Chinese Character Component Segmentation Based on Character Structure Masks

Haiyan Li[1,2] and Fang Yang[1,2]

[1] School of Cyberspace Security and Computer, Hebei University, China
[2] Intelligent graphic Information Processing Institute, Hebei University, China
yangfang@hbu.edu.cn

Abstract. To address the issue where rectangular anchor boxes in object detection-based Chinese character component segmentation cannot segment semi-enclosed Chinese characters, this paper proposes a method for segmenting Chinese character components based on Chinese character structure masks. This method utilizes a U-Net encoder with ResNet as the backbone network, transforming the segmentation of Chinese character components into the generation of Chinese character structure masks. First, this study proposes a Res-CBAM module, which leverages the structural features of Chinese characters by incorporating CBAM into the residual U-Net network, effectively solving the problem of incomplete segmentation of Chinese character components. Secondly, a vector-guided supervision mechanism is designed to guide the training process of the model by designing structure vectors of Chinese characters, effectively addressing the issue of component adhesion in Chinese characters. Experimental results demonstrate that compared to traditional object detection methods, this method can achieve fast and efficient segmentation in lightweight networks by training small datasets.

Keywords: Chinese character component segmentation · ResNet · CBAM · Structure masks.

1 Introduction

Chinese characters, as an integral part of Chinese culture, are the key carriers for recording and transmitting Chinese civilization. As the basic units of Chinese writing, Chinese characters are composed of various components and have complex and diverse structures. Chinese character component segmentation is the process of breaking down a Chinese character into its independent components according to its structure. For example, the character "好", which has a left-right structure, can be divided into two components, "女" and "子". Compared to the character, the number of components is smaller, and the information they contain is more specific. This method of segmenting Chinese characters into components not only simplifies information processing operations but also improves processing accuracy. It holds significant importance for related research areas such as Chinese character recognition [1], Chinese character font conversion [2], and automatic generation of Chinese character styles.

The number of Chinese characters is extremely large, and their font styles are diverse. Component adhesion or overlap often occurs, increasing the difficulty of Chinese character component segmentation. Currently, segmentation of Chinese character components based on object detection has become relatively mature for simple structures such as left-right and top-bottom arrangements. However, the main challenge lies in the segmentation of enclosed and semi-enclosed structures. Examples of Chinese character component segmentation are shown in Fig.1, which includes several representative Chinese characters with enclosed and semi-enclosed structures and their corresponding correct segmentation methods.

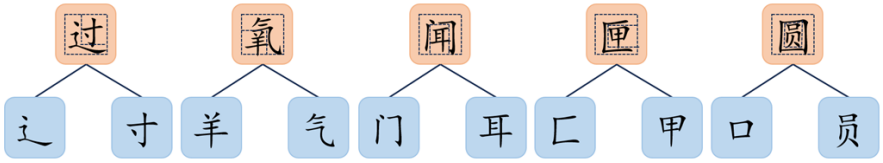

Fig. 1. Example of Chinese character component segmentation.

In recent years, with the continuous development of computer vision and deep learning technologies, image segmentation techniques have made significant progress in various fields, such as autonomous driving [3], remote sensing image analysis [4,5], and medical image segmentation [6,7]. These advancements provide new solutions to the problem of complex Chinese character component segmentation.

Image segmentation techniques primarily involve classifying pixels using semantic labels, segmenting individual objects, or a combination of both to label various object types at the pixel level [8]. Commonly used image segmentation techniques include encoder-decoder-based models and attention-based models. Encoder-decoder models map data points from the input domain to the output domain through a two-stage network. Typical examples include deconvolutional networks [9], SegNet architectures [10], Fully Convolutional Networks (FCNs) [11], and the DeepLab family [12], all of which are based on the encoder-decoder concept. These models extract features from input images using an encoder and gradually restore spatial resolution using a decoder to achieve high-precision segmentation. Similarly, U-Net [13], as a classic image segmentation architecture, also adopts the encoder-decoder concept. It is characterized by a symmetrical U-shaped structure and a strong ability to handle small sample data. [14] combined ResNet with U-Net to address the gradient vanishing and exploding problems in deep neural networks, and subsequent studies [15–17] have adopted this combination. With the development of attention mechanisms, [18,19]introduced attention mechanisms that dynamically adjust the weights of different positions in the feature map, allowing the network to focus more on key areas. Channel attention (SENet) [20], self-attention mechanisms [21], and spatial attention (STN) [21] have been widely used in the field of image segmentation.

In 2018, [22] proposed the BAM attention mechanism, which combines channel and spatial attention. Subsequently, [23] improved upon BAM and proposed the CBAM attention mechanism, further enhancing the model's feature representation capabilities. [24, 25] incorporated the CBAM attention mechanism into residual U-Net networks for image segmentation tasks on small datasets, effectively enhancing the model's grasp of key features and perception of detailed information.

Inspired by the encoder-decoder network architecture, and faced with the extreme irregularity of Chinese character components, we propose a method for segmenting Chinese character components based on Chinese character structure masks to address the current inability to segment semi-enclosed structures. This method uniquely combines the residual U-Net architecture with two attention mechanisms, BAM and CBAM. By leveraging the advantages of the residual U-Net structure on small datasets, the model focuses on both global and local features in the input data, enhancing overall performance. Our model successfully extracts structural features of Chinese characters and achieves accurate segmentation by using the residual U-Net architecture along with BAM and CBAM. Compared to existing models, our proposed method improves accuracy and application range, solving the problem of rectangular bounding boxes in object detection being unable to segment semi-enclosed Chinese characters.

The main contributions of this paper are as follows:

1. We propose a method for segmenting Chinese character components based on Chinese character structure masks, which solves the problem of object detection being unable to segment semi-enclosed Chinese characters.

2. To address the issue of the model recognizing the entire character as a single class, we incorporated BAM and CBAM attention mechanisms into the network to resolve component adhesion problems both globally and locally.

3. To tackle the issue of incomplete segmentation caused by small internal gaps and adhesion within Chinese characters, we designed corresponding structure vectors for Chinese character structures. By controlling the loss reduction between the initial structure vectors and the intermediate generated vectors, we optimized the overall training of the model.

2 Method

2.1 The Proposed Framework

The proposed Chinese character component segmentation method primarily consists of a ResNet encoder and a U-Net decoder, with the network architecture shown in Fig. 2. To address the issue where the model often treats two or three components that should be independent as a single entity during the segmentation process, leading to inaccurate segmentation, we incorporated BAM and CBAM attention mechanisms into ResNet. Additionally, we designed different structure vectors for each type of Chinese character structure. By controlling the loss function descent between the initial vectors and the intermediate vectors,

we optimize the overall training of the model to achieve complete segmentation of Chinese character components.

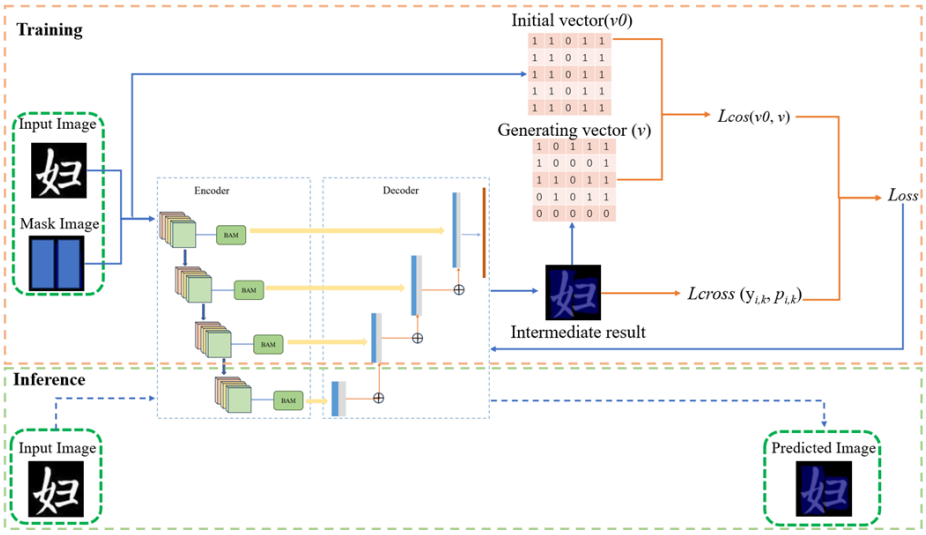

Fig. 2. The proposed framework for image registration. The orange dashed box contains the training process, where the Input Image and Mask Image are the input images and their corresponding masks, respectively. The gray dashed box contains the encoder ResNet and the decoder U-Net. The Intermediate result is the image generated during training, the Generating vector (v) is the vector generated from the Intermiediate result, and the Initial vector (v_0) is the vector of the Input Image. $Lcos(v_0, v)$ represents the cosine similarity loss between v_0 and v, $Lcross$ $(y_{i,k}, p_{i,k})$ represents the original network loss, and $Loss$ is the sum of these two losses. The green dashed box contains the inference process, and the Predicted Image represents the final output.

2.2 Data Preprocessing

In the field of Chinese character segmentation and image processing, correctly parsing the structure of Chinese characters is a crucial step. To this end, we comprehensively analyzed the positional relationships of Chinese character components and, based on these relationships, classified Chinese characters into 13 different structural types [26], including left_right, up_down, up_right, up_left, left_down, up_three, down_three, left_three, surrounded, frame, single_font, up_center_down, and left_center_right. For each structure, we adopted a labeling strategy with structural information to facilitate more accurate training and evaluation of the Chinese character component segmentation network. The specific structural classifications and their corresponding examples are shown in Table 1. This classification helps us to better understand the complex structures of Chinese characters.

Table 1. Chinese character structure classification and corresponding examples.

Serial Number	Label Name	Example Word
1	left_right	
2	up_down	
3	up_right	
4	up_left	
5	left_down	
6	up_three	
7	down_three	
8	left_three	
9	surrounded	
10	frame	
11	left_center_right	
12	up_center_down	
13	single_font	

In the process of Chinese character component segmentation, complex characters often contain multiple nested structures. For example, as shown in Fig.3, the character "哒" not only has a left_right structure but also contains the up_left structure of "达", making accurate segmentation extremely difficult. To address the challenge of segmenting Chinese characters with multiple nested structures, we decompose the characters layer by layer, with each layer belonging to one of the 13 types. By segmenting components at each layer, we achieve the goal of fully segmenting all structures of the Chinese characters.

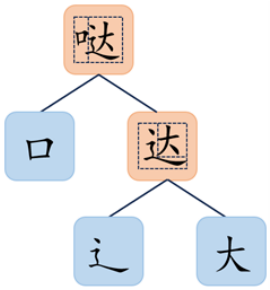

Fig. 3. Structural hierarchy of the Chinese character "哒".

2.3 Encoder-Decoder Framework

In this section, we adopted an improved ResNet-50 and U-Net-based encoder-decoder network for finer segmentation of Chinese character components. The encoder-decoder network structure is shown in Fig.4. The improved ResNet-50 encoder constitutes the "contraction" path of the U-Net, optimizing the capture of global information, while the "expansion" path of the U-Net acts as the decoder, utilizing skip connections to restore detailed information, thereby enhancing segmentation accuracy and efficiency. This structural design is particularly suitable for processing Chinese character images, as the semantics of Chinese characters are relatively simple and their structures are stable. Additionally, Chinese character image datasets typically belong to simple and small sample datasets.

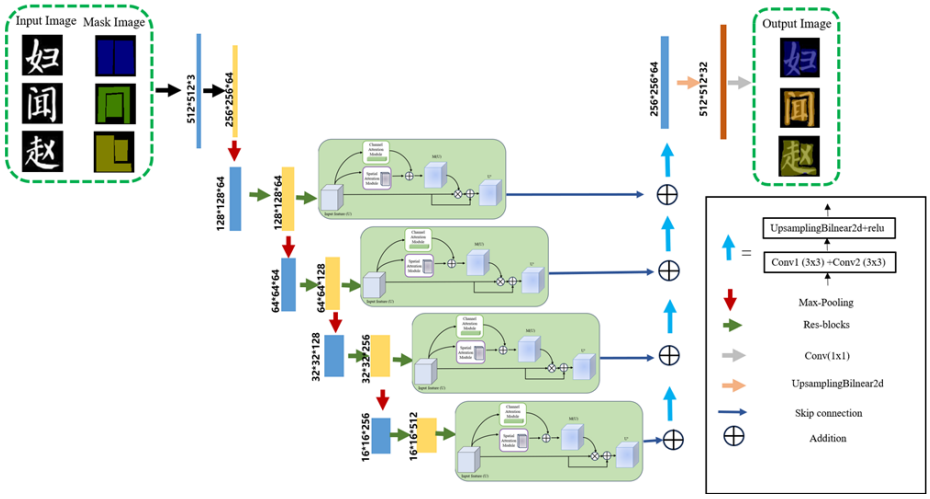

Fig. 4. Encoder-Decoder network structure.

In deep learning networks, efficient and accurate feature extraction is crucial to the overall performance of the network. To address the specific needs of component segmentation, we made key improvements to the standard ResNet-50 architecture to enhance its performance as an encoder in image segmentation tasks. 1) At the end of each stage, we integrated the BAM [22] attention mechanism, placing it on higher-level feature maps to enhance global features and enable the model to better understand the overall structure and context of the image. 2) We added the CBAM [23] attention mechanism after each residual block to optimize the feature maps layer by layer, reinforcing key local features and detailed information.

The BAM network structure consists of parallel spatial and channel attention mechanisms, placed after higher-level feature maps, which are smaller in size

and contain more abstract semantic information. The parallel structure allows BAM to process both spatial and channel information simultaneously, thereby enhancing global features. This enables the model to better understand the overall layout and structure of Chinese characters, accurately identifying their overall contours and correctly labeling their colors.

A standard residual block receives a 256-dimensional feature matrix as input and processes it through three convolutional layers to obtain the processed features. Unlike the standard design, this paper proposes the Res-CBAM module, which adds a CBAM composed of sequential spatial and channel attention mechanisms at the end of each residual block. The design of the residual block and the CBAM network structure are shown in Fig.5. Addressing the issue of small gaps and frequent adhesion between Chinese character components, the CBAM with a sequential structure can capture detailed and local features of Chinese characters after each residual block, enabling the model to accurately segment complex Chinese character components.

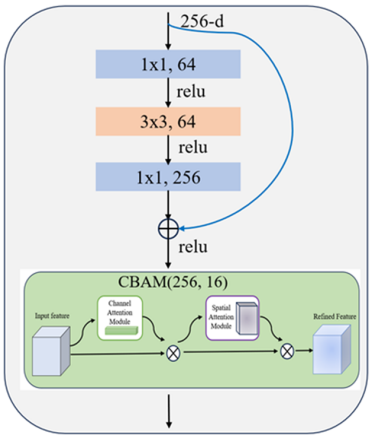

Fig. 5. Res_block structure diagram.

The combined use of BAM and CBAM optimizes the feature extraction process at both global and local levels, enabling the model to accurately recognize complex Chinese character structures and perform precise segmentation. This effectively enhances the model's performance and robustness in the task of Chinese character component segmentation.

2.4 Vector supervision mechanism

To further address the issue of Chinese character component adhesion in images, we introduced a vector supervision mechanism [27] during the training process. These vectors are represented as 5×5 matrices, with each element consisting of 0 or 1, and Chinese character structure vector are shown in Table 2.

Table 2. Chinese character structure vector.

left_right				
1	1	0	1	1
1	1	0	1	1
1	1	0	1	1
1	1	0	1	1
1	1	0	1	1

up_down				
1	1	1	1	1
1	1	1	1	1
0	0	0	0	0
1	1	1	1	1
1	1	1	1	1

up_right				
1	1	1	1	1
1	1	1	1	1
0	0	0	1	1
0	0	0	1	1
0	0	0	1	1

up_left				
1	1	1	1	1
1	1	1	1	1
1	1	0	0	0
1	1	0	0	0
1	1	0	0	0

left_down				
1	1	0	0	0
1	1	0	0	0
1	1	0	0	0
1	1	1	1	1
1	1	1	1	1

up_three				
1	1	1	1	1
1	0	0	0	1
1	0	0	0	1
1	0	0	0	1
1	0	0	0	1

down_three				
1	0	0	0	1
1	0	0	0	1
1	0	0	0	1
1	0	0	0	1
1	1	1	1	1

left_three				
1	1	1	1	1
1	0	0	0	0
1	0	0	0	0
1	0	0	0	0
1	1	1	1	1

surrounded				
1	1	1	1	1
1	0	0	0	1
1	0	0	0	1
1	0	0	0	1
1	1	1	1	1

frame				
1	1	1	1	1
0	0	1	0	0
0	0	1	0	0
0	0	1	0	0
0	0	1	0	0

left_center _right				
1	0	1	0	1
1	0	1	0	1
1	0	1	0	1
1	0	1	0	1
1	0	1	0	1

up_center _down				
1	1	1	1	1
0	0	0	0	0
1	1	1	1	1
0	0	0	0	0
1	1	1	1	1

single_font				
1	1	1	1	1
0	0	0	1	0
0	0	1	0	0
0	1	0	0	0
1	1	1	1	1

In practical applications, we set corresponding initial vectors for different Chinese character structures. For example, for the frame and single_font structures, we designed the shapes of "T" and "Z" respectively. For the left_right, up_down, left_center_right, and up_center_down structures, we fitted the vectors according to the actual layout of the Chinese characters to ensure that each component is adequately represented. For semi-enclosed structures (such as up_right and up_left), we focused on the special non-rectangular parts of the characters. For fully-enclosed structures (surrounded), we emphasized only the surrounding parts. The purpose of this design is to increase the weight of non-rectangular parts, enhance the model's ability to recognize the details and overall layout of Chinese characters, thereby reducing the adhesion of character components and improving the accuracy and stability of the character component segmentation task.The specific process is shown in the training section of Fig.2. The structure vector v is generated by fitting the intermediate result to encode the structural attributes of Chinese characters. To optimize the model's performance, this study employs a cosine-based loss function, as follows:

$$Loss(v_0, v) = 1 - cos(v_0, v) \tag{1}$$

Finally, the obtained loss $Loss(v_0,v)$ is added to the original cross-entropy loss $Lcross(y_{i,k},p_{i,k})$ to form the overall network loss function $Loss$, as follows:

$$Loss = Loss(v_0, v) + Lcross(y_{i,k}, p_{i,k}) \tag{2}$$

This method is designed with the diversity of complex Chinese character structures in mind. By controlling the descent of the overall loss function, the network is supervised and optimized, thereby enhancing the model's accuracy and robustness in handling images of Chinese characters with multiple structures.

3 Experiments

3.1 Dataset

In this study, we selected the works of the renowned calligrapher Yan Zhenqing (709 A.D.-784 A.D., Tang Dynasty) as the dataset. The single-character dataset [28] was obtained through web resources and contains a total of 5,412 calligraphic Chinese character images, with 4,364 images in the test set and 1,048 images in the training set. The selection of the training set was based on the statistical proportions of Chinese character structures [26], and single word image dataset information table are shown in Table 3. Additionally, to more intuitively evaluate the segmentation effects, we assigned unique label colors to each structural type.

Table 3. Single word image dataset information table.

Label Type number	Image percentage	Image name	Label color
left_right	56.52%	592	
up_down	23.30%	244	
up_right	0.79%	8	
up_left	3.89%	41	
left_down	2.62%	27	
up_three	0.90%	9	
down_three	0.10%	2	
left_three	0.17%	2	
surrounded	0.38%	4	
frame	0.55%	6	
left_center_right	0.31%	3	
up_center_down	0.65%	7	
single_font	9.81%	103	
total	**-**	**1048**	**-**

3.2 Implementation Details

We used the Adam optimizer, with specific hyperparameter details provided in Table 4. The model was trained on a server equipped with an NVIDIA GeForce RTX 3080 GPU (8GB of RAM) and an Intel(R) Core(TM) i5-8500 CPU @ 3.00GHz. To fully leverage the advantages of pre-trained models, we employed a transfer learning approach to accelerate the training process and improve the model's generalization ability. The training process was divided into two phases: in the first phase, the backbone network weights were frozen, and the model was trained for 50 epochs to stabilize the foundational features; in the second phase, the weights were gradually unfrozen to fine-tune the network and optimize performance, continuing for another 50 epochs. This process was repeated multiple times, with model performance evaluated every 5 epochs. The model was quantitatively evaluated using three key metrics: accuracy, mean Intersection over Union (mIoU), and mean Pixel Accuracy (mPA). During training, we found that after introducing the structure vectors, the value of $Loss(v_0, v)$ eventually stabilized around 0.6932.The average predicted cost of a picture is 0.068s.

Table 4. Hyperparameter values for the proposed model.

Parameters	Values
Optimizer	Adam
Max Learning rate	$1e^{-4}$
Min Learning rate	$1e^{-4} * 0.01$
Epochs	100
IoU smoothing factor	$1e^{-5}$
Lr_decay_type	cos
Batch size	2
Dice score smoothing factor	$1e^{-5}$
Metrics	**Accuracy, Loss, mIoU, mPA**

3.3 Comparative Experiments

To comprehensively evaluate the performance of the proposed model, we designed two sets of comparative experiments. 1) We used different encoder architectures during training to assess the impact of various encoder architectures on model performance. 2) We compared the performance of our model with two representative models, Faster RCNN [29] and SAM [30].

Comparison Based on Encoder. In this experiment, we used ResNet-50 as the encoder and U-Net as the decoder for structured mask segmentation of Chinese characters. To verify whether ResNet-50 is the optimal choice as an encoder, we compared it with VGG-16 [31], ResNet-50, ResNet-101, and ResNet-152. Each encoder was tested under the same experimental conditions. The visualization results of the tests are shown in Fig.6, clearly demonstrating that

ResNet-50 provides the best performance in the Chinese character segmentation task. It accurately recognizes the structure of Chinese characters and marks them with a single color, whereas other encoders display multicolored results and fail to correctly identify the structure of the characters.

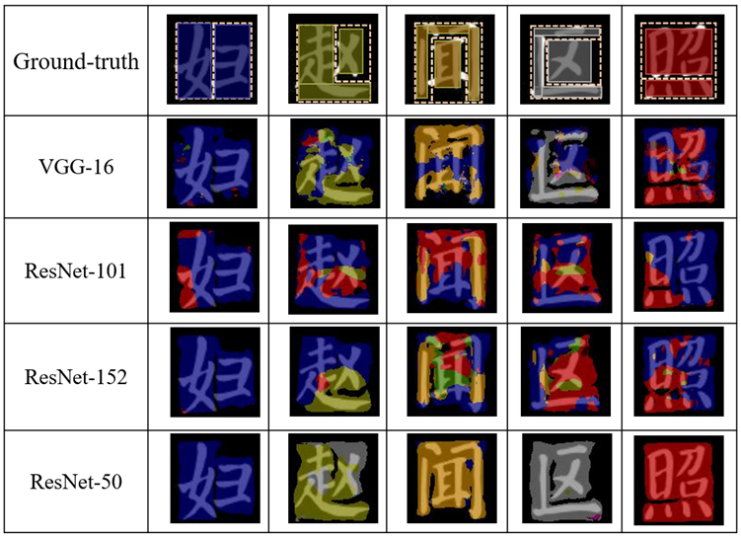

Fig. 6. Schematic diagram of the comparative segmentation of the encoder.

To quantify the model performance, we evaluated it based on specific metrics, with the results presented in Table 5. Comparing the performance of different encoders shows that the configuration using ResNet-50 performed the best in all tests, achieving an accuracy of 77.67%, which is 5.49%, 0.88%, and 1.27% higher than VGG, ResNet-101, and ResNet-152, respectively. Furthermore, the performance of ResNet-50 in terms of mIoU and mPA also confirmed its superiority, which can be attributed to its better feature extraction capability.

Table 5. Encoder performance.

Network structure	Accuracy(%)	mIoU(%)	mPA(%)
VGG-16	72.18	18.05	23.44
ResNet-50	**77.67**	**21.16**	**26.93**
ResNet-101	76.79	19.39	25.11
ResNet-152	76.40	18.90	24.77

Chinese Character Component Segmentation Based on Object Detection Algorithms. We conducted a comparative experiment on Chinese character component segmentation based on Faster RCNN using the same dataset under the same experimental conditions. Similar to our experiment, Faster RCNN performed target detection-based segmentation for Chinese character components, while our method generates semantic segmentation masks. Due to the rectangular detection box of target detection, the segmentable Chinese character structures are relatively simple. Additionally, due to the adhesion between calligraphic fonts, the segmented parts often contain portions of other components. Moreover, we conducted a comparative experiment on Chinese character structure mask segmentation based on SAM, whose main issue is treating the Chinese character as a whole. The visualization results of the comparative experiments are shown in Fig.7.

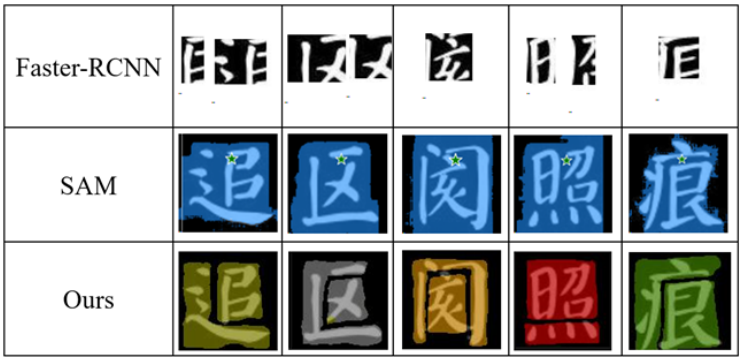

Fig. 7. Schematic diagram of the comparative segmentation of the encoder.

Evaluation Metrics for Various Segmentation Models are shown in Table 6. The experimental results indicate that, under the same number of epochs, our network model demonstrates better performance compared to other backbone networks.

Table 6. Evaluation Metrics for Various Segmentation Models.

Network structure	Accuracy(%)	mIoU(%)	mPA(%)
Ours	86.98	41.02	47.16
Faster RCNN	-	-	-
SAM	83.25	36.18	43.59

Thus, it can be seen that the combination of our model with the attention mechanisms, along with the adopted supervision strategy, forms a complementary bidirectional mechanism. By dividing Chinese characters into 13 structures,

the model becomes more focused on these 13 structures, thereby reducing inference time and training load. The attention mechanisms ensure that the model primarily concentrates on the global and local features of these 13 structures, rather than solely focusing on the entire Chinese character or a single component. Consequently, our model performs exceptionally well in the task of Chinese character segmentation, effectively extracting and processing the detailed features of Chinese characters, significantly improving segmentation accuracy.

3.4 Ablation Study

We used the architecture of ResNet-50 as the encoder and U-Net as the decoder as our baseline model, denoted as A. Each added module was evaluated, and the Performance comparison between different strategies are shown in Table 7. Compared to the baseline, the addition of BAM showed little improvement, while the addition of CBAM increased accuracy by 7.12%, with mIoU and mPA improving by 11.45% and 14.02%, respectively. When both BAM and CBAM were added, there was an improvement compared to the case where only CBAM was added, but the improvement was not significant. However, the performance did not improve when the loss function was added alone, as the structural vectors were introduced after the training, thus the loss function did not have its intended effect. After combining the attention mechanism with the loss function, accuracy, mIoU, and mPA improved by 9.31%, 19.86%, and 20.23%, respectively. This indicates that the attention mechanism alone still struggles to achieve precise segmentation, whereas the combination of these two steps with the supervision of the loss function resulted in the best performance for the model.

Table 7. Performance comparison between different strategies.

Description	Accuracy(%)	mIoU(%)	mPA(%)
A(Baseline)	77.67	21.16	26.93
A+ BAM	77.87	21.55	28.53
A+ CBAM	84.79	32.61	40.95
A+ CBAM + BAM	85.28	38.06	45.01
A+ Loss	77.36	20.60	26.15
A+CBAM+BAM+Loss	**86.98**	**41.02**	**47.16**

For each added module, we provide the corresponding visualization results, as shown in Fig.8. It can be observed that despite the addition of BAM and CBAM attention mechanisms, some adhesion still occurs. However, after introducing the vector supervision mechanism, the segmentation becomes more thorough, and the overall performance reaches its optimal state.

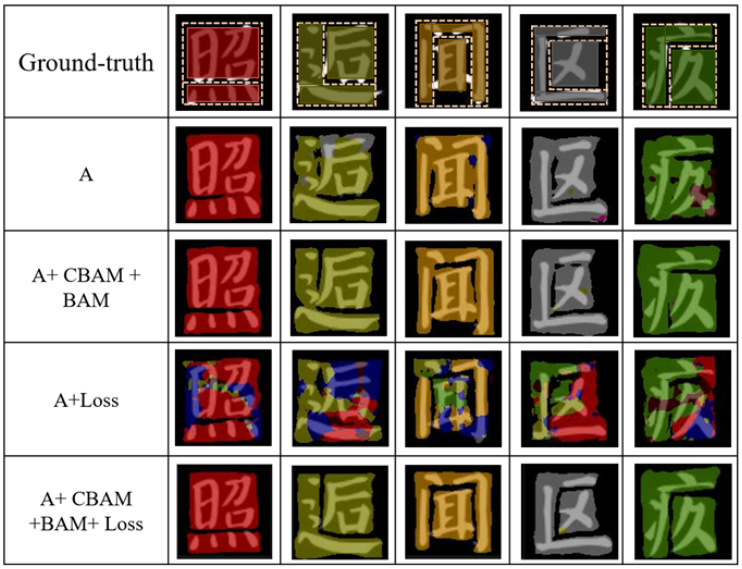

Fig. 8. Visualization Results of Ablation Study.

4 Conclusions

In this paper, we propose a Chinese character component segmentation method based on Chinese character structure masks, which transforms the segmentation of Chinese character components into generating character image masks. This method combines ResNet-50 with U-Net and introduces CBAM and BAM attention mechanisms to accurately capture the details and local components of Chinese characters, effectively improving segmentation accuracy. Additionally, the vector-guided supervision mechanism further optimizes the training process, enabling the model to converge quickly and perform excellently even on small datasets. Experimental results show that our method significantly outperforms other methods in the task of Chinese character segmentation, with higher robustness.

However, during mask generation, there are still cases of incomplete segmentation due to the small gaps in the left_center_right and up_center_down structures. In the future, we will focus on thoroughly segmenting these two structures and continue to optimize the model architecture, attempting to validate its performance in more complex application scenarios.

Acknowledgments. This work was supported and funded by the Science and Technology Project of Hebei Education Department(No.ZD2019131).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. M.-M. Yu, H. Zhang, F. Yin, and C.-L. Liu, "An approach for handwritten chinese text recognition unifying character segmentation and recognition," *Pattern Recognition*, vol. 151, p. 110373, 2024.
2. W. Li, S. Huang, and Y. Shao, "An unsupervised framework for adaptive context-aware simplified-traditional chinese character conversion," in *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, 2024, pp. 1318–1326.
3. J. Li, F. Jiang, J. Yang, B. Kong, M. Gogate, K. Dashtipour, and A. Hussain, "Lane-deeplab: Lane semantic segmentation in automatic driving scenarios for high-definition maps," *Neurocomputing*, vol. 465, pp. 15–25, 2021.
4. Y. Hou, Z. Wu, X. Ren, K. Liu, and Z. Chen, "Bffnet: A bidirectional feature fusion network for semantic segmentation of remote sensing objects," *International Journal of Intelligent Computing and Cybernetics*, vol. 17, no. 1, pp. 20–37, 2024.
5. J. Fan, J. Li, Z. Hua, F. Zhang, and C. Zhang, "Elevation information-guided multimodal fusion robust framework for remote sensing image segmentation," *IEEE Geoscience and Remote Sensing Letters*, 2024.
6. S. Srinivasan, K. Durairaju, K. Deeba, S. K. Mathivanan, P. Karthikeyan, and M. A. Shah, "Multimodal biomedical image segmentation using multi-dimensional u-convolutional neural network," *BMC Medical Imaging*, vol. 24, no. 1, p. 38, 2024.
7. F. J. P. Montalbo, "S3ar u-net: A separable squeezed similarity attention-gated residual u-net for glottis segmentation," *Biomedical Signal Processing and Control*, vol. 92, p. 106047, 2024.
8. H. Chen and K. Kim, "Multi-convolutional channel residual spatial attention u-net for industrial and medical image segmentation," *IEEE Access*, 2024.
9. H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1520–1528.
10. V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.
11. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.
12. L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.
13. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241.
14. H. Zhang, X. Hong, S. Zhou, and Q. Wang, "Infrared image segmentation for photovoltaic panels based on res-unet," in *Chinese conference on pattern recognition and computer vision (PRCV)*. Springer, 2019, pp. 611–622.
15. K. Cao and X. Zhang, "An improved res-unet model for tree species classification using airborne high-resolution images," *Remote Sensing*, vol. 12, no. 7, p. 1128, 2020.

16. H. Guo, Z. Guo, Z. Pan, and X. Liu, "Bilateral res-unet for image colorization with limited data via gans," in *2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*. IEEE, 2021, pp. 729–735.
17. L. Yuan, Y. Li, Y. Si, J. Ren, Y. Yang, Y. Gong, Y. Xia, Z. Tong, and L. Tong, "Multi-objects change detection based on res-unet," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. IEEE, 2021, pp. 4364–4367.
18. Z. Huang, Y. Zhao, Y. Liu, and G. Song, "Gcaunet: A group cross-channel attention residual unet for slice based brain tumor segmentation," *Biomedical Signal Processing and Control*, vol. 70, p. 102958, 2021.
19. A. Mohammed, "Resattunet: detecting marine debris using an attention activated residual unet," *arXiv preprint arXiv:2210.08506*, 2022.
20. J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.
21. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.
22. J. Park, S. Woo, J.-Y. Lee, and I. S. Kweon, "Bam: Bottleneck attention module," *arXiv preprint arXiv:1807.06514*, 2018.
23. S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.
24. H. A. Shah and J.-M. Kang, "An optimized multi-organ cancer cells segmentation for histopathological images based on cbam-residual u-net," *IEEE Access*, 2023.
25. Z. Yang, C. Xu, and L. Li, "Landslide detection based on resu-net with transformer and cbam embedded: Two examples with geologically different environments," *Remote Sensing*, vol. 14, no. 12, p. 2885, 2022.
26. H. Xing, "A statistic analysis of components of the character entries in the hsk graded character list," *Chinese Teaching in the World*, vol. 72, pp. 49–55, 2005.
27. M. Kim, I. Oh, D. Yun, and K. Ko, "Improved semantic segmentation network using normal vector guidance for lidar point clouds," *Journal of Computational Design and Engineering*, vol. 10, no. 6, pp. 2332–2344, 2023.
28. X. Gao, F. Yang, T. Chen, and J. Si, "Chinese character components segmentation method based on faster rcnn," *IEEE Access*, vol. 10, pp. 98 095–98 103, 2022.
29. R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.
30. A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo *et al.*, "Segment anything," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 4015–4026.
31. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.