Stroke Extraction of Chinese Character Based on Deep Structure Deformable Image Registration

Meng Li[1], **Yahan Yu**[1,2], **Yi Yang**[1], **Guanghao Ren**[1], **Jian Wang**[1]*
[1]Institute of Automation, Chinese Academy of Sciences
[2]School of Artificial Intelligence, University of Chinese Academy of Sciences
{meng_li, yuyahan2020, yangyi, guanghao.ren, jian.wang}@ia.ac.cn

Abstract

Stroke extraction of Chinese characters plays an important role in the field of character recognition and generation. The most existing character stroke extraction methods focus on image morphological features. These methods usually lead to errors of cross strokes extraction and stroke matching due to rarely using stroke semantics and prior information. In this paper, we propose a deep learning-based character stroke extraction method that takes semantic features and prior information of strokes into consideration. This method consists of three parts: image registration-based stroke registration that establishes the rough registration of the reference strokes and the target as prior information; image semantic segmentation-based stroke segmentation that preliminarily separates target strokes into seven categories; and high-precision extraction of single strokes. In the stroke registration, we propose a structure deformable image registration network to achieve structure-deformable transformation while maintaining the stable morphology of single strokes for character images with complex structures. In order to verify the effectiveness of the method, we construct two datasets respectively for calligraphy characters and regular handwriting characters. The experimental results show that our method strongly outperforms the baselines. Code is available at https://github.com/MengLi-l1/StrokeExtraction.

Introduction

Stroke extraction of Chinese characters refers to extracting every single stroke of the characters based on the matching with the templates consisting of standard ordered strokes. Stroke extraction is important for certain research on Chinese characters. In the field of character recognition, the experiments in (Xiao et al. 2017; Zhang, Du, and Dai 2020) show that further disassembly of the stroke structure of characters can significantly improve the accuracy of character recognition. In other search fields such as evaluation of calligraphy the important part of traditional Chinese culture (Wang et al. 2016), and character generation (Liu and Lian 2021), the stroke extraction is also of great significance to them.

By analyzing the characteristics of Chinese characters, the difficulties of Chinese character stroke extraction mainly include the following three aspects. First, there are more than 7000 Chinese characters commonly used. Most of them have a complex structure. Second, the shapes of character strokes are simple and only have structural differences. These indistinguishable features make recognition directly of stroke hard for cross strokes within character. Third, the unfixed number of strokes in different Chinese characters makes it difficult to build a stroke extraction model.

Existing research on stroke extraction of Chinese characters, including some deep learning-based methods, mostly focus on image morphological features of strokes and radicals (Kim et al. 2018; Tseng and Chuang 1992; Wang, Jiang, and Liu 2022). Although these methods can realize remarkable results, their cores are only morphological analysis, which exists two drawbacks or constraints: (1) the prior information of the character stroke is rarely used, and (2) the semantic analysis of the stroke is lacking. Rarely using prior information and stroke semantics can lead to errors in the separation of cross strokes and the matching of strokes with the template in the stroke extraction of complex characters.

Inspired by the stroke extraction process of humans, we propose an efficient stroke extraction method of Chinese characters which not only separates strokes but also establishes the matching with the reference template. This method takes semantic features and prior information of strokes into consideration and mainly includes three steps: (1) stroke registration based on the Structure Deformable Image Registration Network (SDNet); (2) stroke segmentation based on the Image Semantic Segmentation Network (SegNet); (3) single stroke high-precision extraction based on the Single Stroke Extraction Network (ExtractNet).

For human cognition, the prior information of Chinese character images refers to the basic knowledge of the position and shape of the strokes. To obtain the prior information, we use the image registration to establish a rough mapping relationship between the reference character strokes and the target character. The transformed reference strokes based on the mapping relationship are used as prior information of the target stroke positions and shapes. The semantic features of the strokes need to be stable during the registration-based transformation for effective prior information. However, the existing image registration methods (Wang, Jiang, and Liu 2022; Haskins, Kruger, and Yan 2020; Sotiras, Davatzikos, and Paragios 2013) usually cause the

*Corresponding authors.

Chinese stroke to be severely distorted when characters have complex structures. To solve the problem, we propose SDNet using multiple linear mapping planes to replace the native single mapping surface. SDNet can maintain the stability of stroke semantic features while transforming deformably stroke structure. Our main contributions are summarized as follows.

- We propose a novel deep learning-based stroke extraction method, which takes semantic features and prior information of strokes into consideration more adequately and achieves significant improvement in the separation of cross strokes and matching of strokes with the template.
- We propose a structure deformable image registration method, which performs better in the registration of image structure.

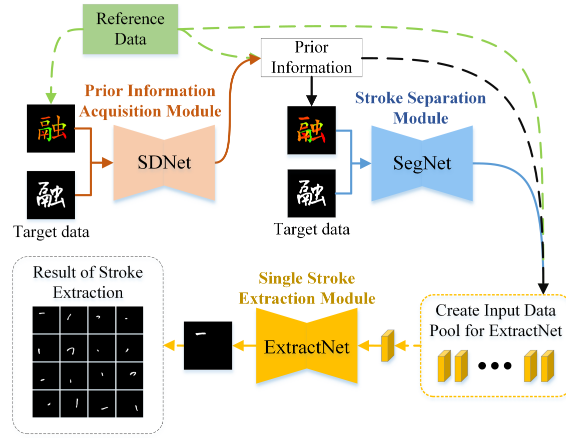

Figure 1: The pipeline of our method. The main modules are marked by different colors.

Related Work

Image Registration

Image registration is a process of establishing pixel-level correspondences between different images with matched image contents. In the past decades, image registration usually extracted and matched feature regions first, such as closed-boundary regions, edges, corners, and other shape features (Zitova and Flusser 2003). By evaluating the transformation model, like elastic model (Davatzikos 1997; Shen and Davatzikos 2002) and discrete model (Dalca et al. 2016; Glocker et al. 2008), the transformation relationship of these feature regions and entire image is established. Later, some researchers began to use deep learning techniques to enhance the feature extraction and matching, based on an iterative framework or reinforcement learning (Cheng, Zhang, and Zheng 2018; Haskins, Kruger, and Yan 2020; Liao et al. 2017). Recently, with the proposal of Spatial Transformer Networks (STN) (Jaderberg et al. 2015), the grid sample block with gradient backpropagation has facilitated the direct application of deep learning (Balakrishnan et al. 2019; Vos et al. 2017), which effectively promotes the improvement of the image registration. However, the existing deep learning-based image registration methods cannot maintain the local stability while the whole structure is freely transformed. It is not suitable for stroke registration of Chinese characters with complex structures.

Stroke Extraction for Chinese Character Image

For most existing stroke extraction methods of Chinese character, analyzing the ambiguous cross region is the core and primary task. By detecting the corners caused by the interlaced strokes, (Sun, Qian, and Xu 2014) disassembled the Kaiti characters into simple strokes for calligraphy robots. (He and Yan 2000) used chained line segments to represent the boundaries of characters and separated interlaced strokes by detecting whether these boundaries are regular. To further improve the accuracy of stroke extraction, some template-based methods are proposed (Cao and Tan 2000; Wang, Jiang, and Liu 2022). The methods use stroke structure matching to establish the correspondence between sub-strokes and template strokes, which is used to merge sub-strokes created by separating strokes with cross region. However, the template of the existing template-based methods is not used for previous steps of cross region detection and separation. In this case, the reference template information (prior information) is insufficiently used. In addition, character structure matching is mostly based on shape analysis and lacks the use of stroke semantic information.

Method

The proposed stroke extraction method, shown in Figure 1, mainly includes the following three modules. (1) The prior information acquisition module that establishes the registration between the reference stroke and the target through SDNet and uses the transformed reference strokes as prior information. (2) The stroke separation module that separates the target character preliminarily by SegNet with the guidance of prior information. (3) The single stroke extraction module that uses ExtractNet to extract every single stroke images of the target one by one according to the order of the reference strokes.

SDNet for Prior Information

Due to the complex stroke structure and various writing styles of Chinese characters, the position and shape of the stroke in the same character may be very different. It is the reason that it needs the high deformable registration model. However, highly distorted strokes can destroy their own shapes and reduce the validity of prior information, which needs the registration model to ensure that the shape of a single stroke is stable before and after transformation. For addressing the problem, we propose a local linear stroke spatial transformation method that constrains the transformation of a single stroke to be linear.

The SDNet uses UNet as the main frame of the registration network similar to (Balakrishnan et al. 2019). The main frame convolves down from size 256×256 to 8×8 and then convolves up to size 256 for output. To improve the analysis of Chinese character features, we add features of the Chinese character recognition of the input characters in the

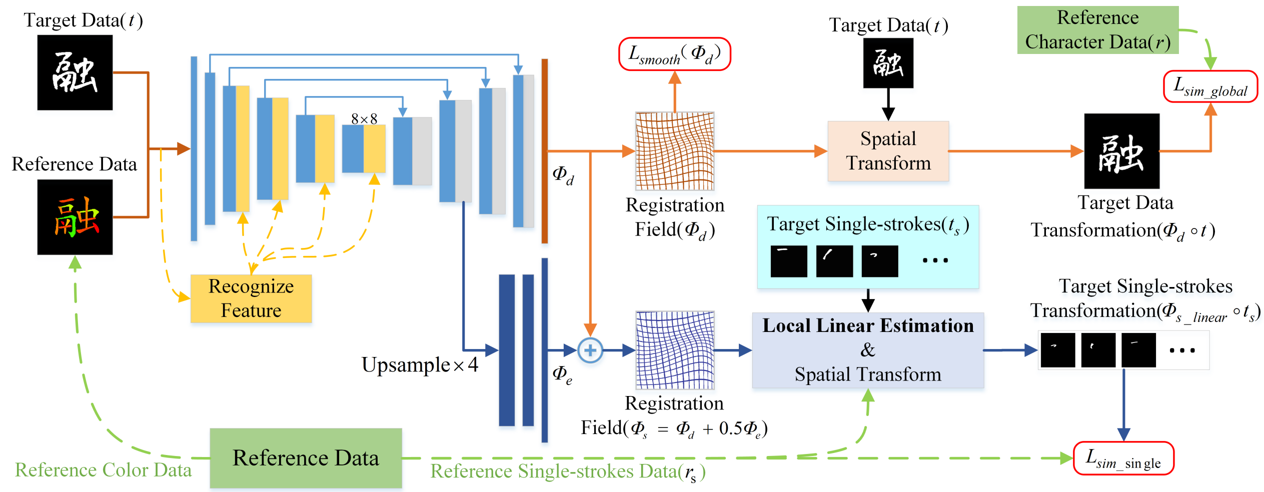

Figure 2: The structure of the SDNet.

last four stages of encoder in the UNet. The Chinese character recognition model is a simple convolution network like VGG (Simonyan and Zisserman 2014). The input of SDNet consists of two parts: target character image input as Target Data t and reference character image marked with different values for different stroke labels input as Reference Data. The output of the model is a prediction of the offset coordinate vector for each pixel, which can be easily constructed as a registration field. The structure of The SDNet is shown in Figure 2.

Based on the existing output registration field Φ_d, as shown in Figure 2, we add a branch in the up-convolution process. This branch upsamples the data at size 32 by a factor of 4 and then upsamples again after one convolution to obtain the registration field Φ_e with the same size as Φ_d. Φ_e is a fine-tuning of the original registration field, and its output weight is only 0.5 of Φ_d. The registration field used to calculate the linear transformation is represented as Φ_s ($\Phi_s = \Phi_d + 0.5\Phi_e$). Due to learning from a smaller size, Φ_e is biased towards th e prediction of the overall offset of the single stroke, which is suitable for the estimation of local linear transformation. During model training, both Φ_d and Φ_s are involved in the loss computation. Φ_s tends to learn the local registration ability of strokes under the constraint, which weakens the learning of global registration. Therefore, Φ_d is used to realize the global registration of character images to make up for the lack of Φ_s.

In actual training, the position and shape of the reference stroke are more stable. We use the reference stroke to mark the local region in Φ_s and calculate the linear transformation estimation of these regions. Therefore, during model training, what we actually learn is the transformation from target to reference. This operation can reduce the noise caused by errors in the linear estimation part, and improve the stability and efficiency of training. During inference, the transformation from reference to target can be obtained by calculating the inverse spatial transformation for every single stroke.

Linear Estimation of Single Stroke Spatial Transformation The existing linear fitting methods cannot be embedded in deep networks to achieve gradient backpropagation. Inspired by the Taylor series, we construct a linear estimation method that can be used in deep neural networks:

$$\begin{aligned} \Phi_{s\_linear} =& mean(\Phi_{s\_local}) + (\boldsymbol{X} - P_x)\times \\ mean(&\frac{\partial \Phi_{s\_local}}{\partial \boldsymbol{X}}) + (\boldsymbol{Y} - P_y) \times mean(\frac{\partial \Phi_{s\_local}}{\partial \boldsymbol{Y}}), \end{aligned} \tag{1}$$

where Φ_{s_local} represents the local region of Φ_s used for linear estimation. $\boldsymbol{X}$ and $\boldsymbol{Y}$ denote coordinate matrices. P_x and P_y denote the coordinates of the centroid of the local region, which can be calculated from the reference stroke. Φ_{s_linear} denotes the linear estimation result. Equation 1 is similar to the Least Squares Linear Fitting method while the slope is directly estimated as the average of the gradients to simplify the calculation. Due to the strong learning ability of deep learning, this simplification will not affect the final estimation result, but it can effectively reduce the computing workload.

Loss for Training The loss of SDNet consists of two parts: the global registration loss L_{global} and the single-stroke registration loss L_{single_linear}. L_{global} includes similarity loss L_{sim_global} and smoothing loss L_{smooth} of Φ_d. L_{single_linear} is the average of all single stroke similarity losses. For the operation of linear transformation estimation, we do not need to calculate the smoothing loss of Φ_s.

Traditional image registration methods usually use Normalization Mutual Information (NCC) to measure the similarity of two images. However, NCC is not suitable for the data with simple shape such as the stroke image. To solve this, we build a ContentNet that is trained to auto-encode stroke images with a simple Encoder-Decoder structure. The similarity loss of the two stroke images is defined as the Euclidean distance of the encoding results with l_2-normalization:

$$S_c(a, b) = Dis[norm_{l_2}(E(a)), norm_{l_2}(E(b))], \tag{2}$$

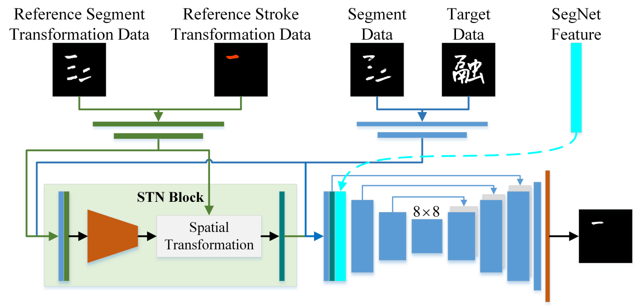

Figure 3: The structure of the ExtractNet.

Name	Description
Target Data	Target character image
Reference Stroke Transformation Data	The corresponding transformed single reference stroke image in prior information
Segment Data	The segmentation result of the category to which the current stroke belongs in the SegNet
Reference Segment Transformation Data	Divide the transformation reference strokes into seven categories similar to the SegNet. And select the category data the current stroke belongs to
SegNet Feature	Feature data at 64 size of up-convolution in the SegNet

Table 1: Detailed descriptions of inputs in the ExtractNet.

where Dis denotes the calculation of Euclidean distance. E denotes the access function of the encoding result of ContentNet.

Final loss in train process is defined as:

$$L_{sum}(t, r, t_s, r_s, \Phi_d, \Phi_s) = \lambda L_{single\_linear}(t_s, r_s, \Phi_s) + L_{sim\_global}(r, t, \Phi_d) + \gamma L_{smooth}(\Phi_d), \tag{3}$$

$$L_{single\_linear}(t_s, r_s, \Phi_s) = \frac{1}{stroke_{num}} \sum_{i=1}^{stroke_{num}} S_c(r_s^i, \Phi_{s\_linear}^i \circ t_s^i), \tag{4}$$

$$L_{sim\_global}(r, t, \Phi_d) = S_c(r, \Phi_d \circ t), \tag{5}$$

$$L_{smooth}(\Phi_d) = mean(\|\frac{\partial \Phi_d}{\partial \boldsymbol{X}}\|^2 + \|\frac{\partial \Phi_d}{\partial \boldsymbol{Y}}\|^2), \tag{6}$$

where $\Phi_{s\_linear}^i$ is the linear transformation estimation result corresponding to the reference single stroke r_s^i and the target single stroke t_s^i. Considering that the global registration result will have a greater impact on Φ_s, we apply a larger weight to Φ_d, especially to L_{smooth} , and set λ and γ to be 0.5 and 5, respectively, in order to ensure a stable and better registration result of Φ_d.

SegNet for Separating Strokes Roughly

There are 32 basic strokes for Chinese characters. However, there is a high similarity between these basic strokes, and the number of strokes is seriously unbalanced. Therefore, we divide the 32 basic strokes into 7 categories artificially based on the following three rules: (1) The number of strokes used in common Chinese characters within the category is as balanced as possible. (2) The similarity of stroke shape is as high as possible within the category and as low as possible between the category. (3) The probability of stroke crossing within the category is as low as possible. We use the network architecture adapted from the Deeplabv3 model (Chen et al. 2017) as the main frame of the SegNet to segment strokes guided by prior information. Considering the cross stroke, we construct the SegNet as a multi-label model. The loss of SegNet is the average of the binary cross-entropy of output and label. The input of SegNet consists of two parts: Target Data and Prior Data. Prior Data is composed of reference single strokes that are linearly transformed by SDNet. Strokes in Prior Data with different categories are marked with different values. In the training process, in order to improve the generalization of SegNet, we apply a random position offset within maximal 5 pixels to every single stroke in Prior Data.

ExtractNet for Single Stroke Extraction

Input Data As shown in Figure 3, ExtractNet has five inputs to provide sufficient prior information and stroke semantic information. Table 1 shows the details of these inputs. For ExtractNet, Segment Data and Reference Stroke Transformation Data provide the major information required for stroke extraction. Considering the possible segmentation errors of SegNet, we add SegNet Feature and Target Data to supplement the information not included in the Segment Data. The Reference Stroke Transformation Data can only roughly mark the location and shape of the target stroke. For the high-precision stroke extraction work, we add Reference Segment Transformation Data to provide relative positional relationship information with Reference Stroke Transformation Data. In this way, through spatial transformation of the STN Block, the transformed reference information can be further registered to the target data, which can provide more accurate prior information of stroke position and shape.

Furthermore, within a Chinese character, the size of different strokes may vary greatly, which will weaken the learning of small-sized strokes. Therefore, we adaptively scale and crop images of input data and labels to eliminate the size difference between strokes.

Structure and Loss The structure of ExtractNet is shown in Figure 3, which mainly includes two parts: STN Block and the simple convolution network used to extract strokes. In the beginning, we quickly compress the input to 1/4 of the original size by two layers of convolution. The STN Block is used to further register the reference information to the target. The output is a single-channel stroke image. We use binary cross-entropy to calculate the loss between the output and label.

Experiments

Datasets and Reference Data

To evaluate our method, we construct two stroke extraction datasets for calligraphy and handwriting, which basically cover the main application fields of the stroke extraction.

- **The Chinese Calligraphy Character Stroke Extraction Dataset (CCSEDB):** CCSEDB has a total of 5000 character data, consisting of calligraphy characters, printed calligraphy characters, and calligraphy characters. We carefully select characters of CCSEDB to maintain a balance between the number of strokes and stroke structure. Each record of data in CCSEDB contains a target image and some singlestroke label images of the target image arranged in reference stroke order.
- **The Regular Handwriting Character Stroke Extraction Dataset (RHSEDB):** We construct RHSEDB referring to (Wang, Jiang, and Liu 2022) based on the online handwriting dataset CASIA-OLHWDB (Liu et al. 2011), which contains 28,080 pieces of handwriting data written by 17 writers in total. The format of each piece of data in RHSEDB is the same as CCSEDB, while the images of writing track of the stroke is normalized to a width of 6 pixels (the size of the stroke image is 256 pixels).
- **Reference data:** We construct reference data for CCSEDB and RHSEDB, respectively. For CCSEDB, due to the large stroke area, we use character images and the corresponding single stroke images of the Kaiti font as reference data. For RHSEDB, due to the thin stroke width, we use the skeleton images and the corresponding single stroke skeleton images of the Kaiti font as reference data, which are also normalized to a width of 6 pixels.

Implementation Detail

In the experiments, for both CCSEDB and RHSEDB, we use 90% of the data for training and 10% for testing. The images of training data of SDNet, SegNet, and ExtractNet have resolution of 256. The training data are binarized, except for a few that need to be marked with three-channel value. The three models are trained progressively with a batch size of 8 and for 40, 10, and 20 epochs respectively. Their learning rates are initialized to 0.0001 and decrease by a factor of 0.5 every 10, 2, and5 epochs respectively.

Stroke Registration Results

In this task, we build experiments on two representative image registration methods: TpsStn (Jaderberg et al. 2015) and VoxelMorph (Balakrishnan et al. 2019). TpsStn uses the idea of STN to realize the TPS registration of two images. Due to the fewer control points, this method well maintains local stability but lacks sufficient deformability. In order to fully verify the registration effect of TpsStn, we evaluate TpsStn with different numbers of control points of 4 and 8 respectively. VoxelMorph is a typical image registration method that can theoretically achieve the maximum deformable transformation for predicting the offset of each pixel. In order to quantify the effect of the prior information constructed by the registration results. For the stroke position, we estimate the qualitative result of position with centroid pixel distance $mDis$ of the single stroke. For the stroke shape that is mainly reflected in the size, we estimate the qualitative result of stroke shape with the IOU $mBIou$ of the bounding box of the single stroke.

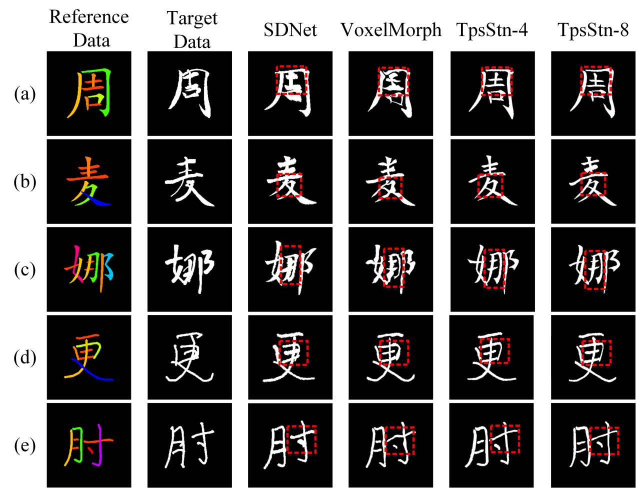

Figure 4: Comparison of stroke registration results for our SDNet and baseline models.

	CCSEDB		RHSEDB	
	$mDis$	$mBIou$	$mDis$	$mBIou$
Reference Data	11.7	0.365	14.763	0.277
SDNet	**4.469**	**0.637**	**5.553**	**0.614**
VoxelMorph	5.352	0.609	7.531	0.527
TpsStn-4	6.616	0.533	7.86	0.479
TpsStn-8	6.271	0.551	7.564	0.487

Table 2: Quantitative results of our SDNet and baseline methods.

$$mDis = \frac{1}{n}\sum_{i=0}^{n} Dis(centroid(rt_s^i), centroid(t_s^i)), \quad (7)$$

$$mBIou = \frac{1}{n}\sum_{i=0}^{n} IOU(box(rt_s^i), box(t_s^i)), \quad (8)$$

In Equation 7 and Equation 8, t_s^i denotes the single stroke of the target character. rt_s^i denotes the single transformation reference stroke in prior information by SDNet. Dis refers to the Euclidean distance. A smaller mDis and a larger mBIou indicate the higher accurate prior information.

As we can see from Table 2, SDNet performs much better than other baseline methods in qualitative results of prior information. This means that our method provides more precise prior information to the target character. As shown in Figure 4, the SDNet has the best results, which is manifested in higher structural deformability and more stable single-stroke morphology. Especially in Figure 4 (d) and (e), this advantage is more prominent when the target structure is quite different from the reference structure. We believe that this is mainly due to the registration branch Φ_e and linear estimation of single stroke spatial transformation. Local linear estimation can construct different transformations

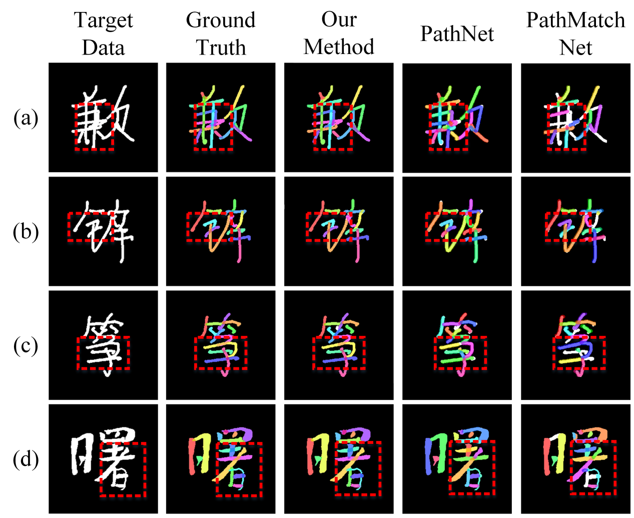

Figure 5: Comparison of stroke extraction results of our method and other baseline methods. Different strokes in the results are marked with different colors. The strokes except the PathNet with the same color mean they match the same single reference stroke.

	CCSEDB		RHSEDB	
	$mIOU_m$	$mIOU_{um}$	$mIOU_m$	$mIOU_{um}$
Our Method	**0.924**	**0.926**	**0.917**	**0.93**
PathNet	\	0.84	\	0.789
PathMatchNet	0.692	0.842	0.654	0.815

Table 3: $mIOU_m$ and $mIOU_{um}$ of our method and baseline methods.

for every single reference stroke even for cross stroke while providing linear constraint. This is the reason for the structure deformable which is enhanced greatly by the registration branch Φ_e. However, TpsStn and VoxelMorph need to balance the deformation and the smoothness because each of them has only one registration field.

Stroke Extraction Results

We find that the morphological analysis method based on deep learning has a great improvement over traditional methods for stroke extraction. Therefore we only compare the recent best deep learning-based stroke extraction methods (Wang, Jiang, and Liu 2022) named Path-MatchNet and (Kim et al. 2018) named PathNet. PathNet separates strokes by predicting the probability that two pixels belong to the same stroke using a deep learning-based image semantic segmentation method. Path-MatchNet adds stroke matching on the basis of PathNet to further improve the accuracy of stroke extraction. Referring to (Kim et al. 2018), we construct two evaluation methods that employ the same evaluation strategy but differ slightly based on whether it needs to match the reference stroke. The evaluation considering matching is defined as :

$$mIOU_m = \sum_{i=0}^{n} IOU(rt_s^i, t_s^i), \tag{9}$$

The evaluation without considering matching is defined as:

$$mIOU_{um} = \sum_{i=0}^{n} IOU(rt_s^i, maxCross(rt_s^i, t_s)), \tag{10}$$

In Equation 9 and Equation 10, t_s^i denotes the single stroke. t_s denotes all of the single strokes of the target character. rt_s^i denotes the single transformation reference stroke in prior information by SDNet. $maxCross$ refers to obtaining the single stroke image of target character with the largest intersection area with rt_s^i in t_s.

As shown in Table 3, our method performs much better in $mIOU_{um}$ and $mIOU_m$ than baseline methods. As shown in Figure 5, our stroke extraction results have higher precision and are almost 100% accurate in matching with reference strokes, which benefit from the use of prior information and stroke semantic information. As shown in Figure 5 (a) and (d), PathNet and PathMatchNet have lower stroke extraction precision in the cross stroke area. That is because they only use deep learning for morphological stroke trajectory analysis, and lack analysis of the semantics. In the part of stroke matching, simple position and morphological feature similarity calculation in PathMatchNet cannot guarantee the accuracy of matching, especially in Chinese characters with a large number of strokes, as shown in Figure 5 (a) and (b).

Ablation Study

To evaluate the effect of prior information and stroke semantic information, we design ablation experiments for SegNet and Extraction.

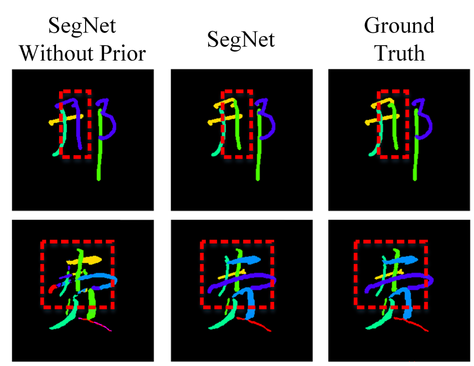

Figure 6: The effect of prior information on semantic segmentation results in the SegNet.

As shown in Table 4 and Figure 6, the prior information has a significant effect on the SegNet because of the high similarity between strokes. Compared to the SegNet, the prior information has a greater effect on the ExtractNet, as shown in Table 5. This is because that the prior information provides major information for analyzing the position

	$mIOU$ of Semantic Segment	
	CCSEDB	RHSEDB
SegNet	**0.917**	**0.908**
SegNet Without Prior	0.735	0.768

Table 4: Effect of prior information in the SegNet.

and shape of the current single stroke in the ExtractNet. Semantic information has a few effect on the ExtractNet. This is because the Target Data can supplement more information when the Segment Data lacks. Using semantic information can further improves the precision of ambiguous strokes, as shown in Figure 7.

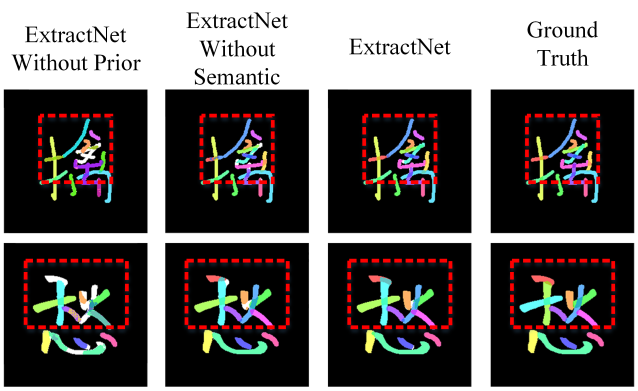

Figure 7: Effect of prior and semantic information on extraction result in the ExtractNet.

	$mIOU_m$	
	CCSEDB	RHSEDB
ExtractNet	**0.924**	**0.917**
ExtractNet Without prior	0.567	0.556
ExtractNet Without Semantic	0.914	0.890

Table 5: Effect of prior and semantic information in the ExtractNet.

Limitations

Finally, as shown in Figure 8, we show typical errors in the results of our method.

- **Scribbled strokes:** Scribbled strokes often lead to densely intersected strokes and indistinguishable stroke morphology, which usually lead to failure of stroke segmentation and single-stroke extraction, such as (a) in Figure 8.
- **Excessive difference in local stroke structure:** As shown in Figure 8 (b), excessive local structure differences between reference and target usually lead to large registration errors, which makes the prior information of these local strokes unusable.

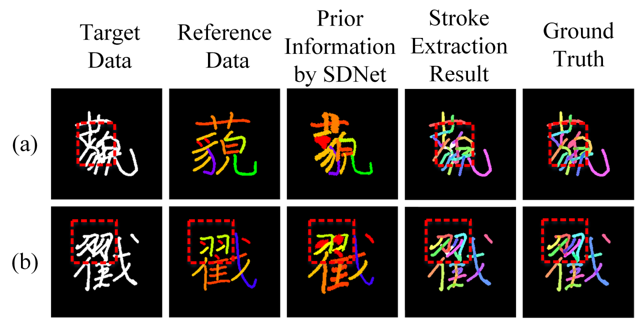

Figure 8: Typical error examples of our method.

Conclusion

In this paper, we propose an efficient stroke extraction model of Chinese characters which takes semantic features and prior information of strokes into consideration efficiently. In our method, SDNet establishes the registration relationship between reference strokes and target strokes to provide the prior information of stroke position and shape to the target character. The prior information can guide SegNet to segment roughly the strokes of the target character. With the prior information and segmentation results, every stroke is extracted high precision through ExtractNet. Furthermore, to solve the registration problem of characters with complex stroke structures, we propose a new method for Chinese character image registration called SDNet. The use of the local linear stroke spatial transformation method in SDNet ensures deformability of stroke structure while maintaining the stability of single stroke shape in transformation.

Experiments show that our method performs better in stroke extraction than baseline methods and can be used for a wide range of characters including calligraphic characters and handwritten characters. And, SDNet performs better in the registration of image structure.

We believe that prior information and stroke semantic information are the keys to the stroke extraction of Chinese characters. In our future work, we will pay more attention to studying new image registration methods based on SDNet and stroke segmentation methods for irregular characters.

References

Balakrishnan, G.; Zhao, A.; Sabuncu, M. R.; Guttag, J.; and Dalca, A. V. 2019. VoxelMorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging*, 38(8): 1788–1800.

Cao, R.; and Tan, C. L. 2000. A model of stroke extraction from chinese character images. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, volume 4, 368–371. IEEE.

Chen, L.-C.; Papandreou, G.; Schroff, F.; and Adam, H. 2017. Rethinking atrous convolution for semantic image segmentation. arXiv:1706.05587.

Cheng, X.; Zhang, L.; and Zheng, Y. 2018. Deep similarity learning for multimodal medical images. *Computer Methods* *in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 6(3): 248–252.

Dalca, A. V.; Bobu, A.; Rost, N. S.; and Golland, P. 2016. Patch-based discrete registration of clinical brain images. In *International Workshop on Patch-based Techniques in Medical Imaging*, 60–67. Springer.

Davatzikos, C. 1997. Spatial transformation and registration of brain images using elastically deformable models. *Computer Vision and Image Understanding*, 66(2): 207–222.

Glocker, B.; Komodakis, N.; Tziritas, G.; Navab, N.; and Paragios, N. 2008. Dense image registration through MRFs and efficient linear programming. *Medical image analysis*, 12(6): 731–741.

Haskins, G.; Kruger, U.; and Yan, P. 2020. Deep learning in medical image registration: a survey. *Machine Vision and Applications*, 31(1): 1–18.

He, R.; and Yan, H. 2000. Stroke extraction as pre-processing step to improve thinning results of Chinese characters. *Pattern Recognition Letters*, 21(9): 817–825.

Jaderberg, M.; Simonyan, K.; Zisserman, A.; et al. 2015. Spatial transformer networks. *Advances in neural information processing systems*, 28.

Kim, B.; Wang, O.; Öztireli, A. C.; and Gross, M. 2018. Semantic segmentation for line drawing vectorization using neural networks. In *Computer Graphics Forum*, volume 37, 329–338. Wiley Online Library.

Liao, R.; Miao, S.; de Tournemire, P.; Grbic, S.; Kamen, A.; Mansi, T.; and Comaniciu, D. 2017. An artificial agent for robust image registration. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31.

Liu, C.-L.; Yin, F.; Wang, D.-H.; and Wang, Q.-F. 2011. CASIA online and offline Chinese handwriting databases. In *2011 International Conference on Document Analysis and Recognition*, 37–41. IEEE.

Liu, Y.; and Lian, Z. 2021. FontRL: Chinese Font Synthesis via Deep Reinforcement Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 2198–2206.

Shen, D.; and Davatzikos, C. 2002. HAMMER: hierarchical attribute matching mechanism for elastic registration. *IEEE transactions on medical imaging*, 21(11): 1421–1439.

Simonyan, K.; and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556.

Sotiras, A.; Davatzikos, C.; and Paragios, N. 2013. Deformable medical image registration: A survey. *IEEE transactions on medical imaging*, 32(7): 1153–1190.

Sun, Y.; Qian, H.; and Xu, Y. 2014. A geometric approach to stroke extraction for the Chinese calligraphy robot. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, 3207–3212. IEEE.

Tseng, L.-Y.; and Chuang, C.-T. 1992. An efficient knowledge-based stroke extraction method for multi-font Chinese characters. *Pattern recognition*, 25(12): 1445–1458.

Vos, B. D. d.; Berendsen, F. F.; Viergever, M. A.; Staring, M.; and Išgum, I. 2017. End-to-end unsupervised deformable image registration with a convolutional neural network. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, 204–212. Springer.

Wang, M.; Fu, Q.; Wang, X.; Wu, Z.; and Zhou, M. 2016. Evaluation of Chinese calligraphy by using DBSC vectorization and ICP algorithm. *Mathematical Problems in Engineering*, 2016.

Wang, T.-Q.; Jiang, X.; and Liu, C.-L. 2022. Query Pixel Guided Stroke Extraction with Model-Based Matching for Offline Handwritten Chinese Characters. *Pattern Recognition*, 123: 108416.

Xiao, X.; Jin, L.; Yang, Y.; Yang, W.; Sun, J.; and Chang, T. 2017. Building fast and compact convolutional neural networks for offline handwritten Chinese character recognition. *Pattern Recognition*, 72: 72–81.

Zhang, J.; Du, J.; and Dai, L. 2020. Radical analysis network for learning hierarchies of Chinese characters. *Pattern Recognition*, 103: 107305.

Zitova, B.; and Flusser, J. 2003. Image registration methods: a survey. *Image and vision computing*, 21(11): 977–1000.