Based on Neural Networks: a Semantic Segmentation Algorithm for Optimization of Distributed Storage of Energy Big Data

Dong Mao, Zhongxu Li, Zuge Chen, Hanyu Rao, Jiuding Zhang and Zehan Liu

June 19, 2022

Based on Neural Networks: A Semantic Segmentation Algorithm for Optimization of Distributed Storage of Energy Big Data

Dong Mao
New Technology Research Center
State Grid Zhejiang Electric
Power Corporation Information
& Telecommunication Branch
Hangzhou, China
641743819@qq.com

Zhongxu Li
Internet Department
State Grid Zhejiang Electric Power Company
Hangzhou, China
147816801@qq.com

Zuge Chen
New Technology Research Center
State Grid Zhejiang Electric
Power Corporation Information
& Telecommunication Branch
Hangzhou, China
chenzuge@163.com

Hanyu Rao
New Technology Research Center
State Grid Zhejiang Electric
Power Corporation Information
& Telecommunication Branch
Hangzhou, China
458994964@qq.com

Jiuding Zhang
New Technology Research Center
State Grid Zhejiang Electric
Power Corporation Information
& Telecommunication Branch
Hangzhou, China
jiuding_zhang@163.com

*Zehan Liu
the Research center of Big Data
and Computational Intelligence
Hangzhou Innovation Institute
of Beihang University
Hangzhou, China
zf2121227@buaa.edu.cn

***Abstract*—There are many kinds of energy data, how to realize unified storage, processing and sharing of energy data is a big problem. As the national energy data center, State Grid aims to build a database that can store distributed heterogeneous asynchronous energy data. The storage of image files in the big energy database will take up a lot of space in the system, but not all parts of the image are needed. Therefore, it is very necessary to accurately segment the effective area of the image to store it so as to achieve the purpose of data compression. This paper proposes the Attention U-Net framework, which combines the traditional semantic segmentation network U-Net with the Attention moudle to focus on the region of interest in the image, emphasize foreground information, and suppress background information. The results show that compared with U-Net, the accuracy is improved by 1.77% and after the segmentation is completed, each image saves an average of 2MB of storage space.**

***Keywords*—Energy data, Semantic Segmentation Network, U-Net, Image segmentation, compression**

I. INTRODUCTION

State Grid Corporation of China, as the national energy big data center, needs to collect the data of all energy users, which requires its service system not only to have high-performance data flow calculation ability[1], resource management and deployment ability to support flow processing, but also to integrate multiple data sources and analyze them efficiently. At present, some researches have put forward the distributed heterogeneous asynchronous data storage of energy big data[2], which can solve the problems of efficient storage and high reliability of energy big data, thus forming a catalogue system of energy big data resources, and finally realizing data aggregation, overall management and organic integration. The random reading operation is greatly reduced, the insertion and query performance of a single data acquisition point is guaranteed to be the best, the system processing capacity is improved, the system storage cost is reduced, and the efficient storage capacity of heterogeneous asynchronous distributed data processing is realized. Through the long-term data sharing and exchange mechanism, we can ensure the data sharing and exchange between units, and provide data services for all units.

For users, heterogeneous asynchronous storage is not a specific storage device, but a collection composed of many storage devices and servers. When users use heterogeneous asynchronous storage, they do not use a specific storage device, but use a data access service provided by the entire heterogeneous asynchronous storage system. So strictly speaking, heterogeneous asynchronous storage is not storage, but storage service. The core of heterogeneous asynchronous storage is the combination of application software and storage device, and the transformation from storage device to storage service is realized through application software.

Traditional cloud-side computing resources have insufficient processing performance and limited storage capacity, which will bring a series of problems such as low data writing efficiency and high data storage cost[3]. In energy big data, image files account for a considerable part of unstructured data[4]. Tens of thousands of images will take up a lot of

The authors would like to thank: State Grid Zhejiang Electric Power Co., Ltd. Technology Project(No.B311XT220006)

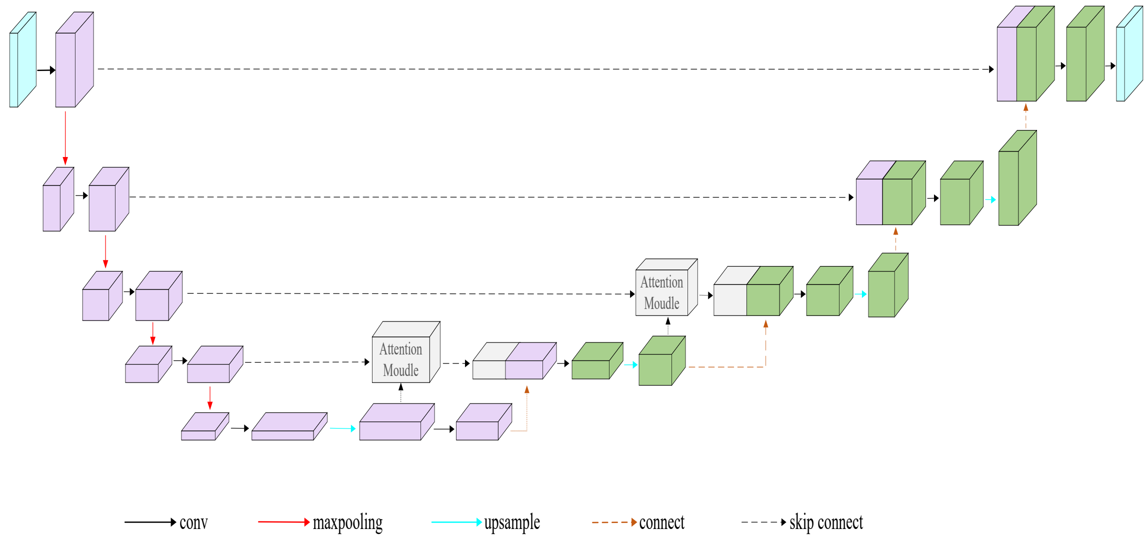

Fig. 1. The overall architecture of our Attention U-Net model, the biggest innovation is the addition of Attention moudle at the shallow skip connection.In this way, more attention is paid to the information of the region of interest and the information of the background region is suppressed.

storage resources of the system, but some parts of a image are not needed by us. Therefore, we want to propose an image segmentation algorithm[5], which is used to accurately segment the required part of each image for re-storage, so as to compress the image file[6] and improve the accuracy of machine identification of energy meter readings. This can greatly reduce the random reading operation ensure the best performance of inserting and querying a single image to the greatest extent, improve the processing capacity of the system, reduce the storage cost of the system, and then realize the efficient storage capacity of heterogeneous asynchronous distributed data.

In this paper, an automatic image segmentation model based on Attention U-Net is proposed, which can accurately segment the required area in the picture, remove the unnecessary part of the image, reduce the memory occupied by the image file, so as to realize the optimization of distributed storage of energy big data. Compared with the traditional segmentation accuracy, dice index is improved by 1.77% and the average memory is saved by about 2MB per image.

II. Related Work

Image segmentation is an important research direction in computer vision. At present, image segmentation has become a hot spot in the field of computer vision, and many researchers have devoted themselves to related fields of image segmentation. The so-called image segmentation refers to dividing the image into several disjoint areas according to some characteristics of the image, such as the color, spatial texture, geometric shape, etc. of the image, so that the same area shows a certain similarity, while different areas show obvious different. Simply put, in an image, the target is separated from the background. From the perspective of algorithm evolution, the current image segmentation algorithms can be roughly divided into two categories: traditional image segmentation algorithms and deep learning image segmentation algorithms. Although most traditional image segmentation algorithms are inferior to deep learning-based methods in segmentation speed and segmentation accuracy, traditional image segmentation algorithms provide a sound theoretical basis, mainly for low-level image information research. The image segmentation algorithm based on deep learning, on the basis of traditional algorithm, fully extracts the high-level features of the image, so that the image segmentation can be applied to more complex scenes.

A. Traditional Methods

Traditional image segmentation methods include the maximum divergence threshold difference method, which performs adaptive threshold segmentation on thermal imaging images of pictures. The hot spot detection method based on Canny edge detection[7] analyzes the distribution of temperature histogram through pre-extracted images to detect hot spots; Through the results of k-means clustering algorithm[8], the straight line of the boundary of the unnecessary area is eliminated, thus achieving the segmentation effect. Limited by the principle of the algorithm, threshold difference method and edge detection method need to filter the background when detecting hot spots, which affects the real-time performance of detection. However, the K value of k-means clustering is difficult to determine and locally optimal.

B. Based on Deep Learning

Image segmentation algorithms based on deep learning are mainly divided into two categories, one is for rough image segmentation, which is called semantic segmentation in academic circles; the other is for high-quality image segmentation, which is called image matting in academic circles. In recent years, image segmentation methods based on deep learning have gradually improved the segmentation effect of traditional algorithms, which mainly depends on the

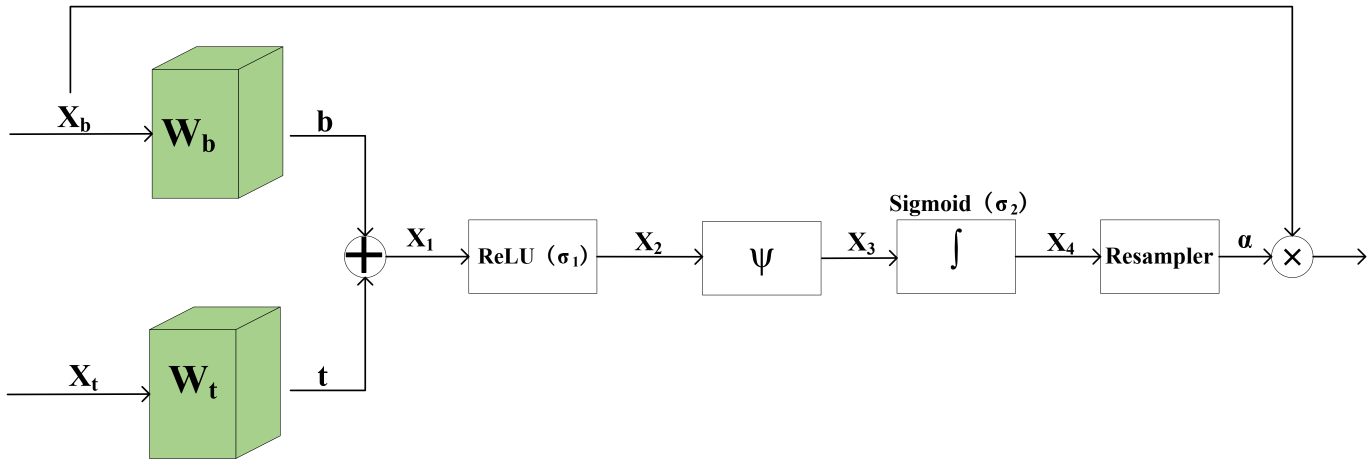

Fig. 2. The main architecture of Attention moudel. X_b and X_t are the two input feature maps of the Attention module, respectively. After operations, the regions of interest of the two feature maps are repeatedly superimposed, so as to deepen the focus on the region of interest.

improvement of computer performance and the emergence of a large number of labeled datasets, such as Pascal VOC[9], MS COCO[10], Ade20K[11] and other datasets containing pixel-wise semantic labels enable deep learning-based image segmentation models to train and converge to a good effect. The emergence of technologies such as the Encoder-decoder basic model, fully convolutional network (FCN)[12], and hole spatial feature pyramid[13] enables neural networks to output higher-resolution image segmentation results. However, these methods have mainly obtained good semantic segmentation results on public datasets, which belong to the field of rough image segmentation, and the segmentation effect of detailed hair and other parts of objects is still too blunt. The technology to solve this problem is currently mainly a high-quality image segmentation and matting algorithm. The matting technology is used to smooth the segmentation effect of the edge of the object, so that the visual effect is better and the segmentation accuracy is higher. However, most of the current matting technologies are interactive and require human input of auxiliary information, such as inputting Trimap and inputting strokes. Trimap is a three-color map that roughly divides the foreground, background, and position areas of the image, while the stroke only needs to define some known areas. In this way, the algorithm's matting efficiency is low, and the effect of real-time video matting cannot be achieved. Therefore, many people are now devoted to the research of fully automatic high-quality image segmentation technology. Although SHM[14], LFM[15] and other algorithms are dedicated to fully automatic matting, due to the complexity of the matting task and the limitations of the data set, these fully automatic matting algorithms are widely used in natural images. The chemistry is not very good.

III. Methods

A. The architecture of Attention U-Net

Fig. 1. shows our proposed Attention U-Net architecture, which is based on the classic U-Net[16] structure.In order to better pay attention to the region of interest in the image, we propose an Attention Moudle[17], which is combined with U-Net to enhance the information of the foreground region of interest and suppress the irrelevant information of the background region.

The network consists of an encoder-decoder structure that considers both global and local information, combining low-resolution global and high-resolution local features via skip connections.In the encoder part, the shallow layer can capture some simple information of the image, such as border and color;the deep layer will capture some abstract features, that is, high-level features, and encourage the generation of more semantic information output because of the increased feeling and the more convolution operations.

B. Attention Moudle

In the deepest structure of the encoder, it often contains the richest semantic information, but because the cascaded convolution[18] is nonlinear, a lot of spatial details will be lost in the output map.This does not work well for us when the region of interest is too small in the whole image.

To solve this problem, we designed an Attention Moudle as shown in Fig. 2. It is used at the junctions of the underlying feature maps to identify relevant spatial information.

It can be seen from Fig. 2. that firstly, X_b and X_t pass through different matrices W_b and W_t, to obtain t and b of the same size, respectively, and then add t to b to obtain X_1. Then, X_1 is subjected to Relu operation to obtain X_2, and X_2 is subjected to ψ operation to obtain X_3. X_3 performs the sigmoid operation to obtain X_4, and X_4 obtains the attention coefficient α through resampling (the attention coefficient is actually the attention weight). Finally, the attention coefficient map is multiplied by the original image X_b, and the output of the entire Attention Moudle is obtained. In this way, the region of interest is further superimposed and the effect of background-independent region information is suppressed.

TABLE I
QUANTITATIVE RESULTS OF ABLATION STUDY ON LOSS

Model	***Dice%***	***Jaccard%***	***Precision%***	***Recall%***	***Acc%***
Attention U-Net + DiceLoss	*81.75±0.15*	*71.93±0.14*	*87.53±0.13*	*81.08±0.83*	*97.55±0.01*
Attention U-Net + CELoss	81.91±0.17	72.37±0.28	88.23±0.11	81.73±0.44	97.63±0.04
Attention U-Net + 0.36*(DiceLoss+CELoss)	82.98±0.05	74.37±0.11	89.09±0.09	82.73±0.23	97.71±0.02

TABLE II
QUANTITATIVE RESULTS OF ABLATION STUDY ON ATTENTION BLOCK POSITION

Model	***Dice%***	***Jaccard%***	***Precision%***	***Recall%***	***Acc%***
1	81.13±0.12	73.14±0.04	88.15±0.16	82.21±0.64	97.58±0.11
2	81.56±0.09	72.43±0.15	87.66±0.13	82.15±0.88	97.46±0.06
1+2	82.98±0.05	74.37±0.11	89.09±0.09	82.73±0.23	97.71±0.02

The formula can be described as follows:

$$q_{Att}^{t} = \psi^{T}(\sigma_1(W_t^T X_{ti} + W_b^T X_{bi}) + h_{X_b}) + h_{\psi} \quad (1)$$

$$\alpha_i^t = \sigma_2(q_{Att}^t(X_{ti}, X_{bi})) \quad (2)$$

where $\sigma_2(X_{ti,c}) = \frac{1}{1+exp(-X_{ti,c})}$correspond to sigmoid activation function. W_b and W_t are linear transformations,which are computed using channel-wise 1x1x1 convolutions for the input feature map tensors. h_{X_b} and h_{ψ} are bias terms.

C. Loss Function

In this paper, our final Loss function[19-20] as follows:

$$Loss = 0.36 * (L_{ce} + L_{dice}) \quad (3)$$

where L_{ce} is Cross-entropy loss and L_{dice} is Dice loss. The formula for Cross-entropy loss as follows:

$$loss(x, class) = -x[class] + log(\sum_j exp(x[j])) \quad (4)$$

where 'class' is the category to be divided into semantic segmentation and 'x' is the input. The formula for Dice loss as follows:

$$DiceLoss = 1 - \frac{2|X \bigcap Y|}{|X| + |Y|} \quad (5)$$

if the 'Dice' coefficient is larger, it indicates that the sets are more similar and the 'Loss' is smaller. vice versa. X represents the set of pixels in the foreground area of the segmented output image,and Y represents the set of pixels in the foreground area of the label.

IV. EXPERIMENTS

According to our investigation, Hangzhou's natural gas data is managed by the Hangzhou Housing and Urban-rural Development Bureau. Although the data generated by each gas meter will be connected to the information collector and uploaded to the cloud, the data transmitted in this way is not 100% normal, and network fluctuations may cause data delays or errors. If the information collector fails It will also lead to data errors, so on average every week or every month, staff will come to the site to read the meters and take photos for backup, which will result in a lot of natural gas meter photos being stored in the database. If you want to use a machine to identify the readings in the photo, the background of each picture will also interfere with the numbers, so we designed an algorithm that can accurately segment the gas meter in the picture to improve the machine's recognition. precision and reduce the storage space occupied by image files.

In this section, we evaluate the performance of the Attention U-Net using our collected dataset. This section is arranged as follows: first, we introduce our collected dataset, experimental settings and evaluation metrics. Second, perform ablation experiments. Last, we compare our method with U-Net on our collected dataset.

A. Dataset

In this experiment, we collected a total of 4,000 natural gas meter pictures and their corresponding natural gas meter labels from universities in Zhejiang province. Among them, the natural gas meter comes from many different brands such as HL, GlodCade, TANCY, etc. In order to make the training model more robust, we divided the entire dataset into training set, validation set and test set, of which the training set is 3000, the validation set is 500, and the test set is 500.

B. Implementation Details and Evaluation Metrics

In this experiment, the input size of all network model is 256×256. All models are trained for 100 epochs when the batchsize is 8. In terms of optimizer, we use Adam as the optimizer for optimizing. The learning rate is 0.001, and the learning rate decay strategy is lr_scheduler.StepLR(), which

TABLE III
QUANTITIVE RESULTS OF DIFFERENT METTHODS ON OUR COLLECTED DATASET (MEAN ± SD)

Model	***Dice%***	***Jaccard%***	***Precision%***	***Recall%***	***Acc%***
U-Net	*81.21±0.14*	*71.75±0.13*	*87.47±0.11*	*81.05±0.92*	*97.54±0.03*
Attention U-Net	82.98±0.05	74.37±0.11	89.09±0.09	82.73±0.23	97.71±0.02

becomes one-tenth of the previous one every 25 rounds. The weight decay is set to 10-9. The deep learning framework uses Pytorch for training on an NVIDIA V100 GPU. To evaluate the performance of the model, we employ several evaluation metrics:

We utilize five commonly used indicators to quantitatively compare different segmentation methods of natural gas meter. They are Dice coefficient (denoted as Dice), Jaccard index (Jaccard similarity coefficient), Recall, Precision and Accuracy.

Accuracy is the most basic metric in image segmentation. The formula of Accuracy is as follows.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{6}$$

where TP, TN, FP, and FN are true positive, true nega - tive, false positive, and false negative respectively.

The formulas for Precision and recall are shown below respectively.

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

Dice is a commonly used indicator in semantic segmentation, and it as follows.

$$Dice = 1 - \frac{2|X \bigcap Y|}{|X| + |Y|} \tag{9}$$

Where X is the prediction pixel and Y is the ground truth.

C. Ablation Study

Ablation study on Loss. In order to choose a suitable LOSS function, we compare the ways of combining Attention U-Net with three LOSS functions. DiceLoss function, CELoss function and 0.36 times DiceLoss+CELoss function, respectively, for a fair comparison, experiments use their original published code and the same settings as ours. The experimental results are shown in Table I. The performance of our model has been improved to some extent. In terms of Dice, it is about 1.23% higher than DiceLoss. Compared with CeLoss, it is improved by about 1.07%. These results show that our proposed 0.36 times DiceLoss+CELoss can be combined with Attention U-Net to improve segmentation performance.

Ablation study on attention block position. From the segmentation network encoder-decoder structure, the deeper the network, the easier it is to extract high-level features. The focus of the Attention Moudle designed in this article is to deepen the foreground information and ignore the background information. Therefore, this article only considers adding the modules we designed in the deep layers, that is, the 1^{st} and 2^{nd} layers are the most suitable. It can be seen from Table II that it is better for us to add two moudles than to add a moudle to layers 1 and 2 respectively. And the effect of adding in the 1^{st} layer is better than that in the 2^{nd} layer, which shows that our moudle added in the deep layer is better than the shallow layer. So the more our moudle adds in the deeper layers the better, the deeper the better. Since the layers in the table are not displayed visually, they are replaced by numbers. The numbers in the table represent the number of layers where the model is located.

D. Comparison with the State-of-the-arts

In order to demonstrate the overall segmentation performance of Attention U-Net, we compare traditional methods (U-Net). For a fair comparison, the experiment use their original release code and the same setup as ours. The experimental results are shown in Table III. The performance of our model has been improved to a certain extent. In terms of Dice, it has been improved by 1.77% compared with U-Net. These results indicate that the Attention module proposed by us can be added to the original U-Net model to improve the segmentation performance.

Besides the quantitative results, the ground truth andsegmentation results of different methods are shown in Fig. 3. The red boxes highlight areas where Attentionl U-Net performs better than other methods. The results show that our method can generate more accurate segmentaion results, which are closer to the ground truth than the comparison models.

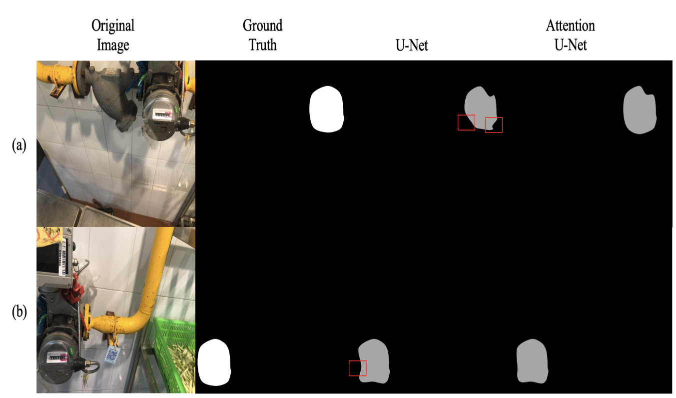

Fig. 3. Segmentation results of two samples (a-b) given by different methods. The red boxes highlight areas where Attention U-Net performs better than other methods.

V. Conclusions

In this paper, we propose a novel Attention U-Net algorithm to achieve accurate segmentation of natural gas meters from a image.Attention U-Net achieves the highest average Dice (83.98%) in this experiment, which is 1.58%-1.96% higher than the comparison methods, and saves 2MB of memory per image on average.

In this information age, more and more data needs to be stored for our use. Our algorithm hopes to reduce the memory occupied by image storage and increase the accuracy of machine recognition of images, thus making a contribution to the construction of a large energy database for the State Grid. However, due to the increase in the number of parameters of our algorithm, the training time of the model increases, which is a major direction that needs to be improved in the future.

Acknowledgment

This work was supported by State Grid Zhejiang Electric Power Co., Ltd. Technology Project(No.B311XT220006).

References

[1] Sharp J A. Data flow computing[J]. 1985.

[2] Chen, Zheyi, et al. "Towards asynchronous federated learning for heterogeneous edge-powered internet of things." Digital Communications and Networks 7.3 (2021): 317-326.

[3] Wu J, Ping L, Ge X, et al. Cloud storage as the infrastructure of cloud computing[C]//2010 International conference on intelligent computing and cognitive informatics. IEEE, 2010: 380-383.

[4] Boulton D, Hammersley M. Analysis of unstructured data[J]. Data collection and analysis, 2006, 2: 243-259.

[5] Haralick R M, Shapiro L G. Image segmentation techniques[J]. Computer vision, graphics, and image processing, 1985, 29(1): 100-132.

[6] Rabbani, Majid, and Paul W. Jones. Digital image compression techniques. Vol. 7. SPIE press, 1991.

[7] Rong W, Li Z, Zhang W, et al. An improved CANNY edge detection algorithm[C]//2014 IEEE international conference on mechatronics and automation. IEEE, 2014: 577-582.

[8] Hartigan J A, Wong M A. Algorithm AS 136: A k-means clustering algorithm[J]. Journal of the royal statistical society. series c (applied statistics), 1979, 28(1): 100-108.

[9] Everingham, Mark, et al. "The pascal visual object classes (voc) challenge." International journal of computer vision 88.2 (2010): 303-338.

[10] Lin, Tsung-Yi, et al. "Microsoft coco: Common objects in context." European conference on computer vision. Springer, Cham, 2014.

[11] Zhou, Bolei, et al. "Scene parsing through ade20k dataset." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[12] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

[13] Chen, Liang-Chieh, et al. "Rethinking atrous convolution for semantic image segmentation." arXiv preprint arXiv:1706.05587 (2017).

[14] Chen, Quan, et al. "Semantic human matting." Proceedings of the 26th ACM international conference on Multimedia. 2018.

[15] Zhang, Yunke, et al. "A late fusion cnn for digital matting." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.

[16] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015.

[17] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.

[18] Zhang, Kaipeng, et al. "Joint face detection and alignment using multitask cascaded convolutional networks." IEEE signal processing letters 23.10 (2016): 1499-1503.

[19] Schorfheide, Frank. "Loss function-based evaluation of DSGE models." Journal of Applied Econometrics 15.6 (2000): 645-670.

[20] Ko, Young-Hyun, Kwang-Jae Kim, and Chi-Hyuck Jun. "A new loss function-based method for multiresponse optimization." Journal of Quality Technology 37.1 (2005): 50-59.