
24. Emmy Prema, C.; Vinsley, S.S.; Suresh, S. Multi Feature Analysis of Smoke in YUV Color Space for Early Forest Fire Detection. *Fire Technol.* **2016**, *52*, 1319–1342. [[CrossRef](#)]
25. Di Lascio, R.; Greco, A.; Saggese, A.; Vento, M. Improving Fire Detection Reliability by a Combination of Videoanalytics. In *Image Analysis and Recognition, Proceedings of the 11th International Conference, ICIAR 2014, Vilamoura, Portugal, 22–24 October 2014*; Campilho, A., Kamel, M., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 477–484.
26. Foggia, P.; Saggese, A.; Vento, M. Real-Time Fire Detection for Video-Surveillance Applications Using a Combination of Experts Based on Color, Shape, and Motion. *IEEE Trans. Circuits Syst. Video Technol.* **2015**, *25*, 1545–1556. [[CrossRef](#)]
27. Borges, P.V.K.; Izquierdo, E. A Probabilistic Approach for Vision-Based Fire Detection in Videos. *IEEE Trans. Circuits Syst. Video Technol.* **2010**, *20*, 721–731. [[CrossRef](#)]
28. Rudz, S.; Chetehouna, K.; Hafiane, A.; Sero-Guillaume, O.; Laurent, H. On the Evaluation of Segmentation Methods for Wildland Fire. In *Advanced Concepts for Intelligent Vision Systems, Proceedings of the 11th International Conference, ACIVS 2009 Bordeaux, France, 28 September–2 October 2009*; Blanc-Talon, J., Philips, W., Popescu, D., Scheunders, P., Eds.; Springer: Berlin/Heidelberg, Germany, 2009; pp. 12–23.
29. Matlani, P.; Shrivastava, M. An Efficient Algorithm Proposed for Smoke Detection in Video Using Hybrid Feature Selection Techniques. *Eng. Technol. Appl. Sci. Res.* **2019**, *9*, 3939–3944. [[CrossRef](#)]
30. Dimitropoulos, K.; Barmpoutis, P.; Grammalidis, N. Spatio-Temporal Flame Modeling and Dynamic Texture Analysis for Automatic Video-Based Fire Detection. *IEEE Trans. Circuits Syst. Video Technol.* **2015**, *25*, 339–351. [[CrossRef](#)]
31. Töreyin, B.U.; Dedeoğlu, Y.; Güdükbay, U.; Çetin, A.E. Computer Vision Based Method for Real-Time Fire and Flame Detection. *Pattern Recognit. Lett.* **2006**, *27*, 49–58. [[CrossRef](#)]
32. Günay, O.; Taşdemir, K.; Uğur Töreyin, B.; Çetin, A.E. Fire Detection in Video Using LMS Based Active Learning. *Fire Technol.* **2010**, *46*, 551–577. [[CrossRef](#)]
33. Wang, D.; Cui, X.; Park, E.; Jin, C.; Kim, H. Adaptive Flame Detection Using Randomness Testing and Robust Features. *Fire Saf. J.* **2013**, *55*, 116–125. [[CrossRef](#)]
34. Guan, Z.; Miao, X.; Mu, Y.; Sun, Q.; Ye, Q.; Gao, D. Forest Fire Segmentation from Aerial Imagery Data Using an Improved Instance Segmentation Model. *Remote Sens.* **2022**, *14*, 3159. [[CrossRef](#)]
35. Zhou, X.; Mahalingam, S.; Weise, D. Modeling of Marginal Burning State of Fire Spread in Live Chaparral Shrub Fuel Bed. *Combust. Flame* **2005**, *143*, 183–198. [[CrossRef](#)]
36. Liu, N.; Zhang, S.; Luo, X.; Lei, J.; Chen, H.; Xie, X.; Zhang, L.; Tu, R. Interaction of Two Parallel Rectangular Fires. *Proc. Combust. Inst.* **2019**, *37*, 3833–3841. [[CrossRef](#)]
37. Çolak, E.; Sunar, F. Evaluation of Forest Fire Risk in the Mediterranean Turkish Forests: A Case Study of Menderes Region, Izmir. *Int. J. Disaster Risk Reduct.* **2020**, *45*, 101479. [[CrossRef](#)]
38. Li, P.; Zhao, W. Image Fire Detection Algorithms Based on Convolutional Neural Networks. *Case Stud. Therm. Eng.* **2020**, *19*, 100625. [[CrossRef](#)]
39. Zhang, Y.; Sun, Y.; Wang, Z.; Jiang, Y. YOLOv7-RAR for Urban Vehicle Detection. *Sensors* **2023**, *23*, 1801. [[CrossRef](#)]
40. Sun, X.; Sun, L.; Huang, Y. Forest Fire Smoke Recognition Based on Convolutional Neural Network. *J. For. Res.* **2021**, *32*, 1921–1927. [[CrossRef](#)]
41. Li, T.; Zhu, H.; Hu, C.; Zhang, J. An Attention-Based Prototypical Network for Forest Fire Smoke Few-Shot Detection. *J. For. Res.* **2022**, *33*, 1493–1504. [[CrossRef](#)]
42. Kim, B.; Lee, J. A Video-Based Fire Detection Using Deep Learning Models. *Appl. Sci.* **2019**, *9*, 2862. [[CrossRef](#)]
43. Lee, Y.; Shim, J. False Positive Decremented Research for Fire and Smoke Detection in Surveillance Camera Using Spatial and Temporal Features Based on Deep Learning. *Electronics* **2019**, *8*, 1167. [[CrossRef](#)]

44. Wang, Z.; Wang, Z.; Zhang, H.; Guo, X. A Novel Fire Detection Approach Based on CNN-SVM Using Tensorflow. In *Intelligent Computing Methodologies, Proceedings of the 13th International Conference, ICIC 2017, Liverpool, UK, 7–10 August 2017*; Huang, D.-S., Hussain, A., Han, K., Gromiha, M.M., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 682–693.
45. Frizzi, S.; Kaabi, R.; Bouchouicha, M.; Ginoux, J.-M.; Moreau, E.; Fnaiech, F. Convolutional Neural Network for Video Fire and Smoke Detection. In Proceedings of the IECON 2016—42nd Annual Conference of the IEEE Industrial Electronics Society, Florence, Italy, 23–26 October 2016; pp. 877–882.
46. Liu, Z.; Zhang, K.; Wang, C.; Huang, S. Research on the Identification Method for the Forest Fire Based on Deep Learning. *Optik* **2020**, *223*, 165491. [\[CrossRef\]](#)
47. Guo, Y.-Q.; Chen, G.; Wang, Y.-N.; Zha, X.-M.; Xu, Z.-D. Wildfire Identification Based on an Improved Two-Channel Convolutional Neural Network. *Forests* **2022**, *13*, 1302. [\[CrossRef\]](#)
48. Qian, J.; Lin, H. A Forest Fire Identification System Based on Weighted Fusion Algorithm. *Forests* **2022**, *13*, 1301. [\[CrossRef\]](#)
49. Xie, C.; Tao, H. Generating Realistic Smoke Images with Controllable Smoke Components. *IEEE Access* **2020**, *8*, 201418–201427. [\[CrossRef\]](#)
50. Ding, Z.; Zhao, Y.; Li, A.; Zheng, Z. Spatial-Temporal Attention Two-Stream Convolution Neural Network for Smoke Region Detection. *Fire* **2021**, *4*, 66. [\[CrossRef\]](#)
51. Yang, J.; Chen, Y. Tender Leaf Identification for Early-Spring Green Tea Based on Semi-Supervised Learning and Image Processing. *Agronomy* **2022**, *12*, 1958. [\[CrossRef\]](#)
52. Wu, X.; Lu, X.; Leung, H. An Adaptive Threshold Deep Learning Method for Fire and Smoke Detection. In Proceedings of the 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Banff, AB, Canada, 5–8 October 2017; pp. 1954–1959.
53. Wang, Y.; Dang, L.; Ren, J. Forest Fire Image Recognition Based on Convolutional Neural Network. *J. Algorithms Comput. Technol.* **2019**, *13*, 1748302619887689. [\[CrossRef\]](#)
54. Zheng, Z.; Zhao, Y.; Li, A.; Yu, Q. Wild Terrestrial Animal Re-Identification Based on an Improved Locally Aware Transformer with a Cross-Attention Mechanism. *Animals* **2022**, *12*, 3503. [\[CrossRef\]](#)
55. Zheng, X.; Chen, F.; Lou, L.; Cheng, P.; Huang, Y. Real-Time Detection of Full-Scale Forest Fire Smoke Based on Deep Convolution Neural Network. *Remote Sens.* **2022**, *14*, 536. [\[CrossRef\]](#)
56. Zhao, Y.; Ma, J.; Li, X.; Zhang, J. Saliency Detection and Deep Learning-Based Wildfire Identification in UAV Imagery. *Sensors* **2018**, *18*, 712. [\[CrossRef\]](#)
57. Yang, F.; Jiang, Y.; Xu, Y. Design of Bird Sound Recognition Model Based on Lightweight. *IEEE Access* **2022**, *10*, 85189–85198. [\[CrossRef\]](#)
58. Yang, J.; Yang, J.; Zhang, D.; Lu, J. Feature Fusion: Parallel Strategy vs. Serial Strategy. *Pattern Recognit.* **2003**, *36*, 1369–1381. [\[CrossRef\]](#)
59. Liu, J.; Fan, X.; Jiang, J.; Liu, R.; Luo, Z. Learning a Deep Multi-Scale Feature Ensemble and an Edge-Attention Guidance for Image Fusion. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 105–119. [\[CrossRef\]](#)
60. Chaib, S.; Liu, H.; Gu, Y.; Yao, H. Deep Feature Fusion for VHR Remote Sensing Scene Classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 4775–4784. [\[CrossRef\]](#)
61. Zeng, N.; Wu, P.; Wang, Z.; Li, H.; Liu, W.; Liu, X. A Small-Sized Object Detection Oriented Multi-Scale Feature Fusion Approach with Application to Defect Detection. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 3507014. [\[CrossRef\]](#)
62. Lu, J.; Behbood, V.; Hao, P.; Zuo, H.; Xue, S.; Zhang, G. Transfer Learning Using Computational Intelligence: A Survey. *Knowl.-Based Syst.* **2015**, *80*, 14–23. [\[CrossRef\]](#)
63. Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A Comprehensive Survey on Transfer Learning. *Proc. IEEE* **2021**, *109*, 43–76. [\[CrossRef\]](#)
64. Li, J.; Huang, R.; He, G.; Liao, Y.; Wang, Z.; Li, W. A Two-Stage Transfer Adversarial Network for Intelligent Fault Diagnosis of Rotating Machinery with Multiple New Faults. *IEEE-ASME Trans. Mechatron.* **2021**, *26*, 1591–1601. [\[CrossRef\]](#)
65. Kaya, A.; Keceli, A.S.; Catal, C.; Yalic, H.Y.; Temucin, H.; Tekinerdogan, B. Analysis of Transfer Learning for Deep Neural Network Based Plant Classification Models. *Comput. Electron. Agric.* **2019**, *158*, 20–29. [\[CrossRef\]](#)
66. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
67. Wang, S.-H.; Fernandes, S.L.; Zhu, Z.; Zhang, Y.-D. AVNC: Attention-Based VGG-Style Network for COVID-19 Diagnosis by CBAM. *IEEE Sens. J.* **2022**, *22*, 17431–17438. [\[CrossRef\]](#)
68. Zhao, Y.; Zhao, L.; Xiong, B.; Kuang, G. Attention Receptive Pyramid Network for Ship Detection in SAR Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 2738–2756. [\[CrossRef\]](#)
69. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [\[CrossRef\]](#)
70. Pan, H.; Badawi, D.; Cetin, A.E. Computationally Efficient Wildfire Detection Method Using a Deep Convolutional Network Pruned via Fourier Analysis. *Sensors* **2020**, *20*, 2891. [\[CrossRef\]](#) [\[PubMed\]](#)
71. Singh, P.; Verma, A.; Alex, J.S.R. Disease and Pest Infection Detection in Coconut Tree through Deep Learning Techniques. *Comput. Electron. Agric.* **2021**, *182*, 105986. [\[CrossRef\]](#)
72. Arora, V.; Ng, E.Y.-K.; Leekha, R.S.; Darshan, M.; Singh, A. Transfer Learning-Based Approach for Detecting COVID-19 Ailment in Lung CT Scan. *Comput. Biol. Med.* **2021**, *135*, 104575. [\[CrossRef\]](#)

73. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
74. 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, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
75. Yin, M.; Lang, C.; Li, Z.; Feng, S.; Wang, T. Recurrent Convolutional Network for Video-Based Smoke Detection. *Multimed. Tools Appl.* **2019**, *78*, 237–256. [[CrossRef](#)]
76. Zhang, J.; Zhu, H.; Wang, P.; Ling, X. ATT Squeeze U-Net: A Lightweight Network for Forest Fire Detection and Recognition. *IEEE Access* **2021**, *9*, 10858–10870. [[CrossRef](#)]
77. 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](#)]

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.