ISSN: 2414-1895 DOI: 10.6919/ICJE.202503_11(3).0010

Research on Scrap Steel Identification and Detection based on SAM and YOLOv9

Zhenzhong Shen[a], Ye Zhu[b], Yao Xiao[c], Youzhi Zhao[d], and Liya Wang[e, *]

College of Science, North China University of Science and Technology, Tangshan 063210, China

[a]shenzhenzhong@stu.ncst.edu.cn, [b]zhuye@stu.ncst.edu.cn, [c]xiao@stu.ncst.edu.cn, [d]2294845993@qq.com, [e], *wangliya@ncst.edu.cn

Abstract

In the context of the "dual carbon" initiative, the steel industry must expedite its efforts to enhance the automation and intelligence of scrap steel recycling, as this is a pivotal step in achieving a low-carbon transformation and ensuring resource circularity. The prevailing methods, which depend heavily on manual labor, are characterized by inefficiencies and suboptimal accuracy. In response to these challenges, this study proposes an innovative approach that integrates the Segment Anything Model (SAM) with the YOLOv9 object detection algorithm for scrap steel classification. Initially, SAM is employed to meticulously delineate scrap steel images, effectively isolating the target regions from their complex backgrounds. Subsequent to this segmentation, the YOLOv9 algorithm is implemented for real-time detection and classification of the segmented regions. Empirical evaluations demonstrate that, in comparison with conventional YOLO-based methodologies, the proposed approach enhances mAP@0.5 by approximately 7% and elevates the recall rate by around 2%. Furthermore, the method demonstrates enhanced robustness and precision when handling scrap steel images with varied morphologies and intricate backgrounds. These results underscore the promise of a segmentation-first strategy in advancing the automated classification of scrap steel, thereby paving the way for more efficient industrial recycling processes.

Keywords

Scrap Steel Classification; Segment Anything Model; YOLOv9; Industrial Intelligence; Multi-scale Detection.

1. Introduction

The iron and steel industry are currently under significant pressure to conserve energy and reduce emissions in the wake of the full implementation of the "dual-carbon" strategy [1]. In this context, scrap recycling emerges as a pivotal strategy for the industry to achieve low-carbon transformation, given its role as a critical component in the iron and steel production process. Efficient recycling of steel scrap has been shown to contribute to energy savings and reduced carbon emissions, while concurrently reducing production costs and enhancing the sustainability of steel production [2]. However, traditional scrap recycling methods rely on manual operations, which are inefficient and difficult to ensure accuracy. The manual sorting process is vulnerable to human factors, and it is both labor-intensive and hazardous, which hinders its ability to satisfy the elevated efficiency and accuracy demands of contemporary manufacturing.

With the rapid development of computer vision technology, intelligent scrap steel recognition technology has brought new breakthroughs to the industry. Liu et al[2] proposed a scrap steel image

 Volume 11 Issue 3, 2025
DOI: 10.6919/ICJE.202503_11(3).0010

classification method based on deep convolutional neural network for the scrap steel recycling problem in steel production. In the study, a dataset containing different kinds of scrap steel images was constructed, and high-precision scrap steel recognition was realized by the improved CNN structure, which provides technical support for the low-carbon transformation and resource recycling in the steel industry. Chen et al[3] proposed a scrap steel (or scrap metal) classification method based on the YOLO target detection algorithm. Through multi-scale feature extraction and region localization of scrap images, real-time detection and classification of different scrap types under complex backgrounds is achieved, providing an effective technical solution for automated scrap sorting. Wang et al [4] used convolutional neural networks combined with migration learning techniques to automatically classify scrap images. By fine-tuning the pre-training model, the researchers achieved high classification accuracy with limited sample data and verified the feasibility of the method in industrial scrap steel recycling. Zhang et al [5] designed a deep neural network incorporating the attention mechanism for accurate classification of targets in scrap steel images. How to utilize channel attention and spatial attention to improve the model's ability to adapt to the complexity and morphological diversity of scrap steel appearance was discussed in detail, and better classification performance than traditional methods was achieved.

The above scholars have achieved remarkable results in the field of steel scrap classification. Then they mostly improve on the model structure to cope with the challenges of polymorphic targets as well as complex backgrounds, with less processing of input data. In this paper, we take an alternative approach to segment the scrap steel using SAM (Segmentation Everything Model), which effectively separates the target region from the background; then we use the YOLOv9 algorithm to recognize and classify the segmented image, which improves the processing accuracy of the scrap steel image, and also accelerates the speed of detection, thus laying the foundation for efficient classification and sorting of scrap steel.

2. Methodology

2.1 SAM

The Segmentation Everything Model (SAM) [6] is a generalized deep learning-based image segmentation framework, which aims to achieve efficient and accurate segmentation of various objects and regions in an image through a small number of hints or complete automation. The model adopts large-scale data pre-training and multi-scale feature fusion techniques, so that it can maintain excellent segmentation results when facing different types and complex scenes.SAM can not only cope with the requirements of specific targets in the traditional segmentation tasks through flexible cueing mechanisms, but also expand to a wider range of application scenarios, showing strong generalization ability. Its design takes into full consideration the real-time and robustness of practical applications, providing strong technical support for image editing, video processing and intelligent surveillance, and exploring a general and efficient solution path for future segmentation tasks.

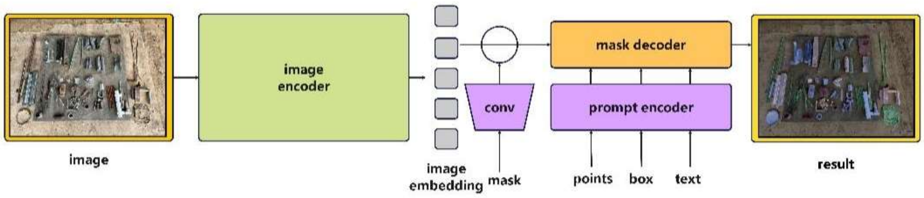

Figure 1. SAM architecture diagra

2.2 YOLOv9

YOLOv9[7], as the latest version of the YOLO series of target detection algorithms, inherits the design concept of efficient end-to-end detection, and optimizes the network structure comprehensively for the detection needs of multi-scale and multi-angle targets. The model adopts an improved backbone network and feature fusion strategy, which effectively improves the model's

ISSN: 2414-1895
DOI: 10.6919/ICJE.202503_11(3).0010

detection accuracy and real-time performance in complex scenes through enhanced data augmentation, multi-scale training, and optimized loss function design. Meanwhile, YOLOv9 has been specially designed in terms of model lightweighting and deployment efficiency, which makes it applicable to embedded systems and mobile devices, providing solid technical support for real-time monitoring, automatic sorting and other tasks in industrial applications.

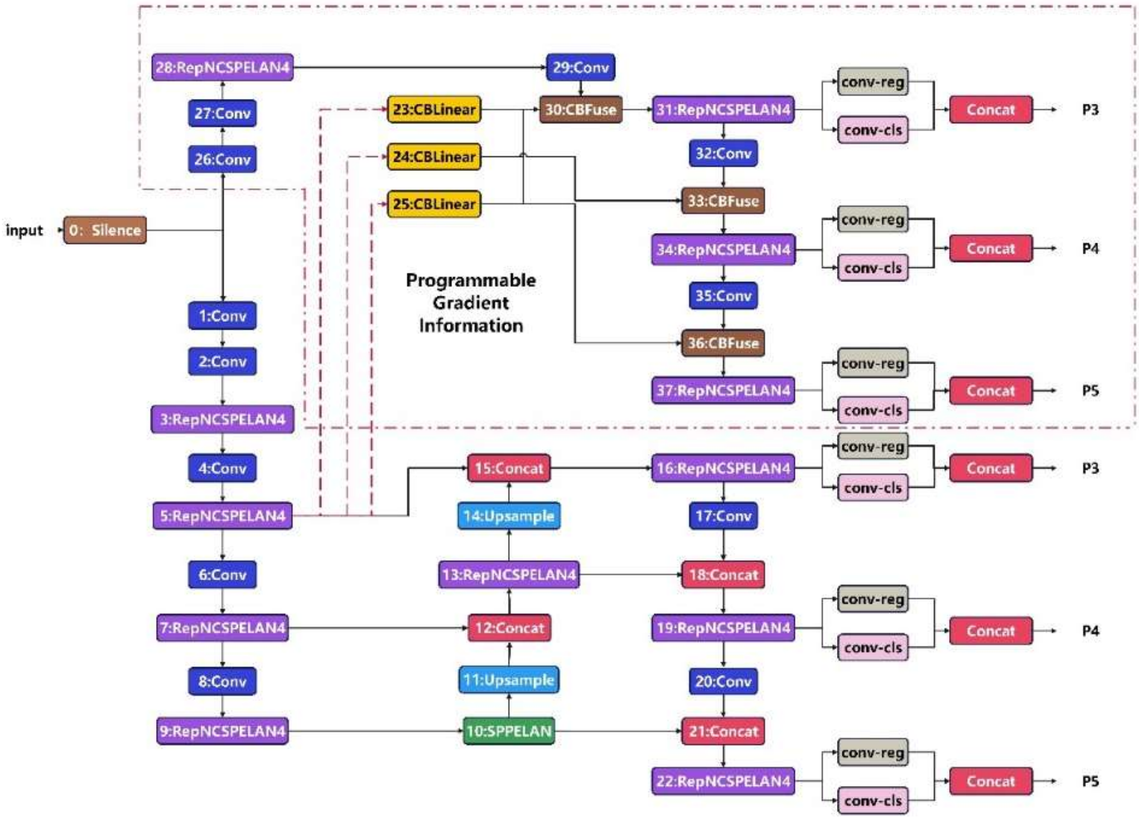

Figure 2. YOLOv9 architecture diagram

In order to further alleviate the information bottleneck and unstable gradient transfer problems commonly found in deep neural networks, YOLOv9 introduces the Programmable Gradient Information (PGI) mechanism. As illustrated in Figure 2, PGI effectively retains crucial information in the deep layers of the network by constructing auxiliary reversible branches and providing stable and reliable gradient feedback to the primary network. This significantly enhances the model's training stability and convergence speed. Additionally, the mechanism employs a multilevel auxiliary information fusion strategy to replan and aggregate gradient signals at different semantic levels. This ensures that both shallow and deep networks have complete access to supervisory information specific to the target task. This innovative approach has two notable advantages. First, it enhances the performance of the lightweight model under limited samples. Second, it establishes a theoretical and practical foundation for high-precision target detection in complex industrial scenarios.

3. Experiments

3.1 Experimental Platform

The experimental environment is configured as follows: the platform operating system is Windows 10 Professional, with an Intel(R) Core(TM) i9-13900K CPU at 3.5 GHz and 128 GiB of RAM. The graphics card is Nvidia GeForce RTX 4070 with 36 GiB of VRAM, the development language is Python 3.10, the deep learning framework is PyTorch 2.1, the CUDA version is 12.2, and the cuDNN version is 8.9.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202503_11(3).0010

3.2 Evaluation Metrics

The experiments will evaluate the model’s performance based on precision, recall, and mean average precision (mAP), with mAP@0.5 used to compute the average precision.

Precision is used to measure how many of the positive class samples predicted by the model are true positive classes. The higher the precision, the higher the reliability of the model in predicting the positive class. The formula is as follows:

$$P = \frac{TP}{TP + FP} \tag{1}$$

Where TP (True Positive) is a positive class sample correctly predicted by the model. FP (False Positive) is a negative class sample that the model incorrectly predicted as positive.

Recall measures how many of the true positive class samples are correctly recognized by the model. The higher the recall, the better the model's ability to capture positive classes. The formula for this is as follows:

$$R = \frac{TP}{TP + FN} \tag{2}$$

Where FN is the sample of positive classes that the model misses (underreporting).

mAP@0.5 (mean Average Precision at IoU=0.5) is a core metric in the target detection task, which represents the mean of the average precision across all categories at an intersection-to-parallel ratio (IoU) threshold of 0.5. The higher mAP@0.5 is, the better is the model's integrated detection performance under loose overlap requirements. Its formula is:

$$mAP@0.5 = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{3}$$

Where N is the total number of categories and AP_i is the average precision of the i th category.

3.3 Crap Dataset

Table 1. Classification of different types of scrap

form	hallmark
Punching material, small material	Overall dimensions: length ≤ 150 mm, thickness ≥ 3 mm; Reinforcing head length ≤ 150mm, reinforcing head diameter ≥ 10mm
Heavy Scrap	Overall dimensions: ≤ 800 mm× 600 mm, thickness ≥ 14 mm. Single weight≤800kg
High quality heavy duty steel scrap	Overall dimensions: ≤ 800 mm× 600 mm, thickness ≥ 20 mm. Single weight≤800kg
medium-sized scrap	Overall dimensions: ≤ 800 mm× 600 mm, thickness ≥ 4 mm - 6 mm. Single weight ≤300kg
Scrap in Shear	Overall dimensions: ≤ 800 mm× 600 mm, thickness ≥ 2 mm - 5 mm. Single weight ≤300kg
Steel plate trimmings	Overall dimensions: ≤ 800mm× 600mm, thickness ≥ 10mm
rebar head	Overall dimensions: steel head length ≤150mm, thickness ≥16mm, ≤350mm

Due to the fact that there are fewer relevant datasets publicly available on the Internet and the amount and categories of data do not meet the needs of model training, our team chose to go to the steel mill for a field trip, took a large number of photos of steel scrap and classified them to produce a steel scrap dataset, and the results are shown in Table 1. All images in the dataset are in standard image format and well labeled to facilitate subsequent computer vision tasks.

4. Result

This experiment uses four models, YOLOv9, YOLOv8, YOLOv7, and our model, all on the same experimental platform, and the performance metrics as well as the recognition accuracies on different scrap steels are shown in Table 2.

Table 2. Classification of different types of scrap

indicators	ours	YOLOv9	YOLOv8	YOLOv7
tube	0.83	0.80	0.75	0,69
billet	0.94	0.88	0.80	0.74
bars	0.79	0.74	0.67	0.65
scrap	0.88	0.89	0.77	0.66
mAP@0.5	0.74	0.67	0.60	0.54
Recall	0.75	0.73	0.69	0.59
billet	0.94	0.88	0.80	0.74

Based on Table 2, in terms of model detection accuracy, our model shows a significant improvement compared to other models, achieving a 7% increase in mAP@0.5 and a 2% improvement in recall relative to the baseline model. The test results for the three models are shown in Figure 3.

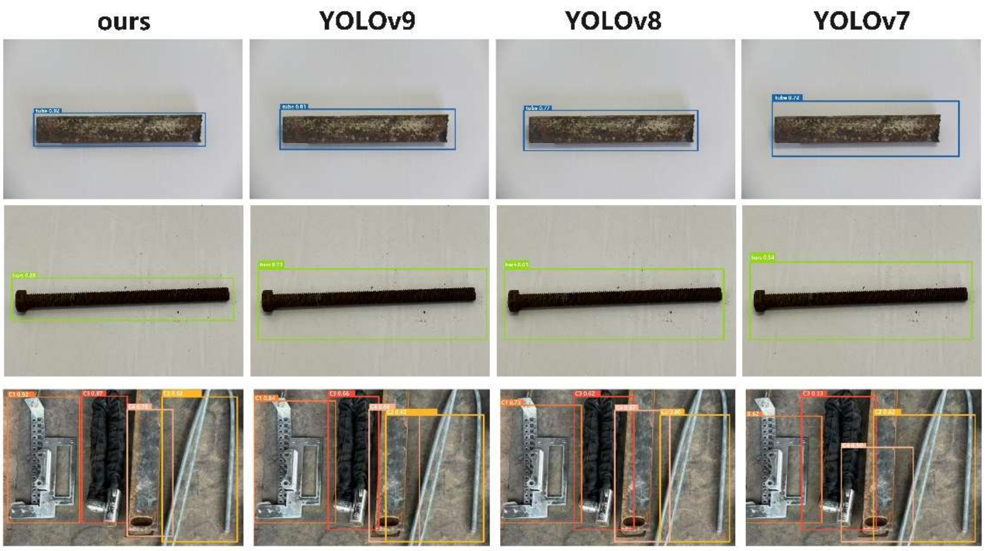

Figure 3. Detection results of different models

5. Conclusion

This study proposes a fusion method that integrates the strengths of both SAM and YOLOv9 to achieve efficient segmentation and classification of scrap steel images. The proposed approach integrates the SAM technique to ensure precise segmentation of scrap steel images, effectively isolating target regions from the background. Subsequently, the YOLOv9 model is employed for real-time detection and classification of the segmented targets. Through experimental evaluations, it has been demonstrated that this approach leads to substantial performance enhancements. Specifically, the proposed method exhibited an approximately 7% increase in mAP@0.5 and a 2% improvement in recall when compared to traditional YOLO series models. These results substantiate the hypothesis that a segmentation-then-detection strategy enhances robustness and accuracy in industrial scenarios characterized by diverse scrap steel morphologies and complex backgrounds.

Notwithstanding the high accuracy and rapid detection speed observed in the experiments, there is still room for improvement in practical applications. For instance, under conditions of extreme lighting, occlusions, and highly complex backgrounds, the model's robustness requires further enhancement. A promising avenue for future research involves the optimization of the information fusion mechanism between the Scale-Adaptive Merging (SAM) and YOLOv9 to reduce computational redundancy and enhance processing efficiency. Furthermore, as the volume and diversity of industrial data continue to expand, augmenting the dataset scale and refining annotation strategies will likely enhance the model's generalization ability and applicability.

In summary, this research on scrap steel classification based on SAM and YOLOv9 offers a novel approach for automated scrap sorting. The integration of precise segmentation with real-time object detection enables efficient and accurate classification in complex industrial environments, thereby establishing a solid technical foundation for advancing intelligent and low-carbon development within the steel industry.

Acknowledgments

Supported by National undergraduate innovation and entrepreneurship training program of China(202410081029);

2024 North China University of Science and Technology Undergraduate Innovation and Entrepreneurship Training Program Project(202410081029).

References

[1] YU Biying, ZHAO Guangpu, AN Runying, et al. Study on China's Carbon Emission Pathway under Carbon Neutral Target[J]. Journal of Beijing Institute of Technology (Social Science Edition),2021, 23(02):17-24.DOI:10.15918/j.jbitss1009-3370.2021.7380.

[2] Liu, Z., et al. (2021). Deep learning-based image classification for scrap steel sorting in steelmaking process. DOI: 10.1016/j.proci.2021.05.023

[3] Chen, H., et al. (2020). A novel YOLO-based method for scrap metal classification and sorting.DOI: 10.1016/j.compind.2020.103215

[4] Wang, Y., et al. (2019). Automated scrap steel classification using convolutional neural networks and transfer learning.DOI: 10.1109/ICMLST.2019.8832389

[5] Zhang, L., et al. (2022). Intelligent scrap steel classification using attention-based deep neural networks.DOI: 10.1016/j.eswa.2022.117923

[6] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., & Girshick, R. (2023). Segment Anything. arXiv preprint arXiv:2304.02643.

[7] Wang, C.-Y., Yeh, I.-Hau, & Liao, H.-Y. M. (2024). YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. arXiv preprint arXiv:2402.13616.