ITC 1/54
Information Technology and Control
Vol. 54 / No. 1/ 2025
pp. 185-197
DOI 10.5755/j01.itc.54.1.37841

YOLOv8-SS: A Method of Localizing Soldiers in Intricate Battlefield Environments

Received 2024/06/30 | Accepted after revision 2024/12/16

HOW TO CITE: Gao, Y., Wang, Y. (2025). YOLOv8-SS: A Method of Localizing Soldiers in Intricate Battlefield Environments. *Information Technology and Control*, 54(1), 185-197. https://doi.org/10.5755/j01.itc.54.1.37841

YOLOv8-SS: A Method of Localizing Soldiers in Intricate Battlefield Environments

Yunlong Gao, Yongjuan Wang

School of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing 210094, Jiangsu, China

Corresponding authors: 18936030961@189.cn

As combat becomes more autonomous and intelligent in the future, and effective military target localization techniques are essential to understanding operational military deployment and target tracking. In this paper, we offer an instance segmentation technique for precise soldier localization in intricate battlefield environments, called YOLOv8-SS. First, in the YOLOv8 backbone network, the C2f module is replaced by the Dual-C2f module, which we created based on DualConv in order to minimize the amount of parameter computation while maintaining accuracy. Second, the feature extraction network is enhanced by import the global attention mechanism (GAM), which increases the cross-dimensional interaction between the channel and spatial information and boosts the model's feature extraction performance. Lastly, the reparameterization module DBB is used to redesign the segmentation head of YOLOv8. Convolutional branches of various sizes and shapes are added to the network's feature representation capacity during the training phase. In the inference phase, the convolutional branches are equivalently replaced with regular convolutional, which increases accuracy while maintaining inference efficiency. Additionally, a dataset for segmenting soldier instances include various battlefield situations is provided in this paper, and experimental validation is carried out using this dataset. The experimental results demonstrate that YOLOv8-SS improves the Box P, Box mAP50, and Box mAP50-95 measures by 2.7%, 2.9%, and 5.1%, Mask P, Mask mAP50, and Mask mAP50-95 improved by 0.7%, 1.7%, and 4.6%,. however, Box R and FPS decreased slightly, by 1.6% and 8.6% in comparison to the baseline model YOLOv8n. As a result, the YOLOv8-SS model performs more accurately when it comes to segmenting soldiers in intricate battlefield environments.

KEYWORDS: YOLOv8; Instance Segmentation; DualConv; Global Attention Mechanism; Diverse Branch Block.

1. Introduction

The form of warfare is evolving towards more intelligence, and unmanned intelligent weapons will play a significant role in the combat system of the future [6]. In this context, the reconnaissance of battlefield posture is a prerequisite for the implementation of effective fire strikes by unmanned intelligent weapons. Tar-get tracking, target precision guidance, and combat situational analysis all depend on accurate and effective military target recognition in complicated battlefield conditions [19]. The effectiveness of firepower is one of the most important factors in winning modern wars. Then, a crucial technology for battlefield situational detection is the identification and location of soldiers in the combat zone [15]. The terrain of the battlefield is unstable and complex, and military objectives are always obstructed by smoke, fire, and other elements. Additionally, their position is constantly obscured by forested areas, mountains, fields, and other complex backdrops [28]. It is very challenging to achieve accuracy in military target detection under the influence of these circumstances [27].

Image-based target detection techniques are being progressively used in the hunt for combat targets as computer vision technology advances. Deep learning has advanced significantly since the proposed of AlexNet [11], and methods like target detection have started to develop. Convolutional neural networks were first used for target identification with R-CNN [16], from which Fast R-CNN [12, 26] was created. A lot of its derivative models are employed in the field of military target detection, such as the detection of knives [8] and weapons [29]. However, the above-mentioned algorithms still have limitations in terms of detection speed. Until the emergence of the series of YOLO [1, 25, 30], deep learning has been more widely used in the field of real-time detection. In the field of military target detection, the YOLO [8] algorithm also achieved better results. Additionally, vision Transformers have become the latest method for computer vision because of its excellent speed performance [5, 9, 34, 35].

The main localization methods for humanoid targets such as soldiers are detection, pose detection [23, 24], and instance segmentation. The dataset for target recognition comprises more interference information be-cause of the variety of tactical maneuvers of the soldiers and the soldier's camouflage proximity to the combat environment [17]. The instance segmentation target localization approach can produce more precise outcomes. Image instance segmentation involves not only instance localization but also pixel-level classification, including semantic segmentation and object detection. Mask RCNN [14] and subsequent YOLACT [2, 20, 36], SOLO [31, 32, 33], and FastInst [13] are examples of instance segmentation algorithms that have developed over time and have increased computing efficiency and accuracy. In recent years, the YOLO series has been used to improve instance segmentation algorithms as a baseline model [3, 4, 16, 21].

To solve the shortcomings of the YOLOv8 method in instance segmentation, together with the features of soldiers in intricate battlefield environments, an improved YOLOv8 algorithm (i.e., YOLOV8-SS) is provided in this paper. The algorithm's primary contributions are as follows:

1. In the YOLOv8 network structure, the DualConv module is used to design the DualC2f module to replace the C2f module. It reduces the network model's parameters and computational complexity without sacrificing accuracy, making the model lighter.
2. Adding the global attention mechanism (GAM) to the feature extraction net-work, which strengthens the channel and spatial information cross-dimensional inter-action and increases the model's capacity to extract meaningful features.
3. The reparameterization module DBB was used to redesign the YOLOv8 seg-mentation head. Convolutional branches of various sizes and shapes are added to the network's feature representation capacity during the training phase. In the inference phase, the convolutional branches are equivalently replaced with regular convolutional, which increases accuracy while maintaining inference efficiency.

This paper's outline is structured as follows: Section 2 describes the improved YOLOv8 model, YOLOV8-SS, and Section 3 describes the experimental details and evaluation metrics. The analysis of the experimental data, including the ablation experiment and the experiment that was compared with other models, is developed in Section 4. A conclusion to the paper is given at the end.

2. Improvements Based on YOLOv8

It was discovered that Yolov8 still have potential to increase effectiveness for the segmentation task of soldiers in intricate battlefield environments. Because of this, YOLOV8-SS is suggested in this study as an improvement over Yolov8, and its network topology is shown in Figure 1. The backbone, neck, and head make up the three main components of the YOLOV8-SS network. The backbone extracts the feature information from the input image; the neck is used to fuse the features that the backbone has extracted; and the head outputs the segmentation results. This paper designs the DualC2f module to replace the C2f module in the backbone and neck in order to minimize the amount of parameter computation without sacrificing feature information. A GAM module is also added to the backbone section to improve the representation of the input features and, hence, increase the detection accuracy while also retaining more feature information. Lastly, the reparameterization module DBB introduces the segment head.

2.1. DualC2f

This paper's DualC2f module is designed based on the DualConv to improve the C2f module from the original YOLOv8 architecture [37]. The DualC2f module has less parameter computation than the original C2f module while maintaining a suitable amount of feature information. The structure of DualC2f is shown in Figure 2.

When the feature map is input to the DualC2f module, it passes through a Conv module, which consists of a Conv2d with a kernel size of 1 × 1 and a step size of 1, a BN layer, and a Silu activation function. These convolution operations help to extract features at different levels in the input data.

Then the input data is separated into two processing branches. While the Dual-Bottleneck module processes the other branch, one branch is sent straight to the output. By improving the network's representational and nonlinear capacities, this branching design helps the network better represent complicated data. The

Figure 1
Network structure of YOLOV8-SS

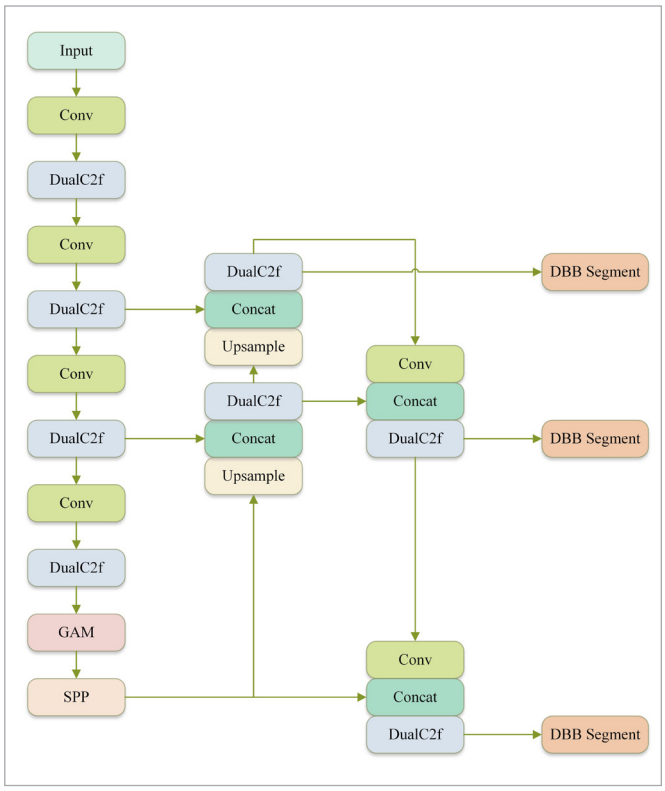
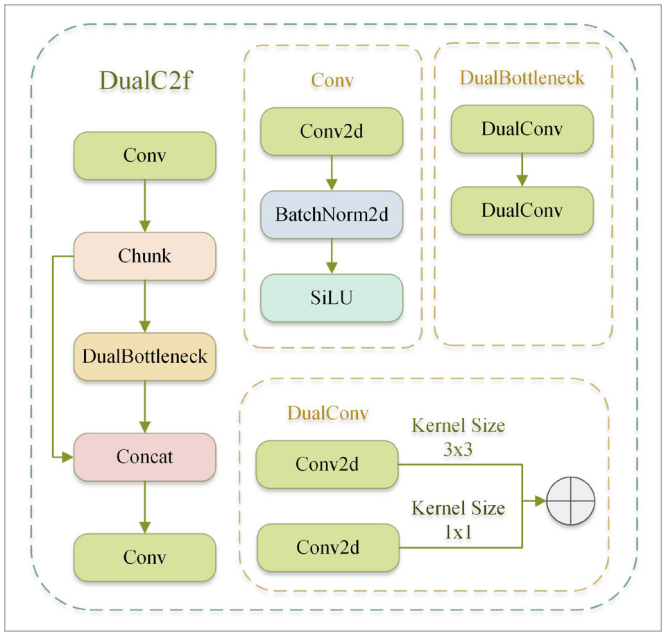


Figure 2
Structure of DualC2f



DualBottleneck mod-ule is composed of two successive DualConv modules. DualConv entails applying a 1×1 point-by-point convolution and a 3×3 set of convolutions to the same input feature maps, followed by their summation. Since applying successive 1×1 convolutions to the input feature map preserves the original information, it can help deeper convolutional layers extract information more efficiently.

Finally, the DualC2f module realizes feature fusion by splicing features from different branches in the channel dimension. The spliced features contain information from different branches, enriching the expressiveness of the features. DualC2f fuses features from two dimensions, which enhances the representation of spatial feature information and helps to effectively distinguish texture features for military targets in complex backgrounds.

2.2. GAM

Numerous research studies have shown how attentional mechanisms can im-prove performance on a range of computer vision tasks. Nonetheless, the conventional attention mechanism grounded in convolutional neural networks primarily concentrates on the examination of the channel domain, taking into account solely the inter-play among feature map channels. This undermines the significance of augmenting cross-dimensional interactions with respect to the preservation of both channel and spatial information. This work adds a global attention module, GAM (Global Attention Module) [22], in the last layer of the feature extraction network, as illustrated in Figure 3, to further improve the model's capacity to extract valuable features. In order to enhance the model's focus on significant features, this module attempts to take into con-sideration the data in both the channel and spatial dimensions. The addition of the GAM module improves feature representation by enabling the model to more precisely express the correlations between various locations and channels. The GAM module enhances the representation of boundary information between the target and the environment in an image by focusing on both channel information and spatial information.

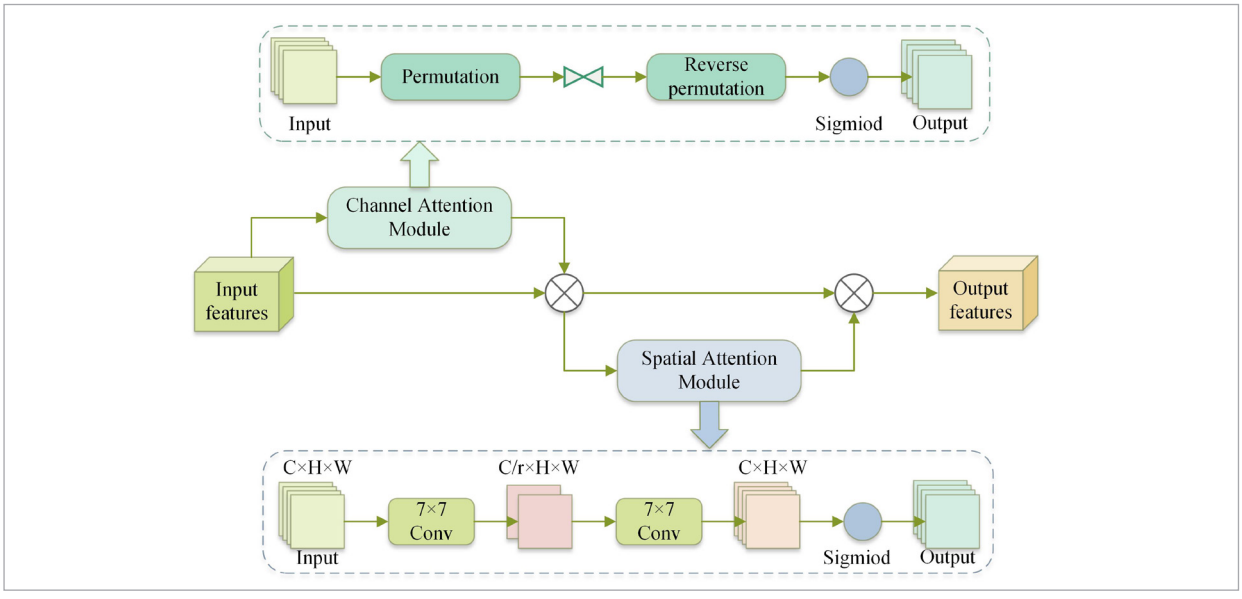
GAM contains two modules, channel attention module and spatial attention. The whole process is shown in Figure 1, given the input feature map $F \in R^{C \times H \times W}$, the intermediate state F_2 and the output F_3 are defined as shown in Equations (1)-(2):

$$F_2 = M_c(F_1) \otimes F_1 \tag{1}$$

$$F_3 = M_s(F_2) \otimes F_2, \tag{2}$$

where M_c and M_s are channel and spatial attention maps, respectively, and $\otimes$ denotes multiplication by elements.

Figure 3
Structure of GAM



The input feature maps are first dimensionally transformed in the channel attention sub-module. After that, they are fed into the MLP (Multi-Layer Perceptron) to amplify the channel-space dependence across dimensions. Lastly, they are transformed back to their original dimensions and the Sigmoid function processes them for output. Two convolutional layers are utilized in the spatial attention sub-module to integrate the spatial information and focus on it. Additionally, the reduction rate (r) is the same as that of the BAM channel attention sub-module.

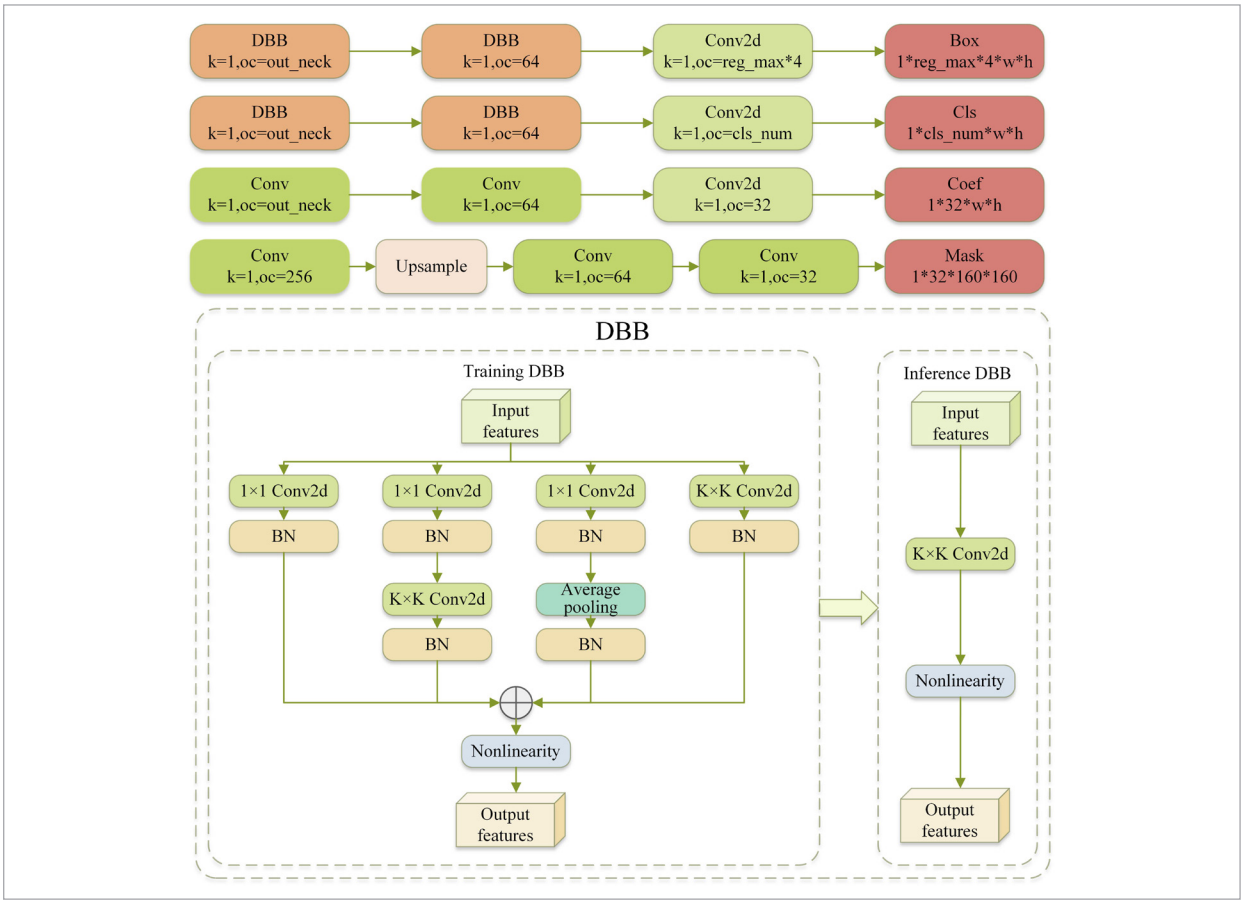
2.3. DBB Segment

The reparameterization module DBB [7] was used to redesign the YOLOv8 seg-mentation head. Convolutional branches of various sizes and shapes are added to the network's feature representation capacity during the training phase. In the inference phase, the convolutional branches are equivalently replaced with regular convolutional, which increases accuracy while maintaining inference efficiency. The structure of DBB segment is shown in Figure 4.

This paper's DBB module employs the following four convolutional deformation techniques in total:

1. Convolutional kernel size 1 × 1 convolutional kernel size conv2d module is connected to the BN module to create;
2. Passing through a module identical to the deformation (1), and then inputs to a conv2d module with convolutional kernel size k × k connected to a BN module;
3. Passing through a module identical to deformation (1) and then feeding into an average pooling module with a BN module;

Figure 4
Structure of DBB segment



4 A conv2d module with a convolutional kernel size of k × k is coupled to the BN module to build the composition.

Following the summation of the four deformation outcomes and nonlinear activation, the feature map is the finally output. In order to improve the feature representation during the inference phase without significantly increasing the number of parameters, regular convolution can be utilized in place of distorted convolution combinations.

3. Experimental Process

3.1. Dataset

The battlefield is always incredibly unpredictable and complex in armed conflict. Because military targets are constantly surrounded by a variety of intricate back-grounds, it might be challenging to identify them quickly. Furthermore, military tar-gets can be concealed in a variety of landscapes, including plains and woodlands, and harder to locate. In order to solve this issue, a soldier segmentation dataset consisting of 820 images is created, as illustrated in the Figure 5, by examining and evaluating various forms of complex background interference in actual combat environments, including night, jungle, mountain, and city. The dataset is split into a training set, a vali-dation set, and a test set in the ratio of 5:3:2 to guarantee an adequate number of samples.

3.2. Experimental Details

All experiments in this paper were conducted on a workstation with the Ubuntu 22.04 operating system. The graphics processor used is a GeForce RTX 3090, while the training and test data are derived from the soldier dataset constructed in this paper.

To ensure consistency in the training process for both ablation and comparison tests, we used uniform hyperparameters for training. Specifically, the batch size is 16, the maximum epoch is set to 800, and the patience is 60. In addition, the loss function is the same as the baseline model.

3.3. Performance Metrics

In this paper, the segmentation performance of the YOLOv8-SS model is evaluated using generalized evaluation criteria. In the evaluation, recall shows the percentage of correctly identified targets, whereas precision measures the model's accuracy in identifying and segmenting targets. Furthermore, we employ

Figure 5
Dataset of soldiers: (a) Forest; (b) Mountain and field; (c) City; (d) Night vision; (5) clean back-ground

(a)

(b)

(b)

(d)

(e)

FSP to gauge the model's inference efficiency and mAP to assess target recognition and segmentation precision in a comprehensive manner. Equations (3)-(6) below illustrate how they are calculated. The performance of the YOLOv8-SS model in the soldier segmentation task can be more thoroughly evaluated by thoroughly examining these assessment measures.

$$P = \frac{TP}{TP + FP} \tag{3}$$

$$R = \frac{TP}{TP + FN} \tag{4}$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{5}$$

$$FPS = \frac{1}{t} \tag{6}$$

4. Results

4.1. Ablation Experiment

To verify the validity of each improvement point, we performed ablation experiments on a self-constructed soldier segmentation dataset to complete the evaluation. To ensure the fairness of the assessment, we set the same parameters for each variable.

As can be seen in Table 1, compared to the baseline model YOLOv8n, YOLOV8-SS improved by 2.7%, 2.9%, and 5.1% in Box P, Box mAP50, and Box mAP50-95 metrics. mask P, Mask mAP50, and Mask mAP50-95 improved by 0.7%, 1.7%, and 4.6%. However, Box R and FPS decreased slightly, by 1.6% and 8.6%, respectively.

The DBB module mainly improved the accuracy and confidence of the model. After removing the DBB module, Box mAP50-95 and Mask mAP50-95 significantly decreased by 1.9% and 2%, respectively, along with a decrease in recall. The GAM module had the greatest impact on mAP, but had a smaller impact on precision and recall. If the DualC2f module is removed, there is a significant decrease in model precision with a value of 5.2%, but there is a 3.3% increase in recall. This shows that the decrease in recall of YOLO-SS is mainly caused by the DualC2f module.

Figure 6 shows a comparison of the results of the ablation experiment of the im-proved model by test on the soldier dataset. The results include Box PR curve, Box F1 curve, Mask PR curve, and Mask F1 curve. As Figure 6 shows, the improved approach completely encircles the curve of the baseline model and is numerically closer to the point (1,1). This illustrates how the improved algorithm performs better than the base-line model and provides several significant advantages.

To provide a more understandable demonstration of the improved model's effectiveness, Figures 7-9 present the processing outcomes of multiple common scenario situations. As demonstrated in Figure 7, the improved model has a higher accuracy and can detect occluded targets, but the baseline model experiences leaky detection when dealing with obscured soldier targets. The soldier is camouflaged to approximate the surroundings in the mountainous setting depicted in Figure 8. The improved model is able to recognize the target, while the baseline model is unable to do so due

Table 1
Results of ablation experiment

Model			Box P	Box R	Box mAP50	Box mAP50-95	Mask P	Mask R	Mask mAP50	Maks mAP50-95	Parameters	Gfloats	FPS
DualC2f	GAM	DBB											
×	×	×	0.892	0.811	0.869	0.616	0.885	0.768	0.841	0.543	3258259	12.0	116
√	√	×	0.891	0.806	0.89	0.638	0.898	0.771	0.863	0.569	4337011	12.0	107
√	×	√	0.917	0.801	0.879	0.626	0.896	0.773	0.845	0.555	4703155	53.9	103
×	√	√	0.867	0.828	0.889	0.641	0.874	0.78	0.845	0.568	6642483	56.0	110
√	√	√	0.919	0.795	0.898	0.657	0.892	0.77	0.859	0.589	4533427	54.7	106

Figure 6
Results of ablation experiment: (a) Box PR curve; (b) Box F1 curve; (c) Mask PR curve; (d) Mask F1 curve

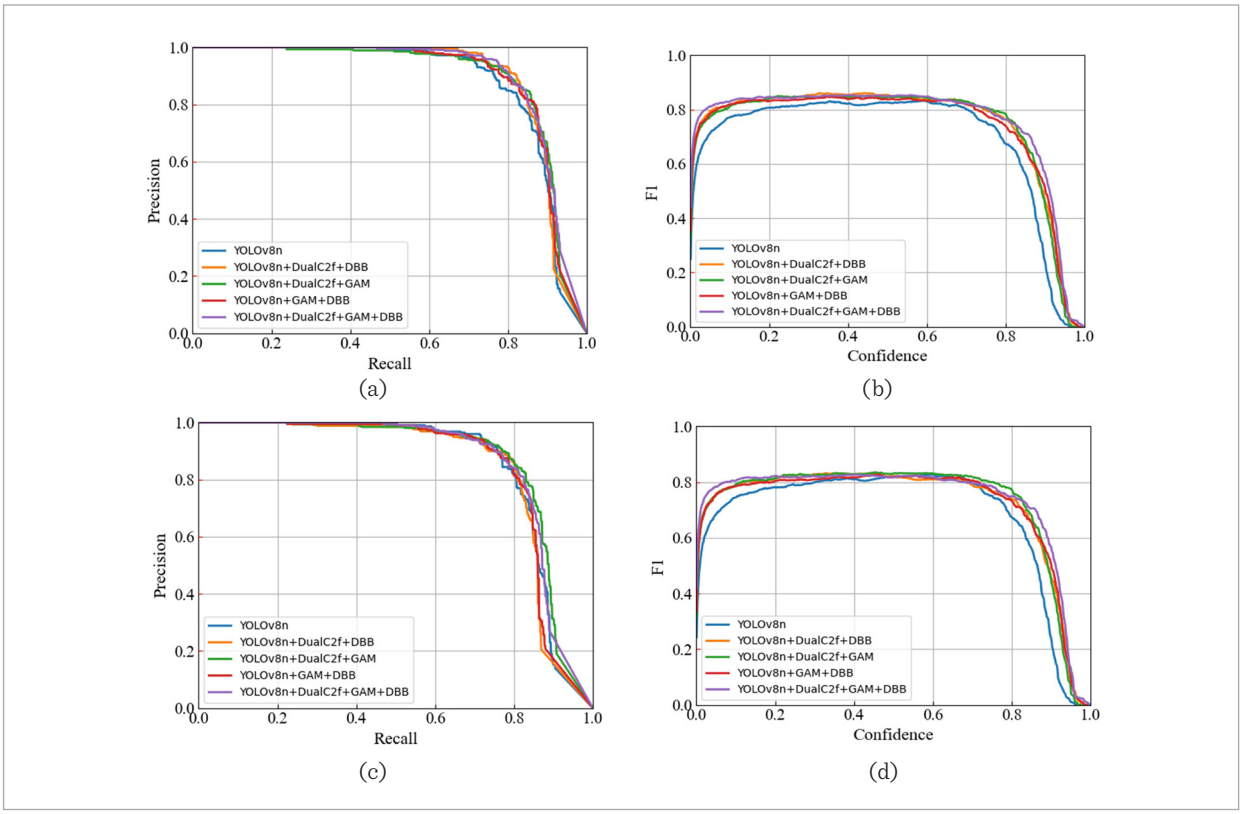


Figure 7
The results of the improved model in scene 1: (a) The original image; (b) YOLOv8n; (c) YOLOv8n+DualC2f+GAM; (d) YOLOv8n+DualC2f+DBB; (e) YOLOv8n+GAM+DBB; (f) YOLOv8n+DualC2f+GAM+DBB

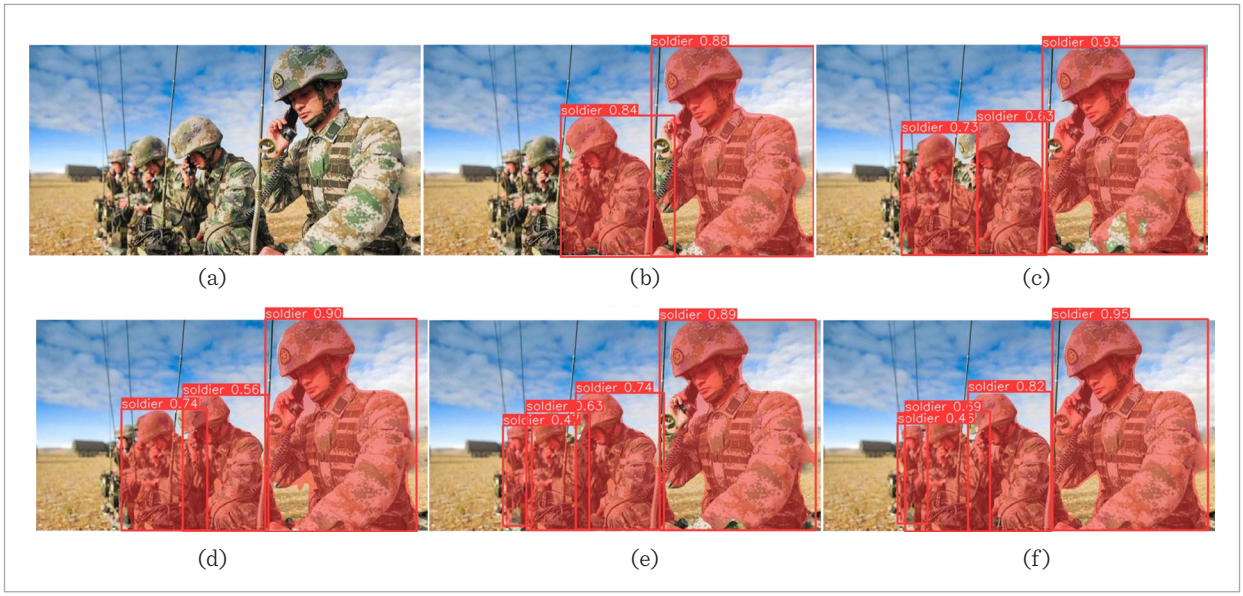
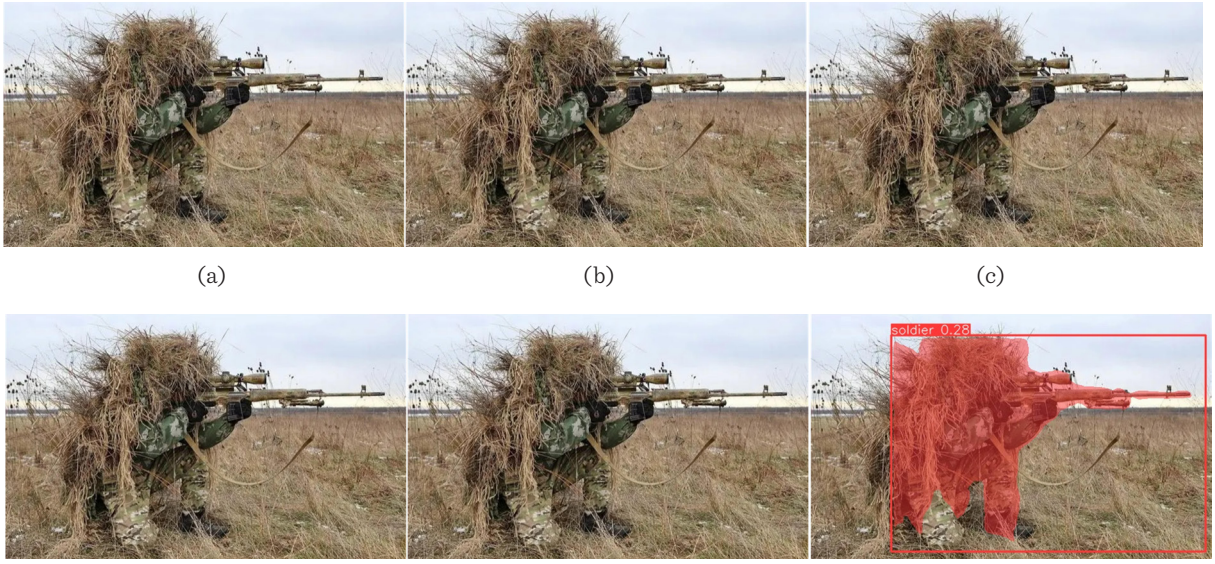


Figure 8

The results of the improved model in scene 2: (a) The original image; (b) YOLOv8n; (c) YOLOv8n+DualC2f+GAM; (d) YOLOv8n+DualC2f+DBB; (e) YOLOv8n+GAM+DBB; (f) YOLOv8n+DualC2f+GAM+DBB

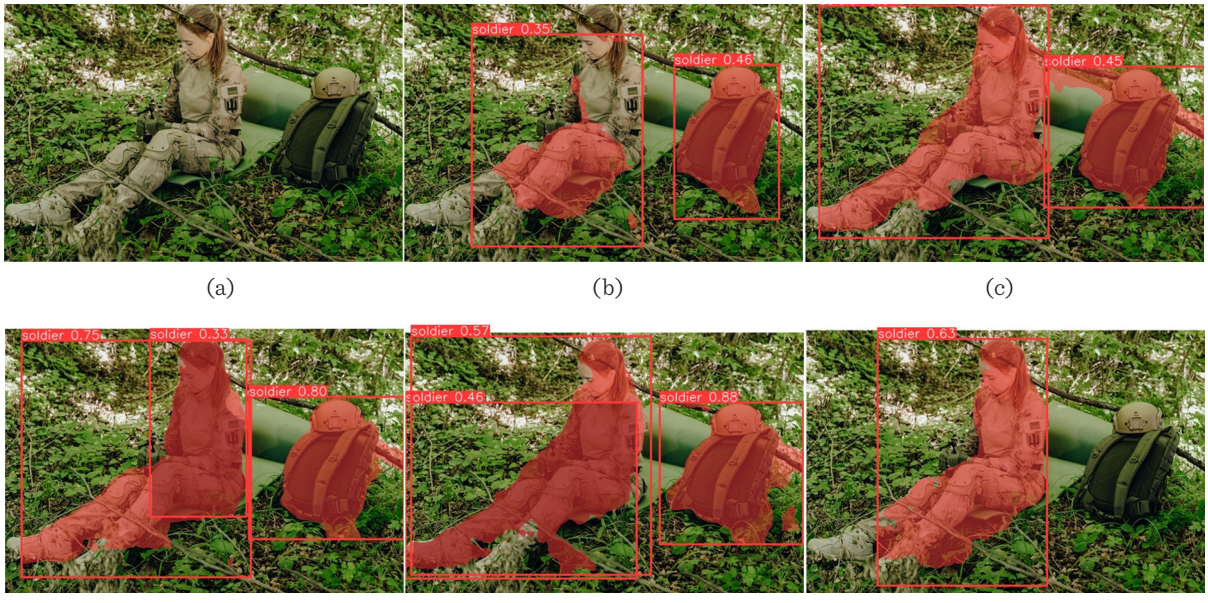


(a) (b) (c)

(d) (e) (f)

Figure 9

The results of the improved model in scene 3: (a) The original image; (b) YOLOv8n; (c) YOLOv8n+DualC2f+GAM; (d) YOLOv8n+DualC2f+DBB; (e) YOLOv8n+GAM+DBB; (f) YOLOv8n+DualC2f+GAM+DBB



(a) (b) (c)

(d) (e) (f)

to camouflage. In the forested battlefield environment illustrated in Figure 9, the baseline model was able to detect the target but mistook the military backpack with the loaded helmet as the target; in contrast, the improved model avoided this misdetection situation.

4.2. Comparison with Other Models

To further confirm the overall detection effectiveness of the improved YOLOv8 model in the soldier dataset, the proposed YOLOV8-SS is compared with three state-of-art instance segmentation algorithms. The comparison algorithms are (1) YOLOv5-seg, (2) Yolact, and (3) YOLOv8n. In order to effectively compare the performance of YOLOV8-SS, the training environments and datasets of the four algorithms are identical.

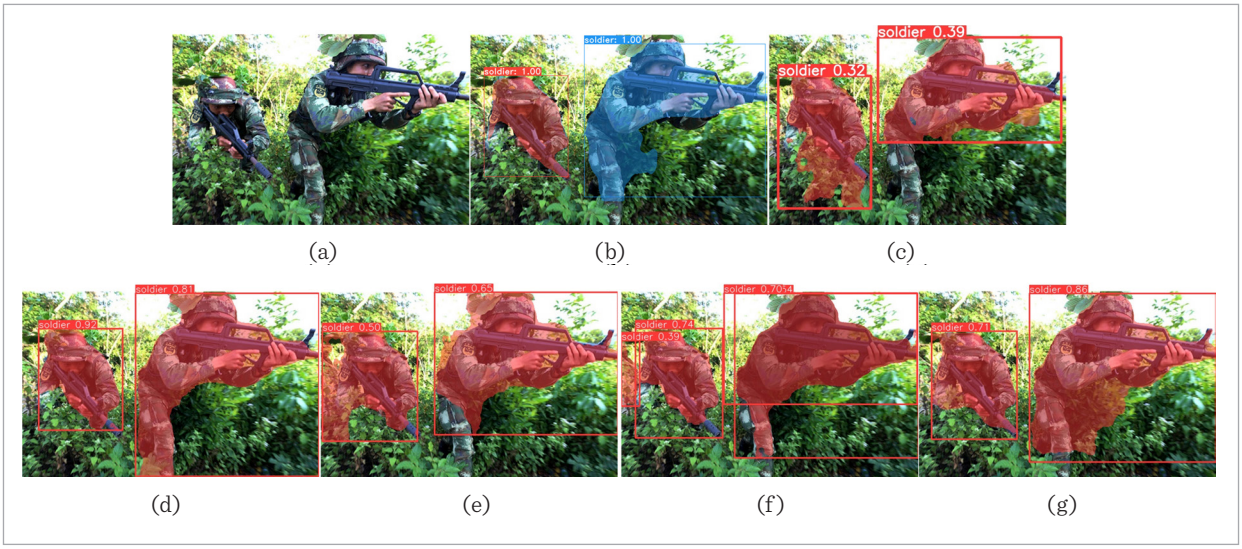
Table 2 shows that the YOLOv8-SS proposed in this paper outperforms the models under comparison in every accuracy-related performance indicator, with just a minor decrease in inference speed.

In order to fully illustrate the adaptability of our model in different scenarios, Figures 10–12 show the processing results of several typical scene examples. Two camouflaged soldiers in forested terrain are seen in Figure 10. While YOLOv8-SS obtains very strong segmentation results, the baseline model suffers from misidentification in a scene with severe interference,

Table 2
Results of comparison with other models

Model	Box P	Box R	Box mAP50	Box mAP50-95	Mask P	Mask R	Mask mAP50	Maks mAP50-95	FPS
Yolact	0.841	0.630	0.844	0.562	0.815	0.599	0.827	0.525	31
Yolov5-seg	0.822	0.781	0.83	0.507	0.822	0.734	0.778	0.473	103
Yolov6-seg	0.884	0.812	0.891	0.644	0.904	0.76	0.867	0.557	108
RTDETR-seg	0.852	0.776	0.832	0.548	0.828	0.761	0.801	0.488	89
Yolov8n	0.892	0.811	0.869	0.616	0.885	0.768	0.841	0.543	116
YOLOv8-SS	0.919	0.795	0.898	0.657	0.892	0.77	0.859	0.589	106

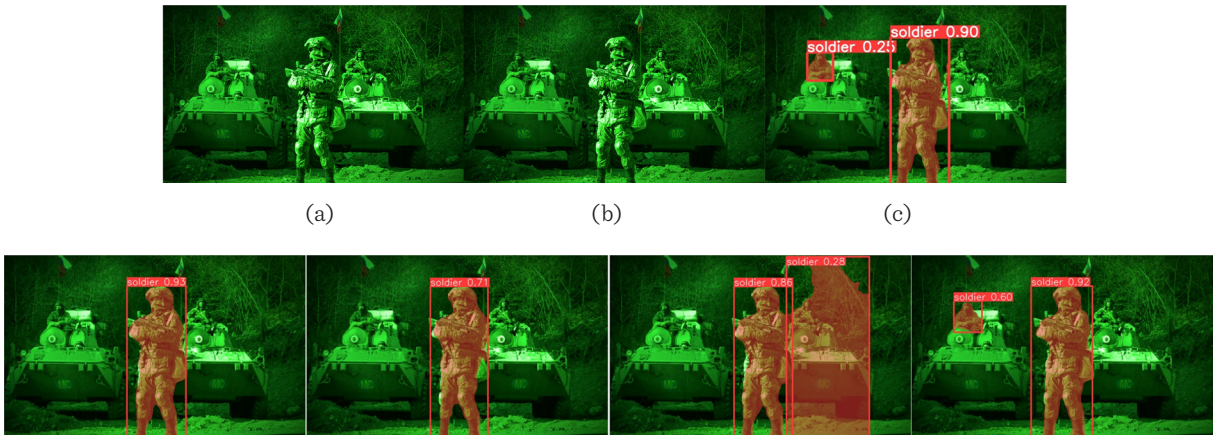
Figure 10
Comparison results of different models in scene 1: (a) The original image; (b) Yolact; (c) YOLOv5-seg; (d) YOLOv6-seg; (e) REDETR-seg; (f) YOLOv8n; (g) YOLOv8n-SS



and the YOLOv5-seg algorithm suffers from inadequate masking, which leads to less accurate localization. Figure 11 shows the situation in a nighttime battlefield environment, where Yolact is unable to localize the target and the baseline model incorrectly identifies the tank target as a soldier, and both YOLO-SS and YOLOv5-seg identify two soldier targets, but YOLO-SS has a higher confidence. The scene in Figure 12 shows a soldier operating a machine gun, and all other models recognize the machine gun and sandbags as soldier targets, but only YOLO-SS accurately locates the soldier target.

Figure 11

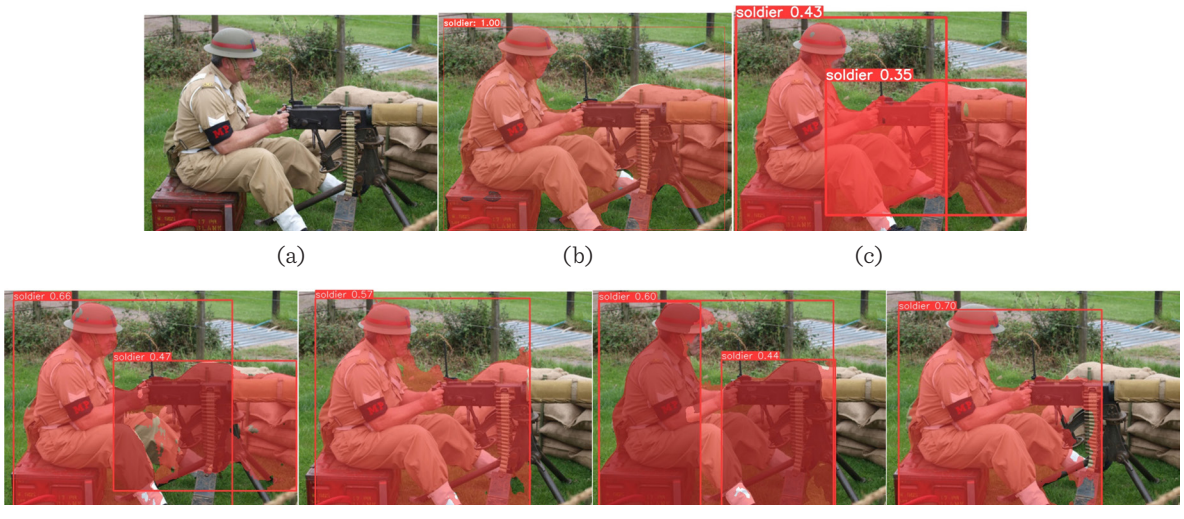
Comparison results of different models in scene 2: (a) The original image; (b) Yolact; (c) YOLOv5-seg; (d) YOLOv6-seg; (e) REDETR-seg; (f) YOLOv8n; (g) YOLOv8n-SS



(a) (b) (c) (d) (e) (f) (g)

Figure 12

Comparison results of different models in scene 3: (a) The original image; (b) Yolact; (c) YOLOv5-seg; (d) YOLOv6-seg; (e) REDETR-seg; (f) YOLOv8n; (g) YOLOv8n-SS



(a) (b) (c) (d) (e) (f) (g)

5. Conclusion

For the detecting and locating of soldiers in complex battlefield environments, this paper proposes a soldier target instance segmentation algorithm, called YOLOV8-SS, based on the improved YOLOv8 algorithm. In this method, the main improvements include three parts: firstly, the DualC2f module is designed based on DualConv to re-place the C2f module in the backbone and neck of YOLOv8; secondly, the global attention module GAM is imported into the feature extraction network; and finally, the reparameterization module is applied to redesign the segmentation head of YOLOv8.

To assess the performance of the YOLOV8-SS model, we performed ablation and comparison experiments to validate the model. Compared to the baseline model YOLOv8n, YOLOV8-SS improved by 2.7%, 2.9%, and 5.1% in Box P, Box mAP50, and Box mAP50-95 metrics, respectively. mask P, Mask mAP50, and Mask mAP50-95 improved by 0.7%, 1.7%, and 4.6%, respectively. However, Box R and FPS decreased slightly, by 1.6% and 8.6%. The experimental results show that the YOLOV8-SS model possesses better performance in the task of segmenting soldiers under intricate battlefield environments.

References

1. Bochkovskiy, A., Wang, C. Y., Liao, H. Y. M. YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv Preprint arXiv:2004.10934, 2020.
2. Bolya, D., Zhou, C., Xiao, F., Lee, Y. J. YOLACT: Real-Time Instance Segmentation. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, 9157-9166. https://doi.org/10.1109/ICCV.2019.00925
3. Cao, L., Zheng, X., Fang, L. The Semantic Segmentation of Standing Tree Images Based on the YOLOv7 Deep Learning Algorithm. Electronics, 2023, 12(4), 929. https://doi.org/10.3390/electronics12040929
4. Cao, X., Su, Y., Geng, X., Wang, Y. YOLO-SF: YOLO for Fire Segmentation Detection. IEEE Access, 2023. https://doi.org/10.1109/ACCESS.2023.3322143
5. Chen, C. F. R., Fan, Q., Panda, R. CrossViT: Cross-Attention Multi-Scale Vision Transformer for Image Classification. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, 357-366. https://doi.org/10.1109/ICCV48922.2021.00041
6. Davis, S. I. Artificial Intelligence at the Operational Level of War. Defence & Security Analysis, 2022, 38(1), 74-90. https://doi.org/10.1080/14751798.2022.2031692
7. Ding, X., Zhang, X., Han, J., Ding, G. Diverse Branch Block: Building a Convolution as an Inception-Like Unit. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, 10886-10895. https://doi.org/10.1109/CVPR46437.2021.01074
8. Du, X., Song, L., Lv, Y., Qiu, S. A Lightweight Military Target Detection Algorithm Based on Improved YOLOv5. Electronics, 2022, 11(20), 3263. https://doi.org/10.3390/electronics11203263
9. Ercolino, S., Devoto, A., Monorchio, L., Santini, M., Mazzaro, S., Scardapane, S. On the Robustness of Vision Transformers for In-Flight Monocular Depth Estimation. Industrial Artificial Intelligence, 2023, 1(1), 1. https://doi.org/10.1007/s44244-023-00005-3
10. Fernandez-Carrobles, M. M., Deniz, O., Maroto, F. Gun and Knife Detection Based on Faster R-CNN for Video Surveillance. Iberian Conference on Pattern Recognition and Image Analysis. Cham: Springer International Publishing, 2019, 441-452. https://doi.org/10.1007/978-3-030-31321-0_38
11. Girshick, R., Donahue, J., Darrell, T., Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, 580-587. https://doi.org/10.1109/CVPR.2014.81
12. Girshick, R. Fast R-CNN. Proceedings of the IEEE International Conference on Computer Vision, 2015, 1440-1448. https://doi.org/10.1109/ICCV.2015.169
13. He, J., Li, P., Geng, Y., Xie, X. FastInst: A Simple Query-Based Model for Real-Time Instance Segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, 23663-23672. https://doi.org/10.1109/CVPR52729.2023.02266
14. He, K., Gkioxari, G., Dollár, P., Girshick, R. Mask R-CNN. Proceedings of the IEEE International Conference on Computer Vision, 2017, 2961-2969. https://doi.org/10.1109/ICCV.2017.322
15. Jensen, B. M., Whyte, C., Cuomo, S. Algorithms at War: The Promise, Peril, and Limits of Artificial Intelligence. International Studies Review, 2020, 22(3), 526-550. https://doi.org/10.1093/isr/viz025

16. Kang, M., Ting, C. M., Ting, F. F., Phan, R. C.-W. ASF-YOLO: A Novel YOLO Model with Attentional Scale Sequence Fusion for Cell Instance Segmentation. Image and Vision Computing, 2024, 147, 105057. https://doi.org/10.1016/j.imavis.2024.105057

17. Kong, L., Wang, J., Zhao, P. YOLO-G: A Lightweight Network Model for Improving the Performance of Military Targets Detection. IEEE Access, 2022, 10, 55546-55564. https://doi.org/10.1109/ACCESS.2022.3177628

18. Krizhevsky, A., Sutskever, I., Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems, 2012, 25.

19. Layton, P. Fighting Artificial Intelligence Battles: Operational Concepts for Future AI-Enabled Wars. Network, 2021, 4(20), 1-100.

20. Liu, H., Soto, R. A. R., Xiao, F., Lee, Y. J. YOLACT-Edge: Real-Time Instance Segmentation on the Edge. 2021 IEEE International Conference on Robotics and Automation (ICRA), 2021, 9579-9585. https://doi.org/10.1109/ICRA48506.2021.9561858

21. Liu, H., Xiong, W., Zhang, Y. YOLO-CORE: Contour Regression for Efficient Instance Segmentation. Machine Intelligence Research, 2023, 20(5), 716-728. https://doi.org/10.1007/s11633-022-1379-3

22. Liu, Y., Shao, Z., Hoffmann, N. Global Attention Mechanism: Retain Information to Enhance Channel-Spatial Interactions. arXiv Preprint arXiv:2112.05561, 2021.

23. Ogundokun, R. O., Maskeli?nas, R., Misra, S., Damasevicius, R. A Novel Deep Transfer Learning Approach Based on Depth-Wise Separable CNN for Human Posture Detection. Information, 2022, 13(11), 520. https://doi.org/10.3390/info13110520

24. Ogundokun, R. O., Maskeli?nas, R., Damaševi?ius, R. Human Posture Detection Using Image Augmentation and Hyperparameter-Optimized Transfer Learning Algorithms. Applied Sciences, 2022, 12(19), 10156. https://doi.org/10.3390/app121910156

25. Redmon, J., Farhadi, A. YOLOv3: An Incremental Improvement. arXiv Preprint arXiv:1804.02767, 2018.

26. Ren, S., He, K., Girshick, R., Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. Advances in Neural Information Processing Systems, 2015, 28.

27. Santos, T., Oliveira, H., Cunha, A. Systematic Review on Weapon Detection in Surveillance Footage Through Deep Learning. Computer Science Review, 2024, 51, 100612. https://doi.org/10.1016/j.cosrev.2023.100612

28. Sharma, P., Sarma, K. K., Mastorakis, N. E. Artificial Intelligence Aided Electronic Warfare Systems-Recent Trends and Evolving Applications. IEEE Access, 2020, 8, 224761-224780. https://doi.org/10.1109/ACCESS.2020.3044453

29. Vallez, N., Velasco-Mata, A., Corroto, J. J., Deniz, O. Weapon Detection for Particular Scenarios Using Deep Learning. Pattern Recognition and Image Analysis: 9th Iberian Conference, IbPRIA 2019, Madrid, Spain, July 1-4, 2019, Proceedings, Part II 9. Springer International Publishing, 2019, 371-382. https://doi.org/10.1007/978-3-030-31321-0_32

30. 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, 2023, 7464-7475. https://doi.org/10.1109/CVPR52729.2023.00721

31. Wang, X., Kong, T., Shen, C., Jiang, Y., Li, L. SOLO: Segmenting Objects by Locations. Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part XVIII 16. Springer International Publishing, 2020, 649-665. https://doi.org/10.1007/978-3-030-58523-5_38

32. Wang, X., Zhang, R., Kong, T., Li, L., Shen, C. SOLOv2: Dynamic and Fast Instance Segmentation. Advances in Neural Information Processing Systems, 2020, 33, 17721-17732.

33. Wang, X., Zhang, R., Shen, C., Kong, T., Li, L. SOLO: A Simple Framework for Instance Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(11), 8587-8601.

34. Yang, B., Wang, X., Xing, Y., Cheng, C., Jiang, W., Feng, Q. Modality Fusion Vision Transformer for Hyperspectral and LiDAR Data Collaborative Classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2024. https://doi.org/10.1109/JSTARS.2024.3415729

35. Zheng, Y., Jiang, W. Evaluation of Vision Transformers for Traffic Sign Classification. Wireless Communications and Mobile Computing, 2022, 2022(1), 3041117. https://doi.org/10.1155/2022/3041117

36. Zhou, C. YOLACT++: Better Real-Time Instance Segmentation. University of California, Davis, 2020.

37. Zhong, J., Chen, J., Mian, A. DualConv: Dual Convolutional Kernels for Lightweight Deep Neural Networks. IEEE Transactions on Neural Networks and Learning Systems, 2022. https://doi.org/10.1109/TNNLS.2022.3151138