THE PUBLISHING HOUSE
OF THE ROMANIAN ACADEMY

PROCEEDINGS OF THE ROMANIAN ACADEMY, Series A,
Volume 25, Number 2/2024, pp. 157–164
DOI: 10.59277/PRA-SER.A.25.2.10

WELD IMAGE SEGMENTATION IN INDUSTRIAL SMOKE SCENE

Xu ZHANG[1], Qingchun ZHENG[1,2,3], Peihao ZHU[2,3], Yangyang ZHAO[1], Jiwei LIU[1]

[1] Tianjin University of Technology, School of Computer Science and Engineering, Tianjin, China
[2] Tianjin University of Technology, Tianjin Key Laboratory for Advanced Mechatronic System Design and Intelligent Control, Tianjin, China
[3] Tianjin University of Technology, National Demonstration Center for Experimental Mechanical and Electrical Engineering Education, Tianjin, China
Corresponding author: Peihao ZHU, E-mail: zhupeihao_gp@163.com

Abstract. Weld recognition is the premise of automatic weld polishing, and weld image segmentation can provide key area information for robots. With the advent of large segmentation model, it will be more convenient to realize weld image segmentation. With the emergence of complex scenes such as smoke, how to achieve high precision weld image segmentation under different smoke concentrations has become a challenge. To solve this problem, we propose a lightweight weld segmentation approach in smoke scenes. The feature transformation can better realize the feature processing of the smoke weld image, and further combine with the large segmentation model to realize the smoke weld image segmentation. The experimental data show that the segmentation accuracy of the weld segmentation approach we proposed achieves 98.18% in everything mode, increasing by 0.67% and 11.64% compared with the typical comparison methods, respectively. And the feature transformation is relatively lightweight.

Keywords*:* image segmentation, weld segmentation, feature transformation, smoke scene.

1. INTRODUCTION

In robotic weld polishing, the accurate positioning of weld region is one of the key steps to ensure high quality polishing. To effectively achieve the function, weld image segmentation has become an indispensable work [1–2]. With the image segmentation technology, the weld can be clearly separated from the surrounding environment, and the robot can be provided with clear information of the weld boundary. The combination of information conversion and image segmentation enables the robot to accurately and efficiently locate the weld area, which provides the necessary basis for subsequent automated grinding.

Many advances have been made in image segmentation, including traditional segmentation methods and deep learning segmentation methods. Traditional methods are mainly based on hand-designed features and mathematical models. And the traditional image segmentation algorithm also plays a certain role in some fields. In recent years, deep learning image segmentation methods have become a research hotspot and have attracted wide attention. Deep learning segmentation methods are mainly divided into semantic segmentation [3–4], instance segmentation [5–8], panoramic segmentation [9–11] and large segmentation model [12]. By constructing complex neural networks, deep learning methods can automatically learn features and patterns [13–15], so that those methods can be more suitable for different weld shapes and complexities. Supervised and unsupervised methods are two main directions in deep learning. With the help of labeled training data, supervised methods learn pixel-level annotations in images through neural networks to achieve accurate segmentation of objects such as welds. This method works well when there is enough data, but the disadvantage is that it requires a large amount of labeled data for training. In contrast, unsupervised methods avoid the dependence on large amounts of labeled data by autonomously learning the latent structures and patterns in images. This is particularly important for industrial automation applications, since obtaining annotated data in real scenarios is often an expensive and time-consuming task.

At present, the research on weld segmentation is still relatively few and scarce existing works mainly focus on X-ray weld segmentation. Rathod *et al*. used three basic segmentation methods to achieve defect detection of ship welds [16]. Wang *et al*. used convolutional neural network and spatial pyramid model to extract weld features and finally achieved 93.70% segmentation accuracy [17]. Zhang *et al*. proposed the boundary label smoothing strategy and combined it with dice loss to improve the segmentation performance [18]. In general, these segmentation studies on welds have an enlightening effect on our research. But the working environment of the weld is usually full of smoke, and the smoke degree of the weld varies with different welding conditions, forming a multi-level and multi-density smoke environment. This characteristic increases the difficulty for weld segmentation, because the smoke may blur the contour of the weld, resulting in the weld boundary being difficult to clearly define. Under such conditions, smoke can cause illumination changes and color distortion, making it difficult to accurately extract edge information. And existing image segmentation algorithms may not perform well when dealing with dense smoke.

With the continuous development of transformer, clip and prompt, large segmentation models began to appear in 2023, and the first model was SAM [12]. Due to the strong versatility of the large segmentation model and the support for multi-functional segmentation with prompt, this technology has become a hot research topic in image segmentation. Large segmentation models have been well applied in the medical field. Ma *et al*. proposed a unified segmentation model MedSAM [19], which was proved to be more robust than the expert model by experiments. Mazurowski et al. similarly work on medical segmentation models and give critical arguments [20]. In 3D scene segmentation, some methods have been tried to put forward, which further benefit fine-grained segmentation and 3D reconstruction [21−22]. In addition, Wang *et al*. proposed an image segmentation algorithm with context reasoning ability [23], which has good performance in the fields of semantic segmentation, panoramic segmentation and video segmentation. Zou *et al*. proposed a large model with interactive segmentation [24], which has the new feature of visual prompt. In weld images, how to solve the weld segmentation under smoke conditions is a difficult problem. The motivation of this paper is to propose a weld segmentation approach that can be applied to smoke environments. The combination of weld segmentation and large segmentation model will be a good solution.

The main contributions of this paper are summarized as follows:

1. In view of the complex working conditions, we propose an approach for weld image segmentation.
2. For weld segmentation in smoke environment, we propose to use feature transformation network for feature processing, and show the detailed network structure.
3. The weld segmentation approach can adapt to different smoke environments and has good adaptability to multiple smoke concentrations.

The remaining chapters of this paper are as follows. In Section 2, we study the overall approach of weld image segmentation. In Section 3, the performance test is studied. In Section 4, the discussion is conducted. In Section 5, we summarized our research progress.

2. WELD SEGMENTATION APPROACH

To realize the weld segmentation under complex working conditions, we propose a weld segmentation approach. The main architecture of the proposed approach is shown in Fig. 1, and the approach can realize the segmentation enhancement of welds in smoke environment. The approach mainly includes two main parts: feature transformation and segmentation net.

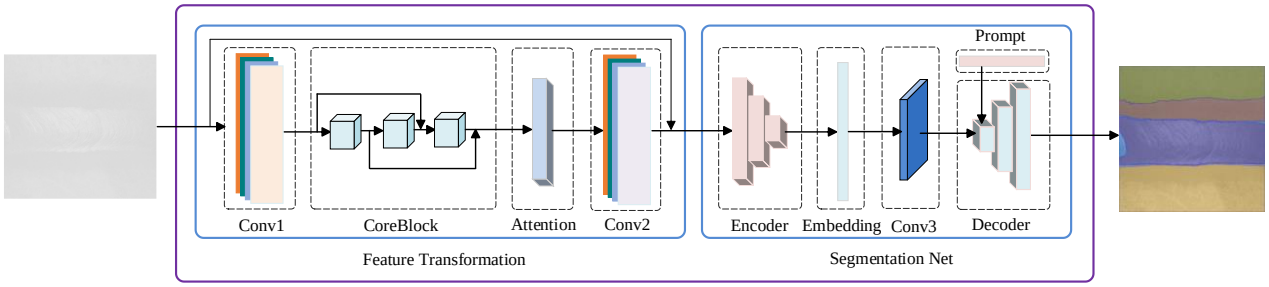
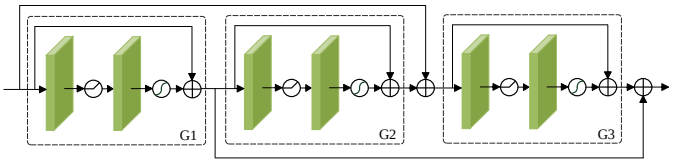


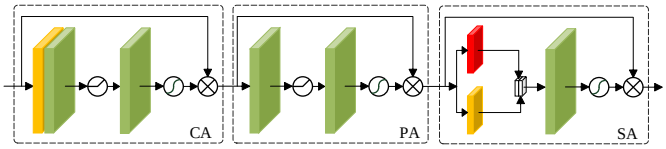
Fig. 1 – Weld segmentation approach.

The main function of feature transformation is to realize the feature processing and feature extraction of smoke weld images. Its network is mainly composed of Conv1, CoreBlock, Attention and Conv2, where CoreBlock is designed as a skip connection form to enhance feature extraction ability and improve gradient propagation. Each unit of CoreBlock has two layers of convolutions with the skip connection inside.

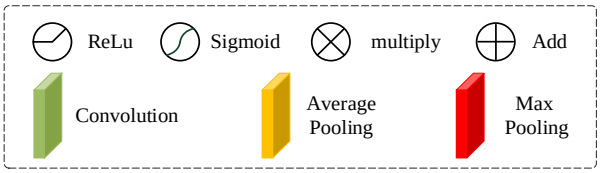
The detailed structure of CoreBlock and Attention in Fig. 1 is shown in Fig. 2. More precisely, Fig. 2a shows the detailed structure of the CoreBlock module, and the CoreBlock is mainly composed of G1, G2 and G3 modules. The structure of these three G modules is the same, and they are all connected by two convolutional layers, a Relu layer and a Sigmoid layer in the form of skip connections. Figure 2b shows the detailed structure of the Attention module. The Attention module is mainly composed of channel attention mechanism, pixel attention mechanism and spatial attention mechanism in series to further refine the extracted features. Fig. 2c shows the relevant operators.



(a) CoreBlock



(b) Attention



(c) Operators

Fig. 2 – Detailed structure of CoreBlock and Attention.

Combined with the structure of feature transformation, its internal core calculation can continue to be described in the form of equations. The feature output of the CoreBlock module is $\boldsymbol{F}$, which first passes through the channel attention mechanism. And the output of the channel attention mechanism is $\boldsymbol{F}_1$, which can be calculated as follows:

$$\boldsymbol{F}_1 = \boldsymbol{F} \otimes \boldsymbol{CA}, \tag{1}$$

$$\boldsymbol{CA} = \sigma(\mathrm{conv}(\lambda(\mathrm{conv}(\boldsymbol{A})))), \tag{2}$$

$$a_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \boldsymbol{x}_c(i,j), \tag{3}$$

where $\otimes$ denotes element-wise multiplication, λ is the ReLu function, σ is sigmoid function, $\boldsymbol{x}_c \in \mathbb{R}^{H \times W}$ is the c-th feature map in $\boldsymbol{F} \in \mathbb{R}^{C \times H \times W}$, a_c is the c-th element in $\boldsymbol{A} \in \mathbb{R}^{C \times 1 \times 1}$, H and W are the height and width of the feature map respectively, and $conv$ is the convolution operation.

The output $\boldsymbol{F}_1$ of the channel attention mechanism is the input of the pixel attention mechanism, and the output $\boldsymbol{F}_2$ of the pixel attention mechanism can be calculated as follows:

$$\boldsymbol{F}_2 = \boldsymbol{F}_1 \otimes \boldsymbol{PA}, \tag{4}$$

$$\boldsymbol{PA} = \sigma(\mathrm{conv}(\lambda(\mathrm{conv}(\boldsymbol{F}_1)))). \tag{5}$$

The output $\boldsymbol{F}_2$ of the pixel attention mechanism is the input of the spatial attention mechanism, and the output $\boldsymbol{F}_3$ of the spatial attention mechanism can be calculated as follows:

$$\boldsymbol{F}_3 = \boldsymbol{F}_2 \otimes \boldsymbol{SA}, \tag{6}$$

$$\boldsymbol{SA} = \sigma(\text{conv}(\text{concat}(\max p(\boldsymbol{F}_2), \text{ave}\, p(\boldsymbol{F}_2)))), \tag{7}$$

where $\max p$ is the max pooling calculation, $\text{ave}\, p$ is the average pooling calculation, and $concat$ is the concatenation operation of the feature map.

The function of segmentation net is to realize the weld features processed by the feature transformation module and realize the weld image segmentation. The network is mainly composed of Encoder, Embedding, Conv3, Decoder, Prompt, which has zero-shot segmentation ability and super generalization ability. Prompt supports everything and box mode. The segmentation net in Fig. 2 adopts the pre-trained model, and the pre-trained parameters are consistent with the literature [25].

3. RESULTS

3.1. Test setup

Figure 1 is a two-stage training approach, and the approach first trains the feature transformation module. The computer hardware is i7-10700U CPU 2.9GHz, 16 G memory, and 2G video memory. The programming software is python version 3.8.5, and the deep learning framework is Pytorch. The data set for one-stage training is RESIDE, the number of training steps is 20000, the initial learning rate is 0.0001, the optimizer is Adam, and the crop size is 100.The segmentation net module is trained in the second stage. To speed up the experiment, the pretrained model is directly used to load the parameters. In total, two types of large pre-trained segmentation model parameters are used, these two models are SAM [12] and FastSAM [25]. For a more detailed setup, please refer to the following link: https://github.com/windrunners/transformation-segmentation.

3.2. Feature transformation training

The feature transformation is the core and the performance evaluation indexes of the network are PSNR and SSIM, which can better reflect the feature transformation performance of the weld smoke image. Figures 3 and 4 show the performance change trend of this network training, where Fig. 3 shows the trend of PSNR, while Fig. 4 shows the trend of SSIM. To speed up the training, the performance parameters are saved for each step when the number of training steps is less than 20, every 10 steps when the number of training steps is more than 20 and less than 200, and every 20 steps when the number of training steps is more than 200.

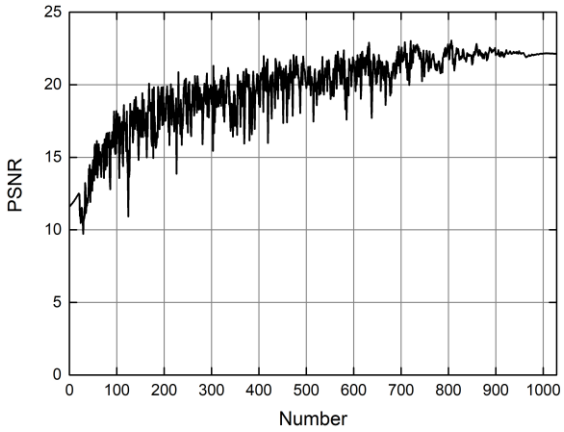


Fig. 3 – PSNR trend during training for feature transformation

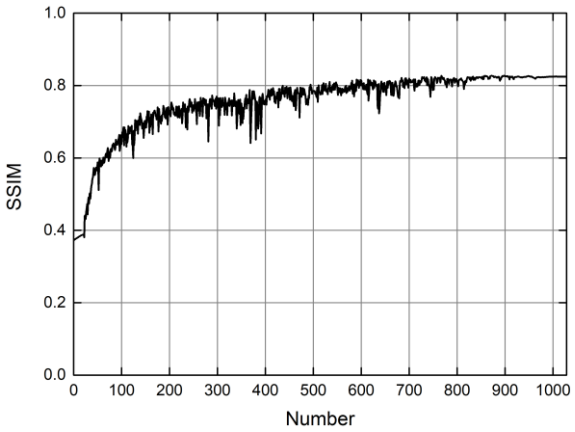


Fig. 4 – SSIM trend during training for feature transformation

It can be seen from Figs. 3 and 4 that both PSNR and SSIM have certain fluctuations in the early training period of the feature transformation module. From the whole training trend, although there are fluctuations, its performance is in a steady upward trend. When the training number reaches 900, the PSNR and SSIM of the network model have become stable.

3.3. Weld segmentation test

To verify the performance of the proposed method, welds with different smoke concentrations are provided, and these weld images are used for segmentation test. The FFA algorithm is used to preprocess the compared algorithms, and then large segmentation models such as SAM and FastSAM are used for further segmentation. The compared methods include FFA-SAM and FFA-FastSAM.

To quantitatively analyze the weld segmentation performance, we provide the average accuracy results of different methods in the two modes of everything and box, as shown in Table 1. It can be seen that in everything mode, compared with the other two methods, the accuracy of our method is improved by 0.67% and 11.64% respectively. At the same time, in box mode, the accuracy of our method is improved by 0.32% and 0.51%, respectively. From the perspective of accuracy, our proposed method has higher segmentation accuracy.

Table 1

Weld segmentation accuracy under different smoke concentrations and different methods

Method	Box mode	Everything mode
FFA-FastSAM	0.9938±0.005	0.9753±0.005
FFA-SAM	0.9919±0.003	0.8794±0.170
ours	0.9970±0.002	0.9818±0.004

For a more detailed presentation, we provide the image segmentation comparison of some weld images under different concentrations and different methods, as shown in Fig. 5. The smoke concentration levels are divided into A to E 5 levels totally, and the smoke concentration increases step by step. The test result shows the performance comparison of different methods using the everything mode for segmentation. It can be seen that the FFA-SAM method has some error at grade A and grade C concentrations, while the error is the largest at grade E concentration. FFA-FastSAM has some errors at grade C and grade D concentrations, and the edge of the segmented weld at grade E concentration is not smooth. However, our proposed method can segment well under all the smoke concentration levels. In general, our method achieves the better weld segmentation performance under everything mode.

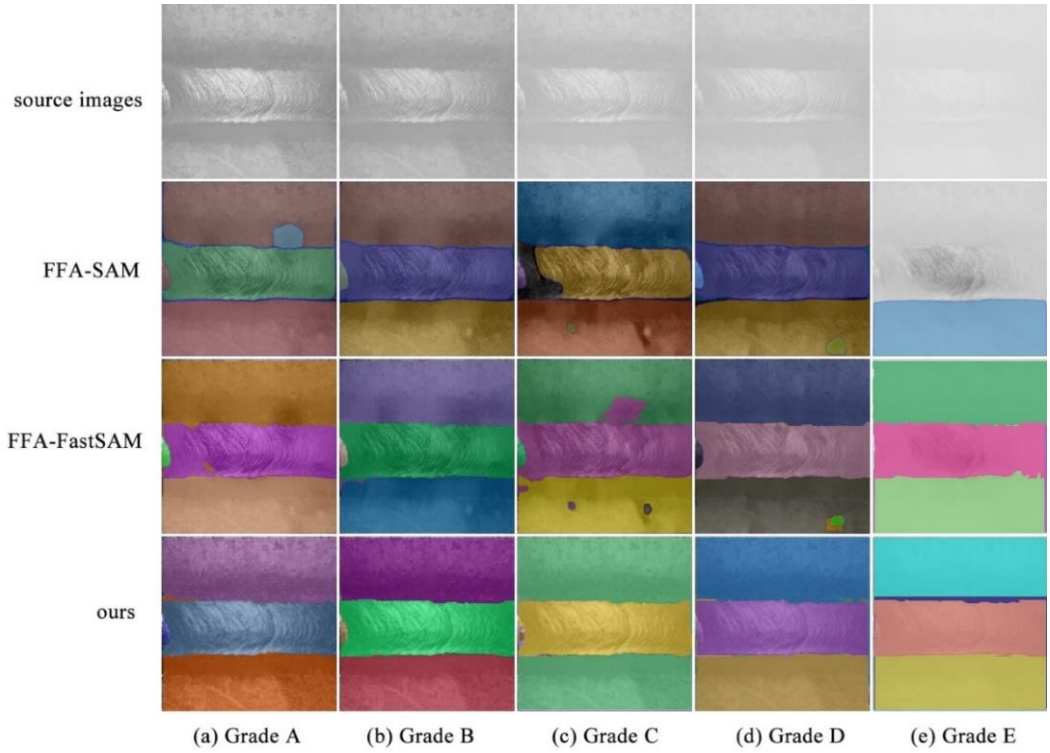


Fig. 5 – Segmentation comparison of different methods and smoke concentration using the everything mode.

3.4. Ablation experiment

To better verify the effectiveness of the proposed method, we continue to conduct ablation experimental studies. The focused network structure is the feature transformation module, and the compared network structures are the network structure in FFA algorithm [26] and the feature transformation structure we proposed. The intermediate features of weld images with different smoke obtained by two different feature transformation networks are shown in Fig. 6. It can be seen that our method has stronger ability to remove smoke in the intermediate link than the FFA method. The performance of feature transformation will further affect the performance of the large segmentation model, which has a great impact on the weld segmentation under the condition of smoke.

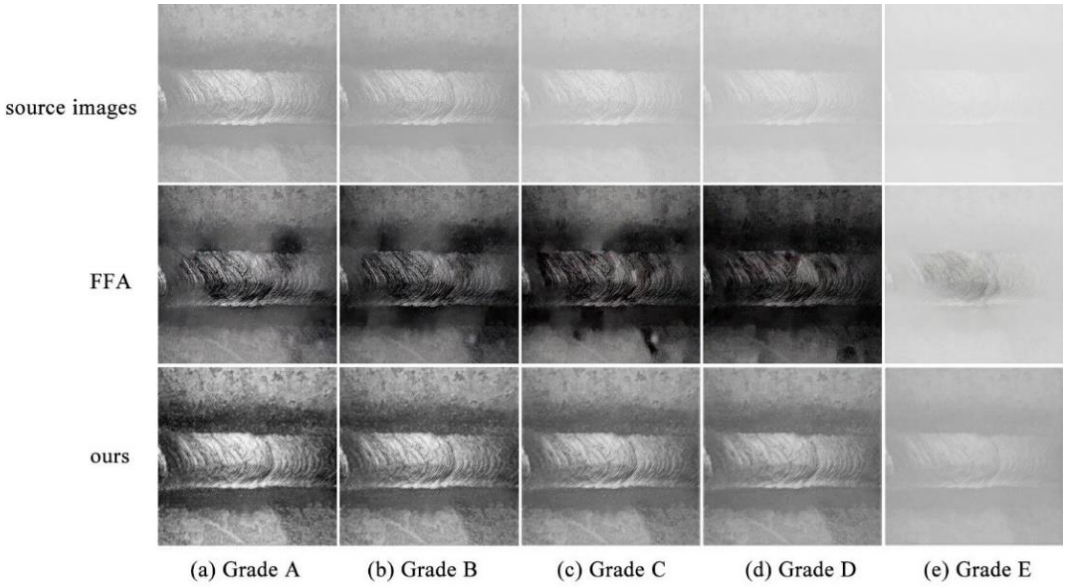


Fig. 6 – Comparison of feature transformation performance of different networks.

4. DISCUSSION

In this paper, an industrial weld segmentation method for complex smoke environment is proposed. The experimental results show that the proposed method is effective. In the experiment, the performance comparison of different methods in the two segmentation modes of everything and box is provided. The experimental results show that the proposed method has better segmentation performance in everything mode than in box mode. In the actual weld segmentation application, the everything mode is more widely used than the box mode.

The presence of smoke is a relatively common phenomenon in the environment of the weld. For weld segmentation of different smoke concentrations, our proposed algorithm is more effective in high smoke concentration than in low smoke concentration.

The reason why our method can work is mainly due to the carefully designed feature transformation module. From the perspective of network parameter size, the network parameter of our proposed method for feature transformation only has 0.26M, which has a good lightweight performance. Therefore, the time of our method for feature transformation of weld image becomes faster.

The weld segmentation task is sensitive to the processing speed. Although the weld segmentation studied in this paper is under complex working conditions, it is worth pursuing to keep the segmentation speed as fast as possible, because it will be beneficial to improve the timeliness of subsequent tasks. We continue to compare average processing time using the comparison method and our method in feature transformation module. The specific performance can be explained as follows. The average feature transformation time of weld images using the FFA network is 3.85 seconds, while it only takes 0.19 seconds using our method. Compared with the comparison method, our method improves the processing time to a greater degree on the same hardware.

5. CONCLUSION

In this paper, a segmentation approach with feature transformation is proposed, which can well deal with weld segmentation in smoke environment. The feature transformation network of the proposed method is relatively lightweight, and the network parameter size is only 0.26M. Compared with FFA-SAM method and FFA-FastSAM method, the proposed method has better weld segmentation accuracy in everything mode. Compared with the FFA network, the average running speed of our method is increased in the feature transformation. In contrast, the weld segmentation performance of our method is better under high smoke concentration.

ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China (Grant No. 62073239).

REFERENCES

[1] Malarvel M, Sethumadhavan G, Bhagi PCR, Kar S, Thangavel S. An improved version of Otsu's method for segmentation of weld defects on X-radiography images. Optik 2017;142:109–118.
[2] Yang L, Song S, Fan J, Huo B, Li E, Liu Y. An automatic deep segmentation network for pixel-level welding defect detection. IEEE Transactions on Instrumentation and Measurement 2021;71:5003510.
[3] Zhou T, Li L, Li X, Feng CM, Li J, Shao L, Group-wise learning for weakly supervised semantic segmentation. IEEE Transactions on Image Processing 2021;31:799–811.
[4] Liu C, Chen LC, Schroff F, Adam H, Hua W, Yuille AL, Li FF. Auto-DeepLab: hierarchical neural architecture search for semantic image segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019, 82–92.
[5] Gu W, Bai S, Kong L. A review on 2D instance segmentation based on deep neural networks. Image and Vision Computing 2022; 120:104401.
[6] Xie E, Sun P, Song X, Wang W, Liu X, Liang D, Shen C, Luo P. Polarmask: single shot instance segmentation with polar representation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020, 12193–12202.
[7] Wang S, Gong Y, Xing J, Huang L, Huang C, Hu W, RDSNet: a new deep architecture for reciprocal object detection and instance segmentation. Proceedings of the AAAI Conference on Artificial Intelligence. 2020, 12208–12215.
[8] Kirillov A, Wu Y, He K, Girshick R, PointRend: image segmentation as rendering. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020, 9799–9808.
[9] Cheng B, Collins MD, Zhu Y, Liu T, Huang TS, Adam H, Chen LC. Panoptic-DeepLab: a simple, strong, and fast baseline for bottom-up panoptic segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020, 12475–12485.
[10] Li Y, Zhao H, Qi X, Wang L, Li Z, Sun J, Jia J. Fully convolutional networks for panoptic segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021, 214–223.
[11] Wang H, Zhu Y, Adam H, Yuille A, Chen LC. Max-DeepLab: end-to-end panoptic segmentation with mask transformers. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021, 5463–5474.
[12] Kirilov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg AC, Lo WY, Dollar P, Girshick. Segment anything. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023, 4015–4026.
[13] Tang Y, Wang Y, Liu C, Yuan X, Wang K, Yang C. Semi-supervised LSTM with historical feature fusion attention for temporal sequence dynamic modeling in industrial processes. Engineering Applications of Artificial Intelligence 2023;117:105547.
[14] Yuan X, Wang Y, Yang C, Gui W. Stacked isomorphic autoencoder based soft analyzer and its application to sulfur recovery unit. Information Sciences 2020;534:72–84.
[15] Bobadilla J, Gutierrez A. Wasserstein GAN-based architecture to generate collaborative filtering synthetic datasets. Applied Intelligence 2024;54:2472–2490.
[16] Rathod VR, Anand RS. A comparative study of different segmentation techniques for detection of flaws in NDE weld images. Journal of Nondestructive Evaluation 2012;31:1–16.
[17] Wang L, Shen Q. Visual inspection of welding zone by boundary-aware semantic segmentation algorithm. IEEE Transactions on Instrumentation and Measurement 2020;70:5001309.
[18] Zhang J, Guo M, Chu P, Liu Y, Chen J, Liu H. Weld defect segmentation in X-ray image with boundary label smoothing. Applied Sciences 2022;12(24):12818.
[19] Ma J, He Y, Li F, Han L, You C, Wang B, Segment anything in medical images. Nature Communications 2024;15:654.
[20] Mazurowski MA, Dong H, Gu H, Yang J, Konz N, Zhang Y. Segment anything model for medical image analysis: an experimental study. Medical Image Analysis 2023;89:102918.
[21] Shen Q, Yang X, Wang X, Anything-3d: towards single-view anything reconstruction in the wild. arXiv 2023: 2304.10261.
[22] Yang Y, Wu X, He T, Zhao H, Liu X, SAM3D: Segment anything in 3D scenes. arXiv 2023: 2306.03908.

[23] Wang X, Zhang X, Cao Y, Wang W, Shen C, Huang T. SegGPT: towards segmenting everything in context. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023, 1130−1140.
[24] Zou X, Yang J, Zhang H, Li F, Li L, Wang J, Wang L, Gao J, Lee YJ, Segment everything everywhere all at once. arXiv 2023: 2304.06718.
[25] Zhao X, Ding W, An Y, Du Y, Yu T, Li M, Tang M, Wang J, Fast segment anything. arXiv 2023: 2306.12156.
[26] Qin X, Wang Z, Bai Y, Xie X, Jia H, FFA-Net: feature fusion attention network for single image dehazing. Proceedings of the AAAI Conference on Artificial Intelligence. 2020, 11908−11915.

Received February 28, 2024