Research on Improved YOLOv3 Fire Detection Based on Enlarged Feature Map Resolution and Cluster Analysis

Conghan Yue, Jun Ye*

School of Computer Science and Cyberspace Security, Hainan University, China

*yejun@hainanu.edu.cn

Abstract: Fire detection can effectively prevent the occurrence of fire. For the current fire detection methods, traditional image processing techniques such as grayscale image processing and feature extraction processing have poor anti-interference ability, weak generalization ability, and the detection results are more sensitive to data fluctuations. At present, based on the concept of deep learning, the proposed convolutional neural network processing of extracted image features has been widely used. On this basis, an improved YOLOv3 fire detection algorithm is proposed in this paper: The K-Means++ algorithm is used for clustering analysis to obtain the corresponding anchor boxes dimension, which reduces the error detection rate caused by the bounding boxes not matching the label; secondly, the resolution of the feature image is improved and the receptive field is enlarged; the image data are sharpened and the contrast is enhanced to make the data features more prominent. The experimental results show that the detection precision of the method is 97.7%, the recall rate is 98.5%, and the fps is 19, effectively solving the problem of high error detection rate and high missing rate of traditional image processing and general CNN networks for suspected flame objects.

Keywords: fire detection; deep learning; YOLOv3 algorithm; data enhancement; cluster analysis

1. Introduction

The occurrence of fires will bring great losses to people's production and life. The use of images to combat fires has become a research hotspot.

Traditional fire detection is generally a method of color processing and artificial design of complex feature extraction, such as, Turgay Çelik et al(2009), proposed a rule-based general color model for flame pixel classification[1]; Bakri et al(2018), used a color pixel classification model to separate the flame pixels from the background and separate the brightness from the chroma of the original image to detect fire[2]; Mahdi Hashemzadeh, Alireza Zademehti(2019) proposed a method of using k-medoids' robust ICA color model to select candidate regions, analyzing fire features based on motion, and distinguishing fire regions with a support vector machine classifier[3]; Chou et al(2017), used block-based analysis of local features including flame color and flame immobility, and then further identified flames through LBP features[4]. Sam G Benjamin et al(2016), used the HSV-YCbCr color space and the five texture features of the Gray Level Co-occurrence Matrix to recognize the flame, and obtained a higher recognition rate[5]. Ouyang et al(2018), used the RGB color model to separate the flame image, and used morphological opening and closing operations and edge extraction operations to...
extract appropriate image edges to identify flames[6]. Amin Khatami et al(2017) used PSO and sample pixels in the image to obtain the weights of the color discrimination transformation matrix, and the resulting transformation matrix can be used for fire detection of different fire images[7]. These methods have certain limitations in recognition rate and robustness.

In recent years, deep learning has received more and more attention from researchers in fire detection all over the world. Zhaochun Liu(2020) proposed a multi-level forest fire detection method based on deep learning[8]; S. Wu and L. Zhang(2018) used classic methods such as R-CNN, YOLO (YOLOv2, YOLOv3) and SSD, and adjusted the tiny-yolo-voc structure of YOLO to improve the accuracy of fire detection[9]; Qi-xing Zhang et al(2017), used a faster R-CNN to detect wild forest fire smoke[10]; Pu Li, Wangda Zhao(2020) discovered that the accuracy of fire detection algorithms based on the object detection network is generally higher than other algorithms, especially YOLOv3[11]; Redmon et al(2016). proposed a new end-to-end object detection and recognition algorithm, named YOLO[12]. It has a low false detection rate and missing detection rate, fast detection speed, and can learn the generalized characteristics of the object. In order to improve the accuracy and recall rate of object positioning, Redmon et al(2017), proposed the YOLOv2 detection algorithm[13]. Compared with the YOLOv1 detection algorithm, YOLOv2 has greatly improved the recognition accuracy, speed and positioning accuracy. And Redmon et al(2018), improved on YOLOv2 and proposed the YOLOv3 algorithm[14]. Compared with the traditional target detection methods such as CNN and R-CNN, the YOLO algorithm is more efficient in terms of detection frequency and precision, which is more appropriate for practical needs. Therefore, this paper is based on YOLOv3 to investigate the fire detection.

2. Models and methods

In this paper, an improved YOLOv3 algorithm is proposed, which uses the K-Means++ cluster analysis algorithm to obtain the corresponding anchor box size and reduce the error detection rate caused by the difference of bounding boxes and labels; secondly, the resolution and contrast of the training sample feature images are modified, data enhancement is performed to expand the acceptance field and enhance the features, so that the training effect is more significant and the network structure is more stable. Finally, the feasibility of the algorithm is verified experimentally.

Fig.1 shows the network structure of YOLOv3, which is mainly composed of feature extraction DarkNet53 and YOLO convolutional layer. The feature extraction layer is mainly composed of a convolutional layer, a BN layer and a jump layer connection. The activation function uses LeakyReLU, DarkNet-53 is a new feature extraction network proposed by the YOLOv3 network.

![YOLOv3 network structure diagram](image-url)
The clustering algorithm is used in YOLOv3, and 9 prediction frames are obtained by clustering, and the obtained 9 prediction frames are used in the detection of 3 different scale features to predict the actual frame of the object. In this paper, K-Means++ clustering is used to initialize anchor boxes, which to improve the selection of initial points on the basis of the means clustering algorithm. Then, uses the K-Means clustering algorithm to cluster the $w$ and $h$ two parameters of the dataset. The current iteration Intersection Union (IOU) parameter generated during training is used to replace the Euclidean distance in the K-Means++ algorithm. The calculation formula is:

$$d(\text{box}, c) = 1 - IOU(\text{box}, c) \quad \text{/* MERGEFORMAT (1)}$$

In the formula: $c$ is the cluster center; box is the true box; IOU is the intersection ratio of the real box and the predicted box.

$$IOU = \frac{\text{box} \text{(Truth)} \cap \text{box} \text{(Prediction)}}{\text{box} \text{(Truth)} \cup \text{box} \text{(Prediction)}} \quad \text{/* MERGEFORMAT (2)}$$

YOLOv3 network structure adopts the feature fusion method of the three scales has an adverse effect on the detection of smaller objects in the fire image, and $13 \times 13$ is easy to cause the loss of small objects. Considering that the resolution of the feature will directly affect the detection of small objects and the overall performance indicators, based on Darknet-53, the three-scale resolution of the original feature map is $13 \times 13$, $26 \times 26$, and $52 \times 52$, modified as the larger-scale resolution is $26 \times 26$ and $52 \times 52$.

### 3. Experiments and Results

#### 3.1. Processing of dataset

The dataset in this article mainly comes from the video intercepted pictures in the fire database of Bilkent University and the flame images on the Internet. Extract fire pictures from the obtained video, and perform data enhancement on these pictures such as adjusting the resolution and sharpening. A total of 2000 pictures and 3471 objects to be detected are obtained.

This paper uses LabelImg to mark the flame objects in the flame image, and randomly divide them into training set, validation set and test set at a ratio of 0.75/0.1/0.15.

#### 3.2. Preparation of bounding box

YOLOv3 uses anchor boxes to predict the bounding box coordinates, and the K-Means++ algorithm to predict 9 anchor boxes, and divides them into 2 larger-scale feature maps to obtain more object edge information. The 9 sets of anchor boxes dimensions calculated in YOLOv3 are respectively $(25,53)$, $(29,34)$, $(33,63)$, $(47,99)$, $(52,45)$, $(67,67)$, $(79,136)$, $(92,99)$, $(119,111)$, therefore, the selected anchor box is shown in Tab. 1.

| Feature map | Anchor box |
|-------------|------------|
| 26×26       | (47,99), (52,45), (67,67) |
| 52×52       | (25,53), (29,34), (33,63) |

**Tab.1** The anchor box of the dataset in this paper.

#### 3.3. Experimental environment and hyper-parameter setting

This paper uses the computer as a laptop, configured as Intel i7-5557U, 16GB RAM, 64-bit Windows 10 operating system; development environment is tensorflow1.14, opencv4.4, python3.5, running in keras environment, with 1500 fire images for the training set, 200 sheets for the validation set and 300 sheets for the test set. In the initial stage, the learning rate is 0.001 and the attenuation coefficient is 0.0005. In order to further converge the loss function, when the number of training iterations is 25000, the learning rate is reduced to 0.0001.

#### 3.4. Result
This paper conducted a set of comparative experiments, using Faster-RCNN, YOLOv3 (unimproved), YOLOv3 (improved) three algorithms for model training and result testing. The experimental results are shown in Tab. 2.

| Network               | mAP  | recall | fps  |
|-----------------------|------|--------|------|
| Faster-RCNN           | 0.932| 0.958  | 2    |
| YOLOv3(unimproved)    | 0.953| 0.963  | 20   |
| YOLOv3 (improved)     | 0.977| 0.985  | 19   |

Tab.2 The results of mAP, recall and fps of three Network

It can be seen from the experimental results that in terms of mAP, the Faster-RCNN algorithm is 0.932, the traditional YOLOv3 is 0.953, and the improved YOLOv3 is 0.977. That is to say, the improved algorithm in this paper is 0.045 and 0.024 higher than the Faster-RCNN algorithm and the YOLOv3 algorithm respectively; The recall rate of the three algorithms is also positively correlated with mAP. The recall rate of the improved YOLOv3 algorithm reached 0.985; from the fps point of view, the YOLO algorithm is much faster than Faster-RCNN, but the improved YOLOv3 algorithm is indeed slightly lower than the unimproved one. But at the application level, such a certain delay is acceptable. Part of the test results of this improved YOLOv3 model are shown in Fig. 2.

4. Conclusion
In this paper, we propose ideas to improve YOLOv3: increasing the resolution scale of the feature map, which has a better ability to describe detailed features, reduce the rate of error and missing detection in the fire detection process; the K-Means++ clustering algorithm can select more appropriate anchor boxes, and the data enhancement operations of sharpening and increasing saturation are beneficial for network training. However, precisely because of the increase in the number of feature channels, the amount of computation increases as well. Compared with the original YOLOv3 method, the improved YOLOv3 method slightly increases the processing time for a single image. Considering the actual needs of the fire alarm situation, increasing the response time by a certain amount has little impact on fire detection and early warning, and improves the accuracy of the method. In other words, the method in this paper can still achieve real-time fire detection and early warning and play a guiding role in intelligent fire detection and prevention.

Acknowledgements
This work is partially supported by the Science Project of Hainan University (KYQD(ZR)20021), the National Natural Science Foundation of China under projects (U19B2044).

References
[1] Turgay Ç and Hasan D 2009 Fire detection in video sequences using a generic color model Fire Safety Journal pp 147-158
[2] BAKRI N S and ADNAN R 2018 A methodology for fire detection using colour pixel classification 2018 IEEE 14th CSPA pp 94-98
[3] Mahdi H and Alireza Z 2019 Fire detection for video surveillance applications using ICA K-medoids-based color model and efficient spatio-temporal visual features Expert Systems with Applications pp 60-78
[4] CHOU K 2017 Bolck-based feature extraction model for early fire detection, 2017 IEEE Symp. series on computational intelligence pp 1-8
[5] Benjamin S G, Radhakrishnan B and Nidhin T G 2016 Extraction of fire region from forest fire
images using color rules and texture analysis 2016 Int. Conf. on emerging technological trends pp 1-7
[6] Ouyang J, Bu L and Zhikai Y 2018 An early flame identification method based on edge gradient feature, 2nd IEEE advanced information management, communicates, electronic and automation control Conf. pp 642–646
[7] Amin K, Saeed M, et al 2017 A new PSO-based approach to fire flame detection using K-Medoids clustering Expert Systems with Applications pp 69-80
[8] Zhaochun L, Kai Z, et al 2020 Research on the identification method for the forest fire based on deep learning Optik.
[9] S Wu and L Z 2018 Using Popular Object Detection Methods for Real Time Forest Fire Detection 2018 11th ISCID pp 280-284
[10] Qi-xing Z, Gao-hua L and et al 2018 Wildland Forest Fire Smoke Detection Based on Faster R-CNN using Synthetic Smoke Images Procedia Engineering pp 441-446
[11] Pu L, Wangda Z 2020 Image fire detection algorithms based on convolutional neural networks Case Studies in Thermal Engineering.
[12] Redmon J, Divvala S, Girshick R and et al 2016 You Only Look Once: Unified, Real-Time Object Detection 2016 IEEE CVPR
[13] Redmon J and Farhadi A 2017 YOLO9000: Better, Faster, Stronger 2017 IEEE CVPR Honolulu pp 6517-6525
[14] Redmon A and Joseph 2018 YOLOv3: An Incremental Improvement, arXiv e-prints arXiv:1804.02767