Computer Systems Science & Engineering
DOI: 10.32604/csse.2023.030849
Article

Tech Science Press

Fine Grained Feature Extraction Model of Riot-related Images Based on YOLOv5

Shaofan Su[1], Deyu Yuan[2,*], Yuanxin Wang[2] and Meng Ding[3]

[1]Department of Information Cyber Security, People's Public Security University of China, Beijing, 100038, China
[2]Key Laboratory of Safety Precautions and Risk Assessment, Ministry of Public Security, Beijing, 102623, China
[3]Public Security Behavioral Science Laboratory, People's Public Security University of China, Beijing, 102623, China
*Corresponding Author: Deyu Yuan. Email: yuandeyu@ppsuc.edu.cn
Received: 03 April 2022; Accepted: 26 May 2022

Abstract: With the rapid development of Internet technology, the type of information in the Internet is extremely complex, and a large number of riot contents containing bloody, violent and riotous components have appeared. These contents pose a great threat to the network ecology and national security. As a result, the importance of monitoring riotous Internet activity cannot be overstated. Convolutional Neural Network (CNN-based) target detection algorithm has great potential in identifying rioters, so this paper focused on the use of improved backbone and optimization function of You Only Look Once v5 (YOLOv5), and further optimization of hyperparameters using genetic algorithm to achieve fine-grained recognition of riot image content. First, the fine-grained features of riot-related images were identified, and then the dataset was constructed by manual annotation. Second, the training and testing work was carried out on the constructed dedicated dataset by supervised deep learning training. The research results have shown that the improved YOLOv5 network significantly improved the fine-grained feature extraction capability of riot-related images compared with the original YOLOv5 network structure, and the mean average precision (mAP) value was improved to 0.6128. Thus, it provided strong support for combating riot-related organizations and maintaining the online ecological environment.

Keywords: Convolutional neural network; YOLOv5; riot-related; fine grained; target detection

1 Introduction

In recent years, with the rapid development of technology, the Internet has been rapidly popularized all over the world, and riot activities have penetrated into the Internet from the original underground. Many riot organizations began to organize riot activities through social media. Therefore, in order to combat the activities of riot organizations, the public security organs should not despise the monitoring needs of riot activities on the Internet. They should not only passively discover the activity track of riot organizations, but also actively discover them. Therefore, in a wide range of media categories [1], we need to pay attention to whether the image data transmitted in the network is related to riot. CNN algorithm is a

convolutional neural network, which is one of the most widely used algorithms. It convolutes the input data according to the matrix requirements, then finds out the characteristics [2] of the data according to different convolution operations [3], and finally outputs the results according to the full connection layer. Its algorithm is very suitable for image data training and processing. In the field of target detection, there are two-stage detection represented by Region-CNN (RCNN) and single-stage detection represented by YOLO series. Double stage detection has high precision, but the speed is relatively slow. Considering the actual needs, this paper adopted YOLO series algorithm as the solution, which has the characteristics of high precision, fast detection and small volume.

In the anti-riot practice, various scenes are complex, so there is more noise [4] in the data set, which has a great impact on the training and recognition of the traditional machine learning model. Therefore, improving the recognition of high noise images is a practical and challenging topic. Based on the YOLOv5 algorithm, aiming at the optimization of its loss function [5] and adaptation algorithm, as well as the optimization processing of the data set, and further parameter optimization through genetic algorithm, this paper designed a riot-related image recognition system based on improved YOLOv5. According to the results, the effect was better than that of the previous generation YOLO and RCNN.

Therefore, we use YOLOv5, the most advanced algorithm in the field of target detection, as the underlying algorithm. After improvement, we completed the classification and detection of riot-related images, and the effect could be applied to practical combat. At present, YOLO series [6] has developed from v1 to v5, but in the field of riot-related image detection, this paper used YOLOv5 to detect the characteristics of riot-related images. With the optimization [7] of model parameters, it is found that the effect is better than the previous generation YOLOv3 and YOLOv4.

The following five sections describe the organization of the full paper. Section 1 introduces the current state of research in the field of riot-related image recognition and the network architecture of YOLOv5. Section 2 describes our improvement strategy for the original YOLOv5. Section 3 describes the general implementation architecture of our fine-grained recognition of riot-related images. Experiments and simulations are presented in Section 4. Finally, Section 5 is the conclusion and outlook.

2 Related Works

With the rapid development of Internet technology, the activities of some riot-related organizations have been gradually transformed from the traditional real scene to the content of the Internet, and through the connectivity of the Internet, they spread violence-related content and are used to spread harmful and healthy ideas, which have brought very bad influence to the society.

Traditional image detection algorithms [8] have gradually fallen out of fashion, while neural networks and deep learning methods are very effective and fast. Throughout the years of the development of the application of deep learning to target detection, deep learning algorithms [9] based on image target detection algorithms can be composed of the following two categories: neural networks and deep learning gradually replaced the traditional image detection algorithms and become the mainstream of target detection methods. Throughout the years of deep learning target detection development, a series of target detection algorithms [10] based on deep learning algorithms can be broadly divided into two major types of algorithms: two-stage algorithm, which first generates candidate regions and then performs CNN classification, represented by the RCNN series. The one-stage algorithm, which directly applies the algorithm to the input image and outputs the category and corresponding localization, is represented by the YOLO series of algorithms.

The R-CNN algorithm is the representative of two-stage, for example, the first one is a region proposal, and then the CNN algorithm is used for classification and recognition. Since the region proposal plays a key

role in the evaluation of the algorithm, the method is named after the initial R of Region plus CNN [11]. Although the two-stage image target detection algorithm is evolving and the detection accuracy is getting higher and higher, there is always a bottle neck of speed in the two-stage process. In some scenarios when real-time target detection is required, the R-CNN family of algorithms is ultimately lacking. The main idea of the YOLO algorithm series is to obtain the class and specific location of the target object directly from the input image, instead of generating candidate regions as in the R-CNN series. The direct effect of this is that it is fast. The YOLO algorithm is used in this paper to investigate the above factors.

So far, YOLO algorithm has five generations. The first generation [12] was proposed by the founder Joseph Redmon in 2015. Until the third generation, Joseph Redmon, the founder of YOLO series algorithms, announced to stop supporting project development. Due to the open source nature of the project, Alexey Bochkovskiy and others developed the fourth generation according to Darknet [13], which has better performance. However, the disadvantage is that the code is miscellaneous and the volume of generated results is too large. Therefore, the fifth generation, developed by the American company ultralytics LLC [14], is the best iterative version of the YOLO series algorithm at present. Based on the latest generation of YOLOv5, this paper is divided into input layer [15], backbone layer (benchmark network), neck layer and output layer. Its network structure is shown in Fig. 1.

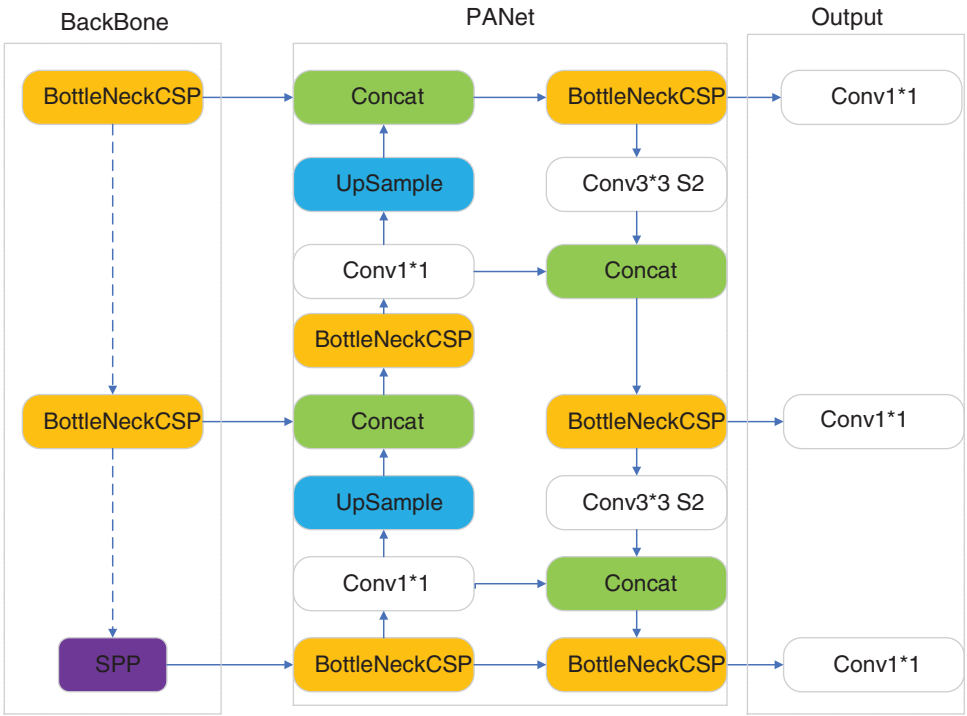

Figure 1: YOLOv5 network architecture

In the input layer, YOLOv5 uses mosaic method to enhance the input data, which is different from the previous generation of cutmix data enhancement. Mosaic data enhancement is the splicing of four pictures by random scaling, random clipping and random arrangement, which can effectively improve the recognition of small targets.

In the backbone layer of YOLOv5, the focus structure [16] is preferentially adopted in YOLOv5, that is, the input image is sliced first, then convoluted through the 32-layer convolution network, and finally the feature map is obtained. Then cross stage partial (CSP) structure [17] is adopted in the backbone network.

In the neck layer of YOLOv5, not only the structure of feature pyramid networks (FPN)+ pixel aggregation network (PAN) [18–26] is adopted, but also the structure of CSP2 is added to enhance the network feature fusion capability.

In the output layer, YOLOv5 uses generalized intersection over union (GIOU_loss) [27] as the loss function to solve the error between the predicted target frame and the real target frame. The loss function is used to measure the inconsistency between the predicted value of the model and the real value. It is a non-negative real-valued function. The smaller the loss function, the better the robustness of the model. And non-maximum suppression (NMS) [28] is used to solve the overlap problem of predicted frames of multiple targets in the same image data.

3 Algorithm

3.1 Selection of Optimization Algorithms

In the YOLOv5 algorithm, the default optimization algorithm is stochastic gradient descent (SGD) [29], that is, stochastic gradient descent algorithm, but it may encounter the problem that the direction of each update has uncertainty, which leads to large fluctuations in the loss function, especially for our specific dataset, where some categories are unevenly classified and setting a uniform learning rate leads to inaccurate recognition of some categories.

Adaptive momentum (Adam) optimization algorithm is an extension of stochastic gradient descent method, which is widely used in computer vision and natural language processing for deep learning applications. Adam combines the optimal performance of adaptive gradient (AdaGrad) and root mean squared propagation (RMSProp) algorithms [30], and it also provides an optimization method to solve sparse gradient and noise problems. The Adam algorithm integrates the first-order momentum of SGD and the second-order momentum of RMSProp. The two most common hyperparameters in optimization algorithms, β_1 and β_2, are included. The former controls the first-order momentum, and the latter controls the second-order momentum. The details are in the following formula:

$$m_t = \beta_1 \times m_{t-1} + (1 - \beta_1) \times g_t \tag{1}$$

$$\mathrm{V}_t = \beta_2 \times \mathrm{V}_{t-1} + (1 - \beta_2) \times g_t^2 \tag{2}$$

It has been proved that ADAM is better than SGD in practice due to the adaptive learning rate, with some parameters optimized, when the data set and model are identical. In addition, ADAM has the advantages of fast convergence and easy parameter tuning.

3.2 A Genetic Algorithm to Further Optimize Hyperparameters

A genetic algorithm, that is stochastic global search and optimization methods developed to mimic biological evolutionary mechanisms in nature, are used for iterative optimization of parameters in models. Since it simulates the recombination of chromosomes during division, it is particularly suitable for optimizing parameters in machine learning algorithms. In this system, the application of genetic algorithm was added to optimize the parameters.

By applying the genetic algorithm at a later stage, the YOLOv5 network was trained and evolved with about 25 hyperparameters for various training settings, and the optimal hyperparameters were selected by 50-fold comparison based on the initial parameters set before, with 100 iterations. And the training accuracy finally reached 0.6128. The specific schematic diagram is shown in Fig. 2.

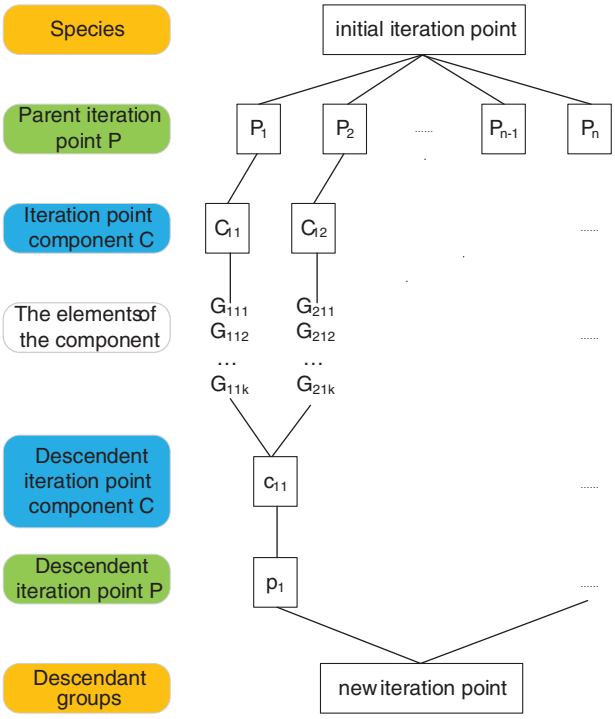

Figure 2: Overall diagram

3.3 Backbone Selection

YOLOv5 adds CSP as a backbone compared to several generations ago, and chooses PyTorch as the framework in the code structure, the size is very small compared to previous generations, and the code structure is clear and easy to learn. It also has the best performance among all previous generations. It is also still open for updates.

However, when the backbone network was modified to SwinTransformer, the feature maps at multiple resolutions could be output. Compared with the original Transformer, the SwinTransformer does Self-Attention for N tokens, and its time complexity is $\mathrm{O}(N^2)$. However, the SwinTransformer divided the N tokens into N/n groups, treats n as a constant, and then processes the n tokens. For attention calculation, the time complexity is reduced to $O(\mathrm{N}{*}n^2)$. Since n is a constant, the time complexity can be approximately equal to $O(N)$.

In addition, based on the native Transformer algorithm, Patch Partion and Linear Embedding layers are used. The two operations together are called the Patch Merging layer, similar to the Pooling layer in CNN. The function of Patch Merging is to realize downsampling of the image, which is realized by the nn.*Unfold* function. The function of this function is to use the sliding window mechanism for the image, which is equivalent to the sampling method in the convolution operation. Therefore, the parameters of this step include the size of the window and the step size of the sliding window. Therefore, the downsampling formula of each channel in the image can be expressed as the formula (3):

$$z^0 = \mathrm{MLP}(\mathit{Unfold}(\mathit{Image})) \tag{3}$$

The role of Patch Merging is similar to the maximum pooling layer of CNN, but the maximum pooling used by CNN to achieve down sampling will discard some information, so using Patch Merging can increase the accuracy of the model.

Another advantage of using Swin Transformer is that the core point of the algorithm uses the Swin Transformer Block, which consists of Window Multi-Head Self-Attention (W-MSA) [31–33] and Shifted-Window Multi-Head Self-Attention (SW-MSA) [34–36], as shown in Fig. 3.

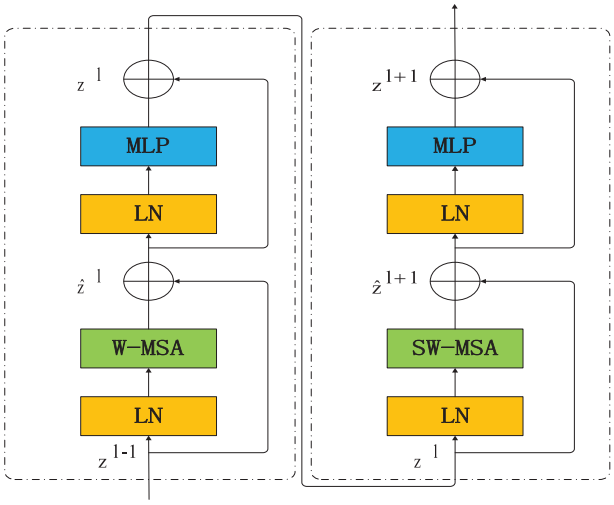

Figure 3: Swin transformer backbone

Therefore, in the core algorithm, the feature z^{l-1} is first normalized through Layer Normalization (LN) [37], and then the feature is learned through W-MSA, and then $\hat{z}^l$ is obtained through the residual operation. After another LN layer, a Multilayer Perceptron (MLP) layer and a residual layer [38], the output feature z^l of this layer is obtained. The structure of the SW-MSA layer is similar to that of the W-MSA layer. Except for the shifted bool value, there is no difference in other parameters. This part can be expressed as the following formula:

$$\hat{z}^l = W - MSA(LN(z^{l-1})) + z^{l-1} \tag{4}$$

$$z^l = MLP(LN(\hat{z}^l)) + \hat{z}^l \tag{5}$$

$$\hat{z}^{l+1} = SW - MSA(LN(z^l)) + z^l \tag{6}$$

$$z^{l+1} = MLP(LN(\hat{z}^{l+1})) + \hat{z}^{l+1} \tag{7}$$

Experimentally, we found that the effect was improved by 2%, and we judged that the application of this system in the field of riot is in line with the real-world needs, so we gave it an application.

4 Experimental Results

4.1 Data Preprocessing

The data were collected from the Internet, including more than 1800 positive samples and 1800 negative samples. The data set consists of internal data set and crawler to crawl the existed images on the Internet, crawl the images by searching the key words “riot organizations” and so on. After crawling the images, manual screening was required. In the process of screening the dataset, we mainly selected images with clearer riot-related elements, and when selecting photos, we should pay attention to exclude similar images of a certain army and so on. In addition, we also selected some pictures of the remains of the scene after the riot attack, and the pictures selected as the dataset were as wide and typical as possible.

The data that meet the conditions are filtered out. In order to simulate the real combat situation, a ratio of 10:1 was used for the test.

By downloading the images related to riot on the Internet, the dataset of images related to riot with significant features, such as blood, flames, dangerous weapons, etc., was constructed according to the requirements of anti-riot. Then, we manually labeled the riot-related elements in the images and output them in Visual Object Classes (VOC) format [39] to facilitate format conversion, and converted the labeled datasets to YOLO format uniformly, and used the datasets for model training.

Before learning and training, we need to pre-process the collected data. Firstly, data cleaning was done to eliminate the poor and inconspicuous features of the data set, which includes labeling errors, too few or too many labeled boxes [40]. Next, we performed data enhancement, using contrast and brightness, and even adding motion blur, changing the perspective, to enhance the existing data, so that the input data can be better applied to the detection system for riot-related images.

4.2 Experimental Results

This section describes the experimental procedure and presents the simulation results. When comparing the original YOLOv5 with SwinTransformer-YOLOv5 (improved YOLOv5), our improved algorithm achieved good results for riot-related image data.

4.2.1 Simulation Training

This project used a custom image dataset and contains many smaller datasets such as guns and masks, so the Adam optimization function is a more suitable choice. At this stage, we could train our own detection system using the built convolutional neural network and the training set of images. The results of the training execution were tested when the default resolution is set to 640, and the best results were achieved when the resolution is set to 320.

After previous experiments, the epochs were around 700 when the threshold value occurred due to overfitting, and the model was observed in real time using Tensorboard visualization. After each training, the model would be saved in the run folder, and the parameters would be continuously adjusted and configured through experiments to get the best model.

4.2.2 Experimental Results

After getting the trained model, we used the test dataset to test the trained model by executing the model test code in cmd.exe under YOLOv5 path to get the test result of the trained model as shown in Fig. 4. After inputting image data, our model annotated riot-related fine-grained features such as suspected dangerous weapons, flame explosions, and thick smoke.

In the YOLOv5 path, executed the real-time detection command of Tensorboard to observe the trained model in the folder run in real time. As a built-in visualization tool, it ploted the log file output of the program and displayed it directly in the browser for visualization purpose.

This paper used the mAP value as an indicator to evaluate the effect of the model. mAP in the detection of multiple categories of objects, each category can draw a curve according to recall and precision, AP is the area under the curve, and mAP is the average of multiple categories of APs. Before mAP was calculated, two concepts need to be used: precision and recall. Among them, the positive example means that there is an object of the corresponding category in the position. A negative example means that there is no object of the corresponding category at the location. Therefore, TP (True Positives) refers to the number of positive samples that are predicted to be positive samples. FP (False Positives): The number of negative samples predicted as positive samples. FN (False Negative): The number of positive samples predicted as negative

samples. TN (True Negative) refers to the number of negative samples that are predicted to be negative samples. Therefore, the formulas for precision, recall values can be derived from the following formulas:

$$\text{Precision} = TP/(TP + FP) \tag{8}$$

$$Recall = TP/(TP + FN) \tag{9}$$

$$Accuracy = (TP + TN)/(TP + FP + TN + FN) \tag{10}$$

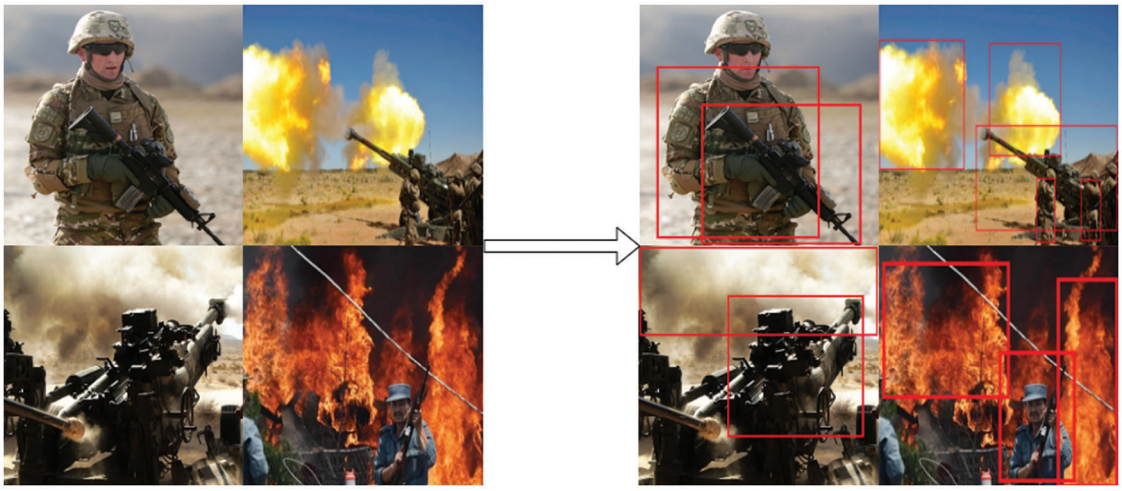

Figure 4: Detection results

AP calculation can be defined as the area of the interpolated precision-recall curve and the X-axis envelope. This method is called: AUC (Area under curve). Among them, $r_1\ r_2\ \ldots\ r_n$ are the recall values corresponding to each segment in the Precision interpolation segment in ascending order. $p_{interp}(r_{i+1})$ refers to the area enclosed by the y-axis (precision) in each interval on the x-axis (recall). The AUC area is then obtained by integration. The AUC curve of the SwinTransformer-YOLOv5 model is shown in Fig. 5. The final mAP value is obtained by calculating the AP value for the K categories separately, and finally summing and averaging. Therefore, the formula for calculating the mAP value is:

$$\text{AP} = \sum_{i=1}^{n-1}(r_{i+1} - r_i)p_{interp}(r_{i+1}) \tag{11}$$

$$\text{mAP} = \frac{\sum_{i=1}^{K} AP_i}{K} \tag{12}$$

As shown in Fig. 6, this paper compared YOLOv5 before the improvement with YOLOv5 after the improvement, and found that the mAP value before the improvement was 0.5327 and after the improvement was 0.6128, an improvement of about 15%.

Tab. 1 shows the comparison of experimental results of different versions of YOLOv5, where Algorithm 1 before improvement is the YOLOv5 directly reproduced using pytorch, and SwinTransformer-YOLOv5 is the improved algorithm for detection of riot-related features. Although the precision value was slightly decreased, the final mAP value was improved under the same data set, which verifies the enhanced feature detection capability of the algorithm.

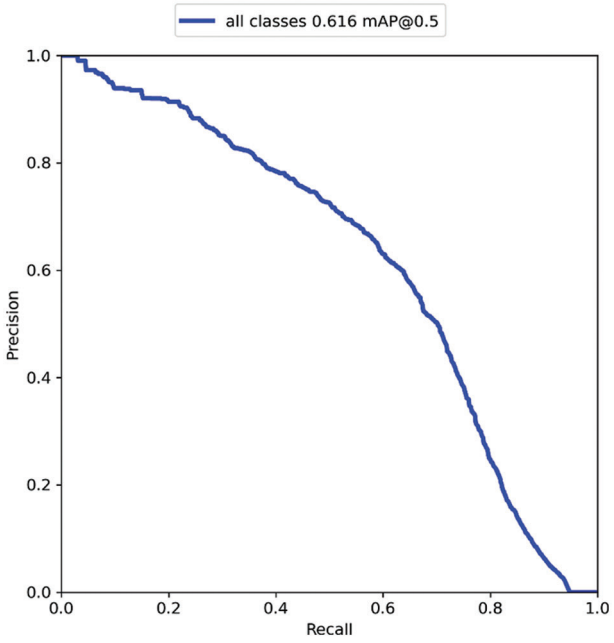

Figure 5: The AUC curve

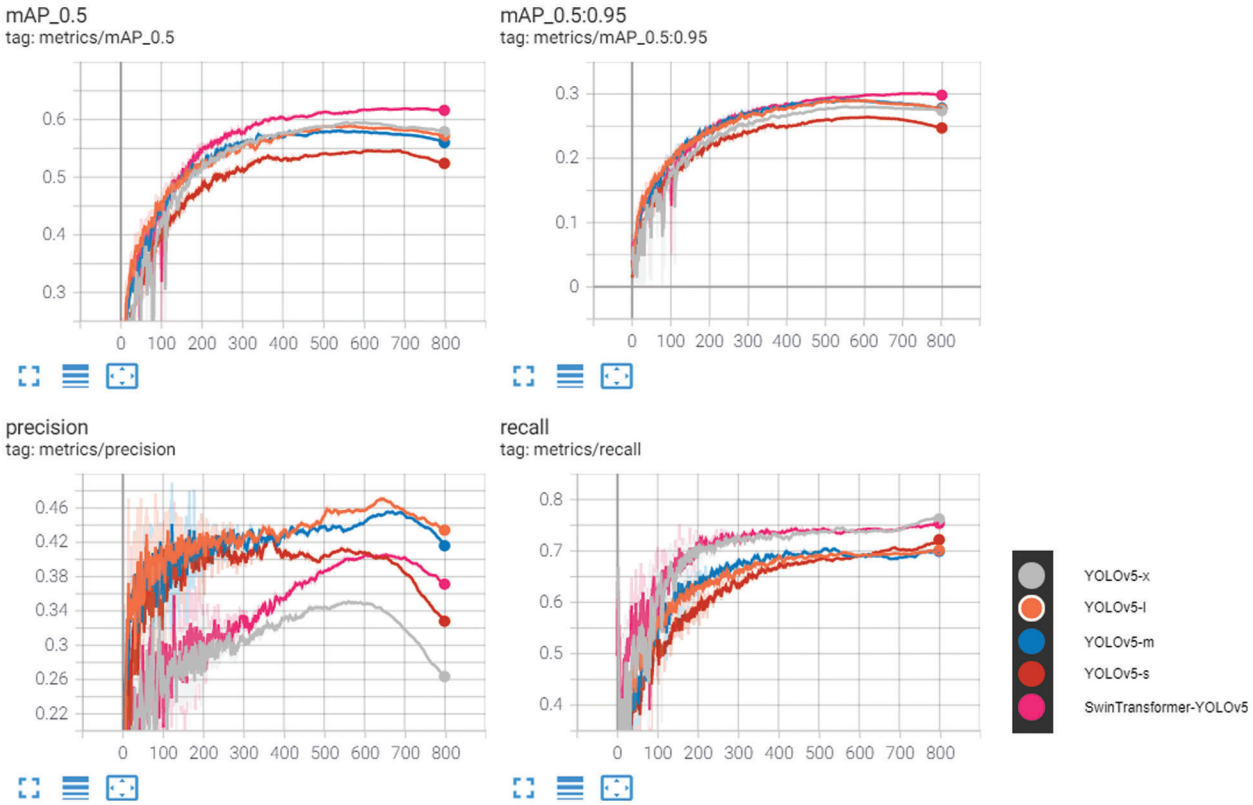

Figure 6: Results comparison chart

Table 1: Comparison of experimental results before and after improvement

name	Optimization algorithm	Genetic algorithm optimization	Backbone	mAP value
YOLOv5-s	SGD	False	CSPNet	0.5327
YOLOv5-m	SGD	False	CSPNet	0.5766
YOLOv5-l	ADAM	False	CSPNet	0.5852
YOLOv5-x	ADAM	True	CSPNet	0.5929
SwinTransformer-YOLOv5	ADAM	True	SwinTransformer	0.6128

In this paper, after several training sessions, the model was found to converge at 650 epochs, so the epoch value chosen was 650. mAP of the network model was significantly improved by replacing the backbone network structure of CSPNet with SwinTransformer and the improved optimization function Adam compared to the traditional Yolov5 algorithm. The effect of modifying the backbone network was more obvious, while the effect of improving the optimization function was slightly weaker, which is related to the role played by the two modules in the model, while the hyperparameter optimization by genetic algorithm made the improvement of mAP value further. From the results, it could be seen that the improvement of the backbone network of the model plays a crucial role in the effectiveness of the fine-grained identification of the present riot-related features, and the mAP value reached 0.6128 for a specific data set, which is sufficient to meet the practical requirements.

5 Summary and Outlook

With the increasing popularity of the Internet, many riot organizations have started to use social media channels to organize riot activities and spread riot ideas, and some riot organizations "Islamic State" have not only used different accounts on many platforms to organize propaganda related to riot, but also released the app "The Dawn of Glad Tidin" as a way to demagogue people and spread riot ideas. Due to its complex social background and historical factors, the threat of international riot the world continues to grow and counter-riot efforts need to be continued. Through the application of YOLOv5 algorithm in the field of fine-grained feature extraction of riot-related images, with the expansion of the data set, and parameters with further iterations of the genetic algorithm, the accuracy of recognition can be further improved for the detection of riot-related information such as images disseminated on the Internet. However, there are still some shortcomings in this system. For example, although the overall mAP values have reached a certain level, the accuracy is not very high in the recognition of certain types of riot-related elements. In the construction of the data set, the data crawled from the Internet using crawlers, the resolution and brightness angle vary, and it is difficult to ensure the quality, and the number of samples is not balanced. All of these have played a big obstacle to the implementation of the model.

In this paper, through the improved YOLOv5 network model, the fine-grained extraction of images with riot-related content crawled from the Internet through crawlers, with an overall mAP of 0.6128, can effectively perform fine-grained extraction of image data in the Internet or in network streams to determine whether there are suspicions of promoting riot, and in addition, if it can be practiced in the field of public security, it can be used in In addition, if it can be practiced in the field of public security, it can be applied in a practical environment, which will be very useful for combating riot crimes, maintaining national stability and safeguarding the network environment from pollution. In this paper, we focused on the fine-grained extraction of riot-related image features based on YOLOv5, with a focus on

researching and summarizing the current algorithms with excellent results and speed in the field of target detection, and improving them to design an Internet content-oriented model. With the development of network and technology, better algorithms and models are bound to appear, which can continuously enrich the technical research in the field of security involving riot content.

There are many directions for improvement. Firstly, in the collection of data set, it is difficult to collect on the open Internet, the sample size is not enough, and the quality and resolution of the samples are not very high. Secondly, there are many loss functions in the model, and we can further optimize our model by modifying different loss functions. In the backbone network, we can add an attention mechanism to compensate for the lack of data set in a self-attentive way.

Funding Statement: This work was supported by Fundamental Research Funds for the Central Universities, People's Public Security University of China (2021JKF215), Key Projects of the Technology Research Program of the Ministry of Public Security(2021JSZ09) and the Fund for the training of top innovative talents to support master's degree program, People's Public Security University of china(2021yjsky018).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] Z. G. Qu, H. R. Sun and M. Zheng, "An efficient quantum image steganography protocol based on improved EMD algorithm," *Quantum Information Processing*, vol. 20, no. 53, pp. 1–29, 2021.

[2] Z. G. Qu, Y. M. Huang and M. Zheng, "A novel coherence-based quantum steganalysis protocol," *Quantum Information Processing*, vol. 19, no. 362, pp. 1–19, 2020.

[3] L. Sun, Y. L. Wang, Z. G. Qu and N. N. Xiong, "BeatClass: A sustainable ECG classification system in IoT-based eHealth," *IEEE Internet of Things Journal*, vol. 9, no. 10, pp. 7178–7195 2021. https://doi.org/10.1109/JIOT.2021.3108792.

[4] Y. Wei, F. R. Yu, M. Song and Z. Han, "User scheduling and resource allocation in HetNets with hybrid energy supply: An actor-critic reinforcement learning approach," *IEEE Transactions on Wireless Communications*, vol. 17, no. 1, pp. 680–692, 2018.

[5] S. Sun, J. Zhou, J. Wen, Y. Wei and X. Wang, "A DQN-based cache strategy for mobile edge networks," *Computers*, *Materials & Continua*, vol. 71, no. 2, pp. 3277–3291, 2022.

[6] A. Kuznetsova, T. Maleva and V. Soloviev, "Detecting apples in orchards using YOLOv3 and YOLOv5 in general and close-up images," in *Int. Symp. on Neural Networks*, Cham: Springer, vol. 12557, pp. 233–243, 2020.

[7] M. K. Eulaers, N. Hahn, S. Berger, T. Sebulonsen, Ø. Myrland *et al.,* "Detecting heavy goods vehicles in rest areas in winter conditions using YOLOv5," *Algorithms*, vol. 14, no. 4, pp. 114–114, 2021.

[8] S. R. Zhou and B. Tan, "Electrocardiogram soft computing using hybrid deep learning CNN-ELM," *Applied Soft Computing*, vol. 86, pp. 105778, 2020.

[9] W. Sun, X. Chen, X. R. Zhang, G. Z. Dai, P. S. Chang *et al.,* "A multi-feature learning model with enhanced local attention for vehicle re-identification," *Computers, Materials & Continua*, vol. 69, no. 3, pp. 3549–3560, 2021.

[10] W. Zhang and X. Nan, "Road vehicle tracking algorithm based on improved YOLOv5," *Journal of Guangxi Normal University (Natural Science Edition)*, vol. 40, no. 1, pp. 49–57, 2021.

[11] Z. Wu, H. Chen, Y. Peng and W. Song, "Visual SLAM for fusing lightweight YOLOv5s in dynamic environments," *Computer Engineering*, vol. 47, pp. 1–11, 2021.

[12] D. Zhang, J. Hu, F. Li, X. Ding, A. K. Sangaiah *et al.,* "Small object detection via precise region-based fully convolutional networks," *Computers, Materials and Continua*, vol. 69, no. 2, pp. 1503–1517, 2021.

[13] N. Li, X. Ye, H. Wang, X. Huang and S. F. Tao, "An improved YOLOv5-based method for SAR image ship detection in complex scenes," *Signal Processing* vol. 37, pp. 1–16, 2021.

[14] J. Yu and S. Luo, "A YOLOv5-based method for unauthorized building detection," *Computer Engineering and Applications*, vol. 57, no. 20, pp. 236–244, 2021.

[15] G. Zhang, W. Li and Y. Zhang, "Traffic sign recognition based on improved YOLOv5 algorithm," in *21 Proc. of the National Conf. on Simulation Technology*, Guiyang, China, pp. 164–231, 2021.

[16] J. Wang, Y. Wu, S. He, P. K. Sharma, X. Yu *et al.,* "Lightweight single image super-resolution convolution neural network in portable device," *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 15, no. 11, pp. 4065–4083, 2021.

[17] W. Sun, G. C. Zhang, X. R. Zhang, X. Zhang and N. N. Ge, "Fine-grained vehicle type classification using lightweight convolutional neural network with feature optimization and joint learning strategy," *Multimedia Tools and Applications*, vol. 80, no. 20, pp. 30803–30816, 2021.

[18] J. Yin, S. Qu, Z. Yao, X. Hu, X. Qin *et al.,* "Traffic sign recognition model based on YOLOv5 in hazy weather," *Journal of Computer Applications*, vol. 35, pp. 1–10, 2021.

[19] Y. S. Li and L. C. Liu, "Lightweight steel detection network with embedded attention mechanism," *Computer Applications*, vol. 41, pp. 1–11, 2021.

[20] S. K. Patnaik, C. N. Babu and M. Bhave, "Intelligent and adaptive web data extraction system using convolutional and long short-term memory deep learning networks," *Big Data Mining and Analytics*, vol. 4, no. 4, pp. 279–297, 2021.

[21] T. Y. Zhou, Q. B. Zhu, M. Huang and X. X. Xu, "Carrier chip defect detection based on lightweight convolutional neural network," *Computer Engineering and Applications*, vol. 58, pp. 1–10, 2021.

[22] J. Ren, P. Chen, Y. Chen, H. Liu and P. Sun, "Deep learning based multi-target motion trajectory prediction algorithm," *Computer Application Research*, vol. 1, pp. 296–302, 2021.

[23] J. Wang, Y. Zou, P. Lei, R. S. Sherratt and L. Wang, "Research on recurrent neural network based crack opening prediction of concrete dam," *Journal of Internet Technology*, vol. 21, no. 4, pp. 1161–1169, 2020.

[24] T. Liu, B. Zhou, Y. Zhao and S. Yan, "Ship detection algorithm based on improved YOLO V5," in *2021 6th Int. Conf. on Automation, Control and Robotics Engineering*, Dalian, China, pp. 483–487, 2021.

[25] Z. Zhong, Y. Xia, D. P. Zhou and Y. Yan, "A lightweight target detection algorithm based on improved YOLOv4," *Computer Applications*, vol. 41, pp. 1–8, 2021.

[26] A. Malta, M. Mendes and J. T. Farinha, "Augmented reality maintenance assistant using YOLOv5," *Applied Sciences*, vol. 11, no. 11, pp. 4758, 2021.

[27] J. Zhang, J. Sun, J. Wang and X. G. Yue, "Visual object tracking based on residual network and cascaded correlation filters," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 8, pp. 8427–8440, 2021.

[28] S. He, Z. Li, Y. Tang, Z. Liao, F. Li *et al.,* "Parameters compressing in deep learning," *Computers, Materials & Continua*, vol. 62, no. 1, pp. 321–336, 2020.

[29] S. Liu and S. S. Agaian, "COVID-19 face mask detection in a crowd using multi-model based on YOLOv3 and hand-crafted features," in *Multimodal Image Exploitation And Learning 2021*, Bellingham, Washington, USA, pp. 11734, 2021.

[30] H. Aung, A. V. Bobkov and N. L. Tun, "Face detection in real time live video using YOLO algorithm based on vgg16 convolutional neural network," in *2021 Int. Conf. on Industrial Engineering, Applications and Manufacturing (ICIEAM)* Sochi, Russia, pp. 697–702, 2021.

[31] D. Zhu, G. Xu, J. Zhou, E. Di and M. Li, "Object detection in complex road scenarios: Improved YOLOv4-tiny algorithm," in *2021 2nd Information Communication Technologies Conf. (ICTC)*, Nanjing, China, pp. 75–80, 2021.

[32] D. Qin, "Real-time drivers' violation detection on mobile terminal based on improved YOLOv4-tiny," *Computer Science and Application*, vol. 11, no. 5, pp. 1291–1300, 2021.

[33] Q. Zhang, Y. Wang, L. Zhu and J. Zhang, "Research on real-time reasoning based on jetson tx2 heterogeneous acceleration YOLOv4," in *2021 IEEE 6th Int. Conf. on Cloud Computing and Big Data Analytics (ICCCBDA)*, IEEE, Chengdu, China, pp. 455–459, 2021.

[34] M. M. Zbek, M. Syed and L. Ksz, "Subjective analysis of social distance monitoring using YOLOv3 architecture and crowd tracking system," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 29, no. 2, pp. 1157–1170, 2021.

[35] B. Li, Z. Gan, E. S. Neretin and Z. Yang, "Object recognition through UAV observations based on YOLO and generative adversarial network," in *IoT as a Service*, Cham: Springer, vol. 346, pp. 439–449, 2021.

[36] H. Jeon, G. Lee, B. Jeong and J. S. Choi, "Design and implementation of real-time vehicle recognition and detection system based on YOLO," in *Advances in Computer Science and Ubiquitous Computing*, Cham: Springer, vol. 715, pp. 25–30, 2021.

[37] S. R. Zhou, M. L. Ke and P. Luo, "Multi-camera transfer GAN for person re-identification," *Journal of Visual Communication and Image Representation*, vol. 59, no. 1, pp. 393–400, 2019.

[38] W. Wang, H. Liu, J. Li, H. Nie and X. Wang, "Using CFW-net deep learning models for X-ray images to detect COVID-19 patients," *International Journal of Computational Intelligence Systems*, vol. 14, no. 1, pp. 199–207, 2021.

[39] W. Wang, Y. Jiang, Y. Luo, J. Li, X. Wang *et al.,* "An advanced deep residual dense network (DRDN) approach for image super-resolution," *International Journal of Computational Intelligence Systems*, vol. 12, no. 2, pp. 1592–1601, 2019.

[40] W. Wang, Y. Yang, J. Li, Y. Hu, Y. Luo *et al.,* "Woodland labeling in chenzhou, China, via deep learning approach," *International Journal of Computational Intelligence Systems*, vol. 13, no. 1, pp. 1393–1403, 2020.