Journal of Image Processing Theory and Applications (2022)
Clausius Scientific Press, Canada

DOI: 10.23977/jipta.2022.050109
ISSN 2560-6239 Vol. 5 Num. 1

Improved Parts Drawing Segmentation Method Based on U-net

Dan Tian*, Hui Yao, Zhiwei Song, Gaohui Zhan, Zhijie Wang

School of Mechatronic Engineering, Xi'an Technological University, Xi'an, Shaanxi, 710021, China
**Corresponding author*

Keywords: Part graph closed contour, Attention mechanism, U-net, Pyramid structure

Abstract: A U-net segmentation network with attention mechanism and pyramid structure is proposed for the problem of low accuracy of closed contour shape recognition of part diagram. The spatial pyramid structure is added before the UpSampling operation of the decoder module of the classical U-net network to expand the perceptual field and reduce the loss of feature details. Meanwhile, the spatial and channel attention mechanisms are added in the middle of UpSampling and convolution of the decoder to extract more significant semantic information. The comparison with the classical U-net analysis shows that this method improves the mean intersection ratio (MIoU) by 7.05%, the category average pixel accuracy (mPA) by 14.63, and Accu by 16.49%, and the experimental results verify that this method can improve the segmentation accuracy of the model in an effective way.

1. Introduction

In this paper, the semantic segmentation method is used to extract the closed contour of the part drawing to improve manufacturing efficiency. Pixels are classified using a semantic segmentation network, and closed contours of different shapes are marked with different colors. The U-net semantic segmentation network has excellent performance and can achieve very good segmentation results on very small datasets. Many scholars have made certain improvements based on the classic U-net network and applied them to other fields. Based on the excellent performance of deep learning in biomedical image segmentation, more and more researchers focus on this direction. Therefore, in recent years, many excellent segmentation networks have appeared in medical modality recognition, the most classic ones are (DUC, HDC) [1] Enet [2], PSPNet [3], SegNet [4], UpperNet [5], GCN [6] and Deeplab [7,8] series. Although many network performances surpass these classic networks, whether it is to propose depthwise separable convolution or hole convolution, or scene perception, the ideas of the predecessors are very worth learning. The U-net-based variant network is a network that improves the segmentation performance by changing the classical U-net network structure. At present, the U-net network variant is usually an improvement of the U-net jumper structure. Such as R2UNet [9], CAggNet [10], MultiResUNet [11], W-Net [12], CSSAMNet [13], U-Net++ [14], etc. are all U-net variants by changing the jumper structure. For example, Chen Songyu [15] proposed an encoder-decoder AFU-net [16] network. A bottom-up,

top-down structure is used based on U-net [17], and dense skip connections are introduced to fuse multi-scale features at different levels. Yu Mingyang [18] et al. proposed an attention-based U-shaped feature pyramid network (AFP-Net) [19], which can focus on building structures of different shapes in high-resolution remote sensing images and achieve efficient extraction of building outlines. In this paper, Li Tao [20] et al. propose a real-time semantic segmentation network combined with a global attention mechanism [21], which significantly improves the segmentation accuracy. The spatial pyramid pooling module (ASPP) is added before the UpSampling operation of the decoder module of the classical U-net network structure, and the average pooling and 1x1 convolution are added to the original spatial pyramid pooling module to increase the semantic information of the global context of the network. One of the improved spatial pyramid structures is shown in Figure 1. Finally, after the experimental validation, it is proved that the model in this paper can achieve better segmentation accuracy with the lowest possible number of parameters.

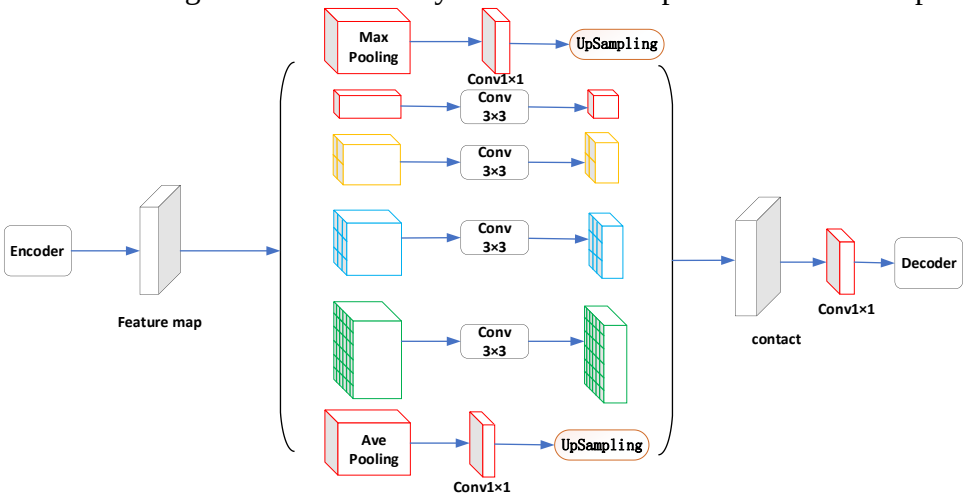


Figure 1: Improved spatial pyramid structure

2. Related Work

CNC machine tools mainly process small batch, complex structure (large) parts, which have complex 2D drawings. Mechanical parts drawing closed contour shapes are many and complex, and industrial manufacturing requires high accuracy of cutting effect when cutting plates. Inspired by the automatic segmentation method of graphic images in medical field, therefore, this paper proposes a U-net based engineering drawing contour segmentation method. Applying the U-net network with fine cutting effect to the study of recognizing the closed contour of mechanical parts diagram can achieve efficient segmentation and extraction of the closed contour of mechanical parts. The improved U-net network in this paper is based on the classical U-net network to improve its jump structure, and on the basis of applying the improved spatial pyramid pooling (ASPP) module [22], we propose to add the upper sampling of the decoder and the inverse The CBAM [23] module (which consists of spatial and channel attention mechanisms) is added between the convolutional layers to give different weights to different parts of different regions in the image, thus finding regions of interest and suppressing regions of disinterest, so that the network focuses on extracting features that are important for recognition results. The attention mechanism uses CBAM (spatial attention combined with channel attention), which gives better segmentation results compared to using only the channel attention mechanism. The spatial attention mechanism increases the weights of specific detection regions, and the channel attention mechanism constrains the semantic information dependencies on the channels to better integrate the semantic information of higher dimensional characteristics with that of lower dimensional features, and to eliminate the semantic

gap to achieve better semantic segmentation results. VGG16 is used as the backbone network to train this model. This method is a novel U-net segmentation network for mechanical part diagrams, and the improved overall network structure is shown in Figure 2.

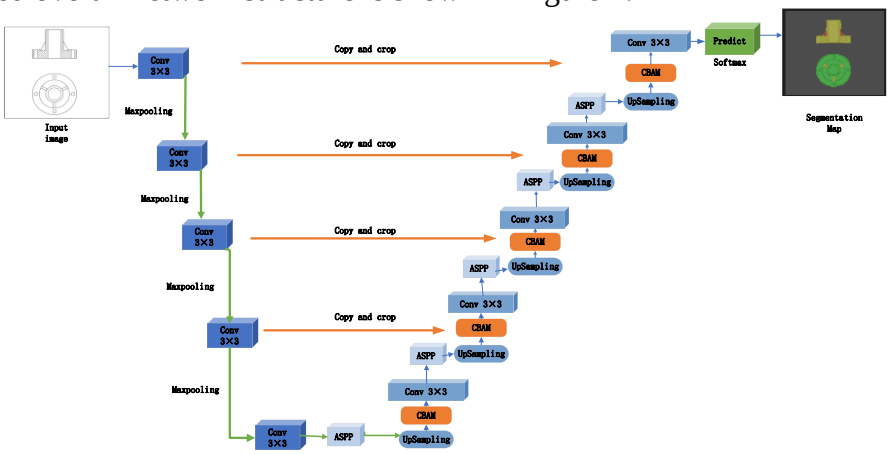


Figure 2: Improved U-net network structure

U-Net semantic segmentation network is named because of its symmetric U-shaped structure. The network belongs to a kind of fully convolutional neural network, which consists of two major parts, encoder and decoder, and fuses deep multidimensional semantic features of image features and shallow low-dimensional semantic features by processing the input image. The left side of the network is the encoder module, which mainly consists of four convolutional blocks, where each convolutional block in turn consists of two convolutional layers of 3×3 convolutional kernels, which are then connected to the convolutional blocks by Max Pooling. After just one convolution block and Max Pooling operation, the extracted feature map can become half of the original feature map, but the number of channels after that operation is doubled. After a total of four identical operations, the size of the original image feature map has been reduced to one-sixteenth of the original size, and the fifth layer still passes through two convolutional layers with 3×3 convolutional kernels and UpSampling (UpSampling) into the decoder module on the right. Similar to the structure of the encoder stage, the decoder stage still consists of four convolutional blocks, each of which is also a convolutional layer of two 3×3 convolutional kernels and connected to the convolutional module through the UpSampling operation. After a total of four convolution blocks, the feature image is then restored to the size of the original input image. The network has a U-shaped structure and is symmetrically distributed on both sides of the central axis, i.e., it has an axisymmetric structure. It is as shown in Figure 2. To overcome this problem and minimize the information missing problem caused by the maximum set, the network fuses the output of each encoder with the input channel of the corresponding decoder, which acts as an information fusion to reduce the information missing problem. Finally, the output image is compressed and plugged into the SoftMax classifier, and the number of channels is changed to the desired number of classes to obtain the probabilities of the classified attributes, and the semantic segmentation task of this image is completed by mapping to the corresponding attributes according to the corresponding probabilities.

The systolic path is mainly responsible for capturing the effective feature information of the image, while the expansive path precisely locates the effective features, i.e., the semantic information is obtained by the DownSampling operation, and the semantic information is located by the UpSampling operation. The structure of U-Net combines deep and shallow feature information by adding cropping and copying operations between the two sides of the path. Therefore, the U-Net network has the superb ability to learn global features and achieve end-to-end training by capturing

as much detailed information as possible with only a small number of images to achieve high-precision semantic segmentation results and high-precision localization requirements.

In this paper, we use a modified U-net network by adding an ASPP module to the front end of each UpSampling of the network decoder and a CBAM module between the UpSampling and the convolution. The ASPP module can this module can obtain convolution kernels of different sizes, which can be used to obtain multi-scale object information and better correlate Contextual global information. The attention mechanism (CBAM) can selectively assign weights to feature information, allowing the network itself to choose to decide to focus on specific feature information. This module enables the network to better capture feature information and allows the network to exhibit better segmentation performance while keeping the number of parameters low. In this paper, the most important core of the improved U-net network composition is the convolutional layer, which is responsible for the basic operation whose role is to extract the image features of the input image. Using the convolution operation, the features of the feature layer or the previous input layer can be generalized and mapped from the original image to the feature space to make the features more abstract, and the output feature layer is made to go to the next convolutional layer operation by the activation function. Finally, the layers obtained by convolution are used for classification or image segmentation.

Convolution layer parameters include boundary fill (pad), perceptual field size (filter) and step size (stride), and the size of the output feature map of the convolution layer is determined by these three parameters together. The boundary fill value indicates the number of fill layers, the fill layers are filled around the feature map, the edge information is extracted more fully, this paper uses the boundary auto-zero fill, so as to obtain more edge information and reduce the impact of insufficient edge extraction by the convolution operation; the size of the perceptual field is smaller than the size of the input image, the larger the perceptual field, the more complex information can be extracted, this network will set the convolution kernel size to 3×3 The step length determines the length of the image elements when the perceptual field scans the adjacent area.

In order to reduce features and parameters, pooling operation is to "simplify" the data by certain rules, thus solving the problem of huge computational effort caused by using image features obtained directly after convolutional operations in each layer of convolutional neural networks. The pooling layer uses a sliding window to obtain new values to replace the original ones, and common pooling methods include average pooling and maximum pooling. Pooling operation can achieve dimensionality reduction, reduce large number of operations, and prevent overfitting to a certain extent, while preserving the most important image features. The image features are invariant, so the generalization of the features during DownSampling of the pooling layer does not lose the original features of the image. According to this principle, the image is reduced and then convolved to reduce the computational effort and expand the perceptual domain, so the convolution layer can extract higher-level image features.

3. Experiment

3.1. Dataset

As we all know, the number of datasets will directly affect the segmentation performance of the network. When the number of samples in the dataset is insufficient, the network may over-fit. The dataset used for training in this paper is a diagram of mechanical parts commonly used in engineering. To avoid the problem of biased training results due to a small dataset, dataset enhancement processing is performed on the graph. The augmentation of the dataset is achieved by cropping, mirroring, deflecting, and adding noise to the dataset, and finally obtains a dataset of 2,700 labeled engineering drawings. Finally, these datasets were modified into training atlases in

VOC2007 format, randomly selected 8:2 according to the number of atlases and used GPU for training and testing. Its segmentation visualization is shown in Figure 3.

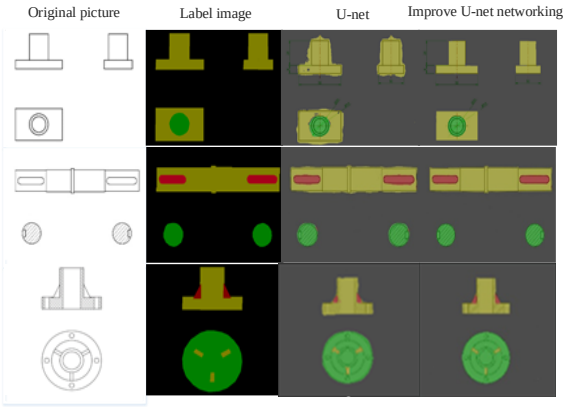
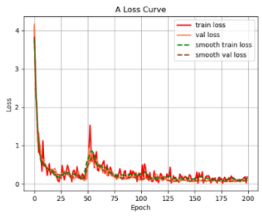


Figure 3: Visualize graph segmentation comparison

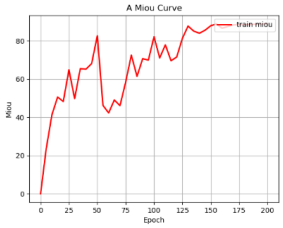
The method in this paper is compared and analysed experimentally with the classic U-net. It can be seen from the visual comparison results that the segmentation method proposed in this paper is better than the classical U-net network in segmenting the outline of the part graph.

3.2. Implementation Details

With VGG16 as the backbone network, the Adam optimizer with internal parameters of 0.9 is used to train the network model. The initial learning rate is set to 1e-4, and the learning rate is decreased by the cos function. On this basis, the combined dynamic scaling cross-entropy loss FL loss [24] is used to reduce the weight of easily distinguishable samples [25], to make the model focus on the weights of those indistinguishable samples during the training process [26]. Average intersection ratio (MIoU) exponential curve, network training loss curve. As shown in Figure 4, it can be seen from the model training curve that the overall network structure converges well, and there is no model explosion [27] and overfitting [28]. It can be seen from Figure 4(a) that the improved U-net model reaches the convergence state at 50 epochs in the process of training and segmenting the mechanical parts diagram. From the graph in Figure 4(b), it can be seen the intersection and union ratio parameter MIoU index in the prediction process is about 85%.



(a) loss graph



(b) Cross-combination ratio graph

Figure 4: Graph of evaluation results

To intuitively get the comparison between the classical U-net network and the network in this paper on the segmentation effect of the same part image, this paper uses the three types of semantic segmentation evaluation indicators mentioned above to compare and analyze the results. The experimental results of semantic segmentation are shown in Table 1.

Table 1: Comparison of experimental results of semantic segmentation

Internet	MIoU/%	mPA	Accu/%
U-net	82.14	80.23	81.87
Improved U-net network structure	89.19	94.86	98.36

From Table 1, it can be seen the intersection-to-merge ratio (MIoU) index of this paper's method for segmenting graphs is about 7.05% higher than that of classical U-net, and the average pixel accuracy and overall accuracy of segmentation are about 14.63 and 16.49% higher than that of classical U-net, respectively. This shows that this method has higher performance and better segmentation accuracy than classical U-net for segmenting mechanical parts graphs.

4. Conclusions

Aiming at the low-precision problem of extracting closed contours of mechanical parts diagrams by classical U-net network segmentation. In this paper, the pyramid structure and attention mechanism modules are added based on the original network. Compared with the segmentation results of the classic U-net, the method of this paper is used to segment the outline of the part graph, and Its MIoU, mPA, and Accu have been significantly improved. Compared with the classical U-net, the method has better performance in mechanical parts graph segmentation.

References

[1] Wang P, Chen P, Yuan Y, et al. (2018) Understanding convolution for semantic segmentation. 2018 IEEE winter conference on applications of computer vision (WACV). IEEE, 1451-1460.

[2] Paszke A, Chaurasia A, Kim S, et al. (2016) Enet: A deep neural network architecture for real-time semantic segmentation. arXiv preprint arXiv: 1606. 02147.

[3] Zhao H, Shi J, Qi X, et al. (2017) Pyramid scene parsing network. Proceedings of the IEEE conference on computer vision and pattern recognition. 2881-2890.

[4] Badrinarayanan V, Kendall A, Cipolla R. (2017) Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence, 39(12): 2481-2495.

[5] Xiao T, Liu Y, Zhou B, et al. (2018) Unified perceptual parsing for scene understanding. Proceedings of the European conference on computer vision (ECCV). 418-434.

[6] Peng C, Zhang X, Yu G, et al. (2017) Large kernel matters--improve semantic segmentation by global convolutional network. Proceedings of the IEEE conference on computer vision and pattern recognition. 4353-4361.

[7] Chen L C, Papandreou G, Kokkinos I, et al. (2014) Semantic image segmentation with deep convolutional nets and fully connected crfs. arXiv preprint arXiv:1412.7062.

[8] Chen L C, Papandreou G, Schroff F, et al. (2017) Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587.

[9] Alom M Z, Hasan M, Yakopcic C, et al. (2018) Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. arXiv preprint arXiv:1802.06955.

[10] Cao X, Lin Y. (2021) Caggnet: Crossing aggregation network for medical image segmentation. 2020 25th International Conference on Pattern Recognition (ICPR). IEEE, 1744-1750.

[11] Ibtehaz N, Rahman M S. (2020) MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. Neural networks, 121: 74-87.

[12] Mohan S, Bhattacharya S, Ghosh S. (2021) Attention W-Net: Improved Skip Connections for better Representations. arXiv preprint arXiv:2110.08811.

[13] Song Z, Yao H, Tian D, et al. (2022) CSSAM: U-net Network for Application and Segmentation of Welding Engineering Drawings. arXiv preprint arXiv:2209.14102.

[14] Zhou Z, Rahman Siddiquee M M, Tajbakhsh N, et al. (2018) Unet++: A nested u-net architecture for medical

image segmentation. Deep learning in medical image analysis and multimodal learning for clinical decision support. Springer, Cham, 3-11.
[15] Song Yuchen, Qiang ZUO, Zhifang WANG. (2022) Semantic Segmentation of Remote Sensing Image Based on U-Net.Radio Engineering, 52(1):168-172.
[16] Chen T, Wang H, Liu H, et al. (2020) An Island Remote Sensing Image Segmentation Algorithm Based on A Fusion Network with Attention Mechanism. Journal of Physics: Conference Series. IOP Publishing, 1693(1): 012179.
[17] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.
[18] Yu M Y, Chen X X, Zhang W Z, et al. (2022) Building extraction on high-resolution remote sensing images using attention gates and feature pyramid structure. Journal of Geo-Information Science, 24(9):1785- 1802.
[19] Wang D, Zhang N, Sun X, et al. (2019). Afp-net: Real-time anchor-free polyp detection in colonoscopy. 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI). IEEE, 636-643.
[20] Li Tao, Gao Zhigang, Guan Shengyuan, et al. (2022) Global attention mechanism with real time semantic segmentation network, DOI:10.11992/tis.202208027.
[21] Liu Y, Shao Z, Hoffmann N. (2021) Global Attention Mechanism: Retain Information to Enhance Channel-Spatial Interactions. arXiv preprint arXiv:2112.05561.
[22] Chen L C, Papandreou G, Kokkinos I, et al. (2016) DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Transactions on Pattern Analysis & Machine Intelligence, 40(4):834-848.
[23] Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). CBAM: Convolutional Block Attention Module. Lecture Notes in Computer Science, 3–19. doi:10.1007/978-3-030-01234-2_1.
[24] Kingma D P, Ba J. (2014) Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
[25] Lin T Y, Goyal P, Girshick R, et al. (2017) Focal loss for dense object detection. Proceedings of the IEEE international conference on computer vision. 2980-2988.
[26] Li X, Sun X, Meng Y, et al. (2019) Dice loss for data-imbalanced NLP tasks. arXiv preprint arXiv:1911.02855.
[27] Chemical Engineering Progress group. (2014). Consequence Analysis Software Models Explosion Risk. Chemical Engineering Progress (10).
[28] Simonyan K, Zisserman A. (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.