A YOLOv3-based Learning Strategy for Real-time UAV-based Forest Fire Detection

Zhentian Jiao[1], Youmin Zhang[2], Lingxia Mu[1], Jing Xin[1], Shangbin Jiao[1], Han Liu[1], Ding Liu[1]

1. Department of Automation and Information Engineering, Xian University of Technology, Xian 710048, China

2. Department of Mechanical, Industrial & Aerospace Engineering, Concordia University, Montreal, QC H3G 1M8, Canada

Abstract: Forest resources safety is of paramount importance for natural and public security. Forest fire detection methods have been attracted much attention recently, but the performance in terms of comprehensiveness, rapidity, and accuracy is still not satisfactory. A deep learning fire detection algorithm is proposed in this paper, aiming at improving the detection accuracy and efficiency by using the unmanned aerial vehicle (UAV). A large-scale YOLOv3 network is firstly developed which can ensure the detection accuracy. The algorithm is then applied to UAV forest fire detection (UAV-FFD) platform, where the fire images can be captured by the UAV and transmitted to the ground-station in real time. The testing results indicate that the recognition rate of the detection algorithm is about 91%, and the frame rate can reach up to 30 FPS (Frames Per Second). It shows strong potential in real-time application for precision forest fire detection.

Key Words: UAV-based forest fire detection (UAV-FFD), large-scale YOLOv3 algorithm, high-precision fire detection

1 Introduction

Wild forest fire is one of the most harmful natural hazards, which would cause massive environment destruction and public security threats. In recent years, because of the climate change such as global warming, a higher probability happened to the heat waves and droughts, resulting in disastrous forest fires. For example, the forest fire happened on March 30th, 2019, in Liangshan, China, led to such heavy losses in the life of fire fighters and the forest resources. If the forest fire could be detected in its early stage, the losses might be restrained.

Early fire detection algorithms based on computer vision have attracted much attention in the past a few years. Traditional detection techniques are mainly based on various color spaces [1] and spatial spaces [2]. Some methods are tending to combine the color and dynamic characteristics of the flame, aiming at a more reliable recognition. Although many researches on fire detection have been carried out, only a few studies take forest environment into consideration.

On the other hand, over the last decade, unmanned aerial vehicle (UAV) becomes a quite efficient platform with application to many tasks due to its flexibility and low-cost [3]. However, few related experiments have used UAVs for forest fire monitoring and detection [4,5]. In this paper, the UAV is used to capture the images of forest environment, by which a wider search area can be covered.

With the evolution of computer technology, especially the graphic processing units (GPUs), deep learning method is rapidly developed. Recently, thanks to the remarkable power on classification and object recognition, deep learning has also been used for forest fire detection [6–10]. Sharma et al. [6] developed a fire detection scheme by combining two pre-trained deep convolutional neural networks (CNNs), VGG16 and Resnet50. Zhang et al. [7] proposed a model consisting of a full image CNN and a local patch NN classifier for forest fire detection, and then used faster region-based CNN to detect wild forest fire smoke [8]. Shen et al. [9] developed flame detection algorithm based on YOLO (You Only Look Once) algorithm. Barmpoutis et al. [10] presented an approach for fire detection from images combining the merits of deep learning and spatial texture analysis.

This paper proposed a deep-learning-based forest fire detection method, aiming at improving the detection precision and efficiency by integrating the flexible UAV platform. A fire detection algorithm using large-scale YOLOv3 network is firstly developed. Then, the algorithm is applied to the UAV-based forest fire detection platform. The images acquired by the UAV is transmitted back to ground station, where the images can be processed by the proposed YOLOv3 algorithm with high precision in real time using the high-performance computer equipped on the ground station.

2 YOLOv3-based Forest Fire Detection Algorithm

YOLOv3 algorithm is well known due to its advantages in object detection application benefitted by its internal network structure. In this section, by analyzing the network principle, the recognition process of YOLOv3 algorithm is presented.

This work is supported in part by National Natural Science Foundation of China under Grants (61833013 and 61573282) and Natural Sciences and Engineering Research Council of Canada (Corresponding Author: youmin.zhang@concordia.ca).

978-1-7281-5855-6/20/$31.00 ©2020 IEEE

2.1 Architecture of YOLOv3

The YOLOv3 model is usually divided into a feature extraction layer and a processing output layer. The feature extraction layer is combined by Darknet-19 and ResNet-like network, while processing output layer draw on the experience of the feature pyramid networks (FPN) [11]. As illustrated in Fig. 1, convolutional unit is the basic component of the Darknet-53 feature extraction layer, which is the conjunction of convolutional layer, batch normalization (BN) layer and leaky ReLU layer. It can effectively prevent over-fitting without dropout layer [12] by using the BN layer [13]. The leaky ReLU layer retains the negative axis information of the training part.

Inspired by residual structure of ResNet, another component in Darknet-53 is residual unit. The use of this configuration makes its network have a deeper structure than Darknet-19 of YOLOv2. As shown in Fig. 2, the basic components of residual unit are convolutional units. In the feature extraction layer of YOLOv3, multi-dimensional residual unit combined with convolutional units are used.

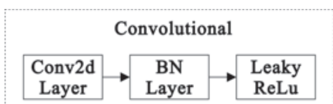

Figure 1: The structure of convolutional unit.

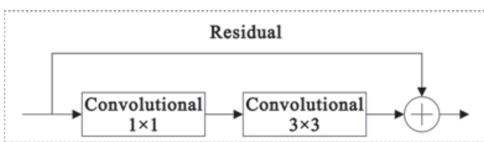

Figure 2: N-dimensional structure of residual unit.

When the input image is given, they would be resized and their features are extracted by the feature extraction layers. There is no pooling layer and full connection layer in the entire Darknet-53 structure. In the process of forward propagation, the size transformation of the tensor is achieved by changing the step size of the convolution filter. The base net of YOLOv3 will pass through 5 convolution layers whose $stride = (2, 2)$, so the size of feature map will be reduced to $\frac{1}{2^5}$ of the original input. When the input is 416×416, the size of the first time of output is $13 \times 13 \times 24$ when the feature map is transfered to the 82nd layer. The second and third feature extractions are performed on the 94th and 106th layers, respectively. In the last layer of Fig. 3, the feature extraction, convolution, upsampling, and feature fusion operations on the block yolo1 are also performed on the blocks yolo2 and yolo3, while the outputs of $26 \times 26 \times 24$ and $52 \times 52 \times 24$ are obtained in the later two blocks, respectively. In this algorithm, 24 represents the size of the network tensor, which can be calculated by Eq. (1). Its network structure is shown as Fig. 3. One can clearly see where the three YOLO layers come from.

$$tensor = N \times N \times [(bounding\ box) \times (offset + object + class)] \tag{1}$$

where $N = 1$. Three bounding boxes are used in each YOLO layer, and each box needs to have five basic parameters $(x, y, w, h, confidence)$ [14], so $offset = 4$, $object = 1$. In order to improve the detection accuracy, we define the forest fire scenes as three scenarios of *smoke*, *fire*, *smoke plus fire* cases, so $class = 3$.

2.2 Identification Method of YOLOv3

YOLOv3 has unparalleled advantages in the speed and accuracy of object detection, mainly due to its superior algorithm. Traditional algorithms generally use artificial selection boxes, which will lead to lower accuracy. The anchor boxes of YOLOv3 is inspired by the anchor mechanism in the faster R-CNN region proposal network (RPN). However, YOLOv3 discards the approach of manual setting and uses the dimensional clustering method used in YOLOv2 instead of determining the anchor boxes, which is obtained by K-means clustering [15]. In our network, the $k = 9$ for prior boxes. This method uses the IOU (intersection over union) score as the final evaluation standard, and more suitable bounding boxes can be found automatically [16]. Through the clustering method, nine anchor boxes are selected based on the average IOU to predict the bounding box. The distance function used by clustering is shown in Eq. (2).

$$d\,(box, centroid) = 1 - IOU\,(box, centroid) \tag{2}$$

The idea of the algorithm for calculating the target bounding box is as follows. When an image is inputted, the target is first selected in the network to determine the center point. Then the input image will be divided into multiple $s \times s$ cells. After the coordinates of the point, which is the center of the cell, are calculated, the predicted bounding box will be produced by the coordinates of the center point. The coordinate calculation is as shown in Eq. (3).

$$\begin{cases} b_x = \sigma\,(t_x) + c_x \\ b_y = \sigma\,(t_y) + c_y \\ b_w = P_w e^{tw} b_h = P_h e^{th} \end{cases} \tag{3}$$

where the coordinates of the center point are (t_x, t_y, t_w, t_h), which represents the coordinates, width and height of the center point of the bounding box. (p_w, p_h) represents the width and height of the cell. (c_x, c_y) indicates the offset coordinate.

$$P_r\,(object) \times IOU\,(b, object) = \sigma\,(t_o) \tag{4}$$

YOLOv3 uses logistic regression when predicting bounding-box. Each time YOLOv3 predicts the bounding-box, it will output $(t_x, t_y, t_w, t_h, t_o)$ and then calculate the absolute (x, y, w, h, c) by Eqs. (3) and (4). Logistic regression is used to make an objectness score on the part enclosed by the anchor.

3 The UAV-based Forest Fire Detection Platform Design

As illustrated in Fig. 4, a UAV-based forest fire detection (UAV-FFD) platform is developed, where the UAV is equipped with a SLR camera for image acquisition. The data transmission system is responsible for the transmission of image data and flight commands. The ground station detects and diagnoses the forest fire after receiving the

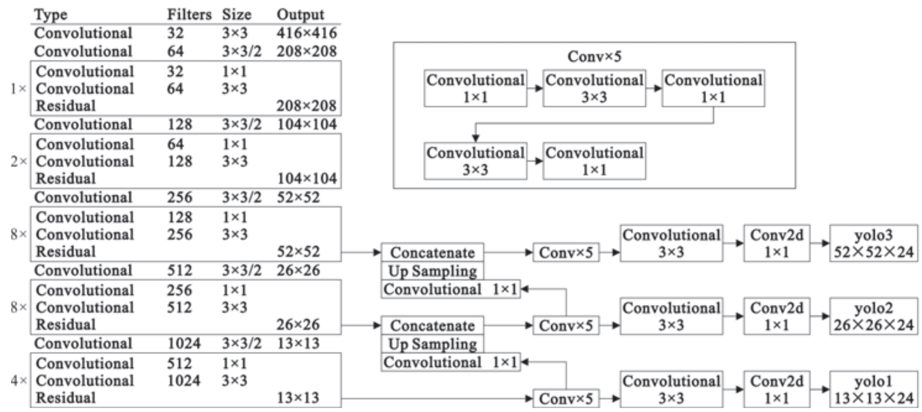

Figure 3: The structure of YOLOv3.

image and location information transmitted by the UAV in real time, and the ground terminal can send commands to the UAV. The platform combines mobility of the UAV with powerful computation ability of the ground station to ensure the detection speed and accuracy.

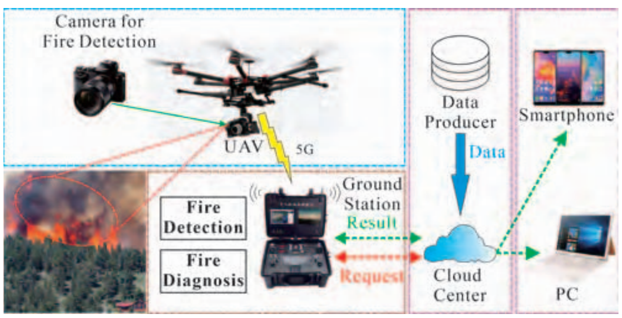

Figure 4: Components and functions of a UAV-based forest fire detection platform.

A six-rotor UAV is used in this platform. The high-definition digital image transmission is adopted, which is adaptable to professional-grade live broadcast equipment. Besides, the delay can be reduced to 50ms, which is reliable for the ground station to process image data in real time. Foreseeably, combined with 5G technology, the HD video transmission and low latency will be achieved in the near future. The ground station is equipped with a video capture card and a high-performance GPU, allowing for image-based fire detection and diagnosis. The computing hardware configuration of the UAV and ground station is shown in Table 1. However, due to small size, light weight, and low power consumption, the onboard computer in a UAV generally cannot afford with complicated detection algorithms for real-time execution. Hence, in the proposed platform, the images are transmitted back to the ground station for processing, which can, on one hand, reduces the onboard computational burden of the UAV, and on the other hand, makes it possible to realize a more complex large-scale deep learning neural network. By this means, the proposed UAV-FFD platform could assign the task reasonably and provide with a real-time, fast, and accurate scheme for forest fire detection.

Table 1: Configuration of the UAV-FFD platform

Platform	Parameter	Description
UAV	Flight controllerr	DJI A3
	Image transmission	DJI LightBridge2
	Camera	Sony A7-R
Gound Station	CPU	i7-9700K
	RAM	16GB
	Graphics card	RTX2080-8G
	System environment	Ubuntu 18.04LTS

In the forest fire detection system, it is important to use a data center to aggregate the information detected by all units and perform big data processing. A large amount of image data could be generated throughout the inspection process, which will undoubtedly become a challenge. Since the transmission of large amounts of data in the network may cause congestion, the cloud computing model is hardly suitable for that numerous video processing.

In this regard, edge computing is proposed in the whole system, as shown in Fig. 4. The UAV can be regarded as a user node, and the ground station and the servers are considered as edge nodes. In edge computing, cloud center decentralizes related requests, each edge node processes the request in conjunction with local video data, and then returns only the relevant results to the cloud center, which can reduce network traffic. Using the edge computing architecture, the response time of the system to image recognition can be improved, and the data can be reduced by 20 times in terms of integration and migration. In addition, after some of the computing tasks are off-loaded from the cloud to the edge, the energy consumption of the entire system is reduced by 30% to 40%.

After the cloud system collects the detection information of the forest edge computing, the data is analyzed by big data center, and the current weather, temperature, humidity and historical data are combined to predict the probability of current forest fire occurrence. If the cloud computing center receives a forest fire detection alarm or calculates a high forest fire occurrence probability index, it will send an alarm message to the client (PC or smartphone). Therefore, the system uses UAVs, ground station, and edge computing network to form a forest fire detection system, and can independently judge the forest fire and return it to the user with the detection result. The system will greatly save labor and provide a reliable guarantee for the prevention of forest fires.

4 Experiments

The YOLOv3 model is firstly trained and tested on the desktop. It has been trained for 57,000 steps, and 64 images are used in each step. The network performance results are shown in Table 2. The YOLOv3 network has very high FPS performance on the proposed UAV-FFD platform, which can provide reliable real-time detection. The results are shown in Table 3. Because of the software acceleration, the FPS of processing is much faster than the frame rate.

Table 2: The network performance results

Parameter	Value
mAP@0.5	78.92%
mAP@0.75	66.11%
Average_IOU	73.78%
Precision	0.84
Recall	0.78
F1-score	0.81

Table 3: The network performance results

Video name	Resolution	Frame rate	FPS of processing
Test1.mp4	400×240	29.97	82.4
Test2.avi5	1920×1080	25.00	82
Test3.mp4	1920×1080	30.10	82
Test4.mp4	1140×544	26.47	80
USB-camera	640×480	30.0	30

Using 60 images to test the YOLOv3 model, the detection rate can reach 91%. Some of the test results are shown in Fig. 5. It can be seen from Fig. 5 that the algorithm has an ideal performance. Because the algorithm of YOLOv3 uses dimensional clustering to predict coordinates and multi-scale prediction, it can detect small targets quickly and accurately. The developed algorithm is not only sensitive to the large area fire and smoke in the image, but also able to detect small objects.

Figure 5: Results using the proposed detection algorithm under UAV-FFD platform.

5 Conclusions

This paper proposed a forest fire detection algorithm using deep learning network with application in a developed UAV-FFD platform through this work. A large-scale YOLOv3 network is developed to process the images received from the UAV in real time. By using the high-performance computer on the ground-station of the UAV-FFD platform, the YOLOv3 detection algorithm shows high precision with low time cost. This shows strong potential in real-time application for precise forest fire detection even with small sparks. However, the fire can be successfully detected or not partly depends on the transmission ability between the UAV and ground-station. Hence, in our future work, further enhancement on the detection reliability of the forest fire will be one of the tasks.

REFERENCES

[1] K. Dimitropoulos, F. Tsalakanidou, and N. Grammalidis, “Flame detection for video-based early fire warning systems and 3d visualization of fire propagation,” in *13th IASTED International Conference on Computer Graphics and Imaging (CGIM 2012), Crete, Greece*, 2012.

[2] N. Grammalidis, E. Çetin, K. Dimitropoulos, F. Tsalakanidou, K. Kose, O. Gunay, B. Gouverneur, D. Torri, E. Kuruoglu, S. Tozzi *et al.*, “A multi-sensor network for the protection of cultural heritage,” in *2011 19th IEEE European Signal Processing Conference*, 2011, pp. 889–893.

[3] C. Yuan, Y. M. Zhang, and Z. X. Liu, “A survey on technologies for automatic forest fire monitoring, detection, and fighting using unmanned aerial vehicles and remote sensing techniques,” *Canadian Journal of Forest Research*, vol. 45, no. 7, pp. 783–792, 2015.

[4] C. Yuan, Z. X. Liu, and Y. M. Zhang, “Aerial images-based forest fire detection for firefighting using optical remote sensing techniques and unmanned aerial vehicles,” *Journal*

of Intelligent & Robotic Systems, vol. 88, no. 2-4, pp. 635–654, 2017.

[5] C. Yuan, Z. X. Liu, and Y. M. Zhang, "Learning-based smoke detection for unmanned aerial vehicles applied to forest fire surveillance," *Journal of Intelligent & Robotic Systems*, vol. 93, no. 1-2, pp. 337–349, 2019.

[6] J. Sharma, O.-C. Granmo, M. Goodwin, and J. T. Fidje, "Deep convolutional neural networks for fire detection in images," in *International Conference on Engineering Applications of Neural Networks*. Springer, 2017, pp. 183–193.

[7] Q. J. Zhang, J. L. Xu, L. Xu, and H. F. Guo, "Deep convolutional neural networks for forest fire detection," in *2016 International Forum on Management, Education and Information Technology Application*. Atlantis Press, 2016.

[8] 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.

[9] D. Q. Shen, X. Chen, M. Nguyen, and W. Q. Yan, "Flame detection using deep learning," in *2018 4th International Conference on Control, Automation and Robotics (ICCAR)*. IEEE, 2018, pp. 416–420.

[10] P. Barmpoutis, K. Dimitropoulos, K. Kaza, and G. Nikos, ""fire detection from images using faster r-cnn and multidimensional texture analysis"," *Proceedings - ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing*, 2019.

[11] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2117–2125.

[12] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[13] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[14] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[15] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 7263–7271.

[16] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 779–788.