Academic Journal of Computing & Information Science
ISSN 2616-5775 Vol. 7, Issue 10: 67-75, DOI: 10.25236/AJCIS.2024.071010

An optimized Yolov8n approach and its application to object detection problems

Shouhao Zheng[1], Feng Tian[2]

[1]College of Intelligent and Information Engineering, Shandong University of Traditional Chinese Medicine, Jinan, 250399, China
[2]School of Data Intelligence, Yantai Institute of Science and Technology, Yantai, 265600, China

Abstract: *Aiming at the problems of low detection accuracy and low detection efficiency of Yolov8s algorithm, an improved Yolov8n algorithm with high accuracy and few parameters is proposed. Specifically, to address the problem of a large number of parameters in the basic Yolov8s, its variant Yolov8n algorithm is introduced to improve the model training efficiency. Subsequently, for the problem that the original CIOU function cannot be accurately localised, the WIOU loss function is introduced to consider the degree of overlap between different regions and accurately mask the boundaries, which further improves the segmentation performance. Finally, an SGD optimiser is introduced on this basis to improve the model training efficiency. The improved Yolov8n algorithm is applied to the VOC2007 dataset for training and validation. The experimental results show that the proposed improved Yolov8n algorithm has higher training accuracy and fewer parameters with high training efficiency, reaching 85.6% for P, 76.9% for R and 83.3% for map on the dataset, which has high detection accuracy and validates the model. Compared with other models, the number of parameters is reduced, and the detected maps are improved by 0.6% ~ 17.1% and significantly improved on several detection categories, which verifies the superiority of the model. In summary, the improved yolov8n algorithm proposed in this paper can meet the functional requirements of high accuracy and high efficiency of traffic object detection in road traffic equipment as well as self-driving traffic roads.*

Keywords: *Yolov8n, WIOU, SGD, VOC2007*

1. Introduction

In recent years, with the rapid development of driving technology, self-driving cars continue to move towards practicality. In the field of automated driving, traffic device detection is very important to transmit the recognition results to the automated driving decision module to ensure that the vehicle can drive safely according to the traffic rules [1]. However, with the large number and variety of transport vehicles, manual monitoring is bound to have problems such as low efficiency and high leakage rate. Therefore, there is an urgent need to develop an efficient traffic target detection method that can quickly and accurately detect transport vehicles on the road, improve the safety of road traffic and reduce the accident rate [2].

Currently, target detection algorithms mainly include single-stage algorithms and two-stage algorithms. Ross Girshick [3] proposed an RCNN algorithm for the high accuracy detection needs in autonomous driving scenarios, but the algorithm convolutional network is time-consuming and the training steps are cumbersome, and there is a problem of low detection accuracy under large-scale datasets. Considering the existing problems, Girshick improves on the R - CNN and SPPne models for the high-precision detection needs in autonomous driving scenarios, and proposes the Fast - RCNN [4] optimisation model, which no longer needs extra space to save the feature vectors extracted by the CNN network but separates the target detection network, which has the problem of detecting the feature vectors with low accuracy under large-scale datasets and problem of low accuracy.

In comparison, the one-stage algorithm has the advantages of faster computation speed, lower computational cost, and better real-time performance than the two-stage algorithm.Joseph Redmon [5] proposed the YOLOv3 algorithm for the high-precision detection needs in the autonomous driving scenario.The algorithm has high training efficiency, but there is the problem of low detection accuracy under large-scale datasets.Glenn-Jocher [6] proposed the yolov5 algorithm for the demand of high-precision detection in unmanned scenarios, but the algorithm still has the problem of low detection accuracy under large-scale datasets.Subsequently, Glenn - Jocher further proposed the yolov8 [7]

ISSN 2616-5775 Vol. 7, Issue 10: 67-75, DOI: 10.25236/AJCIS.2024.071010

algorithm for the high-precision detection needs in automatic driving scenarios, which is characterised by fast training efficiency, but the detection accuracy still needs to be improved under special large-scale datasets. Therefore, there is an urgent need for more efficient target detection algorithms to meet the functional requirements of high-precision detection for autonomous driving.

Kingma combined the advantages of AdaGrad and RMSprop algorithms and proposed the Adam [8] algorithm.Wang [9] applied the ADAM optimiser in the hyper-parameter optimisation of the yolov5 model, which improves the training efficiency of the model, but the ADAM optimisation belongs to the stochastic gradient information, which is easy to fall into the local optimal solution species, resulting in low training accuracy;Sun [10] applied the ADAMW optimiser to the hyperparameter optimisation of the YOLOv8 model, although it can effectively deal with the sparse gradient problem to a certain extent, the ADAM optimiser has a large memory footprint, more parameter choices and is sensitive to small batches of data, resulting in lower applicability in some scenarios;and Cheng [11] et al. applied the RMSprop algorithm optimiser to the hyperparameter optimisation of YOLOv3 model, which can improve the training efficiency of the model.The Yolov3 model in hyperparametric optimisation overcomes the problem of excessive reduction of the AadmGrad learning rate to a certain extent, but the optimiser's adjustment of the learning rate may be too drastic leading to oscillations during the training process, and it may not be able to effectively cope with the situation in which the gradient changes in some directions are large while the gradient changes in other directions are small. Therefore, there is an urgent need to adopt optimisation algorithms with stronger applicability and outstanding optimisation search capability, which in turn can improve the model training efficiency and detection accuracy.

In summary, this paper proposes an improved YOLOv8n algorithm, which is characterised by high accuracy, small number of parameters and fast detection speed. Specifically, to address the high number of parameters in the basic YOLOv8s, its variant YOLOv8n algorithm is introduced to improve the efficiency of model training. Subsequently, to address the problem that the original CIOU function cannot accurately measure the loss due to the aspect ratio difference between predicted and real frames when the aspect ratio difference grows or decreases in a linear proportion, the WIOU loss function is introduced to take into account the degree of overlap between different regions and accurately mask the boundaries to further improve the segmentation performance. Finally, the SGD optimiser is introduced to improve the model training efficiency. Taken together, the improved yolov8n algorithm proposed in this paper can meet the demand for high-performance intelligent detection for autonomous driving.

2. Basic YOLOV8s model network structure

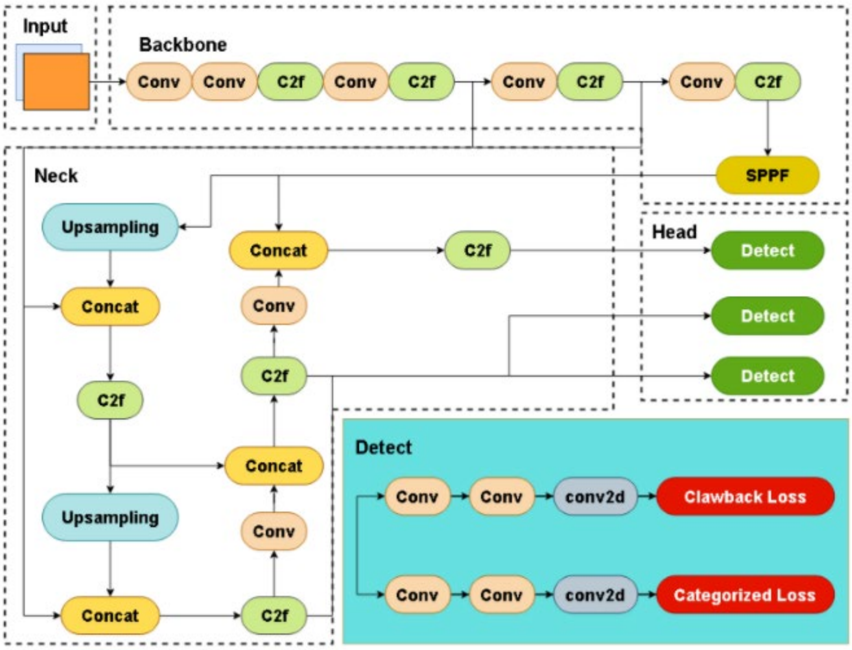

Figure 1: YOLOv8n network main structure

Yolov8 [12] is the eighth version of the YOLO model, designed to provide higher detection accuracy and faster inference. Depending on the depth and width of the network, it can be classified into several

ISSN 2616-5775 Vol. 7, Issue 10: 67-75, DOI: 10.25236/AJCIS.2024.071010

different versions: YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x, which have the same network structure, the only difference is the depth and width of the network. The YOLOv8n algorithm provides a new SOTA model that can be used for tasks such as target detection, image classification, instance segmentation and target tracking in computer vision. The structure of the YOLOv8n model consists of three main parts: the backbone part (Backbone), the neck network (Neck) and the detection head (Head). Its network structure is shown in Fig. 1.

Backbone is mainly responsible for extracting features from the input image, and then three valid feature layers are selected in Backbone and input into YOLOv8s Neck, which consists of multiple Conv, C2f modules and SPPF in the tail. Referring to the residual structure of the C3 module and the ELAN idea of YOLOv7 [13], YOLOv8 designs the C2f structure that can obtain richer gradient flow information while guaranteeing light weight, and adjusts the number of channels according to the model scale, which significantly improves the model performance. In particular, the CBS in the C2f structure consists of 3 parts: the 2D Convolution and the 2D BatchNorm, and the SiLU activation function, where the activation of the SiLU is computed by multiplying the Sigmoid function with its inputs. The structural diagrams of the C2f, as well as of the Bottleneck and the CBS modules are shown in Figs. 2 and 3 below, respectively.

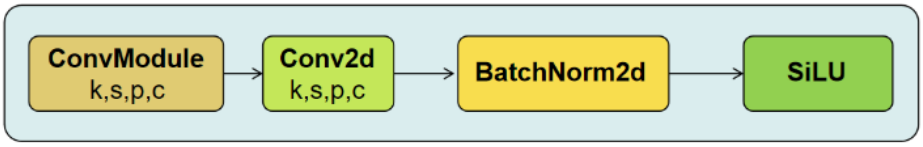

Figure 2: C2f and the Bottleneck module section

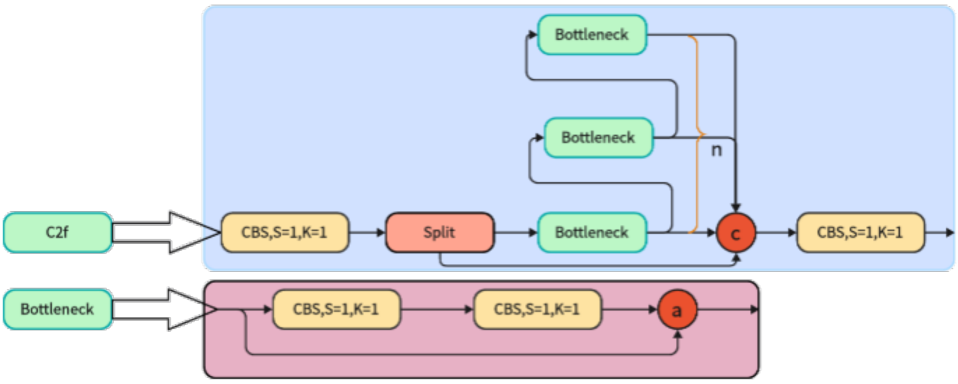

Figure 3: CBS module components

The main task of Neck is to fuse the multi-scale features and generate the feature pyramid to provide more information for the subsequent tasks. The PANet4 structure is adopted, which consists of two parts: feature pyramid network (FPN) [14] and path aggregation network (PAN) [15]. FPN also makes full use of the background information of the bottom layer for fusion, so that the target information in the high-level feature map can be retained, while the background information in the bottom layer feature map can be fully utilised. PAN is a network structure used to enhance the performance of FPNs, which introduces the concept of path aggregation and fully preserves spatial feature information by extracting information from feature maps of different layers and fusing them through a bottom-up structure.

For the inspection head section, the YOLOv8n uses a split inspection head without anchors, which contributes to improved inspection accuracy and a more efficient inspection process compared to anchor frame-based methods. Head part is replaced by the mainstream decoupled head result (Dccoupled - Head), Dccoupled - Head uses two convolutions to do classification and regression respectively; refer to the YOLOX algorithm, Anchor - Free is used instead of Anchor - Based, which is more advantageous in the face of the target with irregular length and width, as shown in Fig. 4.

ISSN 2616-5775 Vol. 7, Issue 10: 67-75, DOI: 10.25236/AJCIS.2024.071010

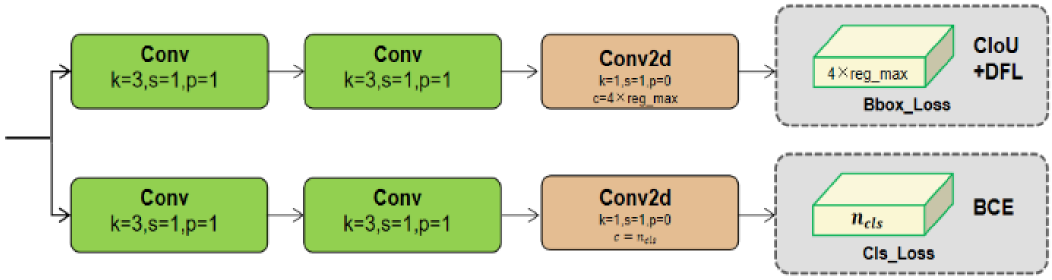

Figure 4: YOLOv8 Decoupled Head

Yolov8n also suffers from the following shortcomings:(1) The backbone network has insufficient ability to extract locally important information features in special environments such as fuzzy, dark, bright, etc., and poor generalization ability. (2) The C2f module of the newly proposed YOLOv8n stacks more Bottleneck structures, which not only reduces the real-time effect of the model, but also increases the computation of the model. (3) The model is difficult to detect correctly and accurately in the mobile detection where the environment is complex and the detection accuracy is required.

3. Improved YOLOv8n algorithm

3.1 Improvement of the loss function

For the original YOLOv8 model, its classification loss is Vari focal Loss (VFL) and its regression loss is in the form of CIoU Loss + DFL (Distributed Focusing Loss), in order to better improve the recognition accuracy, this study considers to improve the loss function by replacing the CIoU Loss with the WIoU Loss [16].

VFL proposes an asymmetric weighting operation in the following form:

$$VFL(p,q) = \begin{cases} -q\left(\begin{array}{c} q\log(p) \\ +(1-q)\log(1-p) \end{array}\right) \\ -\alpha p^{\gamma}\log(1-p) \end{cases} \tag{1}$$

where q is the intersection ratio of predicted and labelled frames and p is the probability of predicted and labelled frames.

The intersection over union (IoU) used by YOLOv8 and the loss function is CIoU, which takes into account the overlap area between the predicted box and the real box, the center point distance and the aspect ratio [17]. However, there is still some ambiguity in the description of aspect ratio. The formula defined for the original CloU Loss is shown here:

$$L_{\mathrm{CIoU}} = L_{\mathrm{IoU}} + \frac{(x-x_g)^2 + (y-y_{gt})^2}{(W_g^2 + H_g^2)} + \alpha\nu, \tag{2}$$

$$L_{\mathrm{IoU}} = 1 - IoU = 1 - \frac{W_i H_i}{wh + w_{gt}h_{gt} - W_i H_i}, \tag{3}$$

$$\alpha = \frac{\nu}{L_{\mathrm{IoU}} + \nu}, \tag{4}$$

$$\nu = \frac{4}{\pi^2}\left(\arctan\frac{w}{h} - \arctan\frac{w_{\mathrm{gt}}}{h_{\mathrm{gt}}}\right)^2, \tag{5}$$

where, L is the degree of overlap between the metric prediction frame and the real frame, α is the balance parameter, v is the consistency of the aspect ratio, and the rest of the parameters are defined as shown in Fig. 5:

ISSN 2616-5775 Vol. 7, Issue 10: 67-75, DOI: 10.25236/AJCIS.2024.071010

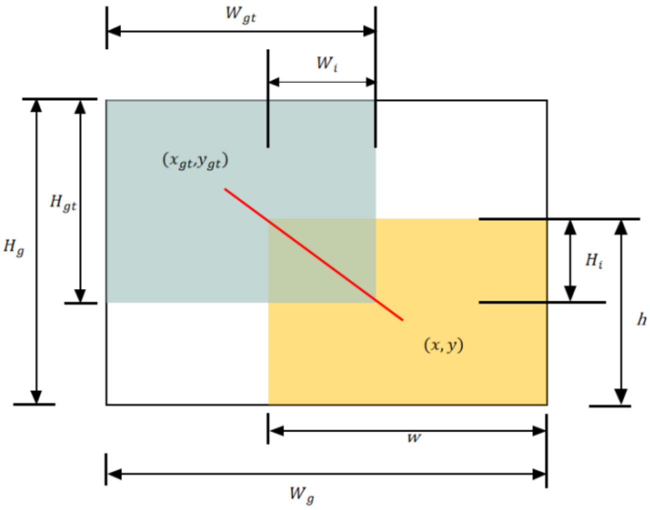

Figure 5: Definition of relevant parameters

CloU Loss focuses on enhancing the fitting ability of bounding box losses, ignoring the deleterious effects of low quality data on model performance. To solve this problem, WIoU Loss is chosen instead of CIoU Loss. WloU Loss reduces the competitiveness of high-quality bounding boxes while reducing the harmful gradient of low-quality data, allowing the model to focus on average-quality anchor frames, which improves the overall performance of the model.

The degree of abnormality in the prediction box is denoted by β, defined as follows:

$$\beta = \frac{L_{\mathrm{IoU}}^{*}}{\overline{L_{\mathrm{IoU}}}} \in [0, +\infty), \tag{6}$$

where, L_{IoU}^{*} is the constant into which the variable $\overline{L_{\mathrm{IoU}}}$ is converted, and $\overline{L_{\mathrm{IoU}}}$ is the running mean of the momentum m. The following is a summary of the results of the study, $m = 1 - \sqrt[tn]{0.5}, tn > 7000$.

Constructing a dynamic nonmonotonic focusing mechanism using β and combining it with an attention-based bounding box loss yields a WIoU Loss [18] that can readily assign gradient gains that match the prevailing situation, as defined below:

$$L_{\mathrm{WIoU}} = \frac{\beta}{\delta \alpha^{\beta - \alpha}} R_{\mathrm{WIoU}} L_{\mathrm{IoU}}, \tag{7}$$

$$R_{\mathrm{WIoU}} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{W_g^2 + H_g^2}\right) \tag{8}$$

where, α, β are learning parameters, $R_{\mathrm{WIoU}} \in [1, \mathrm{e})$, significantly amplifies the L_{IoU} of normal quality anchor frames.

3.2 Optimizer Selection

Stochastic Gradient Descent (SGD)[19] is a commonly used optimizer and one of the most basic optimization algorithms in deep learning models. It is an implementation of gradient descent algorithm and is often used for weight updating of neural networks. The specific parameter update formula is as follows:

In the gradient descent algorithm, the prediction function is:

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_n x_n \tag{9}$$

Using the squared loss function you can get the loss function of this function:

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{m}(h_\theta(x) - y)^2 \tag{10}$$

In order to minimise the loss function of this function, the partial derivatives of each θ_j : hyperparameter can be found for the current round of gradient, and then the loss function is updated along the opposite direction of the gradient, with continuous iterative updating, which leads to a globally optimal solution for the hyperparameters. Using the chain derivation method, the mathematical process can be expressed as:

$$\begin{aligned}\frac{\partial}{\partial\theta_j}J(\theta) &= \frac{\partial}{\partial\theta_j}\frac{1}{2}(h_\theta(x) - y)^2 = (h_\theta(x) - y)\cdot\frac{\partial}{\partial\theta_j}(h_\theta(x) - y)\\ &= (h_\theta(x) - y)\cdot\frac{\partial}{\partial\theta_j}\left(\sum_{i=0}^{n}\theta_i x_i - y\right) = (h_\theta(x) - y)x_j\end{aligned} \tag{11}$$

Getting the gradient for each round of iterations, plus the learrningrate α, gives the complete formula for gradient descent:

$$\theta_j{}' = \theta_j - \alpha(h_\theta(x) - y)x_j \tag{12}$$

Thus the stochastic gradient descent algorithm is obtained:

$$\theta_j{}' = \theta_j + \alpha(y^{(k)} - h_\theta(x^{(k)}))x_j^{(k)} \tag{13}$$

The core idea of SGD is to estimate the gradient by randomly selecting one sample (or a small batch of samples) in each iteration, and to achieve online learning by avoiding the use of the entire dataset by calculating the gradient and updating the weights on each training sample. This makes the computational process more efficient, especially when the dataset is large. In addition, since randomness introduces a certain amount of noise, SGD is able to escape from local minima, which helps the model to explore more of the parameter space.

4. Experimental Results and Discussion

4.1 Data sets and experimental environments

This experiment uses 5011 images from the VOC2007 [20] dataset as the training dataset. A modified Yolov8n model is used to train on this dataset and its accuracy in detecting buses, private cars, pedestrians, trains, bicycles and motorbikes is evaluated. The model parameters are shown in Table 1.

Table 1: The model parameter Configuration

Parameter	Configuration
epochs	150
batch-size	8
imgsz	640
worker	8
optimizer	SGD
lr0	0.01
momentum	0.937
warmup_epochs	3.0
warmup_momentum	0.8
lrf	0.01

The experiments aimed to enhance the algorithms using the PyTorch framework on a local host system and to evaluate the performance of various algorithms. The experimental setup configuration is detailed in Table 2.

ISSN 2616-5775 Vol. 7, Issue 10: 67-75, DOI: 10.25236/AJCIS.2024.071010

Table 2: The configuration of the experimental environment

Parameter	Configuration
CPU	AMD Ryzen 7 5800H w
GPU	NVIDIA GeForce RTX 3060
Operating System	Windows 10
Model	Improved Yolov8n
Acceleration Environment	CUDA 11.1
Language	Python 3.8.0

4.2 Analysis of experimental results and comparison of algorithms

The evaluation metrics of this model mainly include the metrics of mean accuracy rate (mAP), frames per second (FPS), precision rate (P), recall rate (R), etc., where the formulas of P, R and mAP are shown in the following equation:

$$Precision = \frac{TP}{TP + FP} \tag{14}$$

$$Recall = \frac{TP}{TP + FN} \tag{15}$$

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_i \tag{16}$$

where TP is the number of correctly classified positive samples, FP is the number of incorrectly classified negative samples, FN is the number of incorrectly classified positive samples, N is the total number of classifications detected, and mAP is the area enclosed under the PR curve, which is a combined evaluation index of P and R."mAP@0.5" denotes the mAP when the IOU (intersection-to-parallel ratio) is 0.5;t1 is the image preprocessing time, t2 is the image inference time, and t3 is the post-processing time, and FPS is the measure of the number of frames of the transmitted image per second, and its value The larger it is, the faster it represents the detection speed of the network.

In this paper, 511 images are divided into sequence set and val set in 8:2 ratio.A modified Yolov8n algorithm is used to detect buses, private cars, pedestrians, trains, bicycles and motorcycles categories in the dataset. The improved algorithm outperforms both the original algorithm.The accuracy of the algorithm for different categories is shown in Table 3, where model 1 is Yolov8n, model 2 is Yolov8n + SIOU + Adamx, model 3 is Yolov8n + SIOU + SGD, model 4 is Yolov8n + WIOU + Adamx and model 5 is Yolov8n + WIOU + SGD.

Table 3: Recognition accuracy of different models for each class

	Data	Model 1	Model 2	Model 3	Model 4	Model 5
Train	P	85.6%	67.9%	80.1%	79.6%	84.7%
	R	76.9%	60.3%	71.9%	70.1%	79.8%
	mAP@0.5	82.7%	66.2%	78.6%	79.0%	83.3%
	Bus	89.70%	72.90%	86.20%	84.10%	90.50%
	Car	91.40%	85.20%	89.80%	90.80%	92.50%
	Person	88.20%	81.40%	86.00%	86.90%	89.80%
	Train	86.30%	76.10%	82.80%	86.10%	89.90%
	Bicycle	89.30%	77.50%	88.10%	87.30%	91.50%
	Motorbike	88.90%	73.20%	86.00%	84.10%	90.70%
Val	P	81.2%	65.7%	81.3%	80.5%	80.6
	R	70.4%	61.3%	71.4%	69.7%	71.0%
	mAP@0.5	78.8%	63.4%	75.5%	75.8%	80.9%
	Bus	84.30%	69.90%	83.10%	82.50%	88.20%
	Car	88.70%	83.60%	86.10%	87.30%	90.10%
	Person	85.50%	76.20%	84.30%	81.70%	86.40%
	Train	83.10%	72.50%	80.20%	83.30%	87.40%
	Bicycle	84.20%	74.30%	86.70%	82.90%	89.30%
	Motorbike	84.60%	71.90%	83.60%	81.70%	88.50%

The data in the above table indicates that the enhanced Yolov8n + WIOU + SGD algorithm shows improvements across the categories of bus, private car, pedestrian, train, bicycle, and motorcycle, outperforming the other three combined algorithms. According to Table 3, this improved model achieves the best performance compared to the other three models, with Yolov8n + WIOU + Adamx coming in second. Figure 6 illustrates the model's accuracy for each category, demonstrating an increase in recognition accuracy across the board.

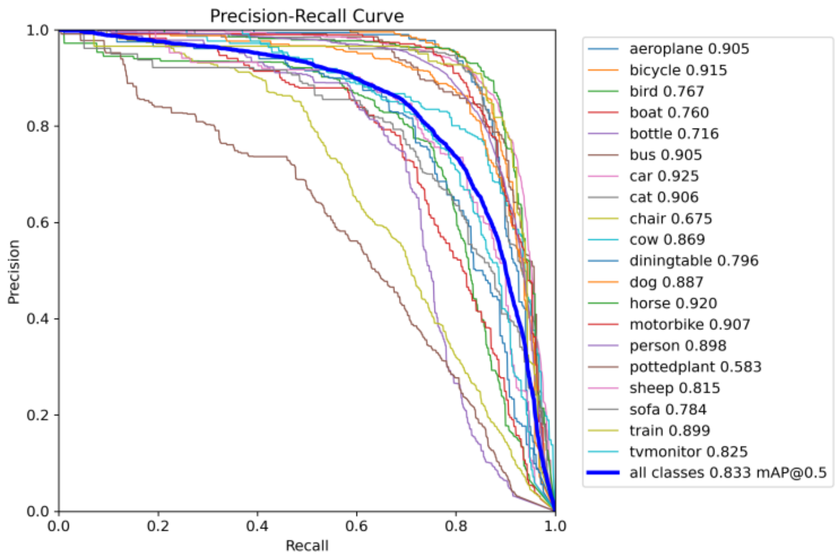

Figure 6: The accuracy of each class in the improved Yolov8n

The improved Yolov8n algorithm achieved higher performance than the original Yolov8s algorithm, which may be due to several factors. Firstly, the VOC2007 dataset contains images taken under different weather and lighting conditions, which makes detecting objects challenging due to complex lighting and blurred targets.

The Wise IoU loss function discards the aspect ratio penalty from CIoU and balances the influence of both high-quality and low-quality anchor boxes on model regression; it further improves the segmentation performance by taking into account the degree of overlap between different regions and accurately masking the boundaries, which enhances the model's generalisation ability and improves the overall performance of the model, and the model's detection accuracy of the traffic target is significantly improved.

Replacing the original optimiser with the SGD optimiser is a further improvement to the underlying Yolov8n model. Compared with other traditional optimisers, SGD usually performs computation and convergence faster and requires fewer parameters, which greatly decreases the time and resources needed for training, enhancing training efficiency. In addition to this, it is also able to automatically escape from saddle points and poorer local optimal points, further improving the training accuracy. In addition to this, it is also able to automatically escape saddle points and poorer local optimal points to further improve training accuracy. In the target detection task, the model's rapid convergence enables it to quickly learn effective features for target detection, thus improving the final accuracy.

5. Conclusions

To address the issues of low detection accuracy and efficiency in Yolov8s, this paper introduces an improved Yolov8n algorithm, which is characterised by high accuracy and a small number of parameters. Specifically, to address the problem of more parameters in the basic Yolov8s, its variant Yolov8n algorithm is introduced to improve the efficiency of model training. Subsequently, to address the problem that the original CIOU function cannot be accurately localised, the WIOU loss function is introduced to consider the degree of overlap between different regions and accurately mask the boundaries, which further improves the segmentation performance. Finally, the SGD optimiser is introduced on the basis of the above to improve the efficiency of model training.

The proposed improved Yolov8n algorithm is applied to the VOC2007 dataset for training and validation. The experimental results show that the proposed improved Yolov8n algorithm has high

ISSN 2616-5775 Vol. 7, Issue 10: 67-75, DOI: 10.25236/AJCIS.2024.071010

training accuracy with fewer parameters and high training efficiency, reaching 82.3 % for P, 78.2 % for R, and 83.3 % for map on the dataset, which has high detection accuracy and verifies the effectiveness of the model. Compared with other models, the number of parameters is reduced, and the map after detection is increased by 0.6 % to 17.1 %, and the categories of buses, private cars, pedestrians, trains, bicycles, and motorcycles are significantly improved. The superiority of the model has been verified.In summary, the improved yolov8n algorithm proposed in this paper is able to meet the functional requirements for high-precision and high-efficiency detection of road traffic equipment and traffic targets in self-driving traffic roads.

References

[1] Jiang H. Overview of autopilot technology in industry[J]. Logistics technology and application, 2024, 29(03): 106-107.
[2] Kamal U, Tonmoy T I, Das S, et al. Automatic traffic sign detection and recognition using SegU-Net and a modified Tversky loss function with L1-constraint[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 21(4): 1467-1479.
[3] Girshick R, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[J]. Computer Vision and Pattern Recognition , 2014(1): 580-587.
[4] Girshick R. Fast R-CNN[C]. Proceedings of the IEEE international conference on computer vosion, 2015: 1440-1448.
[5] Liu X, Cheng J, Cheng Y, et al. Distance posts detection and character sequences recognition method in video images acquired from camer in moving vehicle[J]. Computer Engineering and Applications, 2023, 59(8): 175-81.
[6] Liu H, Li J. Vehicle type detection algorithm based on YOLOv5s model[J]. Journal of Lang fang Normal College (Natural Science Edition), 2022, 22(3): 24-28.
[7] Li S, Shi T, Jing F. Improved road damage detection algorithm for YOLOv8[J]. Computer Engineering and Application, 2023, 59(23): 165-174.
[8] Kingma D, Ba J. Adam: A method for stochastic optimization in: Proceedings of the 3rd international conference for learning representations (iclr'15)[J]. SanDieg0, 2015, 500.
[9] Wang S, Ton W, Huang J, et al. Tile surface defect detection based on YOLOv5[J] . Packaging Engineering, 2022, 43(09): 217-224.
[10] Sun Y, Li W, Zhou N. Study on improved YOLOv5m method for cargo category detection[J]. Computer Engineering and Applications, 2023, 59(16): 205-211.
[11] Shang C, Wang M, Ning J, et al. Identification of high similarity dairy goats under joint loss optimization[J]. Chinese Journal of Image and Graphics, 2022, 27(04): 1137-1147.
[12] Safaldin M, Zaghden N, Mejdoub M. Moving object detection based on enhanced Yolo-V2 model[C]. IEEE, 2023: 1-8.
[13] Wang C, Bochkovskiy A, Liao H. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]. Vancouver: IEEE, 2023: 7464-7475.
[14] Lin T, Dollar P, Girshck R, et al. Feature pyramid networks for object detection[C]. Proceedings of the IEEE: 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), IEEE, 2017: 936-944.
[15] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]. Proceedings of the IEEE: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). IEEE, 2018: 8759-8768.
[16] Zhang Y, Ren W, Zhang Z, et al. Focal and efficient IoU loss for accurate bounding box regression [J]. Neurocomputing, 2022, 506: 146-157.
[17] Zheng Z, Wang P, Ren D, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation[J]. IEEE Trans-actions on Cybernetics, 2022, 52(8): 8574-8586.
[18] Li M, Xiao Y, Zong W, et al. A lightweight chestnut fruit recognition method based on improved Y0L0v8 model[J]. Journal of Agricultural Engineering, 2024, 40(1): 201-209.
[19] Jian W, Gauri J. Cooperative SGD: A unified framework for the design and analysis of local-update SGD algorithms[J]. Journal of Machine Learning Research, 2021, 22(1): 213-223.
[20] Evringham M, Van C, Williams C, et al. The pascal visual object classes (voc) challenge[J]. International journal of computer vision, 2010, 88(2): 303-338.