Early Thermal Forest Fire Detection using UAV and Saliency map

B. Dilli
Dept. of Mechanical Engineering
Thiagarajar College of Engineering
Madurai, Tamil Nadu, India
dilli@student.tce.edu

M. Suguna
Dept. of Computer Science Engineering
Thiagarajar College of Engineering
Madurai, Tamil Nadu, India
mscse@tce.edu

2022 5th International Conference on Contemporary Computing and Informatics (IC3I) | 979-8-3503-9826-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/IC3I56241.2022.10072674

***Abstract*— Wildfire is a dangerous disaster that threatens not only animals but also humans. Traditionally, Ground crew inspections form the basis for firefighting monitoring systems, which have several limitations. With the help of UAVs, before a wildfire becomes out of control, it is simple to spot it in its early stages. The last few years have seen a surge in interest in autonomous wildfire early detection using various deep learning algorithms based on UAV-based visual data. But the area of night vision surveillance is unattended. In this research, we concentrate on detecting the early stage of wildfires using YOLO model, a DL based object detection library on thermal images from UAVs and also the drawbacks in using the thermal images are solved using saliency maps fused with thermal images, which is then compared with the benchmark of the YOLO model trained only on thermal images. The proposed methodology is found to be capable enough to provide technical support in night surveillance to reduce the catastrophic losses of wild forest fires in their early stages and in terms of forest resources, and human and animal lives.**

Keywords*— *Saliency Map, YOLOv7, Forest fire, Unmanned Aerial Vehicle

I. INTRODUCTION

Wild Forest Fires being the most dangerous natural disaster puts both human and animal life in danger. Recently, there have been a number of large-scale wildfires across the globe. One of them is the Australian wildfires of 2020 where 28 people died as a result of this fire, which burned more than 17.9 million acres all throughout Australia's six states [1]. Another more recent example is the wildfires in California, Oregon, and other western US states which burned millions of acres. According to the BBC, 6.7 million acres have burned as of September 17, 2020, and more than 30 people have dead [2]. The main causes of forest fires are human negligence, lightning, and complete exposure to extremely hot temperatures. But the main cause is due to human activity [3]. The campfire that is left over during the night serves as the main cause of wildfires. This makes early Forest fire detection a hot topic for research to reduce the risks and losses of life. Moreover, It aids the firefighters in their efforts to put out the fire in its early stages

Today's detection methods include topics like satellite photography, watching towers, long-distance video recording, etc. However, these do not provide a solution to improve the effectiveness of the detection of forest fires. Watchtowers being an olden method have a number of drawbacks, including a narrow field of view, expensive construction, and are highly vulnerability to fire damage, which would incur added expenses. It is difficult to notice fire areas at the right time when using satellites with a very wide field of view because of costs, flexibility, and spatial/temporal imaging resolution constraints. [4].

Recent developments in early wildfire detection using UAVs and DL based object detection techniques have shown promising alternatives for monitoring wildfires [5,6]. Since optical visible cameras have a disadvantage in night vision. Thermal drones with vision imaging cameras would capture videos and images of any object or material by detecting heat, which has a variety of useful applications; it serves over visual images in low light surveillance. But thermal images are less distinct, because of thermal reflections, blurry edges, high noise, and UAV jitter which causes image blur and the object detection models suffer from false detections. To solve this problem we suggest fusing saliency maps with thermal images to improve the image data.

Since the UAV has a limitation of low computing hardware configuration but proposed Deep learning approach requires high hardware configuration. The UAV transmits the thermal image data to the ground station where the proposed Wildfire Detection is carried out. Figure 1 shows the required platform based on its risk calculation.

II. RELATED WORK

Traditionally, forest fire monitoring systems involve the use of watchtowers and satellite photography. Recent years have seen a greater success in the development of new technologies in wildfire monitoring systems, such as remote sensing-based warning systems and computer vision algorithms. UAVs equipped with optical sensors serve as a better monitoring system over other systems which have cost and efficiency limitations. Computer Vision algorithms based on DL can be used to identify fire and smoke in forests and natural areas. Wildfire detection using flame has been the subject of some investigations [7][8]. Since the fire may be disguised in its early stages, especially in deep forests, some targeted fire detection by smoke [9] appears more appropriate for early detection. Recent research has mostly focused on simultaneously detecting smoke and flame in order to get over the restrictions associated with focusing on just one object. They can be accomplished using the object detection approach and the image classification method. The authors Srinivas and Dual [10] suggested using a straightforward CNN. Using a structure akin to AlexNet's, they were able to attain a 95% accuracy. Zhang et al. [11] have proposed a DL approach to identify and locate the precise location of wildfires using a pair of CNN architectures, the first of which has the propensity to identify the fire and the second of which has the propensity to determine whether the input image contains a fire or not. The authors of Barmpoutis et al. [12] proposed DCNN model using UAV and satellite based imagery from a 360 degree RGB camera datatset which achieve an F1-Score score of 94.8%, 93.9%, and 94.6% for flame, smoke, and combined flame and smoke, respectively.

979-8-3503-9826-7/22/$31.00 ©2022 IEEE

YOLO, a single-stage detector, is used in many wildfire monitoring systems. Alexandrov et al. [13] compared YOLOv2, Faster RCNN and SSD for aerial detection of fire and smoke. Among these models, with FPS = 6, YOLOv2 attained a F1-score of 99.14%, and accuracy of 98.3%. Yadav [14], in his study on the detection of wildfires, used different versions of YOLO. Out of which YOLOv3-SPP had a slightly higher mAP of 97.81% than YOLOv3 (97.6%) on the single flame dataset. Jiao et al.[15], [6] proposed an improved version of YOLOv3-tiny by adding four individual Darknet convolutional layer, Batch Normalization layer, and Leaky ReLU layer and compared it with YOLOv3 to detect both flame and smoke from real time UAV equipped with DJI MANIFOLD onboard computer with 6.5FPS and produced an 82% precision rate. Goyal et al.'s [7] use of YOLOv3 yielded an F1-Score of roughly 91% for locating wildfires and alerting the proper authorities. Zhao et al.[16] made Fire-net, a Deep Convolutional Neural Network to detect wildfire from UAV imagery with 98% accuracy, and used a saliency detection algorithm to find the fire spots and segment fire areas. Though they are good at wildfire detection in early stages during the day time, they can't detect the wildfire early during the night time. Thus this paper completely focused on detecting wildfires in early stages during night time.

III. PROPOSED WORK

Human interventions are the main reason for 95% of forest fires [3]. So, the need for forest surveillance is not only necessary in the daytime but also at night time. In this research, we use the idea of fused thermal images generated from saliency maps of their respective thermal images to distinguish the main object from a complicated background. We hope that such a system will perform more effectively, particularly when wildfires in thermal images are harder to distinguish from their surroundings due to thermal reflections. And since saliency maps ignore all textural data available in the thermal images we add the saliency maps to their respective thermal images. We implement this by replacing the matching saliency maps for one duplicate channel in the three-channel thermal images as shown in Figure 2. Thus the fusion of saliency maps greatly helps in retaining the textural information of the image while distinguishing the wildfire from the background. We then train our YOLOv7 model on (i) only Thermal images (ii) Saliency Maps generated using the BAS-Net model (iii) Fused thermal images.

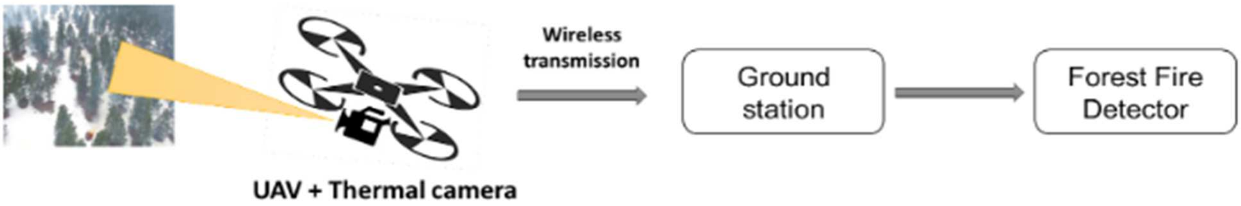

Fig. 1: Platform for UAV based thermal forest fire detection

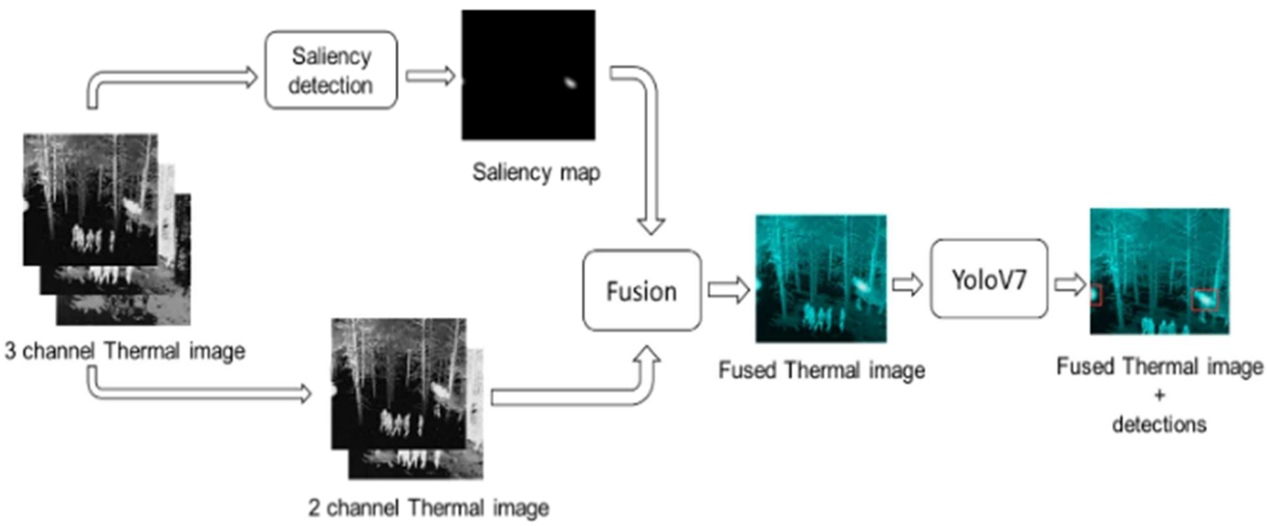

Fig. 2: Procedure for proposed thermal image forest fire detection with saliency maps

A. *Overview of YOLOv7*

The most recent work in the YOLO series is the YOLOv7. The YOLOv7 algorithm combines a number of several bag-of-freebies methods to significantly improve detection accuracy while utilizing less processing power. Using around 40% of the function parameters, this model produces results that are superior to all other object detection models without sacrificing accuracy. In the original paper [17], E-ELAN is proposed as the general architecture, and expand, shuffle, and merge cardinality are used to achieve the capacity to continuously improve the learning power of the network

without deviating from the initial gradient path. E-ELAN can direct various computational block groups to learn a wide range of features. In order to preserve the model's original features and optimal structure, the research also suggests a compound model scaling approach. The paper introduces dynamic label assignment and model re-parameterization, analyzes their current issues, and enhances them in terms of network optimization strategy. Because RepConv [18] has identity connections, the author believes that direct access to a ResNet [19] or DenseNet [20] cascade will break the network's structure by providing more gradient diversity for various characteristic graphs. The identity connection in RepConv was dropped since the author realized that re-parameterized convolution and other networks worked well together; as a result, they developed the reparameterized convolution they had in mind. The study addresses the latter utilising the idea of deep supervision [21] and adds a second auxiliary head structure as an auxiliary loss in the middle layer of the network to modify the weight of the shallow network. For this structure, a new label assignment technique is created.

B. *Deep Saliency Network: BAS-Net (Boundary-Aware Salient Object Detection Network)*

BAS-Net [22] is a Boundary-Aware Segmentation Network that comprises a predict-refine architecture and a hybrid loss, for highly accurate image segmentation. A densely supervised encoder-decoder network and a residual refinement module make up the predict improve architecture, which is used to both predict and refine a segmentation probability map. Figure 3 shows the architecture of BASNet model. It consists of a Predict module which is similar to the U-shape-Net (Olaf et al., 2015) and a Residual Refining module. The model starts with an encoder phase that consists of a CNN layer and six fundamental res-blocks taken from ResNet-34. The module then performs a six-phase decoding phase that preserves symmetry. Each stage begins with a CNN layer followed by batch normalization (BN) and a ReLU activation function. The concatenated feature maps of the output from the stage before it and the stage that corresponds to that stage in the encoder serve as the input for each stage. This module generates an unreliable saliency map with erroneous object boundaries. The residual refinement module then learns the discrepancy between the anticipated saliency map and the ground truth, which is used to further refine the saliency map of the prediction module. Both an encoder phase and a decoder phase exist in this approach. Both the encoder and the decoder have four stages, in contrast to the predict module. The resampling stages employ non-overlapping max pooling and bilinear interpolatio. The improved saliency map combined with the initial thermal image, is the result.

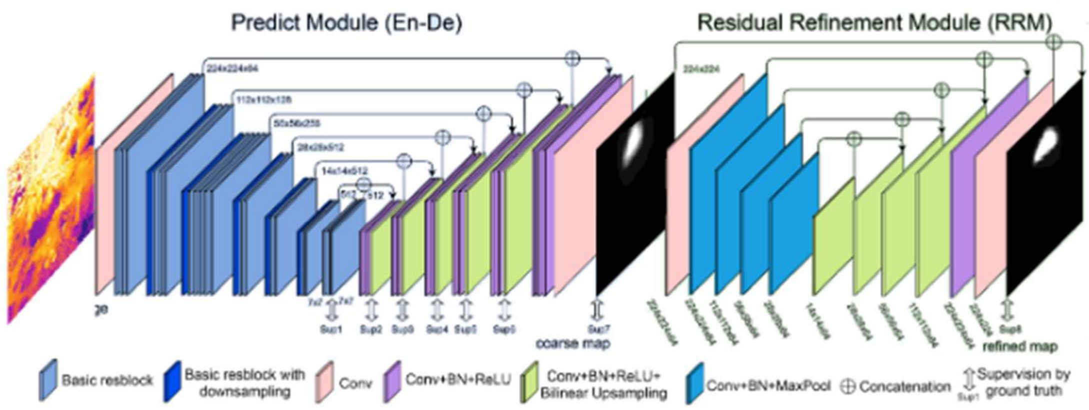

Fig. 3: Architecture of BAS-Net Saliency Detection Model

IV. DATASET AND EVALUATION PROTOCOLS

For training the deep neural network, data samples are needed in large numbers. But, currently, the thermal dataset for Wildfire from the perspective of UAV is limited. In our work, we make use of the unannotated thermal images from the FLAME Dataset [23]. These videos are taken from a test conducted in a ponderosa pine forest with assistance from the Flagstaff (Arizona) Fire Department, who burned brush piles on city-owned land on Observatory Mesa. The prescribed test was conducted on January 16th, 2020, when the temperature was 43°F (6°C), there was no wind, and the sky was partially cloudy. Zenmuse X4S and the Phantom 3 camera were used to capture the normal spectrum palette on Phantom 3 Professional and Matrice 200 drones and the Forward Looking Infrared Vue Pro R (FLIR) camera was used to capture the other thermal and IR outputs. All videos were recorded at 30 frames per second and 640 x 512 resolution. The FLAME dataset's entire library of images, videos, and data is accessible on IEEE-Dataport [22]. By sampling a single frame per second from the thermal video of Fusion palettes we get 1436 images from which we select 862 images for the training set and 574 for the testing set. For salient object recognition, binary masks of the salient objects are also required. To develop wildfire saliency masks, we annotate these thermal images using VGG Image Annotator [24] that can be used to train the BAS-Net model. The new FLAME Saliency Detection Dataset's sample images and annotations are displayed in Figure 4.

The Yolov7 Wildfire detector is evaluated using mean Average Precision (mAP) of detections at IOU = 0.5 with the

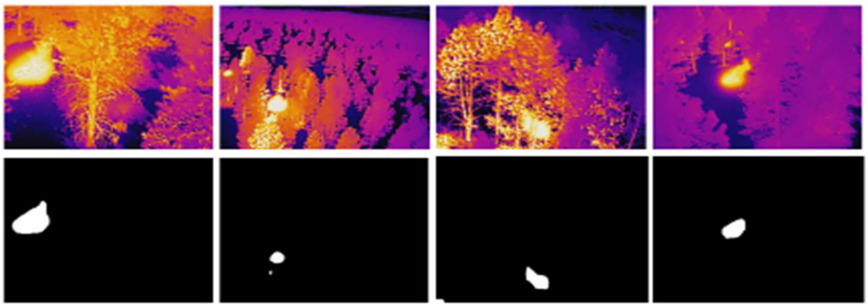

Fig. 4: Sample annotations from our FLAME thermal Dataset

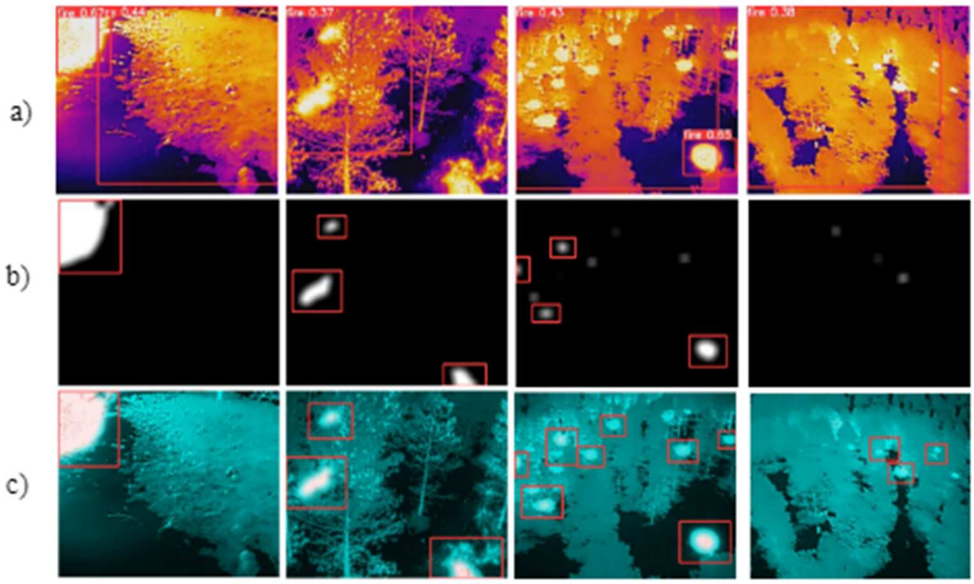

Fig. 5: Sample results from forest fire detection on images: (a) Thermal images, (b) BAS-NET Saliency, (c) BAS-Net Saliency + Thermal

ground truth box. To assess the performance of the wildfire saliency detector, we use weighted harmonic mean of precision and recall which is the F-measure (F_β score), and the average absolute difference between predicted saliency map's pixel values and corresponding ground truth saliency map's pixel values which is the Mean absolute error (MAE score). The higher the score of F_β and MAE the better the model. The specific formula for F_β and MAE is

$$F_\beta = \frac{(1+\beta^2) Precision \times Recall}{\beta^2 Precision + Recall} \quad (1)$$

where, β^2 is 0.3

$$\text{MAE} = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} |S(x,y) - \bar{S}(x,y)| \quad (2)$$

where width and height of the images are represented as W and H respectively.

$S(x,y)$ are $\bar{S}(x,y)$ the pixel values of saliency maps and their ground truth images.

V. IMPLEMENTATION AND RESULT

A. *Benchmarkark for Wildfire detection using Yolov7*

The YOLOv7 model is trained on normal thermal images with a 0.001 learning rate and a batch size of 8. The training is done using the NVIDIA Tesla P100 GPU with 16GB video memory. To optimize the model, we use an SDS of 0.001 initial learning rate, 0.0005 weight decay, and 0.825 momentum. The output of the non-maximum suppression (NMS) operation with a 0.5 IoU threshold is the final prediction result.

B. Saliency Map Generation

The original architecture of BAS-Net mentioned in the research paper [22] is used to create saliency maps. Initial feature extraction network parameters are set using the ResNet-34 network's weights. The training is done from 0 with a learning rate of 0.01, SGD with a learning rate of 0.001, momentum 0.9, and weight decay 0.0005 on an NVIDIA Tesla P100 GPU with 16GB of video memory. Finally, we resize the output to the original image size using bilinear interpolation

C. Fused Thermal images for wildfire detection

Single channel saliency map is fused by eliminating the red channel in 3-channel thermal image to create the fused thermal image. Figure 3 shows that the fusion image preserves the textural information while enhancing the salient portions of the image. We train the YOLOv7 model on the NVIDIA Tesla P100 GPU. The pre-trained model on the thermal images with a 0.001 learning rate and batch size of 8 is fed with fused thermal images as input for 150 epochs. We use an SDS of 0.001 initial learning rate, weight decay of 0.0005, and 0.825 momentum to optimize the model. The final predictions are retired after the non-maximum suppression (NMS) operation is performed with the 0.5 IoU threshold.

D. Evaluation of Deep Saliency Networks on the FLAME thermal image dataset

We evaluate the performance of the BAS-Net on the test data of our annotated FLAME dataset. Figure 5 (b) shows the Saliency masks generated by these networks. From Table 1, it is inferred that a good saliency detection performance is achieved.

Model	F_{β}	MAE
BAS-Net	0.739	0.0075

TABLE 1. DEEP SALIENCY NETWORKS' PERFORMANCE ON OUR ANNOTATED TEST DATASET

E. Evaluation and Analysis of Wildfire Detections in the Thermal Images using Saliency Maps

Figure 5 illustrates the detection results of the Yolov7 model. The first row has the results of a model trained only on thermal images, the second row has the results of the model trained only on saliency maps, and the third row has the results of the model trained on fused images. It can be inferred that the model trained on normal thermal images (a) suffers from thermal reflections and the model trained on saliency maps (b) also might miss some objects, as shown in the 4th column. These drawbacks are overcome by fusion images which contribute to increased performance. It can be observed in the 3rd and 4th columns that wildfires have been precisely detected. This proves the hypothesis that the saliency map fusion improves the object detection accuracy of thermal images. Table 2 shows the MAP score on different datasets.

Model	On only Thermal images	On only Saliency maps	On Fused images
MAP score	0.932	0.943	0.899

TABLE 2. PERFORMANCE OF YOLOV7 MODEL ON THERMAL, SALIENCY, FUSED IMAGE DATASET

VI. CONCLUSION

In this paper, we conclude that the fused thermal images produced from the deep saliency network outperform the normal thermal images when trained on Yolov7 with the FLAME dataset since the core part is been highlighted. This reduces the computational burden, unlike in coupled visible thermal image systems. The proposed system being more accurate during thermal reflections it can be implemented in a wide range. Hence extended it to UAV monitoring of early wildfire detection. However, it would be interesting to see an onboard early wildfire detection with a less weight deep saliency object detection module by including saliency maps in the networks via a saliency proposal stage like SDS R-CNN [25]. And also the idea of Autonomous Multi Drone Wildfire Monitoring systems would also increase the performance efficiency of Wildfire Monitoring systems.

REFERENCES

[1] Jessie Yeung, “Australia’s deadly wildfires are showing no signs of stopping. Here’s what you need to know.”- CNN, 2020

[2] The Visual and Data Journalism Team: California and Oregon 2020 wildfires in maps, graphics and images - BBC News.

[3] AJ Willingham, “6 important things to know about wildfires” - CNN, 2018.

[4] Huang, Q., Razi, A., Afghah, F., & Fule, P. (2020, August). Wildfire spread modeling with aerial image processing. In *2020 IEEE 21st International Symposium on" A World of Wireless, Mobile and Multimedia Networks"(WoWMoM)* (pp. 335-340). IEEE.

[5] Alkhatib, A. A. (2014). A review on forest fire detection techniques. *International Journal of Distributed Sensor Networks*, *10*(3), 597368.

[6] Jiao, Z., Zhang, Y., Xin, J., Mu, L., Yi, Y., Liu, H., & Liu, D. (2019, July). A deep learning based forest fire detection approach using UAV and YOLOv3. In *2019 1st International conference on industrial artificial intelligence (IAI)* (pp. 1-5). IEEE.

[7] Goyal, S., Shagill, M. D., Kaur, A., Vohra, H., & Singh, A. (2020). A yolo based technique for early forest fire detection. *Int. J. Innov. Technol. Explor. Eng.(IJITEE) Vol*, *9*, 1357-1362.

[8] Novac, I., Geipel, K. R., de Domingo Gil, J. E., de Paula, L. G., Hyttel, K., & Chrysostomou, D. (2020, January). A Framework for Wildfire Inspection Using Deep Convolutional Neural Networks. In *2020 IEEE/SICE International Symposium on System Integration (SII)* (pp. 867-872). IEEE.

[9] Alexandrov, D., Pertseva, E., Berman, I., Pantiukhin, I., & Kapitonov, A. (2019, April). Analysis of machine learning methods for wildfire security monitoring with an unmanned aerial vehicles. In *2019 24th conference of open innovations association (FRUCT)* (pp. 3-9). IEEE.

[10] Srinivas, K., & Dua, M. (2019, August). Fog computing and deep CNN based efficient approach to early forest fire detection with unmanned aerial vehicles. In *International Conference on Inventive Computation Technologies* (pp. 646-652). Springer, Cham.

[11] Zhang, Q., Xu, J., Xu, L., & Guo, H. (2016, January). Deep convolutional neural networks for forest fire detection. In *2016 International Forum on Management, Education and Information Technology Application* (pp. 568-575). Atlantis Press.

[12] Barmpoutis, P., Stathaki, T., Dimitropoulos, K., & Grammalidis, N. (2020). Early fire detection based on aerial 360-degree sensors, deep convolution neural networks and exploitation of fire dynamic textures. *Remote Sensing*, *12*(19), 3177.

[13] Alexandrov, D., Pertseva, E., Berman, I., Pantiukhin, I., & Kapitonov, A. (2019, April). Analysis of machine learning methods for wildfire security monitoring with an unmanned aerial vehicles. In *2019 24th conference of open innovations association (FRUCT)* (pp. 3-9). IEEE.

[14] Yadav, R. (2020). Deep Learning Based Fire Recognition for Wildfire Drone Automation. *Can. Sci. Fair J.*, *3*(2), 1-8.

[15] Jiao, Z., Zhang, Y., Mu, L., Xin, J., Jiao, S., Liu, H., & Liu, D. (2020, August). A yolov3-based learning strategy for real-time uav-based forest fire detection. In *2020 Chinese Control And Decision Conference (CCDC)* (pp. 4963-4967). IEEE.

[16] Zhao, Y., Ma, J., Li, X., & Zhang, J. (2018). Saliency detection and deep learning-based wildfire identification in UAV imagery. *Sensors*, *18*(3), 712.

[17] Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696*.

[18] Ding, X., Zhang, X., Ma, N., Han, J., Ding, G., & Sun, J. (2021). Repvgg: Making vgg-style convnets great again. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 13733-13742).

[19] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[20] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).

[21] Lee, C. Y., Xie, S., Gallagher, P., Zhang, Z., & Tu, Z. (2015, February). Deeply-supervised nets. In *Artificial intelligence and statistics* (pp. 562-570). PMLR.

[22] Qin, X., Zhang, Z., Huang, C., Gao, C., Dehghan, M., & Jagersand, M. (2019). Basnet: Boundary-aware salient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 7479-7489).

[23] Shamsoshoara, A., Afghah, F., Razi, A., Zheng, L., Fulé, P. Z., & Blasch, E. (2021). Aerial imagery pile burn detection using deep learning: The FLAME dataset. *Computer Networks*, *193*, 108001.

[24] A. Dutta, A. Gupta, and A. Zissermann. VGG image annotator (VIA). http://www.robots.ox.ac.uk/ vgg/software/via/, 2016. Version: 1.0.6, Accessed:03-01-2019.

[25] Brazil, G., Yin, X., & Liu, X. (2017). Illuminating pedestrians via simultaneous detection & segmentation. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 4950-4959).