A similarity-based method for evaluating single-stroke regions

Jiayun Yu[a], Jinghong Wang*[b], Zhanyang Xu[b], Dingyu Li[b], Wei Lin[c]
[a]School of Educational Science, Nanjing Normal University, Nanjing 210023, Jiangsu, China;
[b]School of Software, Nanjing University of Information Science &Technology, Nanjing 210044, Jiangsu, China; [c]Jiangsu Shao Er Chun Internet Education Technology Co., Ltd., Nanjing Technology R&D Center, Nanjing 211899, Jiangsu, China

ABSTRACT

With the advancement of calligraphy education, an array of intelligent handwriting evaluation systems has emerged to support teaching. However, existing handwriting evaluation systems fall short in analyzing and evaluating individual strokes of Chinese characters. To address this issue, we introduce SSRSE, a Single-Stroke Regional Segmentation Evaluation method for hard-pen regular script. Initially, we gather handwritten character images featuring 22 commonly used strokes in hard-pen regular script from primary and secondary schools. These strokes are then categorized into four regions based on their writing patterns, forming a stroke regional segmentation dataset through adjustments in size and region marking. We propose a single-stroke regional segmentation method based on an inverted residual U-net. This involves incorporating an inverse residual structure into the U-network and employing a loss function based on Intersection over Union(IoU) to tackle the imbalance in sample categories. Subsequently, Hu moments are computed for each region post-segmentation of both the copy strokes and template strokes. After mathematical conversion, the extent of image variance is measured, and stroke similarity levels are determined based on weighted assignments. Experimental findings demonstrate the superiority of our segmentation method over traditional approaches in terms of average pixel accuracy. Moreover, the region similarity calculation method yields results akin to manual evaluation, showcasing high feasibility.

Keywords: Stroke evaluation, stroke segmentation, inverted residual block, Hu moments

1. INTRODUCTION

Chinese characters, as the main official script of China throughout the ages, are one of the oldest scripts in the world, with a history of more than 6,000 years. Chinese calligraphy, as a unique traditional art of Chinese characters, is expressed in various forms and styles. Among the many fonts, there are a number of sub-divisions, among which the Regular Script, which is a typical square font, is the first choice of beginners because of its rigorous rules and verticality. Chinese calligraphy is mainly written with brushes, and with the development of society, pens have gradually become the main writing tool. However, with economic globalization and the advancement of information technology, keyboards, touch screens and other electronic tools have gradually replaced the function of the pen in daily life, and many students and schools no longer pay attention to the correctness and beauty of writing, which has led to the frequent occurrence of mistakes and forgetfulness of words in the lives of many people, which has greatly hindered the progress and development of calligraphy.

In recent years, with the focus of the state and the release of relevant documents, calligraphy education has gradually been integrated into the compulsory education stage, which has also made the majority of educators pay attention to calligraphy education and promoted the development of calligraphy culture. However, in the process of calligraphy education program, primary and secondary school teachers are inexperienced and subjective. Also, the teaching methods are not universal, and it is difficult to give targeted advice and guidance to each student, the teaching effect is inevitably unsatisfactory[1].

With the development of computer technology, the use of computer-assisted teaching gradually came into people's view. In the study of calligraphy, there are also attempts to use computer-assisted calligraphy education to carry out work[2,3]. However, in the early days, these researches only stayed in the teaching stage, aiming at assisting teachers to carry out

*1984648801@qq.com; phone 13276688717

International Conference on Optics, Electronics, and Communication Engineering (OECE 2024),
edited by Yang Yue, Proc. of SPIE Vol. 13395, 133954G · © 2024 SPIE · 0277-786X · Published
under a Creative Commons Attribution CC-BY 3.0 License · doi: 10.1117/12.3049017

teaching tasks and reducing the burden of teachers, and not evaluating the quality of the students' written. In the context of vigorously developing traditional Chinese culture, simple teaching aids are increasingly unable to meet the requirements of the quality of calligraphy education, and calligraphy educators have begun to explore smarter teaching aids to improve the level of calligraphy education. In calligraphy education, the most important thing is to give effective evaluation and guidance to calligraphy learners. So, scholars focus more on "learning" rather than "teaching", and have conducted various studies on the analysis and evaluation of handwriting quality.

However, in the existing handwriting evaluation systems or methods, scholars' evaluation targets mainly focus on the whole character, and rarely analyze the strokes in detail. For beginners in calligraphy, the practice of strokes, as the basic component unit of Chinese characters, is particularly important, and it is difficult to carry out the subsequent study of calligraphy if one cannot grasp the structure and layout of the strokes of Chinese characters. Therefore, this paper expects to find an evaluation method for the handwritten strokes of hard-pen regular script to satisfy the needs of the hard-pen calligraphy study.

2. RELATED WORK

In traditional graphical methods, the stroke features of handwriting need to be defined and extracted by the researcher and then calculated to get quality results, which has the advantage that quality analysis can be realized in an offline environment, and the features can be defined and rules can be formulated according to the research needs. However, the definition of stroke features under the traditional method is often based on the known characteristics of stroke structure and direction, and some potential information has not been explored.

Zhuang et al.[4] proposed a method to extract the eigenvalues of thirty-six types of strokes in combination with the nine-square grid to match the similarity with the template to judge the neatness of Chinese characters' writing. Although the method has carried out an exhaustive study on the categorization of writing characteristics of strokes, the definition of the features is simple, and the definition of the eigenvalues is ambiguous, and thus the evaluation result is not ideal. Wang et al.[5] proposed a fuzzy analysis method of the Chinese characters writing quality, using the key point of the stroke as the stroke feature, establishing a fuzzy feature vector and a fuzzy affiliation function, defining and calculating the fuzzy feature proximity of the Chinese character's strokes in order to rate the quality. The method is based on the key point of the stroke, but the extracted features are relatively simple, and it cannot give a more detailed evaluation as well as a targeted guidance. Chen[6] divided Chinese strokes into straight and curved strokes and discussed separately. Slopes, angles and curvatures are calculated to evaluate the quality of the strokes. Although this method can judge the specific gaps in various aspects, the number of items concerned is too small, and it cannot really express all the characteristics of the strokes, which is a limitation.

Compared with traditional methods, machine learning methods are more intelligent and efficient, which are characterized by the fact that researchers no longer need to manually define and extract image features, but can allow computers to automatically extract multi-dimensional features and high-level semantic features that are difficult for humans to capture in an image, and to complete the task based on self-learning and evolution. According to the basic calligraphy rules, Geng et al.[7] took the abstract features of complexity, fullness and morphological structure as the distance relationship between pixel change and circumscribed rectangle, calculated the standard error of each attribute by using BP neural network, and drew a conclusion. Sun et al.[8] proposed a method using neural network to evaluate the handwriting aesthetic, and proposed 22 global shape features as aesthetic features according to the classical rules of calligraphy, invited subjects to evaluate the aesthetic quality of calligraphic works of different qualities and formed a database of aesthetic evaluation of Chinese calligraphy, after which used the neural network to evaluate the calligraphic works, and obtained results comparable to the manual evaluation. Jiang et al.[9] proposed indicators such as direction of stroke, two-dimensional shape, length ratio, orientation, and combination relationship, and determined the weight relationship after comparative experiments, which improved the consistency between correctness evaluation and manual evaluation. Yan et al.[10] proposed a Gabor feature extraction method for evaluating the writing quality of Chinese characters, using the Gabor function to extract the texture features of Chinese characters, and then using SVM to classify their images, with good performance in judging the clarity of the strokes and the consistency of the style. Xiao et al.[11] summarized the existing handwriting evaluation methods, and concluded that the current Chinese character quality evaluation is mainly based on four categories of methods: rule-based methods, feature similarity calculation, fuzzy matrix, and machine learning; rule-based methods are easy to implement, but they need to specify the rules for each feature and need to be constantly updated, which is a large amount of workload; fuzzy matrix methods can solve the problem of the fuzzy concepts in handwritten Chinese characters, but the acquisition of data relies on on-line equipment and lacks detailed evaluation; the feature similarity calculation method can give specific evaluation for a certain aspect, but its accuracy depends on the quality of the extracted features, so it requires higher quality of feature definition and feature extraction; the method based on machine learning has the advantages of fast speed and high accuracy, but most of this kind of method focus on the global features, and it is easy to ignore the local features .

Combining the above related researches, this paper prepares to combine machine learning and similarity calculation and proposes SSRSE, a single-stroke regional similarity evaluation method for the handwritten characters in hard-pen regular script, in order to meet the needs of current calligraphy beginners for calligraphy learning, and also to provide evaluation norms and methods for the related intelligent calligraphy evaluation system, which has a good application prospect.

In the study of Chinese character strokes, most researchers regard a single stroke as a whole, mine and calculate the characteristics of this whole to get useful concrete information on which to base their judgement. However, in the process of writing Chinese strokes, whether it is traditional brush writing or hard-pen writing, the information is much more than just the length, width, height, center of gravity, position, etc. shown on the image; the direction of the tip of the brush, the staccato, the turn, and the connection between the segments are all information to be grasped, so this paper intends to study the stroke segments of a single stroke. How to divide the strokes into meaningful segments becomes one of the main points of research. As early as 2005, Batuwita et al.[12] proposed an offline handwritten English character segmentation algorithm to split a single English character into meaningful segments based on the skeleton, which defines the primary and secondary starting points of the character skeleton, and obtains the coordinates of the primary and secondary starting points and all the pixel points of the skeleton by traversing the pixels of the character skeleton, and then the character is divided into multiple segments according to the change of traversal direction, which is a good method for handwritten English characters. This method works well for handwritten English characters, but Chinese characters have many strokes and are complicated to write, in which there are many combinations of straight lines and curves, so this method is not applicable to Chinese characters' strokes. Liu et al.[13] proposed a stroke segmentation method based on fast penalty dynamic programming, which measures the reasonableness of each point as a split point by calculating the penalty degree, then calculates the minimum fitting error of the sub-segments and the corresponding fitting type, and finally calculates the minimum fitting error of the whole segment from the bottom up to construct the optimal solution. Compared with the traditional stroke segmentation method that requires prior knowledge of the number and type of segments, this method combines the global optimal and local optimal strategies to achieve the stroke segmentation task without prior knowledge.

The above methods all use the refined character skeleton for segmentation, and the key lies in the determination of the segmentation point and the distinction between straight line segments and curved segments. However, after the refinement of Chinese character strokes, the important features contained in the original character image will be lost, which is not conducive to the subsequent evaluation of the quality of writing strokes. We hope to complete the segmentation of the strokes on the basis of no loss of the original information, so this paper proposes to use the neural network method to segment the strokes.

Since the information of Chinese character stroke images is simple and without redundant information, the Unet model[14] for medical image segmentation is considered as the basis for the study. Medical images are similar to Chinese character stroke images in that the main composition of the image is gray pixels, the image content is simple, and there are no multiple targets or backgrounds to affect the segmentation task. Because of the simple semantics of medical images, both their shallow and deep information are important, and the low-dimensional and high-dimensional features splicing in the Unet model can well grasp the features at different levels, and retrieve the missing edge features of the deep network through splicing to achieve better segmentation results. Meanwhile, the simple network structure can also avoid the model overfitting problem.

In summary, the main contributions of this paper are as follows:

(1) Four meaningful stroke segments are defined based on their writing patterns for hard-pen regular handwriting strokes;

(2) Based on the defined types of stroke segments, a single stroke region segmentation task is proposed to implement the training and prediction of automatic stroke region segmentation using an improved U-shaped network structure;

(3) Aiming at the corresponding stroke area after the segmentation of the copy strokes and the template strokes, a similarity judgement method based on Hu moments is proposed, and the degree of similarity of the corresponding area is calculated.

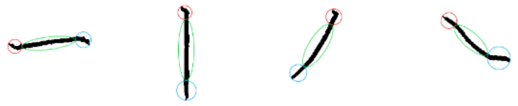
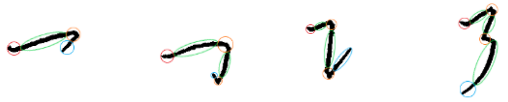
3. SINGLE STROKE REGIONAL SIMILARITY EVALUATION METHOD (SSRSE)

In this section we describe the region segmentation task for a single stroke, the inverse residual U-net based model for region segmentation of a single stroke, and the method for evaluating the similarity of the segmented stroke regions.

3.1 Single stroke segmentation task

According to the writing order of Chinese character strokes, we divide Chinese character strokes into four regions[15]. The four regions are "starting region", "smoothing region", "turning region", and "ending region". We believe that each region of a stroke contains the characteristics of the stroke during the writing process, including the direction of the stroke, the structure of each region, and the connection between the regions. After obtaining each region, the similarity of each region between the copy stroke and the template stroke can be calculated separately to achieve finer stroke evaluation. The definition of several types of common stroke regions is shown in Table 1.

Table 1. Definition of stroke regions.

Strokes	Schema	Regions
Heng, Shu, Pie, Na		Includes starting, smoothing and ending regions.
Henggou, Hengzhegou, Hengzheti, Hengzhezhepie		Includes starting, smoothing, turning and ending regions.

3.2 Single stroke region segmentation model IRB-Unet

In the Unet model, the maximum pooling method is used in the down sampling stage to achieve image size reduction in order to obtain feature maps of different sizes from which different dimensions of image features can be uncovered. The advantage of this is that the image size can be reduced quickly, which reduces the number of parameters and the amount of computation, and can improve the training and testing speed of the model. Since maximum pooling only focuses on the largest value in the target region, it has the invariance of rotation and translation, i.e., no matter where the target features in the feature map is located, the output of maximum pooling is always the same, so maximum pooling can extract the most significant features in the image, which helps to improve the robustness and generalization ability of the model, and makes the model have a better performance in different datasets or different tasks. However, because maximum pooling only retains the maximum value in the pooling area and ignores other feature values, this can lead to the loss of other features in the image, thus affecting the final result. When maximum pooling focuses on the most salient information in the feature map, the model is highly sensitive to such information, which can lead to model overfitting.

In this paper, the inverted residual structure is introduced with the aim of solving the possible problems in the Unet model described above. The inverted residual structure is shown in Figure 1. It is the opposite of ResNet's residual structure, which is first uplifted and then down lifted, and is divided into an extension layer, a feature extraction layer, and a projection layer. Its essence is to first use 1×1 convolution to upgrade the feature map, mapping the low-dimensional features to the high-dimensional space, expanding the dimensionality, and facilitating the extraction of the subsequent features; after that, the features are extracted by using the 3×3 depth separable convolution, which decomposes the convolution operation into the depth convolution and the point-by-point convolution, greatly reduces the amount of computation and the number of parameters, and greatly improves the speed of the model. At the same time, it makes the model learn data at a finer granularity, and enhances the expressive ability of model; finally, the 1×1 convolution block is used to reduce the dimensionality, which remaps the high-dimensional information to the low-dimensional, making the network shrink again. In order to limit the range of activation values, the original ReLU activation function is replaced by the ReLU6 function as follows to eliminate the effect of gradient explosion and gradient disappearance on the model, and the linear activation function is chosen to be used after the projection layer due to the fact that the nonlinear activation function will cause the loss of low-dimensional information.

$$y = \mathrm{ReLU}\,6(x) = \min(\max(x,0),6) \tag{1}$$

In addition, the inverted residual structure incorporates jump connections similar to ResNet, where the inputs and outputs are superimposed when and only when the convolution step is 1. This is done to ensure that the model does not become worse when the model layers are deepened.

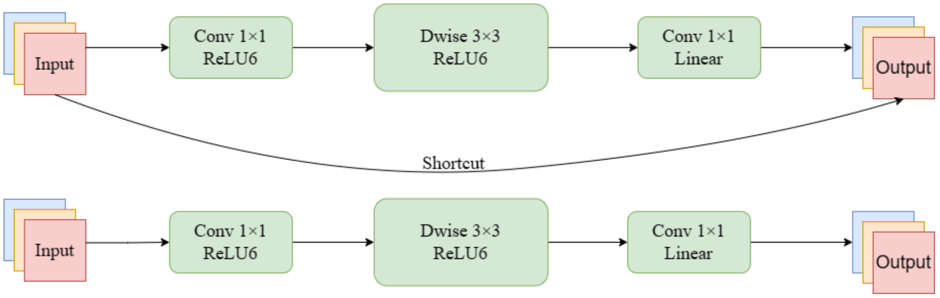


Figure 1. Structure of the inverted residual block.

Based on the Unet model, this paper introduces the inverted residual block model in MobileNetV2[16], proposes IRB-Unet (Inverted Residual Block Unet) model, which replaces the down sampling part of Unet with the inverted residual block, and adopts the idea of full-size feature fusion in the Unet++ model to enhance the feature extraction capability by up-sampling the higher levels and splicing the lower levels to construct the middle layer[17], reduce the feature loss caused by maximum pooling in the down-sampling, and improve the segmentation accuracy. The structure of the model is shown in Figure 2, and the layers of the network are defined in the Table 2.

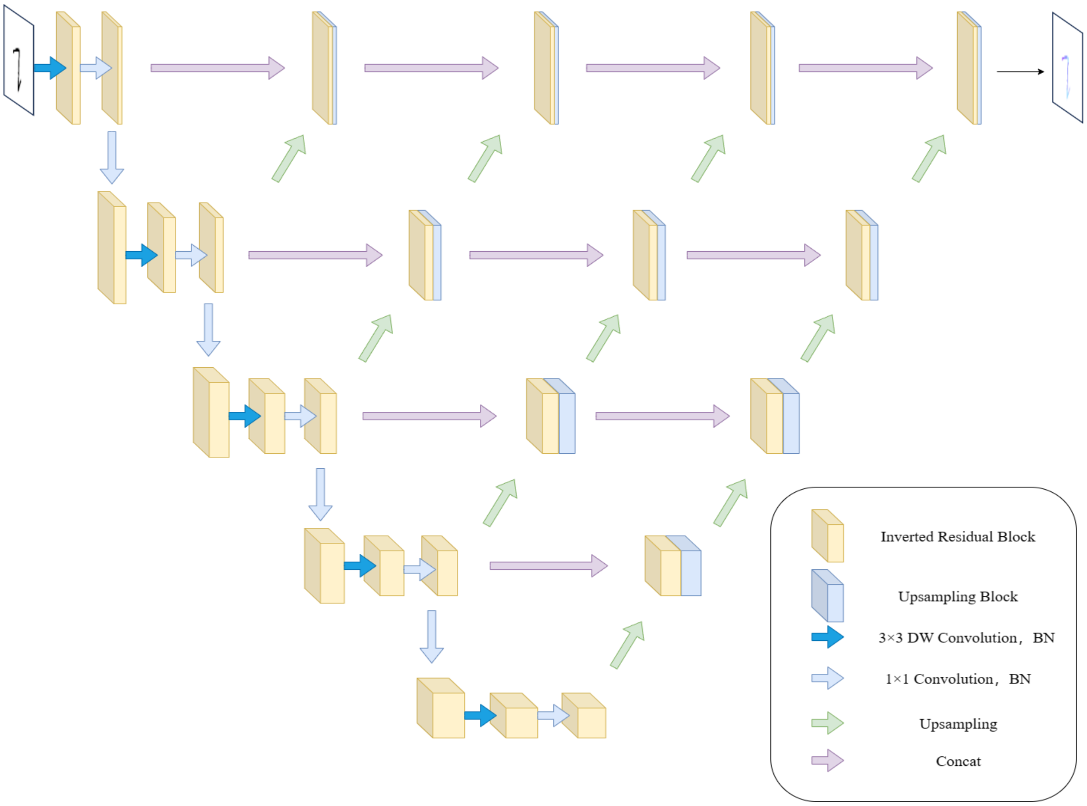


Figure 2. Schematic diagram of the IRB-Unet model structure.

Table 2. Definitions of network layers.

Layer	Input	Upsample	Output	Output size
Input Image	-	-	-	1×256×256
E1	Input Image	-	X1	32×256×256
E2	X1	X1_1	X2	64×128×128
E3	X2	X2_1	X3	128×64×64
E4	X3	X3_1	X4	256×32×32
E5	X4	X5	X5	512×16×16
M1_1	(X1_1, X1)	-	X1_2	32×256×256
M1_2	(X2_2, X1_2)	-	X1_3	32×256×256
M1_3	(X2_3, X1_3)	-	X1_4	32×256×256
M2_1	(X2_1, X2)	X2_2	X2_2	64×128×128
M2_2	(X3_2, X2_2)	X2_3	X2_3	64×128×128
M3_1	(X3_1, X3)	X3_2	X3_2	128×64×64
D1	(X5, X4)	X6	X6	256×32×32
D2	(X6, X3)	X7	X7	128×64×64
D3	(X7, X2)	X8	X8	64×128×128
D4	(X8, X1_4)	-	X9	32×256×256
Output Image	X9	-	output	5×256×256

In the proposed single-stroke region segmentation method, the up-sampling portion uses a sampling method of bilinear interpolation to gradually restore the down-sampled feature map to its original size, and concatenates the feature map of the up-sampled portion with the down-sampled portion for feature fusion to improve segmentation efficiency.

In the stroke segmentation data, the proportion difference between the target and background in the image is significant, and the sizes of the regions of the parts in the same stroke are also different. Since the dataset for the single-stroke segmentation task has the problem of sample and category imbalance, in the proposed single-stroke region segmentation method, we use the Lovász Softmax[18] function to evaluate the gap between the model prediction and the true value. Lovász Softmax has better sample-category balancing ability than the traditional cross-entropy loss function and can better handle the boundary blurring problem. If $y*$ denotes the segmentation result, and y denotes the labelled result, then the evaluation metrics of the IoU-based segmentation task is:

$$J_c(y^*, y) = \frac{\left|\{y^* = c\} \bigcap \{y = c\}\right|}{\left|\{y^* = c\} \bigcup \{y = c\}\right|} \tag{2}$$

Its loss function is expressed as:

$$\Delta J_c(y^*, y) = 1 - J_c(y^*, y) \tag{3}$$

For the output of the above model, we use the Softmax function to map the output to a probability distribution, assuming that $f_i(c) \in [0,1]$ is the probability that the ith pixel belongs to category c , then the category c has an pixel error $m_i(c)$:

$$m_i(c) = \begin{cases} 1 - f_i(c), & \text{if } c = y_i* \\ f_i(c), & \text{other} \end{cases} \tag{4}$$

Then the Jaccard’s coefficient of category c is:

$$loss(f(c)) = \overline{\Delta J_c}(m(c)) \tag{5}$$

Average and get the Lovász Softmax Loss:

$$loss(f) = \frac{1}{|C|}\sum_{c\in C}\overline{\Delta J_c}(m(c)) \tag{6}$$

3.3 Regional similarity evaluation

In Chinese character strokes, the same stroke is in different forms in different structures, and there are also different stroke forms in different types of calligraphy, the stroke characteristics will be completely lost after the stroke refinement, so this paper analyses the similarity of the corresponding regions using the invariant moments of the images from the original stroke images. Hu Moments[19] is a feature description method used in image processing and computer vision, and its seven numerical features have the invariance of rotation, translation and scaling, which can give accurate similarity judge results for images.

The Hu moments are computed for the corresponding stroke regions in the copy stroke image and the template stroke image after region segmentation, and the Hu moments are computed as follows:

$$h_0 = \eta_{20} + \eta_{02} \tag{7}$$

$$h_1 = \left(\eta_{20} - \eta_{02}\right)^2 + 4\eta_{11}{}^2 \tag{8}$$

$$h_2 = \left(\eta_{30} - 3\eta_{12}\right)^2 + \left(3\eta_{21} - \eta_{03}\right)^2 \tag{9}$$

$$h_3 = \left(\eta_{30} + \eta_{12}\right)^2 + \left(\eta_{21} + \eta_{03}\right)^2 \tag{10}$$

$$\begin{aligned} h_4 = {} & \left(\eta_{30} - 3\eta_{12}\right)\left(\eta_{30} + \eta_{12}\right)\left[\left(\eta_{30} + \eta_{12}\right)^2 - 3\left(\eta_{21} + \eta_{03}\right)^2\right] \\ & + \left(3\eta_{21} - \eta_{03}\right)\left(\eta_{21} + \eta_{03}\right)\left[3\left(\eta_{30} + \eta_{12}\right)^2 - \left(\eta_{21} + \eta_{03}\right)^2\right] \end{aligned} \tag{11}$$

$$h_5 = \left(\eta_{20} - \eta_{02}\right)\left[\left(\eta_{30} + \eta_{12}\right)^2 - \left(\eta_{21} + \eta_{03}\right)^2\right] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \tag{12}$$

$$\begin{aligned} h_6 = {} & \left(3\eta_{21} - \eta_{03}\right)\left(\eta_{30} + \eta_{12}\right)\left[\left(\eta_{30} + \eta_{12}\right)^2 - 3\left(\eta_{21} + \eta_{03}\right)^2\right] \\ & - \left(\eta_{30} - 3\eta_{12}\right)\left(\eta_{21} + \eta_{03}\right)\left[3\left(\eta_{30} + \eta_{12}\right)^2 - \left(\eta_{21} + \eta_{03}\right)^2\right] \end{aligned} \tag{13}$$

where, η_{ij} denotes the normalized center distance obtained by normalizing the center distance of each order of the image using the zero-order center moment:

$$\eta_{ij} = \frac{u_{ij}}{u_{00}{}^{r}}(r = \frac{i+j}{2}) \tag{14}$$

$$u_{ij} = \sum_{x=1}^{C}\sum_{y=1}^{R}\left(x - x_0\right)^i\left(y - y_0\right)^j f\left(x, y\right) i, j = 0,1,2... \tag{15}$$

where C, R denote the columns and rows of the image, $f(x, y)$ denotes the gray value of the point (x, y), and x_0, y_0 denote the horizontal and vertical coordinates of the center of mass:

$$x_0 = \frac{M_{10}}{M_{00}} \tag{16}$$

$$y_0 = \frac{M_{01}}{M_{00}} \tag{17}$$

$$M_{ij} = \sum_X \sum_Y x^i y^j f(x, y) \tag{18}$$

For the calculated invariant moments $h_0 \sim h_6$, using the log transformation to distribute the values into the same range:

$$H_i = -sign(h_i) log\,|\,h_i\,| \tag{19}$$

Defining the degree of difference $D(A,B)$ to measure the difference between shapes A and B , giving a regional similarity score based on the difference:

$$D(A,B) = \sum_{i=0}^{6} |\,H_i^B - H_i^A\,| \tag{20}$$

4. EXPERIMENTS

4.1 Experimental data

In this paper, 22 handwritten character sets of commonly used strokes were collected from primary and secondary schools, and the original data is shown in Figure 3. In this paper, 5 of them were selected for region annotation. Since the single-stroke segmentation task requires that the stroke regions in the dataset are complete and free of redundant pixels, the traditional tool LabelMe is unable to do the labelling at the pixel level, this experiment uses the PS tool to label to ensure the accuracy and completeness of the stroke regions at the pixel level. A total of 960 images were created for this experiment, of which 160 were used as the test set, 640 as the training set, 160 as the validation set, and the stroke region labelling schematic is shown in Figure 4.

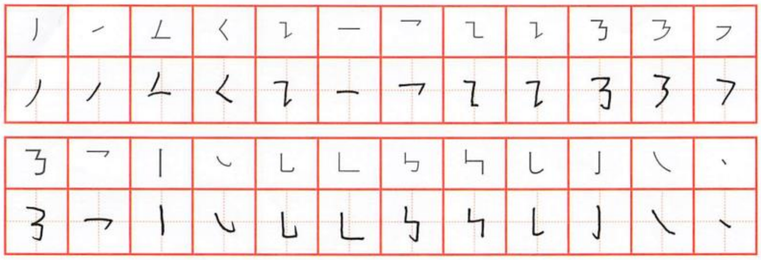


Figure 3. Schematic diagram of the original data.

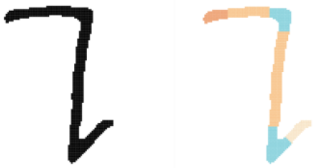


Figure 4. Schematic diagram of the original stroke and labelling stroke.

4.2 Experimental setup

The device used in this experiment is a personal server based on 64-bit Ubuntu system with i7 9700K processor at 3.6 GHz, 32 GB of RAM, and Nvidia GeForce RTX 2080Ti GPU. The network is optimized with Adam optimizer to ensure fast convergence, the number of training rounds is 100, the Batch Size is 2, the initial learning rate is 0.001, weight decay is 1×10^{-8}, and momentum is 0.9.

4.3 Evaluation metrics

Common performance evaluation metrics in image segmentation tasks are Accuracy, Precision, Recall, IoU, Dice coefficient, etc. The main purpose is to assess the similarity between the segmentation results predicted by the network and the real values labelled by human beings, and the data required is the confusion matrix computed from the segmentation results and the real value images. The computed confusion matrix contains the following four values shown in Figure 5: TP, FP, TN, FN, which represent the positive values predicted as positive, negative values predicted as positive, negative values predicted as negative, and positive values predicted as negative, respectively.

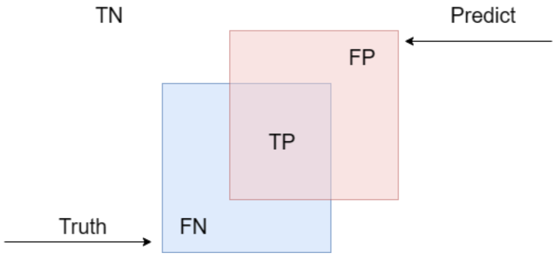


Figure 5. TP, FP, TN, FN.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{21}$$

$$Precision = \frac{TP}{TP + FP} \tag{22}$$

$$Recall = \frac{TP}{TP + FN} \tag{23}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{24}$$

$$Dice = \frac{2TP}{2TP + FP + FN} \tag{25}$$

In the multi-target segmentation task, due to the existence of multiple categories, its confusion matrix contains more ranks, so the evaluation metrics are slightly different. In this paper, we choose the category average pixel accuracy (MPA), average intersection ratio (MIoU), average Dice coefficient (MDice) three metrics to evaluate the model, the specific calculation is as follows, N is the number of categories, and i denotes the *i*th category.

$$CPA = \frac{TP}{TP + FP} \tag{26}$$

$$MPA = \frac{\sum_{i}^{N} CPA}{N} \tag{27}$$

$$MIoU = \frac{\sum_{i}^{N} IoU}{N} \tag{28}$$

$$MDice = \frac{Dice_i}{N} \tag{29}$$

5. EXPERIMENTAL RESULTS AND ANALYSES

In this section, the proposed method and its variants were trained separately using the same dataset, and were compared with traditional segmentation networks UNet and UNet++. The loss curves of the validation and training sets for each model during training are shown in Figures 6-9. Here, “DS” denotes the introduction of deep supervision mechanism, and “G” indicates the use of global inverted residual, where the upsampling part is replaced with inverted residual blocks.

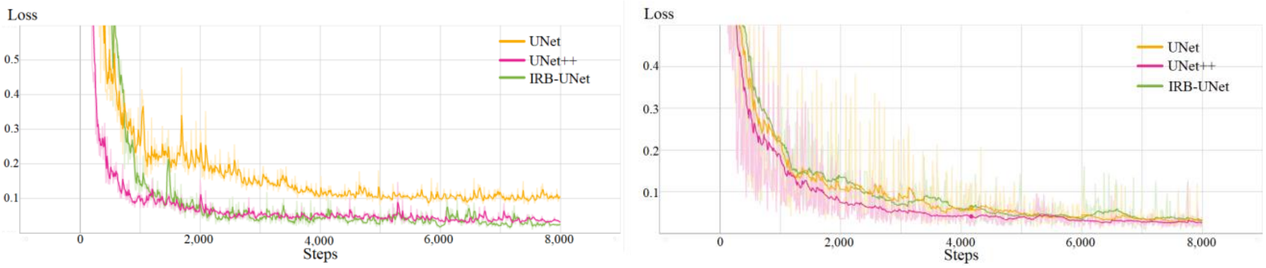


Figure 6. Comparison of validation and training set losses for the three models.

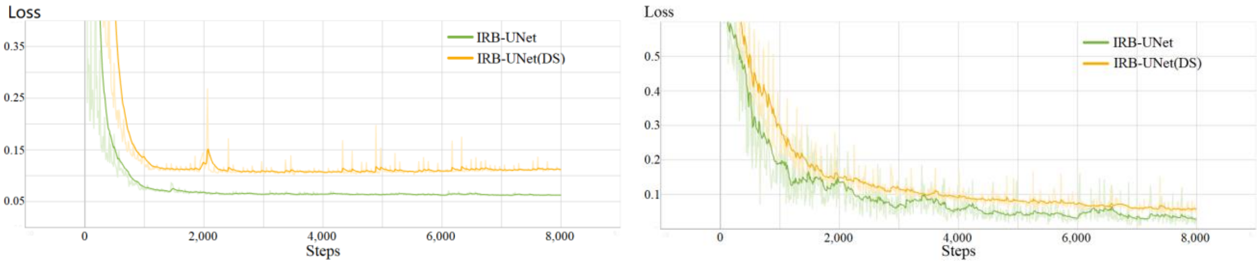


Figure 7. Comparison of validation and training set losses after adding deep supervision.

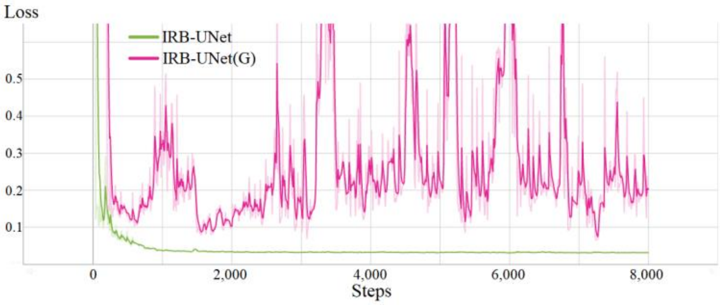


Figure 8. Comparison of validation set losses for global inverted residual.

As shown in Figure 7, after introducing the deep supervision mechanism, the convergence of loss for IRB-UNet on both the validation and training sets did not improve. This is believed to be because deep supervision adds an additional convolutional layer to each node at the top layer, generating extra predictions for each top-layer node. These predictions at different levels cause the network to overly focus on certain features of intermediate layers, which affects the final results. As shown in Figure 8, after incorporating global inverted residual, IRB-UNet exhibits significant fluctuations in validation set loss, while convergence on the training set remains normal. This is believed to be due to overfitting of the training data by the network model after the introduction of global inverted residual, resulting in unstable convergence on the validation set. Therefore, considering introducing deep supervision in IRB-UNet with global inverted residual to enhance learning in intermediate layers.

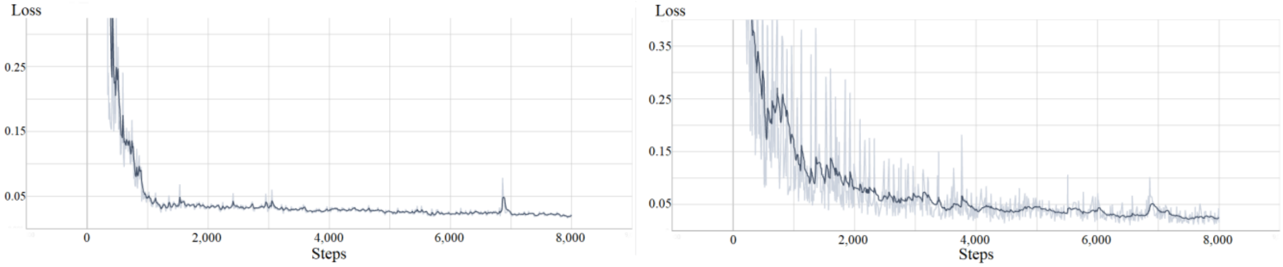


Figure 9. Validation and training set losses for IRB-UNet (G+DS).

As depicted in Figure 9, after incorporating deep supervision into IRB-UNet (G), the losses on both the validation and training sets stabilize and converge. This indicates that deep supervision enhances training stability and has a certain limiting effect on overfitting. Figure 10 displays the segmentation results of three types of strokes, with different regions segmented and labeled using various colors. This paper compares the proposed model and its variants with traditional segmentation networks using the evaluation metrics MPA, MIoU, and MDice. Tables 3 and 4 respectively present the comparison results of various models under different loss functions, while Tables 5 and 6 show the ablation results under different loss functions.

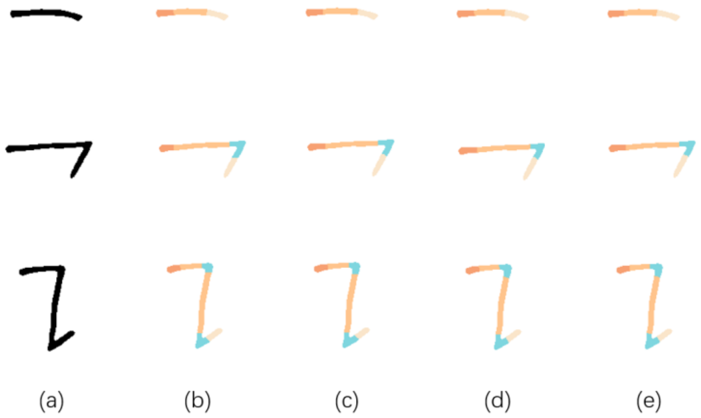


Figure 10. (a) Original Stroke, (b) Ground Truth, (c) IRB-Unet, (d) Unet++, (e) Unet.

Table 3. Comparison of methods (cross entropy loss).

Network	MPA	MIoU	MDice
Unet	0.895	0.82	0.890
Unet++	0.909	0.844	0.901
Unet++ (DS)	0.916	0.850	0.907
IRBUnet	**0.921**	**0.852**	**0.913**

Table 4. Comparison of methods (Lovász Softmax loss).

Network	MPA	MIoU	MDice
Unet	0.925	0.86	0.916
Unet++	0.929	0.874	0.920
Unet++ (DS)	0.931	**0.882**	0.922
IRBUnet	**0.937**	0.866	**0.924**

Table 5. Ablation experiment (cross entropy loss).

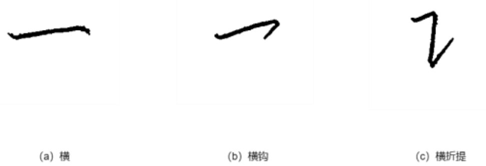
Network	MPA	MIoU	MDice
IRB-UNet	**0.921**	**0.852**	**0.913**
IRB-UNet (G)	0.882	0.808	0.878
IRB-UNet (DS)	0.912	0.846	0.894
IRB-UNet (G+DS)	0.915	0.819	0.882

Table 6. Ablation experiment (Lovász Softmax loss).

Network	MPA	MIoU	MDice
IRB-UNet	**0.937**	**0.866**	**0.924**
IRB-UNet (G)	0.917	0.855	0.897
IRB-UNet (DS)	0.920	0.861	0.903
IRB-UNet (G+DS)	0.919	0.858	0.899

As shown in Table 3, when using the traditional cross-entropy loss function, IRB-UNet achieves better performance in all three metrics. As indicated in Table 4, the segmentation results of each model are improved when using the Lovász Softmax loss function. IRB-UNet achieves 0.937 and 0.924 in average pixel accuracy and average Dice coefficient, respectively, both of which are higher than the traditional segmentation networks UNet and UNet++. However, traditional networks have a slight advantage in average IoU, which the paper attributes to the class imbalance issue in the samples. As shown in Tables 5 and 6, IRB-UNet with the addition of global inverted residual does not achieve better results in the final test set prediction, consistent with the loss convergence performance during training. However, with the addition of deep supervision, the segmentation accuracy improves, indicating that the deep supervision mechanism can improve its stability to some extent. Nevertheless, due to the significant increase in parameters and computational complexity introduced by the global inverted residual, the training speed of the model is greatly reduced, making it unsuitable for practical applications.

After IRB-Unet segmentation and regional similarity calculation this study invited 50 participants to rate the handwritten strokes in the test, including 45 students and 5 teachers. For each region of the same stroke, the score is given according to the similarity with the corresponding region of the template. Template stroke is shown in Figure 11. Details of average scoring of the three strokes are shown in Tables 7-9.



(a) 横

(b) 横钩

(c) 横折提

Figure 11. Template strokes.

Table 7. Stroke “Heng” regional similarity scores.

Strokes	Similarity rating	Manual scoring (mean)
	75.5	75.4
	75.0	70.1
	86.0	91.5

Table 8. Stroke “Henggou” regional similarity scores.

Strokes	Similarity rating	Manual scoring (mean)
	70.7	80.8
	65.6	67.8
	56.5	60.3

Table 9. Stroke “Hengzheti” regional similarity scores.

Strokes	Similarity rating	Manual scoring (mean)
	79.2	86.9
	69.9	76.3
	67.0	63.6

From the above table, the mean scores of subjects’ scores are similar to the results of similarity score calculation, which proves that the similarity evaluation method possesses high feasibility and can be used in the stroke evaluation part of the evaluation system or procedure of handwriting evaluation system of hard-pen regular script, or to provide evaluation metrics for handwriting evaluation system.

6. CONCLUSIONS

After researching and analyzing Chinese characters, this paper proposes a similarity evaluation method for handwritten strokes in hard-pen regular script, defines four regions of strokes according to the writing process and characteristics of strokes, and proposes IRB-Unet, a U-type network based on the inverted residual structure, to segmentation regions of single strokes. Compared with the traditional method, the proposed network retains the features of small size and fast operation of the U-type network, and at the same time reduces the loss of features in the down sampling part, improves the segmentation accuracy, and has a good application prospect. For the segmented stroke region, a similarity evaluation method based on Hu moments is proposed, which is experimentally proved that the score obtained by the method is similar to the manual score, and can provide evaluation metrics for the subsequent overall evaluation of handwriting, which is applicable to the evaluation system or software of handwriting in hard-pen regular script.

There is still room for improvement in the proposed segmentation network, and we expect to get a model with higher accuracy and faster speed. In the future, we will conduct further research on this model, increase the number of datasets, and introduce multiple styles of calligraphy, such as semi-cursive script, cursive script, etc. to realize the segmentation of multiple types of characters, and provide more help for calligraphy education.

REFERENCES

[1] Deng, C., [Research on the Current Situation of Calligraphy Education in Primary and Secondary Schools], Chongqing: Southwest University, Master’s Thesis, (2015).

[2] Liu, Y. and He, K. K., “Research on Computer aided Chinese character writing teaching-development of a writing Chinese character library generation system,” Journal of Chinese Information Processing 4, 34-42 (1994).

[3] Wang, S. K. and Zhao, X. W., “Research on computer-aided stroke order writing of Chinese characters,” Journal of Inner Mongolia Normal University (Natural Science Edition) 39, 428-432 (2010).

[4] Zhuang, C. B. and Jin, L. W., “An intelligent evaluation algorithm for the correctness, error, and neatness of online Chinese character writing,” 12th National Signal Processing Academic Annual Conference 281-284 (2005).

[5] Wang, Q. Z., Dai, Y., Fan, L. and Sun, G. W., “Fuzzy analysis method for the quality of Chinese character writing,” Computer Engineering and Applications 49, 180-185 (2013).

[6] Chen, H. M., “Writing quality evaluation system,” Electronic Technology & Software Engineering 4, 192-193 (2014).

[7] Geng, X. Y., Xu, W. S. and Wu, J. W., “A quantitative evaluation model for Chinese character quality based on neural networks,” Computer and Modernization 1, 96-99+120 (2014).

[8] Sun, R., Lian, Z., Tang, Y., et al., “Aesthetic visual quality evaluation of Chinese handwritings,” IJCAI, Buenos Aires, Argentina 2510-2516 (2015).

[9] Jiang, J., Wu, J. Y., Han, Q. and Li, Y., “Implementation and effect testing of a comprehensive scheme for evaluating the accuracy of handwritten Chinese characters,” E-Education Research 40, 50-58 (2019).

[10] Yan, W. Y., Guo, M. T., Wang, Z. X. and Zhang, J. L., “Exploiting Gabor feature extraction method for Chinese character writing quality evaluation,” Journal of Physics: Conference Series 1575, 012065 (2020).

[11] Xiao, X. and Li, C. C., “Research progress on evaluation methods for handwritten Chinese characters,” Computer Engineering and Applications 58, 27-42 (2022).

[12] Batuwita, K. and Bandara, G., “New segmentation algorithm for individual offline handwritten character segmentation,” International Conference on Fuzzy Systems and Knowledge Discovery 215-229 (2005).

[13] Liu, W. Y., Tong, L., Yu, Y. J., Shuang, L. and Rui, Z., “Online stroke segmentation by quick penalty-based dynamic programming,” IET Computer Vision 7, 311-319 (2013).

[14] Olaf, R., Fischer, P. and Brox, T., “U-net: Convolutional networks for biomedical image segmentation,” Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference 234-241 (2015).

[15] Fan, K., [Research on Fine Grained Writing Control and Copying Methods of Strokes in Robot Calligraphy], Chongqing: Southeast University, Master’s Thesis, (2018).

[16] Mar, S., Andrew, H., Zhu, M. L., Andrey, Z. and Chen, L. C., “Mobilenetv2: Inverted residuals and linear bottlenecks,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 4510-4520 (2018).

[17] Zhou, Z. W., Siddiquee, R., Mahfuzur, M., Nima, T. and Liang, J. M., “Unet++: A nested u-net architecture for medical image segmentation,” Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Granada, Spain, 4, 3-11 (2018).

[18] Maxim, B., Rannen, T. A. and Matthew, B., “The Lovász-Softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Utah, USA, 4413-4421 (2018).

[19] Hu, M. K., “Visual pattern recognition by moment invariants,” IRE Transactions on Information Theory 8, 179-187 (1962).