Proceedings of the 4th International Conference on Mechatronics and Smart Systems
DOI: 10.54254/2755-2721/2026.32254

Improved YOLOv8s-Pose for Base Fire Keypoint Localization

Zhichun Yang

School of Mechatronic Engineering, Changchun University of Science and Technology, Changchun, China
y1371960393@163.com

Abstract. To address the issue of keypoint deviation in existing primer keypoint localization methods, an improved YOLOv8s-Pose primer detection and keypoint localization algorithm is proposed. Firstly, the standard convolution in the C2f module of the YOLOv8s-Pose backbone network is replaced with GhostConv, which improves computational efficiency and reduces the number of parameters. Secondly, the SimAM attention mechanism is integrated; this parameter-free attention mechanism not only does not add extra parameters but also effectively enhances the feature expression capability of small targets. Finally, the keypoint regression head in the detection head network is replaced with SimCC to improve the localization accuracy of primer crater keypoints. Experimental results demonstrate that on a self-built primer dataset, the improved algorithm achieves an mAP@0.5 of 98.5%, an increase of 11.6% compared to the original YOLOv8s-Pose algorithm, and an mAP@0.5:0.95 of 89.2%, an increase of 21.9%, while reducing computational load and number of parameters, meeting the precision requirements for primer identification and keypoint localization in industrial inspection scenarios.

Keywords: YOLOv8s-Pose, primer keypoint localization, deep learning, attention

1. Introduction

With the arrival of the "Made in China 2025" strategy, the shift towards intelligent transformation in the defense manufacturing sector has become a key national development goal. In automated ammunition assembly production lines, primers, as the core component of the ignition device, require precise recognition and positioning, which directly affects assembly efficiency and safety [1]. However, primer components are usually small in size and have weak features, making high-precision positioning extremely challenging in real production environments. Traditional machine vision methods such as template matching, edge detection, and feature point extraction rely heavily on manually designed features, making it difficult to meet the modern industry's requirements for high-precision positioning [2]. Based on the precise detection requirements for primer components, this method improves the detection accuracy of primer key points by optimizing the Backbone module, introducing the SimAM attention mechanism, and refining the keypoint regression head, based on YOLOv8s-Pose [3].

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.32254

2. Improved YOLOv8-Pose model

YOLOv8 offers models of various sizes, including n, s, m, l, and x, providing multiple options between high-performance detection and lightweight design [4]. To meet the task requirements of primer recognition and keypoint localization detection, due to the small size and weak features of primers, this paper uses the YOLOv8s-Pose model as the baseline, ensuring lightweight design while maintaining accuracy. This paper is based on the YOLOv8 detection framework and proposes an improved YOLOv8 object detection model to address its shortcomings in practical applications, as shown in Figure 1.

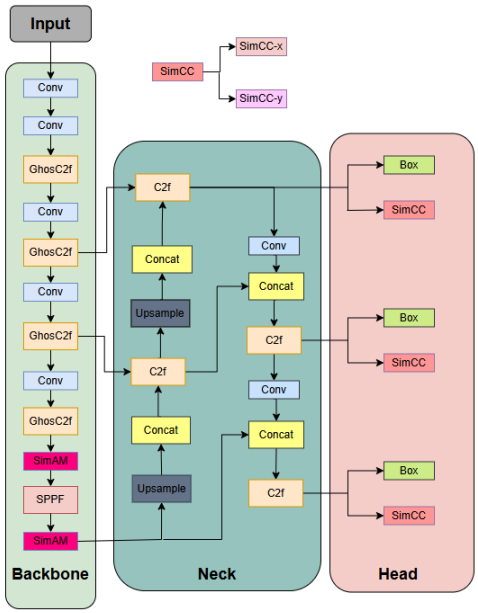

Figure 1. Improved YOLOv8s-Pose network structure

In this article, all standard convolutions inside the C2f modules in the Backbone are replaced with GhostConv, which can improve computational efficiency and reduce the number of parameters. In the Backbone, the SimAM attention mechanism is added after the last C2f module and after the SPPF module. Attention is not introduced after the first three C2f modules because shallow feature maps have high resolution, which is more suitable for extracting fine-grained features, and excessive attention can easily introduce noise enhancement. The Neck module has limited ability to fuse detailed information and mainly focuses on multi-scale feature fusion. Introducing SimAM attention has little effect on enhancing tiny targets and will increase unnecessary computational load. imCC transforms the keypoint localization problem from continuous coordinate regression into a discretized one-dimensional coordinate classification task, which differs from models that directly perform pixel coordinate regression. Common issues in traditional regression methods, such as training instability, difficulty in gradient convergence, and sensitivity to noise and outliers, can be effectively avoided by this approach. At the same time, SimCC achieves sub-pixel level prediction accuracy through high-resolution discretized coordinate axes. Therefore, to improve the localization accuracy of the three primer pit keypoints and the overall robustness of the network, this paper replaces the keypoint regression head of YOLOv8s-Pose with SimCC.

DOI: 10.54254/2755-2721/2026.32254

3. Experiment implementation and evaluation metrics

3.1. Experimental setup

Since there is currently no publicly available dataset of primer keypoint detection images, this study used an industrial camera to collect a total of 1,500 raw images of primers. To enhance the diversity of the dataset, data augmentation was applied to expand the primer images to 2,000. Then, using a stratified sampling method, the dataset was divided into a training set of 1,400 images, a validation set of 300 images, and a test set of 300 images in a 7:2:1 ratio.

The model training and experiments in this study were carried out using Python as the primary programming language and implemented based on the PyTorch deep learning framework. The training experiments were conducted on a high-performance computing server to verify the performance of the improved YOLOv8s-Pose in detecting primer and borehole keypoints. The experimental platform was set up in an Anaconda environment, with hardware configurations including an Intel i7-12700H processor and an NVIDIA RTX 3060 graphics card. The operating system chosen was Windows 10. In terms of the software environment, Python version 3.12.7, PyTorch version 2.8.1, and CUDA version 11.8 were used, ensuring efficient operation of deep learning tasks and stability of model training.

This article makes some parameter settings for the training environment, uniformly adjusting the resolution of all input images to 640×640 pixels, setting the initial number of training epochs to 200, and enabling early stopping. Training will stop early if the performance on the validation set does not improve over several consecutive epochs, preventing underfitting or overfitting.

3.2. Key evaluation indicators

For the task of locating the three key points on the primer surface, commonly used evaluation metrics are mAP, keypoint similarity OKS, number of parameters, and computational cost GFLOPs which can assess the model's performance, accuracy, and efficiency [5].

OKS is a measure of the normalized distance between predicted keypoints and ground truth keypoints, calculated as follows:

$$OKS_p = \frac{\sum_i exp\left(-\frac{d_{ip}^2}{2s_p^2k_i^2}\right)\delta(v_i>0)}{\sum_i \delta(v_i>0)} \tag{1}$$

$$\delta\left(x>0\right) = \begin{cases} 1 & x>0 \\ 0 & x\le 0 \end{cases} \tag{2}$$

In the formula, i represents the detection category of the firing pin keypoint, p is the firing pin number in the detection image, v_i is the visibility code of the keypoint, d_{ip} is the Euclidean distance between the predicted keypoint and the ground truth keypoint, s_p is the scale of the target region, k_i is the normalization factor for each keypoint, used to adjust the weight of different keypoints. $\delta(v_i>0)$ is an indicator function, which is 1 when the keypoint is visible and 0 otherwise.

DOI: 10.54254/2755-2721/2026.32254

The OKS value ranges between 0 and 1. The closer the value is to 1, the closer the predicted keypoint position is to the actual keypoint position, indicating higher estimation accuracy of the model. Let t be the threshold; when the OKS is greater than this threshold, the keypoint is considered successfully detected. In the evaluation of the primer dataset, the OKS values are used to calculate the AP values:

$$AP@t = \frac{\sum_p \delta\left(OKS_p > t\right)}{\sum_p 1} \tag{3}$$

mAP is the mean of AP calculated at different thresholds, which is used to evaluate the performance of the model:

$$mAP = \frac{1}{\mathrm{N}} \sum_{\mathrm{t} \in \mathrm{T}} AP@t \tag{4}$$

Here, N is the total number of thresholds, T is the set of thresholds, usually ranging from 0.5 to 0.95. AP@t is the AP value calculated at different thresholds.

In addition, this paper chooses GFLOPs and the number of parameters as evaluation criteria to measure the model's size.

4. Experimental results and analysis

4.1. Comparison of experimental results and analysis

In the experiment, by comparing the performance of SimpleBaseline, HRNet-W32, YOLOv8s-Pose, YOLOv11s-Pose, and the improved YOLOv8s-Pose model proposed in this study on the primer dataset, the effectiveness of primer pit keypoint detection was analyzed using mAP@0.5, mAP@0.5:0.95, parameter count, and computational cost as evaluation metrics, exploring the strengths and weaknesses of different algorithms in primer recognition and keypoint localization.

Table 1. Comparison of primer keypoint detection results

Model	mAP@0.5(%)	mAP@0.5:0.95(%)	Parameters /M	GFLOPs/G
SimpleBsaeline	82.5	53.6	28.5	41.2
HRNet-W32	85.6	56.8	42.1	39.8
YOLOv8s-Pose	86.9	67.2	11.6	30.2
YOLOv11s-Pose	91.3	72.7	9.9	26.0
Improve model	98.5	89.2	11.3	29.1

As shown in Table 1, the improved YOLOv8s-Pose model presented in this paper performs best in the primer keypoint localization task, achieving an mAP@0.5 of 98.5%. Compared to the SimpleBaseline, HRNet-W32,YOLOv8s-Pose, and YOLOv11s-Pose models, it shows improvements of 16%, 12.9%, 11.6%, and 7.2%, respectively. The mAP@0.5:0.95 reaches 89.2%, representing gains of 35.6%, 32.4%, 22%, and 16.5% over the respective models. In terms of parameters and computational cost, it is higher only than the YOLOv11s-Pose model and lower than the other models, fully demonstrating that the proposed model achieves lightweight design while improving detection accuracy.

DOI: 10.54254/2755-2721/2026.32254

4.2. Ablation experiment results and analysis

To verify the detection performance of the improved YOLOv8s-Pose model in the primer keypoint localization task, a systematic ablation study was conducted on the primer dataset constructed in this study. The purpose of the experiments was to evaluate the actual improvement of the overall model after optimizing the backbone module, introducing the SimAM attention mechanism, and optimizing the keypoint regression head. In the experimental setup, Group 1 used the standard configuration of YOLOv8s-Pose as the baseline, with a network depth coefficient of 0.33 and a width coefficient of 0.50. Based on this, Group 2 replaced all standard convolutions within each C2f module in the backbone with GhostConv. Group 3 was based on Group 2 and added the SimAM attention mechanism at the output of the last C2f module in the backbone and after the SPPF module. Group 4 was based on Group 3 and replaced the keypoint regression head with SimCC. All improvements were then compared with the baseline model, focusing on analyzing the impact of different combinations on the localization accuracy of primer crater keypoints. The comparison metrics included four performance evaluations: precision, recall, mAP@0.5, and mAP@0.5:0.95.

For clarity, in the ablation experiments, the module numbering is defined as follows: A represents the optimized Backbone module, B represents the introduction of the SimAM attention module, and C corresponds to replacing the keypoint regression head with SimCC, as shown in Table 2.

Table 2. Module number and corresponding name

Module ID	Module Name
A	Optimized Backbone module
B	Introduced SimAM attention
C	Replaced keypoint regression head with SimCC

Table 3 shows the ablation experiment results of different models in the task of keypoint detection for the three primer positions. The baseline model performed the lowest in both mAP@0.5 and mAP@0.5:0.95. The second and third group models showed gradual improvements. The fourth group model, after integrating all the needed improvements or introduced modules, achieved an mAP@0.5 of 98.5%, which represents an increase of 11.6%, 6.8%, and 5.3% compared to the first three groups, respectively. The mAP@0.5:0.95 reached 89.2%, which is 21.9%, 16.8%, and 7.4% higher than the first three groups, respectively. The parameters and computation cost also decreased compared to the first three groups, achieving both higher accuracy and a more lightweight model. Although the third group increased in parameters and computation, its accuracy improved, fully validating the effectiveness of each improvement module in the keypoint detection task for the three primer positions.

Table 3. Ablation experiment results for primer keypoint detection

Group	A	B	C	Parameters /M	GFLOPs /G	mAP@0.5(%)	mAP@0.5:0.95(%)
1				11.6	30.2	86.9	67.3
2	✓			11.4	29.4	91.7	72.4
3	✓	✓		11.4	29.5	93.2	81.8
4	✓	✓	✓	11.3	29.1	98.5	89.2

DOI: 10.54254/2755-2721/2026.32254

4.3. Results and analysis of the reasoning experiment

To verify the improvement in the accuracy of keypoint detection for the three primer positions in the proposed improved YOLOv8s-Pose algorithm model, typical primer images from the test set were selected as experimental subjects for inference comparison experiments. In the experiments, keypoint detection on the test samples was performed using both the original YOLOv8s-Pose model and the improved YOLOv8s-Pose model proposed in this study, and the inference results of the two models were compared and analyzed both qualitatively and quantitatively.

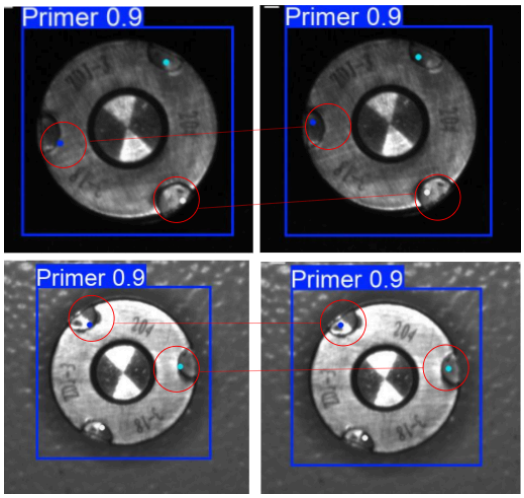

Figure 2. Comparison of model inference results

As shown in Figure 2, the figure compares the inference results of the model before and after improvement. By visually comparing the detection results, it can be seen that the model before improvement is prone to issues such as keypoint displacement and unstable positioning. In contrast, the improved model achieves more accurate and stable positioning of the keypoints at the three pit locations, with a more concentrated distribution of keypoints and significantly reduced displacement. This effectively enhances the reliability of detecting the primer keypoints, providing more precise visual information support for the subsequent grabbing of primers.

5. Conclusion

This paper proposes an improved YOLOv8s-Pose scheme, which replaces the standard convolution inside the C2f module with GhostConv in the Backbone, introduces the SimAM attention mechanism after the last C2f module and SPPF, and replaces the keypoint regression head in the detection head network with SimCC, achieving both improved keypoint detection accuracy and model lightweighting. Based on practical work scenarios, a dataset of 2,000 primer images was constructed. Comparative experiments were conducted between the proposed algorithm and SimpleBaseline, HRNet-W32, and YOLOv11s-Pose. Ablation and inference experiments were also performed according to the proposed improvement strategies. The results show that the improved model achieves an mAP@0.5:0.95 of 89.2% for primer keypoint detection, an improvement of 21.9% over the baseline model, The mAP@0.5 reached 98.5%, an improvement of 11.6% over the baseline model.while also reducing the number of parameters and computational cost. Additionally, the training converges faster and achieves lower loss. These results fully validate the effectiveness of the proposed improvement strategies in high-precision keypoint localization, providing reliable

DOI: 10.54254/2755-2721/2026.32254

visual information support for subsequent robotic primer grasping and serving as a reference for small-target keypoint detection.

References

[1] Gong Y, Luo J, Shao H, et al. Automatic Defect Detection for Small Metal Cylindrical Shell Using Transfer Learning and Logistic Regression [J]. Journal of Nondestructive Evaluation, 2020, 39(1): 1-13.
[2] Gao S H, Cheng M, Zhao K, et al. Res2net: A new multi-scale backbone architecture [J]. IEEE transactions on pattern analysis and machine intelligence, 2019, 43(2): 652-662.
[3] Jeon, Hyung-Joon, et al. An Integrated Real-time Monocular Human Pose & Shape Estimation Pipeline for Edge Devices [C]. //IEEE International Conference on Robotics and Biomimetics (ROBIO). Koh Samui, Thailand, December 4-9, 2023. IEEE, 2023: 1-6.
[4] Ultralytics.YOLOv8 [EB/OL].https: //github.com/ultralytics/ultralytics, accessed 2023.
[5] Lin T Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context [C]// Computer Vision-ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. Springer International Publishing, 2014: 740-755.