Research on Tiny Object Detection Optimization based on YOLOv7

Bi kaiyu, Sun Baoshan

Abstract—**In this paper, we propose a new object detection model, YOLOv7-TinyObject, which is an upgrade of YOLOv7. Two technologies, SE (Squeeze-and-Excitation) and NWD (Normalized Gaussian Wasserstein Distance), are introduced into YOLOv7 to improve the detection ability of tiny objects. Although YOLOv7 has a great improvement over the previous YOLO series, there are still many shortcomings for detecting tiny objects. The detection of tiny objects has always been a difficult and challenging task in object detection tasks, because tiny objects usually have low resolution and fuzziness, and are easily disturbed by noise and occlusion. To address these issues, we propose a novel optimization method that combines the SE attention(SE) with the NWD technique. First, we analyze the limitations of YOLOv7 in tiny object detection. Specifically, YOLOv7 is prone to the problem of positioning error and missing targets in tiny object detection. To overcome these problems, we introduce the SE. The SE improves the representation ability of important features by adaptively learning the relationship between feature channels. By adjusting the weights of feature channels, the SE can enhance the modeling of important features, thereby improving the accuracy of object detection algorithms. Based on this, we further introduce the NWD to enhance the perception ability of the algorithm for tiny targets. The NWD provides richer semantic information by modeling the contextual information around the target. It considers context relations such as domain information, texture features and edge information of the object and is added to the object detection model to enhance the localization and detection accuracy of tiny objects. By increasing the contextual semantic understanding ability of the target, NWD helps to reduce the problem of localization error and missed detection, and improves the performance of tiny target detection. To validate our proposed optimization method, we use a wide range of datasets for experiments and evaluation. Experimental results show that compared with the traditional YOLOv7 model, our proposed improved method achieves significant performance improvement in tiny object detection. By introducing the SE and NWD, our algorithm exhibits higher accuracy and robustness in tiny target scenarios. Our study shows that the SE attention mechanism can improve the representation ability of important features and the NWD can enhance the perception of tiny targets. With these optimization methods, our algorithm achieves higher accuracy and robustness in the task of tiny object detection. This is of great significance for improving the application performance of object detection algorithms in complex scenes. Future research can further optimize and extend these methods to address challenges in more real-world applications.**

Index Terms—**YOLOv7,TinyObject Detection,SE,NWD**

I. INTRODUCTION

As a core task in computer vision, object detection has been widely concerned and studied. In recent years, object detection algorithms based on deep learning have made great breakthroughs and have been widely used in various fields. However, the detection of tiny objects has always been a challenging problem in practice. The features of tiny objects are usually small and fuzzy, and they are easily disturbed by image noise, occlusion and illumination change, which makes their detection and localization more difficult. Therefore, improving the performance of tiny object detection has always been one of the hot spots in the research of object detection algorithms.

To solve the problem of tiny object detection, researchers [2,3,4,5,6] have proposed many optimization methods. Among them, YOLOv7[1] has attracted much attention for its efficient detection speed and accuracy. However, despite the excellent performance of YOLOv7 in large object detection, there are still some limitations in the detection of tiny objects. In the YOLOv7 model, due to the characteristics of its feature extraction backbone network, it is weak in the detail and context modeling of tiny objects. In order to improve the detection of tiny objects based on YOLOv7, SE[7] and NWD[8] are introduced in this paper, aiming to improve the accuracy and robustness of the algorithm. The SE is a lightweight attention mechanism that enhances the representation ability of features by learning the relationship between feature channels. By adaptively weighting the feature channels, the SE can highlight important features, so as to improve the attention and representation ability of the object detection algorithm for tiny objects. On the other hand, the NWD enhances the perception of tiny objects by modeling the contextual information around the object. NWD not only considers the characteristics of the target itself, but also

Manuscript received July 23, 2023.

Bi kaiyu, School of Computer Science and Technology,Tiangong University,Tianjin,China

Sun Baoshan, School of Computer Science and Technology,Tiangong University,Tianjin, China

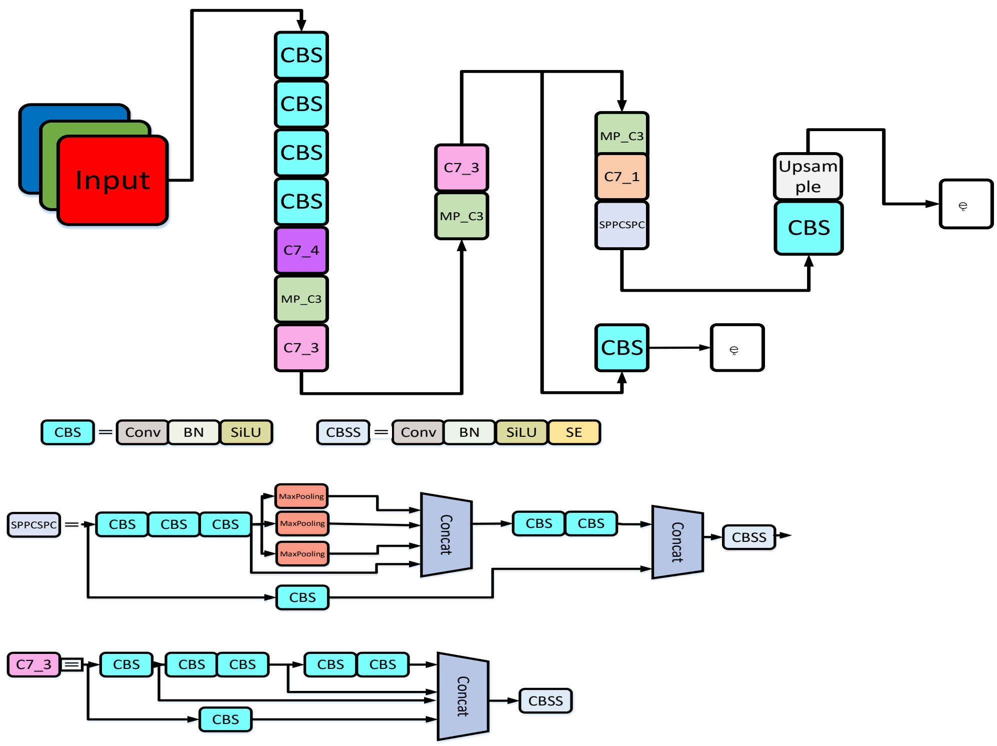


Figure 1. This figure shows the structure of YOLOv7-TinyObject modifications. Add SE after the CBS module to become the CBSS module, and then add the CBSS module after the SPPCSPC module and the C7_3 module to become YOLOv7-TinyObject.

analyzes the domain information, texture features and edge information around the target. By combining these contextual information with object features, the NWD can provide more comprehensive semantic information, which can help reduce the localization error and missed detection problems in tiny object detection.

In this paper, we aim to improve the tiny object detection algorithm based on YOLOv7[1] to improve its accuracy and robustness in complex scenes. We recognize that the detection of tiny objects is challenging and therefore need to combine efficiency and accuracy to design optimization methods. To this end, we propose improvements that introduce SE[7] and NWD[8]. The main purpose of introducing the SE is to enhance the representation ability of important features. The traditional YOLOv7 model may ignore some important features in tiny object detection, resulting in a decrease in the accuracy of detection. Through the SE, we can adaptively learn the relationship between feature channels and highlight important information by weighting features. In this way, our algorithm can better focus on tiny object regions and improve the accuracy of object detection. In addition, we introduce the NWD to enhance the perception of tiny targets. NWD provides more semantic information by modeling the contextual information around the target. This context information can include the domain information of the target, texture features, edge information, etc., which can enhance the understanding and detection of tiny targets. By adding the NWD, our algorithm can better capture the semantic context of the area where the tiny target is located, thus improving the accuracy of localization and detection.

To verify the effectiveness of our proposed method, we conduct experiments and evaluation on datasets with both tiny and large targets. We will compare with the traditional YOLOv7[1] algorithm to evaluate the performance of our optimized model in the task of tiny object detection. We will use common evaluation metrics such as precision, recall, and F1 score to evaluate the accuracy and robustness of the algorithm.

The structure of this paper is as follows: The second part will introduce the related research work and technical background; The third part will introduce our proposed optimization method for tiny object detection based on YOLOv7 in detail. Section IV will show the experimental results and performance evaluation. Finally, Section V will conclude our work and outline future research directions

II. RELATED WORK

Since the release of YOLOv7[1], many researchers [9,10,11,12] have tried to improve it. As of this writing, no other researchers have been found to have conducted exactly the same research on YOLOv7 as the improvement work in this paper. In the improvement work presented in this paper, some related techniques are used. Here is a brief overview of related work on these techniques to better understand how YOLOv7-TinyObject works.

A. *Related work on SE*

In [7], the authors studied another aspect of network design: the relationship between channels. They introduced a new architectural unit called squeeze and excitation (SE) block, which aims to improve the quality of representations produced by the network by explicitly modeling the interdependencies between channels in convolutional features.

This mechanism allows the network to learn to use global information to selectively emphasize informative features and suppress less useful ones, enabling feature recalitioning.

A SE network (SENet) [7]can be built by simply stacking multiple SE blocks. In addition, SE blocks can also be used as replacement blocks for original blocks with different depths in the network architecture. Although the template for building blocks is generic, SE blocks play different roles at different depths throughout the network. In earlier layers, SE blocks motivate informative features in a class-agnotic manner, enhancing the shared underlying representation. While in later layers, the SE blocks become increasingly specialized and respond to different inputs in a highly class-specific manner. Thus, the feature recalivity effect of the SE block can be accumulated throughout the network.

B. Related work on NWD

Literature [8] proposed an innovative tiny object detection evaluation method based on Wasserstein distance. Specifically, the authors first represent BBox as two-dimensional Gaussian distributions, and then introduce a new metric, namely Normalized Wasserstein distance (NWD), to measure the quality of object detection results by calculating the similarity between their corresponding Gaussian distributions. This NWD metric can be easily embedded into the allocation, non-maximum rejection, and loss functions of any anchor-based detector to replace the conventional IoU metric. The authors evaluate the metric on AI-TOD, a new dataset dedicated to tiny object detection, where the average object size is much tinyer than existing object detection datasets. Extensive experimental results show that the proposed method has potential application value in improving the accuracy of tiny object detection when using NWD metric..

III. METHODS

A. Introduce the SE

First, we could consider adding SE[7] to different layers of backbone. Backbone usually consists of multiple convolutional and pooling layers to extract image features. By adding the SE module after each convolutional layer, we can enhance the important information representation of different layers in the feature map and further improve the expressive power of the network. The benefit of this is that the SE module is able to introduce critical information at an early stage of feature extraction, allowing the network to focus on useful features earlier, thus improving performance.

Second, we can add the SE[7] module to the head part, which is the last few layers of the network. In this case, we pay more attention to applying SE on the feature fusion and decision layer before the network output. By adding the SE module to the head section, we are able to mine the information in the feature map more pertinently and improve the recognition ability of the network for the target. The benefit of this is that the SE module can pertinently strengthen the features that have been extracted, so that the network can judge the target more accurately, thus improving the detection accuracy.

Furthermore, we can add SE[7]modules to both backbone and head to further improve performance. Through this double reinforcement method, we are able to mine and optimize the important information in the feature map in an all-round way, so as to obtain higher object recognition accuracy. The benefit of this is that the SE module can simultaneously strengthen the feature extraction process and the feature fusion process, so as to better capture the key features of the target and improve the expression ability of the whole network.

The specific situation can be referred to Figure 1.

B. Add the NWD

NWD[8] can also be used when the objects to be detected in the dataset are not entirely tiny objects. In this paper, both IoU and NWD are used. In order to better deal with tiny targets, we introduce the coefficient iou_ratio. This factor is critical to balance the weight of IoU and NWD in the total loss. Suppose we set iou_ratio=0.5, meaning that IoU and NWD each account for 50% of the total loss.

First, we consider the case when the dataset is full of tiny objects. Since tiny targets tend to have tiny regions, the intersection and union ratio between the predicted box and the true target box may be relatively low. Therefore, we can set the iou_ratio to 0, so that the loss of NWD[8] will be more important. By paying more attention to the distance between the predicted box and the true target box, the algorithm will locate the tiny target more accurately and improve the detection accuracy.

However, the number of tiny targets accounts for about half of the total in my dataset. This means that a certain percentage of medium and large targets still exist in the dataset. If we completely ignore the importance of the IoU metric, we may not be able to take full advantage of the overlap between the predicted box and the actual target box. Therefore, in order to balance the performance of the object detection algorithm on different object sizes, we set iou_ratio to 0.5. With this setting, we add the losses of IoU and NWD to the total loss with equal weight. This practice allows the algorithm to achieve a reasonable balance between tiny objectives and other objectives. It makes the algorithm more accurate in locating tiny targets, and can make full use of the intersection and union ratio information of medium and large targets.

Using iou_ratio to balance the loss of IoU and NWD[8] is very important when dealing with datasets containing tiny targets. By properly setting the value of iou_ratio, we are able to optimize the performance of the object detection algorithm on different object sizes. This method takes full account of the characteristics of tiny targets, but also retains the accuracy requirements for other targets, so as to improve the overall object detection performance. In practical applications, we can dynamically adjust the value of iou_ratio according to the characteristics and requirements of the data set to further improve the effect of the algorithm.

Here's the core code for NWD:

```
def wasserstein_loss(pred, target, eps=1e-7, constant=12.8):
    """Implementation of paper `Enhancing Geometric Factors into
    Model Learning and Inference for Object Detection and Instance
    Segmentation <https://arxiv.org/abs/2005.03572>`_.
```

```
    Code is modified from https://github.com/Zzh-tju/CIoU.
    Args:
        pred (Tensor): Predicted bboxes of format (x_center, y_center, w, h),
            shape (n, 4).
        target (Tensor): Corresponding gt bboxes, shape (n, 4).
        eps (float): Eps to avoid log(0).
    Return:
        Tensor: Loss tensor.
    """
    center1 = pred[:, :2]
    center2 = target[:, :2]
    whs = center1[:, :2] - center2[:, :2]
    center_distance = whs[:, 0] * whs[:, 0] + whs[:, 1] * whs[:, 1] + eps #
    w1 = pred[:, 2]  + eps
    h1 = pred[:, 3]  + eps
    w2 = target[:, 2] + eps
    h2 = target[:, 3] + eps
    wh_distance = ((w1 - w2) ** 2 + (h1 - h2) ** 2) / 4
    wasserstein_2 = center_distance + wh_distance
    return torch.exp(-torch.sqrt(wasserstein_2) / constant)
nwd = wasserstein_loss(pbox, tbox[i]).squeeze()
iou_ratio = 0.5
lbox += (1 - iou_ratio) * (1.0 - nwd).mean() + iou_ratio * (1.0 - iou).mean()  # iou loss
```

IV. EXPERIMENT

A. *The Dataset*

In this experiment, we used the Windmill dataset as the experimental dataset. The dataset contains one category, namely windmill, and the dataset is divided into training, val and test. The train contains 2522 images, the val contains 240 images, and the test contains 123 images. Multiple windmills are present in each image, half of which are larger size windmills and half of which are tinyer size windmills, which corresponds to the aforementioned iou_ratio=0.5. In addition, the image contains some targets that are not windmills, such as high-voltage line towers, and there are some windmills that are difficult to recognize due to occlusion.

The reasons for choosing the Windmill dataset as the experimental dataset are manifold. Firstly, this dataset can comprehensively test the performance of the improved YOLOv7-TinyObject in tiny object detection, and verify whether the improved method affects the detection effect of large objects. Secondly, the dataset contains multiple complex situations that are common in actual scenes, such as windmill with different sizes and occlusions. Therefore, by conducting experiments on this dataset, the performance of the algorithm can be evaluated more comprehensively and accurately.

The experimental results show that the improved YOLOv7-TinyObject has achieved significant improvement in tiny object detection, and it will not have a negative impact on the detection effect of large objects. This indicates that the improved method can detect tiny objects without error and maintain good detection results for large objects. These experimental results further verify the effectiveness and robustness of the improved YOLOv7-TinyObject in dealing with the problem of tiny object detection.

B. *Training*

During training, we used an NVIDIA GEFORCE RTX3080 GPU as the hardware device due to limited computing resources. To find the right number of epochs, we first set the epoch to 100 and ran the training. However, the detection accuracy after the training is not high, from which it is inferred that there is underfitting. To fix this,we increased the number of epoch to 300 and did the training again. However, satisfactory detection accuracy is still not achieved after training. By querying the records of the training process, we found that overfitting was occurring. Based on this,we set the number of epochs to 200 and did the final training.

The final training results show that after setting the training number of 200 epochs, the model reaches a good fitting state, and the detection accuracy of the target can achieve the most ideal state. I trained both the original model and the improved YOLOv7-TinyObject with the same training parameters, and the results will be described in the next section. Figure 2 shows a part of the training results during the training process, and it can be observed that the model can accurately identify large and tiny objects, and still perform effective detection when it meets object occlusion, object blur, or object part missing. Through these training results, we can conclude that the improved YOLOv7-TinyObject model can achieve accurate and robust object detection after training.

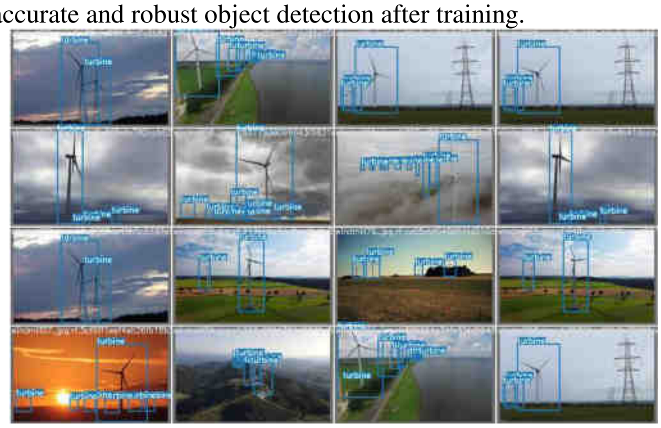


Figure 2. This figure shows some of the training results during the training process.

C. Experimental Results

a. Training results

After training, we looked at the training curve, which included the F1 curve. The F1 curve is a visual tool used to evaluate the performance of binary classification models at different thresholds. It helps us understand the overall performance of the model by plotting Precision, Recall, and F1 score for different thresholds. The F1 score is the harmonic mean of precision and recall, taking into account both performance metrics. Through the F1 value curve, we can be sure to find the balance between different precision and recall rates and choose the best threshold. Figure. 3 shows the F1 curve.

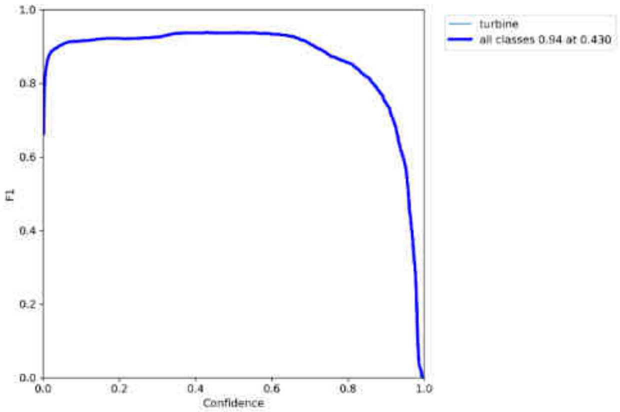


Figure 3. This plot shows how the F1 changes during training.

In addition, the training results also contain four equally important curves: Box, Objectness, val Box and val Objectness. Box indicates that the prediction is the mean of the loss function, and a tinyer value indicates that the prediction box is more accurate. Objectness is the mean of the object detection loss, with tinyer values indicating more accurate object detection; val Box is the bounding box loss on the validation set and val Objectness is the object detection loss on the validation set. The four training curves are shown in FIG. 4. Both the F1 curve and these additional curves provide useful information for evaluating model performance as well as for selecting the best threshold.

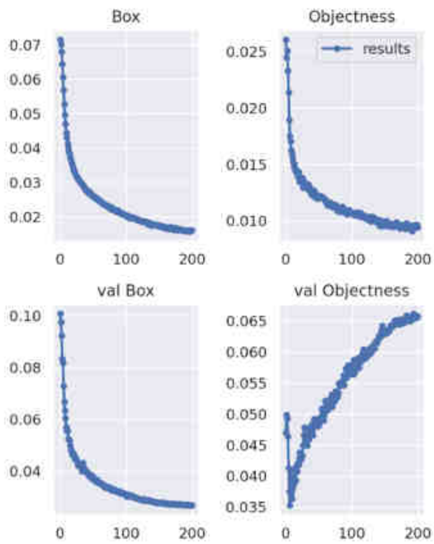


Figure 4. This figure shows how Box, Objectness, val Box, val Objectness changes during training.

b. Comparison with the original model

We compared the training results with the original model by four metrics: Precision, Recall, mAP@0.5 and mAP@0.5:0.95. Precision refers to the proportion of the number of samples correctly predicted as positive examples among all the samples predicted as positive examples. It can be seen from Figure 5 that compared with the original YOLOv7-TinyObject, the detection accuracy of YOLOv7 for tiny objects is improved from 0.83 to 0.94. Recall refers to the proportion of the number of samples that are correctly predicted as positive examples to the number of samples that are actually positive examples. Figure 6 shows that the recall rate of YOLOv7-TinyObject is also improved compared with the original YOLOv7. Precision curves are often used in conjunction with Recall curves to provide a more comprehensive analysis of classifier performance and to help evaluate and compare the performance of different models. Recall is also called Sensitivity or True Positive Rate.

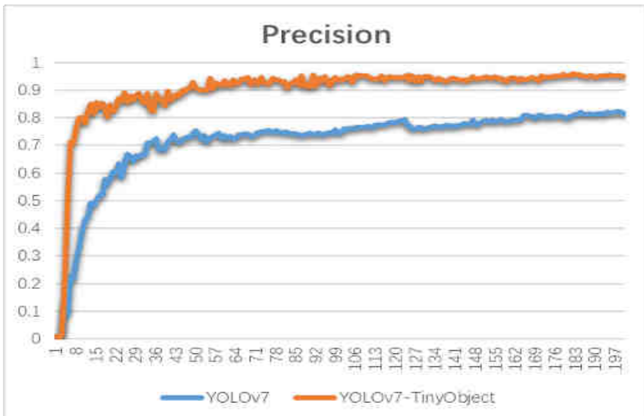


Figure 5. This graph shows how the accuracy of the two models compares. From this figure, it can be seen that YOLOv7-Tinyobject has a significant improvement in the detection accuracy of small objects compared with the original YOLOv7.

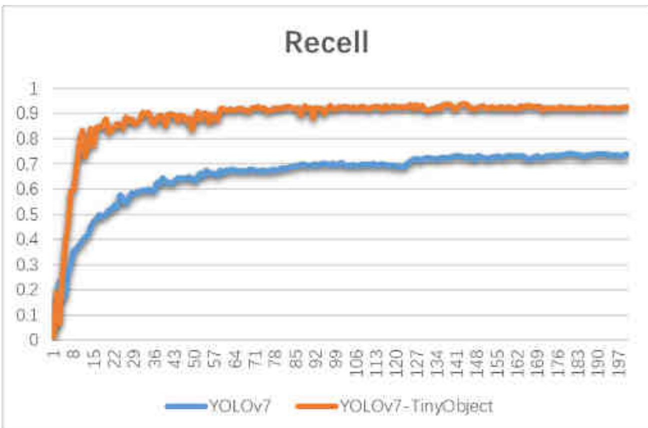
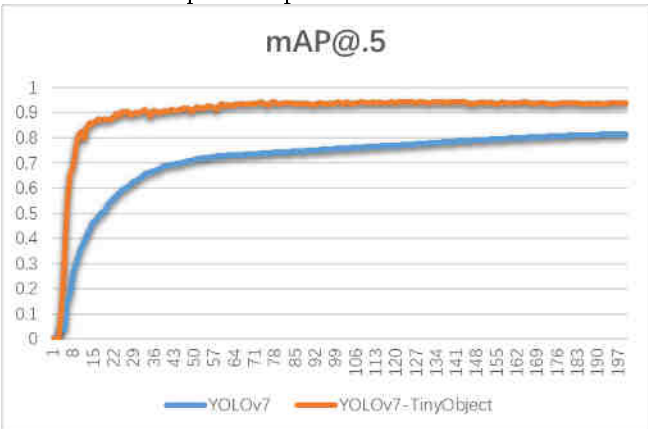
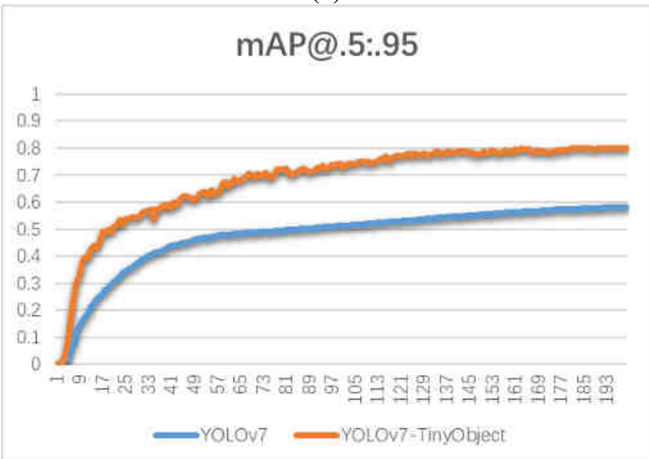


Figure 6. This plot shows how the recall of the two models compares. From this plot, you can see that YOLOv7-Tinyobject has a significant improvement in recall over the original YOLOv7.

The mAP (mean Average Precision) is an evaluation metric obtained by plotting a curve on the Precision and Recall axes and calculating the area under the curve. mAP is usually calculated based on different IoU thresholds. IoU is the criterion to determine whether the predicted box is a positive example by comparing the overlap degree between the predicted box and the true labeled box. mAP@0.5:0.95 indicates that the average mAP value is calculated at different IoU thresholds ranging from 0.5 to 0.95 with a step size of 0.05. map@ [.5:.95] represents the mAP value at different IoU thresholds for each 0.05 interval in the interval 0.5 to 0.95. While mAP@0.5 represents the average mAP value when the threshold is greater than 0.5. Through 7, it can be seen that both maps are improved.



(a)



(b)

Figure 7. (a) and (b) show the boost effect of mAP@.5 and mAP@.5:.95, respectively.

Through comparison, we can find that after introducing SE and NWD into YOLOv7, the detection ability of tiny targets is significantly improved, and the detection ability of large targets is not significantly weakened. Therefore, the experiment shows that our improvement of YOLOv7 is successful.

V. Conclusion and Prospect

This paper aims to improve the tiny object detection algorithm based on YOLOv7, and improve the accuracy and robustness of the algorithm in tiny object scenes by introducing the SE and NWD. Through experiments and evaluation, we verify the effectiveness and superiority of the proposed method.

Firstly, by introducing the SE, our algorithm is able to adaptively learn the relationship between feature channels, highlight important features and improve the accuracy of object detection. By enhancing the representation ability of tiny objects, our algorithm is able to better focus on tiny object regions and reduce localization errors and missed detection problems. Secondly, by introducing the NWD, our algorithm is able to make full use of the contextual information around the target and provide more semantic information to enhance the perception ability of tiny targets. By integrating object features and context information, our algorithm is able to locate and detect tiny objects more accurately and improve the overall detection performance.

Through experiments and evaluation, we find that compared with the traditional YOLOv7 algorithm, our proposed improved method achieves significant performance improvement in tiny object detection. The introduction of SE and NWD can improve the representation and perception ability of the target, so as to improve the detection accuracy of tiny targets. This is of great significance for the detection of tiny targets in practical applications.

Despite the achievements of our work, there are still some room for improvement and expansion. In future research, we can further optimize the algorithm performance and robustness for more complex scenarios and more challenging tiny targets. In real-world applications, tiny targets usually suffer from different disturbances, such as occlusion, illumination changes, and complex backgrounds. Therefore, we can further improve the algorithm to cope with the challenges in these complex scenarios. For example, incorporating more contextual information, adopting more sophisticated feature fusion strategies, or introducing attention mechanisms to weight specific regions are all directions that can be explored. Multi-scale strategies can also be considered to improve the performance of tiny object detection. Tiny objects usually have different scale variations, so considering object detection at different scales can cover a variety of object sizes more comprehensively and improve the accuracy of detection. Other advanced object detection techniques can also be combined to further improve the performance. The research in the field of object detection is changing rapidly, and new algorithms and techniques are emerging one after another. We can try to combine other advanced object detection algorithms and techniques, such as EfficientNet, RetinaNet, etc., with our proposed improved method to further improve the performance of tiny object detection. Validation and evaluation on larger datasets can likewise be a future research direction. Although we have conducted a series of experiments and evaluations in the present paper, the dataset for validation is still limited. Further expansion of the dataset size and validation in more domains and various complex scenarios will enable a more comprehensive understanding of the performance and application scope of our improved method.

Acknowledgment

This work is partically supported by Natural Science Foundation of China Grants(61972456， 61173032)and Tianjin Natural Science Foundation (20JCYBJC00140).

REFERENCES

[1] Wang C Y, Bochkovskiy A, Liao H Y M, “YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,”Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023, pp. 7464-7475.

[2] Redmon J, Divvala S, Girshick R, et al, “ You only look once: Unified, real-time object detection,”Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788.

[3] Redmon J, Farhadi A, “Yolo9000: better, faster, stronger,”Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 7263-7271.

[4] Li X, Jiang Y, Liu C, et al, “Playing Against Deep-Neural-Network-Based Object Detectors: A Novel Bidirectional Adversarial Attack Approach,” IEEE Transactions on Artificial Intelligence, vol. 3, no. 1, pp. 20-28, Feb. 2022.

[5] Wang C Y, Bochkovskiy A, Liao H Y M,“Scaled-yolov4: Scaling cross stage partial network,”Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 13029-13038.

[6] Yung N D T, Wong W K, Juwono F H, et al, “Safety Helmet Detection Using Deep Learning: Implementation and Comparative Study Using YOLOv5, YOLOv6, and YOLOv7,” 2022 International Conference on Green Energy, Computing and Sustainable Technology (GECOST), Miri Sarawak, Malaysia, 2022, pp. 164-170.

[7] Hu J, Shen L, Sun G, “Squeeze-and-Excitation Networks,”Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 7132-7141.

[8] Wang J, Xu C, Yang W, et al, “A normalized Gaussian Wasserstein distance for tiny object detection,”. arXiv preprint arXiv:2110.13389, 2021, unpublished.

[9] Liu S, Wang Y, Yu Q, et al, “CEAM-YOLOv7: Improved YOLOv7 Based on Channel Expansion and Attention Mechanism for Driver Distraction Behavior Detection,” IEEE Access, vol. 10,, 2022pp. 129116-129124.

[10] Baghbanbashi M, Raji M, Ghavami B, “Quantizing YOLOv7: A Comprehensive Study,” 2023 28th International Computer Conference, Computer Society of Iran (CSICC), Tehran, Iran, Islamic Republic of, 2023, pp. 01-05.

[11] Liu Y, Wang X, “SAR Ship Detection Based on Improved YOLOv7-Tiny,” 2022 IEEE 8th International Conference on Computer and Communications (ICCC), Chengdu, China, 2022, pp. 2166-2170.

[12] Hong X, Wang F, Ma J, “Improved YOLOv7 Model for Insulator Surface Defect Detection,” 2022 IEEE 5th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Chongqing, China, 2022, pp. 1667-1672.

[13] Zhu L, Wang X, Ke Z, et al, “BiFormer: Vision Transformer with Bi-Level Routing Attention,”Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023, pp. 10323-10333.

Bi kaiyu,Graduate student. His research interests include computer vision and object detection

Sun Baoshan, Associate Professor Sun Baoshan, Doctor of Engineering, Master Supervisor, Deputy Head of Department. The country sent abroad to study in the UK visiting scholar, CCF Member of China Computer Society. Graduated from Tianjin University with a master's degree in computer application technology, and graduated from Tianjin Polytechnic University in computer testing Ph.D. graduate. Selected into the "Outstanding Young Teachers Funding Program" of Tianjin Universities. Leading the research team in related research fields and achieved a series of scientific research results in research topics. And has been in foreign SCI journals, EI journals, domestic important journals and international more than 20 high-level academic papers have been published at the conference.