

# Location and segmentation of important railway signs based on improved segmentation

Improved  
SegNeXt

Wang Zengqing, Zheng Yu Xie and Jiang Yiling

*School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China*

21

Received 28 October 2023  
Revised 15 November 2023  
Accepted 24 November 2023

## Abstract

**Purpose** – With the rapid development of railway-intelligent video technology, scene understanding is becoming more and more important. Semantic segmentation is a major part of scene understanding. There is an urgent need for an algorithm with high accuracy and real-time to meet the current railway requirements for railway identification. In response to this demand, this paper aims to explore a variety of models, accurately locate and segment important railway signs based on the improved SegNeXt algorithm, supplement the railway safety protection system and improve the intelligent level of railway safety protection.

**Design/methodology/approach** – This paper studies the performance of existing models on RailSem19 and explores the defects of each model through performance so as to further explore an algorithm model dedicated to railway semantic segmentation. In this paper, the authors explore the optimal solution of SegNeXt model for railway scenes and achieve the purpose of this paper by improving the encoder and decoder structure.

**Findings** – This paper proposes an improved SegNeXt algorithm: first, it explores the performance of various models on railways, studies the problems of semantic segmentation on railways and then analyzes the specific problems. On the basis of retaining the original excellent MSCAN encoder of SegNeXt, multiscale information fusion is used to further extract detailed features such as multihead attention and mask, solving the problem of inaccurate segmentation of current objects by the original SegNeXt algorithm. The improved algorithm is of great significance for the segmentation and recognition of railway signs.

**Research limitations/implications** – The model constructed in this paper has advantages in the feature segmentation of distant small objects, but it still has the problem of segmentation fracture for the railway, which is not completely segmented. In addition, in the throat area, due to the complexity of the railway, the segmentation results are not accurate.

**Social implications** – The identification and segmentation of railway signs based on the improved SegNeXt algorithm in this paper is of great significance for the understanding of existing railway scenes, which can greatly improve the classification and recognition ability of railway small object features and can greatly improve the degree of railway security.

**Originality/value** – This article introduces an enhanced version of the SegNeXt algorithm, which aims to improve the accuracy of semantic segmentation on railways. The study begins by investigating the performance of different models in railway scenarios and identifying the challenges associated with semantic segmentation on this particular domain. To address these challenges, the proposed approach builds upon the strong foundation of the original SegNeXt algorithm, leveraging techniques such as multi-scale information fusion, multi-head attention, and masking to extract finer details and enhance feature representation. By



© Wang Zengqing, Zheng Yu Xie and Jiang Yiling. Published in *Smart and Resilient Transportation*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence maybe seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

Smart and Resilient  
Transportation  
Vol. 6 No. 1, 2024  
pp. 21-31  
Emerald Publishing Limited  
e-ISSN: 2632-0495  
p-ISSN: 2632-0487  
DOI 10.1108/SRT-10-2023-0010

---

doing so, the improved algorithm effectively resolves the issue of inaccurate object segmentation encountered in the original SegNeXt algorithm. This advancement holds significant importance for the accurate recognition and segmentation of railway signage.

**Keywords** Railway, Semantic segmentation, Characteristics of small objects, Railroad marking, Safety protection, SegNeXt

**Paper type** Research paper

---

## 1. Introduction

At present, with the development of high-speed railroads, the safety problem is gradually highlighted, and the railroad-intelligent video technology is currently on the railroad can be timely detection of abnormalities on the line (Li, 2022), to protect the safety of an important method, by the industry and academia's extensive attention. In railroad-intelligent video technology, scene understanding is the first condition to realize the technology, and semantic segmentation is the key component of scene understanding (Lu *et al.*, 2023). Therefore, considering an efficient and convenient semantic segmentation technique can greatly complement the existing security intelligent protection system and provide a more efficient and convenient security protection method.

Nowadays, with the development of deep learning technology, the application of semantic segmentation technology on the railroad is gradually increasing. Song and Ge (2023), based on transformer and attention mechanism, proposed a network structure TransMANet that enhances the semantic information of shallow network with a two-branch decoder that fuses local and global contexts, to solve the problems of insufficient correlation of MANet algorithm with the semantic information of the image, insufficient global feature extraction and low segmentation accuracy. Zhai *et al.*, based on the classical MaskRCNN instance segmentation network architecture, designed a new feature extraction architecture DseNet, the combination of which obtained DseNet-MaskRCNN, which focuses on strengthening the accuracy at the edge of the image in the network architecture, to improve the accuracy of railroad image segmentation so that it can be used as a basic vision task for the unmanned railroads and the work of the equipment such as the safe, fast and accurate detection (Zhai and Li, 2021). Zhang *et al.* proposed an improved DeepLabV3+ algorithm that combines different attention mechanisms, to address the problem that the backbone network structure of the original model is more complex and requires more computational resources for training, the backbone network is replaced with ResNet50, which solves the problem that existing semantic segmentation models for self-driving road scenarios do not extract features sufficiently in the feature extraction stage and ignore the differences in feature importance in different levels of feature maps, thus leading to poor segmentation results (Zhang *et al.*, 2023). However, from the results of the above paper, its for long-distance small target is still deficient, and now the railroad environment railroad marking is more cluttered, and marking in the current railroad environment is manifested as a small target, small target segmentation at present is still relatively lack of Ma *et al.* (2023).

Overall, the railroad scene currently suffers from a number of detailed markings that are difficult to accurately segment, such as signals, distant tracks and narrow roads. For the above identification, existing algorithms perform poorly, which is difficult to carry out for deep understanding of railroad scenarios, so we explored the major algorithmic models in the hope of investigating a semantic segmentation model for railroads to better solve such problems.

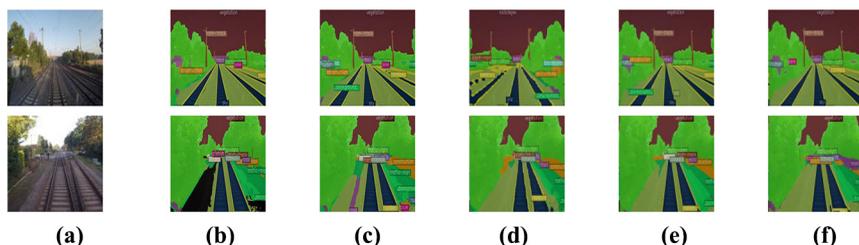
## 2. Methodology and rationale

### 2.1 Comparison of algorithms in railroad scenarios

Most of the algorithms are compared and improved on existing publicly available data sets, and continue to be problematic in railroad applications. In response to this, this paper conducts experiments based on the RailSem19 data set, a complementary and improved public data set of railroads, to experimentally compare the algorithms of the existing models SegNeXt, san, DeepLabV3+ and resnest+DeepLabV3. The results of the experiment are shown in Figure 1 and Table 1.

In this table, time denotes the time required to test one round of pictures, and one round of pictures is 50 pictures, which can be used to increase the running speed by using multithreading in the actual railroad operation. However, after our investigation, multithreaded processing in the actual railroad scenario can be used in three-threaded operation: thread 1 is used to push the stream, thread 2 is used for backbone operation and encoding and decoding are performed in addition; however, it is also necessary to run a single server to achieve more than five frames per second before real-time computing and decoding output (Wang et al., 2023).

The study found that SegNeXt performed the best and had the lowest frame per second (FPS). And there is a phenomenon worth noting: the experimental investigation of the railroad scene in the target scale varies greatly, such as near and far railroad tracks and bridges. It is also clear from the test images that the other models fall short of SegNeXt for the distance. The multiscale feature extraction mechanism in SegNeXt can improve the



**Notes:** (a) Original figure; (b) annotation figure; (c) segnext; (d) san; (e) deepLabV3+; (f) resnest

**Source:** Figure created by authors

**Figure 1.**  
Schematic  
representation of  
railroad data for each  
algorithmic model

Model	Resolution ratio	ACC	MACC	IOU				Time
				Traffic light	Road	Pole	MIOU	
SegNeXt	$512 \times 512$	88.77	73.02	85.42	75.47	39.91	62.85	0.0953
	$1,080 \times 1,080$	89.84	79.67	89.95	76.31	50.39	69.85	0.2808
San	$512 \times 512$	86.10	73.44	82.21	72.08	37.92	60.53	0.1886
	$1,080 \times 1,080$	85.25	69.72	80.53	69.67	38.88	55.14	0.6272
DeepLabV3+	$512 \times 512$	90.56	79.69	86.63	77.32	48.51	69.41	0.2511
	$1,080 \times 1,080$	86.72	73.12	82.46	65.88	44.41	61.64	0.5793
Resnest+DeepLabV3	$512 \times 512$	87.79	72.67	83.19	74.74	42.04	62.79	0.2994
	$1,080 \times 1,080$	82.31	56.26	78.05	57.93	41.37	48.63	0.7706

**Source:** Table created by authors

**Table 1.**  
Specific performance  
of each detection  
model

recognition of targets at different scales, and the strip conv in SegNeXt has a better segmentation effect for the wires, indicator poles and other markers in the railroad representations, but they are generally less accurate for the detection of status quo objects, which is not suitable to be applied in the actual railroad scenarios (Liu *et al.*, 2023).

But as the image resolution gradually increases, SegNeXt grows much smaller than the other models for the time used. This is because the computational complexity of the self-attention mechanism is related to the resolution squared. SegNeXt uses the convolution operation to avoid such problems to a certain extent. For the railroad scene, the role of high-resolution images is huge, so the next work in this paper will be aimed at SegNeXt algorithm for the railroad scene to improve (Guo *et al.*, 2022).

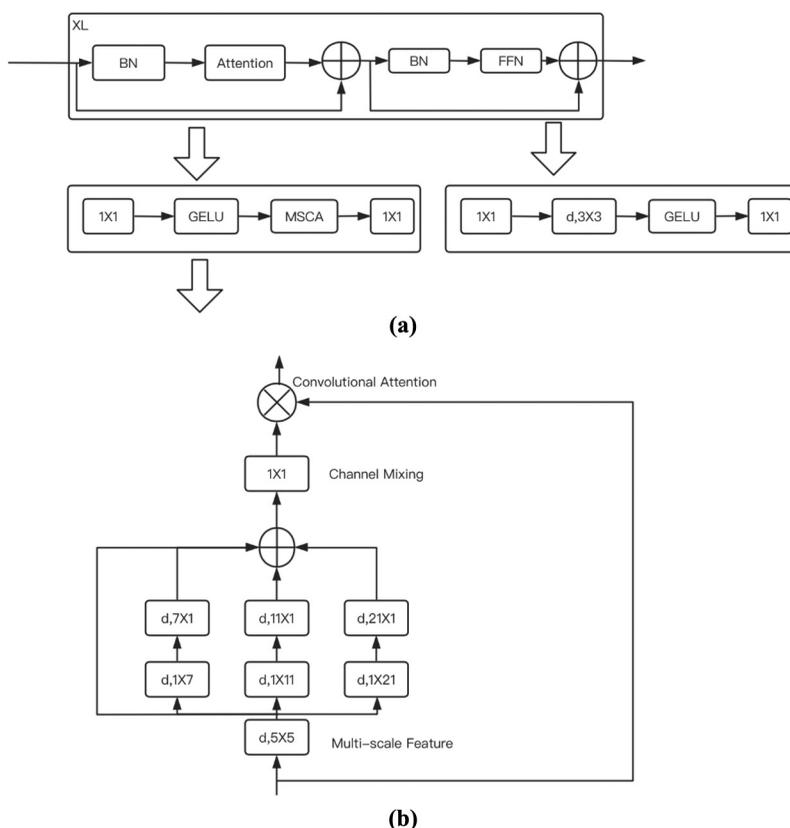
### 2.2 Improvement of SegNeXt algorithm

In response to the work done in the first subsection, this paper finds that SegNeXt performs better on railroads. It is well known that in the era of deep learning, encoder and decoder are the two major parts of the mainstream segmentation model (Xu *et al.*, 2023); and for the encoder, mainstream networks such as DeepLabV3+ (Su *et al.*, 2023; Chang *et al.*, 2022) and Mask2Former (Cheng *et al.*, 2022) are using the popular classification network (resnet) (Zhang *et al.*, 2022), which is not specialized for specific scenarios. However, it needs to be shown that semantic segmentation is an intensive prediction task; in some ways, it is different from categorization. And SegNeXt proposes a convolutional neural network (CNN)-based model that obtains larger sensory fields, multiscale information and adaptivity through a simple multiscale convolutional attention (MSCA) mechanism. Therefore, it appeared to be the best performer in the experiments of this paper. However, it cannot be ignored that none of these models have been customized for railroad scenarios, thus limiting the performance of the models in their application. As verified by the experiments in the first subsection of this chapter, when oriented to the railroad environment, the major models struggle with complex strip targets (pole, rail track, etc.), and although there exists a strip conv in SegNeXt that can help to extract the strip targets, mean intersection over union (MIOU) = 34 is far from enough for field applications. Therefore, in this paper, we improve the SegNeXt network based on this and change the original decoder to that of Mask2Former (Chen *et al.*, 2018).

*2.2.1 Encoder design.* In the encoder, this paper follows the design of the original authors of SegNeXt and uses the MSCA module; the MSCA consists of three parts: a deep convolution for aggregating local information, a multibranch deep strip convolution for capturing multiscale contexts and a  $1 \times 1$  convolution for modeling the relationship between different channels. The output of the  $1 \times 1$  convolution is directly used as the attention weights to reweight the input of the MSCA. In math, MSCA can be written as:

$$\text{Att} = \text{Conv}1 \times 1 \left( \sum_{i=0}^3 \text{Scale}_i(\text{DW} - \text{conv}(F)) \right) \quad (1)$$

where  $F$  denotes the input features, Att and Out are the attention and output, respectively,  $\otimes$  denotes the matrix multiplication operation and DW-Conv denotes the deep convolution, Scale,  $i \in \{0, 1, 2, 3\}$ . Each branch uses two depth-direction band convolutions to approximate the standard depth-direction convolution of the large kernel. Banded convolution is complementary to network convolution and helps in extracting features in the form of strips. A series of building blocks are stacked to form the Convolutional Encoder MSCAN as shown in Figure 2. Each stage contains a downsampling block and a bunch of building blocks. The

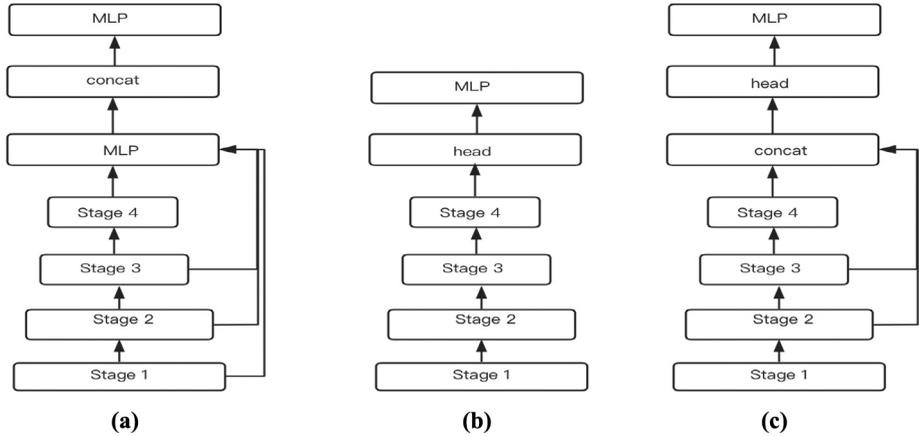


**Notes:** (a) A stage of MSCAN; (b) MSCA  
**Source:** Figure created by authors

**Figure 2.**  
 MSCAN structure  
 diagram

downsampling block has a convolution with a step size of 2 and a kernel size of  $3 \times 3$ , followed by a Batch Norm layer. Note that in each building block of MSCAN, the authors use the batch norm instead of the layer norm because the authors found that the batch norm provides a greater gain in segmentation performance.

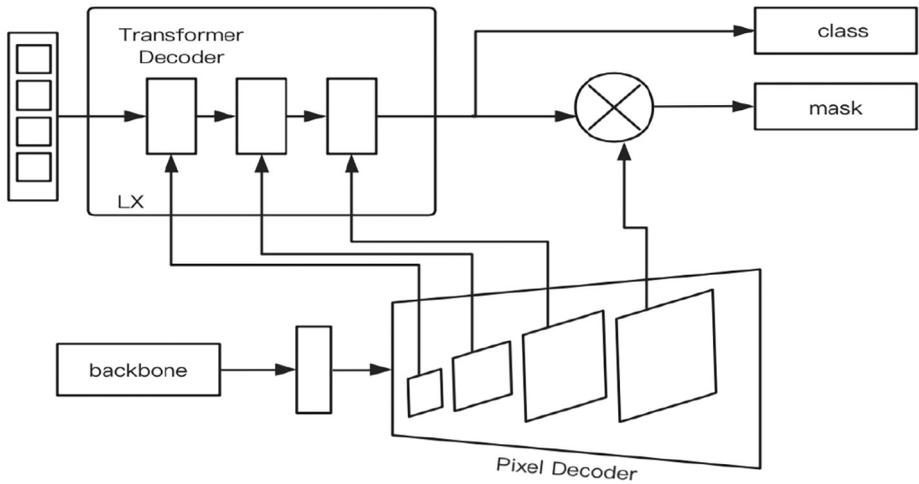
*2.2.2 Decoder design.* In the field of semantic segmentation, encoders are mostly pretrained on data sets and to capture high-level semantics, it is usually necessary to adapt a decoder to it. In the original SegNeXt work, the authors discussed three simple decoder structures as shown in Figure 3. The first one, adopted in SegFormer, is a structure based purely on multilayer perceptron (MLP). The second one mainly uses CNN-based models. In this structure, the output of the encoder is directly used as an input to heavy decoder headers such as Atrous spatial pyramid pooling (ASPP), pyramid scene parsing (PSP) and DANet. The last one is the structure used in SegNeXt. However, in this paper, it is found that the lightweight decoder used in SegNeXt is not able to satisfy the logo segmentation in complex railroad scenarios.



**Figure 3.**  
Three common  
decoder  
configurations

**Notes:** (a)Decoder 1; (b)Decoder 2; (b)Decoder 3  
**Source:** Figure created by author

Through experiments, it is found that the decoder in Mask2Former has significant enhancement for small objects. Therefore, this paper explores the structure of Mask2Former and finds that Mask2Former uses a pixel decoder, as shown in Figure 4, to gradually up-sample low-resolution features from the output of the backbone feature extractor to generate high-resolution per-pixel embeddings. In Mask2Former, the pixel decoder is usually an anticonvolutional network that gradually restores the resolution of the feature map to the size of the original image using an anticonvolution operation. The features output from the encoder are converted into pixel-level predictions to improve the accuracy of image segmentation.



**Figure 4.**  
Pixel decoder

**Source:** Figure created by authors

For the shallowest features output by MSCAN, do addition with the second shallow features output by Pixel Decoder after up-sampling, and then get the mask features after full connectivity, and continue the processing of the mask features, Pixel Decoder's output features, which fully use the multiscale features. And the cross-attention mechanism with mask is executed first, and then the regular operations such as self-attention mechanism are executed to accelerate the convergence of the model. The specific structure is shown in Figure 5.

### 3. Experimental comparison

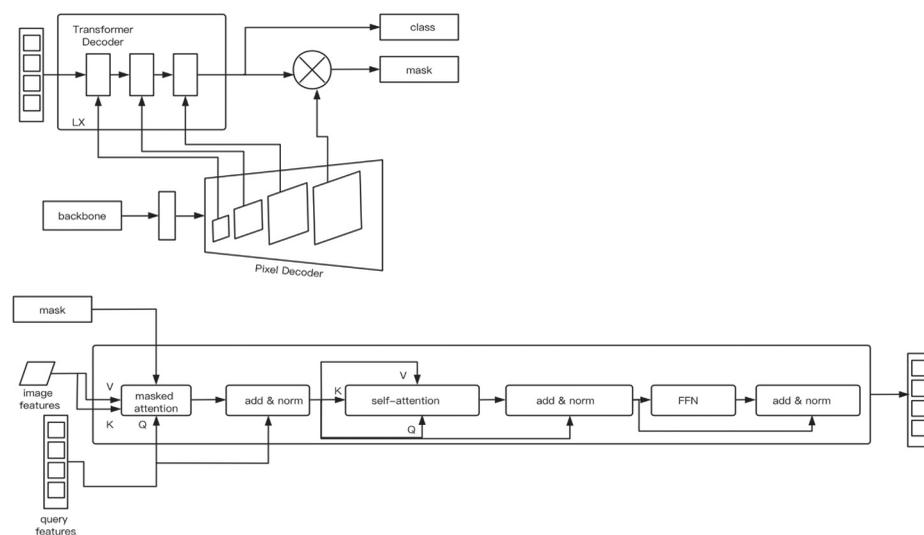
For the improved SegNeXt network in this paper, this paper mainly carries out three stages of experiments: supplementing and adjusting the data set, dividing the images of high, medium and low resolutions; feeding the improved network for training and fine-tuning; and validating and analyzing the performance and practicability of the improved network.

#### 3.1 Experimental setup and evaluation parameters

To make the algorithmic model performance of this paper representative and comparative, this paper uses the same equipment as well as the same strategy and hyperparameters for the experiments. The experimental hardware and software environments are shown in Table 2.

The rest of the hyperparameters are set the same, such as the stochastic gradient descent optimization algorithm is used in all of this paper, the momentum is used 0.9, the initial learning rate is used 0.01, the weight\_decay rate is used 0.0005, the number of training iterations is 60,000, the batch-size is 4, image resolution are  $640 \times 640$  and so on.

In terms of evaluation parameters, to accurately evaluate the performance of the model, this paper uses accuracy (ACC), mean accuracy (MACC), intersection over union (IOU), MIOU, FPS, memory used by hardware and training time. Definitions are shown in equations 1 through 4:



Source: Figure created by authors

Figure 5.  
Mask2former decoder  
framework

$$\text{acc} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (2)$$

$$\text{IoU} = \sum_i \frac{n_{ii}}{t_i + \sum_j n_{ji} - n_{ii}} \quad (3)$$

$$\text{mIoU} = \frac{1}{n_{\text{cls}}} \cdot \sum_i \frac{n_{ii}}{t_i + \sum_j n_{ji} - n_{ii}} \quad (4)$$

### 3.2 Experimental data

The data set selected for this paper is an adapted and improved one based on the publicly available data set RailSem19 (Zendel *et al.*, 2019), which consists of 8,500 short sequences of train driving perspectives, including images of more than 1,000 railroad crossings and 1,200 trolley scenes. However, because the RailSem19 data set is a semisupervised automatic labeling data set, it is still insufficient relative to manual labeling. This paper improves and optimizes the labeling for this data set and improves it in accordance with the standards of the national railroads to make the data set labeling distribution more reasonable and suitable for practical scenarios. To address the problem that this data set has many categories but a small amount of data, this paper supplements the data with 3,000 images covering the categories of the data set RailSem19 by conducting experiments in the railroad test line several times. To improve the generalization ability of the model, data enhancement is performed on the cropped images through the processing of flip transform, color dithering, noise addition, etc. (Wang *et al.*, 2019). Finally, the images are divided into 9,200 training sets and 2,300 test sets according to the ratio of 8:2. RailSem19 data set samples are displayed as shown in Plate 1.

### 3.3 Experimental results and analysis

To comprehensively measure the level of detection of signs on railroads by the model built in this paper, this paper compares the original SegNeXt with the improved SegNeXt in this paper, and models with multiple network depths are selected for comparative experiments. The improved method in this paper has a clear enhancement for the detection of line objects on the adjusted RailSem19 data set, and the experimental results are shown in Table 3.

The experimental results are shown in Figure 6, and it can be seen that the method proposed in this paper improves more for the aspect of linear features.

From the details of Figure 6, it can be seen that SegNeXt is still defective for distant linear features, and jaggedness is obvious (Wang *et al.*, 2019); compared to SegNeXt, the method MIOU proposed in this paper improves 8 points in the case of image input resolution are  $640 \times 640$ , and for long targets, such as pole, rail track and other line objects have a huge

**Table 2.**  
Experimental  
software and  
hardware  
environment

Operating system	Ubuntu
CPU	Genuine Intel(R) Core(TM)i7-7820 CPU @3.60GHZ
GPU	NVIDIA GTX 1080 Ti
CUDA	CUDA 10.2
PYTORCH	PYTORCH1.12.1
PYTHON	PYTHON 3.8
MMCV	2.0.0

**Source:** Table created by authors

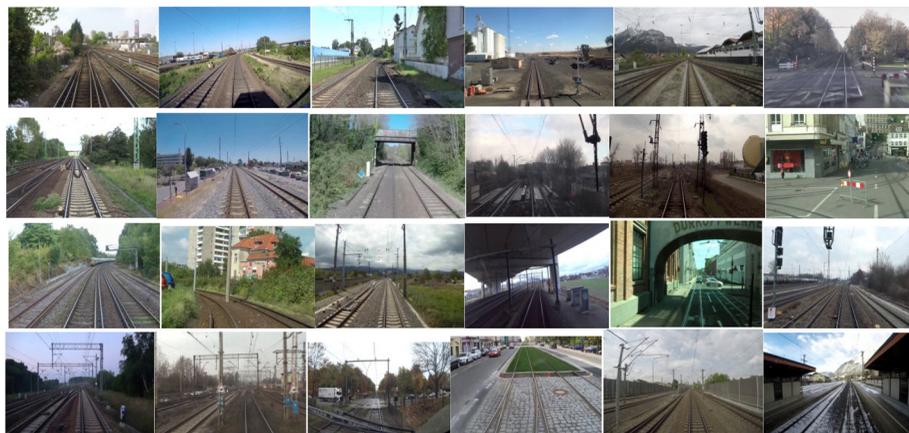


Plate 1

RailSem19 data set  
samples

Source: Figure courtesy of Zendel *et al.* (2019)

Model	Resolution ratio	ACC	MACC	Traffic sign	IOU		FPS
					Road	Pole MIOU	
SegNext	512 × 512	88.77	73.02	65.56	75.47	39.91 62.85	13
	640 × 640	89.95	79.69	69.55	76.34	50.96 70.19	4
	1,080 × 1,080	89.84	79.67	69.45	76.31	50.39 69.85	2
Proposed model	512 × 512	93.22	88.2	74.29	78.79	60.64 68.38	6
	640 × 640	92.86	87.69	77.19	82.94	65.35 78.23	2
	1,080 × 1,080	92.86	87.69	73.19	76.96	63.27 70.27	1

Note: Better performance is shown in red

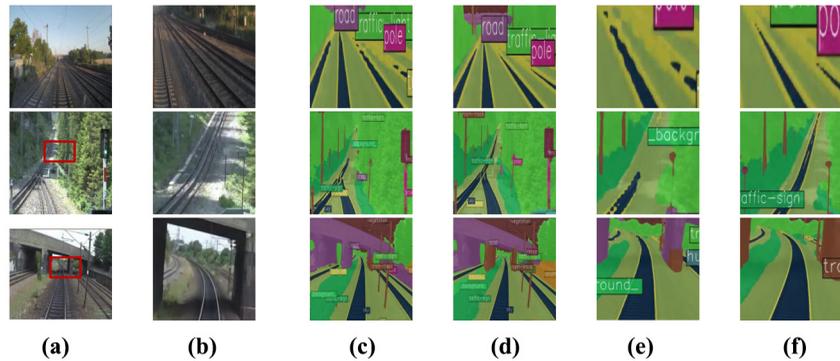
Source: Table created by authors

Table 3.  
Comparison of  
improvement results

improvement, such as pole's IOU increased by 14.39%, ACC increased by 15.3%; for the rail track the MIOU is boosted by 11.16%. Therefore, the method proposed in this paper has been greatly improved in the railroad.

#### 4. Conclusion

This article proposes an improved SegNeXt algorithm: first, it explores the performance of various models on railways, studies the problems of semantic segmentation on railways and then analyzes the specific problems. On the basis of retaining the original excellent MSCAN encoder of SegNeXt, multiscale information fusion is used to further extract detailed features such as multihead attention and mask, solving the problem of inaccurate segmentation of current objects by the original SegNeXt algorithm. The improved algorithm is of great significance for the segmentation and recognition of railway signs. However, it is found through the results (as shown in Table 3) that the model developed in this paper has accuracy differences when segmenting images of different scales. Further experiments in this paper found that the proposed model was improved only with the help of multiscale information for learning and training, but did not pay sufficient attention to the contextual



**Notes:** (a) Original figure; (b) Detailed drawing ( $\times 4$ ); (c) Segnext; (d) Our; (e) Detailed drawing; (f) Detailed drawing. The region selected by the red box in (a) is zoomed in to show the detail region; (b) is shown by zooming in 4 times for the region picture; (c) and (d) are shown for the whole picture segmentation; (e) and (f) are shown for the semantic segmentation of the detail region before and after the improvement

**Source:** Figure created by authors

**Figure 6.**  
Experimental effect  
diagram

information; so in the future, we will continue to improve and optimize this phenomenon to improve the performance of the model in different resolution scenarios.

## References

- Chang, Z., Xiaoka, Y., Lu, Y. and Hao, Z. (2022), "A study on change detection in high-resolution remote sensing images based on improved DeepLabv3+", *Laser and Optoelectronics Progress*, Vol. 59 No. 12, 1228006-1228006-12.
- Chen, L.C., Zhu, Y., Papandreou, G., Schro, F. and Hartwig, A. (2018), "Encoder-decoder with atrous separable convolution for semantic image segmentation", *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 801-818.
- Cheng, B., Misra, I., Schwing, A.G., Kirillov, A. and Girdhar, R. (2022), "Masked-attention mask transformer for universal image segmentation", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1290-1299.
- Guo, M.H., Lu, C.Z., Hou, Q., Liu, Z.N., Cheng, M.M. and Hu, S.M. (2022), "Segnext: rethinking convolutional attention design for semantic segmentation", *Advances in Neural Information Processing Systems*, Vol. 35, pp. 1140-1156.
- Li, C. (2022), *Research on Railroad Perimeter Segmentation and Anomaly Perception Based on Computer Vision*, Jiaotong University, Beijing, doi: [10.26944/d.cnki.gbfju.2022.003365](https://doi.org/10.26944/d.cnki.gbfju.2022.003365).
- Lu, T., Zujun, Y., Bao-qing, G. and Tao, R. (2023), "Semantic segmentation of railroad scenes with mesh multiscale and two-way channel attention", *Transportation Systems Engineering and Information*, Vol. 23 No. 2, p. 233.
- Liu, L., Guangdi, H. and Jianyao, H. (2023), "Lane line detection method based on instance segmentation", *Computer and Digital Engineering*, Vol. 51 No. 7, pp. 1620-1625.
- Ma, W.Q., Shi, J. and Wu, H.J. (2023), "A review of deep convolutional neural network semantic segmentation", *Microelectronics and Computers*, Vol. 2023 No. 9, pp. 55-64, doi: [10.19304/J.ISSN1000-7180.2022.0825](https://doi.org/10.19304/J.ISSN1000-7180.2022.0825).

- 
- Song, X. and Ge, H. (2023), "TransMANet-based semantic segmentation algorithm for remote sensing images", *Advances in Lasers and Optoelectronics*, pp. 1-21, available at: <http://kns.cnki.net/kcms/detail/31.1690.TN.20231108.0935.026.html> (accessed 14 November 2023).
- Su, Z., Li, J., Jiang, J., Lu, Y. and Zhu, M. (2023), "A semantic segmentation method for remote sensing images based on improved DeepLabV3+", *Laser and Optoelectronics Progress*, Vol. 60 No. 6, pp. 0628003-0628003-8.
- Wang, C.Y., Ni, H.Y. and Shang, Z.D. (2019), "Semantic segmentation of autonomous driving scenes using convolutional neural networks", *Optics and Precision Engineering*, Vol. 27 No. 11, pp. 2429-2438.
- Wang, B., Han, Y., Cui, H., Liu, Y., Ren, M., Gao, W., Chen, S., Liu, J. and Cui, Y. (2023), "Image-based road semantic segmentation detection method", *Journal of Shandong University (Engineering Edition)*, Vol. 5, pp. 37-47, available at: <http://kns.cnki.net/kcms/detail/37.1391.T.20231020.1005.008.html> (accessed 27 October 2023).
- Xu, M., Zhang, Z., Wei, F., Hu, H. and Xiang, B. (2023), "SAN: Side adapter network for open-vocabulary semantic segmentation", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2945-2954.
- Zendel, O., Murschitz, M., Zeilinger, M., Steininger, D., Abbasi, S. and Beleznai, C. (2019), "Railsem19: a dataset for semantic rail scene understanding", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*.
- Zhai, X. and Li, X. (2021), "Research on railroad instance segmentation based on deep learning", *Journal of Jiamusi University(Natural Science Edition)*, Vol. 39 No. 6, pp. 116-118+147.
- Zhang, H., Wu, C., Zhang, Z., Yi, Z., Lin, H., Zhang, Z., Sun, Y., He, T., Mueller, J., Manmatha, R., Li, M. and Smola, A. (2022), "Resnest: split-attention networks", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2736-2746.
- Zhang, J., Yuan, J., Hui, Y., Hu, Y. and Zhang, Y. (2023), "Research on semantic segmentation method based on autonomous driving road scene", *Mechanical Engineering and Automation*, Vol. 2023 No. 6, pp. 15-18.

### Further reading

- Yang, W., Liqiang, Z., Zujun, Y. and Bao-qing, G. (2019), "Segmentation and recognition algorithm for high-speed railroad scenes", *Acta Optica Sinica*, Vol. 39 No. 6, p. 610004.

### Corresponding author

Zheng Yu Xie can be contacted at: [xiezhengyu@bjtu.edu.cn](mailto:xiezhengyu@bjtu.edu.cn)

---

For instructions on how to order reprints of this article, please visit our website:

[www.emeraldgrouppublishing.com/licensing/reprints.htm](http://www.emeraldgrouppublishing.com/licensing/reprints.htm)

Or contact us for further details: [permissions@emeraldinsight.com](mailto:permissions@emeraldinsight.com)