Expert Systems With Applications 182 (2021) 115125

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

DeepSmoke: Deep learning model for smoke detection and segmentation in outdoor environments

Salman Khan [a,b], Khan Muhammad [j,*], Tanveer Hussain [i], Javier Del Ser [c,d], Fabio Cuzzolin [b], Siddhartha Bhattacharyya [e], Zahid Akhtar [f], Victor Hugo C. de Albuquerque [g,h]

[a] Visual Analytics for Knowledge Laboratory (VIS2KNOW Lab), Department of Software, Sejong University, Seoul 143-747, Republic of Korea
[b] Visual Artificial Intelligence Laboratory (VAIL), School of Engineering, Computing and Mathematics, Oxford Brookes University, OX33 1HX Oxford, United Kingdom
[c] TECNALIA, Basque Research and Technology Alliance (BRTA), 48160 Derio, Spain
[d] University of the Basque Country (UPV/EHU), 48013 Bilbao, Spain
[e] Rajnagar Mahavidyalaya, Rajnagar, Birbhum, India
[f] Department of Network and Computer Security, State University of New York Polytechnic Institute, Utica, NY 13502, USA
[g] Graduate Program on Teleinformatics Engineering, Federal University of Ceará, Fortaleza, Fortaleza/CE, Brazil
[h] Graduate Program on Electrical Engineering, Federal University of Ceará, Fortaleza, Fortaleza/CE, Brazil
[i] Department of Software, Sejong University, Seoul 143-747, South Korea
[j] Visual Analytics for Knowledge Laboratory (VIS2KNOW Lab), School of Convergence, College of Computing and Informatics, Sungkyunkwan University, Seoul 03063, Republic of Korea

ARTICLE INFO

Keywords:
Smoke detection and segmentation
Semantic segmentation
Foggy surveillance environment
Wildfires
Disaster management

ABSTRACT

Fire disaster throughout the globe causes social, environmental, and economical damage, making its early detection and instant reporting essential for saving human lives and properties. Smoke detection plays a key role in early fire detection but majority of the existing methods are limited to either indoor or outdoor surveillance environments, with poor performance for hazy scenarios. In this paper, we present a Convolutional Neural Network (CNN)-based smoke detection and segmentation framework for both clear and hazy environments. Unlike existing methods, we employ an efficient CNN architecture, termed EfficientNet, for smoke detection with better accuracy. We also segment the smoke regions using DeepLabv3+, which is supported by effective encoders and decoders along with a pixel-wise classifier for optimum localization. Our smoke detection results evince a noticeable gain up to 3% in accuracy and a decrease of 0.46% in False Alarm Rate (FAR), while segmentation reports a significant increase of 2% and 1% in global accuracy and mean Intersection over Union (IoU) scores, respectively. This makes our method a best fit for smoke detection and segmentation in real-world surveillance settings.

1. Introduction

Disaster management is a wide domain of research to which many researchers have contributed in recent years (Finney, 2020; Muhammad, Rodrigues, Kozlov, Piccialli, & Albuquerque, 2020) because of its direct relevance to human lives and properties. There are several categories of disasters (Lu, Shi, Wang, & Jia, 2019; Muhammad, Hussain, Tanveer, Sannino, & de Albuquerque, 2019), including e.g. flood and fire, which need to be detected and monitored at their early stages to allow preventive actions. Among these disasters, fire is the most dangerous and can bring enormous damages (Bilbao, Ser, Salcedo-Sanz, & Casanova-Mateo, 2015; Cui, 2020). Thus, its automatic detection in IoT environments plays a vital role in the early handling of fire (Abbas, Zhang, Taherkordi, & Skeie, 2018). The identification of smoke is a primary sign of fire and its early detection is an effective way of averting damages caused by fire (Tian, Li, Wang, & Ogunbona, 2014). To prevent damages caused by fire disaster, several traditional and vision sensor-based fire and smoke detection methods (Muhammad et al., 2018; Muhammad, Khan, & Baik, 2020) have been proposed. Among these methods, vision-based smoke detection systems have attracted much

* Corresponding author.
E-mail addresses: Salmank@ieee.org (S. Khan), khan.muhammad@ieee.org, khanmuhammad@sju.ac.kr (K. Muhammad), Tanveer445@ieee.org (T. Hussain), Javier.delser@tecnalia.com (J.D. Ser), Fabio.cuzzolin@brookes.ac.uk (F. Cuzzolin), dr.siddhartha.bhattacharyya@ieee.org (S. Bhattacharyya), akhtarz@sunypoly.edu (Z. Akhtar), victor.albuquerque@ieee.org (V.H.C. de Albuquerque).

https://doi.org/10.1016/j.eswa.2021.115125
Received 15 July 2020; Received in revised form 4 December 2020; Accepted 25 April 2021
Available online 28 April 2021
0957-4174/

attention within the research community.

Towards smoke detection through vision sensors, the employed methods can be divided into two major categories: traditional and deep learning-based approaches. The mainstream methods belonging to the first category use different features such as color, shape, and texture to identify smoke regions. These features identify smoke without involving any learning process. Furthermore, some of these methods use low-level features and passes them to different classifiers or clustering techniques for smoke prediction. These methods have several limitations including high False Alarm Rate (FAR), limited accuracy, and lower detection range.

In contrast to the traditional approaches for smoke detection, the deep learning-based methods utilize learned features for the identification and segmentation of smoke patterns. These methods are reinforced by convolutions, pooling, and fully connected layers for learning visual notions. Several CNN-based smoke detection methods exist in the literature, which employ different models for both smoke detection and segmentation. Unfortunately, the huge model size and higher computational complexity of these modeling choices restrict their usage in real-time IoT scenarios (Liu, Luo, Zhou, & Chen, 2019; Muhammad, Hussain, Tanveer, Sannino, & de Albuquerque, 2019).

To tackle these issues effectively, we pose this novel research for smoke detection and segmentation using efficient CNNs that are functional in real-world IoT environments. Specifically, we propose an efficient smoke detection and semantic segmentation method for outdoor clear and hazy environments. The proposed framework comprises two modules: the first one is smoke detection and the second one is semantic segmentation of smoky regions. In the first module, we follow a classification strategy for the identification of smoke inside video frames. To this purpose, we use a smoke detection dataset in a hazy environment, consisting of four different classes: "smoke", "non-smoke", "smoke with fog", and "non-smoke with fog". For classification, we fine-tune a pre-trained EfficientNet architecture (Tan & Le, 2019), that is highly efficient and precise when compared to state-of-the-art CNN models. After the successful detection of smoke, the next step is to segment the smoke regions from the detected frames. We tackle this problem through a recent semantic segmentation CNN model, DeepLabv3+ (Chen, Zhu, Papandreou, Schroff, & Adam, 2018). The latter has efficient encoder and decoder layers, followed by a pixel-wise classifier for localization of smoke regions that are useful for disaster management. The main contributions of this paper can be summarized as follows:

- We present "DeepSmoke" (a deep learning framework for smoke detection and segmentation), by incorporating two essential modules for intelligent disaster management, smoke detection, and segmentation. The key ingredient of our framework is its high-level of adaptability to clear and hazy surveillance environments for smoke detection and localization.
- Smoke segmentation plays a key role in fire scene contextual analysis and its respective reporting. Its relevant literature lacks publicly available challenging and real-world datasets, which restrict the contributions from researchers to the mentioned domain. To handle this issue, we created our own smoke segmentation dataset by manually labelling the smoke regions. The dataset is made publicly available for research community (https://github.com/salmank255/deepsmoke).
- Two sets of experiments are carried out to show the effectiveness of DeepSmoke: a) an evaluation of different CNN models, and b) a comparison with existing smoke detection and segmentation techniques from different perspectives i.e., FAR, accuracy, and mean Intersection over Union. Finally, we advocate for the usage of EfficientNet for smoke detection and DeepLabv3+ for segmentation as the optimum CNN models for the said problem.
- Unlike the recently devised smoke segmentation technique (Yuan et al., 2019) that relies only on localization without focusing on smoke detection, we tackle both tasks using efficient and state-of-the-art CNN models. Automatic smoke detection followed by segmentation can provide a complete functional system with the advantage of providing instant alert generation in IoT networks and ensuring reduced FAR.

The rest of this paper is organized as follows. Section 2 explains the detailed literature of smoke detection and segmentation based on both traditional and deep learning-based methods. Section 3 provides the complete details of our proposed framework for smoke detection and segmentation in both indoor and outdoor environments. Extensive experiments of our framework on both surveillance and wildfires data are explained in Section 4. Finally, Section 5 concludes the paper by concentrating on the main building blocks of the proposed method with future research directions.

2. Related works

In this section, we thoroughly review existing smoke detection and segmentation techniques. The overall section is divided into traditional (Section 2.1) and deep learning-based (Section 2.2) smoke detection techniques.

2.1. Traditional smoke detection

Several smoke detection methods based on low-level features i.e., color, shape, and motion have been reported in the last decade. For instance, the authors in Nguyen and Kim (2013), Tung and Kim (2011) and Yuan (2008) extracted motion features using optical flow, color, and image energy to identify smoke patterns inside an image. The classifiers used in these methods included back-propagation neural network and fuzzy C-mean for the clustering of smoke patterns. Shape and texture features were also explored for the detection of smoke. For example, the authors in Chen, Wang, Tian, and Huang (2013) and Ye et al. (2015) resorted to shape and texture features for the recognition of smoke patterns in visual data. The motion features employed were smoke contours, approximate median, and accumulative motion orientation model, whereas those related to texture were gray level co-occurrence matrices, wavelet, and local binary patterns. The major drawbacks of these smoke detection methods are high FAR, thereby a limited accuracy in the presence of small-sized smoke or from a distance.

To cope with these issues, several intelligent smoke detection methods have been proposed thereafter. For example, the authors in Yuan, Fang, Wu, Yang, and Fang (2015) presented a smoke detection technique using statistical and Haar-like features with a dual threshold and an Ada-boost classifier. Similarly, another work (Dimitropoulos, Barmpoutis, & Grammalidis, 2017) introduced a "higher-order linear dynamical system" for early smoke identification in videos. To improve the classification accuracy, they used particle swarm optimization for the textural analysis of smoke patterns. The main limitations of these methods are their low detection rate, higher computational complexity, and low frame rate.

2.2. Deep learning-based smoke detection

In the past decade, deep learning methods have gained a significant advancement in several areas of computer vision e.g., violence detection (Ullah, Ullah, Muhammad, Haq, & Baik, 2019), objects detection, and action/activity recognition (Saha, Singh, & Cuzzolin, 2017; Singh, Saha, Sapienza, Torr, & Cuzzolin, 2017; Ullah, Muhammad, Del Ser, Baik, & de Albuquerque, 2019). Similarly, deep learning based methods are also emerged as the standard modeling approach for fire/smoke detection and segmentation (Muhammad, Khan, Elhoseny, Hassan Ahmed, & Wook Baik, 2019; Muhammad, Khan, Palade, Mehmood, & De Albuquerque, 2019) in current times. For instance, a deep normalization and CNN for smoke detection in still images was presented in Yin, Wan, Yuan, Xia, and Shi (2017). They replaced simple convolutional layers

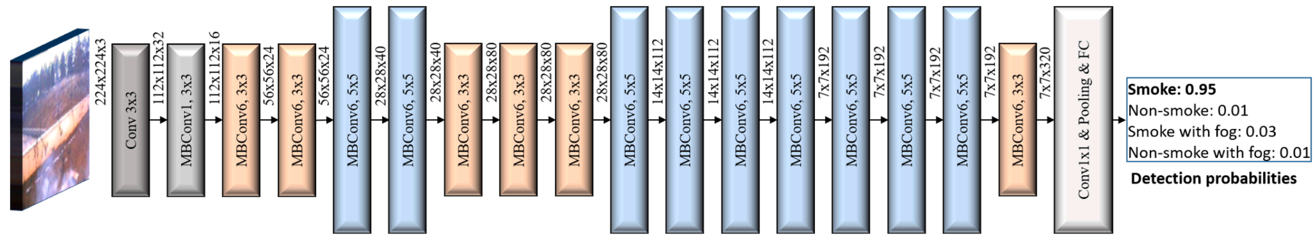

Fig. 1. A single frame of size 224 × 224 × 3 pixels is passed through our employed smoke detection architecture (Tan & Le, 2019) for acquiring output probabilities. The maximum prediction probability 0.95 for class "smoke" is the final output.

with normalized convolutional layers to boost the training process. Similarly, Hu and Lu (2018) utilized spatial–temporal CNN features for real-time smoke detection. Recently, Lin, Zhang, Xu, and Zhang (2019) presented a hybrid RCNN and 3D-CNN-based smoke detection technique to recognize smoke in a sequence of video frames. The major limitations of these methods are their limited accuracy and higher FAR in hazy environments. A recent attempt to address these limitations was made in Khan, Muhammad, Mumtaz, Baik, and de Albuquerque (2019), where a smoke detection method was proposed for both clear and hazy environments. Their method was specially trained on foggy data to differentiate between smoke and foggy regions.

In summary, several smoke detection and segmentation techniques have been put forward, but this domain is still far from maturity for a number of reasons: 1) the FAR, smoke detection accuracy, and running time still require a significant improvement to enable early fire detection for disaster management using resource-constrained devices, 2) benchmark smoke segmentation datasets are not available, thereby limiting future research in this area, and 3) smoke segmentation accuracy levels must also be significantly enhanced for detailed scene analysis without compromising the overall performance and efficiency of deployable solutions, while considering all constraints imposed by disaster management scenarios (origin of fire, size of fire, and burning degree etc.).

As shown in the following sections, our work addresses all these issues. Specifically, details on the proposed methods are given in Section 3, whereas experimental validation is discussed in Section 4.

3. Proposed DeepSmoke framework

In order to detect fire at its early stages, we propose a novel lightweight smoke detection and segmentation framework coined as DeepSmoke. The overall pipeline of our method is divided into two modules. The first module performs smoke detection using a lightweight CNN model, followed by a second module addressing semantic smoke segmentation. In the following subsections we delve into the technical details of each of these modules:

3.1. Smoke detection

To detect smoke in real-time video IoT surveillance, we studied several state-of-the-art CNN architectures and tested them experimentally. These architectures include AlexNet (Krizhevsky, Sutskever, & Hinton, 2012), GoogleNet (Szegedy et al., 2015), VGG (Simonyan & Zisserman, 2014), MobileNetV2 (Sandler, Howard, Zhu, Zhmoginov, & Chen, 2018), and EfficientNet (Tan & Le, 2019). They were tested against a set of evaluation metrics, including their accuracy, computational complexity, and FAR. After intensive experiments, we found that EfficientNet is the best candidate solution to our problem.

3.1.1. Architecture

EfficientNet utilizes the concept of CNNs scaling in a novel and effective manner, and provides an appropriate balance among width, depth, and resolution of the proposed CNN model. Tan and Le (2019) first tested their scaling method over existing networks and finally developed their own mobile-size baseline "EfficientNet" network. The main building blocks of this network are mobile inverted bottleneck (MBConv) layers with squeeze-and-excitation optimization (Hu, Shen, & Sun, 2018). Following this, our employed network architecture contains a series of MBConv layers, followed by a GlobalMaxPooling, dropout, fully connected layer, and a Softmax classifier. The input to our fine-tuned CNN architecture is a 224 × 224 RGB image that is first processed by a Conv3 × 3 layer, followed by MBConv1, 3 × 3 with 112 × 112 × 16 dimensions. The output from MBConv1 is further processed by a series of MBConv layers with varying dimensions. The modified network architecture, where input leading to output with details about the intermediate layers and their dimensions is presented in Fig. 1 in an easy and understandable way.

Table 1
Comparison of different CNN models in terms of efficiency, size, top-1, and top-5 accuracy. The higher values of different parameters are highlighted in **bold**.

Model	MFLOPS/ image	Model size (MB)	Top-1 accuracy (%)	Top-5 accuracy (%)
AlexNet (Krizhevsky et al., 2012)	720	219	57.1	80.2
GoogleNet (Szegedy et al., 2015)	1500	39.66	69.8	89.3
VGG (Simonyan & Zisserman, 2014)	20,000	930	70.5	91.2
MobileNetV2 (Sandler et al., 2018)	**300**	**13.23**	71.8	91.0
ResNet-152 (He, Zhang, Ren, & Sun, 2016)	11,000	–	77.8	93.8
ResNeXt-101 (Xie, Girshick, Dollár, Tu, & He, 2017)	32,000	–	80.9	95.6
EfficientNet (Tan & Le, 2019)	4200	77.85	**83.0**	**96.3**

The original publicly available EfficientNet model was trained over millions of images on the ImageNet dataset (Russakovsky et al., 2015) for 1000-class classification. In our paper, we fine-tuned the existing weights, whose parameters are already familiar with diverse types of image categories, following the well-known transfer learning methodology. We then attached a Softmax classifier to train our desired data with four classes i.e., "smoke", "non-smoke", "smoke with fog", and "non-smoke with fog".

3.1.2. Motivation for model selection

We selected a model for smoke detection based on different parameters, including performance, size, and computational complexity. To achieve this goal, we investigated several state-of-the-art CNN models and concluded that EfficientNet is the optimal one. The detailed performance of the different models on the large-scale ImageNet dataset is given in Table 1.

3.2. Smoke semantic segmentation

The second module of our method performs smoke region

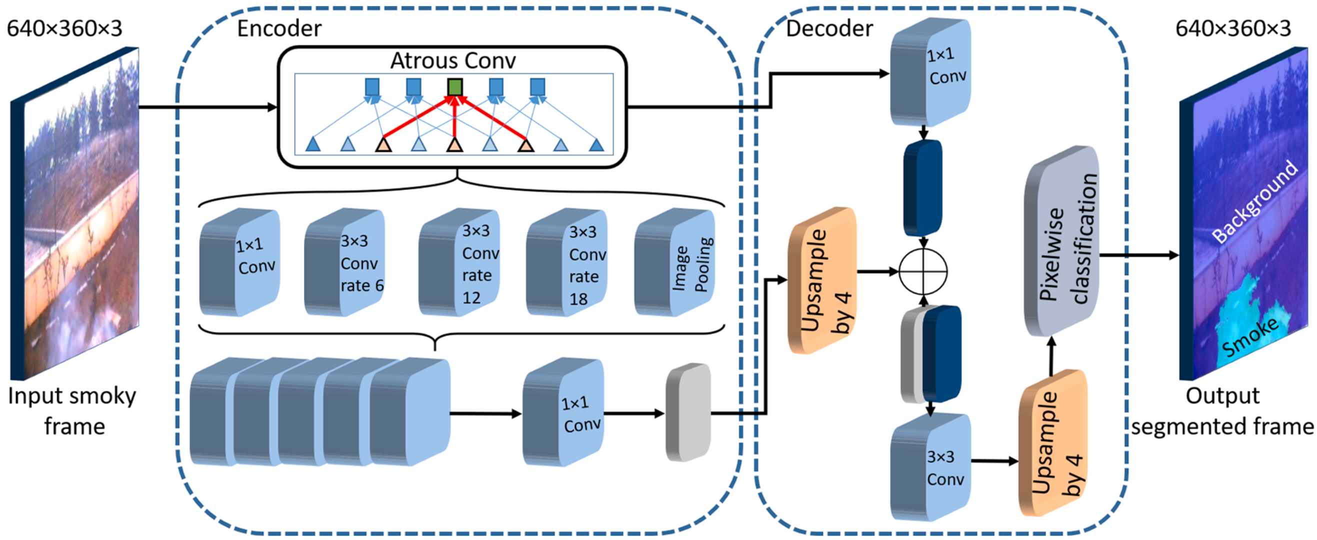

Fig. 2. Architecture (Chen, Zhu, et al., 2018) of our proposed smoke segmentation module. The input frame shown in this figure is passed through the module, resulting in an image with segmented background and smoke regions.

Table 2
Statistics for comparison of different existing models with our employed DeepLabV3+ segmentation model. The values in **bold** correspond to the best performance.

Model	Mean IoU	Boundary F1-score
DeepLab (Chen, Papandreou, et al., 2018)	54.74	29.33
FCN (Long et al., 2015)	51.96	33.18
DeconvNet (Noh et al., 2015)	59.77	40.79
SegNet (Badrinarayanan et al., 2017)	60.10	46.84
DeepLabV3+ (Chen, Zhu, et al., 2018)	**63.02**	**65.05**

segmentation, given the smoky image acquired as an output from the first module. To this end, we investigated an encoder-decoder approach with atrous separable convolutions for smoke segmentation, adopted from Chen, Zhu, et al. (2018). The original architecture uses an encoder and decoder deep network followed by depth-wise separable convolutions (or group convolutions) for semantic segmentation. The architecture is trained and tested over the popular PASCAL VOC 2012 (Everingham et al., 2015) and Cityscapes (Cordts et al., 2016) datasets. As semantic segmentation is about the pixel-wise labelling of each object in an image, we updated the given architecture for the segmentation of smoke and background. To achieve this goal, we utilized the pre-trained DeepLabv3+ (Chen, Zhu, et al., 2018) model and updated the weights of the pixel-wise classifier by transfer learning.

3.2.1. Architecture

In our employed architecture, we have three main building blocks: encoder with atrous convolution at multiple scales, effective decoder, and a pixel-wise classifier, as visualized in Fig. 2. In the encoders, an atrous spatial pyramid pooling module investigates the convolutional features at multiple scales by employing atrous convolution at different rates. In our method, we use the default configuration as in Chen, Zhu, et al. (2018). It outputs feature maps with 256 channels that possess rich semantic information. As discussed in Chen, Zhu, et al. (2018), the size of these features can be adapted to the available computational resources. On the decoder side, these features are bilinearly upsampled by a factor of 4 and concatenated with low-level features acquired from the encoder module. A series of convolutional operations are applied over the concatenated features for refining, followed by another bilinear upsampling by a factor of 4. The finally obtained feature maps from the decoder are fed to the Softmax layer for the classification of each pixel.

3.2.2. Motivation for model selection

Model selection is a critical step while dealing with the applications related to disaster management. To select the best model for smoke segmentation, we examined numerous semantic segmentation models such as DeepLab (Chen, Papandreou, Kokkinos, Murphy, & Yuille, 2018), FCN (Long, Shelhamer, & Darrell, 2015), DeconvNet (Noh, Hong, & Han, 2015), SegNet (Badrinarayanan, Kendall, & Cipolla, 2017), and selected the DeepLabv3+ (Chen, Zhu, et al., 2018) due to its favorable and supportive structure to gauge between computational complexity and classification accuracy. A detailed empirical comparison of our employed model with the other state-of-the-art models over the CamVid dataset (Brostow, Fauqueur, & Cipolla, 2009) is given in Table 2.

3.3. Loss function

The loss function used in our employed architectures is cross-entropy loss (Zhang & Sabuncu, 2018). It is the most popular loss function used in different CNN architectures. Cross-entropy, also known as Softmax loss, has been shown to perform quite better in removing outliers in smoke detection and segmentation.

4. Experiments

This section describes the experimental details including dataset and simulation environment used for the testing. Next, a comparison of our model with the other state-of-the-art deep learning architectures is presented. Finally, our "DeepSmoke" is extensively compared with recent smoke detection and segmentation methods from various perspectives.

4.1. Experimental details

To validate the performance of our proposed method, we performed experiments on a computer having an Nvidia GetForce RTX 2080 GPU with 12 GB on-board memory. For smoke detection implementation, we used Python 3.5 with Keras Deep Learning framework, while for the smoke segmentation module, we used MATLAB with the support of the Deep Learning and Computer Vision toolboxes. Furthermore, we used three different datasets for the comparison of our method with other state-of-the-art smoke detection and segmentation approaches.

4.1.1. Dataset 1 (DS1) (Khan et al., 2019)

The first dataset consists of four different classes: "smoke", "non-

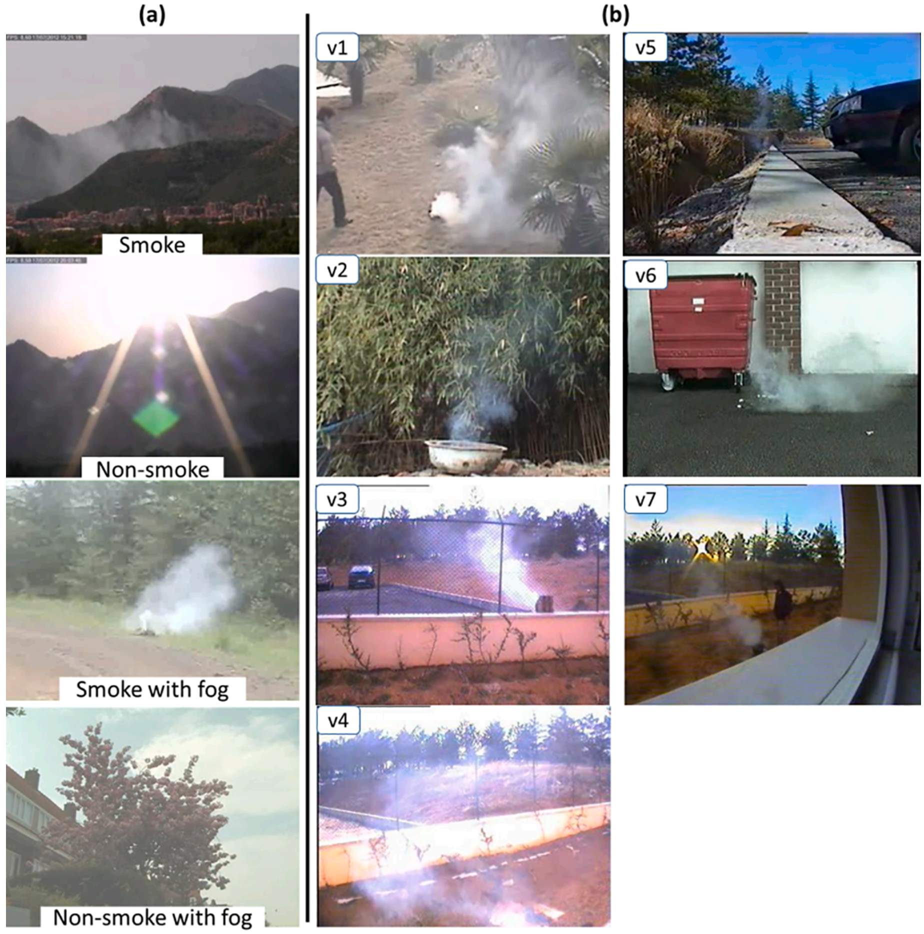

Fig. 3. Representative frames of smoke detection datasets, where (a) represents the sample frames of DS1 from all classes i.e., “smoke”, “non-smoke”, “smoke with fog”, and “non-smoke with fog” and (b) shows sample frames from each video of DS2.

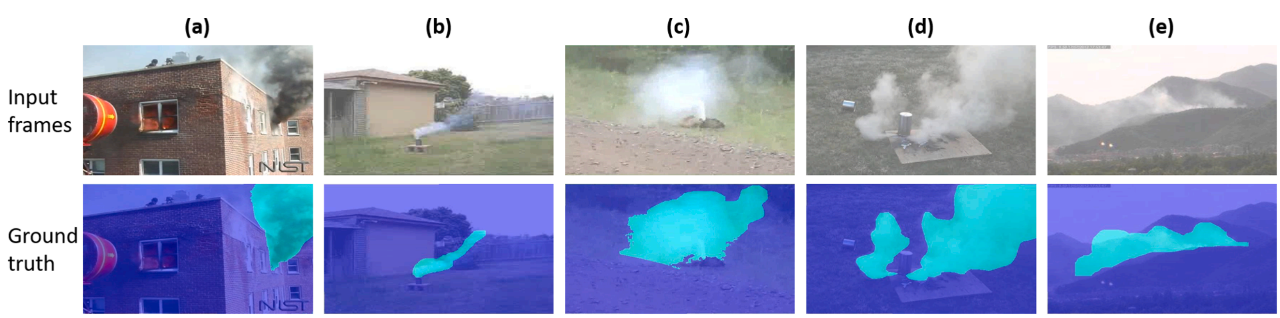

Fig. 4. Sample images from our DS3 dataset for smoke segmentation in clear and hazy environments. (a), (b), and (c): smoke in a clear environment, (d) and (e) smoke in a hazy (foggy) environment.

smoke”, “smoke with fog”, and “non-smoke with fog”. The total number of images in this dataset is 72,012 with 18,532 for each “smoke” and “smoke with fog” and 17,474 for “non-smoke” and “non-smoke with fog”, individually. This dataset follows a 20%, 30%, and 50% split for training, validation, and testing data, respectively. Sample frames from this dataset are visualized in Fig. 3 (a).

4.1.2. Dataset 2 (DS2) (Filonenko, Hernandez, & Jo, 2018)

This dataset comprises seven smoky videos, captured in different challenging outdoor environments. The environments include smoke at long distance, smoke in parking lot with other moving objects, smoke on a cotton rope, etc. These seven videos are only used for testing purposes and are not used in training process. Despite this, our method outperforms our competitors on DS2. Sample frames from these videos are shown in Fig. 3 (b).

4.1.3. Dataset 3 (DS3)

This is our own proprietary dataset and is used for the evaluation of the semantic segmentation module. It comprises 252 images annotated from DS1 by considering only two classes: “smoke” and “smoke with fog”. This dataset is divided into 60% and 40% for training and testing, respectively. Its images are annotated using the MATLAB “Image-Labeler” application with pixel ROI labels setting. Sample images from both the environments are visualized in Fig. 4, where the smoke is

Table 3
Confusion matrix of our DeepSmoke detection module using test set of DS1.

	Smoke (%)	Non-smoke (%)	Smoke with fog (%)	Non-smoke with fog (%)
Smoke	**98.85**	0.63	0.5	0.02
Non-smoke	1.79	**98.21**	0	0
Smoke with fog	1.24	0	**97.62**	1.14
Non-smoke with fog	0.13	0	1.8	**98.07**

depicted in light blur, and the background in a darker shade of blue.

4.2. Comparison with state-of-the-art CNNs

In this section, we compared our architectures for smoke detection with our competitors using DS1 and DS2, and did the same for semantic segmentation over DS3. Our DeepSmoke detection module is evaluated using five different metrics, i.e., FAR, frames per second (FPS), precision, recall, and F-measure. The confusion matrix for smoke detection over the test set of DS1 is given in Table 3. The models used for the smoke detection comparison are AlexNet (Krizhevsky et al., 2012), GoogleNet (Szegedy et al., 2015), VGG (Simonyan & Zisserman, 2014), and MobileNetV2 (Sandler et al., 2018). The overall results for all these models are given in Table 4. It can be observed that AlexNet and GoogleNet performed the worst across all the evaluation metrics on DS1 and DS2. The FPS of GoogleNet is better than AlexNet due to its lower computational complexity. VGG achieved better performance compared to both AlexNet and GoogleNet. In comparison to these models, MobileNetV2 performed better in terms of all evaluation metrics except recall on DS1 and precision on DS2. Finally, it can be observed from Table 4 that our DeepSmoke outperformed all the existing models in terms of all evaluation metrics, showing the effectiveness of our method.

We also compared our DeepSmoke semantic segmentation approach with other state-of-the-art semantic segmentation architectures. The results are evaluated using our own proprietary dataset (DS3) for smoke semantic segmentation, as explained in Section 4.1. The comprehensive results are given in Table 5 using several evaluation metrics such as global accuracy, mean accuracy, mean IoU, weighted IoU, and Mean Boundary F1-score (BF score). The first two metrics are related to accuracy. Global accuracy represents the ratio of correctly classified pixels to the total number of pixels, whereas mean accuracy is the ratio of correctly classified pixels in each class to the total number of pixels. The next two metrics based on IoU are referred to as mean and weighted IoU, where mean IoU is the average IoU of all the classes as given in (1), where *PR* and *GT* represent the predicted output and ground truth, respectively. The weighted IoU is the average IoU of all the classes, weighted by the number of pixels in each class. The last evaluation metric is the mean of the BF score, which is defined as the harmonic mean of precision and recall (2). The results using these evaluation metrics are compared with three state-of-the-art segmentation architectures i.e., DeepLab (Chen, Papandreou, et al., 2018), FCN (Long et al., 2015), DeconvNet (Noh et al., 2015), and SegNet (Badrinarayanan et al., 2017). The results obtained suggest the superiority of the proposed smoke segmentation module over the existing state-of-the-art deep semantic segmentation models.

$$MeanIoU = \frac{1}{n}\sum_{i=1}^{n}\frac{PR_i \cap GT_i}{PR_i \cup GT_i} \quad (2)$$

$$BGScore = 2\frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

4.3. Comparison with existing smoke detection methods

This section investigates the performance of our smoke detection module in contrast to state-of-the-art smoke detection methods. We have used a common dataset DS2 for the comparison of our method. The evaluation metrics used for assessment of our method are FAR, FPS, accuracy, and input image resolution used for the experimentations. The detailed statistics for each method are given in Table 6. The results show that the method due to Avgerinakis, Briassouli, and Kompatsiaris (2012) performed worst in terms of FAR and FPS, but its accuracy is still higher than that of Yuan et al. (2015). The method by Barmpoutis, Dimitropoulos, and Grammalidis (2014) attained an average accuracy but its FAR and FPS are still inferior to those of other methods. The methods by Dimitropoulos et al. (2017) and Tian, Li, Ogunbona, and Wang (2018) achieved almost similar results with average FPS and accuracy. The method introduced in Filonenko et al. (2018) achieved the best FPS among all the methods, but its FAR and accuracy are still not as good as

Table 5
Comparative analysis of various smoke segmentation architectures over our DS3 dataset.

Model	Global Accuracy	Mean Accuracy	Mean IoU	Weighted IoU	Mean BF Score
DeepLab (Chen, Papandreou, et al., 2018)	0.8397	0.8193	0.7687	0.8073	0.4371
FCN (Long et al., 2015)	0.8124	0.7952	0.6912	0.7439	0.4064
DeconvNet (Noh et al., 2015)	0.8617	0.8432	0.7144	0.8116	0.4833
SegNet (Badrinarayanan et al., 2017)	0.9045	0.9167	0.7468	0.8392	0.4456
DeepSmoke	**0.9134**	**0.9333**	**0.7786**	**0.8546**	**0.5076**

Table 6
Time complexity analysis of different smoke detection methods with different parameters such as FAR, FPS, accuracy, and resolution.

Method	FAR	FPS	Accuracy (%)	Resolution
(Avgerinakis et al., 2012)	15.92	5	80.08	–
(Barmpoutis et al., 2014)	6.63	5.7	90.87	240 × 320
(Yuan et al., 2015)	5.0	25	47.71	320 × 240
(Dimitropoulos et al., 2017)	–	5.2	94.81	320 × 240
(Tian et al., 2018)	–	3.25	84.47	256 × 500
(Filonenko et al., 2018)	4.29	**61**	84.85	320 × 240
(Yin et al., 2017)	2.44	30.73	–	48 × 48
DeepSmoke	**1.98**	32.57	**98.18**	224 × 224

Table 4
Detailed comparison of various CNN models for smoke detection on the DS1 and DS2 datasets.

Model	DS1					DS2			
	FAR	FPS	Precision	Recall	F-measure	FAR	Precision	Recall	F-measure
AlexNet (Krizhevsky et al., 2012)	3.39	17	0.96	0.95	0.96	4.21	0.94	0.93	0.92
GoogleNet (Szegedy et al., 2015)	3.17	23	0.96	0.96	0.96	3.58	0.94	0.95	0.94
VGG (Simonyan & Zisserman, 2014)	2.30	31.33	0.97	**0.98**	0.97	3.11	0.97	0.95	0.96
MobileNetV2 (Sandler et al., 2018)	2.06	**39.78**	0.98	0.97	**0.98**	2.81	0.97	0.96	0.96
DeepSmoke	**1.98**	32.57	**0.99**	**0.98**	**0.98**	**0.98**	**0.98**	**0.97**	**0.98**

Table 7
A detailed comparative analysis of various methods over the seven videos of DS2 using mean execution time (MET), precision (P), recall (R), and f-measure (F).

Video	(Yuan et al., 2015)				(Filonenko et al., 2018)				(Khan et al., 2019)				**DeepSmoke**			
	MET	P	R	F	MET	P	R	F	MET	P	R	F	MET	P	R	F
V1	73.21	0.97	0.99	0.98	16.30	1	0.79	0.88	30.29	0.98	0.96	0.97	28.14	0.99	0.98	0.99
V2	62.25	1	0.93	0.96	17.45	0.99	0.94	0.97	31.5	0.99	0.95	0.97	29.62	0.98	0.95	0.97
V3	63.33	0.2	0.32	0.3	21.31	0.93	0.96	0.95	41.28	0.93	0.92	0.92	41.32	0.96	0.97	0.97
V4	67.76	0.86	0.57	0.69	27.60	0.99	0.77	0.87	57.11	0.99	0.97	0.98	55.52	0.99	0.99	0.99
V5	69.44	0	0	–	18.09	0.97	0.8	0.88	39.76	0.99	0.98	0.99	36.45	0.98	0.99	0.98
V6	67.62	0.96	0.79	0.43	17.30	0.97	0.93	0.95	37.77	0.99	1	1	39.21	0.98	1	0.98
V7	66.56	0	0	–	22.02	0.85	1	0.92	48.08	1	0.91	0.95	46.97	0.99	0.94	0.97
Average	67.16	0.57	0.51	0.48	**20.01**	0.95	0.88	0.91	40.82	**0.98**	0.95	0.96	39.43	**0.98**	**0.97**	**0.98**

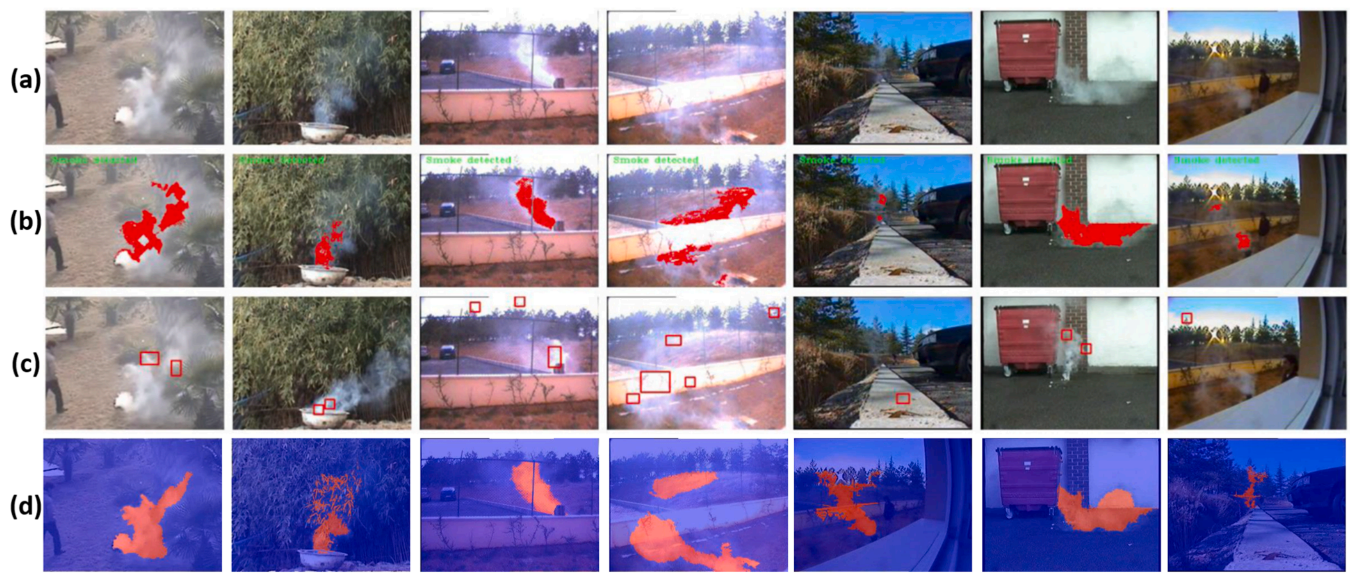

Fig. 6. Sample segmentation results of DeepSmoke where row (a) portraits sample input frames, (b) and (c) are the corresponding segmentation results by Filonenko et al. (2018) and Yuan et al. (2015), respectively, and (d) shows the visual results of the proposed DeepSmoke approach.

the proposed method. Summarizing, it can be seen that DeepSmoke outperformed all the existing approaches in terms of FAR and accuracy, making it a more suitable aspirant for deployment in real IoT scenarios.

To further scrutinize the performance of our smoke detection module, we compared it with recent smoke detection methods using the seven videos of DS2. The evaluation metrics used for the comparison include mean execution time (MET) in milliseconds for each frame, precision, recall, and F-measure. MET is used to evaluate the execution time, while the other metrics are used for performance evaluation. The collected results are given in Table 7, where each row represents the results of each video followed by an average score of all the videos. We can see that the method by Yuan et al. (2015) performed worst as compared to all other methods. The next method by Filonenko et al. (2018) performed best in terms of processing time compared to other methods but its remaining factors are worse than that of Khan et al. (2019) and the proposed DeepSmoke. Lastly, the method by Khan et al. (2019) and DeepSmoke method attained almost similar results but our proposed DeepSmoke is more efficient, as it processes up to 33 frames per second, which is enough for real-time processing in IoT environments.

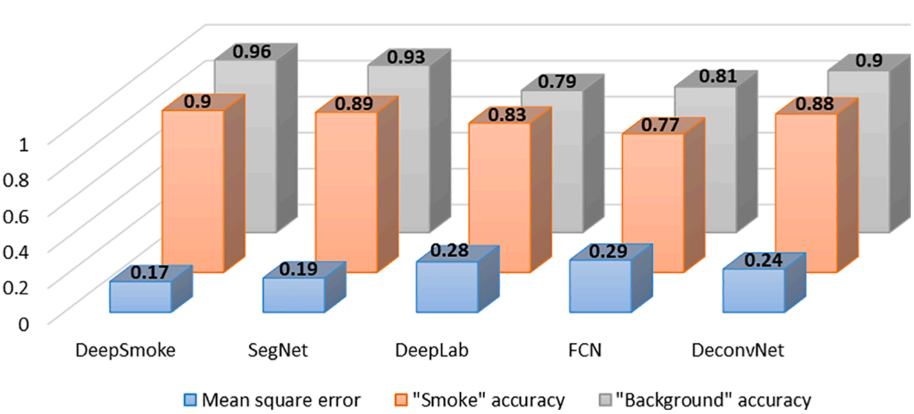

Fig. 7. Comparison of our DeepSmoke segmentation module over DS3 with state-of-the-art semantic segmentation models i.e., SegNet (Badrinarayanan et al., 2017), DeepLab (Chen, Papandreou, et al., 2018), FCN (Long et al., 2015), and DeconvNet (Noh et al., 2015).

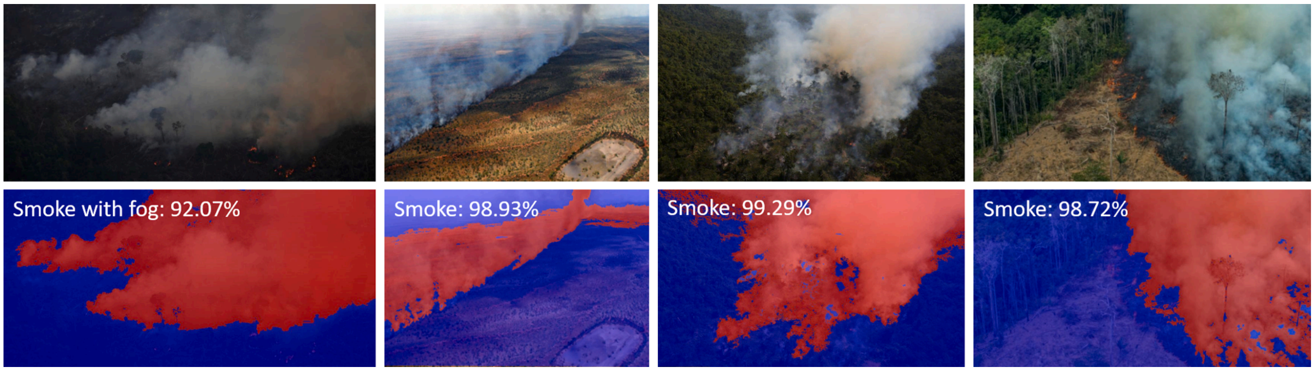

Fig. 8. Exposure of our "DeepSmoke" detection and segmentation module to wildfire images captured from arial view.

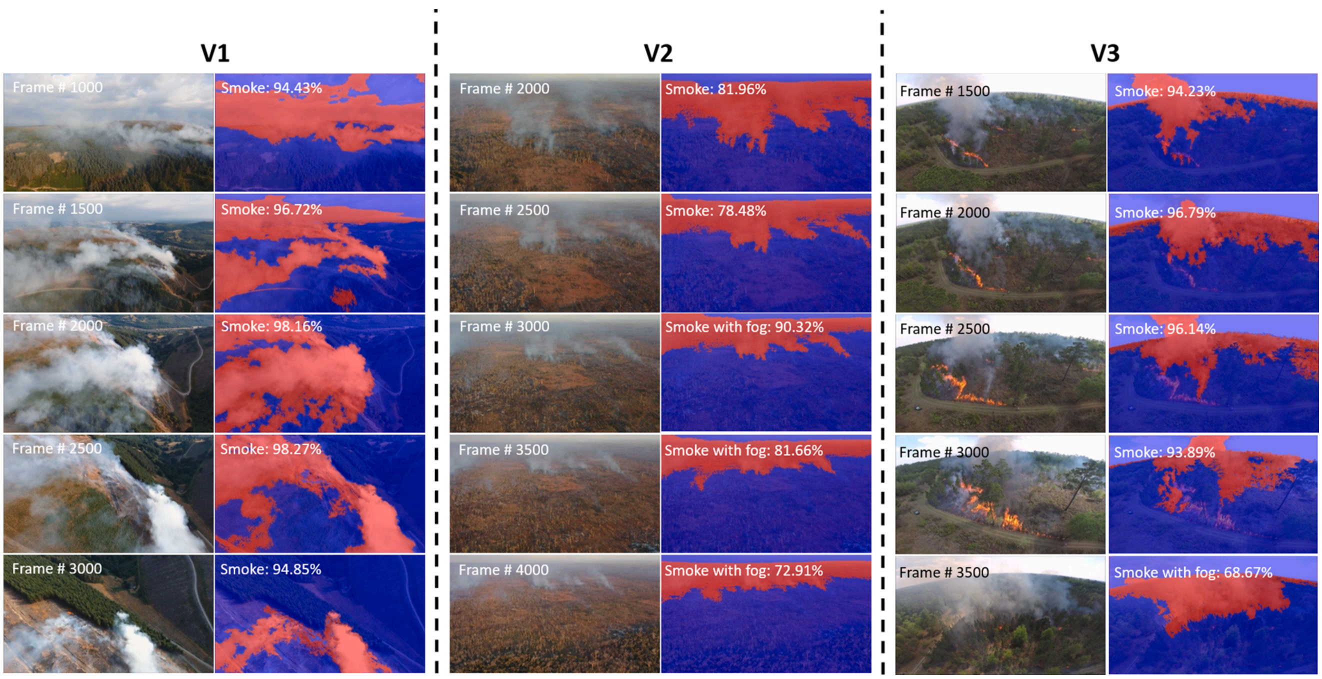

Fig. 9. Qualitative results of our smoke detection and segmentation module over three drone videos (v1, v2, and v3) recorded in different places. All the videos contain unique type of smoke with different sizes and distances. Only five frames with a gap of 500 (frame number is added to sample frames) are given in the figure as an example and best viewed in color and zoom in. The links to the original videos are added to our GitHub (https://github.com/salmank255/deepsmoke).

4.4. Comparison with existing smoke segmentation methods

To demonstrate the superiority of our smoke segmentation module, we compared our results with recent methods in the literature. Some representative visual results from DS2 are given in Fig. 6, where (a) represents the input frames of the seven videos and (b) shows the results achieved by the authors in Filonenko et al. (2018). The segmentation results of the method reported in Filonenko et al. (2018) indicate that this method performs well when the smoke is larger in size or near to camera, while detection accuracy is worse if the smoke is smaller in size or farther away. Afterward (c) achieves very limited accuracy and higher FAR by segmenting the non-smoky areas inside the frame. Finally, our proposed DeepSmoke method outperformed both the existing methods by correctly segmenting smoke near to camera and even small-sized smoke at larger distances.

The dominance of our DeepSmoke segmentation module is further confirmed by comparing it with the state-of-the-art semantic segmentation models. The evaluation metrics used for the comparison are mean squared error (MSE) of smoke region and accuracy for both the classes i.e., smoke and background individually. The overall results are visualized in Fig. 7, where our DeepSmoke outperformed all the existing models in terms of all evaluation metrics.

4.5. Performance on wildfire/smoke detection and segmentation

Wildfires in different countries around the globe i.e., America (Moeini et al., 2020), Australia (Wills, Liddelow, & Tunsell, 2020), and Spain (Martín, Pierna, & Díez, 2006) motivated us to conduct experiments on wildfires to show the adaptiveness of our system. In this direction, we downloaded 20 images and three videos (captured using a drone camera) from Google Images and YouTube, respectively. Subsequently, we passed them through our proposed smoke detection and segmentation modules.

In our experiments, we used both images and videos with the main goal of analyzing our method's response towards both static and dynamic scenes. The results achieved from wildfire images resulted to be of better quality than the videos due to the static nature of the scenery and the clear view. Some sample results are given in Fig. 8, where the detection label with the highest prediction probabilities are given over each image. The segmentation results are visualized in two colors i.e., blue and red, for background and smoke, respectively. Following the same setup, the results of the wildfire videos are also visualized in Fig. 9, wherein the smoke detection classes are predicted accurately except for a few cases, where the model is confused between "Smoke" and "Smoke with fog". Nevertheless, both classes in confusion are refereeing to fire. On the other hand, the segmentation module segments the smoke regions efficiently. However, due to the challenging background it also segments the clouds as smoke in some cases and misses to delimit the actual smoke regions. The main reason for these mis-classifications and segmentations is the change of domain (dataset shift), as this type of data is totally new to our models and none of the samples from these data is used in our training process. In this section, we only discuss qualitative results, due to the unavailability of annotated wildfires image datasets. In summary, the results yielded by our developed model when processing real imagery are satisfactory considering the complex scene information in the images with respect to the datasets from where the model was trained.

5. Conclusion

Mainstream vision-based smoke detection techniques are based on low-level features and statistical learning, limited to indoor or outdoor scenarios without any focus on hazy environments. In this paper we fine-tuned an EfficientNet architecture for detection of smoke, non-smoke, smoke with fog, and non-smoke with fog. Our system significantly improved the False Alarm Rate by reporting a reduction of 1.98%, and by boosting the accuracy to 98.18% when compared to the state-of-the-art models and existing smoke detection methods. The smoke detection results were passed to a semantic segmentation module, where a pre-trained DeepLabv3+ based on deep encoders and decoders along with pixel-wise classifier was utilized. Our segmentation results reported a significant increase of 2% and 1% in global accuracy and mean IoU, respectively. Furthermore, bearing in mind the need for early wildfire detection and segmentation, we also conducted an ablation study to show the performance and adaptiveness of our method to large-scale wildfires. These results made our system a realistic option for early fire detection and disaster prevention.

In future, we are trying to improve the segmentation accuracy, which is comparatively less in certain cases. Furthermore, we are planning to train an end-to-end deep learning model for both smoke detection and localization. Finally, our method can be extended to include a contextual analysis of the fire scene to extract and exploit additional information.

CRediT authorship contribution statement

Salman Khan: Conceptualization, Methodology, Software, Writing - original draft. **Khan Muhammad:** Conceptualization, Methodology, Software, Writing - original draft, Formal analysis, Investigation, Validation, Writing - review & editing, Supervision. **Tanveer Hussain:** Methodology, Software, Writing - original draft. **Javier Del Ser:** Methodology, Investigation, Validation, Writing - review & editing. **Fabio Cuzzolin:** Investigation, Validation, Writing - review & editing, Supervision. **Siddhartha Bhattacharyya:** Formal analysis, Investigation, Writing - review & editing. **Zahid Akhtar:** Formal analysis, Investigation, Writing - review & editing. **Victor Hugo C. de Albuquerque:** Conceptualization, Writing - review & editing, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The work of Salman Khan and Fabio Cuzzolin has received funding from the European Union's Horizon 2020 research and innovation programme, under grant agreement No. 964505 (E-pi). J. Del Ser acknowledges funding support from the Basque Government through the ELKARTEK program (3KIA project, KK-2020/00049) and the consolidated research group MATHMODE (ref. T1294-19).

References

Abbas, N., Zhang, Y., Taherkordi, A., & Skeie, T. (2018). Mobile edge computing: A survey. *IEEE Internet of Things Journal, 5*(1), 450–465. https://doi.org/10.1109/JIOT.2017.2750180

Avgerinakis, K., Briassouli, A., & Kompatsiaris, I. (2012). Smoke detection using temporal HOGHOF descriptors and energy colour statistics from video. *Paper presented at the International workshop on multi-sensor systems and networks for fire detection and management*.

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(12), 2481–2495.

Barmpoutis, P., Dimitropoulos, K., & Grammalidis, N. (2014). Smoke detection using spatio-temporal analysis, motion modeling and dynamic texture recognition. *Paper presented at the Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European*.

Bilbao, M. N., Ser, J. D., Salcedo-Sanz, S., & Casanova-Mateo, C. (2015). On the application of multi-objective harmony search heuristics to the predictive deployment of firefighting aircrafts: A realistic case study. *International Journal of Bio-Inspired Computation, 7*(5), 270–284.

Brostow, G. J., Fauqueur, J., & Cipolla, R. (2009). Semantic object classes in video: A high-definition ground truth database. *Pattern Recognition Letters, 30*(2), 88–97.

Chen, J., Wang, Y., Tian, Y., & Huang, T. (2013). Wavelet based smoke detection method with RGB Contrast-image and shape constrain. *Paper presented at the 2013 Visual Communications and Image Processing (VCIP)*.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 40*(4), 834–848.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. *Paper presented at the Proceedings of the European conference on computer vision (ECCV)*.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., … Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. *Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition*.

Cui, F. (2020). Deployment and integration of smart sensors with IoT devices detecting fire disasters in huge forest environment. *Computer Communications, 150*, 818–827.

Dimitropoulos, K., Barmpoutis, P., & Grammalidis, N. (2017). Higher order linear dynamical systems for smoke detection in video surveillance applications. *IEEE Transactions on Circuits and Systems for Video Technology, 27*(5), 1143–1154.

Everingham, M., Eslami, S. M. A., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2015). The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision, 111*(1), 98–136.

Filonenko, A., Hernandez, D. C., & Jo, K.-H. (2018). Fast smoke detection for video surveillance using CUDA. *IEEE Transactions on Industrial Informatics, 14*(2), 725–733.

Finney, M. A. (2020). The wildland fire system and challenges for engineering. *Fire Safety Journal, 103085*.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition*.

Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. *Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition*.

Hu, Y., & Lu, X. (2018). Real-time video fire smoke detection by utilizing spatial-temporal ConvNet features. *Multimedia Tools and Applications, 77*(22), 29283–29301.

Khan, S., Muhammad, K., Mumtaz, S., Baik, S. W., & de Albuquerque, V. H. C. (2019). Energy-efficient deep CNN for smoke detection in Foggy IoT environment. *IEEE Internet of Things Journal, 6*(6), 9237–9245.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Paper presented at the Advances in neural information processing systems*.

Lin, G., Zhang, Y., Xu, G., & Zhang, Q. (2019). Smoke detection on video sequences using 3D convolutional neural networks. *Fire Technology, 55*(5), 1827–1847.

Liu, J., Luo, K., Zhou, Z., & Chen, X. (2019). ERP: Edge Resource Pooling for Data Stream Mobile Computing. *IEEE Internet of Things Journal, 6*(3), 4355–4368. https://doi.org/10.1109/JIOT.2018.2882588

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition*.

Lu, C., Shi, J., Wang, W., & Jia, J. (2019). Fast abnormal event detection. *International Journal of Computer Vision, 127*(8), 993–1011. https://doi.org/10.1007/s11263-018-1129-8

Martín, S. D., Pierna, C. S., & Díez, E. G. (2006). Fires during 2005 in Spain: A forest fire in Guadalajara and an urban fire in Windsor building (Madrid). *Forest Ecology and Management, 234*, S72.

Moeini, O., Tarasick, D. W., McElroy, C. T., Liu, J., Osman, M. K., Thompson, A. M., … Oltmans, S. J. (2020). Estimating wildfire-generated ozone over North America using ozonesonde profiles and a differential back trajectory technique. *Atmospheric Environment: X*, Article 100078.

Muhammad, K., Ahmad, J., Lv, Z., Bellavista, P., Yang, P., & Baik, S. W. (2018). Efficient deep CNN-based fire detection and localization in video surveillance applications. *IEEE Transactions on Systems, Man, and Cybernetics Systems, 99*, 1–16.

Muhammad, K., Hussain, T., Tanveer, M., Sannino, G., & de Albuquerque, V. H. C. (2019). Cost-effective video summarization using Deep CNN with hierarchical weighted fusion for IoT surveillance networks. *IEEE Internet of Things Journal, 7*(5), 4455–4463.

Muhammad, K., Khan, S., & Baik, S. W. (2020). 3 Efficient Convolutional neural networks for fire detection in surveillance applications. *Deep Learning in Computer Vision: Principles and Applications, 63*.

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

Muhammad, K., Khan, S., Palade, V., Mehmood, I., & De Albuquerque, V. H. C. (2019). Edge intelligence-assisted smoke detection in Foggy surveillance environments. *IEEE Transactions on Industrial Informatics, 16*(2), 1067–1075.

Muhammad, K., Rodrigues, J. J. P. C., Kozlov, S., Piccialli, F., & Albuquerque, V. H. C.d. (2020). Energy-efficient monitoring of fire scenes for intelligent networks. *IEEE Network, 34*(3), 108–115.

Nguyen, T. K. T., & Kim, J.-M. (2013). Multistage optical smoke detection approach for smoke alarm systems. *Optical Engineering, 52*(5), 057001. https://doi.org/10.1117/1.OE.52.5.057001

Noh, H., Hong, S., & Han, B. (2015). Learning deconvolution network for semantic segmentation. *Paper presented at the Proceedings of the IEEE international conference on computer vision*.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., … Fei-Fei, L.i. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision, 115*(3), 211–252.

Saha, S., Singh, G., & Cuzzolin, F. (2017). Amtnet: Action-micro-tube regression by end-to-end trainable deep architecture. *Paper presented at the Proceedings of the IEEE International Conference on Computer Vision*.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). Mobilenetv 2: Inverted residuals and linear bottlenecks. *Paper presented at the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Singh, G., Saha, S., Sapienza, M., Torr, P. H., & Cuzzolin, F. (2017). Online real-time multiple spatiotemporal action localisation and prediction. *Paper presented at the Proceedings of the IEEE International Conference on Computer Vision*.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., … Rabinovich, A. (2015). Going deeper with convolutions. *Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition*.

Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. arXiv preprint arXiv:1905.11946.

Tian, H., Li, W., Ogunbona, P. O., & Wang, L. (2018). Detection and separation of smoke from single image frames. *IEEE Transactions on Image Processing, 27*(3), 1164–1177.

Tian, H., Li, W., Wang, L., & Ogunbona, P. (2014). Smoke detection in video: An image separation approach. *International journal of computer vision, 106*(2), 192–209.

Tung, T. X., & Kim, J.-M. (2011). An effective four-stage smoke-detection algorithm using video images for early fire-alarm systems. *Fire Safety Journal, 46*(5), 276–282.

Ullah, A., Muhammad, K., Del Ser, J., Baik, S. W., & de Albuquerque, V. H. C. (2019). Activity recognition using temporal optical flow convolutional features and multi-layer LSTM. *IEEE Transactions on Industrial Electronics, 66*(12), 9692–9702.

Ullah, F. U. M., Ullah, A., Muhammad, K., Haq, I. U., & Baik, S. W. (2019). Violence detection using spatiotemporal features with 3D convolutional neural network. *Sensors, 19*(11), 2472.

Wills, A. J., Liddelow, G., & Tunsell, V. (2020). Wildfire and fire mosaic effects on bird species richness and community composition in south-western Australia. *Fire Ecology, 16*(1), 5.

Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. *Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition*.

Ye, W., Zhao, J., Wang, S., Wang, Y., Zhang, D., & Yuan, Z. (2015). Dynamic texture based smoke detection using Surfacelet transform and HMT model. *Fire Safety Journal, 73*, 91–101.

Yin, Z., Wan, B., Yuan, F., Xia, X., & Shi, J. (2017). A deep normalization and convolutional neural network for image smoke detection. *IEEE Access, 5*, 18429–18438. https://doi.org/10.1109/ACCESS.2017.2747399

Yuan, F. (2008). A fast accumulative motion orientation model based on integral image for video smoke detection. *Pattern Recognition Letters, 29*(7), 925–932.

Yuan, F., Fang, Z., Wu, S., Yang, Y., & Fang, Y. (2015). Real-time image smoke detection using staircase searching-based dual threshold AdaBoost and dynamic analysis. *IET Image Processing, 9*(10), 849–856.

Yuan, F., Zhang, L., Xia, X., Wan, B., Huang, Q., & Li, X. (2019). Deep smoke segmentation. *Neurocomputing, 357*, 248–260. https://doi.org/10.1016/j.neucom.2019.05.011

Zhang, Z., & Sabuncu, M. (2018). Generalized cross entropy loss for training deep neural networks with noisy labels. *Paper presented at the Advances in neural information processing systems*.