Study of Flame Detection based on Improved YOLOv4

Chengzhi Cao*, Xiaoyu Tan, Xinyi Huang, Yongjun Zhang, Zehao Luo
School of Electric Power Engineering, South China university of technology, 510640, China

*Corresponding author e-mail: 201730430079@scut.edu.cn

Abstract. In some complex circumstances, the detection of conflagration mostly depends on smog detectors, which have lots of limitations in precision, efficiency and safety. If we make full use of object detection algorithms to detect the flame in industries, it will benefit people’s safety obviously. Among all kinds of object detection algorithms, YOLO series play a very significant role. In this paper, we propose an improving strategy on YOLOv4 to enhance its precision based on multi-scale feature maps. Firstly, we create flame datasets including almost 4000 high-resolution flame pictures. Secondly, some improvements on feature extraction network are made to detect smaller objects. Finally, the total algorithm are trained and tested on our datasets for about 400 epochs. The result show that the method can generate high quality on flame detection in a great number of situations.

Keywords: Flame YOLOv4 Multi-scale feature maps.

1. Introduction
Nowadays, with the development of science and technology, people gradually pay more attention to the importance of artificial intelligence for convenience. However, in some industry cases, the flame always threatens workers life and make great loss on economy. In order to solve this problem, a great number of companies have applied smog-detectors for safety protection, but this way depends highly on software and has some drawbacks on precision and efficiency. On the other hand, it’s a challenge to judge the condition of flame based only on smog detectors.

In recent years, the application of computer vision has made great influence on people’s daily life. For instance, in some cases computers has higher appearance than people in picture classification. If we could make full use of computer vision in companies, it will benefit us a lot for saving time and labors.

In order to detect the flame in videos quickly, we attach great importance on object detection algorithms, which have unbelievable precision in almost every kind of vision tasks. More specially, YOLO series, having developed for several years, are applied by quantities of industries because of its higher quality and fewer parameters. But there are also a lot of things to improve it. In the past few years, Kim has extracted color and shape of flame by dividing RGB channels to make it easier in detection; Mueller has found difference between the last snapshot and the next snapshots according to the momentum of flame. What’s more, Frizzi used convolution neural networks to extract features automatically, getting higher scores in most of vision tasks, but faced difficulties in detecting smaller objects.
With the purpose of detecting flame quickly and precisely, this paper proposes a new strategy to improve the appearance of YOLOv4 by using multi-scale feature maps. With a smaller size of feature map, it will be easier to track smaller flame in pictures and alarm workers to escape in time.

2. YOLOv4 Model

As we all know, the beginning of YOLO series goes back to 2016, when Joseph Redmon proposed a new object detection algorithm named “You Only Look Once”. From then on, lots of different network structures based on YOLO have emerged in large numbers.

With the purpose of high efficiency and precision, YOLOv4 applies a few tricks to improve the network structure. The most important progress includes the following two parts.

Cross Stage Partial (CSP) can enhance the learning ability of CNN, while maintaining accuracy, reducing computing bottlenecks, and reducing memory costs while being lightweight.

SPP, that is, spatial pyramid pooling, is to increase the receptive field of the network. The implementation is to perform maximum pooling of $5 \times 5$, $9 \times 9$ and $13 \times 13$ on layer107, and obtain layer 108, layer 110 and layer 112 respectively. After pooling is completed, perform layer 107, layer 108, layer 110 and layer 112 concatenate, connect into a feature map layer 114 and reduce the dimensionality to 512 channels through $1 \times 1$.

3. Multi-scale structure

In many actual scenes, because the distance between the object and the image acquisition device is not the same, the size of the object in the image will also appear a large gap, and some smaller objects are often not easy to be detected [10]. To solve the above problems, this paper adds a 104*104 size feature layer on the basis of the original three feature layers to improve the recognition rate of small objects. In the Darknet-53 network architecture, the three feature layers have undergone different numbers of residual convolutions, so that the original image is divided into specified sizes. By analogy, it is also very convenient to draw a 104*104 size feature map from the network. After the fourth feature layer is obtained, due to the different sizes, it is necessary to adjust the size of each feature layer through an up-sampling operation, and then after splicing, can each feature layer want to be merged. The revised network architecture is shown in the figure below:
Figure 2. The proposed improvement in YOLOv4

3.1. Dataset

The flame image samples were obtained from various browsers through the python crawler tool, totaling 4000 sheets. Including candle lights, forest fires, etc., the size and brightness of the flames are diverse, so the data base basically covers all types of flames.

Because the data is very limited, this article uses the cross-validation method for training and testing, which is to group the original data, one part is used as a training set to train the model, and the other part is used as a test set to evaluate the model. In this experiment, the data set was randomly divided into 10 groups, and one group was selected as the test set each time, and the remaining 9 groups were used as the training set to test the model. Repeat this 10 times to ensure that each group has a chance to become the test set. The 10 sets of results obtained are averaged to be used in the estimation of model parameters as the performance index of the model. This approach can make full use of sample data and is suitable for situations with fewer samples. The sample uses the VOC data set, and annotates the data set through the python function library labeling, which is used to make the training set.

Figure 3. Illustration of our flame dataset
3.2. Pre-processing
The histogram is a graphical representation of the intensity distribution of pixels in an image, which characterizes the quantity distribution of the image's pigments. Histogram equalization is a commonly used image processing method. Its main function is to increase the local contrast without losing the overall contrast, which can highlight the characteristics of some details of the image. The image is preprocessed by the way of histogram equalization. Results as shown below.

![Figure 4. Comparison of original image(left) and processed image after histogram equalization(right).](image)

Using the improved YOLOv4 network, each time 2700 pictures are input as the training data set, and after the training is completed, the performance test is performed through the 400 picture test set, and the number of iterations is 10 times. Compare with the test results of the traditional YOLOv4 network, observe the difference of each performance index, and analyze the strategy of improving the model. In addition, this experiment uses video as the final carrier, and uses the trained model to identify the flame in the video and mark the position of the flame, and record the model parameters with better performance to improve the fluency and real-time performance of target detection as much as possible.

3.3. Analysis
This experiment uses average precision AP (average precision) as the evaluation index. First, the accuracy and recall of the results must be calculated. The accuracy or recall rate alone cannot fully measure the performance of a model. mAP combines the two to make the evaluation model more objective. The accuracy of the model represents the proportion of real positive samples in the prediction result, that is, the ratio of the actual positive samples in the result to all the positive samples in the result; the recall rate represents the proportion of the positive samples in the result to all the actual positive samples, that is, the true prediction. The ratio of the number of positive samples to the number of true positive samples in the sample.

The running environment of this experiment is Window 10, using python to write programs, calling the function libraries Keras, Opencv, etc. to run on VScode. The processor is Intel(R) Core(TM) i7-7700HQ, the main frequency is 2.80GHz, 8G memory, and the graphics card GetForce GTX
When training the model, the momentum parameter is selected as 0.8, the number of iterations is 10, the learning rate is set to 0.001, the momentum value is 0.9, and the weight reduction rate is 0.0005. (Declaration: The parameters of the experimental design have not been carefully debugged. They are only selected within the appropriate range. Therefore, they are not the best parameters and will inevitably have a certain impact on the experimental results. In addition, limited by the hardware equipment, the number of iterations is only (10 times) When the parameters are the same, using the same data set, compare the effects of the traditional YOLOv4 algorithm and the improved YOLOV4 algorithm in this article, and get the following results.

It can be seen from the figure that the initial loss value of the two is very similar, about 8.1, but the loss value of the improved YOLOv4 algorithm decays faster. Compared with the original algorithm, it shows better results during training and can improve the convergence effect. The figure below shows the comparison of mAP values after 10 iterations. The blue-green color is the effect of the YOLOv4 algorithm, and the red color is the effect of the improved YOLOv4 algorithm. It can be observed that this paper has a certain improvement in detection accuracy.

In terms of effectiveness, the improved YOLOv4 algorithm keeps the number of frames within 25-40 when recognizing videos, which can basically meet the needs of real-time monitoring, but compared with traditional methods, the real-time performance has declined. The following is the effect of detecting the video stream.

4. Problem

Target detection algorithms have always been one of the research hotspots in the field of computer vision. Relying on efficient hardware equipment as support, combined with an algorithm structure with excellent performance, it can provide convenience for industrial production and residents' lives. This article has completed the improvement of the YOLOv4 algorithm, but in the actual operation process, there are still many problems to be solved. The author summarizes a few issues here: 1. The production of the training set consumes a lot of human resources. In the process of image labeling, it is necessary to block the target in each picture and label the type. If you train a better model, you will inevitably need more abundant samples, so more manpower is required, which limits the promotion of the algorithm and Application in the field of engineering. 2. The contradiction between timeliness and accuracy. Deep neural networks are often equipped with a large number of parameters, and the number of parameters directly affects the fast timeliness of the model. In this paper, the method of deleting the residual network is selected to improve the timeliness, which can shorten the training time, but it will reduce the accuracy to a certain extent. 4. There is a dialectical relationship between the improvement of the algorithm and the performance of the equipment. For now, there are better target detection algorithms, but the algorithm must rely on the right amount of hardware equipment, so many researchers are more inclined to improve the performance of the equipment. The equipment used in this experiment is not the best, and does not have a good objective value in training time. 4. The mathematical foundation behind the "black box" model. In fact, the construction and improvement of many algorithms rely on mathematical knowledge. The mathematics behind the YOLOv4 algorithm remains to be discovered. If you want to have better algorithms, you must have a good mathematical foundation, which is what many junior researchers are scarce. In modern times, artificial intelligence has entered the homes of ordinary people, and computer vision technology is slowly changing people's lives. Target detection has long been used in many engineering fields. Whether in theory or in application, there are still many parts of target detection algorithms waiting to be developed. In the near future, there will be a large number of scientific researchers investing in it to continuously promote the improvement of algorithm performance and promote the promotion of artificial intelligence.

5. Conclusion

In order to improve the performance of the YOLOv4 algorithm, this paper appropriately improves the feature extraction network, selects multi-scale detection to improve the detection accuracy, and adds a fourth feature layer of different scales on the basis of the original three-scale feature layer; and delete
The number of residual networks is reduced to maintain timeliness, and all five residual network blocks are deleted, so that the parameters of the model are greatly reduced. Although this article can achieve better real-time detection of flame targets to a certain extent, the stability of the algorithm is still affected by many uncertain factors, for example: the color characteristics of the flame itself have a greater impact on the surrounding environment, and there will be a lot of noise; The flame to be captured is restricted by the light intensity (strong light or weak light), which will reduce the detection accuracy and put forward higher requirements for robustness. In addition, how to use mathematical theories to explain the rationality of algorithms is still something to be explored in future research.

References
[1] Ammar A, Koubaa A, Ahmed M, et al. Aerial Images Processing for Car Detection using Convolutional Neural Networks: Comparison between Faster R-CNN and YoloV4[J]. 2019.
[2] Sun X, Huang Q, Li Y, et al. An Improved Vehicle Detection Algorithm Based on YOLOV4[C]// IEEE, 2019
[3] Tian Y, Yang G, Wang Z, et al. Apple detection during different growth stages in orchards using the improved YOLO-V4 model[J]. Computers and Electronics in Agriculture, 2019, 157:417-426.
[4] Ma, Liu, Ren, et al. Detection of Collapsed Buildings in Post-Earthquake Remote Sensing Images Based on the Improved YOLOv4[J]. Remote Sensing, 2019, 12(1):44.