Transforming Limitations into Advantages: Improving Small Object Detection Accuracy with SC-AttentionIoU Loss Function

Mingle Zhou[1,2], Changle Yi[1,2], Min Li[1,2], Honglin Wan[3], Gang Li[1,2](✉), and Delong Han[1,2]

[1] Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center (National Supercomputer Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences), Jinan, China
lig@qlu.edu.cn

[2] Shandong Provincial Key Laboratory of Computer Networks, Shandong Fundamental Research Center for Computer Science, Jinan, China

[3] College of Physics and Electronic Science, Shandong Normal University, Jinan, China

Abstract. Small object detection is widely used in industries, military, autonomous driving and other fields. However, the accuracy of existing detection models in small object detection needs to be improved. This paper proposes the SC-AttentionIoU loss function to stress the issue. Due to the less features of small objects, SC-AttentionIoU introduces attention within the true bounding box, allowing the existing detection models to focus on the critical features of small objects. Besides, considering attention perhaps ignore non-critical features, SC-AttentionIoU proposes an adjustment factor to balance the critical and non-critical feature areas. Using the YOLOv5s model as a baseline, compared with the widely used CIoU, SC-AttentionIoU achieved an average improvement of 1% in mAP@.5 on the SSDD dataset and an average improvement of 1.47% in mAP@.5 on the PCB dataset in this experiment.

Keywords: Object detection · Small objects · Loss function · Attention

1 Introduction

The object detection has been developed significantly with the emergence of a large number of models such as Swin-Transformer [8] and DAMO-YOLO [12]. Small object detection is important in many fields such as industry, military, and autonomous driving. However, small objects themselves have few features while detection scenes often require high accuracy and real-time performance. Therefore how to balance them is a challenge to researchers [4].

L. Iliadis et al. (Eds.): ICANN 2023, LNCS 14260, pp. 221–232, 2023.
https://doi.org/10.1007/978-3-031-44195-0_19

There are some models such as Next-ViT [7], PPYOLOE [11] and PP-PicoDet [14], have been proposed with high detection accuracy and lightweight weights. The existing improvement of loss functions has focused on more accurate calculation of the deviation between predicted and ground-truth bounding boxes. There is few work on improving loss functions based on the characteristics of detection targets. Existing loss functions include IoU [15], GIoU [10], DIoU [19], CIoU [19], EIoU [17], and Alpha-IoU [5]. Researchers have explored factors such as intersection over union, center point distance, aspect ratio, and orientation, to make the loss function more accurate. By providing more accurate guidance during training, the detection model can improve its performance.

In this paper, the novel attention-based loss function called SC-AttentionIoU is presented which guides the network to pay more attention to the key regions of the detection target. This loss function identifies the spatial region with significant features of the detection target and incorporates this prior knowledge. Loss function attention determines the weight of the internal ground-truth bounding box. High weight is assigned to the key feature regions of small targets, while low weight is assigned to irrelevant or defective regions. In summary, this paper presents the following innovations:

1. The attention-based loss function SC-AttentionIoU is proposed. It incorporates attention into the loss calculation and guides the model to focus on the key features of small objects to improve detection accuracy.
2. A novel weight generation strategy is proposed in SC-AttentionIoU. Different weight regions are generated within the predicted box based on the distribution characteristics of small objects, which calculates the loss value of the SC-AttentionIoU more accurately.
3. The factor is proposed to address the potential problems caused by introducing attention into the loss function. This factor allows the model to not only focus on key features but also consider other regions during training, resulting in more accurate prediction box locations.

This paper is organized as follows. Section 2 provides a brief review of related methods for loss functions. Section 3 describes the proposed methods. Section 4 presents experimental results, and Sect. 5 provides conclusions.

2 Related Work

2.1 Small Object Detection

Small object detection has always been a challenging problem. The definition of small objects usually refers to objects with a size in the image that is very small, even less than 10 pixels. In this case, small objects are often affected by various factors such as image blur, noise, and scale variation, making it difficult to accurately detect and recognize them. To address the problem of small object detection, Facebook AI proposed a Transformer-based small object detection method called DETR [2], which uses deformable convolution and feature pyramid techniques to effectively solve the problem of object size and scale variation in small object detection. Many researchers have also attempted to improve small object detection by improving network structures and feature extraction methods. Yang et al. combined spatial attention, channel attention, and self attention to improve the detection effect of small objects in intelligent transportation [13]. The model has faster convergence speed and higher accuracy. Zhao et al. proposed a detection model [18] based on feature fusion and anchor point improvement, which reduces the missed detection rate of small targets in complex backgrounds and has considerable detection speed.

2.2 Loss Funcation

The most important in the loss function for object detection is the bounding box regression which is used to calculate the offset from the detection box to the ground-truth box. IoU [15] was first proposed to calculate the intersection-over-union ratio between the detection box and the ground-truth box. GIoU [10] used the minimum enclosing rectangle to calculate the offset between the predicted box and the ground-truth box based on IoU. DIoU [19] replaced the minimum enclosing rectangle with the Euclidean distance between the center points, which more accurately calculates the distance between the predicted box and the ground-truth box. CIoU [19] adds an aspect ratio penalty to DIoU to solve the problem of IoU being insensitive to shape. EIoU [17] further improved the aspect ratio by using the ratio of the lengths of the width and height sides, resulting in more accurate results. SIoU discovered the influence of the direction between the detection box and the ground-truth box.

3 Method

The SC-AttentionIoU proposed in this paper is focused on the coarseness of object detection annotation and the lack of features in small objects. It is well known that in annotated images, the detected object does not always occupy the entire area of the bounding box [9]. There are often other unrelated regions such as other objects or backgrounds within the bounding box, as shown in Fig. 1. These unrelated regions interfere with the model training. Since small objects contain fewer features, it is feasible to improve detection accuracy by focusing on the key features of small objects. To reduce the impact of unrelated regions, this paper introduces attention within the bounding box to assign different regions within the bounding box with different weights. This allows the model to focus more on regions with higher weights during bounding box regression, and reduces the interference of unrelated regions on the model. However, if the model focuses too much on the key features, it may generate predicted boxes that are closer to the distribution of key features but do not match the ground truth boxes. To avoid this issue, this paper further improves AttentionIoU and proposes SC-AttentionIoU.

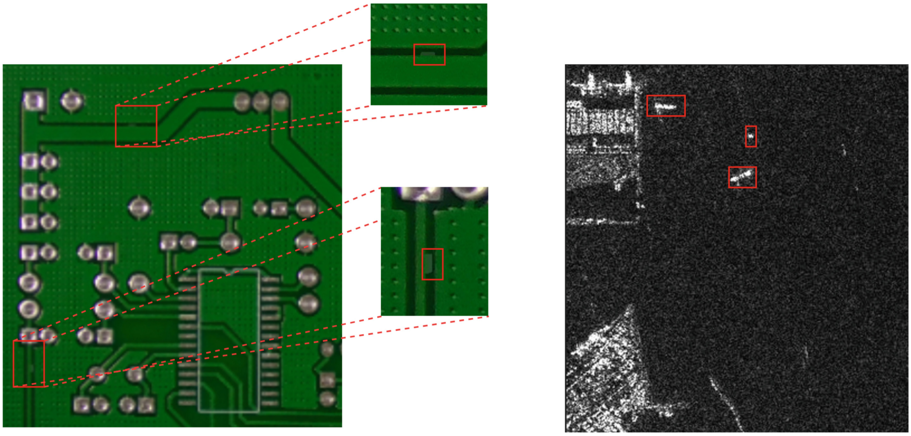

Fig. 1. PCB data set and SSDD data set small object data set picture.

3.1 AttentionIoU

In IoU calculations, the weights of the internal regions of the predicted and detected boxes were averaged. When the overlap area remains constant and the sizes of the detection box and true box are consistent, predicted boxes of varying shapes have an identical intersection over union ratio to the true box. This result is insensitive to shape variations. In the initial training stage, the averaging of the weights of the internal regions of the true box leads to equal attention being given by the model to both irrelevant regions and critical features within the true box, which is disadvantageous for model training.

When the target itself contains a large number of features, it is difficult to select the key features, as the importance of the detection target's features for model training cannot be determined based on prior knowledge. Conversely, small targets with fewer features can be more easily identified with key features, thus allowing the model to focus on critical features and improve detection accuracy. As shown in Fig. 1, it is evident that the key features are located at the center position of the true box.

When applying weights to the key features of detection targets, the attention-based network [3,6] is used to add pixel-level weight annotations to the object within the true box. However, the approaches greatly increases the annotation cost. Therefore, this paper combines the distribution of the key feature regions to adjust the true box and generate a novel weight box. The weight box is located inside the true box and is closer to the distribution of key feature regions. The weight box annotates the critical features of small targets to match high-weight areas. The region between the weight box and the true box is referred to as the low-weight area, primarily comprising irrelevant regions, as depicted in Fig. 2.

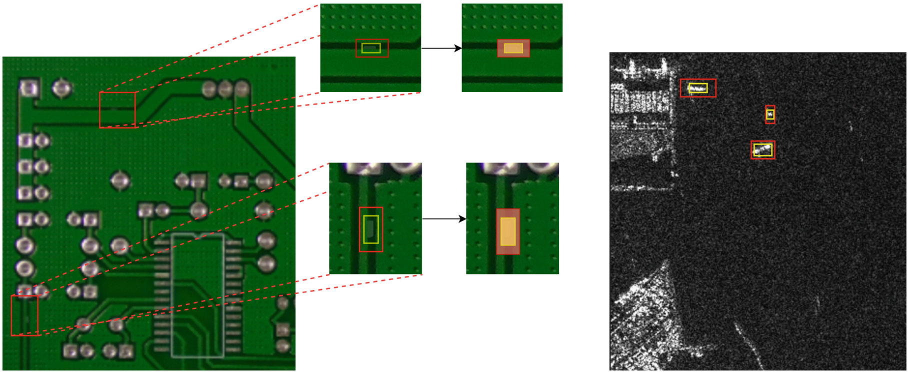

Fig. 2. The red box is the annotation box, and the yellow box is the weight box. (Color figure online)

The formula of AttentionIoU is

$$AttentionIoU = \frac{(\theta \cap \rho) + \varepsilon(\exists \cap \rho)}{(\theta + \rho + \varepsilon^{*}\exists) - (\theta \cap \varphi)} \tag{1}$$

where ρ denotes a prediction box, θ denotes a real box, $\exists$ denotes a weight box, and ε denotes the weight factor of the weight box that ε in [1,5]

The weight box is located inside the true box, and the intersection between the weight box and the predicted box is contained within the intersection between the true box and the predicted box. When calculating IoU, the intersection of the internal regions of the weight box is increased by a factor $1{+}\varepsilon$. This increases the weight of the weight box region. The formula to generate the coordinates of the weight box is

$$\exists_{xi} = x_i^{gt} + (-1)^{i+1} * \frac{w^{gt}}{\tau} \tag{2}$$

$$\exists_{yi} = y_i^{gt} + (-1)^{i+1} * \frac{h^{gt}}{\mu} \tag{3}$$

where τ is the generation factor for the x-coordinate of the weight box. μ is the generation factor for the y-coordinate of the weight box. Both generation factors are based on prior knowledge of the detection targets. $\exists_{xi}$ denotes the x-coordinates of the top-left and bottom-right corners of the weight box, $\exists_{yi}$ denotes the y-coordinates of the top-left and bottom-right corners of the weight box, (x_1, y_1) and (x_2, y_2) are the top-left and bottom-right coordinates of the true box, respectively, w^{gt} is the width of the true box, and h^{gt} is the height of the true box. x_i^{gt} denotes the x-coordinate of the true box. The variable y_i^{gt} represents the y-coordinate of the true box, where i lies within the range of [1, 2].

3.2 SC-AttentionIoU

Adding a weight box inside the real box allows the model to focus on the key features of the small object as much as possible. However, this will also make the detection box more closely related to the size and shape of the weight box. If the difference between the weight box and the real box is large, it will affect the detection accuracy. To solve this problem, this paper uses aspect ratio as the weight of AttentionIoU and proposes SC-AttentionIoU. The formula for SC-AttentionIoU is

$$SC - AttentionIoU = 1 - \partial * AttentionIoU + \frac{\rho^2(b, b^{gt})}{c^2} \tag{4}$$

where ∂ is the weight calculated based on aspect ratio. After using AttentionIoU, the IOU value becomes sensitive to the shape of the predicted box. In exploring the aspect ratio of the predicted box, CIoU and EIoU respectively explored two ways of calculating the aspect ratio penalty term. However, since AttentionIoU itself is sensitive to shape, the previous approach of separating the IOU from the aspect ratio factor is no longer feasible. This paper proposes to use aspect ratio as the weight for IoU value, and combine it with AttentionIoU for calculation. The formula for ∂ is

$$\partial = 1 - \frac{\left(\frac{\rho^2(w,w^{gt}))}{(w^c)^2} + \frac{\rho^2(h,h^{gt}))}{(h^c)^2}\right)}{T} \tag{5}$$

where ∂ cannot be 0. $T \geq 2$ is a regulation factor for the value range of ∂. In this paper, the value range of ∂ is set to [(T−2)/T,1]. w^{gt} denotes the width of the annotated box, h^{gt} denotes the height of the annotated box. w^c is the width of the bounding rectangle between the real box and the predicted box. h^c is the height of the bounding rectangle between the real box and the predicted box.

SC-AttentionIoU already takes into account the effect of aspect ratio, so no aspect ratio penalty is added in SC-AttentionIoU. In CIoU when IoU is 0, the loss is guided by Euclidean distance instead of the aspect ratio. The proposal in loss function can speed up the convergence of the model training. The comparison of different loss functions and the proposed loss function on the PCB dataset are shown in Fig. 3. The results show that SC-AttentionIoU converges faster and has a lower loss convergence value.

4 Experiment

4.1 Experiment Setting

In this paper, the YOLOv5 was used as the baseline and compared with existing loss functions on two datasets: ship remote sensing dataset [16] released by the Department of Electronic and Information Engineering of the Naval Aeronautical and Astronautical University, and printed circuit board defect dataset (PCB dateset) released by the Intelligent Robot Open Laboratory of Peking University.

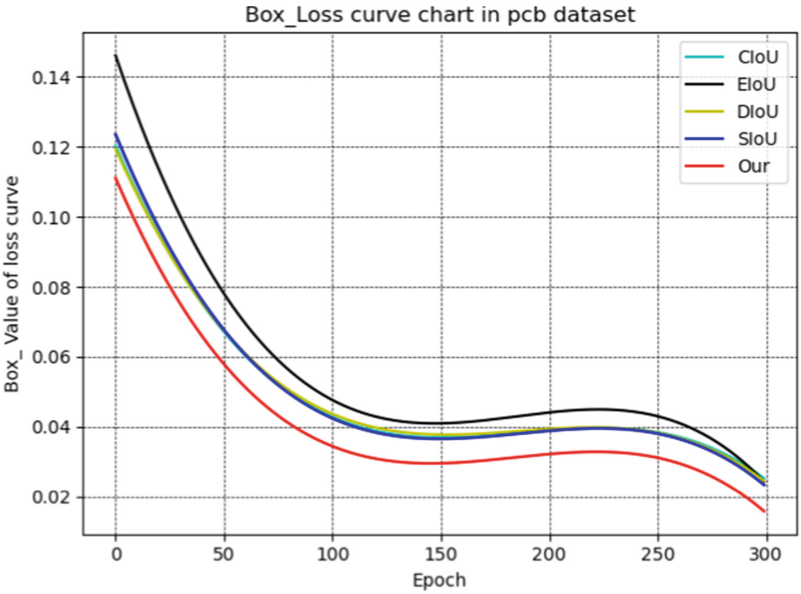

Fig. 3. Boundary box regression Loss function image.

The former dateset consists mostly of small targets. The latter dateset contains 1,386 images with six types of small defect targets.

In training, the image size is adjusted to 640 × 640, and data augmentation techniques such as mosaic [1], translation, and flipping are used. The optimizer used is SGD, and the learning rate follows a cosine annealing scheme with an initial value of 0.01. Results were obtained after running 300 epochs on a Tesla A100-SXM4 GPU.

4.2 Experimental Indicators

According to the needs of small target detection scenarios, this article selects precision, recall, and mAP as detection indicators. Precision mainly measures the accuracy of the model in identifying targets. The recall rate mainly measures the model's ability to identify all true targets. mAP reflects the comprehensive level of accuracy and recall rate. mAP@0.5 primarily focuses on the average precision of the model at relatively lenient thresholds, while mAP@0.5–0.95 specifically focuses on the average precision of the model at stricter thresholds.

4.3 Comparative and Ablation Experiments

Table 1 shows the experimental results of SC-AttentionIoU and other loss functions on the SSDD [16] dataset. This loss function is designed based on the characteristics of small objects. The SSDD dataset has a simple background,

Table 1. SSDD dataset loss function comparative test

Method	Precision	Recall	mAP@.5	mAP@.5-.95
DIoU	95.4%	90%	94.6%	59.4%
CIoU	96%	90%	94.1%	58.5%
EIoU	96%	90%	94%	59.4%
SIoU	93.6%	90.4%	94.8%	59.3%
Our	**95.5%**	**92%**	**95.3%**	**61.8%**

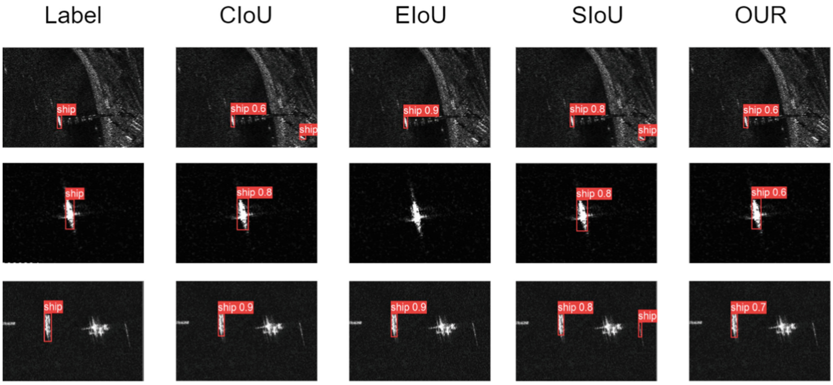

Fig. 4. Comparison of partial detection results with different loss functions.

few categories, and contains a small number of relatively large objects. The experiment results show that the mAP@.5 score of this loss function exceeds DIoU by 0.7%, CIoU by 1.2%, EIoU by 1.3%, and SIoU by 0.5%. It demonstrate that this loss function can maintain effectiveness in the presence of relatively large objects interference. The test results are shown in Fig. 4.

Table 2 shows the experiments of SC-AttentionIoU and other loss functions on the PCB defect dataset in this paper. The PCB defect dataset has more types of defects and various defect shapes compared to SSDD, which verifies the improvement effect of the proposed loss function in complex backgrounds. The experimental results show that the proposed loss function outperforms DIoU by 6.6%, CIoU by 4.5%, EIoU by 6.2%, and SIoU by 3.6% in terms of recall rate. This demonstrates that the proposed loss function can improve the detection ability of small targets in complex backgrounds. In terms of mAP@.5-.95, the proposed loss function is 1.85% higher than the average level of other loss functions. Therefore, the proposed loss function can effectively improve the detection ability of the model on small target datasets. The test results are shown in Fig. 5.

Table 3 shows the experiments of SC-AttentionIoU and other detection models on the PCB defect dataset in this paper. We compare the advanced models such as PPYOLOE, YOLOv6, PP-PicoDet, combined with SC-AttentionIoU

Table 2. PCB dataset loss function comparative test

Method	Precision	Recall	mAP@.5	mAP@.5-.95
DIoU	97.2%	83.4%	92%	49.4%
CIoU	88.1%	85.5%	90%	49.1%
EIoU	96.5%	83.8%	91.4%	49.5%
SIoU	93.1%	86.4%	89.9%	49.4%
Our	**91%**	**90%**	**92.3%**	**51.1%**

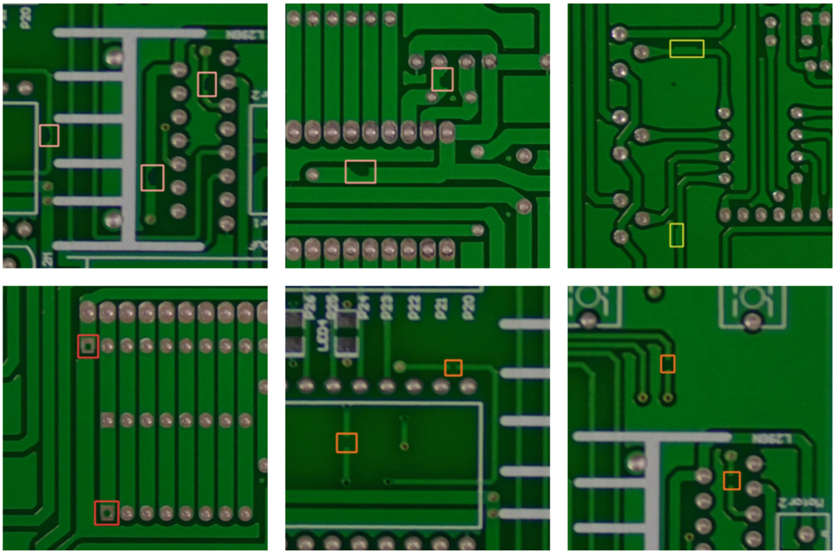

Fig. 5. PCB Dataset Detection Results.

proposed in this paper with the widely used CIoU to verify the robustness of SC-AttentionIoU in different detection models. * indicates the use of the proposed loss function in this paper. The experimental results show that after being combined with SC-AttentionIoU, the YOLOv6 model improves the mAP@.5-.95 score by 1.4%, the PPYOLOE-s model improves by 4.3%, and the PP-PicoDet model improves by 2.6%. Therefore, the proposed loss function has strong robustness on different detection models.

Table 4 shows the experiments of AttentionIoU and other Loss Funcation on the SSDD ship dataset in this paper. We combine the widely used CIoU, EIoU, SIoU, and attentionIoU and replace the IoU in the loss function with attentionIoU to validate the effectiveness and robustness of the proposed attentionIoU idea. * denotes the use of attentionIoU proposed in this paper. Experimental results show that when combined with attentionIoU, CIoU* improves MAP@.5-.95 by 2% compared to CIoU, EIoU* improves MAP@.5-.95 by 2% compared to EIoU, and SIoU* improves MAP@.5-.95 by 2% compared to SIoU. Therefore,

Table 3. Ablation experiments of SC-AttentionIoU on different detection models

Method	Precision	Recall	mAP@.5	mAP@.5-.95
YOLOv6	95.4%	83.6%	86.4%	45.4%
YOLOv6*	**90.7%**	**85.9%**	**86.9%**	**46.8%**
PPYOLOE	92.9%	76.2%	83.5%	37.5%
PPYOLOE*	**91.7%**	**80.2%**	**84.1%**	**41.8%**
PP-PicoDet	67.6%	64%	67.1%	29%
PP-PicoDet*	**67%**	**65%**	**67.2%**	**31.6%**

Table 4. Ablation Experiments of Attention IoU on Different Loss Functions

Method	Precision	Recall	mAP@.5	mAP@.5-.95
CIoU	96%	90%	94.1%	58.5%
CIoU*	**94.5%**	**90.7%**	**94.2%**	**59.8%**
EIoU	96%	90%	94%	59.4%
EIoU*	**93.2%**	**91.6%**	**94.7%**	**60.2%**
SIoU	93.6%	90.4%	94.8%	59.3%
SIoU*	**92.9%**	**91.1%**	**94.8%**	**60.5%**

the proposed attentionIoU demonstrates strong robustness and effectiveness. The formula for CIoU*, EIoU*, and siou* is shown below.

$$\mathrm{CIoU}^* = 1 - \mathit{AttentionIoU} + \frac{\rho^2(\mathbf{b}, \mathbf{b}^{gt})}{c^2} + \alpha v \tag{6}$$

$$\mathrm{EIoU}^* = 1 - \mathit{AttentionIoU} + \frac{\rho^2\left(\mathbf{b}, \mathbf{b}^{\mathrm{gt}}\right)}{\mathrm{c}^2} + \frac{\rho^2\left(\mathbf{w}, \mathbf{w}^{\mathrm{gt}}\right)}{\mathrm{C}_{\mathrm{w}}^2} + \frac{\rho^2\left(\mathbf{h}, \mathbf{h}^{\mathrm{gt}}\right)}{\mathrm{C}_{\mathrm{h}}^2} \tag{7}$$

$$\mathrm{SIoU}^* = 1 - \mathit{AttentionIoU} + \frac{\rho^2(\mathbf{b}, \mathbf{b}^{gt})}{c^2} + \frac{\Delta + \Omega}{2} \tag{8}$$

α is a trade-off parameter, and v is a parameter used to measure the consistency of the aspect ratio. $\frac{\rho^2\left(\mathbf{w},\mathbf{w}^{\mathrm{gt}}\right)}{\mathrm{C}_{\mathrm{w}}^2} + \frac{\rho^2\left(\mathbf{h},\mathbf{h}^{\mathrm{gt}}\right)}{\mathrm{C}_{\mathrm{h}}^2}$ represents the width and height losses of the prediction frame and the real frame. Δ represents distance loss. Ω represents shape loss. The above variables are not discussed due to the limited length of the article.

In summary, SC-AttentionIoU has shown a certain improvement in small object detection datasets. This paper also points out the potential problems caused by changing the weights of the internal regions of the ground truth boxes and proposes corresponding solutions. This proves that exploring the weights of the internal regions of ground truth boxes is beneficial. Our work provides a new direction for exploring loss functions in small object detection and is beneficial for the development of future object detection technologies.

5 Conclusion

This paper proposed the AttentionIoU loss which extends the IoU loss in the bounding box regression to stress small object detection accuracy. Moreover, to address the issue that changing the internal weights of the predicted boxes affects the aspect ratio loss in AttentionIoU, the SC-AttentionIoU loss function is presented. Finally, the proposed loss function is validated on two small object datasets and compared with other existing loss functions, demonstrating its effectiveness in small object detection.

Acknowledgements. This work was supported by Key R&D Program of Shan dong Province, China (2022RZB02012), the Taishan Scholars Program (NO. tscy2 0221110).

References

1. Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M.: YOLOv4: optimal speed and accuracy of object detection (2020)
2. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers (2020)
3. Chu, X., et al.: Twins: revisiting the design of spatial attention in vision transformers (2021)
4. Ge, Z., Liu, S., Wang, F., Li, Z., Sun, J.: YOLOX: exceeding YOLO series in 2021 (2021)
5. He, J., Erfani, S., Ma, X., Bailey, J., Chi, Y., Hua, X.S.: Alpha-IoU: a family of power intersection over union losses for bounding box regression (2022)
6. Hou, Q., Zhou, D., Feng, J.: Coordinate attention for efficient mobile network design (2021)
7. Li, J., et al.: Next-ViT: next generation vision transformer for efficient deployment in realistic industrial scenarios (2022)
8. Liu, Z., et al.: Swin transformer: hierarchical vision transformer using shifted windows (2021)
9. Prathima, G., Lakshmi, A.Y.N., Kumar, C.V., Manikanta, A., Sandeep, B.J.: Defect detection in PCB using image processing. Int. J. Adv. Sci. Technol. **29**(4) (2020)
10. Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., Savarese, S.: Generalized intersection over union: a metric and a loss for bounding box regression (2019)
11. Xu, S., et al.: PP-YOLOE: an evolved version of YOLO (2022)
12. Xu, X., Jiang, Y., Chen, W., Huang, Y., Zhang, Y., Sun, X.: DAMO-YOLO: a report on real-time object detection design (2023)
13. Yang, L., Zhong, J., Zhang, Y., Bai, S., Li, G., Yang, Y., Zhang, J.: An improving faster-RCNN with multi-attention ResNet for small target detection in intelligent autonomous transport with 6G. IEEE Trans. Intell. Transp. Syst., 1–9 (2022). https://doi.org/10.1109/TITS.2022.3193909
14. Yu, G., et al.: PP-PicoDet: a better real-time object detector on mobile devices (2021)
15. Yu, J., Jiang, Y., Wang, Z., Cao, Z., Huang, T.: UnitBox: an advanced object detection network. In: Proceedings of the 24th ACM International Conference on Multimedia, pp. 516–520 (2016). https://doi.org/10.1145/2964284.2967274

16. Zhang, T., et al.: SAR Ship Detection Dataset (SSDD): official release and comprehensive data analysis. Remote Sensing **13**(18), 3690 (2021). https://doi.org/10.3390/rs13183690
17. Zhang, Y.F., Ren, W., Zhang, Z., Jia, Z., Wang, L., Tan, T.: Focal and efficient IOU loss for accurate bounding box regression (2022)
18. Zhao, W., Kang, Y., Chen, H., Zhao, Z., Zhai, Y., Yang, P.: A target detection algorithm for remote sensing images based on a combination of feature fusion and improved anchor. IEEE Trans. Instrum. Meas. **71**, 1–8 (2022). https://doi.org/10.1109/TIM.2022.3181927
19. Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., Ren, D.: Distance-IoU loss: faster and better learning for bounding box regression. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34(07), pp. 12993–13000 (2020). https://doi.org/10.1609/aaai.v34i07.6999