IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 40, No. 5, October, 2023, pp. 2267-2273
Journal homepage: http://iieta.org/journals/ts

Enhanced Campus Security Target Detection Using a Refined YOLOv7 Approach

Fengyun Cao*, Shuai Ma

School of Computer Science and Technology, Hefei Normal University, Hefei 230601, China

Corresponding Author Email: caofengyun@hfnu.edu.cn

https://doi.org/10.18280/ts.400544

Received: 22 May 2023
Revised: 18 August 2023
Accepted: 25 August 2023
Available online: 30 October 2023

Keywords:
transfer learning, attention mechanism, deformable convolution, object detection

ABSTRACT

In many educational institutions, safety management traditionally depends upon manual video surveillance, leading to potential delays in the identification and alerting of perilous activities, notably the possession of controlled knives and smoking behaviors exhibited by students. These activities possess significant consequences for both the psychological and physical well-being of students. Recognizing this pressing need, an augmented object detection method for campus security, rooted in YOLOv7, is presented. The EIoU (Efficient Intersection over Union) loss function has been substituted to expedite model convergence and heighten detection fidelity. Additionally, the integration of the CBAM (Convolutional Block Attention Module) attention mechanism with the DCNv2 (Deformable ConvNets v2) deformable convolutional kernel not only mitigates the challenge of information inundation but also enhances feature extraction capabilities, facilitating adjustments to geometric deformations. Experimental findings indicate that this proposed method achieves a detection accuracy of 92.6% across various categories on a dataset comprising three categories, spanning a total of 4500 images, and attains an mAP of 96.4%. In comparison to the conventional YOLOv7 algorithm, enhancements in detection accuracy and mAP by 6.9% and 6.6%, respectively, have been observed, affirming the efficacy of the presented algorithm.

1. INTRODUCTION

With the evolution of societal dynamics, emphasis on campus security has progressively intensified. Educational campuses, hosting a multitude of students and faculty, frequently witness activities leading to safety breaches and emergent situations, encompassing illicit possession of knives, smoking, altercations, and other deviant behaviors. Significant risks to campus harmony, order, and student welfare are posed by such occurrences. Traditional monitoring systems on campuses have been found to predominantly depend on manual scrutiny and video reviews for anomaly detection, rendering them prone to inefficiencies and errors.

Nevertheless, advances in computer vision and deep learning have substantially matured object detection and recognition technologies. It has been observed that by harnessing these technologies for early identification and intervention of campus hazards, automation and management levels of monitoring systems can be enhanced, effectively mitigating campus safety breaches and consequently ensuring the protection of both educators' and learners' lives and assets.

The realm of campus target detection and recognition, rooted in computer vision, seeks to implement real-time identification and preemptive alerts for on-campus targets via video surveillance. It has been noted that with the escalating progression of artificial intelligence, rapid advancements have been witnessed in campus anomaly detection. Within deep learning, predominantly two methodologies for target detection have been identified: two-stage and one-stage detection methods. The former, a two-stage detection, initially manifests candidate regions, subsequently leveraging convolutional neural networks for target classification and localization. Renowned algorithms within this methodology encompass R-FCN [1], Fast RCNN [2], and Mask-RCNN [3]. For instance, a face detection system based on R-FCN was proposed by Ruan et al. [4]. Despite achieving commendable accuracy in intricate backgrounds, its detection latency was found to be extensive. In contrast, one-stage detection, representative algorithms of which include the YOLO series [5, 6], SSD series [7, 8], and RetinaNet [9-11], directly employ convolutional neural networks for target classification and location prediction. They have been observed to offer more expedient detection, aligning better with real-time requirements. For instance, an improved YOLOv5 algorithm tailored for airport security was implemented by Guo et al. [12], enhancing detection velocity and precision upon loss function alteration. Similarly, a cigarette detection model founded on YOLOv5, with an integrated SENet attention mechanism, was presented by Li et al. [13], showcasing efficient detection for minute targets such as cigarettes.

An algorithm geared towards campus safety object detection, rooted in the YOLOv7 single-stage object detection framework, is elucidated in the subsequent sections. Emphasis has been laid on the detection and classification of three pivotal categories: faces, knives, and cigarettes. Relative to its predecessor, YOLOv5, the YOLOv7 framework displays enhanced accuracy and velocity but showcases limitations in petite object detection. Modifications including CIoU-Loss replacement with EIoU-Loss [14], and integration of the CBAM attention mechanism [15-17] and deformable convolutions [18-20], have been observed to collectively bolster the algorithm's efficacy in campus safety object

detection.

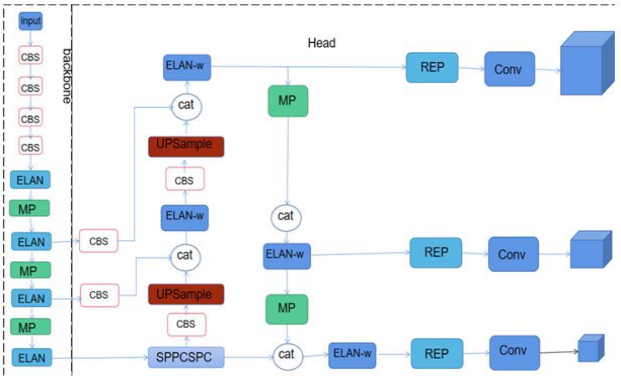
The contributions of this research comprise: (1) Construction of a dataset mirroring genuine campus management scenarios, encompassing three primary categories and aggregating 4500 real-life instances. (2) Integration of strategies such as the EIoU loss function, CBAM, and the DCNv2, resulting in considerable enhancement of model detection accuracy. (3) Implementation of ablation studies to ascertain the influence of each introduced modification on model efficiency.

The ensuing sections of this study are structured as follows: Section 2 delineates the related work; Section 3 explicates the method's design and data architecture; Section 4 undertakes experimental comparative analyses; Section 5 encapsulates the methodology's strengths and areas of enhancement, paving the path for future exploration.

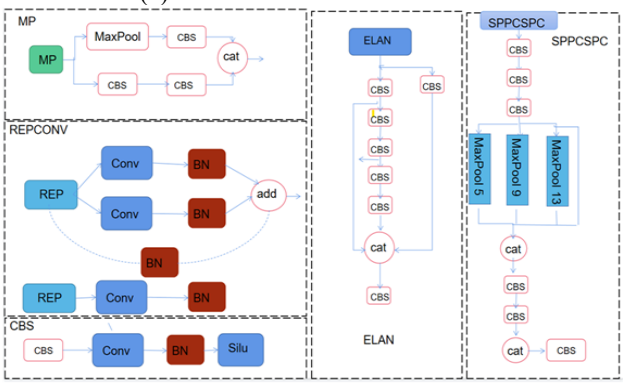
2. PRELIMINARIES

2.1 YOLOv7 architecture

The YOLO algorithm, recognized as a quintessential one-stage target detection approach, capitalizes on neural networks for object recognition and positioning. Notably, it boasts rapid execution, facilitating its integration into real-time systems. As the latest iteration in the YOLO series, YOLOv7 [21, 22] is observed to surpass extant detectors in terms of speed and accuracy across the spectrum of 5-160 FPS.



(a) Overall network architecture



(b) Network architecture of each module

Figure 1. Structure of YOLOv7

As depicted in Figure 1, the YOLOv7 architecture is trifurcated into three distinct components: the input, the backbone, and the head networks. Initially, images undergo a preprocessing stage wherein their dimensions are conformed to a standard 640×640×3. Subsequent to this preprocessing, images are channeled into the backbone network for salient feature extraction. This network then predicts three core tasks associated with image detection: classification, front and back background classification, and border. The culmination of this pipeline produces the detection results. The Backbone network, intricately designed, comprises 50 layers. These layers encompass CBS composite modules, the efficient converging structure known as ELAN, and the MP modules, with SiLU serving as the activation function. In the MP module, a Maxpool layer is superimposed atop the CBS layer, bifurcating the design into an upper and a lower branch. This design culminates in a Concat operation that synergizes features extracted from both branches, thereby enhancing the network's feature extraction prowess. The ELAN module, an amalgamation of assorted convolutions, regulates both the shortest and the longest gradient paths, ensuring that deeper networks assimilate and converge effectively. The head network, on the other hand, is an intricate blend of the SPPCSPC module, a suite of CBS modules, an UPSample module, a REPCONV module, and the Elan-w module. This ensemble ultimately yields three feature maps of varying dimensions. Within the SPPCSPC module, four different maxpools exist, specifically tailored for 5, 9, 13, and 1-sized entities, facilitating the handling of objects of disparate sizes. The REPCONV module bifurcates into training and deployment segments. Through a judicious layering of these models, YOLOv7 exhibits a commendable blend of precision and efficiency.

2.2 Loss function of YOLOv7

YOLOv7 adopts CIoU-LOSS [23-25] as the loss function of the detection frame, and its formulation is provided below:

$$L_{CIoU} = 1 - IoU + \frac{\rho^2\left(\mathrm{b}, \mathrm{b}^{gt}\right)}{c^2} + \alpha\nu \tag{1}$$

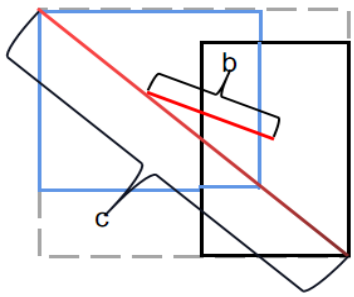


Figure 2. Schematic diagram of CIoU

Figure 2 elucidates the concept further, where $\boldsymbol{\rho^2(b, b^{gt})}$ signifies the Euclidean distance between the center points of the real frame and the prediction frame. Herein, c denotes the diagonal span of the smallest enclosing region encompassing both the prediction box and the real box.

$$\alpha = \frac{\nu}{(1 - IoU) + \nu} \tag{2}$$

$$\nu = \frac{4}{\pi}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \tag{3}$$

where, v is an parameter elucidating the consistency of aspect ratios between the prediction frame and the real frame. h^{gt} and w^{gt} represent the height and width of the real frame, while h and w encapsulate the height and width of the prediction frame.

3. RECOGNITION OF CAMPUS SECURITY TARGET

3.1 Target categories and sample compilation

As depicted in Figure 3, the experimental dataset was categorized into three distinct classes: faces, knives, and cigarettes. The images within each class were compiled from three primary sources: (1) Downloaded from the internet; (2) Captured using camera devices; (3) Modified by rotations, pans, zooms, and brightness adjustments to the aforementioned images. Each class encompasses 1500 images, from which 1200 were used for training purposes and the remaining 300 were designated for verification.

Figure 3. Some of the samples

3.2 Dataset annotations and enhancement

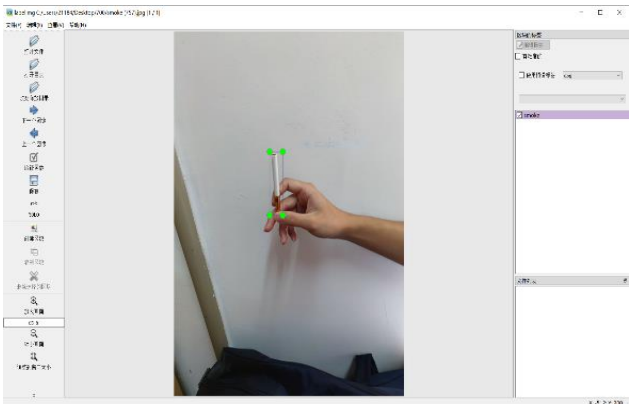


Figure 4. Dataset annotations

LabelImage, an open-source annotation tool, was employed to tag the three aforementioned classes. As illustrated in Figure 4, txt tag files in accordance with the YOLO format were generated. The tag parameters are divided into five tuples: (class_id, X, Y, W, H), defined as follows:

$$X = \frac{1}{2w}(x_1 + x_2) \tag{4}$$

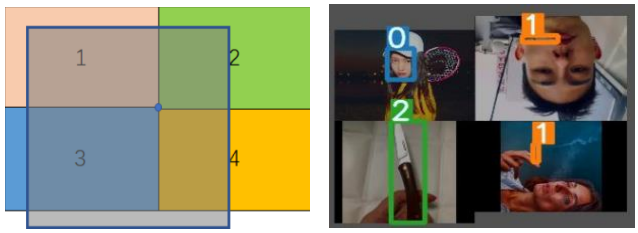
$$Y = \frac{1}{2h}(y_1 + y_2) \tag{5}$$

$$W = \frac{1}{w}(x_1 - x_2) \tag{6}$$

$$H = \frac{1}{h}(y_2 - y_1) \tag{7}$$

where, (x_1, y_1) and (x_2, y_2) represents the coordinates of the upper left and lower right corners of the labeled inspection frame, respectively. h and w indicate the height and width of the image, whilst class_id denotes the specific category under training.

For YOLOv7, a data enhancement technique termed "Mosaic data enhancement" was implemented. This method amalgamates four images through arbitrary zooms, crops, and layouts to generate a singular, novel image. Such an enhancement considerably diversifies the detection dataset. Notably, the random scaling introduces numerous diminutive targets, subsequently augmenting the network's stability. Simultaneously, it reduces GPU memory usage and facilitates direct computation on the quartet of images.



(a) Concept Map (b) design sketch

Figure 5. Mosaic data enhancement

Figure 5 offers a visual representation of this method. Four distinct images were randomly selected and colour-coded to delineate their individuality. These images underwent arbitrary zooms, crops, and arrangements to seamlessly fit within a specified frame, whilst extraneous sections were discarded.

3.3 Model improvement and training

Within this study, an augmented object detection algorithm, centred upon YOLOv7, was devised specifically for the campus environment. Initial steps involved the substitution of the EIoU loss function to expedite model convergence and elevate detection precision. The integration of the CBAM attention mechanism and the Deformable ConvNets v2 was subsequently pursued. This aimed to effectively circumvent information overload issues whilst bolstering the capacity for feature extraction and adaptation to geometric deformities.

3.3.1 Improvement of the loss function

Though CIoU-Loss has demonstrated considerable improvements in terms of convergence speed and detection accuracy, the parameter v, which ascertains the aspect ratio's consistency between the prediction and real boxes, has not been adequately defined. Here, parameter v merely indicates the difference in aspect ratio, neglecting the genuine relationship between the width and height of the prediction box in comparison with the target box. Such an oversight results in potential non-simultaneous regression of the prediction box's width and height once they converge to a linear ratio.

To better address the limitations inherent to CIoU-Loss, the study introduced EIoU-Loss as a replacement. EIoU-Loss is partitioned into three primary components: (1) Overlap loss

between the prediction frame and the real frame; (2) Loss of center distance between the prediction frame and the real frame (L_{dis}); (3) The width and height losses of the prediction box and the real box (L_{asp}). The formula can be written as:

$$L_{EIoU} = L_{IoU} + L_{dis} + L_{asp} = 1 - IoU + \frac{\rho^2\left(b, b^{gt}\right)}{c^2} + \frac{\rho^2\left(w, w^{gt}\right)}{c_w^2} + \frac{\rho^2\left(h, h^{gt}\right)}{c_h^2} \quad (8)$$

where, the first two components of EIoU-Loss retain the CIoU_Loss framework. Here, the aspect ratio's loss term is split into the difference between the prediction frame's width and height and that of the enclosing minimal frame. Such partitioning has been observed to foster accelerated convergence of the prediction frame and ameliorate regression accuracy.

3.3.2 Attention mechanism

Given the observation that the categories of face, cigarette, and knife typically constitute only a minor portion of the image, with the predominant region being natural background, the CBAM attention mechanism was embedded into YOLOv7's backbone network (see Figure 6). The objective was to diminish extraneous feature information unrelated to the primary target and enhance the feature extraction efficacy of the object detection model. Emulating the human visual attention system, CBAM re-calibrates extracted target features, automatically sieving out inconsequential data, thus optimizing visual information processing efficiency and precision.

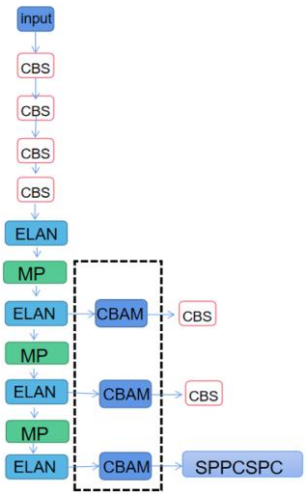


Figure 6. Location of improvements in CBAM attention mechanism

Further, the comprehensive CBAM attention mechanism is depicted in Figure 7, comprising both a Channel Attention Module and a Spatial Attention Module.

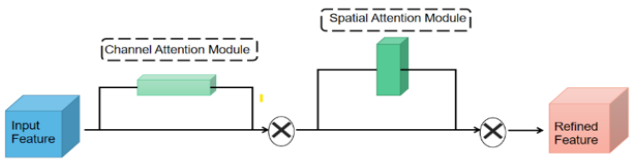


Figure 7. Structure of CBAM attention mechanism

For the Channel Attention Module, a 1D convolution was applied to the input feature map $F \in R^{C*H*W}$ to yield $M_c \in R^{C*1*1}$. Subsequently, this output was multiplied with the original feature map. The outcome of this process then served as input for the Spatial Attention Module. Here, a 2D convolution was employed to obtain $M_s \in R^{C*1*1}$, which was then multiplied with the original feature map. The formulas for calculating this sequence are:

$$F' = M_C(F) \otimes F \quad (9)$$

$$F'' = M_S(F') \otimes F' \quad (10)$$

where, $\otimes$ represents the multiplication of two matrices.

3.3.3 Deformable convolution

In deep learning, convolutional kernels are instrumental in feature extraction. However, conventional kernels, being static in size, often exhibit limited generalization capabilities, constraining adaptability to unforeseen modifications. Conversely, deformable convolutional kernels append direction parameters to each constituent element of the original structure. This flexibility broadens the extraction range for target features during training, enhancing the network's ability to adjust to geometric deformations. This methodology has been incorporated into the primary ELAN module of the backbone network, supplanting the conventional 3×3 convolution, as illustrated in Figure 8.

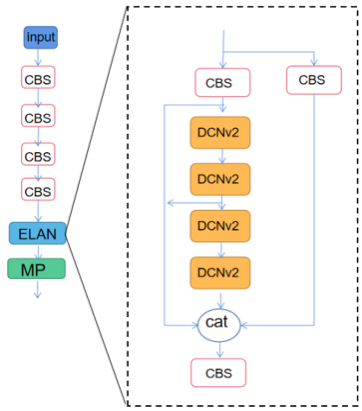


Figure 8. Improved location of DCNv2

The equation is:

$$y(P_0) = \sum_{P_n \in R} w(P_n) \bullet x(P_0 + P_n + \Delta P_n) \bullet \Delta m_k \quad (11)$$

DCNv2's innovation lies in its introduction of a moderating factor Δm_k, determining weights for the offset of input feature map sampling points, effectively eliminating irrelevant contextual information.

4. EXPERIMENTAL RESULTS

4.1 Experimental environment and parameters

Experiments were conducted within the deep learning framework, Pytorch. Table 1 shows the specific configuration

of the training environment. Prior to training, internal parameters of YOLOv7 were adjusted, with the corresponding experimental parameters detailed in Table 2.

Table 1. Experimental environment

Name	Parameters
CPU	AMD Ryzen9 4900H CPU @ 3.30GHz
GPU	NVIDIA GeForce GTX 2080TI 11G
Framework	Pytorch1.8.0
Programming language	Python3.8

Table 2. Experimental Parameters

Name	Parameters
Epochs	200
Batch-size	32
Weights	yolov7.pt
Image-size	640×640

4.2 Evaluation indicators

For performance assessment, the mean average precision (mAP) was adopted, calculated using the following formula:

$$mAP = \frac{\sum_{i=1}^{k} P(k) \Delta R(k)}{k} \tag{12}$$

where, k signifies the number of samples, P denotes accuracy, and ΔR reflects the change in recall rate, the formulas are:

$$P = \frac{TP}{TP + FP} \tag{13}$$

$$R = \frac{TP}{TP + FN} \tag{14}$$

In these equations, *TP* (True Positive) indicates instances where positive samples were correctly classified as positive; *FN* (False Negative) when positive samples were mis-classified as negative; *FP* (False Positive) where negative samples were wrongly predicted as positive; and *TN* (True Negative) when negative samples were aptly classified as negative.

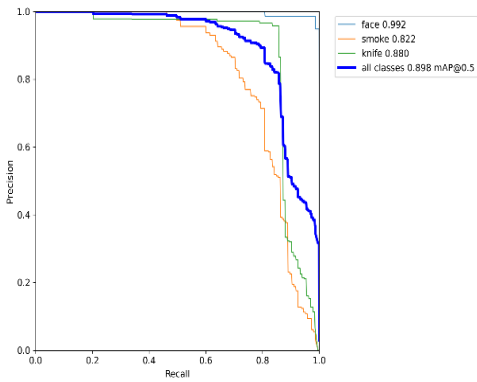
4.3 Comparison and analysis of results

To ascertain the effectiveness of the proposed algorithm, an ablation comparison experiment was undertaken, examining the impact of individual improvements on model performance. Three distinct improvements, namely the EIoU-Loss function, the CBAM attention mechanism, and the DCNv2 deformable convolution, were sequentially integrated into the YOLOv7 model. Under consistent experimental conditions, 200 iterations were performed, adopting mAP as the performance benchmark. The outcomes are summarized in Table 3, where a √ indicates the application of a specific improvement strategy.

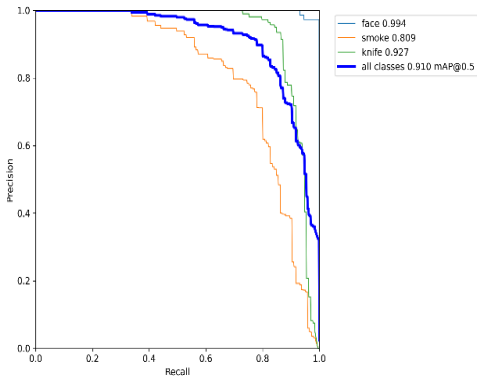
From Table 3, it can be inferred that YOLOv7-1 represents the original YOLOv7 algorithm with an mAP of 89.8%. For the YOLOv7-4 iteration, CIoU-Loss was substituted with EIoU-Loss, followed by the inclusion of the CBAM attention mechanism, and the final phase encompassed the integration of the deformable convolutional kernel DCNv2 in lieu of the original 3×3 convolution. It was observed that each incremental improvement bolstered performance. Compared to the foundational algorithm, the mAP surged by 6.6%, underscoring the enhancements' effectiveness. Notably, despite a marginal increase in detection time (0.4 ms), the criteria for real-time detection were still met. The training process is shown in Figure 9.

Table 3. Performance comparison of different improvement algorithms

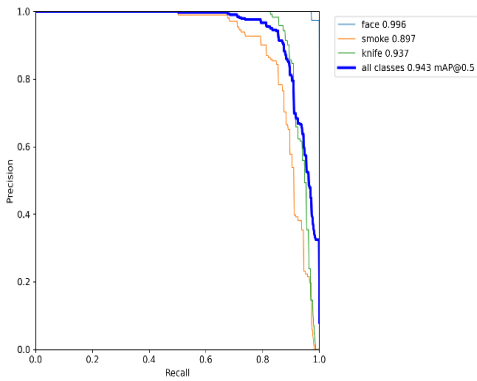
Model	EIoU-Loss	CBAM	DCNv2	mAP/%	Speed/ms
YOLOv7-1	×	×	×	89.8	2.2
YOLOv7-2	√	×	×	91.0	1.8
YOLOv7-3	√	√	×	94.3	2.4
YOLOv7-4	√	√	√	96.4	2.6



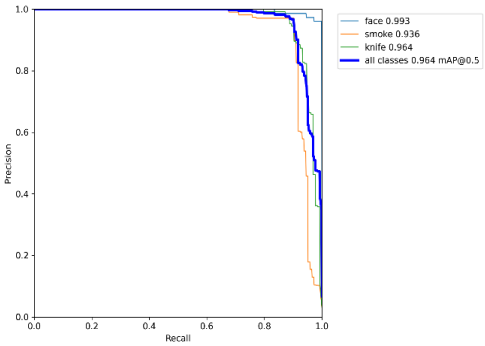
(a) YOLOv7-1



(b) YOLOv7-2



(c) YOLOv7-3



(d) YOLOv7-4

Figure 9. Training process of each model

Figure 10 offers a comparative view of detection effects before and after the improvement. (a) is the detection effect of the original YOLOv7 algorithm, and (b) is the detection effect of the improved YOLOv7 algorithm. It becomes evident that the improved algorithm not only heightens detection accuracy but also mitigates false detection and omissions.

(a) Test results of YOLOv7

(b) Test results of the proposed method

Figure 10. Comparison of effects before and after algorithm improvement

4.4 Comparison of mainstream algorithms

For a comprehensive assessment, the enhanced YOLOv7 model was juxtaposed with established models such as YOLOv3, Fast R-CNN, and YOLOv5s. Table 4 showcases the results, adopting mAP and average detection speed as evaluation metrics. As per the findings in Table 4, the enhanced YOLOv7 model surpasses YOLOv3 and YOLOv5s in terms of detection speed and accuracy. It's noteworthy that while Faster R-CNN exhibits a marginally superior accuracy, its detection speed is considerably languid, rendering it unsuitable for real-time detection requirements.

Table 4. Comparison of mainstream algorithms

Model	mAP/%	Speed/ms
Faster R-CNN	97.72	168.9
YOLOv3	90.07	7.2
YOLOv5S	92.76	5.4
The proposed method	96.4	2.6

5. CONCLUSIONS

To address the prevailing challenges of campus security management, an enhanced target detection algorithm predicated on YOLOv7 has been presented. Distinct modifications were incorporated into the foundational YOLOv7 structure. Firstly, the CIoU loss function was substituted with the EIou loss function, which was observed to expedite convergence speed and elevate detection accuracy. Subsequently, in response to the challenges posed by inconspicuous feature information during the detection of diminutive targets, a CBAM attention mechanism was integrated into the backbone network. Such an inclusion facilitated automatic delineation of pivotal target areas, mitigating distractions from extraneous background elements and augmenting the proficiency and precision of visual data processing. Lastly, the 3×3 convolutions in the initial Elan module of the backbone network were replaced with deformable convolutions, enhancing the algorithm's adaptability to target geometric variations and its learning efficacy.

Through a series of comparative experiments, the effectiveness of these modifications was conclusively validated. Notably, enhancements led to a discernible increase in detection speed, complemented by a marked improvement in accuracy. Looking ahead, to further hone its applicability to the domain of campus security, considerations for future studies include the introduction of a broader array of potential security threats and behaviors to the dataset. Furthermore, optimization of the network structure is envisaged, with an emphasis on streamlining network parameters to achieve a lightweight design. Such refinements aim to build upon the current findings, aspiring to further augment detection precision and speed.

ACKNOWLEDGMENT

This paper is funded by Anhui University collaborative innovation project (Grant No.: GXXT-2022-045). School level key scientific research projects of Hefei Normal University (Grant No.: 2021KJZD12). Project supported by the National Natural Science Foundation of China (Grant No.: 11905247).

REFERENCES

[1] Debauche, O., Elmoulat, M., Mahmoudi, S., Bindelle, J., Lebeau, F. (2021). Farm animals' behaviors and welfare analysis with IA algorithms: A review. Revue d'Intelligence Artificielle, 35(3): 243-253. https://doi.org/10.18280/ria.350308

[2] Mohammed, H., Tannouche, A., Ounejjar, Y. (2022). Weed detection in pea cultivation with the faster RCNN ResNet 50 convolutional neural network. Revue d'Intelligence Artificielle, 36(1): 13-18. https://doi.org/10.18280/ria.360102

[3] Yang, F., Wang, M. (2021). Deep learning-based method for detection of external air conditioner units from street view images. Remote Sensing, 13(18): 3691. https://doi.org/10.3390/rs13183691

[4] Ruan, S., Tang, C., Zhou, X., Jin, Z., Chen, S., Wen, H., Liu, H.B., Tang, D. (2020). Multi-pose face recognition

based on deep learning in unconstrained scene. Applied Sciences, 10(13): 4669. https://doi.org/10.3390/app10134669
[5] Padmanabula, S.S., Puvvada, R.C., Sistla, V., Kolli, V.K.K. (2020). Object detection using stacked YOLOv3. Ingénierie des Systèmes d'Information, 25(5): 691-697. https://doi.org/10.18280/isi.250517
[6] More, B., Bhosale, S. (2023). A comprehensive survey on object detection using deep learning. Revue d'Intelligence Artificielle, 37(2): 407-414. https://doi.org/10.18280/ria.370217
[7] Liang, X., Xiao, H. (2023). Lightweight strip defect real-time detection algorithm based on SDD-YOLO. China Measurement and Test.
[8] Wang, Z. (2022). Automatic and robust hand gesture recognition by SDD features based model matching. Applied Intelligence, 52(10): 11288-11299. https://doi.org/10.1007/s10489-021-02933-y
[9] Luo, Y.T., Jiang, P.F., Duan, C., Zhou, B. (2021). Small object detection oriented improved-RetinaNet model and its application. Computer Science, 48(10): 233-238. https://doi.org/10.11896/jsjkx.200900172
[10] Yang K., Li R., Luo L., Xie, L.M. (2022). Research on train key components detection based on improved RetinaNet. Laser & Optoelectronics Progress, 59(12): 294-301. https://doi.org/10.3788/LOP202259.1215006
[11] Jiao J.F, Jin G.W, Xiong X., Luo, Y.L. (2020). SAR images nearshore ship detection based on RetinaNet algorithm with rotated rectangular box. Journal of Geomatics Science and Technology, 37(6): 603-609. https://doi.org/10.3969/j.issn.1673-6338.2020.06.009
[12] Guo, S., Chai, X.H., Hong, Y. (2022). Security inspection system algorithm based on improved YOLOv5. Journal of Shenyang University (Natural Science), 34(6):453-460,420.
[13] Li, Z., Jiang, X., Shuai, L., Zhang, B., Yang, Y., Mu, J. (2022). A real-time detection algorithm for sweet cherry fruit maturity based on YOLOX in the natural environment. Agronomy, 12(10): 2482. https://doi.org/10.3390/agronomy12102482
[14] Li D.N., Luan J., Mu J.Q. (2023). Cigarette object detection algorithm based on YOLOv5. Software Guide, 22(1): 229-235.
[15] Su, H., Wang, X., Han, T., Wang, Z., Zhao, Z., Zhang, P. (2022). Research on a U-Net bridge crack identification and feature-calculation methods based on a CBAM attention mechanism. Buildings, 12(10): 1561. https://doi.org/10.3390/buildings12101561
[16] Chen, L., Yao, H., Fu, J., Ng, C. T. (2023). The classification and localization of crack using lightweight convolutional neural network with CBAM. Engineering Structures, 275: 115291. https://doi.org/10.1016/j.engstruct.2022.115291
[17] Li, Z., Li, B., Ni, H., Ren, F., Lv, S., Kang, X. (2022). An effective surface defect classification method based on RepVGG with CBAM attention mechanism (RepVGG-CBAM) for aluminum profiles. Metals, 12(11): 1809. https://doi.org/10.3390/met12111809
[18] Gong S.Y., Xu S.J., Zhou L.F., Zhu, J., Zhong, S. (2022). Deformable atrous convolution nearshore SAR small ship detection incorporating mixed attention. Journal of Image and Graphics, 27(12): 3663-3676.
[19] Kee, K.W., Lim, K.H., Lim, C.H., Lim, W.L., Yap, H.E. (2023). Cracks identification using mask region-based denoised deformable convolutional network. Multimedia Tools and Applications, 82(3): 4387-4404.
[20] Yu R.Y., Lin F.Y., Gao N.W., Li, J. (2021). Passenger demand forecast model based on deformable convolution spatial-temporal network. Journal of Software, 32(12): 3839-3851. https://doi.org/10.13328/j.cnki.jos.006115
[21] Gallo, I., Rehman, A.U., Dehkordi, R.H., Landro, N., La Grassa, R., Boschetti, M. (2023). Deep object detection of crop weeds: Performance of YOLOv7 on a real case dataset from UAV images. Remote Sensing, 15(2): 539. https://doi.org/10.3390/rs15020539
[22] Zhao, Y.L., Shan, Y.G., Yuan, J. (2023). Wearing mask pedestrian tracking based on improved YOLOv7 and DeepSORT. Computer Engineering and Applications, 59(6): 221-230. https://doi.org/10.3778/j.issn.1002-8331.2210-0479
[23] Chen, C., Wu, B., Zhang, H. (2023). An image recognition technology based on deformable and CBAM Convolution Resnet50. IAENG International Journal of Computer Science, 50(1): 1-8.
[24] Xue, J., Cheng, F., Li, Y., Song, Y., Mao, T. (2022). Detection of farmland obstacles based on an improved YOLOv5s algorithm by using CIoU and anchor box scale clustering. Sensors, 22(5): 1790. https://doi.org/10.3390/s22051790
[25] Zhang, G., Du, Z., Lu, W., Meng, X. (2022). Dense pedestrian detection based on YOLO-V4 network reconstruction and CIoU loss optimization. In Journal of Physics: Conference Series, 2171(1): 012019. https://doi.org/10.1088/1742-6596/2171/1/012019