

Research Article

Computer Vision-Based Wildfire Smoke Detection Using UAVs

Ehab Ur Rahman ¹, Muhammad Asghar Khan ², Fahad Algarni ³, Yihong Zhang,¹
M. Irfan Uddin ⁴, Insaf Ullah,² and Hafiz Ishfaq Ahmad⁵

¹College of Information Science & Technology, Donghua University, Shanghai, China

²Hamdard Institute of Engineering & Technology, Islamabad 44000, Pakistan

³College of Computing and Information Technology, University of Bisha, Bisha, Saudi Arabia

⁴Institute of Computing, Kohat University of Science and Technology, Kohat, Pakistan

⁵Faculty of Engineering, UniversitiTeknologi Malaysia, Johor, Malaysia

Correspondence should be addressed to M. Irfan Uddin; irfanuddin@kust.edu.pk

Received 21 March 2021; Revised 14 April 2021; Accepted 20 April 2021; Published 29 April 2021

Academic Editor: Dr. Dilbag Singh

Copyright © 2021 Ehab Ur Rahman et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents a new methodology based on texture and color for the detection and monitoring of different sources of forest fire smoke using unmanned aerial vehicles (UAVs). A novel dataset has been gathered comprised of thin smoke and dense smoke generated from the dry leaves on the floor of the forest, which is a source of igniting forest fires. A classification task has been done by training a feature extractor to check the feasibility of the proposed dataset. A meta-architecture is trained above the feature extractor to check the dataset viability for smoke detection and tracking. Results have been obtained by implementing the proposed methodology on forest fire smoke images, smoke videos taken on a stand by the camera, and real-time UAV footages. A microaverage F1-score of 0.865 has been achieved with different test videos. An F1-score of 0.870 has been achieved on real UAV footage of wildfire smoke. The structural similarity index has been used to show some of the difficulties encountered in smoke detection, along with examples.

1. Introduction

Wildfire is a colossal threat to damaging the human and wildlife ecosystem. Statistics show that wildfires in Northern California in the United States caused more than 40 deaths and about 50 missing individuals in 2015 [1–3]. There were some major wildfire outbreaks in several countries around the world in the year 2019. It was seen to be the most unfortunate year for such incidents. Moreover, 3500 square miles of the Amazon rainforest have been burnt down by wildfires. A forest fire recently caused 89 fatalities in Australia and burned 3500 homes. It became of such incidents of great importance to detect wildfires accurately in advance when it turns into chaos. Traditional methods of wildfire detection, which are mainly based on human observation from watchtowers, are inefficient. The inefficiency is primarily due to the spatiotemporal connection. Unmanned aerial vehicles (UAVs) have been extensively used as

monitoring tools for the last couple of years [4–8]. A high-definition and lightweight cameras can generate an aerial photograph with specific location information when connected to UAVs along with global positioning systems (GPSs) [9]. Besides, cost-effectively, a well-organized swarm of UAVs can easily accomplish a complex task.

Image recognition attained a state-of-the-art performance using deep convolutional neural networks (DCNNs); its architecture and learning scheme leads to an effective extractor of sophisticated, high-level features that are highly robust to input transformations [10]. However, implementing deep learning and computer vision techniques in the application for wildfire smoke detection is scarce. Besides, the limitations and difficulties in such kinds of techniques are not widely discussed. Object detectors mainly based on video fire detection methods can be categorized into two classes, i.e., flame detection and smoke detection. Since the smoke generated by forest fires is observable before

the flames, video smoke detection acquires more attention for early fire alarm in forest fire protection engineering. The traditional video smoke detection methods mainly emphasize the combination of static and dynamic features for smoke detection. The typical features of smoke contain color, texture, motion orientation, and so on [11]. These different characteristics can get better performances in specific images dataset [12]. However, due to the poor robustness of algorithms, the performances incline to be unfavorable in different images dataset, and those approaches can barely remove sophisticated interference in real engineering applications.

Currently, object detection achieved a lot of progress due to the use of GA, PSO, ANN, and DCNNs [13–18]. Modern object detectors founded on these networks—such as Faster R-CNN [19], SSD [20], and YOLOv3 [21]—are now robust enough to be deployed in customer products (e.g., Google Photos and Pinterest Visual Search), and some are fast enough to be run on mobile devices such as MobileNet. Most of these object detectors are deployed in different applications [22]. However, it can be challenging for practitioners to select what architecture is more appropriate to their application. Standard accuracy metrics do not clarify the entire options, such as mean average precision (mAP); for practical deployment of computer vision systems, running time and memory usage are also important. For example, mobile devices in many cases need a small memory footprint, and self-driving cars need real-time executions. SSD achieves a good trade-off between speed and precision. SSD runs a convolutional network on input image a single time and calculates a feature map following a small 3×3 sized convolutional kernel on this feature map to predict the bounding boxes and categorization probability. SSD also uses anchor boxes on various aspect ratios similar to Faster-RCNN and learns the counterbalance to a definite extent than learning the box. To handle the scale, SSD predicts bounding boxes after multiple convolutional layers. Since every convolutional layer function at a diverse scale, it can detect objects of varying scales [23].

There are several ways of comparing images as if they are identical or near-identical such as structure similarity index measure (SSIM), mean square error (MSE), normalized color histogram, and local sensitivity hashing. These methods have various benefits over one another. SSIM tries to model the modification of the image's structural details. SSIM is more robust capable of disclosing changes in the image structure rather than just the perceived change [24].

1.1. Contributions. In this work, we present a dataset, grouping several images from different sources such as thin, dense with different color, and texture smoke images, taken from different scenarios such as wildfire and other emergency conditions such as building fires and fires from an explosion. The SSD Inception-V2 state-of-the-art models are trained, and their different parameters such as dropout, batch normalization, and learning rate are tuned to choose the best model for real-time fire detection in videos. Comparisons of the results are obtained on several wildfire

smoke videos taken by a UAV and smoke images with different kinds of backgrounds and lighting conditions. Also, some of the limitations and difficulties are discussed found during the study of this research.

1.2. Organization of the Paper. The other sections of the paper are presented as follows. Section 2 discusses the different datasets followed by the training of the object detector in Section 3. Section 4 presents details on the results and concludes the work, and Section 5 concludes the entirety of the paper.

2. Material and Methods

2.1. Real Smoke Training Images. The dataset used to train the model consists of 14096 images of real smoke, both comprised of thin and dense smoke. The smoke is typically generated in different scenarios. One set of smoke is generated from burning dry leaves and small bushes, which is one of the fuel causes of igniting a forest fire and smoke. The images are shown in Figures 1(a) and 1(b) as dense and thin smoke, respectively. Another set of images consists of smoke images taken in different light conditions such as yellow and white light as it affects smoke color and texture, as shown in Figures 1(c) and 1(d). These images have been taken from [11], which is also comprised of some other images having added smoke to the forest background. Examples of these images are presented in Figures 1(e) and 1(f). A third set consists of smoke images taken from various open sources from the Internet that present real-time emergencies such as an apartment on fire or a vehicle. Such images are shown in Figures 1(g) and 1(h).

2.2. Test Images. The proposed methodology and object detector are evaluated based on test smoke images to check the generalization ability of the trained object detection model. These images are a collection of smoke images taken from the UAV with a camera of 12 MP [11], and also some of them are taken from phone cameras specifically using an iPhone 6s camera with a 12 MP rear camera, while the rest are selected from the open-source datasets available on the Internet. Figure 2 illustrates such kinds of images.

2.3. Test Videos. With static camera videos as well as forest smoke videos in real-time taken by UAV, the performance of various object detection models is also tested. The focus of this work is on choosing an object detector that for real-time object detection that has a better trade-off between speed and precision. Table 1 displays the technical specifications of the UAV, along with the UAV image (see Figure 3) [25], which can be used to capture the smoke images.

3. Training and Detection

DCNNs have presently dominated computer vision tasks in which region-based object detection methods are state-of-the-art. These methods have different advantages, such as removing the gruesome work of manual feature extraction.



FIGURE 1: Training images taken from different sources.



FIGURE 2: Test images from different sources.

The network learns patterns from the images without needing any preprocessing. Currently, several different architectures of feature extractors are available. Selecting one for a specific application is a trivial subject. According to [26], different speed versus accuracy comparison has been presented. Figure 4 illustrates such speed versus accuracy trade-offs between the current state-of-the-art object detection models. Speed and accuracy are both of keen importance for real-time smoke detection. From Figure 4, it is clear that single shot detectors (SSDs) achieve a better trade-off in the aspect of swiftness and accuracy. Also, SSD

provides a good precision in detection objects of different sizes as compared with Faster-RCNN architecture.

In this work, a pretrained feature extractor has been trained by transfer learning with a custom classifier with two fully connected layers and a final log softmax classifier to classify the proposed dataset into two classes with one class of smoke including both dense and thin smoke images and another class having fire images. The classification task was aimed to check the feasibility of the proposed dataset. The promising results have been presented in the results section of this paper, along with some examples. SSD as a meta-

TABLE 1: Technical parameters of the UAV.

Technical parameters	Value
Camera lens	FOV 81.9° 25 mm
Take-off weight	300 g
Video resolution	FHD: 1920 × 1080 30 p
Max hovering time	15 minutes
Max flight speed	31 mph
Endurance	16 minutes
Positioning system	GPS/GLONASS
Obstacle sensing range	11–16 ft (0.2–5 m)
Operating temperature range	320 to 1040 F
Transmitter power(EIRP)	2.4 GHz
Max transmission distance	100 m (distance), 50 m (height)
Dimensions	143 × 143 × 55 mm
Operating frequency	2.412–2.462 GHz; 5.745–5.825 GHz
Camera sensor	½.3" CMOS; effective pixels: 12MP
ISO range	Video: 100–3200 photo: 100–1600
Electronic shutter speed	2–1/8000 s
Max vidoe bitrate	24 mbps
Photo format	JPEG
Video format	MP4
Controllable range	Pitch: –850 to 00
Stabilization	2-axis mechanical (pitch, roll)
Velocity range	36 kph at 2 m above ground
Altitude range	0–8 m
Operating range	0–30 m

architecture and Inception-V2 as a feature extractor have been chosen to be more suitable for real-time smoke detection as they offer better speed versus accuracy trade-off, as shown in Figure 4 taken from the study by Huang [26].

In the proposed work, different images taken from different sources are collected together to increase the richness of the training data primarily focusing on images having thin smoke as it serves as an alarm before a fire starts, and there is an immense need to detect the smoke at this starting stage to prevent the ignition and spreading of the wildfire. Before testing the proposed model, we trained the model on more than 14000 training image samples comprised of both dense smoke and thin smoke and smoke in a different light, color, and texture as well as different backgrounds. For validation of the model, more than 3100 images were used. This approach was aimed to improve the generalization ability of the smoke detection model. We then tested the model using wildfire smoke images taken from a drone, mobile phone camera, and some from Internet open sources which present real-time scenarios of both forest fire and in day-to-day life emergencies so that the proposed approach can be used in any kind of situation for future projects.

Test videos taken from standby cameras of both thin and dense smoke with different backgrounds, lighting conditions, and from different distances are used for inferencing. Also, real footage of wildfire and smoke taken by a drone is tested. Results have been presented in Section 4 for analysis, along with some discussions on the limitations and difficulties found in the research of this study of wildfire smoke detection. The overall workflow of the experiments is presented in Figure 5.

First, the feature extractor Inception-V2 pretrained on the COCO dataset is trained by transfer learning using the proposed dataset with the object detector SSD for making predictions of smoke. The trained model is then tested with different test videos and images to check the feasibility of the model.

Classification results of the wildfire smoke classifier model trained by using the dataset comprising of both smoke and fire images are presented in Table 2. Metrics used for evaluation are presented in the study by Bashir and Porikli [27] and Forman and Scholz [28]. The metrics are True positive (TP), False Positive (FP), True Negative (TN), False Negative (FN), and False alarm rate (FAR) along with detection rate recall precision and F-score. Such kind of metrics are currently used in the computer vision community for evaluation object detection models. Another metric that is used for comparing the similarity in the structure formation of two images is SSIM [24]. Formulas of the metrics are presented as follows:

$$\text{recall (Re)} = \frac{TP}{(TP + FN)}, \quad (1)$$

$$\text{precision (Pr)} = \frac{TP}{(TP + FP)}, \quad (2)$$

$$F - \text{score} = \frac{2 \cdot (Pr \cdot Re)}{(Pr + Re)}, \quad (3)$$

$$\text{false alarm rate (FAR)} = \frac{FP}{(TP + FP)}, \quad (4)$$

$$\text{accuracy} = \frac{(TP + TN)}{FN}, \quad (5)$$

$$\text{tracker detection rate (TDR)} = \frac{TP}{TG}, \quad (6)$$

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}. \quad (7)$$

Equation (7) presents a comparison of two windows, i.e., small subsamples despite the whole image, leading to a better approach that can sense for changes in the structure of the image. The parameters of equation (7) confine the (x, y) location of the $N \times N$ window in each image, the mean of the pixel intensities in the x and y direction, and also the variance of intensities in the x and y direction, along with the covariance.

4. Research Findings and Discussion

As introduced in Section 3, we trained a classifier on the smoke image dataset along with another set of fire images. Some of the classification results are shown in Figure 4 along with top k class probabilities. Figure 6(a) presents a test image taken from a thin smoke dataset, Figure 6(b) presents a test image taken from dense smoke images, and Figure 6(c) presents a test image from a fire image.

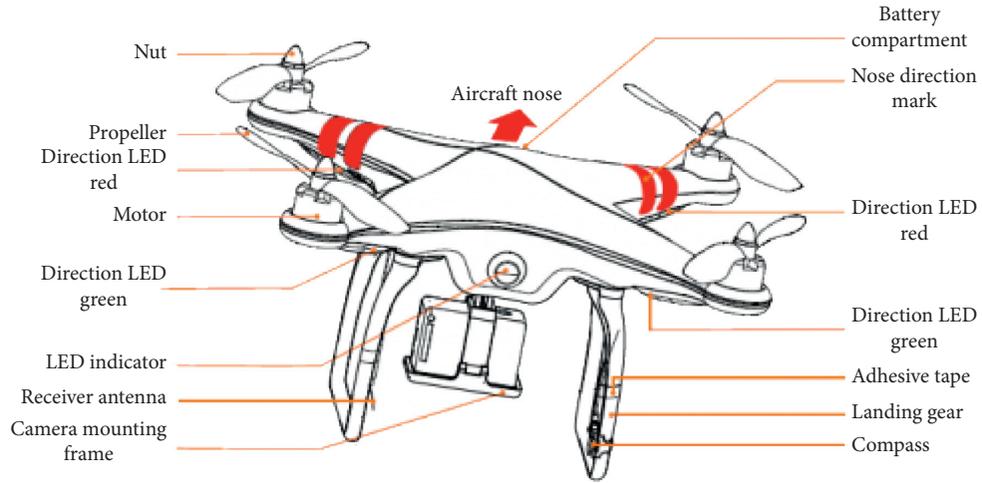


FIGURE 3: Main on-board components of the UAV (quadcopter) [25].

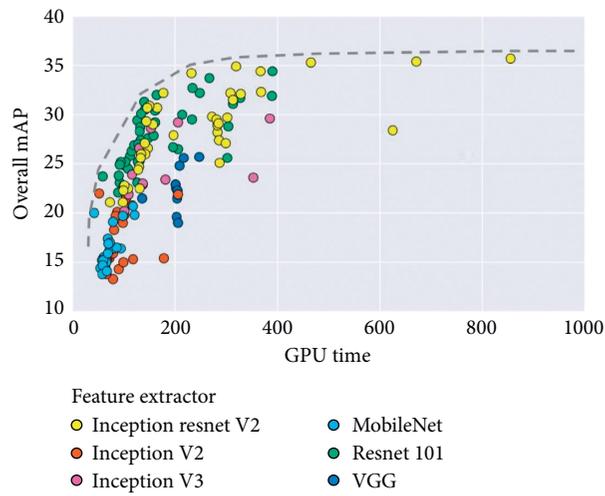


FIGURE 4: Speed vs. accuracy trade-off.

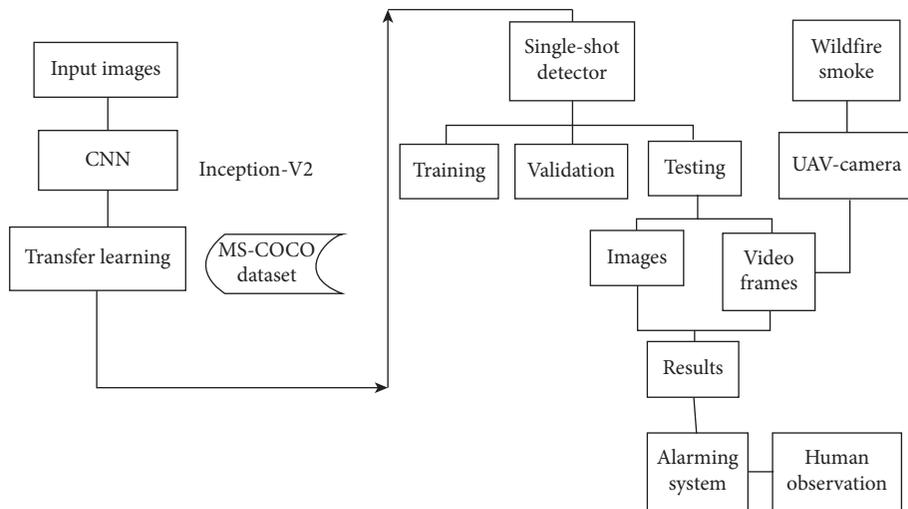


FIGURE 5: Diagram of the proposed methodology for smoke detection.

TABLE 2: Two class classification results of the classifier.

Test set	Samples	TP	FP	TN	FN	FAR	Recall	Precision	F-score
Smoke	20	18	2	0	0	0.1	1.00	0.90	0.947
Fire	20	18	0	2	0	0	1.00	1.00	1.00

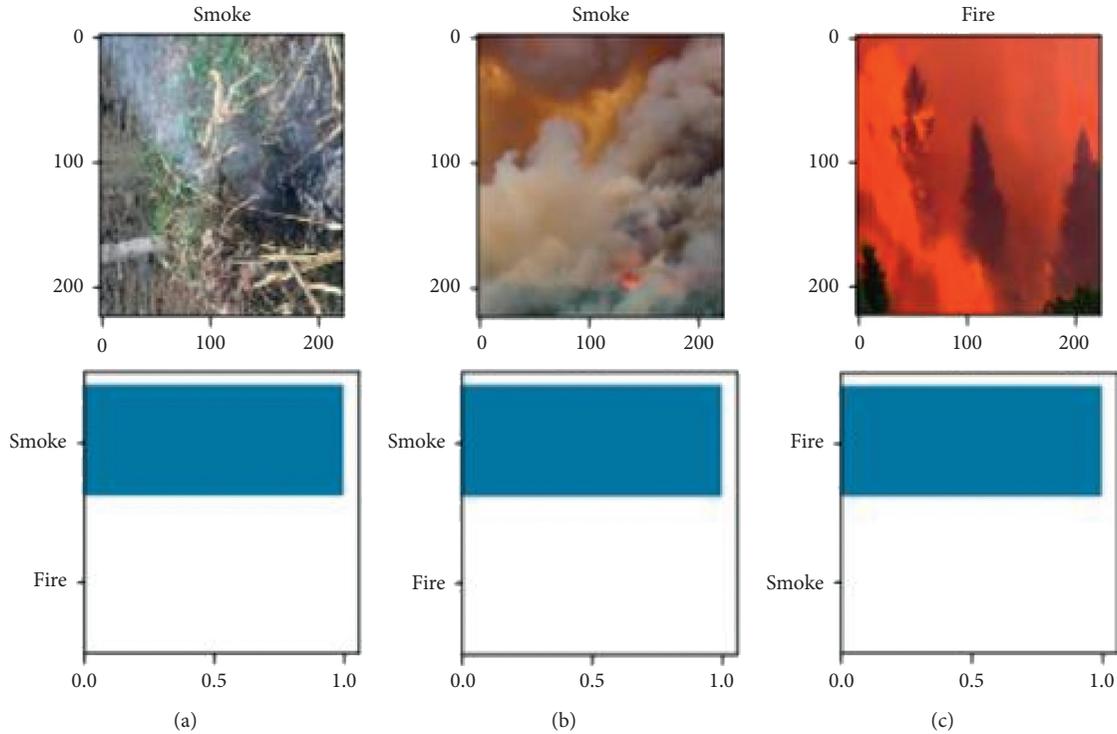


FIGURE 6: Classification result of feature extractor with high-class probabilities.

Table 2 presents obtained results of the classification of the dataset into two classes, i.e., fire and smoke. The results from Table 2 show that the feature extractor is generalizing well and is learning different smoke patterns, i.e., thin, dense, white, and so on. This small experiment was aimed to prove the viability of using the dataset in the training of the object detector for real-time smoke detection.

The trained wildfire smoke detector model, i.e., SSD Inception-V2 has been tested with images taken from a UAV shown in Figure 7(a) along with the detection score and bounding boxes. Figure 7(b) presents the frames captured from video of thin smoke and dense smoke generated by burning dry leaves and shrubs. These videos are recorded on a mobile phone camera. Frames captured of different instances from real-time UAV footage of wildfire and smoke are presented in Figure 7(c) with respective bounding boxes. Figure 7(e) presents a frame comprising of both dense and thin smoke along with detections.

The test results are presented in Table 3. It is observed that Video 2, i.e., a dense smoke video has the highest F-score among all that is because of the rich texture, shape, and color. The lowest F-score is observed from thin video samples and that is because of light color and features captured by the camera, but still, the model achieves an accuracy of 64% and an F-score of 0.747, proving the

feasibility of the dataset on thin smoke also. Another good result is from the drone footage sample which is the main focus of this study to detect smoke in such kind of scenario achieved 83.2% accuracy and an F-score of 0.870 which proves the suitability of this study. The results from Table 3 show that this approach is feasible to implement in real-time applications.

Figure 8 presents the mean average recall (mAR) and mean average precision (mAP) of each test dataset that has been evaluated with the wildfire smoke detection model. These metrics are famous for evaluating the overall generalization ability of object detection models. The respective figure proves that our smoke detection model efficiently generalized to different datasets which comprise distinctive image sets with different properties such as thin smoke and dense smoke, and an image of smoke taken from different angles and approaches.

Some of the limitations have been observed while acquiring the results. Figure 9(a) shows smoke due to wildfire, and the frame is captured from real footage. The texture and color of this image are nearly similar to the texture and color of the cloud shown in Figure 9(b). Such a kind of coincidence makes it difficult for the object detector to differentiate between them. The structural similarity between the two images has been calculated

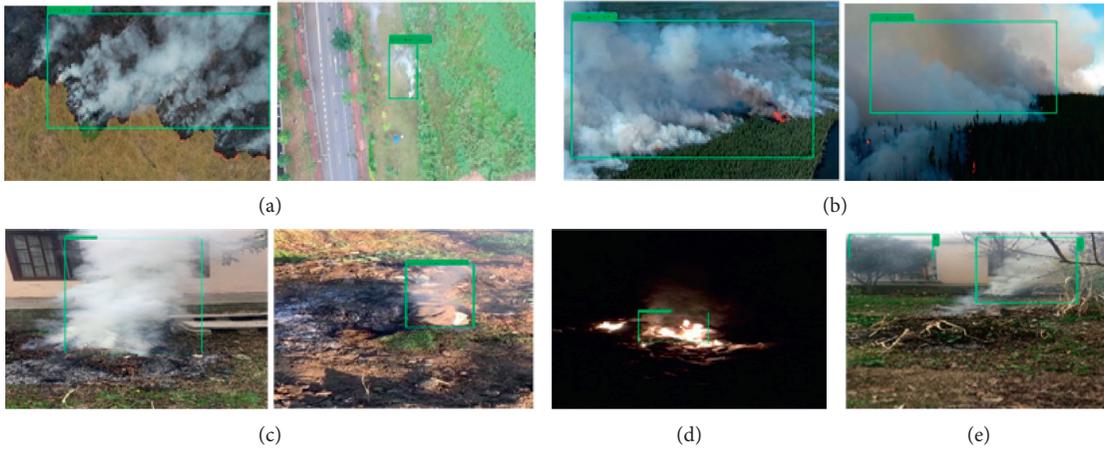


FIGURE 7: Sample's result taken from different test videos along with detection boxes.

TABLE 3: Results of real smoke video samples.

Test set	Smoke sample	Nonsmoke samples	TP	FP	TN	FN	FAR	Accuracy	Detection rate/recall	Precision	F1-score
Images	74	30	50	21	32	1	0.296	0.766	0.980	0.704	0.818
Video 1 (dense smoke near)	816	0	599	216	0	1	0.265	0.734	0.998	0.735	0.847
Video 2 (dense smoke far)	307	0	301	6	0	0	0.019	0.980	1.000	0.980	0.989
Video 3 (thin + dense smoke)	313	0	245	68	0	0	0.217	0.782	1.000	0.783	0.878
Video 4 (drone footage)	3983	1479	3099	884	1447	32	0.221	0.832	0.989	0.778	0.870
Video 5 (thin smoke)	308	0	184	124	0	0	0.402	0.600	1.000	0.597	0.747
Video 6 (thin smoke)	289	0	185	104	0	0	0.359	0.640	1.000	0.640	0.780

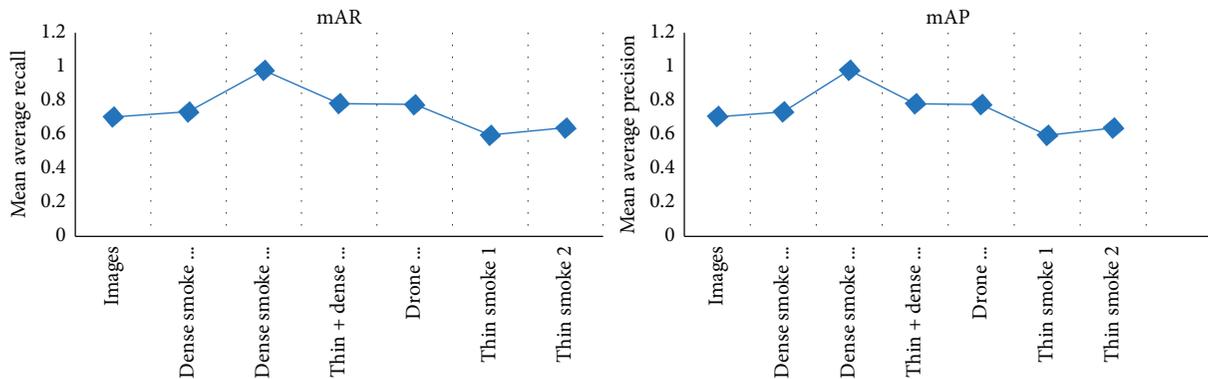


FIGURE 8: Mean average recall (mAR) and mean average precision (mAP) for different test datasets.

using the structure similarity index measure (SSIM). The value of SSIM calculated for these images shows that they have a 63% structure similarity. Both of them have been detected as smoke by the trained object detector. The confidence score and bounding box are shown in Figure 9(c). Also, in Figure 9(d), a video of the fire taken at

night in the dark is also tested with the trained object detector. Though the smoke detector does detect some of the frames, the overall accuracy is unsatisfactory. This was meant to give intuition to the readers for such kinds of difficulties, which may be addressed in future research by designing new approaches.

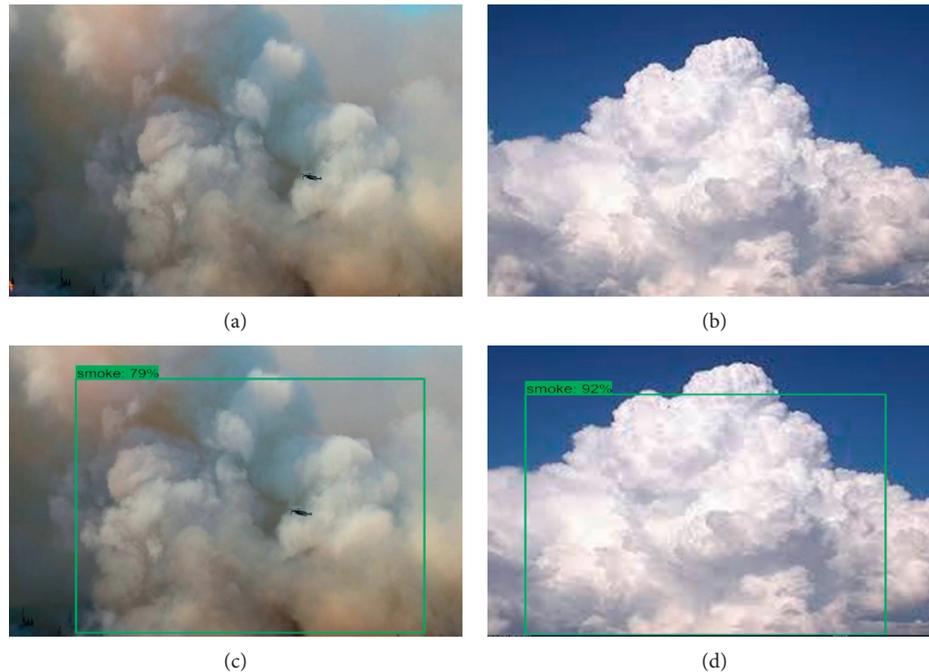


FIGURE 9: Smoke vs. cloud similarity comparison with SSIM value.

5. Conclusion

In this paper, SSD Inception-V2 was chosen to be a viable detector of wildfire smoke in videos taken by UAVs both in terms of accuracy and speed. Different smoke image datasets such as one generated by using a synthetic process and another from real smoke images are used to train the model. One of the significant solutions is presented to detect thin and dense smoke in videos taken by UAVs as previous methods comprise images or static camera videos. The test results promise of extending the solution to real-time drone surveillance. An F1-score of 0.784 and 0.747 has been achieved on test videos of thin smoke surpassing the previous literature. Limitations and difficulties found in the study are discussed along with an example using structural similarity index as a quantifying parameter. In the future, the proposed solution can be extended to detect smoke in real-time UAV footage in different light and weather conditions along with designing a fire alarm. The performance of the model on thin smoke can be further improved by enriching the thin smoke image dataset mainly taken by UAVs in different weather and light conditions.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] A. D. Syphard, V. C. Radeloff, N. S. Keuler et al., "Predicting spatial patterns of fire on a southern California landscape," *International Journal of Wildland Fire*, vol. 17, no. 5, pp. 602–613, 2008.
- [2] N. Nauslar, J. Abatzoglou, and P. Marsh, "The 2017 north Bay and southern California fires: a case study," *Fire*, vol. 1, no. 1, p. 18, 2018.
- [3] T. John and A. Abatzoglou, "Impact of anthropogenic climate change on wildfire across western US forests," *Park Williams Proceedings of the National Academy of Sciences*, vol. 113, no. 42, pp. 11770–11775, 2016.
- [4] I. Colomina and P. Molina, "Unmanned aerial systems for photogrammetry and remote sensing: a review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 92, pp. 79–97, 2014.
- [5] S. S. V. Vijayakumar, C. S. Kumar, V. Priya, L. Ravi, and V. Subramaniaswamy, "Unmanned aerial vehicle (UAV) based forest fire detection and monitoring for reducing false alarms in forest-fires," *Computer Communications*, vol. 149, pp. 1–16, 2020.
- [6] F. Noor, M. A. Khan, A. Al-Zahrani, I. Ullah, and K. A. Al-Dhlan, "A review on communications perspective of flying ad-hoc networks: key enabling wireless technologies, applications, challenges and open research topics," *Drones*, vol. 4, no. 65, pp. 1–14, 2020.
- [7] B. Alzahrani, O. S. Oubbati, A. Bernawi, A. Atiquzzaman, and D. Alghazzawi, "UAV assistance paradigm: state-of-the-art in applications and challenges," *Journal of Network and Computer Applications*, vol. 166, no. 102706, pp. 1–44, 2020.
- [8] M. A. Khan, I. M. Qureshi, and F. A. Khanzada, "Hybrid communication scheme for efficient and low-cost deployment of future flying ad-hoc network (FANET)," *Drones*, vol. 3, no. 16, pp. 1–20, 2019.

- [9] X. Li, Y. Zhao, J. Zhang, and Y. Dong, "A hybrid PSO algorithm based flight path optimization for multiple agricultural UAVs," in *Proceedings of 2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI)*, San Jose, CA, USA, November 2016.
- [10] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, June 2014.
- [11] Q.-x. Zhang, G.-h. Lin, Y.-m. Zhang, G. Xu, and J.-j. Wang, "Wildland forest fire smoke detection based on faster R-CNN using synthetic smoke images," *Procedia Engineering*, vol. 211, pp. 441–446, 2018.
- [12] E. U. Rahman, Y. Zhang, S. Ahmad, H. I. Ahmad, and S. Jobaer, "Autonomous vision-based primary distribution systems porcelain insulators inspection using UAVs," *Sensors*, vol. 21, no. 3, p. 974, 2021.
- [13] Z. Ali and T. Mahmood, "Complex neutrosophic generalised dice similarity measures and their application to decision making," *CAAI Transactions on Intelligence Technology*, vol. 5, no. 2, pp. 78–87, 2020.
- [14] T. Sangeetha and G. M. Amalanathan, "Outlier detection in neutrosophic sets by using rough entropy based weighted density method," *CAAI Transactions on Intelligence Technology*, vol. 5, no. 2, 2020.
- [15] C. Zhu, W. Yan, X. Cai, S. Liu, T. H. Li, and G. Li, "Neural saliency algorithm guide bi-directional visual perception style transfer," *CAAI Transactions on Intelligence Technology*, vol. 5, no. 1, pp. 1–8, 2020.
- [16] C.-F. J. Kuo, J.-M. Liu, M. L. Umar, W.-L. Lan, C.-Y. Huang, and S.-S. Syu, "The photovoltaic-thermal system parameter optimization design and practical verification," *Energy Conversion and Management*, vol. 180, pp. 358–371, 2019.
- [17] M. Safa, M. Ahmadi, J. Mehrmashadi et al., "Selection of the most influential parameters on vectorial crystal growth of highly oriented vertically aligned carbon nanotubes by adaptive neuro-fuzzy technique," *International Journal of Hydromechatronics*, vol. 3, no. 3, p. 238, 2020.
- [18] B. R. Murlidhar, R. K. Sinha, E. T. Mohamad, R. Sonkar, and M. Khorami, "The effects of particle swarm optimisation and genetic algorithm on ANN results in predicting pile bearing capacity," *International Journal of Hydromechatronics*, vol. 3, no. 1, p. 69, 2020.
- [19] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: towards real-time object detection with region proposal networks," in *Proceedings of Advances in Neural Information Processing Systems*, Montreal, Canada, December 2015.
- [20] W. Liu, D. Anguelov, D. Erhan et al., "Ssd: single shot multibox detector," in *Proceedings of European Conference on Computer Vision*, Amsterdam, Netherlands, October 2016.
- [21] H. Shah, *YOLO Vs. SSD: Choice of a Precise Object Detection Method*, <https://technostacks.com/blog/yolo-vs-ssd>, 2020.
- [22] E. U. Rahman, Y. Zhang, S. Ahmad, H. I. Ahmad, and S. Jobaer, "Autonomous vision-based primary distribution systems porcelain insulators inspection using UAVs," *Engineering*, 2020, preprint.
- [23] H. Shah, *What is the Main Difference Between Yolo and Ssd?*, Technostacks Infotech Pvt. Ltd., Ahmedabad, India, 2018, <https://technostacks.com/blog/yolo-vs-ssd/>.
- [24] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [25] "General thoughts on putting together your own quadcopter", Available online: <https://unleashthebot.com/best-drone-kits/>, 2021.
- [26] J. Huang, "Speed/accuracy trade-offs for modern convolutional object detectors," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, June 2017.
- [27] F. Bashir and F. Porikli, "Performance evaluation of object detection and tracking systems," in *Proceedings 9th IEEE International Workshop on PETS*, New York, NY, USA, June 2006.
- [28] G. Forman and M. Scholz, "Apples-to-apples in cross-validation studies," *Acm Sigkdd Explorations Newsletter*, vol. 12, no. 1, pp. 49–57, 2010.