Design of Fish School Behavior Pattern Recognition Model SPD-YOLOv10n

Hanlin XU[1], Shiyu WU[2], Guochao DING[1*]

1. College of Information and Electrical Engineering, Heilongjiang Bayi Agricultural University, Daqing 163319, China; 2. Harbin Sandi Industrial Control Engineering Co., Ltd., Harbin 150001, China

Abstract A common but flawed design in existing CNN architectures is using strided convolutions and/or pooling layer, which will result in the loss of fine-grained feature information, especially for low-resolution images and small objects. In this paper, a new CNN building block named SPD-Conv was used, which completely eliminated stride and pooling operations and replaced them with a space-to-depth convolution and a non-strided convolution. Such new design has the advantage of downsampling feature maps while retaining discriminant feature information. It also represents a general unified method, which can be easily applied to any CNN architectures, and can also be applied to strided conversion and pooling in the same way.

Key words Fish; Group behavior; Behavior recognition; Deep learning; YOLOv10

DOI:10.19759/j.cnki.2164-4993.2025.01.018

The YOLO (You Only Look Once) model family[1] has been widely used in the field of fish school behavior recognition because of its efficient target detection performance. Although YOLOv10[2] has higher detection accuracy and faster reasoning speed than previous models, it still faces certain challenges in fish school behavior pattern recognition, especially in low-resolution images. In this study, in order to solve these problems, SPD-YOLOv10n, an improved model based on YOLOv10, which further optimized the feature extraction module and significantly improved the detection precision and calculation efficiency of the model in complex water environment, was proposed. Moreover, SPD-YOLOv10n further improved the adaptability of the model to dynamic fish school behavior by enhancing the training data and improving the loss function.

Design of SPD-YOLOv10n model

SPDConv module

The SPDConv layer enhances the representation capacity of spatial pattern by special spatial division (downsampling operation) on input tensor[3]. Through four different spatial slices of the input tensor, the spatial structure information of the input image in different regions was captured. Then, the information was spliced together to provide richer spatial context information for subsequent convolution operations. The SPD component downsampled the feature maps inside CNN and across the whole CNN as shown below. When considering the intermediate feature map X with an arbitrary size of S × S × C1, the sub-feature map sequences were sliced as following:

$f_{0,0} = X[0:S:scale, 0:S:scale]$, $f_{1,0} = X[1:S:scale, 0:S:scale]$, ...,

$f_{scale-1,0} = X[scale-1:S:scale, 0:S:scale]$;

$f_{0,1} = X[0:S:scale, 1:S:scale]$, $f_{1,1}$, ..., $f_{scale-1,1} = X[scale-1:S:scale, 1:S:scale]$; ...,

$f_{0,scale-1} = X[0:S:scale, scale-1:S:scale]$, $f_{1,scale-1}$, ...,

$f_{scale-1,scale-1} = X[scale-1, S:scale, scale-1:S:scale]$

The schematic diagram of SPD-Conv is shown in Fig. 1.

Improved SPD-YOLOv10n

The structure of optimized SPD-YOLOv10n algorithm was obtained by replacing the original convolution with an SPD-Conv building block. The structure diagram of the optimized SPD-YOLOv10n algorithm is shown in Fig. 2.

Performance Evaluation of SPD-YOLOv10n Fish School Behavior Pattern Recognition Model

Multi-index analysis of model training verification process

The training loss (train/box_loss, train/cls_loss, train/dfl_loss)[4] and verification loss (val/box_loss, val/cls_loss, val/dfl_loss)[5] in the loss indicators showed a continuous downward trend in the training and verification process, indicating that the model was constantly improving and the prediction error was gradually decreasing. Metrics/precision(B) (precision) and metrics/recall(B) (recall)[6] in the classification indicators rose rapidly at the initial stage of model training, and then tended to be stable, and their values were close to 1.0. The model had high precision and recall. MAP (average precision) metrics/mAP50(B) (mAP@0.5) and metrics/mAP50-95(B) (mAP@0.5-0.95)[7] in target detection and segmentation indicators all showed a rapid growth and tended to be stable, indicating that the curve trend of the model still maintained good performance under the stricter evaluation index mAP@0.5-0.95 (Fig. 3).

Analysis of model training results

Through the comparison table of model results, it could be seen that the improved SPD-YOLOv10n was greatly improved compared with the original model (Table 1).

Received: November 16, 2024 Accepted: January 19, 2025

Hanlin XU (1999 -), male, P. R. China, devoted to research about fishery digitization.

* Corresponding author.

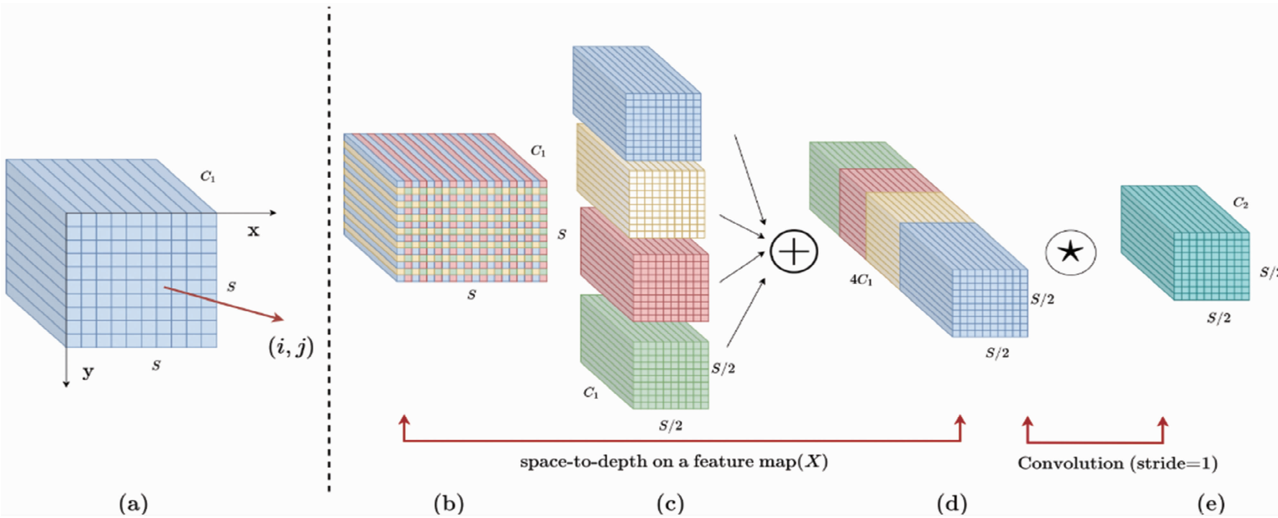

Fig. 1 Schematic diagram of SPD-Conv

Table 1 Comparison of model results

Model	P	R	MAP50	MAP50-95
YOLOv10n	0.904	0.850	0.852	0.891
SPD-YOLOv10n	0.908	0.854	0.859	0.906

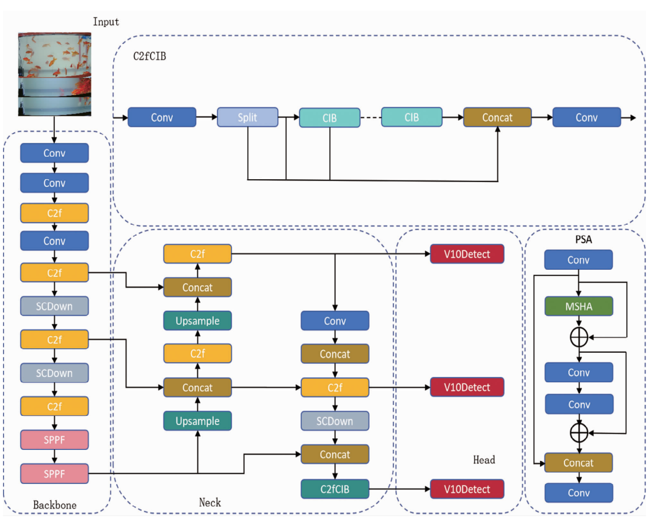

Fig. 2 Structure diagram of optimized SPD-YOLOv10n algorithm

As shown in Table 1, the accuracy of YOLOv10n was 0.904, and that achieved by SPD-YOLOv10n was 0.908, showing an increase of 0.004. The variation showed that the SPD-YOLOv10n model was improved slightly in reducing false positive and could detect the targets more accurately. The recall of YOLOv10n was 0.850, and that of SPD-YOLOv10n was 0.854, showing an increase of 0.004. The improvement of recall meant that the SPD-YOLOv10n model could identify more true positive when detecting targets, and reduce false positive. In terms of the mAP50 index, the value of YOLOv10n was 0.852, and that of SPD-YOLOv10n increased to 0.859, an increase of 0.007. This improvement meant that SPD-YOLOv10n could provide better positioning and matching in target detection, especially when IoU was greater than 0.5. From the perspective of the mAP50-95 index, the value of YOLOv10n was 0.891, and that of SPD-YOLOv10n was 0.906, increasing by 0.015. This increase was more obvious under a strict IoU threshold, which showed that the SPD-YOLOv10n model had improved its ability to accurately locate targets and was adapted to higher accuracy requirements. Compared with YOLOv10n, SPD-YOLOv10n was improved to a certain extent in various evaluation indexes (accuracy, recall, mAP50 and mAP50-95), especially in positioning accuracy (improvement of mAP50-95). Although the improvement ranges seem small, in practical application, these minor improvements can significantly improve the reliability and robustness of the model, especially in more complex detection tasks. It shows that the SPD-YOLOv10n model can provide higher detection accuracy while maintaining low computational complexity, and is suitable for efficient deployment in the environment with limited resources.

In the early stage of training, the loss of YOLOv10n-improve rose slightly faster than that of YOLOv10n, but after about 150 epoches, the loss of YOLOv10n increased rapidly, while YOLOv10n-improve maintained a more stable growth trend. YOLOv10n-improve was more stable in the convergence process than YOLOv10n, which might indicate that the improved algorithm optimization strategy was more effective and avoided over-fitting or gradient explosion. The loss value of YOLOv10n fluctuated and rose greatly in the later period, which might reflect the instability of training. The improved algorithm had obvious advantages in this respect. The improved model exhibited a consistent improvement in the performance of mAP@0.5, especially in the later stage of training (after 150 epoches). Both of them quickly reached a high mAP value (>0.7) in the first 50 epoches, and were then gradually stabilized. In the later period of training (200 – 300 epoches), the mAP curve of YOLOv10n-improve was obviously more stable than YOLOv10n, showing that it had stronger generalization ability and was not easily affected by data noise or over-fitting. The improved model was also superior to YOLOv10n in the performance of mAP@0.5:0.95, especially in the later period (200 – 300 epoches). mAP@0.5:0.95 considered the average precision under multiple IoU thresholds, which could reflect the robustness of the detection model in different target sizes and positions. The improvement of this index showed that the improved algorithm had significantly optimized performance in multi-scale target detection and fine-grained prediction. The curves of Loss, mAP@0.5 and mAP@0.5:0.95 are shown in Fig. 4.

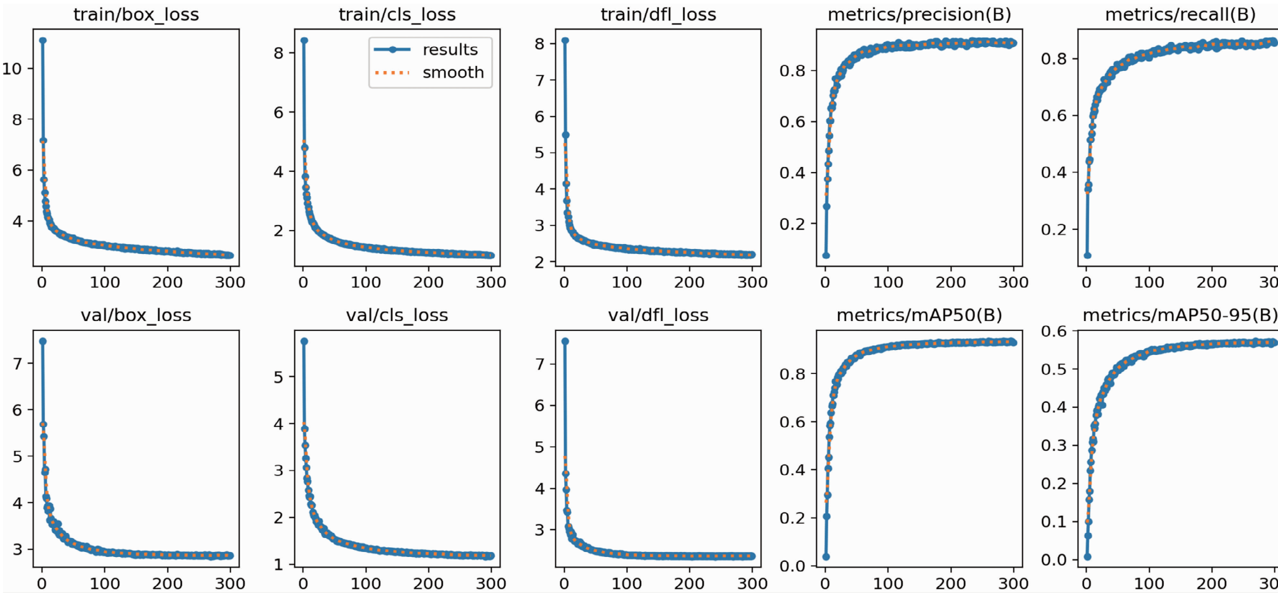

Fig. 3 Curves of multiple indexes in training and verification

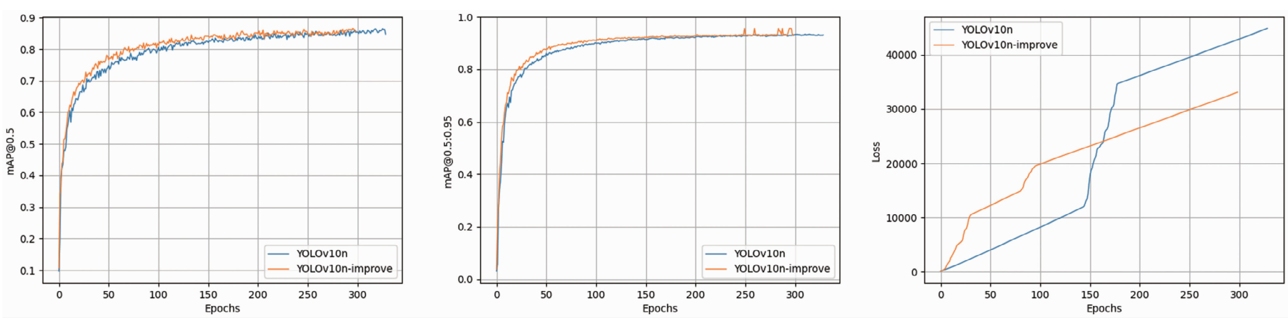

Fig. 4 Curves of Loss, mAP@0.5 and mAP@0.5:0.95

Conclusions and Discussion

In this study, the improved SPD-Conv convolution structure was used to replace the feature extraction layer in the traditional YOLOv10 model, and a new efficient feature extraction module was constructed, which effectively improved the feature expression capacity and the overall performance of the model. The experimental results on the self-built data set showed that the improved SPD-YOLOv10n was improved to a certain extent compared with YOLOv10n in terms of various evaluation indexes (precision, recall, mAP50 and mAP50-95), especially in positioning precision (improvement of mAP50-95). Although the improvements may be insignificant, in practical application, these small improvements can significantly enhance the reliability and robustness of the model, especially in more complex detection tasks.

References

[1] REDMON J, DIVVALA S, GIRSHICK R, *et al*. You only look once: Unified, real-time object detection[C/OL]//2016 IEEE Conference on computer vision and pattern recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016: 779-788.

[2] TIAN Q, HUO Y, YAO M, *et al*. A method for detecting dead fish on large water surfaces based on improved YOLOv10[J]. arXiv preprint arXiv:2409.00388, 2024.

[3] GU Z, ZHU K, YOU S. YOLO-SSFS: A method combining SPD-Conv/STDL/IM-FPN/SIoU for outdoor small target vehicle detection[J]. Electronics, 2023, 12(18): 3744.

[4] WITT J, RASING S, DUMANČIĆ S, *et al*. A divide-align-conquer strategy for program synthesis[A/OL]. arXiv, 2023[2024-12-19]. http://arxiv.org/abs/2301.03094.

[5] ZHAO ZQ, ZHENG P, XU ST, *et al*. Object detection with deep learning: A review[J/OL]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(11): 3212-3232.

[6] WANG R, WANG Z, XU Z, *et al*. A real-time object detector for autonomous vehicles based on YOLOv4[J/OL]. Computational Intelligence and Neuroscience, 2021, 2021(1): 9218137.

[7] WU C, WANG D, WU M, *et al*. Real-time detection and identification of fish skin health in the underwater environment based on improved Yolov10 model[J]. Available at SSRN 5044386.

Editor: Yingzhi GUANG Proofreader: Xinxiu ZHU