2022 IEEE 6th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC) | 978-1-6654-5864-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/IAEAC54830.2022.9929889

Pedestrian fall detection based on AC-YOLOv5s

Guoxin Shen[1], Ziqin Wei[2], Xuerong Li[3], Yi Wei[1], Ke Li[4]

1.School of Automation, Wuhan University of Technology, Wuhan, China
2.School of Electrical Engineering, Zhengzhou University, Zhengzhou, China
3.College of mechanical and electrical engineering, Shanghai Donghai Vocational and Technical College, Shanghai, China
4.School of Food Engineering, Qingdao Institute of Technology, Qingdao, China
1523900304@qq.com, 705111765@qq.com, 793184137@qq.com, weiyi@whut.edu.cn, 1173586824@qq.com
Corresponding Author: Xuerong Li 793184137@qq.com

Abstract—**In view of the serious occlusion phenomenon of pedestrian fall detection, the difficulty of extracting small target details, and the slow detection speed, this paper proposes a high-precision lightweight detection network AC-YOLOv5s. First, the convolution module in the backbone is replaced by ACBConv, and the C3 module is replaced by ACBC3 to improve the detailed feature extraction capability. Secondly, a small target detection layer is added to the feature fusion network (FPN) to improve the detection accuracy of small targets. Finally, use Alpha IoU loss replaces CIoU loss to improve the loss and regression accuracy of the High IoU target. Finally, compared with the original YOLOv5s, the network in this paper improves the mAP by 2.33%, and the FPS reaches 21 during detection. The experimental results show that our network achieves better results than other networks.**

Keywords—ACBConv; ACBC3; YOLOv5s; Alpha IoU loss

I. Introduction

The study of pedestrian fall detection in the context of open field of view with high density of human traffic has great difficulties [1]. Because there are more overlapping parts between people in this scenario, the model detection is very prone to inaccurate detection due to unclear boundaries of the fall target. Therefore, it is of great interest to study lightweight fall detection models.

In recent years, target detection models based on deep learning have been widely used in varIoUs aspects. Such as SSD [2], R-CNN [3], Fast RCNN [4], Faster RCNN [5], YOLO series [6-9], and other target detection models are widely used in unmanned vehicles, automatic navigation, pose detection, and so on. Target detection algorithms are classified into one-stage and two-stage according to the detection steps. two-stage algorithms are typically represented by R-CNN, Fast RCNN, and Faster RCNN, which is an algorithm that divides detection into two steps, first dividing the region of interest on the feature map, and then predicting the feature map by the model, each step needs to be done separately, which will lead to slow detection speed. The one-stage algorithm is a direct target detection by regression, which will improve the detection speed, but requires some means to improve the detection accuracy, the typical representatives are YOLO series and SSD algorithm. In this paper, we decided to use YOLOv5s as the base detection network after weighing the detection accuracy and speed. To solve the problems of unbalanced accuracy and computational cost and insufficient model generalization of YOLO series target detection algorithms, Xiuyi Zhang et al.[10] proposed YOLO-Day Night and Fast (YOLO-DNF), a high-precision and fast vehicle and pedestrian detection model that can meet the target detection needs under different lighting scenes. To address the problems of low accuracy of nighttime small-target infrared pedestrian detection in assisted driving, the large memory space occupied by the network model, and the difficulty of the detection speed to meet the real-time detection requirements, Zifen He et al.[11] proposed a lightweight nighttime infrared image pedestrian detection neural network YOLO-Person. Pedestrian targets in subway scenes have problems such as different sizes, different degrees of occlusion and blurred targets caused by too dark environment, which largely affect the accuracy of pedestrian target detection. To address the above problems, Xiu-Zai Zhang et al.[12] proposed an improved YOLOv5s target detection algorithm to enhance the detection of pedestrian targets in subway scenes. In dense pedestrian detection scenarios, mutual occlusion overlap between targets causes degradation in the detection performance of the YOLOv3 model. Xiang Li et al.[13] proposed a clustering loss function to make the prediction frames belonging to the same target more compact by optimizing the variance and mean of the prediction frame coordinates, and thus reduce the false positive rate. To address the problems of low accuracy and low recall in pedestrian detection with the Tiny YOLOv4 target detection algorithm, Xuan-Yong et al.[14] improve the feature extraction network and the prediction network. In the feature extraction network, a depth-separable convolutional network is used instead of the traditional convolutional network, and an attention mechanism module is added to the feature extraction network to enhance the region of interest of the detected target.

Although the existing detection algorithms have made some progress in the pedestrian fall detection problem, the

978-1-6654-5864-1/22/$31.00 ©2022 IEEE

detailed feature extraction for small objects and the multiple occlusion problems are still not well solved. In addition, in public scenarios, the pedestrian fall detection algorithm should take into account both accuracy and real-time performance. In response to the above problems, this paper proposes a lightweight AC-YOLOv5s network, which includes: (1) Replace the convolution module in the backbone with ACBConv, and replace the C3 module with ACBC3 to improve the ability to extract details. The implementation of ACBConv is by using three convolution fusions with convolution kernel sizes of 3*3, 1*3 and 3*1, so it can also cleverly solve the problem of multiple occlusions without increasing the computational overhead, ensuring the detection accuracy. real-time. (2) A small target detection layer is added to the feature fusion network (FPN) to improve the detection accuracy of small targets. (3) Use Alpha IoU loss instead of CIoU loss to improve the loss and regression accuracy of the High IoU target.

II. THIS PAPER WORKS

A. *The network structure of this paper*

The overall network structure of this paper consists of three parts: backbone, neck (FPN), and head. In the backbone part, a 3*3conv, 1*3conv and 3*1conv parameters are fused into a convolution kernel parameter, and then activated with the silu activation function. The above constitutes the ACBConv module of this article; two ordinary convolutions and shortcuts are composed Bottleneck, multiple ACBConv and Bottleneck form ACBC3. In the feature fusion network (neck) part, this paper adds a small target detection layer, that is, a feature fusion network containing 4 layers, and the size of the feature map is 20*20, 40*40, 80*80, 160*160. No changes are made to the Head section. The specific structure diagram is shown in the following figure:

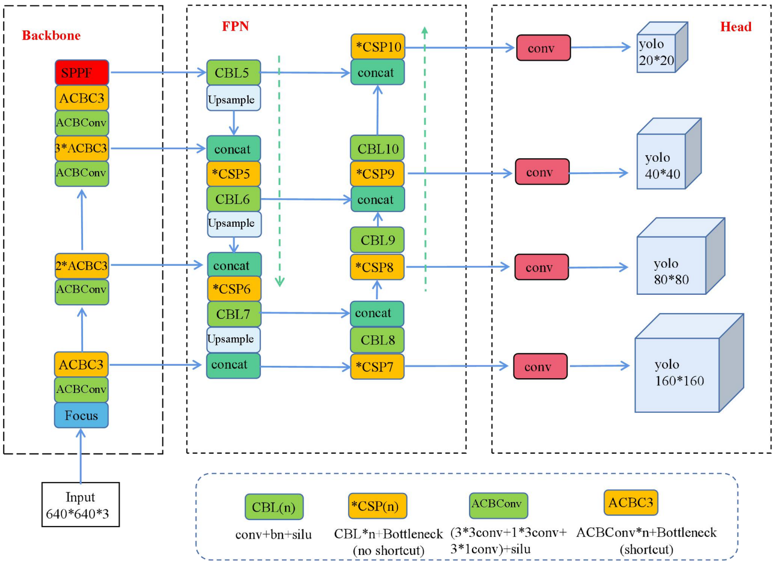

Fig. 1. Network Framework Diagram.

B. *ACBConv and ACBC3*

Xiaohan Ding [15] et al. proposed a new convolution structure, Asymmetric Convolution, which performs convolution expansion during training. During the training process, three convolutions with different convolutional kernel sizes are used for parallel training, and the parameters of the three convolutions are superimposed during the inference phase. The ACBConv in this article adopts this idea, combining the three convolutions of 1*3, 3*1 and 3*3 into one convolution for training. After the training, parameter fusion is performed. This part is mainly for the three volumes. The nuclei are fused. In the implementation process of this part, the fused convolution kernel parameters are used to initialize the existing network. Therefore, in the inference stage, the network structure is exactly the same as the original network, but the network parameters use the ones with stronger feature extraction ability. The parameters are the fused convolution kernel parameters, so there will be no increase in the amount of computation in the inference stage. So this is also a way to lighten the network, bringing improvements without increasing inference overhead. The ACBC3 module in this article is to replace the convolution in the original C3 module with ACBConv. It should be noted here that ACBConv can only be effective by replacing the 3*3 convolution, and cannot replace the 1*1 convolution.

C. *Small target detection layer*

Among the detection problems, small object detection is usually a difficult problem to solve. In this paper, a more common way is adopted, which is to increase the detection layer. Small objects will lose their texture features as the number of convolutions increases, so this paper adds a detection layer with a feature map size of 160*160 to improve the detection effect of small objects.

D. IoU loss improvement

1) Alpha IoU loss

CIoU and DIoU [17] are treated equally for each IoU target, so high_IoU target cannot achieve high-precision regression. Therefore, this paper uses Alpha IoU Loss [18] instead of CIoU loss. Alpha IoU Loss can adaptively weight the target loss and gradient according to the size of IoU, which is beneficial to improve the regression accuracy of the High_IoU target. Alpha IoU Loss is defined as:

$$l_{\alpha-IoU}=\frac{1-IoU^{\alpha}}{\alpha},\alpha>0 \tag{6}$$

Where, α is an adjustable parameter, when $\alpha\rightarrow 0$, $l_{\alpha-IoU}=-\log(IoU)$,when $\alpha\nrightarrow 0$, $l_{\alpha-IoU}=1-IoU^{\alpha}$ .If the penalty term is added to the above formula, $\alpha-IoU$ can be extended to a more general form：

$$l_{\alpha-IoU}=1-IoU^{\alpha_1}+p^{\alpha_2}(B,B^{gt}) \tag{7}$$

In this case, GIoU, DIoU, CIoU can be represented by compression by $\alpha-IoU$. At the same time, it has an important property, Since α is an adjustable exponential parameter, Then when α is greater than 1, It will magnify the loss weight of the High_IoU target, which will help the detector to pay more attention to the High_IoU target, that is, it is more sensitive to the High_IoU loss target. In order to improve the regression accuracy of the High_IoU target, this paper uses Alpha IoU Loss instead of CIoU loss, experimentally tested, α equal 3 gives the best results.

III. Experiments

The deep learning environment and framework used in this experiment are shown in Table 1, and are applied to other contrastive networks with the same configuration.

TABLE I. EXPERIMENTAL ENVIRONMENT CONFIGURATION TABLE

Configuration name	**Configuration parameters**
Operating system	Windows11
CPU	AMD RYZEN7-5800H
GPU	Nvidia GeForce RTX 3060 Laptop 6GB
RAM	16GB
Software	Anaconda、Pycharm
Deep Learning Framework	Pytorch
GPU acceleration library	CUDA11.3

The detection target in this paper is a specific target, so the data set used in this experiment is a data set we independently produced. The data set has a total of 8000 pictures, of which 4800 are used as training set, 800 are used for validation set, and 2400 are used for testing. Set, the ratio is 6:1:3. Mosaic data augmentation was used before training began.

A. Network Hyperparameter Settings

First, the size of the image input by the network is uniformly resized to 640*640, the bacthsize is set to 8, and the stochastic gradient descent method is used as the optimizer of the network in this paper. The initial learning rate lr0 is 0.01, the cyclic learning rate lrf is 0.1, the learning rate momentum is 0.937, the weight_decay is 0.0005, the warmup_epochs is 3.0, the warmup momentum is 0.8, the IoU_t is 0.40, the probability of Mosaic is 1.0, the fl_gamma: is 0.0, and the obj_pw is 1.0, the IoU loss weights of the small, medium, and large feature layers are set to 4.0:1.0:0.4.

B. Experimental results

In this paper, we first compare the proposed AC-YOLOv5s algorithm with other existing algorithms in the validation set. The detection result is considered as True positive (TP) if the target category is correctly detected and the frame center coordinates and frame dimension are within a certain range, False positive (FP) if the target category is incorrectly identified or the frame is not within the set threshold, and False negatives (FN) if the number of incorrectly classified negative samples.

$$P=\frac{TP}{TP+FP},R=\frac{TP}{TP+FN} \tag{8}$$

$$\text{AP}=\int_0^1 PdR,\text{mAP}=\frac{1}{N}\sum_{i=1}^{N}AP_i \tag{9}$$

The meaning of AP is the area enclosed by the P-R curve. Keeping the hyperparameters and operating environment unchanged, this paper uses YOLOv3-SPP, YOLOv4, YOLOv5s, EfficientNet-YOLOv5s and AC-YOLOv5s in this paper for training, and the mAP0.5 drawing curve of the training result is shown in the figure.

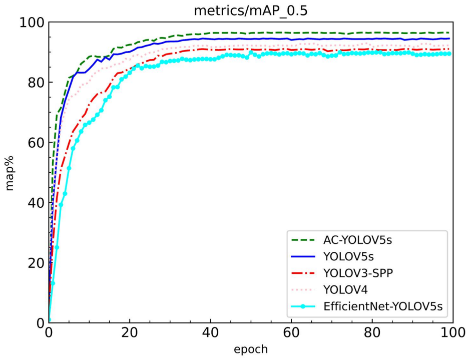

Fig. 2. mAP curve

As shown in the figure, our network achieves better results than other networks. The original YOLOv5s achieved a mAP value of 95.23%, and the AC-YOLOv5s in this paper achieved a mAP of 97.56, which was 2.33% higher than the original YOLOv5s, and the convergence speed was faster during the training process. Compared with YOLOv3-SPP, YOLOv4, EfficientNet-YOLOv5s and other networks, there are different degrees of improvement, and the model parameters are smaller, and the FPS during detection reaches 21.

We compared the results of the above model with those of the model in this paper, using P, R, mAP0.5, and the number of parameters, respectively, and the results of the comparison are shown in the following table. From the table, we can see that the number of parameters of the model in this paper has almost no increase compared to the original YOLOv5, but the accuracy has increased.

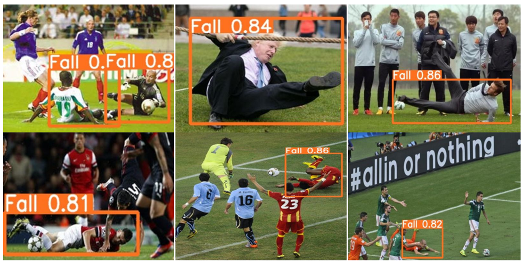

Fig. 3. Test result graph

TABLE II. COMPARISON OF RESULTS

Mdel	P(%)	R(%)	mAP0.5(%)	size	Params(M)
YOLOv3-SPP	**87.12**	**89.31**	**88.52**	**640**	**61.53**
YOLOv4	**91.43**	**91.84**	**94.57**	**640**	**52.5**
YOLOv5s	**91.46**	**92.16**	**95.23**	**640**	**7.02**
EfficientNet-YOLOv5s	**85.17**	**88.45**	**87.75**	**640**	**3.02**
AC-YOLOv5s	**92.56**	**93.87**	**97.56**	**640**	**7.67**

In order to compare the influence of different improvements, this paper does ablation experiments for comparison. The specific results are shown in the table below.

TABLE III. ABLATION EXPERIMENT TABLE

Num	ACBConv and ACBC3	Small target detection layer	Alpha IoU	mAP0.5(%)	Params(M)
1	✓	×	×	96.68%	7.03
2	×	✓	×	96.15%	7.67
3	×	×	✓	95.64%	7.03
4	×	×	×	95.23%	7.03

It can be seen from the experimental results that ACBConv and ACBC3 are the key to the improvement of the network model in this paper. When they are used alone, they can bring a 1.55% increase in mAP to the network, and the number of parameters does not increase compared to the original YOLOv5s, which is beneficial on its design structure. The detection results of AC-YOLOv5s are shown in the figure below.

REFERENCES

[1] Shi Xin, Lu Hao, Qin Pengjie, Leng Zhengli. A long-distance pedestrian small target detection method [J/OL]. Journal of Instrumentation: 1-11 [2022-0619]. http://kns.cnki.net/kcms/detail/11.2179.TH.20220601.1329.008.html.

[2] LIU W, Dragomir Anguelov2 and Dumitru Erhan. SSD: Single Shot MultiBox Detector[C]//Proceedings of European Conference on Computer Vision. Amsterdam: Springer International Publishing. 2016: 21-37.

[3] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE. 2014: 580-587.

[4] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE Conference on International Conference on Computer Vision. Boston: IEEE, 2015: 1440-1448.

[5] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards realtime object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis \& Machine Intelligence, 2017, 39(6): 1137-1149.

[6] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, realtime object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.

[7] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.

[8] Redmon J, Farhadi A. YOLOv3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.

[9] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.

[10] Zhang Xiu Yi, Chen Chang Xing, Du Juan, Li Jia, Cheng Kuan Hong. Vehicle and pedestrian detection model YOLO-DNF for assisted driving [J/OL]. Computer Engineering and Applications:1-15 [2022-08-09]. http://kns.cnki.net/kcms/detail/11.2127.TP.20220726.1607.010.html.

[11] He Zifen, Chen Guangchen, Chen Junsong, Zhang Yinhui. Multi-scale feature fusion for lightweight nighttime infrared pedestrian detection in real time [J/OL]. China Laser:1-18[2022-08-09]. http://kns.cnki.net/kcms/detail/31.1339.TN.20220713.1838.044.html.

[12] Zhang Xiu Zai, Qiu Ye, Zhang Chen. Improved YOLOv5s algorithm for pedestrian target detection in subway scenes[J/OL]. Advances in Lasers and Optoelectronics:1-20[2022-08-09]. http://kns.cnki.net/kcms/detail/31.1690.TN.20220713.1944.609.html.

[13] Li Xiang, He Miao, Luo Haibo. An improved YOLOv3 algorithm for occlusion-oriented pedestrian detection[J]. Journal of Optics,2022,42(14):160-169.

[14] Zheng Ge, Jianfeng Wang, Xin Huang, Songtao Liu, Osamu Yoshie.LLA: Loss-aware label assignment for dense pedestrian detection[J]. Neurocomputing, Volume 462, 2021, Pages 272-281, ISSN 0925-2312.

[15] Xiaohan Ding, Yuchen Guo, Guiguang Ding, et al.ACNet: Strengthening the Kernel Skeletons for Powerful CNN via Asymmetric Convolution Blocks[C]//Proceedings of the IEEE Conference on International Conference on Computer Vision. Seoul: IEEE, 2019: pp. 1911-1920.

[16] Hamid Rezatofighi, Nathan Tsoi and JunYoung Gwak. Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 658-666.

[17] ZHENG Z H, WANG P and LIU W. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression [C]//Proceedings of the AAAI Conference on Artificial Intelligence. New York: Springer International Publishing. 2020:1-7.

[18] HE J B, Sarah Erfani and MA X J. Alpha-IoU: A Family of Power Intersection over Union Losses for Bounding Box Regression[C]//Proceedings of the Conference and Workshop on Neural Information Processing Systems. NeurIPS, 2021: 1-10.