International Journal of Innovative Computing, Information and Control
Volume **19**, Number **1**, February **2023**

ICIC International ©2023 ISSN 1349-4198
pp. 87–100

RESEARCH ON CURVE IMAGE DATA RECONSTRUCTION METHOD BASED ON MULTI-TASK JOINT LEARNING

Shuanzhu Sun*, Jiewen Zhu, Chunlei Zhou, Guoqiang Xu and Tian Shi

China Jiangsu Frontier Electric Technology Co., Ltd.
No. 58, Suyuan Avenue, Nanjing 211102, P. R. China
{ 18795886863; 13851845492; 15905166681; 19895870424 }@163.com
*Corresponding author: 15905166613@139.com

Received June 2022; revised October 2022

Abstract. *Data visualization facilitates the overall analysis and in-depth mining of data. However, visualized data in the form of images are not conducive to machine understanding, which means it cannot meet the requirements of subsequent automated intelligent systems for practical engineering projects. The traditional method of curve image data reconstruction relying on manual features obtains the reconstructed data by point selection method or image pre-processing combined with manual input of coordinate axis information. Its accuracy and efficiency are low, and it does not have the characteristics of automation and high-volume processing. To solve this problem, we propose an end-to-end convolutional neural network model based on a deep learning approach. The model adopts a multi-task co-encoding-independent decoding structure, and combines an attention mechanism to effectively improve the information quality during the jump connection process to achieve multi-task joint learning for curve segmentation and text detection. We validate the effectiveness of the model on a curve image dataset, where the accuracy of curve segmentation and text detection achieved 94.0% and 99.5%, respectively.*
Keywords: Curve image data reconstruction, Convolutional neural network, Multi-task joint learning, Curve segmentation, Text detection

1. **Introduction.** With the rapid development and popularity of the Internet, images have become one of the most abundant mediums for carrying information. Image is the basis of human vision and objective response to natural things. It is an important source for humans to understand the world and themselves. In practical production life, in order to facilitate the analysis and comparison of a large amount of data, visualization of data in the form of images has become a common practice nowadays. Data visualization makes communication more clearly and reasonably. It can assist people to dissect data more strongly. The essence of data visualization is visual effect session. Data in the form of images can not only present people with a richer information world from the perspective of visual perception, but also make data more intuitive, easy to analyze and grasp. However, it is common that only the image is retained but the original data is lost or not traceable. The data in the form of image is not conducive to machine understanding and cannot meet the requirements of subsequent automation systems in actual engineering projects. Therefore, reconstruction of curve image data is an objective and inevitable realistic demand.

Currently, there are two methods for reconstruction of curve image data: one is software tools and the other is traditional image processing. Software tools, such as getData, engage digitizer and originpro, project raw curve data based on the type of manually selected axis and the range of values set for the axis, as well as the selected key nodes. The advantages

DOI: 10.24507/ijicic.19.01.87

of this method are that the selection of points is subjective and easy to operate. It can handle graphs in various coordinate systems. The disadvantage is that these software tools can only calculate other point data based on the selected points, resulting in curve data relying on manually selected points for reconstruction. This method is not effective enough for high volume production requirements.

Traditional image processing methods, such as extracting curves from digital images by image grayscale transform, binary transform, and morphological operations, combined with manual definition of the range of coordinate axes, achieve data reconstruction of curved images. In the work of [1], after filterings denoising and binarizing the image, they find the proportional relationship between the true coordinates and the coordinates of the grayscale matrix. They multiply this relationship by a scale factor to obtain the true coordinate value of each point on the curve. This method has a small error margin, but the actual accuracy depends on the number of pixels in the original image. Tan et al. extracted the curve data by traversing and scanning the curves in the image according to the hand-designed features. Based on this method, they combined the segmented scanning method to complete the extraction of multiple curves [2]. Xu obtained the edge-processed image of the transmission line by binarization and projection, and tracked it in real time. The edge-processed images were then iteratively calculated and modified to obtain the reconstructed curve data [3]. Although the traditional image processing method avoids the problem of large errors in the process of manual selection of coordinate points, the accuracy of the reconstructed curve image depends on the ability of engineers to judge the image features and debug errors. In addition, the traditional method cannot realize the automatic processing of the range of coordinate axes taking values. It means that the traditional image processing method to reconstruct curve image data essentially does not solve the problem of relying on manual processing. There are still significant limitations in the accuracy of curve image data reconstruction.

To solve the problems above, this paper proposes to construct a convolutional neural network with multi-task co-encoding-independent decoding based on deep learning methods. And the attention mechanism is used to effectively suppress the transmission of noisy information during the hopping connection process to realize the end-to-end reconstruction of curved image data. Different from traditional image processing methods, the deep learning methods learn image features through deep convolutional neural networks. Deep convolutional neural networks can acquire advanced semantic features in images and improve the flexibility and adaptability of the image processing process to obtain the accuracy closer to the original real data. Deep convolutional neural networks can realize end-to-end processing of images and simultaneously learn the features of curves and coordinate axis values in the input curve images. It fundamentally solves the problem of relying on manual features, and provides the basis for being able to realize automated and intelligent processing in the real sense in practical engineering.

The main contributions of this paper include the following three points.

1) We propose to construct a convolutional neural network model to realize end-to-end curve image data reconstruction. Combined with a multi-task joint learning approach, we realize the joint training of curve segmentation task and text detection task models.

2) The model is based on a U-shaped structure design and uses an attention mechanism to suppress noisy information in low-order features and improve the quality of the transmitted features.

3) To the best of our knowledge, this paper is the first approach to using semantic segmentation combined with text detection techniques to solve the curve image data reconstruction task with higher performance and accuracy than traditional methods.

The remaining contents of the paper are as follows: Section 2 presents the related work of the research content; Section 3 introduces the structure of the proposed model, attention mechanism and loss function in detail; Section 4 demonstrates the specific setup of the experiment, the validation procedure and the results; Section 5 concludes the content of this paper, including the analysis of existing shortcomings and work prospects for future.

2. Related Work.

Reconstruction of curve image data usually consists of two parts: the extraction of the curve and the extraction of the coordinate axis values. Since different techniques are used for these two parts, we will introduce the existing techniques in two parts.

2.1. Semantic segmentation.

The semantic segmentation method is to cluster the image parts belonging to the same object together to achieve pixel-level classification. The goal of this approach is to provide pixel-level image understanding in the way that humans perceive it. At present, semantic segmentation has been widely used in various fields, such as image understanding for autonomous driving [4], medical image segmentation [5], and surface image paddy field segmentation [6]. This paper mainly wants to use the semantic segmentation method to extract the curve in the curve image, which is used to solve the problem of curve image data reconstruction. In the image understanding for autonomous driving problem, the latest research results show that effective modules help to improve the performance of the model. Sagar and Soundrapandiyan proposed to integrate the sum of multi-scale feature accuracy at different scales, encode more context information with the attention module, and enhance the receiving field of the network, so as to carry out effective pixel-level segmentation of road scene images [7]. Fan et al. proposed a new network structure called short-term density concatenate (STDC) network. It uses the detail aggregation module to integrate the learning of spatial information into the bottom layer in a single-stream way, and combines the bottom features and deep features to predict the final segmentation result [8]. An et al. believe that lightweight models are the key to efficient semantic segmentation. To solve the problem of limited convolution and small receiving field, they proposed two complementary schemes, self-attention distillation and hierarchical context distillation, to supplement the context information of small networks [9].

In the medical image segmentation problem, segmented objects usually include specific organ parts, tumors, and retinal vessels. These objects generally present small features which have higher requirements for the structure of the segmentation network. Olaf et al. proposed a U-shaped network structure, which consists of a contraction path for context capture and a symmetric expansion path for precise localization. U-net is a classical network structure for medical image segmentation, which can realize network training in the case of a small number of images and obtain great results [10]. After further research and analysis, the curve image features in this paper have a high degree of commonality with the retinal blood vessel image features, mainly manifested in the small and relatively long target object. The latest research on retinal vessel image segmentation methods shows that the effect of U-shaped network structure for retinal vessel segmentation task is still worthy of attention. Boudegga et al. proposed a new method for retinal vessel segmentation based on deep learning. The network is a U-shaped structure and uses lightweight convolutional blocks to maintain higher segmentation performance while reducing computational complexity [11]. Based on the U-shaped network structure, Sathananthavathi and Indumathi proposed a retinal vessel segmentation method using encoder to enhance the hole structure. The depth stitching process was improved by adding layers to the encoder part to improve the retinal blood vessel segmentation accuracy [12]. Zhang et al. proposed a context-sensitive U-net retinal vessel segmentation network based on block loss weight mapping, named BridgeNet. The network combines recurrent neural network (RNN) with convolutional neural network (CNN) to convey context, and corrects the unbalanced distribution of thick and thin vessels based on the weight loss mapping of plaques [13]. So, the network structure used for image curve extraction in this paper is built based on U-net.

2.2. **Text detection.** Text detection is a subproblem of object detection direction, which aims to find a reliable method as the front end of text recognition technology, and its goal is to locate the position of text in an image. The process of obtaining text information in an image through text detection and recognition is defined as optical character recognition (OCR). Text detection is the first and essential step in OCR. This paper mainly wants to adopt text detection method to extract the value of the coordinate axis in curve images, which is used to solve the problem of data reconstruction of curve images. In the work of [14,15], traditional text detection methods crop characters out through image preprocessing, layout analysis and other methods to form single characters for subsequent text recognition. The disadvantages of traditional methods mainly include two aspects: on the one hand, the error rate of manual character cropping is high; on the other hand, the text is easy to deform in the case of serious interference in the background, which leads to the difficulty of manual discrimination. These shortcomings cause the accuracy of traditional text detection methods to be greatly limited.

With the emergence of deep learning convolutional neural networks, text detection methods break through the traditional practice. This method can not only significantly improve the detection accuracy, but also adapt to complex background, such as variant text, and rotated text. Text detection methods based on deep learning can be divided into segment-based methods, candidate box-based methods and hybrid methods. The segment-based method is to perform pixel-level semantic segmentation of text objects in images, and then construct text lines according to the segmentation results. The candidate box-based method uses multiple default boxes to generate a large number of candidate text boxes, and then obtains the final detection result by non-maximum suppression (NMS). Among the existing regression-based methods [16,17], most of them use regression on masks or contour points of text regions to model text instances. This method performs well for the interference of complex background. However, regression of complete masks requires high training complexity and performs poorly for detection of highly curved texts. The method based on candidate box is weak in the interference of complex background, but the training process is simple, and the detection effect of curved text is better. Text detection methods based on hybrid segmentation and candidate boxes are more widely adopted in practical applications due to their complementarity. The classic network architecture based on mixing two detection methods is EAST, which employs a simple yet powerful pipeline. The pipeline directly predicts words or text lines with any orientation and quadrilonoid shape in the whole image through a single neural network, eliminating unnecessary intermediate steps such as candidate aggregation and word partitioning [18]. In the latest research on text detection methods [19-21], the design of network structure is also aimed at realizing end-to-end text detection with low training complexity. Therefore, the proposed image text detection network structure selection is constructed based on the classical model EAST.

3. **Proposed Method.** In this section, we detail the proposed end-to-end multi-task jointly trained convolutional neural network with five key components: network structure, semantic segmentation, text detection, attention mechanism, and loss function.

3.1. **Network structure.** The model proposed in this paper consists of three parts: encoder, text detection decoder and semantic segmentation decoder. The overall model structure is shown in Figure 1. The encoder uses a partial VGG16 network structure in addition to the fully connected layer [22]. It is divided into five stages according to the criteria of pooling layer, which are Conv1-2, Conv2-2, Conv3-3, Conv4-3, and Conv5-3 in order. The model extracts the input curve image features by the encoder, and then decodes them with the text detection decoder and the semantic segmentation decoder, respectively. In this model, the extraction of curve and coordinate axis values is done simultaneously.

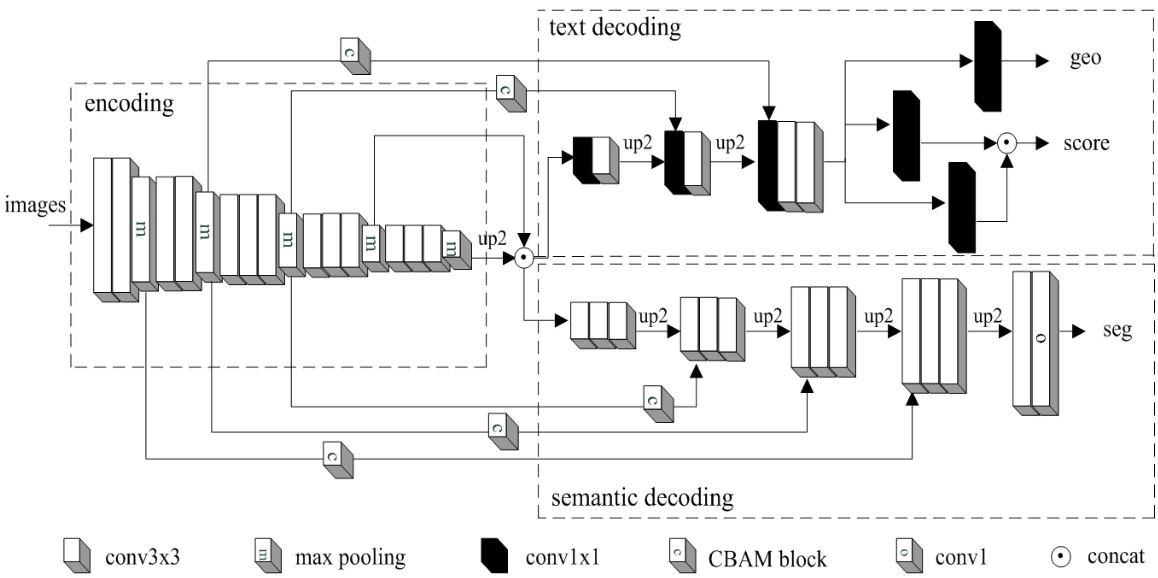


FIGURE 1. The network structure of the proposed method

3.2. **Semantic segmentation.** As a key step in curve image data reconstruction, the effect of curve segmentation directly affects the final data reconstruction result. The design idea of curve segmentation decoder in this paper is consistent with U-net. The features in the encoding stage are added to the upsampling process in the decoding stage, and the cross-layer concatenation is used to improve the utilization of features. This cross-layer connection and feature splicing fusion method not only enables the network to learn more feature information, but also makes use of the shallow texture features to supplement the high-level semantic features and avoid the loss of small features, especially edge features, during the learning process. In addition, the structure is simple and stable. Its advantages are more obvious in the case of small target objects and small data volume. The upsampling process of the curve segmentation decoder part proposed in this paper restores the resolution of the feature map to the same resolution as the input.

3.3. **Text detection.** The detection of values on the curve axes is another necessary step to achieve data reconstruction of the curve image. Text detection aims to locate the position of the value on the coordinate axis in the curve image. The value is then further obtained through the recognition network. The decoder part of the model proposed in this paper is consistent with the decoder part of EAST, which uses the idea of U-shaped structure to gradually fuse the features at different levels. Since the size of text regions varies greatly, locating large text requires deeper features, i.e., a larger receptive domain, while locating small text requires shallower features, i.e., a smaller receptive domain.

Therefore, the selection of features at different levels becomes a necessary requirement. In addition, since merging a large number of channels after a large feature map increases the computational overhead, the decoder uses a smaller upsampling branch, i.e., the decoder does not upsample the feature map to the same size as at the input, and the output image size is 1/4 of that at the input.

3.4. **Attention mechanism.** It is undoubtedly effective to pass features from the encoding process as supplementary information to the decoding process through jump connections. A large number of network structures have been designed using this approach because it helps to recover the information lost due to deep convolution operations [23-27]. Shallow features mainly contain geometric information such as texture and shape of the object, which is important for object boundary delineation. However, shallow features also contain noisy information, and jump connections, while recovering the information lost in deep convolution, also pass shallow noisy information to deep layers, which adversely affects task segmentation and recognition. So, an attention mechanism is employed to suppress the propagation of noisy information in the shallow layer and prevents the interference of noisy information, while ensuring the effective information transfer to the deep layer.

The convolutional block attention module is a lightweight general-purpose module that consists of two parts: channel attention and spatial attention which is proposed by Woo et al. The structure of convolutional block attention module is shown in Figure 2. This module derives the attention map sequentially from two independent dimensions, channel and spatial, and then multiplies the attention map with the input feature map for adaptive feature refinement [28]. The channel attention module uses the channel relationship of features to generate the channel attention graph, which focuses on "what" is meaningful given an input image. In order to calculate the channel attention feature effectively, the spatial dimension of the input feature map is compressed by means of channel superposition. The spatial attention module uses the combination of average pooling and maximum pooling to infer finer channel attention, because the average pooling layer can effectively learn the range of target objects, and the maximum pooling layer can obtain unique object features. The combination of these two parts can greatly improve the representation ability of the network.

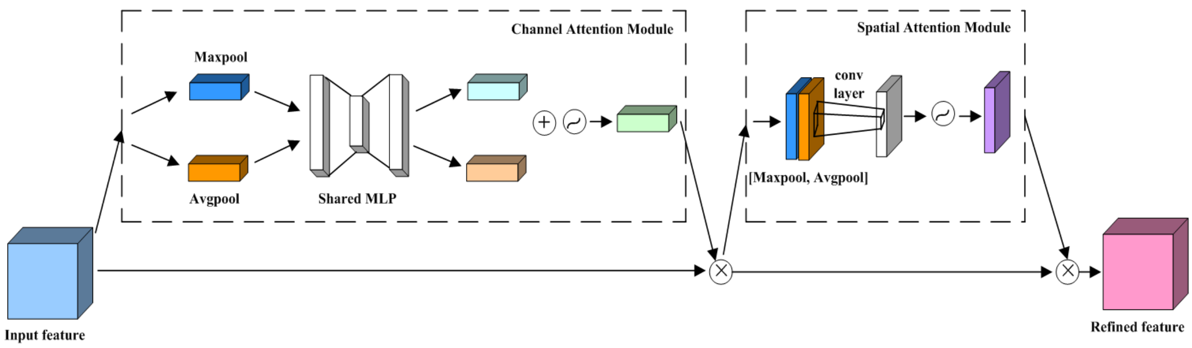


FIGURE 2. The structure of convolutional block attention module (CBAM)

3.5. **Loss function.** The calculation of the loss function in this model is divided into two parts: semantic segmentation and text detection. The loss functions used in the semantic segmentation task are cross-entropy loss and dice loss. Cross-entropy loss uses an interclass competition mechanism and is good at learning inter-class information. However, it only cares about the accuracy of the prediction probability of correct labels and ignores the differences of other mislabels, which leads to the scattering of learned features. Dice loss is a region-dependent loss. The loss and gradient values of a pixel point are related not only to the labeled and predicted values of that point, but also to the labeled and predicted values of other points. It has good performance for scenarios with severe imbalance between positive and negative samples. Hence, dice loss and cross-entropy loss are used in combination to play a complementary role in this paper, which is calculated as follows:

$$\text{seg\_loss} = -(n\log(p) + (1-n)\log(1-p)) + 1 - \frac{2 * gt * pre}{gt + pre} \tag{1}$$

where n denotes the number of categories, p denotes the probability of the predicted category, gt denotes the true label, and pre denotes the predicted value.

The loss function used for the text detection task is divided into three components: classification loss, position loss and angle loss, which correspond to the three outputs of the text detection model, i.e., score map, position map and angle map, respectively. The loss function is usually in the order of fractions, which makes the model learn the classification task easily and the error detection rate of the trained model is low. The position loss is calculated using intersection over union (IOU) loss, i.e., the loss is calculated using the IOU between the predicted frame and the true labeled frame; the larger the IOU value is, the closer the predicted frame is to the true labeled frame, and the smaller the loss value is. Since the value range of IOU is $[0, 1]$, the loss of this last part is taken as a positive value after taking the logarithm of the IOU result. The angular loss is calculated using the cosine function. Because the cosine function is an even function, it is not necessary to take the absolute value of the angle difference. Therefore, when the difference of two angles is smaller, the value of the cosine function is larger, and its value range is $[0, 1]$. So the angular loss is the value of 1 minus cos. The total loss function of the text detection task is calculated as follows:

$$\text{txt\_loss} = 1 - \frac{2 * gt * pre}{gt + pre} + \frac{-\log((iner + 1)/union + 1) * gt}{gt} + \frac{(1 - \cos(angle)) * gt}{gt} \tag{2}$$

where gt is the true label, pre is the predicted value, $iner$ is the intersection of the true label box and the predicted box, $union$ is the union set, and $angle$ is the difference between the true angle and the predicted angle.

4. **Experiment.** In this section, we first introduce the dataset used to train and test the performance of the proposed model. Next, the experimental settings are described. Then, the compared experimental results of the proposed model are given. Finally, we take the ablation experiment and the loss function matching coefficient selections.

4.1. **Dataset.** To verify the effectiveness of the proposed method, training and testing were conducted on a dataset containing 2292 curve images and corresponding semantic segmentation labels and text detection labels, of which 1911 were used as training data and 381 were used as testing data. The labels for semantic segmentation and text detection are independent of each other and participated in the training process together. The curve image in the dataset is a multi-color image containing a single curve, and the colors of the coordinate axes and curves contain a lot of kinds. The shape of the curve is mainly complex, but contains a few simple shapes. Some curve images from the dataset are selected and shown in Figure 3.

4.2. **Implementation.** The proposed network encoder uses the structure of the VGG16 network except for the fully connected layer. We initialize the network encoder using the pre-trained weights on the ImageNet dataset. In the joint training process, a total of 80 epochs are performed with 1×10^{-5} as the learning rate of the model. For the use of

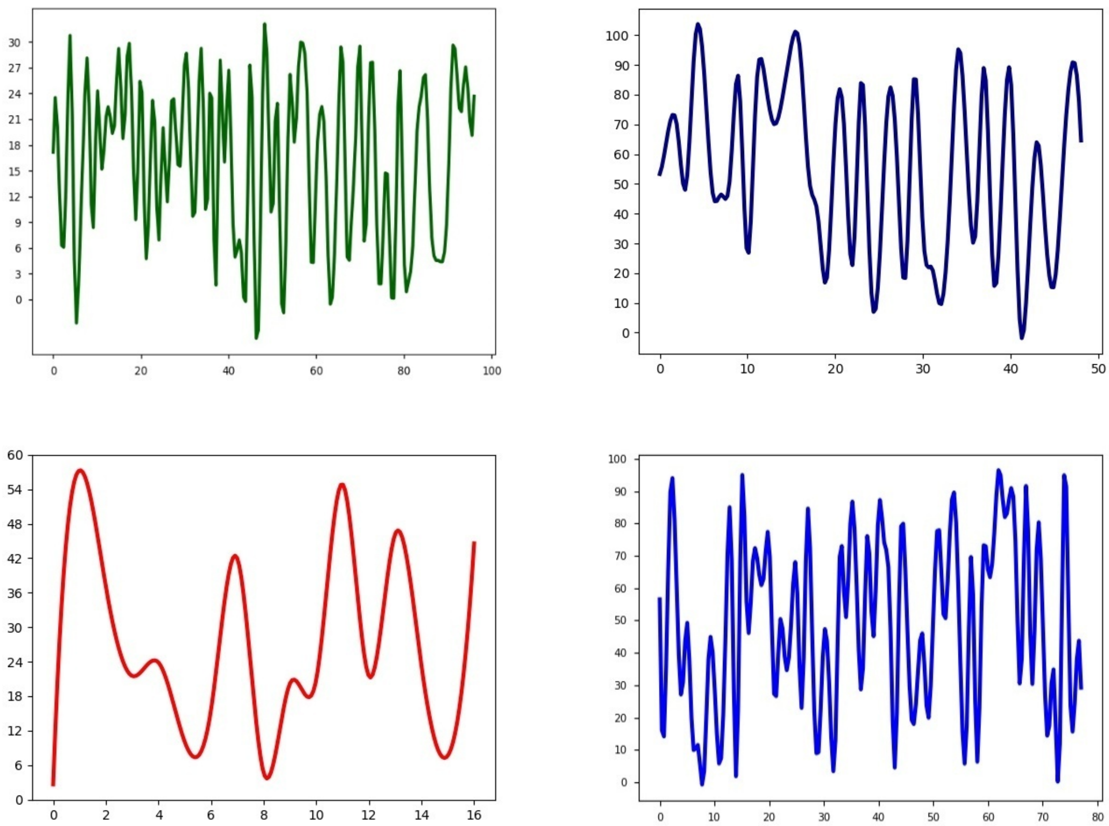


FIGURE 3. Samples from curve dataset

loss functions, different loss functions are used for the semantic segmentation and text detection tasks. The former uses cross-entropy loss and dice loss, while the latter uses dice loss, IOU loss and cosine loss. In addition, the loss function values of both are combined in a 5 : 5 ratio during the training process.

4.3. **Results.** To effectively evaluate the performance of the proposed model, evaluation metrics are selected for the semantic segmentation task and the text detection task to evaluate and compare the effectiveness of the proposed model in this paper. For the semantic segmentation task, the selected metrics include Accuracy, Sensitivity (SE), Specificity (SP) and F1-score, while for the text detection task, Precision, Recall and Hmean are selected as evaluation metrics. The specific evaluation metrics meanings and their calculations are as follows:

1) Accuracy, i.e., precision, represents the ratio of samples that are paired to the total number of samples, and is calculated as

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

2) Specificity (SP), i.e., specificity, indicates the probability that a negative sample from the original sample is correctly predicted as a negative sample, and is calculated as

$$specificity = \frac{TN}{TN + FP} \tag{4}$$

3) Sensitivity (SE), i.e., sensitivity, has the same meaning as Recall, i.e., recall, and indicates the probability that a positive sample from the original sample is correctly

predicted as a positive sample, which is calculated as follows:

$$sensitivity = recall = \frac{TP}{FN + TP} \tag{5}$$

4) Both *F1-score* and *Hmean* are the summed mean values of correctness and recall, reflecting the relationship between the two. The calculation formula is as follows:

$$F1\text{-}score = hmean = \frac{2 \times precision \times recall}{precision + recall} \tag{6}$$

where TP and TN indicate that the predicted result is the same as the actual result, and FP and FN indicate that the predicted result is the opposite of the actual result.

The curve segmentation and text detection results are tested on the curve image dataset in Table 1, where the segmentation accuracy is 94.0% and the text detection precision is 99.5%. In order to demonstrate the curve segmentation and text detection effects of the model in this work more intuitively, the test results of both are visualized in Figure 4. Figure 4(a) shows the result of curve segmentation. The segmented curve is represented in gray, and the black is the background of the non-curve part. Figure 4(b) shows the result of text detection. The detection results of values on the coordinate axes are represented by red boxes. Note that the original images corresponding to the results shown in 4(a) and 4(b) are in dataset introduction above.

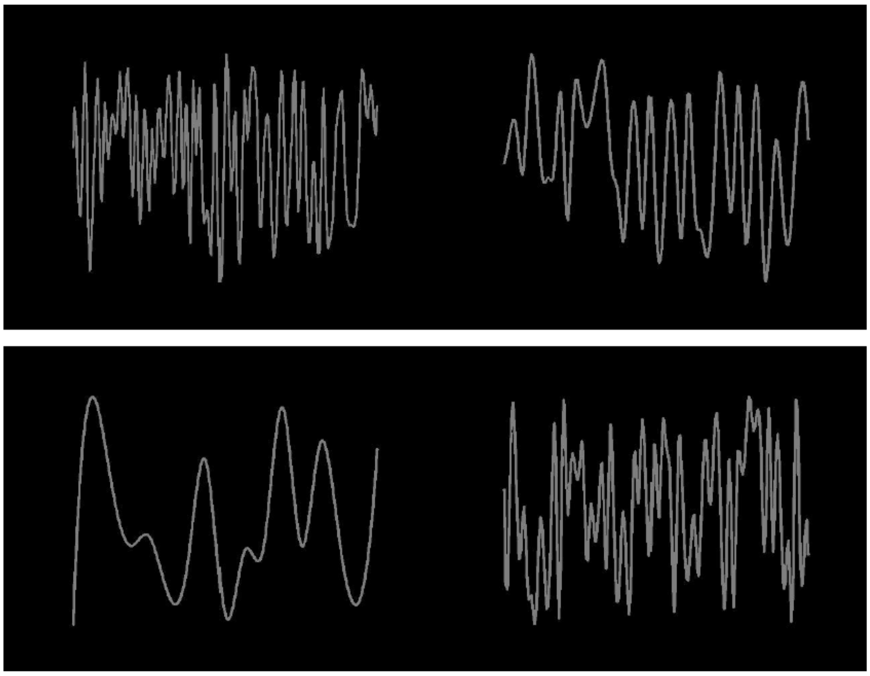
Table 1. Test results on curve image dataset

	Criterion	Segmentation result	Detection result
Curve segmentation	accuracy	0.940	/
	specificity	0.945	/
	sensitivity	0.869	/
	F1-score	0.612	/
Text detection	precision	/	0.995
	recall	/	0.996
	hmean	/	0.996

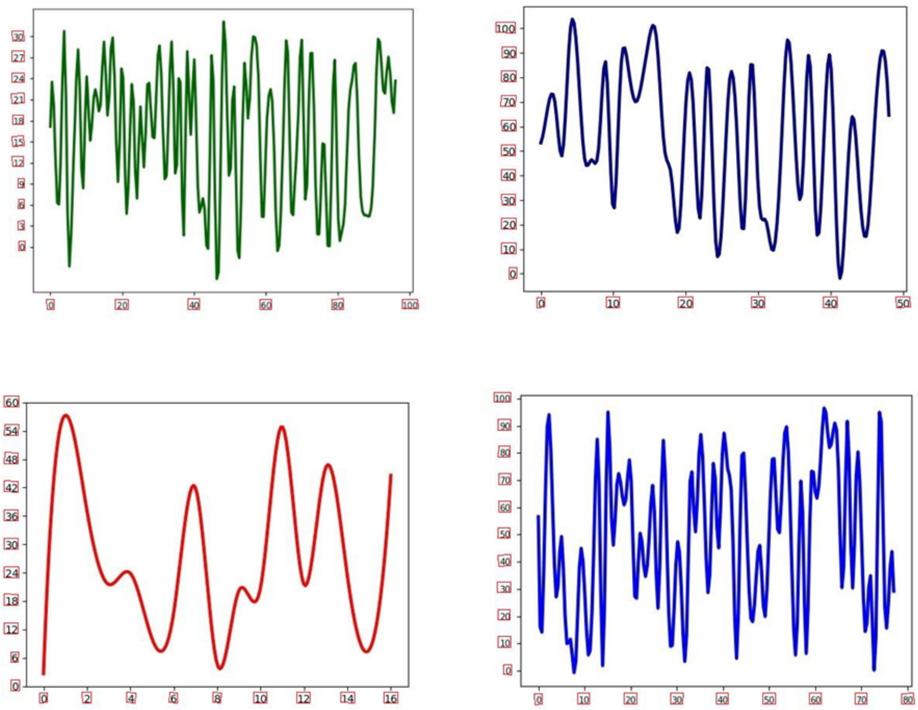
4.4. **Ablation.** In order to illustrate the necessity of the network structure and its components proposed in this paper, as well as the reasonableness of the loss function rationing coefficient setting, we conducted the corresponding ablation experiments, and the specific experimental procedure is described as follows.

1) **Joint Training.** Based on the structure design of co-encoding and independent decoding of curve segmentation and text detection tasks, this paper uses a multi-task joint training strategy to complete the model training of the two tasks. This joint training method can not only reduce the training time and improve the training efficiency, but also learn richer features in the process of joint encoding to further improve the performance of the model. To demonstrate the effectiveness of the joint training strategy adopted in this paper, ablation experiments of the model training are conducted. We conduct experiments on two separate tasks of curve segmentation and text detection, and compare them with the results of joint training. The experimental results show that the joint training of curve segmentation and text detection tasks is better than the two tasks trained separately in Table 2.

2) **Attention Module.** This experiment was conducted to demonstrate the effectiveness of the attention mechanism used in the proposed model. As shown in Table 3, we compare the effects of the model before and after using the CBAM module. Experimental results show that the use of CBAM module in the network structure encoder proposed in



(a) The result of curve segmentation



(b) The result of text detection

Figure 4. Visualization of test results

this paper can achieve the purpose of suppressing interference information and improving the accuracy of the model. After using CBAM module, the performance of three indicators is improved for curve segmentation task, and all indicators are improved for text detection task. In addition, we also tried other attention mechanism of the SE attention mechanism. Experimental results show that the SE attention mechanism helps to improve the accuracy of semantic segmentation tasks, but it interferes with the information

Table 2. Ablation result of joint training

	Criterion	Segmentation only	Detection only	Segmentation &detection
Segmentic segmentation	accuracy	0.931	/	**0.940**
	specificity	0.919	/	**0.945**
	sensitivity	0.882	/	0.869
	F1-score	0.593	/	**0.612**
Text detection	precision	/	0.991	**0.995**
	recall	/	0.966	**0.996**
	hmean	/	0.979	**0.996**

Table 3. Ablation result of attention module

	Criterion	With SE	Without CBAM	With CBAM
Segmentic segmentation	accuracy	**0.940**	0.937	**0.940**
	specificity	0.943	0.938	**0.945**
	sensitivity	**0.886**	0.812	0.869
	F1-score	**0.616**	0.609	**0.612**
Text detection	precision	0.990	0.994	**0.995**
	recall	0.950	0.995	**0.996**
	hmean	0.969	0.995	**0.996**

flow in text detection tasks and reduces the accuracy of text detection. Considering the prediction effect of the multi-task model for each sub-task, we finally choose the CBAM attention mechanism.

4.5. **Loss function coefficient.** The loss function is used to show the difference between the predicted value and the actual data. The smaller the value of the loss function, the better the performance of the model. Due to the higher complexity of curve segmentation task, the learning process of segmentation model is relatively slower, and the loss function value decreases more slowly. In order to train these two tasks jointly, it is necessary to find a balance coefficient of loss function so that the model training of the two tasks can get the optimal solution.

The balance coefficient of loss function involved in multi-task model training is usually the sum of the loss function values of multiple tasks in a certain proportion, which is used as the basis for the convergence of the whole model training. In this paper, the loss function values of the curve segmentation task and the text detection task are added in a $5 : 5$ ratio to obtain the total loss function values of the model. We also conducted experiments on other ratio settings to verify the rationality of the ratios. The experimental results are shown in Table 4. The data shows that the ratio adopted in this paper is more reasonable and conducive for the performance improvement.

5. **Conclusions.** In this paper, a convolutional neural network model based on deep learning method is proposed. The model adopts a multi-task co-encoding independent decoding structure, and uses the attention mechanism to improve the quality of information transmission in the jump connection process, so as to realize the joint training of curve segmentation and text detection tasks. To verify the effectiveness of the proposed model, we conduct experiments on the curve image dataset. Experimental results show that the accuracy of curve segmentation is 94.0%, and the precision of text detection is 99.5%. Compared with the traditional curve image data reconstruction methods, the proposed

Table 4. Experimental results of loss function coefficient

	Criterion	1 : 5	2 : 5	3 : 5	4 : 5	5 : 5
Segmentic segmentation	accuracy	0.936	0.937	0.934	**0.941**	0.940
	specificity	0.942	0.940	0.937	0.936	**0.945**
	sensitivity	0.840	0.881	**0.889**	0.853	0.869
	F1-score	0.588	0.601	0.593	0.609	**0.612**
Text detection	precision	0.987	0.984	0.969	0.982	**0.995**
	recall	0.989	0.979	0.971	0.981	**0.996**
	hmean	0.961	0.981	0.955	0.979	**0.996**

method gets rid of the dependence on artificial features, and achieves end-to-end curve image data reconstruction with higher accuracy. However, the proposed model has many areas for further optimization. For example, the information flow of both can be further considered to complement each other to achieve higher accuracy. In the future work, this paper will continue to study and improve.

REFERENCES

[1] K. Fu, B. Zheng and X. Li, Matlab based image curve data extraction method, *Journal of Shantou University: Natural Science Edition*, vol.25, no.2, pp.51-55, 2010.

[2] Z. Tan, S. Liu and X. Jiang, Data extraction method of complex curve image, *Electronic Measurement Technique*, vol.39, no.12, pp.159-163, 2016.

[3] R. Xu, Single transmission line curve extraction method based on digital image processing, *China Equipment Engineering*, 2018.

[4] H. Sheng, R. Cong, D. Yang, R. Chen, S. Wang and Z. Cui, UrbanLF: A comprehensive light field dataset for semantic segmentation of urban scenes, *IEEE Transactions on Circuits and Systems for Video Technology*, doi: 10.1109/TCSVT.2022.3187664, 2022.

[5] K. Rajamani, S. D. Gowda, V. N. Tej and S. T. Rajamani, Deformable attention for semantic image segmentation, *The 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society*, pp.3781-3784, doi: 10.1109/EMBC48229.2022.9871439, 2022.

[6] A. A. S. Gunawan, Fanny and E. Irwansyah, Development of paddy field mapping from satellite image using semantic segmentation method in Central Borneo, *ICIC Express Letters*, vol.15, no.9, pp.941-949, doi: 10.24507/icicel.15.09.941, 2021.

[7] A. Sagar and R. Soundrapandiyan, Semantic segmentation with multi scale spatial attention for self driving cars, *Proc. of the IEEE/CVF International Conf. on Computer Vision*, pp.2650-2656, 2021.

[8] M. Fan, S. Lai, J. Huang, X. Wei, Z. Chai, J. Luo and X. Wei, Rethinking biSeNet for real-time semantic segmentation, *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, pp.9716-9725, 2021.

[9] S. An, Q. Liao, Z. Lu and J.-H. Xue, Efficient semantic segmentation via self-attention and self-distillation, *IEEE Transactions on Intelligent Transportation Systems*, pp.1-11, 2022.

[10] R. Olaf, P. Fischer and T. Brox, U-net: Convolutional networks for biomedical image segmentation, in *Medical Image Computing and Computer-Assisted Intervention C MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science*, N. Navab, J. Hornegger, W. Wells and A. Frangi (eds.), Springer, Cham, 2015.

[11] H. Boudegga, Y. Elloumi, M. Akil, M. H. Bedoui, R. Kachouri and A. B. Abdallah, Fast and efficient retinal blood vessel segmentation method based on deep learning network, *Computerized Medical Imaging and Graphics*, vol.90, 2021.

[12] V. Sathananthavathi and G. Indumathi, Encoder enhanced atrous (EEA) Unet architecture for retinal blood vessel segmentation, *Cognitive Systems Research*, vol.67, pp.84-95, 2021.

[13] Y. Zhang, M. He, Z. Chen, K. Hu, X. Li and X. Gao, Bridge-Net: Context-involved U-net with patch-based loss weight mapping for retinal blood vessel segmentation, *Expert Systems with Applications*, vol.195, 2022.

[14] W. Huang, Z. Lin, J. Yang and J. Wang, Text localization in natural images using stroke feature transform and text co-variance descriptors, *Proc. of the IEEE International Conf. on Computer Vision*, pp.1241-1248, 2013.

[15] X. Yin, X. Yin, K. Huang and H.-W. Hao, Robust text detection in natural scene images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.36, pp.970-983, 2014.
[16] M. Liao, Z. Zou, Z. Wan, C. Yao and X. Bai, Real-time scene text detection with differentiable binarization and adaptive scale fusion, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, doi: 10.1109/TPAMI.2022.3155612, 2022.
[17] Z. Raisi, M. A. Naiel, G. Younes, S. Wardell and J. S. Zelek, Transformer-based text detection in the wild, *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp.3162-3171, 2021.
[18] X. Zhou, C. Yao, H. Wen, Y. Wang, S. Zhou, W. He and J. Liang, EAST: An efficient and accurate scene text detector, *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.5551-5560, 2017.
[19] T. Ali, M. F. Siddiqui, S. Shahab and P. P. Roy, GMIF: A gated multi-scale input feature fusion scheme for scene text detection, *IEEE Access*, vol.10, pp.93992-94006, 2022.
[20] J. Kang, M. Ibrayim and A. Hamdulla, MR-FPN: Multi-level residual feature pyramid text detection network based on self-attention environment, *Sensors*, vol.22, no.9, 2022.
[21] W. Zhang, Y. Qiu, M. Liao, R. Zhang, X. Wei and X. Bai, Scene text detection with scribble line, *International Conference on Document Analysis and Recognition*, pp.79-94, 2021.
[22] S. Karen and A. Zisserman, Very deep convolutional networks for large-scale image recognition, *arXiv Preprint*, arXiv: 1409.1556, 2014.
[23] J. Liang, F. Xu and S. Yu, A multi-scale semantic attention representation for multi-label image recognition with graph networks, *Neurocomputing*, vol.491, pp.14-23, 2022.
[24] T. Wu, J. Huang, G. Gao, X. Wei, X. Wei, X. Luo and C. Liu, Embedded discriminative attention mechanism for weakly supervised semantic segmentation, *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, pp.16765-16774, 2021.
[25] X. Ding, C. Shen, Z. Che, T. Zeng and Y. Peng, SCARF: A semantic constrained attention refinement network for semantic segmentation, *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, pp.3002-3011, 2021.
[26] Y. Zhu and J. Du, TextMountain: Accurate scene text detection via instance segmentation, *Pattern Recognition*, vol.110, 2021.
[27] Y. Wang, Z. Xu, X. Wang, C. Shen, B. Cheng, H. Shen and H. Xia, End-to-end video instance segmentation with transformers, *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, pp.8741-8750, 2021.
[28] S. Woo, J. Park, J.-Y. Lee and I. S. Kweon, CBAM: Convolutional block attention module, *Proc. of the European Conf. on Computer Vision*, pp.3-19, 2018.

Author Biography

Shuanzhu Sun received the B.Sc. degree in Production Process and Automation from North China Electric Power University, China, 1995; the M.Sc. degree in Power Plant Thermal Power Engineering from North China Electric Power University, China, 1998.

Mr. Sun is currently a fourth-level employee of China Jiangsu Frontier Electric Technology Co., Ltd. He mainly researches on the energy saving and emission reduction monitoring technology, power big data analysis application and related environmental protection products in the power industry.

Jiewen Zhu received the B.Sc. degree in Energy and Power Engineering from Southeast University, China, 2017; the M.Sc. degree in Power Engineering and Engineering Thermophysics from Southeast University, China, 2020.

Ms. Zhu is currently an employee of China Jiangsu Frontier Electric Technology Co., Ltd. She mainly researches on the energy conservation and emission reduction monitoring technology, carbon emission monitoring technology application and related products in the power industry. She has a deep interest in artifificial intelligence, image data reconstruction and abnormal data detection.

Chunlei Zhou received the B.Sc. degree in Environmental Monitoring from Nanjing University of Science and Technology, China, 1995; the M.Sc. degree in Computer Application from Donghua University, China, 1998.

Ms. Zhou is currently an employee of China Jiangsu Frontier Electric Technology Co., Ltd. She mainly researches on the power big data analysis, data mining technology application and related products. She has a deep interest in artificial intelligence, image data reconstruction and abnormal data detection.

Guoqiang Xu received the B.Sc. degree in Computer Science and Technology from East China University of Science and Technology, China, 2008.

Mr. Xu is currently an employee of China Jiangsu Frontier Electric Technology Co., Ltd. He mainly researches on the development and management of continuous smoke emission monitoring system, environmental protection monitoring platform, energy conservation and emission reduction system. He has a deep interest in artificial intelligence, image data reconstruction and abnormal data detection.

Tian Shi received the B.Sc. degree in Thermal Energy and Power Engineering from Southeast University, China, 2016; the M.Sc. degree in Power Engineering and Engineering Thermophysics from Southeast University, China, 2019.

Mr. Shi is currently an employee of China Jiangsu Frontier Electric Technology Co., Ltd. He mainly researches on the environmental protection of power supplies. He has a deep interest in artificial intelligence, image data reconstruction and abnormal data detection.