



Article

Adversarial Fusion Network for Forest Fire Smoke Detection

Tingting Li ^{1,2} , Changchun Zhang ^{1,2}, Haowei Zhu ^{1,2} and Junguo Zhang ^{1,2,*} 

¹ School of Technology, Beijing Forestry University, Beijing 100083, China; litingting@bjfu.edu.cn (T.L.); zhangchangchun@bjfu.edu.cn (C.Z.); haoweiz@bjfu.edu.cn (H.Z.)

² BFU Research Center for Biodiversity Intelligent Monitoring, Beijing 100083, China

* Correspondence: zhangjunguo@bjfu.edu.cn; Tel.: +86-010-6233-7736

Abstract: Recent advances suggest that deep learning has been widely used to detect smoke for early forest fire warnings. Despite its remarkable success, this approach has a number of problems in real life application. Deep neural networks only learn deep and abstract representations, while ignoring shallow and detailed representations. In addition, previous models have been trained on source domains but have generalized weakly on unseen domains. To cope with these problems, in this paper, we propose an adversarial fusion network (AFN), including a feature fusion network and an adversarial feature-adaptation network for forest fire smoke detection. Specifically, the feature fusion network is able to learn more discriminative representations by fusing abstract and detailed features. Meanwhile, the adversarial feature adaptation network is employed to improve the generalization ability and transfer gains of the AFN. Comprehensive experiments on two self-built forest fire smoke datasets, and three publicly available smoke datasets, validate that our method significantly improves the performance and generalization of smoke detection, particularly the accuracy of the detection of small amounts of smoke.

Keywords: forest fire smoke detection; adversarial feature adaptation; shallow network; feature fusion network; attention mechanism



Citation: Li, T.; Zhang, C.; Zhu, H.; Zhang, J. Adversarial Fusion Network for Forest Fire Smoke Detection. *Forests* **2022**, *13*, 366. <https://doi.org/10.3390/f13030366>

Academic Editor: David R. Weise

Received: 6 January 2022

Accepted: 21 February 2022

Published: 22 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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

Smoke detection has attracted much attention, as it is an important element of early fire warning and fire prevention [1]. Depending on the acquisition level, there are three widely used smoke monitoring systems, namely satellite, terrestrial, and aerial. Terrestrial and aerial systems tend to be more efficient than satellite systems in terms of resolution and response time to early wildfire incidents [2,3]. Therefore, our work focuses mainly on RGB smoke images captured by terrestrial systems and aerial systems. Earlier image-based smoke detection methods [4,5] rely on specialist prior knowledge to extract detailed and shallow features such as color, texture, and contour. Although shallow representation has a pivotal role in smoke detection, it is difficult for supervised classifiers trained with these features to simultaneously guarantee high accuracy and a low false alarm rates [6]. Recently, deep fire-smoke detection methods [7–9] have significantly boosted detection performance. These methods automatically extract deep and abstract smoke features to achieve highly accurate end-to-end fire smoke detection.

Despite deep detection methods enjoying a promising performance in benchmark smoke datasets, these methods may fail in the face of complicated forest environments. Specifically, some disturbances, such as the mistaken identification of clouds, fog, and haze, cause false alarms due to their similarity to smoke in their color and contours [10]. Deep convolutional neural networks (CNNs) are not suitable for complex forest scenarios, as they mainly focus on abstract features and ignore detailed features. Moreover, ImageNet-trained CNNs tend to classify objects by using local textures instead of global object contours and color [11,12]. To extract more discriminative smoke features, many conventional

strategies, combined with deep-learning methods [8,9] and dual-channel convolutional neural networks [13–15], have been proposed for fire smoke detection.

Another key issue with smoke detection is that performance is considerably degraded in forest environments that are different from the training set [16]. The success of these deep methods can be partially attributed to a large number of annotated data [17]. However, forest fire smoke images or videos are very hard to capture in real life. There are also few forest fire smoke images in benchmark datasets. Therefore, the above methods are not particularly suitable for forest scenarios with much fewer training samples, due to the domain shift on visual features [18]. In order to alleviate the weak generalization in domains other than training, several transfer-learning methods have been proposed. Many deep-learning smoke detection methods learn features by directly employing ImageNet-trained CNNs, and fine-tune the weights on the smoke training dataset [19]. Domain adaptation (DA) is a commendable learning paradigm for tackling the above issue. DA methods aim at reducing the difference between the covariance matrices of source and target domains.

To further improve the performance and robustness of forest fire smoke detection methods, motivated by domain-adversarial neural networks [20], we propose a dual-channel convolutional neural network with domain-adversarial training named the adversarial fusion network (AFN). Firstly, feature fusion network, which contains a densely dilated convolutional network (DDCN) and an attention-based skip connection network (ASCN), is designed to bring down high false alarm rates. The DDCN is used to generate deep and abstract features, while the ASCN specializes in extracting shallow and detailed features. Specifically, to obtain better receptive fields, we select dilated convolutional instead of common convolutional in the DDCN. The multiscale-channel attention module (MSCAM) [21] and skip connection in the ASCN are adopted to improve the representation capacity of global information. In this way, the fused features of DDCN and ASCN are more discriminative. Secondly, we introduce the adversarial feature adaptation network, which narrows the differences between the source and target domains to alleviate the domain shift. Moreover, the mixed dataset of base smoke and stylized data is utilized as the training dataset for AFN, to increase the shape bias of learning feature representations. Extensive experiments demonstrate the effectiveness and generalization of our method on two self-built forest fire smoke datasets and three publicly available smoke datasets.

2. Related Work

2.1. Fire Smoke Detection

Previous image-based fire smoke detection methods train supervised classifiers using manually extracted features [22–24]. These models greatly depend on feature selection to improve detection performance. These methods can be applied well to fixed scenarios, but have weak generalization to changing environments. Recently, deep fire smoke detection methods have achieved superior performance via generation of deep and abstract features. Specifically, Mao et al. [8] proposed a novel fire smoke recognition method based on a multichannel convolutional neural network, to overcome the deficiencies of machine-learning-based methods. A novel deep normalization and convolutional neural network (DNCNN) with 14 layers was introduced in [9] to enable end-to-end automated smoke feature extraction and detection. Although these methods have improved the accuracy of smoke detection compared with conventional methods, they probably fail against a complex forest environment. To extract more discriminative smoke features, many dual-channel frameworks have been developed. Gu et al. [15] designed a new deep dual-channel neural network (DCNN) for smoke detection, of which the first subnetwork is good at extracting the detail information, and the second subnetwork can capture the base information. In [14], a dual-channel convolutional neural network (DC-CNN) using transfer learning for detecting smoke images was proposed. These methods produce state-of-the-art results on public smoke datasets.

2.2. Attention Mechanism

Attention mechanisms enhance the discriminability of network representations through focusing on important features and ignoring unnecessary information, which is inspired by the human visual perception process [25]. The attention mechanism was first applied to natural language processing (NLP) in [26], and is now widely used in computer vision tasks. For example, Hu et al. [27] concentrated on the channel relationship and proposed the squeeze-and-excitation (SE) block, which can improve the representational power of a network and bring significant performance improvements without increasing the computational cost. To capture more sophisticated channel-wise dependencies, a number of improved SE blocks have been proposed [28–30]. Moreover, many researchers have combined spatial attention and channel attention to design more sophisticated attention modules. Woo et al. [31] presented the CBAM, a new efficient architecture that uses both spatial and channel-wise attention. A convolutional-triplet attention module (CTAM), capable of producing cross-dimensional interactions between spatial attention and channel attention, was proposed by Misra et al. [32]. Recently, researchers have started to take the scale issue of attention modules into account, in order to improve the feature discriminability. In [21], a multiscale-channel attention module (MS-CAM) that aggregates local and global feature contexts inside the module was proposed. Experiments on different benchmark datasets have demonstrated the superiority of these methods over using only channel-wise attention.

2.3. Domain Adaptation

Domain adaptation (DA) aims to diminish the difference between source and target domains in the feature space [33]. Some methods have used maximum mean difference (MMD) to mitigate domain shift [34], whereas others have employed an adversarial approach to reducing domain shift. Domain-adversarial networks, which consist of a feature generator, a classifier and a discriminator, have been widely applied in computer vision tasks. They have achieved great success, and have effectively alleviated the consumption of training sample annotation [35–37]. Tzeng et al. [38] introduced an adaptation layer and an additional domain-confusion loss to learn a representation that is both semantically discriminative and domain invariant. WDGRL [39] was proposed to learn domain invariant features by introducing Wasserstein Distance with adversarial training. Different from the application of the above feature adaptation methods, Brochu et al. [40] adopted domain-adversarial networks to improve the contour bias and model generalization capabilities of CNNs. Recently, DA has also been applied to fire smoke detection [41,42]. To extract a powerful feature representation of smoke, a deep architecture based on domain adaptation, to confuse the distributions of features extracted from synthetic and real smoke images, was presented by Xu et al. [43].

3. Materials and Methods

In this section, we elaborate on the details of the adversarial fusion network (AFN) for forest fire smoke detection, and we describe the experimental setting, including two self-built forest fire smoke datasets, implementation details, and evaluation metrics.

The general framework of AFN is shown in Figure 1. Our method consists of three components: feature fusion network G_f , label predictor G_l , and adversarial feature-adaptation network G_d . Firstly, we present a dual-channel feature fusion network constructed from a densely dilated convolutional network (DDCN) and an attention-based skip connection network (ASCN). Secondly, we introduce the adversarial feature adaptation network, training the base dataset and the stylized dataset to mitigate the discrepancy between the two domains. Finally, an efficient iterative optimization method is proposed. The goal of our method is to improve the robustness of the detection network via learning more image representations.

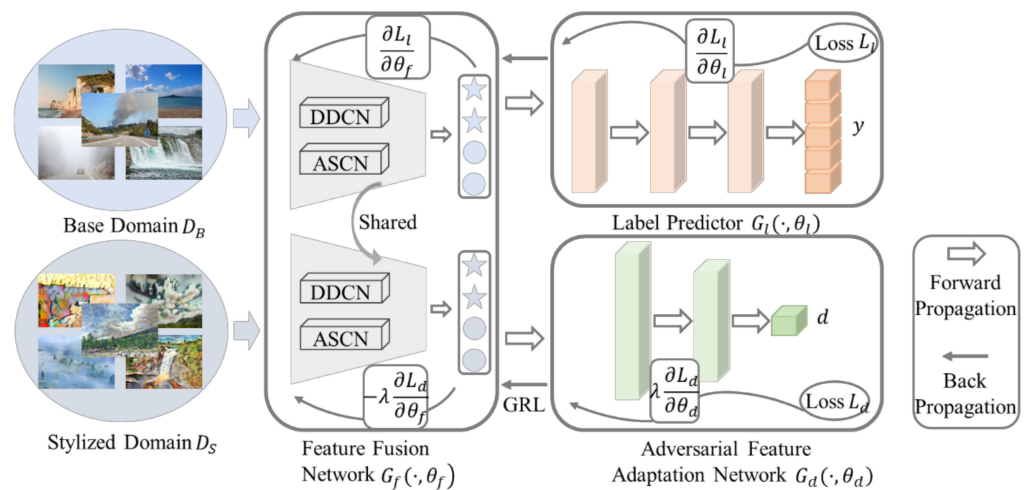


Figure 1. The whole architecture of our method.

Given a labeled smoke dataset as the base domain $(D_B)^{n_B} \sim \{(x_i, y_i)\}_{i=1}^{n_B}$, where x_i indicates the input image, $y_i = \{0, 1, \dots, m-1\}$ represents the corresponding label, and n_B and m denote the total number of samples and the number of classes, respectively. The stylized smoke dataset is treated as the stylized domain $(D_S)^{n_S} \sim \{x_j\}_{j=1}^{n_S}$. Note that the base domain and stylized domain are drawn from different joint distribution and share the same label space. DDCN and ASCN extract abstract features F_1 and detailed features F_2 across domains. Then, the fused features $G_f(x, \theta_f)$ of F_1 and F_2 are delivered to the label predictor $G_l(\cdot, \theta_l)$ and adversarial feature adaptation network $G_d(\cdot, \theta_d)$ for class prediction y and domain classification d , respectively.

3.1. Feature Fusion Network

The feature fusion network is a dual-channel CNN which contains both the DDCN and the ASCN. The DDCN is employed to extract deep and abstract features from smoke images. The backbone of the DDCN is ImageNet-trained Densenet-169 [44]. This backbone not only has the advantage of alleviating vanishing-gradient and enhancing feature propagation, but also achieves significant improvements over the state-of-the-art in various computer vision tasks [45,46]. Moreover, to expand the receptive field without increasing the computational cost, the original convolutional layers are replaced by dilated convolutional layers. However, ImageNet-trained CNNs are biased towards learning texture features rather than color and counters [47]. It is necessary to learn shallow and detailed representation by adding a shallow neural network.

The ASCN specializes in extracting shallow and detailed features, such as color and contours. Inspired by AlexNet [48] and the original domain-adversarial neural network [20], we build a new convolutional neural network consisting of five convolutional layers, two batch normalization layers, five activation function layers, three max-pooling layers, and an average-pooling layer. A batch normalization layer replaces the local normalization layer in AlexNet to accelerate convergence and prevent overfitting. The activation function is a nonlinear function called the rectified linear unit (ReLU). In addition, the multiscale-channel attention module (MS-CAM) [21] and skip connection in ASCN act to improve the representation capacity of global information and share feature information, respectively.

The DDCN takes a gray-scale image of x as input and the ASCN uses the original RGB image x as input. The feature connection of two subnetworks is the output $G_f(x, \theta_f)$ of the feature fusion network, where θ_f denotes the learning parameter. The base sample feature representation is $B(G_f) = \{G_f(x_i, \theta_f) | x_i \in D_B\}$ and the stylize sample feature representation $S(G_f) = \{G_f(x_j, \theta_f) | x_j \in D_S\}$.

The label predictor is similar to the label predictor of the original domain-adversarial neural network. We also add dropout regularization after each activation function layer to prevent overfitting. The output $G_l(G_f(x, \theta_f), \theta_l)$ of the label predictor is the probability of detection of x . The classification loss is calculated with negative log-probability and is expressed as follows:

$$L_l(G_l(G_f(x, \theta_f), \theta_l), y) = \log \frac{1}{G_l(G_f(x, \theta_f), \theta_l)_y}. \tag{1}$$

3.2. Adversarial Feature Adaptation Network

A domain-adversarial training network is employed to further increase contour-learning capability and the generalization capabilities of the smoke detection network. This method not only improves the representation capabilities of our network, but also forces the network to integrate spatial information over long distances. Inspired by the Proxy A -distance [49], we learn an adversarial feature adaptation network that contains two fully connected layers, a batch normalization layer, an activation function layer, and a classifier layer. The activation function and classifier are the same as the label predictor. The output $G_d(G_f(x, \theta_f), \theta_d)$ of the adversarial feature adaptation network is the probability that x comes from the base domain or stylized domain. The domain classification loss is calculated by the focal loss [50] function and is expressed as follows:

$$L_d(G_d(G_f(x, \theta_f), \theta_d), d) = d\alpha(1 - G_d)^\gamma \log \frac{1}{G_d} + (1 - d)(1 - \alpha)G_d^\gamma \log \frac{1}{1 - G_d}, \tag{2}$$

where $\gamma \geq 0$ denotes the focusing parameter, α denotes the balance factor, and d denotes a binary variable. If $d = 0$, x belongs to the base domain, and vice versa, $d = 1$ belongs to the stylized domain. The objective function of adversarial feature adaptation is as follows:

$$E(\theta_f) = \max_{\theta_d} \left[-\frac{1}{n_B} \sum_{i=1}^{n_B} L_d^i(G_d(G_f(x_i, \theta_f), \theta_d), d_i) - \frac{1}{n_S} \sum_{j=1}^{n_S} L_d^j(G_d(G_f(x_j, \theta_f), \theta_d), d_j) \right]. \tag{3}$$

3.3. Model Optimization

The total loss of the AFN consists of the label predictor loss and the adversarial feature adaptation network loss. Therefore, the complete objective function of our algorithm is as follows:

$$\begin{aligned} E(\theta_f, \theta_l, \theta_d) &= \frac{1}{n_B} \sum_{i=1}^{n_B} L_l^i(G_l(G_f(x_i, \theta_f), \theta_l), y_i) - \lambda E(\theta_f, \theta_d) \\ &= \frac{1}{n_B} \sum_{i=1}^{n_B} L_l^i(G_l(G_f(x_i, \theta_f), \theta_l), y_i) - \lambda \left(\frac{1}{n_B} \sum_{i=1}^{n_B} L_d^i(G_d(G_f(x_i, \theta_f), \theta_d), d_i) + \frac{1}{n_S} \sum_{j=1}^{n_S} L_d^j(G_d(G_f(x_j, \theta_f), \theta_d), d_j) \right), \end{aligned} \tag{4}$$

where the parameters of the label predictor are updated via the minimization objective function, and the parameters of the domain classifier are updated via the maximization objective function:

$$(\hat{\theta}_f, \hat{\theta}_l) = \underset{\theta_f, \theta_l}{\operatorname{argmin}} E(\theta_f, \theta_l, \hat{\theta}_d), \tag{5}$$

$$(\hat{\theta}_d) = \underset{\theta_d}{\operatorname{argmax}} E(\hat{\theta}_f, \hat{\theta}_l, \theta_d). \tag{6}$$

To optimize the above objective function, stochastic gradient descent (SGD) is used to update the learning parameters with the following equations:

$$\theta_f \leftarrow \theta_f - \mu \left(\frac{\partial L_l^i}{\partial \theta_f} - \lambda \frac{\partial L_{d_i}^i}{\partial \theta_f} - \lambda \frac{\partial L_{d_j}^j}{\partial \theta_f} \right), \tag{7}$$

$$\theta_l \leftarrow \theta_l - \mu \frac{\partial L_l^i}{\partial \theta_l}, \quad (8)$$

$$\theta_d \leftarrow \theta_d - \mu \lambda \frac{\partial L_d^i}{\partial \theta_d} - \mu \lambda \frac{\partial L_d^j}{\partial \theta_d}, \quad (9)$$

where μ denotes the learning rate. Moreover, a gradient reversal layer (GRL) is inserted between the feature extractor and the adversarial feature adaptation network to implement the adversarial. GRL acts as an identity switch in forward propagation, while in backward propagation, it changes the gradient sign by multiplying by $-\eta$. The forward and backward propagation “pseudo-functions” of the GRL [20] are defined as follows:

$$R(x) = x, \quad (10)$$

$$\frac{dR}{dx} = -\eta I, \quad (11)$$

where I denotes the identity matrix. For this reason, the objective function of our algorithm is reformulated as follows:

$$\tilde{E}(\theta_f, \theta_l, \theta_d) = \frac{1}{n_B} \sum_{i=1}^{n_B} L_l^i(G_l(G_f(x_i, \theta_f), \theta_l), y_i) - \lambda \left(\frac{1}{n_B} \sum_{i=1}^{n_B} L_d^i(G_d(R(G_f(x_i, \theta_f))), \theta_d), d_i) + \frac{1}{n_S} \sum_{j=1}^{n_S} L_d^j(G_d(R(G_f(x_j, \theta_f))), \theta_d), d_j) \right). \quad (12)$$

3.4. Experimental Setting

3.4.1. Forest Fire Smoke Dataset

The forest fire smoke dataset (FF_Smoke dataset) is one of our self-built forest fire smoke datasets. The FF_Smoke dataset is a mixed dataset from the base domain and stylized domain. The base domain D_B contains 5000 RGB images from surveillance cameras, publicly available wildfire smoke datasets, and the web, consisting of five categories: smoke, cloud, fog, trees, and cliffs. Each category contains 1000 images of dimensions 224×224 . In addition, AdaIN style transfer [51] is applied to construct a stylized smoke dataset as the stylized domain D_S . We replace the texture of original images with ten randomly selected painting styles, which are derived from the Kaggle’s Painter by Numbers dataset [52]. The advantages of this stylization method are the variety of stylized images that can be created and the speed of transformation. The base dataset and the stylized dataset are randomly divided into a training set and a test set in the ratio of 8:2, respectively. Figure 2 shows examples from the base domain and the stylized domain. It can be seen that the stylized images are able to retain the global contours.

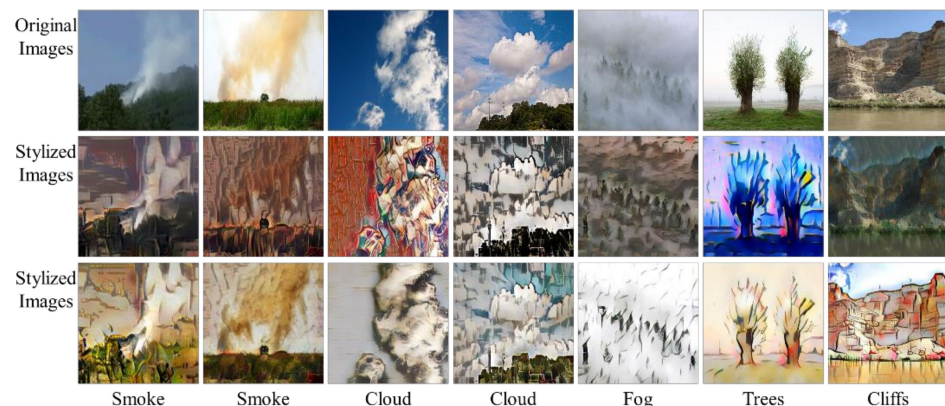


Figure 2. Sample images from the base domain and the stylized domain.

3.4.2. Early Wildfire Surveillance Dataset

To further evaluate the robustness and generalization capabilities of our method, an early forest fire surveillance dataset (EWS_Smoke dataset) is constructed that has never been seen in the FF_Smoke dataset. Early forest fire smoke is often captured by long-range cameras and accounts for only a small portion of the image. This dataset contains 221 smoke images from different surveillance sites. The sample images are shown in Figure 3.

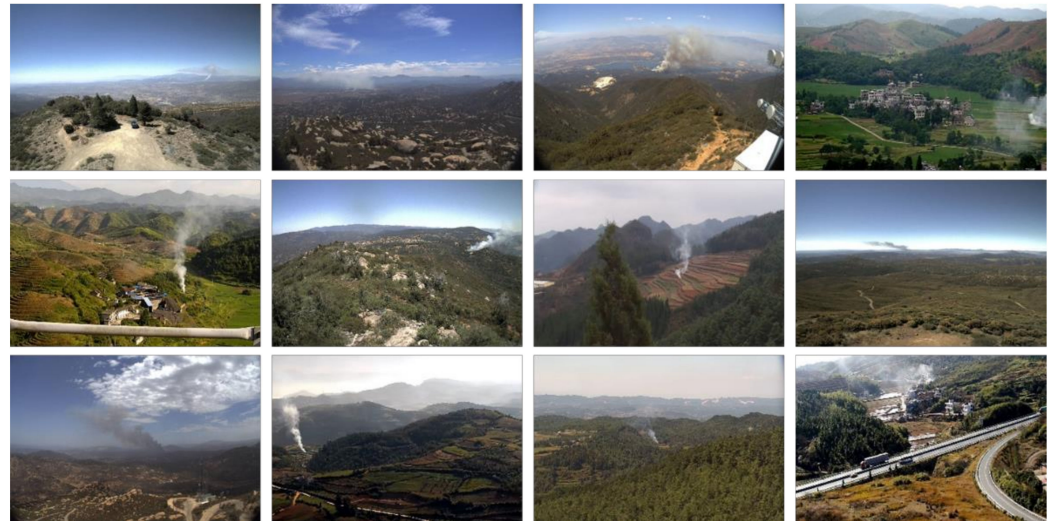


Figure 3. Sample images from real surveillance video.

3.4.3. Implementation Details and Evaluation Metrics

We implement our model with PyTorch and train all models on a single NVIDIA GeForce RTX 2080ti GPU device. The label predictor of our method contains three fully connected layers, two batch normalization layers, two activation function layers and a classifier layer, where the activation function and classifier are ReLU and softmax. The model is trained via Adam optimizer for 300 epochs using a cosine decay learning rate, with a batch size of 64, and an initial learning rate that subsequently decreases by 0.1 every 30 epochs. Augmentation and regularization strategies are then adopted to mitigate overfitting in training.

To fairly evaluate the performance of our method, accuracy rate (AR), false alarm rate (FAR), detection rate (DR), recall rate (RR), and F1-score (F1) [8] are adopted as evaluation criterion for forest fire smoke detection.

4. Results and Discussion

In this section, the performance of our method is first evaluated on the FF_Smoke dataset. Then, the robustness and generalization of our method is evaluated on the EWS_Smoke dataset and three real-world public smoke datasets. Finally, the effectiveness of the key components of our method is verified.

4.1. Main Results

4.1.1. Selection of Backbone Networks

We analyze the effect of color feature on the performance of ImageNet-trained CNNs. To eliminate color information, RGB images from the FF_Smoke dataset were converted to gray-scale images. Pre-trained CNNs were fine-tuned on original images and grayscale images, respectively, and then tested on original images. The implementation details of the two experiments are completely consistent. The ARs of 100 tests are shown in Table 1. The numbers in parentheses are the increase or decrease in AR compared to original image training. It can be seen that AlexNet (84.78%) and DenseNet-169 (79.45%) achieved the highest top-1 AR on the original dataset and gray dataset, respectively. In addition, lacking

color information deteriorates networks performance and results in a maximum difference in AR of 7.21%.

Table 1. Detection accuracy of pre-trained convolutional neural networks on FF_Smoke dataset.

Model	RGB–RGB		Gray–RGB	
	Top-1 (%)	Top-3 (%)	Top-1 (%)	Top-3 (%)
AlexNet	84.78 ± 0.86	97.85 ± 0.44	77.59 ± 1.20 (−7.19)	96.27 ± 0.62 (−1.58)
ResNet-50	83.99 ± 0.42	96.41 ± 0.39	77.45 ± 0.88 (−6.54)	90.48 ± 0.81 (−5.93)
ResNet-101	83.92 ± 0.47	95.58 ± 0.61	79.00 ± 1.12 (−4.92)	93.64 ± 0.64 (−1.94)
DenseNet-121	83.05 ± 0.42	95.63 ± 0.39	75.84 ± 1.20 (−7.21)	91.82 ± 0.59 (−3.81)
DenseNet-169	84.04 ± 0.51	96.53 ± 0.41	79.45 ± 0.76 (−4.59)	90.39 ± 0.79 (−6.14)

To further analyze the effect of color information on performance, we also conducted experiments with CNNs trained from scratch. The ARs of 100 tests are shown in Table 2. It is clear that AlexNet (81.45%), trained from scratch, achieved the highest top-1 AR on the original dataset and gray dataset. In addition, the difference in AR can reach a maximum of 26.15%.

Table 2. Detection accuracy of scratch-trained convolutional neural networks on FF_Smoke dataset.

Model	RGB–RGB		Gray–RGB	
	Top-1 (%)	Top-3 (%)	Top-1 (%)	Top-3 (%)
AlexNet	81.45 ± 0.86	97.70 ± 0.34	73.97 ± 1.04 (−7.48)	95.95 ± 0.53 (−1.75)
ResNet-50	79.22 ± 1.27	96.41 ± 0.39	65.30 ± 1.13 (−13.92)	91.89 ± 0.87 (−4.52)
ResNet-101	78.59 ± 1.06	96.72 ± 0.61	59.52 ± 1.29 (−19.07)	90.58 ± 0.75 (−6.14)
DenseNet-121	79.85 ± 0.82	96.20 ± 0.53	53.70 ± 0.87 (−26.15)	85.05 ± 0.69 (−11.15)
DenseNet-169	79.99 ± 0.86	96.14 ± 0.67	55.59 ± 0.89 (−24.40)	83.24 ± 1.14 (−12.90)

For sample-limited datasets, pre-training not only helps CNNs to converge quickly, but also achieves the desired AR. Since the maximum difference number in Table 2 is three times greater than in Table 1, it is assumed that the CNN trained from scratch is more sensitive to color information compared to pre-trained networks. Therefore, we combined the advantages of pre-trained and scratch-trained networks to construct a dual-channel feature extractor based on CNNs. Pre-trained DenseNet-169 was used as the first sub-backbone to extract abstract features, and scratch-trained AlexNet was applied as the second sub-backbone to learn detailed features, such as color and contours.

4.1.2. AFN with Different Loss Functions

We also compared the performance of the AFN with different loss functions. The cross-entropy loss function and the focal loss function were employed for the adversarial feature adaptation network, and the test results are shown in Figure 4. Figure 4a,b show the confusion matrices of the cross-entropy loss function and focal loss function, respectively. It can be seen that the focal loss function not only increases the DR (+4%) of smoke, but it also significantly reduces the FAR (−4%) of smoke. In addition, the DR of fog (+6%), cloud (+16%), and cliffs (+2%) are significantly improved. The focal loss function is proposed to address the problem of imbalance between positive and negative samples and the uneven proportion of hard and easy samples. Therefore, the focal loss function is adopted as the adversarial feature-adaptation network-loss function in this paper.

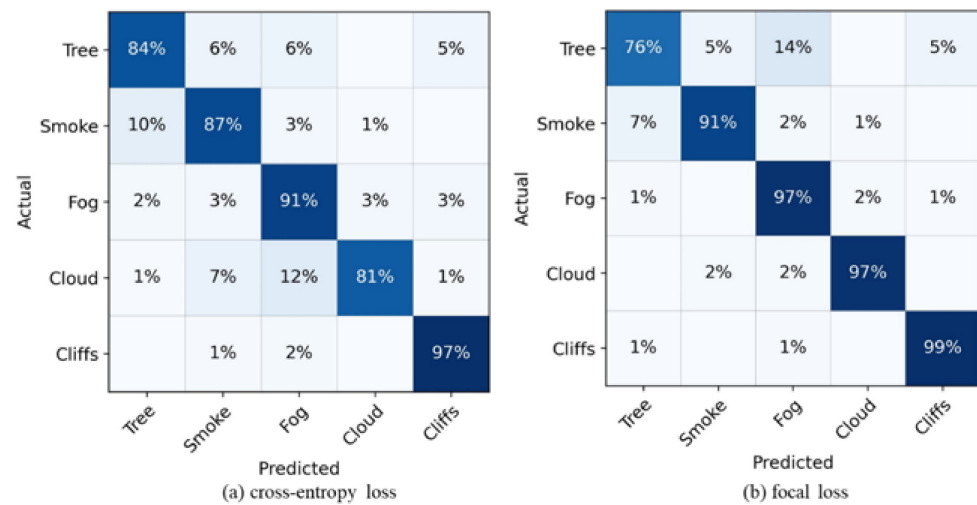


Figure 4. Test results of (a) cross-entropy loss function and (b) focal loss function on FF_Smoke dataset.

4.2. Comparison with State-of-the-Arts

4.2.1. Performance on Self-Built Smoke Datasets

We tested the AFN on the early forest fire surveillance images. The performance of our method is compared with the Dual-Net, DDCN, ASCN, pre-trained DenseNet-169, and scratch-trained AlexNet. Figure 5 shows the AR of 100 tests. The performance of networks with domain-adversarial training is significantly better than that of networks without on different datasets. It can be seen that our proposed algorithm achieved the highest AR (91.85% and 92.96%) on the FF_Smoke dataset and the EWS_Smoke dataset, respectively. The test results present satisfactory performance on EWS_Smoke dataset, demonstrating the decent generalization capability of our proposed algorithm.

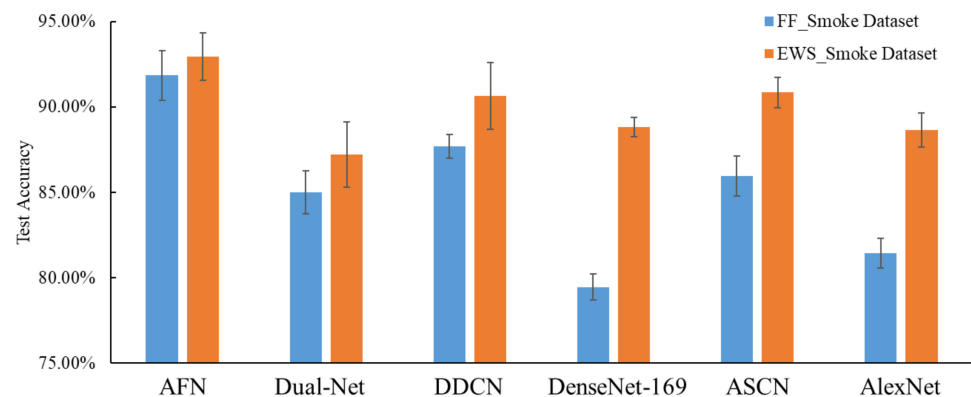


Figure 5. Average test accuracy on self-built smoke datasets. The error line represents the standard deviation.

4.2.2. Performance on Publicly Available Smoke Datasets

To verify the robustness and generalization of our proposed algorithm, we compared it with other state-of-the-art algorithms on three publicly available smoke datasets, including Yuan_Smoke [9], USTC_SmokeRS [53], and Fire_Smoke [54]. The Yuan_Smoke dataset and the Fire_Smoke dataset are all derived from living-fire and urban-fire scenarios, while the USTC_SmokeRS is a satellite imagery smoke dataset covering various areas across the world. Our method was compared with state-of-the-art algorithms on different datasets, respectively. HLTPMC [55] is a conventional machine-learning approach based on hand-crafted features, whereas ZF-Net [56], MCCNN [8], DNCNN [9], and SmokeNet [53] are

end-to-end detection methods based on CNNs. Dual channel CNN [15] is also used to compare performance. The results of the evaluation are shown in Table 3. Our approach achieves the highest AR (99.78%) on the Yuan_Smoke dataset. Although our DR (99.64%) is slightly lower than HLTPMC (99.82%) and MCCNN (99.82%), we have the lowest FAR (0.12%) and are on par with advanced dual-channel CNNs. In addition, RR is also employed to evaluate the performance of our algorithm, and when RR and AR are in conflict, performance is evaluated by the F1. Our algorithm reaches the highest at RR (99.82%) and F1 (99.73%) on the Yuan_Smoke dataset, respectively. It is clear that the FAR (7.59%) of SmokeNet is more than twice that of ours (3.21%) on USTC_SmokeRS dataset. The evaluation results on publicly available smoke datasets achieved a desirable performance, demonstrating the good stability and generalization ability of our method. Our method is not only applicable for forest fire smoke detection, but also for urban fire smoke detection. Moreover, it can also be used for satellite imagery smoke detection.

Table 3. Performance comparison with state-of-the-art methods on publicly available smoke datasets.

Dataset	Model	AR (%)	DR (%)	FAR (%)	RR (%)	F1
Yuan_Smoke [9]	HLTPMC [55]	98.48	99.82	2.41	96.50	98.13
	ZF-Net [56]	97.18	94.02	0.72	98.86	96.38
	MCCNN [8]	99.71	99.82	0.36	99.46	99.64
	DNCNN [9]	97.83	95.29	0.48	99.25	97.23
	DCNN [15]	99.71	99.46	0.12	99.82	99.64
	AFN	99.78	99.64	0.12	99.82	99.73
USTC_SmokeRS [53]	SmokeNet [53]	92.75	94.68	7.59	68.99	79.82
	AFN	96.98	98.07	3.21	84.23	90.62
Fire_Smoke [54]	AFN	96.67	96.00	3.00	94.12	94.05

4.3. Ablation Studies

4.3.1. AFN with Different Attention Modules

To explore the impact of different attention modules on our approach. We compared the performance of the spatial attention module (SAM), the channel attention module (CAM), the squeeze-and-excitation network (SENet) [27], the convolutional-block attention module (CBAM) [31], the criss-cross attention module (CCNet) [30], the convolutional-triplet attention module (CTAM) [32], and the multiscale-channel attention module (MS-CAM) [21]. The AR and loss of 100 tests are shown in Table 4. It is clear to see that the attention module can improve the performance of our method. The AFN with MS-CAM achieved the highest accuracy (91.85%) and the lowest loss (0.44).

Table 4. Performance comparison of different attention modules.

Model	Accuracy		Loss
	Top-1 (%)	Top-3 (%)	
AFN_Without	89.58 ± 1.47	98.05 ± 0.71	0.50 ± 0.05
AFN_CBAM	89.89 ± 3.13	98.46 ± 0.85	0.80 ± 0.43
AFN_CCNet	90.21 ± 2.16	98.65 ± 0.27	0.47 ± 0.05
AFN_CTAM	91.69 ± 1.57	99.25 ± 0.17	0.44 ± 0.04
AFN_MS-CAM	91.85 ± 0.76	98.92 ± 0.27	0.44 ± 0.03
AFN_SENet	90.37 ± 0.89	98.49 ± 0.43	0.68 ± 0.38
AFN_CAM	90.10 ± 1.52	96.91 ± 1.16	0.97 ± 0.01
AFN_SAM	91.72 ± 0.60	98.84 ± 0.24	0.46 ± 0.02

In order to evaluate the class discriminatory ability of the attention module, we also used gradient-weighted class activation mapping (Grad-CAM) and guided Grad-CAM [57] to visualize the regions of the smoke image that provide support for a particular prediction. The visualization results are shown in Figure 6, wherein Grad-CAM (1) and Grad-CAM

(2) represent the attention maps of AFN with MS-CAM and AFN without the attention module, respectively. The AFN with MS-CAM shows precise smoke localization to support predictive performance, while the Grad-CAM can even localize small amounts of smoke (Figure 6d–f).

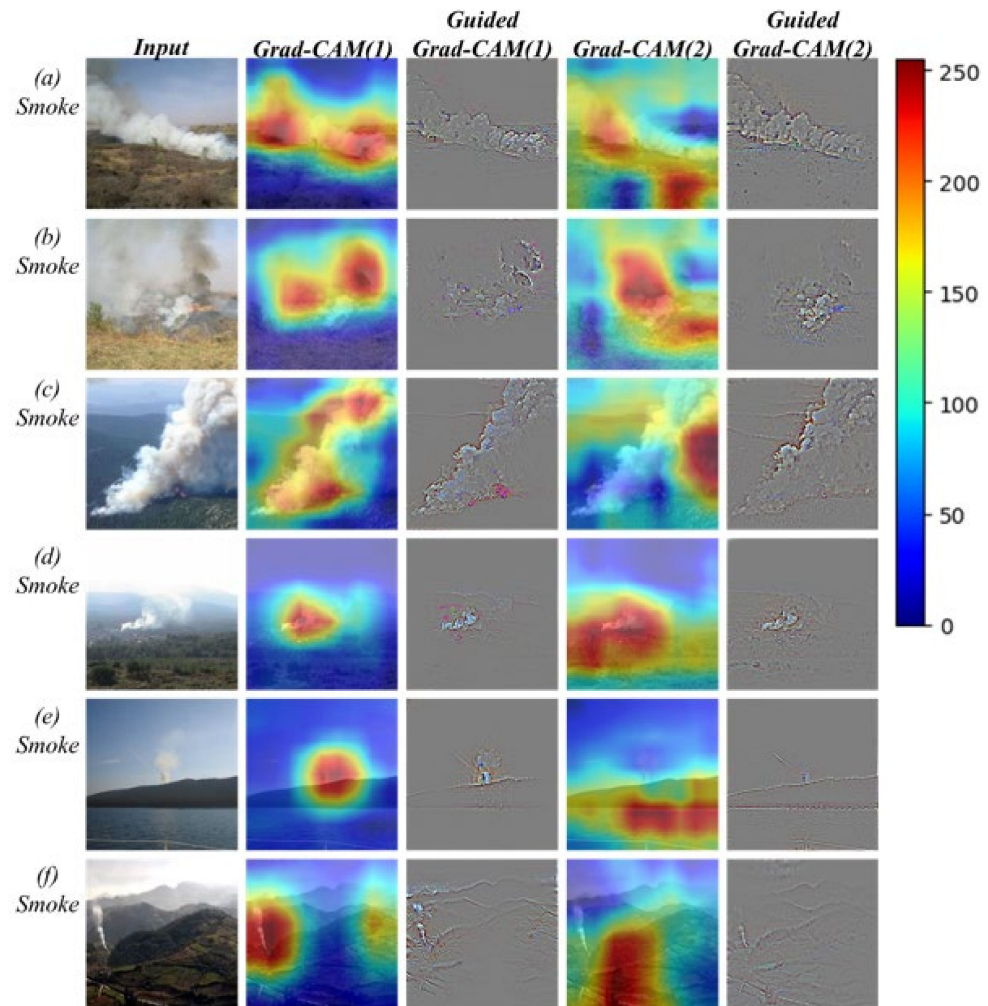


Figure 6. Grad-CAM and guided Grad-CAM visualization of randomly selected (a–f) smoke images.

4.3.2. Impact of Adversarial Feature Adaptation Network

To evaluate the effectiveness of the domain-adversarial training method, we compared the performance of different domain-adversarial CNNs. The training and test results are shown in Figure 7, where Figure 7a is the training accuracy curve and Figure 7b is a box plot of the accuracy for 100 tests. It can be seen that our method starts to converge after 5 epochs, which is significantly faster than the network trained from scratch. AFN, DDCN, and ASCN train with adversarial feature adaptation network, whereas the other methods train without. Networks with domain-adversarial training methods outperform those without, and our method achieved the highest AR ($91.85 \pm 0.76\%$). The AR of the Dual-Net ($85.02 \pm 1.26\%$) is higher than that of the pre-trained DenseNet-169 ($79.45 \pm 0.76\%$) and scratch-trained AlexNet ($81.45 \pm 0.86\%$).

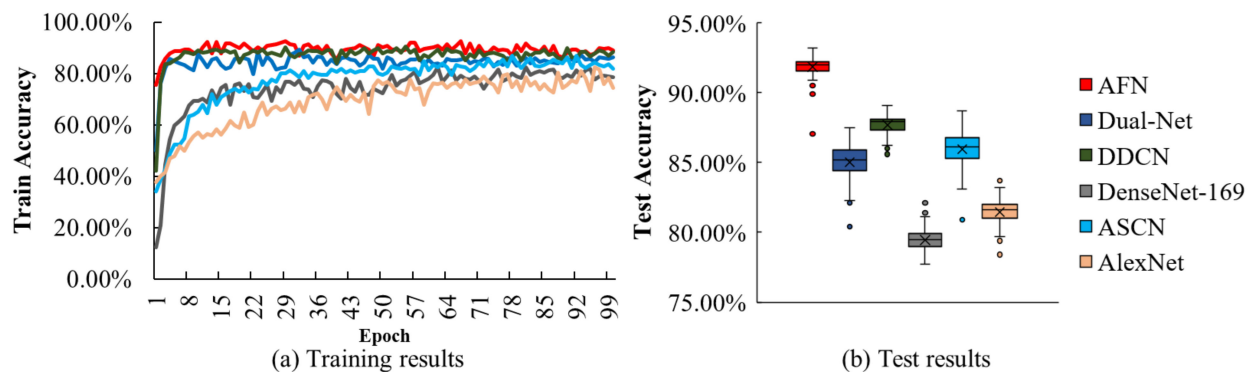


Figure 7. Results of (a) training curves and (b) test box plot on FF_Smoke dataset. The line in the box of Figure 7b shows the median; the cross indicates the mean; the box extends from the first quartile to the third quartile; the whisker extends from the box by an inter-quartile range of 1.5; and the outliers are marked with a circle.

The above experimental results show that the Dual-Net can learn more representative features than classical CNNs. Moreover, the domain-adversarial training method can also further improve the detection accuracy. Therefore, we propose a novel dual-channel convolutional neural network with domain-adversarial training for forest fire smoke detection.

5. Conclusions

In this paper, we propose an adversarial fusion network for forest fire smoke detection. Unlike conventional deep smoke detection methods, our approach is able to produce both abstract features and detailed features through the feature fusion network. Moreover, the adversarial feature adaptation network is employed to eliminate discrepancies between the base domain and the stylized domain. Extensive experiments on smoke datasets obtained from terrestrial and satellite systems show that our method achieves excellent robustness and generalization compared to existing deep-learning approaches.

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

Funding: This research was funded by the SINOMACH Intelligence Technology Research Institute Co., Ltd., China (Grant No. TC210H00L-40), and the National Key R&D Program of China (Grant No. 2020YFC1511601).

Data Availability Statement: Forest fire smoke datasets, as well as source codes of adversarial fusion network and data visualization, are available from the authors upon request.

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

References

1. Liu, T.; Cheng, J.H.; Du, X.Y.; Luo, X.B.; Zhang, L.; Cheng, B.; Wang, Y. Video smoke detection method based on change-cumulative image and fusion deep network. *Sensors* **2019**, *19*, 5060. [[CrossRef](#)]
2. 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](#)]
3. Barmpoutis, P.; Papaioannou, P.; Dimitropoulos, K.; Grammalidis, N. A Review on Early Forest Fire Detection Systems Using Optical Remote Sensing. *Sensors* **2020**, *20*, 6442. [[CrossRef](#)] [[PubMed](#)]
4. Qureshi, W.S.; Ekpanyapong, M.; Dailey, M.N.; Rinsurongkawong, S.; Malenichev, A.; Krasotkina, O. QuickBlaze: Early fire detection using a combined video processing approach. *Fire Technol.* **2016**, *52*, 1293–1317. [[CrossRef](#)]
5. Li, S.; Shi, Y.S.; Wang, B.; Zhou, Z.Q.; Wang, H.L. Video smoke detection based on color transformation and MSER. *Trans. Beijing Inst. Technol.* **2016**, *36*, 1072–1078.

6. Yin, M.X.; Lang, C.Y.; Li, Z.; Feng, S.H.; Wang, T. Recurrent convolutional network for video-based smoke detection. *Multimed. Tools. Appl.* **2019**, *78*, 237–256. [[CrossRef](#)]
7. Lin, G.H.; Zhang, Y.M.; Xu, G.; Zhang, Q.X. Smoke detection on video sequences using 3d convolutional neural networks. *Fire Technol.* **2019**, *55*, 1827–1847. [[CrossRef](#)]
8. Mao, M.T.; Wang, W.P.; Dou, Z.; Li, Y. Fire recognition based on multi-channel convolutional neural network. *Fire Technol.* **2018**, *54*, 531–554. [[CrossRef](#)]
9. Yin, Z.J.; Wan, B.Y.; Yuan, F.N.; Xia, X.; Shi, J.T. A deep normalization and convolutional neural network for image smoke detection. *IEEE Access* **2017**, *5*, 429–438. [[CrossRef](#)]
10. Jeong, M.; Park, M.J.; Nam, J.; Ko, B.C. Light-weight student LSTM for real-time wildfire smoke detection. *Sensors* **2020**, *20*, 5508. [[CrossRef](#)]
11. Geirhos, R.; Rubisch, P.; Michaelis, C.; Bethge, M.; Wichmann, F.A.; Brendel, W. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. In Proceedings of the International Conference on Learning Representations (ICLR), Vancouver, BC, Canada, 30 April–3 May 2018.
12. Brendel, W.; Bethge, M. Approximating CNNs with bag-of-local-features models works surprisingly well on ImageNet. In Proceedings of the International Conference on Learning Representations (ICLR), New Orleans, LA, USA, 6–9 May 2019.
13. Pundir, A.S.; Raman, B. Dual deep learning model for image based smoke detection. *Fire Technol.* **2019**, *55*, 2419–2442. [[CrossRef](#)]
14. Zhang, F.; Qin, W.; Liu, Y.B.; Xiao, Z.T.; Liu, J.X.; Wang, Q.; Liu, K.H. A dual-channel convolution neural network for image smoke detection. *Multimed. Tools. Appl.* **2020**, *79*, 34587–34603. [[CrossRef](#)]
15. Gu, K.; Xia, Z.F.; Qiao, J.F.; Lin, W.S. Deep dual-channel neural network for image-based smoke detection. *IEEE Trans. Multimed.* **2020**, *22*, 311–323. [[CrossRef](#)]
16. Asadi, N.; Sarfi, A.M.; Hosseinzadeh, M.; Karimpour, Z.; Eftekhari, M. Towards shape biased unsupervised representation learning for domain generalization. *arXiv* **2019**, arXiv:1909.08245.
17. Xue, W.Q.; Wang, W. One-shot image classification by learning to restore prototypes. In Proceedings of the Association for the Advance of Artificial Intelligence (AAAI), New York, NY, USA, 7–12 February 2020; pp. 6558–6565.
18. Liu, S.; Sun, Y.; Zhu, D.F.; Ren, G.H.; Chen, Y.; Feng, J.S.; Han, J.Z. Cross-domain human parsing via adversarial feature and label adaptation. In Proceedings of the Association for the Advance of Artificial Intelligence (AAAI), New Orleans, LA, USA, 2–7 February 2018.
19. Luo, Y.M.; Zhao, L.; Liu, P.Z.; Huang, D.T. Fire smoke detection algorithm based on motion characteristic and convolutional neural networks. *Multimed. Tools. Appl.* **2018**, *77*, 15075–15092. [[CrossRef](#)]
20. Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; Marchand, M.; Lempitsky, V. Domain-adversarial training of neural networks. *J. Mach. Learn. Res.* **2016**, *17*, 2030–2096.
21. Dai, Y.M.; Gieseke, F.; Oehmcke, S.; Wu, Y.Q.; Barnard, K. Attentional feature fusion. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), Snowmass Village, CO, USA, 1–5 March 2020.
22. Prema, C.E.; 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](#)]
23. Zhao, Y.Q.; Li, Q.J.; Gu, Z. Early smoke detection of forest fire video using CS Adaboost algorithm. *Optik* **2015**, *126*, 2121–2124. [[CrossRef](#)]
24. Li, X.L.; Song, W.G.; Lian, L.P.; Wei, X.G. Forest fire smoke detection using back-propagation neural network based on MODIS data. *Remote Sens.* **2015**, *7*, 4473–4498. [[CrossRef](#)]
25. Mnih, V.; Heess, N.; Graves, A.; Kavukcoglu, K. Recurrent models of visual attention. In Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS), Montreal, QC, Canada, 8–13 December 2014; pp. 2204–2212.
26. Bahdanau, D.; Cho, K.H.; Bengio, Y. Neural machine translation by jointly learning to align and translate. In Proceedings of the International Conference on Learning Representations (ICLR), Banff, AB, Canada, 14–16 April 2014.
27. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E.H. Squeeze-and-excitation networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *42*, 2011–2023. [[CrossRef](#)]
28. Wang, Q.L.; Wu, B.G.; Zhu, P.F.; Li, P.H.; Hu, Q.H. ECA-net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
29. Qin, Z.Q.; Zhang, P.Y.; Wu, F.; Li, X. FcaNet: Frequency channel attention networks. In Proceedings of the 2021 IEEE International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021.
30. Huang, Z.L.; Wang, X.G.; Huang, L.C.; Huang, C.; Wei, Y.C.; Liu, W.Y. CCNet: Criss-cross attention for semantic segmentation. In Proceedings of the 2019 IEEE International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019.
31. 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.
32. Misra, D.; Nalamada, T.; Arasanipalai, A.U.; Hou, Q.B. Rotate to attend: Convolutional triplet attention module. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), Snowmass, CO, USA, 1–5 March 2020.
33. Zhang, C.C.; Zhao, Q.J.; Wang, Y. Hybrid adversarial network for unsupervised domain adaptation. *Inf. Sci.* **2020**, *514*, 44–55. [[CrossRef](#)]
34. Gretton, A.; Smola, A.; Huang, J.; Schmittfull, M.; Borgwardt, K.; Schölkopf, B. *Covariate Shift and Local Learning by Distribution Matching*; MIT Press: Cambridge, MA, USA, 2019; pp. 131–160.

35. Sankaranarayanan, S.; Balaji, Y.; Castillo, C.D.; Chellappa, R. Generate to adapt: Aligning domains using generative adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 8503–8512.
36. Long, M.; Cao, Z.; Wang, J.; Jordan, M. Conditional adversarial domain adaptation. Processing of the 32nd Conference on Neural Information Processing Systems (NIPS), Red Hook, NY, USA, 3–8 December 2018; pp. 1640–1650.
37. Zhao, P.; Zhang, W.H.; Liu, B.; Kang, Z.; Bai, K.; Huang, K.Z.; Xu, Z.L. Domain adaptation with feature and label adversarial networks. *Neurocomputing* **2021**, *439*, 294–301. [[CrossRef](#)]
38. Tzeng, E.; Hoffman, J.; Darrell, T.; Saenko, K. Simultaneous deep transfer across domains and tasks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 4068–4076.
39. Shen, J.; Qu, Y.; Zhang, W.; Yu, Y. Wasserstein distance guided representation learning for domain adaptation. In Proceedings of the Association for the Advance of Artificial Intelligence (AAAI), Vancouver, BC, Canada, 2–7 February 2018; pp. 4058–4065.
40. Brochu, F. Increasing shape bias in ImageNet-trained networks using transfer learning and domain-adversarial methods. *arXiv* **2019**, arXiv:1907.12892.
41. Jia, Y.; Chen, W.G.; Yang, M.J.; Wang, L.W.; Liu, D.C.; Zhang, Q.X. Video smoke detection with domain knowledge and transfer learning from deep convolutional neural networks. *Optik* **2021**, *240*, 166947. [[CrossRef](#)]
42. Liu, Z.; Yang, X.P.; Liu, Y.; Qian, Z.H. Smoke-detection framework for high-definition video using fused spatial- and frequency-domain features. *IEEE Access* **2019**, *7*, 89687–89701. [[CrossRef](#)]
43. Xu, G.; Zhang, Y.M.; Zhang, Q.X.; Lin, G.H.; Wang, J.J. Deep domain adaptation based video smoke detection using synthetic smoke images. *Fire Saf. J.* **2017**, *93*, 53–59. [[CrossRef](#)]
44. Huang, G.; Liu, Z.; Laurens, V.D.M.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269.
45. Gardner, J.R.; Kusner, M.J.; Li, Y.; Upchurch, P.; Weinberger, K.Q.; Hopcroft, J.E. Deep manifold traversal: Changing labels with convolutional features. In Proceedings of the International Conference on Learning Representations (ICLR), San Juan, Puerto Rico, 2–4 May 2016.
46. Gatys, L.; Ecker, A.; Bethge, M. A neural algorithm of artistic style. *J. Vis.* **2016**, *16*, 326. [[CrossRef](#)]
47. Geirhos, R.; Temme, C.R.M.; Rauber, J.; Schutt, H.H.; Bethge, M.; Wichmann, F.A. Generalization in humans and deep neural networks. In Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS), Red Hook, NY, USA, 3–8 December 2019; pp. 7549–7561.
48. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]
49. David, S.B.; Blitzer, J.; Crammer, K.; Pereira, F. Analysis of representations for domain adaptation. In Proceedings of the International Conference on Neural Information Processing Systems (NIPS), Vancouver, BC, Canada, 4–7 December 2006; pp. 137–144.
50. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.M.; Dollár, P. Focal loss for dense object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327. [[CrossRef](#)]
51. Huang, X.; Belingie, S. Arbitrary style transfer in real-time with adaptive instance normalization. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 1510–1519.
52. Available online: <https://www.kaggle.com/c/painter-by-numbers/> (accessed on 16 December 2019).
53. Ba, R.; Chen, C.; Yuan, J.; Song, W.G.; Lo, S. SmokeNet: Satellite Smoke Scene Detection Using Convolutional Neural Network with Spatial and Channel-Wise Attention. *Remote Sens.* **2019**, *11*, 1702. [[CrossRef](#)]
54. Available online: <https://github.com/DeepQuestAI/Fire-Smoke-Dataset> (accessed on 31 August 2020).
55. Yuan, F.N.; Shi, J.T.; Xia, X.; Fang, Y.M.; Fang, Z.L.; Mei, T. High-order local ternary patterns with locality preserving projection for smoke detection and image classification. *Inf. Sci.* **2016**, *372*, 225–240. [[CrossRef](#)]
56. Zeiler, M.D.; Fergus, R. Visualizing and understanding convolutional networks. In Proceedings of the European Conference on Computer Vision (ECCV), Zurich, Switzerland, 6–12 September 2014; pp. 818–833.
57. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedanta, R.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. *Int J. Comput. Vis.* **2020**, *128*, 336–359. [[CrossRef](#)]