Chinese Character Components Segmentation Method Based on Faster RCNN

XIANG GAO, FANG YANG, TIAN CHEN, AND JIANHUI SI
School of Cyber Security and Computer, Hebei University, Baoding 071002, China
Hebei Machine Vision Engineering Research Center, Hebei University, Baoding 071002, China
Corresponding author: Fang Yang (yangfang@hbu.edu.cn)

This work was supported in part by the Science and Technology Project of Hebei Education Department under Grant ZD2019131, and in part by the “One Province, One University” of Hebei University under Grant 521000981155.

ABSTRACT To solve the component segmentation problem caused by the sticking and overlapping of parts in incoherent handwritten calligraphy characters, we propose a Chinese character part segmentation method based on Faster RCNN. The method utilizes the advantages of Faster RCNN on multi-scale and small targets to solve difficult problems in component segmentation. The hierarchical features of the components were used in our proposed method to identify each layer of the Chinese character structure to obtain the components. Qualitative and quantitative calculations were used to test the segmentation effect of the proposed method. The experimental results demonstrate the accurate segmentation effectiveness of our method for adhering and overlapping components. In addition, these components could be retrieved accurately in the retrieval system, and the mean Average Precision of the top 30 retrieval results reached 95.7%. A better retrieval accuracy reflects a better segmentation effect from the side, which proves the effectiveness of the proposed method.

INDEX TERMS Chinese character, chinese character components segmentation, chinese character structure, faster RCNN.

I. INTRODUCTION
A component is a character unit composed of strokes with the function of combining Chinese characters, which can be obtained by splitting the structure of the Chinese characters [1]. As the name implies, Chinese character component segmentation is the division of a complete Chinese character into several parts, as shown in Figure 1. “相” is the left and right structure, which can be divided into the “木” component and “日” component. The structure of Chinese character components contains more information than strokes, is simpler than the whole Chinese character, and the number of Chinese character components is much less than the number of Chinese characters. Therefore, the accurate segmentation of Chinese character components is significant for the research related to the Chinese character image word stock [2], Chinese character recognition [3], and Chinese character font conversion [4]. The current research on Chinese character images focuses on the research of strokes and segmentation of Chinese characters. Strokes are the basic elements of Chinese characters and parts of the Chinese character component, which play an important role in the study of Chinese characters. Researchers have made great efforts in strokes, and have achieved certain results. These researches include stroke writing trajectory [5],

FIGURE 1. Structure schematic.
shape context [6], and so on. In addition, Liu et al. [7] proposed a structure-aware image resizing method for Chinese characters, and this method can resize a given Chinese character into arbitrary aspect ratios. Chinese character segmentation refers to the splitting of the whole Chinese character from the text lines. In the early days, Yao et al. [8] used a dynamic programming algorithm to obtain the optimal split path. Faced with irregularly arranged Chinese characters, Yang et al. [9] designed the enclosing circle according to the connected domain to achieve segmentation. In addition, Su et al. [10] and Zhang et al. [11] both proposed a segmentation method for handwritten Chinese characters. Although the existing character segmentation methods have reference significance for component segmentation, the effect of these methods on component segmentation is not significant in practical applications.

In the research of component segmentation, Liu et al. [4] proposed to use the method of matching the skeleton points of the target word and the skeleton points of Kaiti to complete the component segmentation on the target word, and the segmentation effect of this method is closely related to the extraction of target word skeleton points. Liu et al. [12] used a minimum enclosing box to construct a Chinese character block model to segment the components. However, additional steps or manual interaction are required for the glued components to ensure the accuracy of the segmentation results. Although the above methods have a certain segmentation effect, they require manual intervention and a large workload. Lin et al. [13] proposed a new Non-negative Matrix Factorization algorithm (NMF) to segment the components by decomposing the binary image. But, since the NMF algorithm is to bring all elements in the matrix as close to 0 or 1 as possible, there will be still some components left in the result. In summary, although some results have been achieved in the research of component segmentation, there are still some related problems that need to be solved. Due to the compact structure of the components contained in Chinese characters, the following problems exist in the current component segmentation. (1) The position of the same component in different Chinese characters is different, as shown in Figure 2(a). (2) The same component has different sizes and various aspect ratios in different Chinese characters, as shown in Figure 2(b). (3) In the process of components segmentation, an important reason why handwritten fonts are different from printed fonts is that there are adhesions and laps between the components of handwritten fonts, which would directly lead to the inability to accurately determine which strokes belong to which components. The above problems cannot be well solved by traditional methods such as manual feature extraction segmentation of components, and there are problems of poor segmentation effect, and a huge workload.

In recent years, neural networks [14], [15] have developed rapidly, and the research on Chinese characters has achieved more results with the rise of neural networks. For example, Zeng et al. [16] and Zhang et al. [17] applied neural networks to generate Chinese characters, and Zeng et al. [16] also used stroke encoding information to ensure the quality of the generated Chinese characters. Furthermore, neural networks can also implement Chinese character recognition [18], [19] and text conversion [20] in different scenarios. The process of segmenting components on a Chinese character image is essentially the process of finding a specific target on the image. So, the component segmentation problem can be converted into a target detection direction. Neural networks also have excellent performance in the field of target detection. RCNN [21], Fast RCNN [22], Faster RCNN [23], and Mask RCNN [24], which are representatives of two-stage target detection algorithms. It is worth noting, that Faster RCNN has better performance on multi-scale and small target problems. Salvador et al. [25] applied Faster RCNN to instance retrieval and achieved good retrieval results on different datasets. Due to the unique advantages of Faster RCNN, on the one hand, many researchers have improved it to improve the detection of small-scale objects [26], [27]. On the other hand, researchers have applied it to research fields such as remote sensing [28], face detection [29], and medicine [30]. In component segmentation, it is a problem that the area of some Chinese character components occupies a small proportion of a whole character image, such as in Figure 2 (b). In addition, there is also another problem that the position and size of the components are uncertain in the component segmentation. The characteristics of these problems are similar to those described above. Therefore, this paper proposes a Chinese character components segmentation method based on Faster RCNN. We design the Chinese character structure type and select the characters according to the proportion [31] and then we use Faster RCNN to recognize and segment the components to achieve component segmentation of canonical handwritten calligraphic characters. In addition, we use a retrieval system to check the segmentation effect of the components.

In Section 1, we discuss the importance of components, possible problems during segmentation, and some related research. The rest of our paper is organized as follows. In Section 2, we introduce the research method of this paper in detail. Then in Section 3, we describe the experimental data and results in detail. Finally, Section 4 provides a
II. COMPONENT SEGMENTATION METHOD BASED ON FASTER RCNN

Components are hierarchical, and they can be obtained by splitting the structure of Chinese characters. Aiming at the uncertainty of the position and size of the same component in different Chinese characters, and the problem of component deformation in incoherent handwritten calligraphic characters, this paper proposes a component segmentation method based on Faster RCNN. First, we annotate the Chinese characters images in the dataset labeled with the structure information layer by layer according to the structure of Chinese characters. There are thirteen kinds of structural labels, such as left_right, up_down, etc. Then use the labeled Chinese character images and the label information corresponding to each image as the training set to train Faster RCNN. Finally, the trained Faster RCNN network is used to recognize and segment the test set images to obtain the components. To test the segmentation effect, this study uses the map search system to test the component dataset and evaluate the component segmentation effect by the retrieval results of the system. The overall framework is shown in Figure 3.

In Figure 3, the image first goes through a series of convolutional layers, relu layers, and pooling layers to obtain feature maps. The feature map will be passed into RPN and ROI Pooling. In RPN, the feature map will generate anchors through Reshape and Softmax. These anchors will be judged whether they contain targets, and the bounding box regression parameters will be generated to adjust the anchor coordinates, and finally synthesized into a proposal. The proposal and feature map are then passed into ROI Pooling for final category judgment, and the target category and category probability contained in the picture is obtained.

A. SELECTION OF CHINESE CHARACTER IMAGES

In this paper, we refer to the thirteen Chinese character structures used in the literature [31] to label each layer of the Chinese character structure. The labels include left_right, up_down, up_right, up_left, left_down, up_three, down_three, left_three, surrounded, frame, left_center_right, up_center_down, and single_font. As shown in Table 1. Among the thirteen types of structures, the single_font and the frame have only one layer of structural information, and the first layers of the structure are labelable. The sample characters are shown in Table 1 as the “Example Word.”
TABLE 1. Structure schematic table.

| Serial Number | Label Name       | Example Word |
|---------------|------------------|--------------|
| 1             | left_right       | 煮油站        |
| 2             | up_down          | 章呈志        |
| 3             | up_right         | 氢氧金        |
| 4             | up_left          | 瘦庙届        |
| 5             | left_down        | 阳阖闯        |
| 6             | up_three         | 坐山画        |
| 7             | down_three       | 区臣图        |
| 8             | left_three       | 长承称        |
| 9             | surrounded       | 娜街街        |
| 10            | frame            | 九乙土        |
| 11            | left_center_right|              |
| 12            | up_center_down   |              |
| 13            | single_font      |              |

**FIGURE 5. Faster RCNN diagram.**

corresponding to single_font and frame. The rest of the structure types have more than two layers of annotation information. The specific annotation process is shown in Figure 4.

**B. FASTER RCNN**

Girshick proposed Faster RCNN in 2016. Based on Fast RCNN, he replaced the Selective Search algorithm and Edge Boxes algorithm in Fast RCNN with Region Proposal Network (RPN) and shared feature maps. These improvements reduce the number of repetitive computations, speed up training and prediction and improve operational efficiency. The schematic of Faster RCNN is shown in Figure 5.

We can roughly divide Faster RCNN into feature extraction layer, Region Proposal Network (RPN) layer, ROI pooling layer, and classifier regression layer. The specific steps are as follows:

1) In the feature extraction layer, the backbone network extracts the feature from the input image to generate feature maps. The backbone network usually consists of certain convolutional layers, pooling layers, etc. Currently, widely used backbone networks include VGG, ResNet, etc. The feature mapping maps generated in the feature extraction are shared in the subsequent region suggestion layers.

2) Pass the feature map into the RPN to generate anchors. The anchors are then subjected to binary classification and bounding box regression. The former is used to determine whether the anchor contains targets, and the latter is used to generate bounding box regression parameters to adjust anchor coordinates. These operations are the premise of proposals.

3) The proposals are mapped onto feature maps and fed into the ROI pooling layer for pooling. Proposals of different sizes are uniformly scaled to a fixed size, and then go through a fully connected layer to get the class results.

The RPN in Faster RCNN can accept input images of arbitrary size and the output is a set of proposals containing scores. In the shared feature map, k anchors of different sizes are generated after scanning each position of the feature map using an n × n sliding window. Each anchor has two operations, classification and bounding box regression. As shown in Figure 6. Thus, each anchor will get two classification scores and four bounding box regression parameters. The two classification scores are the foreground probability and the background probability of the anchor, which are obtained by comparing the overlap of the anchor with the ground truth. At the same time, the anchor gets a foreground or background label. The four bounding box regression parameters are the adjustments made by RPN to the predicted proposal to bring the proposal closer to the ground truth. In this paper, the sliding window is 3 × 3, and 9 anchors are collected at each pixel position.

The anchors use three different aspect ratios {1:1, 1:2, 2:1} and three different sizes {128^2, 256^2, 512^2} to ensure better prediction of objects of different sizes and positions. As shown in Figure 7. “a” in the “b”, which is a
Among the remaining anchors, we choose M components according to the Chinese character structure. RPN filters the resulting anchors, ignoring those that are out of bounds. For example, the anchor with an aspect ratio of 2:1 in Figure 7. Among the remaining anchors, we choose M randomly and use them to calculate the RPN loss. When the ratio is insufficient, negative samples are used to fill. A positive sample is defined as (a): The anchor with the largest IoU value among the anchors intersecting with the ground truth. (b): The anchor with an IoU value greater than 0.7 with any of the ground truth. IoU, which means intersection ratio, mainly measures the degree of overlap between the bounding box predicted by the model and the ground truth. The RPN loss function formula is as follows:

\[
L ((p_i^s, {t_i}) = \frac{1}{N_{cls}} \sum_i L_{cls} (p_i, p_i^s) + \frac{1}{N_{reg}} \sum_i p_i^s L_{reg} (t_i, t_i^s)
\]

The loss functions are divided into classification loss and bounding box regression loss. \(i\) is the i-th anchor in training. \(\lambda\) is the weighting factor. \(L_{cls}\) is implemented using Softmax_Cross_Entropy as shown in Equation (2), which indicates the error between the predicted and the ground truth of the i-th anchor. \(p_i\) is the probability that the i-th anchor is predicted to be the foreground. \(p_i^s\) represents the value of the ground truth label (0 for negative samples and 1 for positive samples).

\[
L_{cls} = - \left[ p_i^s \log (p_i) + (1 - p_i^s) \log (1 - p_i) \right]
\]

\(L_{reg}\) uses the \(smooth_{L_i}\) function as shown in Equation (3). \(t_i\) is the bounding box regression parameter predicted by the i-th anchor, and \(t_i^s\) is the bounding box regression parameter of the true value corresponding to the i-th anchor. The \(smooth_{L_i}\) function is shown in Equation (4), where \(x\) represents the input.

\[
L_{reg} (t_i, t_i^s) = \sum_i smooth_{L_i} (t_i, t_i^s)
\]

\[
smooth_{L_i} (x) = \begin{cases} 0.5x^2 & |x| < 1 \\ |x| & \text{other} \end{cases}
\]

C. SEARCH SYSTEM

Different from the traditional retrieval methods, this study uses the Milvus [32], which is a new vector database, to build a map search system. The system uses the neural network model VGG as the feature extraction part to obtain the features of the component images, which transforms unstructured data into high-dimensional vector data and preserves the image features better. The approximate nearest neighbor algorithm (ANN) is used to calculate the similarity and improve the retrieval speed. The system has a high retrieval performance with a retrieval time of about one second. The structure of the map search retrieval system is shown in Figure 8.

III. EXPERIMENTS AND RESULTS ANALYSIS

A. EXPERIMENTAL ENVIRONMENT AND DATA

1) EXPERIMENTAL ENVIRONMENT

To verify the effectiveness of the method proposed in this paper on the image segmentation task of Chinese characters, this paper uses the YOLOv3spp network as a comparative experiment. The proposed method and YOLOv3spp are both trained in the same experimental environment and using the same data set, and further comparisons have been achieved to prove the effectiveness of the method in this paper. The experimental environment of the two networks is the same: training using NVIDIA GeForce RTX2080 graphics processor (GPU), 8G of RAM, Intel(R) Xeon(R) Silver 4215 CPU @2.50GHz, and Ubuntu 16.04.7 LTS for the operating system, Python3.8, and Pytorch1.6. Use Resnet50 + FPN officially provided by Pytorch as the feature extraction network of Faster RCNN network (including pre-training weights), fifteen epochs, and fourteen categories (thirteen Chinese character structures and the background). Both batch size and dataloader workers are 2, and the learning rate is 0.00015. YOLOv3spp uses the Darknet53 network as the backbone. Batch size is four, thirty epochs, dataloader is 4, and the learning rate is 0.00001.

2) EXPERIMENTAL DATA

The single character dataset used in this paper is the single character images of calligrapher Yan Zhengqing (709 A.D.-784 A.D., Tang Dynasty) crawled online. There are a total of 5412 calligraphic Chinese character images. Among them, there are 4411 images in the test set and 1001 images in the training set. The training set is the single character images selected according to the proportion of the first layer structure of Chinese characters as counted in the literature [31], as shown in Table 2. Meanwhile, all the Chinese character structures were labeled for each single character image using the annotation tool according to the Pascal VOC.
dataset standard. Finally, the Chinese character component images are obtained by segmentation using the test set. The proposed method obtains 20900 Chinese character component images, and YOLOv3spp obtains 115119 images.

In Table 2, the first column shows the names of the thirteen structures designed in this study. The second column is the ratio of the thirteen structures. The third column is the number of Chinese characters selected according to the ratio in the second column, and the first structure of each character represents itself. The fourth column is the number of labels for the thirteen structures.

In this study, several predictions are obtained using the trained model in the test set and these predictions are the structure of a certain Chinese character. The next step is to segment the components of the Chinese character image based on these structures, as shown in Figure 9. All the obtained components are used as a component dataset for retrieval to evaluate the effectiveness of the proposed method.

B. ANALYSIS OF EXPERIMENTAL RESULTS

Use the test set to test the trained Faster RCNN. In this paper, when segmenting Chinese character images according to the prediction results, the prediction labels are saved as images separately. In this way, each prediction result of each picture can be known more clearly. All these component images are uploaded to the map search system. When testing the segmentation effect, this paper adopts qualitative and quantitative evaluation methods.

The same test set is used to test the comparative experiment YOLOv3spp. In the segmentation results, most of the targets contained in the obtained segmented images are incomplete, and there are also a large number of useless images with wrong predictions. Therefore, this paper adopts a qualitative method to compare with the proposed method. analyze.

1) EVALUATION INDICATORS

The proposed method is evaluated in terms of both visual effect and retrieval accuracy. The visual effect is evaluated by manual evaluation, and the retrieval accuracy is assessed using two evaluation metrics: Average Precision (AP) and mean Average Precision(mAP).

Equation (5) is the calculation formula for image retrieval accuracy, where \( k \) is the top \( k \) images returned when querying an image. \( precision_i \) is the precision of the top \( i \) images in the returned results. The \( i \) is the i-th image.

\[
AP_k = \frac{1}{k} \sum_{i=1}^{k} precision_i
\]  

Equation (6) is the precision calculation formula, where \( I_s \) is the actual number of similar images returned in top \( k \) images.

\[
precision_i = \frac{I_s}{i}
\]  

Equation (7) is the mean Average Precision calculation formula, where \( c \) is the number of queries. The mAP is higher to indicate better retrieval results. Meanwhile, the side
reflects the better segmentation effect.

\[ \text{mAP} = \frac{1}{c} \sum_{i=1}^{c} \text{AP}_c \]  

(7)

2) ANALYSIS OF VISUAL RESULTS

YOLO [33] is one of the representatives of the one-stage algorithm. The characteristic of the algorithm is that it does not need to generate a candidate region stage, and directly generates the class probability and position coordinate value of the object. In this way, the final detection result can be directly obtained after a single detection, so it has a faster detection speed. In terms of the time spent per image, the comparison group predicts an image of about 0.009 seconds on average, while the proposed method takes about 0.044 seconds. It can be seen that the speed is better than the proposed method. However, from the segmentation results, the speed improvement also has the loss of recognition accuracy and accuracy. Although YOLOv3 improves the recognition accuracy and accuracy at the expense of a small part of the running time. However, in this actual segmentation result, the recognition effect of the comparison group is still not as good as that of the proposed method.

In the segmentation results, the number of Chinese character components identified in the control group was abnormal in most cases, and it ranged from twenty to one hundred. Under normal circumstances, the number of parts contained in a Chinese character is between one and fifteen, and fewer Chinese characters will exceed this range. As shown in Figure 10, YOLOv3spp has identified a lot of repeated components and useless components. The segmentation results of several Chinese characters shown in the figure are all more than 30, which may be the reason for the abnormal number of components. The YOLO series is more suitable for real-time detection due to its fast detection speed but is not suitable for the scene of identifying Chinese characters.

For the proposed method, we randomly selected 100 whole words to view the segmentation results, and some of the results are in Figure 11. As can be seen from the figure, the method in this paper has a better segmentation effect when the adhesion lap is not serious, like “结”, “颜”, and others. In the case of the more serious lap of the components, the method can still be divided out of the corresponding components. Such as “艹” and “礻” in the words “艹” and “礻” left and right sides.

According to the literature [34], we select 20 components that compose a relatively large number of Chinese characters and then submit these component images sequentially to the retrieval system. The retrieval system will retrieve images similar to these components. Figure 12 shows the partial retrieval results of the image search system. In the figure, (a) is the image of the search component, and (b) to (f) are the first five search results. Figure 13 shows the first five search results for “艹” from (b) to (f) and their corresponding images of the original Chinese character. In Figure 12, the segmentation results are great for the same components on different Chinese characters. From Figure 13, the method in this paper is also effective in the case of uncertainty of the component’s position, and the segmentation effect is also good. For the characters, A to D, “艹” is written about the upper left. But in character E, its position is the lower left. And in character F, it’s in the middle.

3) QUANTITATIVE COMPARATIVE ANALYSIS

In this paper, we retrieve similar components in the component dataset by using a map search system to check the
effectiveness of component segmentation. The level of Average Precision can reflect the segmentation effect side-by-side. As shown in Figure 14, this paper calculates the Average Precision for the top 10, top 20, and top 30 retrieval results for the selected 20 components according to the experimental evaluation index. As can be seen from the figure, the Average Precision calculated for the retrieval results of all three cases is above 94%, and the Average Precision of the vast majority of components is above 90%. The mean Average Precision calculated for the retrieval results of each of the three cases, is above 95%, as shown in Table 3. From the above high retrieval accuracy, we can see that the method in this paper has a good performance in solving the difficult problem of component segmentation.

**IV. CONCLUSION**

Chinese character components are hierarchical and can be obtained by dividing Chinese characters layer by layer. However, due to the compact internal structure of Chinese characters, the location, size, and aspect ratio of the same component in different Chinese characters are different. There are problems with adhesion and overlap between components, which makes it hard to divide Chinese character components. In this paper, we propose a component segmentation method based on Faster RCNN. First, we design thirteen Chinese character structures and select single character images according to a specific scale. The segmentation of Chinese character components can be transformed into the process of finding the target of the image. We take advantage of Faster RCNN in the target detection to identify the Chinese character structures and implement the component segmentation. In testing the segmentation effect, this paper takes both qualitative and quantitative approaches for evaluation. First, we randomly select 100 Chinese characters and then segment these characters using the method in this paper, and judge the obtained segmentation results in terms of visual effects. Then we refer to “Specification of Common Modern Chinese Character Components and Component Names” to select 20 components that appear more frequently. Then we use Milvus to build a map search system to verify the accuracy of the method segmentation. The more accurate the segmentation, the higher the retrieval accuracy. The experimental results show that the average retrieval accuracy can reach more than 95%. It proves that the method of this paper can still achieve good results in the case of the complex structure of Chinese characters, different positions and sizes of components, and the existence of adhesions and laps. From the analysis of the results of the comparison group experiment, it can be seen that YOLOv3spp is not suitable for the scene of segmenting Chinese characters. It is faster than the method proposed in this paper. However, the recognition accuracy cannot achieve the desired effect, which is not as good as the method in this paper.

The method in this paper uses rectangular boxes as component boundaries in segmentation, so over-segmentation and under-segmentation may occur in the segmentation process. Next, we will consider how to process the inaccurate components to achieve higher quality and more accurate segmentation.

**REFERENCES**

[1] L. Jingnian, “A study on the structure of Chinese character,” Ph.D. thesis, School Humanities, Jilin Univ., Changchun, China, 2011.

[2] F. Warren and J. Lianwen, “Hierarchical Chinese character database based on radical reuse,” Comput. Appl., vol. 26, no. 3, pp. 714–716, 2006. [Online]. Available: https://kns.cnki.net/kcms/detail/detail.aspx?FileName=JSJY200603064&DbName=CSSQ2006

[3] X.-Y. Zhang, Y. Bengio, and C.-L. Liu, “Online and offline handwritten Chinese character recognition: A comprehensive study and new benchmark,” Pattern Recognit., vol. 61, pp. 348–360, Jan. 2017, doi: 10.1016/j.patcog.2016.08.005.

[4] L. Chengdong, L. Zhourui, T. Yingmin, and X. Jianguo, “Automatic system to generate high-quality Chinese font libraries based on component assembling,” Acta Scientiarum Naturalium Universitatis Pekinensis, vol. 54, no. 1, pp. 35–41, 2018, doi: 10.13209/j.0479-8023.2017.082.

[5] Z. Lian, B. Zhao, and X. Xiao, “Automatic generation of large-scale handwriting fonts via style learning,” SIGGRAPH Asia Tech. Briefs, pp. 1–4, Nov. 2016, Art. no. 12, doi: 10.1145/3005358.3005371.

[6] Z. Xu, Y. Liang, Q. Zhang, L. Dong, and E. Izquierdo, “Decomposition and matching: Towards efficient automatic Chinese character stroke extraction,” in Proc. Vis. Commun. Image Process. (VCIP), 2016, pp. 1–4, doi: 10.1109/VCIP.2016.7805484.

[7] C. Liu, Z. Lian, Y. Tang, and J. Xiao, “Structure-aware image resizing for Chinese characters,” in Proc. Int. Conf. Multimedia Model., Cham, Switzerland: Springer, 2017, pp. 379–390, doi: 10.1007/978-3-319-51811-4_31.
[8] Y. Zhengbin, D. Xiaoqing, and L. Changsong, “Segmentation of online handwritten characters based on stroke merging and dynamic programming,” J. Tsinghua Univ. Sci. Technol., vol. 44, no. 10, pp. 1417–1421, 2004, doi: 10.16511/j.cnki.qdxb.2004.10.031.

[9] Y. Xieliu, N. Xihui, and L. Wenfeng, “Character segmentation method for irregularly arranged text in Chinese,” J. Comput.-Aided Des. Comput. Graph., vol. 31, no. 9, pp. 1542–1548, 2019. [Online]. Available: https://kns.cnki.net/kcms/detail/detail.aspx?FileName=JSJF201909010&DbName=CJFQ2019&JID=JSJF201909010.

[10] T. H. Su, T.-W. Zhang, D.-J. Guan, and H.-J. Huang, “Off-line recognition of realistic Chinese handwriting using segmentation-free strategy,” Pattern Recognit., vol. 42, no. 1, pp. 167–182, 2009, doi: 10.1016/j.patcog.2008.05.012.

[11] C. Zhang and W. Li, “Recognizing handwritten Chinese daily and month words by combining a holistic method and a segmentation-based method,” Neural Comput. Appl., vol. 23, no. 6, pp. 1661–1668, 2013, doi: 10.1007/s00521-012-1125-9.

[12] L. Min, Z. Huanian, L. Xiaohui, and H. Jiajia, “Morphing technology for Chinese characters based on skeleton graph matching,” J. Beijing Univ. Aeronaut. Astronaut., vol. 41, no. 2, pp. 364–368, 2015, doi: 10.13700/j.hh.1001-9565.2014.0469.

[13] C. Y. Lin, L.-W. Kang, T.-Y. Huang, and M.-K. Chang, “A novel non-negative matrix factorization technique for decomposition of Chinese characters with application to secret sharing,” EURASIP J. Adv. Signal Process., vol. 2019, no. 1, pp. 1–8, 2019, doi: 10.1186/s13634-019-0636-2.

[14] A.-K. F. Lui, Y.-H. Chan, and M.-F. Leung, “Modelling of destinations for data-driven pedestrian trajectory prediction in public buildings,” in Proc. IEEE Int. Conf. Big Data, Dec. 2021, pp. 1709–1717, doi: 10.1109/BigData52589.2021.9671813.

[15] Y. Hu, G. Wen, M. Luo, D. Dai, W. Cao, Z. Yu, and W. Hall, “Inner-imaging networks: Put lenses into convolutional structure,” IEEE Trans. Cybern., vol. 52, no. 8, pp. 8547–8560, Aug. 2022, doi: 10.1109/TCYB.2020.3034605.

[16] J. Zeng, Q. Chen, Y. Liu, M. Wang, and Y. Yao, “StrokeGAN: Reducing mode collapse in Chinese font generation via stroke encoding,” 2020, arXiv:1208.06857.

[17] X.-Y. Zhang, F. Yin, Y.-M. Zhang, C.-L. Liu, and Y. Biegno, “Drawing and recognizing Chinese characters with recurrent neural network,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 4, pp. 849–862, Apr. 2018, doi: 10.1109/TPAMI.2017.2695539.

[18] Z. Cao, J. Lu, S. Cui, and C. Zhang, “Zero-shot handwritten Chinese character recognition with hierarchical decomposition embedding,” Pattern Recognit., vol. 107, Nov. 2020, Art no. 107488, doi: 10.1016/j.patcog.2020.107488.

[19] J. Zhang, J. Du, and L. Dai, “Radical analysis network for learning hierarchies of Chinese characters,” Pattern Recognit., vol. 103, Jul. 2020, no. 10,7305, doi: 10.1016/j.patcog.2020.107305.

[20] Q. Yang, J. Huang, and W. Lin, “SwapText: Image based texts transfer in scenes,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2020, pp. 14700–14709, doi: 10.1109/CVPR42600.2020.01471.

[21] R. Girshick, J. Donahue, and T. Richar, “Tierch hierarchies for accurate object detection and semantic segmentation,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2014, pp. 580–587, doi: 10.1109/cvpr.2014.81.

[22] R. Girshick, “Fast R-CNN,” in Proc. IEEE Int. Conf. Comput. Vis., Dec. 2015, pp. 1440–1448, doi: 10.1109/ICCV.2015.169.

[23] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” in Proc. Adv. Neural Inf. Process. Syst., vol. 28, 2015, pp. 1–9, doi: 10.1109/TPAMI.2016.257031.

[24] K. He, G. Glkixari, P. Dollar, and R. Girshick, “Mask R-CNN,” in Proc. IEEE Int. Conf. Comput. Vis., Feb. 2017, pp. 2961–2969, doi: 10.1109/TPAMI.2018.2844475.

[25] A. Salvador, X. Giró-i-Nieto, and F. Marqués, “Faster R-CNN features for instance search,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops. Jun. 2016, pp. 9–16, doi: 10.1109/CVPRW.2016.56.

[26] B. Wei, K. Hao, L. Gao, and E. Al, “Detecting textile micro-defects: A novel and efficient method based on visual gain mechanism,” Inf. Sci., vol. 541, pp. 60–74, Dec. 2020, doi: 10.1016/j.ins.2020.06.035.

[27] Y. Xiao, X. Wang, P. Zhang, F. Meng, and F. Shao, “Object detection based on faster R-CNN algorithm with skip pooling and fusion of contextual information,” Sensors, vol. 20, no. 19, p. 5490, doi: 10.3390/s20195490.