



A Residual Network with Multi-Scale Dilated Convolutions for Enhanced Recognition of Digital Ink Chinese Characters by Non-Native Writers

Huafen Xu^{1,2*}, Xiwen Zhang^{2*}

¹ College of Computer, North China Institute of Science and Technology, 065201 Langfang, China

² College of Information Science, Beijing Language and Culture University, 100083 Beijing, China

* Correspondence: Huafen Xu (xhf@ncist.edu.cn); Xiwen Zhang (zxw@blcu.edu.cn)

Received: 06-10-2024

Revised: 07-19-2024

Accepted: 07-25-2024

Citation: H. F. Xu and X. W. Zhang, “A residual network with multi-scale dilated convolutions for enhanced recognition of digital ink Chinese characters by non-native writers,” *Int J. Knowl. Innov Stud.*, vol. 2, no. 3, pp. 130–146, 2024. <https://doi.org/10.56578/ijkis020302>.



© 2024 by the author(s). Published by Acadlore Publishing Services Limited, Hong Kong. This article is available for free download and can be reused and cited, provided that the original published version is credited, under the CC BY 4.0 license.

Abstract: Digital ink Chinese character recognition (DICCR) systems have predominantly been developed using datasets composed of native language writers. However, the handwriting of foreign students, who possess distinct writing habits and often make errors or deviations from standard forms, poses a unique challenge to recognition systems. To address this issue, a robust and adaptable approach is proposed, utilizing a residual network augmented with multi-scale dilated convolutions. The proposed architecture incorporates convolutional kernels of varying scales, which facilitate the extraction of contextual information from different receptive fields. Additionally, the use of dilated convolutions with varying dilation rates allows the model to capture long-range dependencies and short-range features concurrently. This strategy mitigates the gridding effect commonly associated with dilated convolutions, thereby enhancing feature extraction. Experiments conducted on a dataset of digital ink Chinese characters (DICCs) written by foreign students demonstrate the efficacy of the proposed method in improving recognition accuracy. The results indicate that the network is capable of more effectively handling the non-standard writing styles often encountered in such datasets. This approach offers significant potential for the error extraction and automatic evaluation of Chinese character writing, especially in the context of non-native learners.

Keywords: Digital ink Chinese character recognition (DICCR); Multi-scale context; Residual network; Dilated convolution

1 Introduction

With the continuous development and maturity of handwriting input devices such as tablets and digital pens, DICCs are constantly generated in learning, work, and life. There are already datasets stemming from native Chinese speakers with correct and standardized Chinese character writing, such as the datasets CASIA-OLHWDB [1] and SCUT-COUCH2009 [2]. Great progress has been made in DICCR for these native-speaker datasets [3].

<https://doi.org/10.56578/ijkis020302>

Nowadays, more and more foreign students are using handwriting devices for their studies and daily lives, and DICCs are generated constantly. The recognition of DICCs by foreign students is of great significance for Chinese international education. Recognizing handwritten Chinese characters can facilitate the evaluation of their correctness and provide standardized guidance. This makes it possible for computers to guide students in writing Chinese characters. However, foreign students often unconsciously apply their native thinking and writing habits to Chinese character learning, leading to writing errors and various non-standardization, which pose great challenges to their DICCR:

(1) Stroke errors:

- * Extra strokes: Unnecessary additional strokes that do not belong to the standard form.
- * Missing strokes: Omissions of required strokes.
- * Connected strokes: Strokes that should be separate but are incorrectly joined.
- * Broken strokes: Strokes that should be continuous but are interrupted.
- * Incomplete strokes: Partially written strokes that do not reach their intended endpoints.

(2) Incorrect stroke relationships:

- * Stroke order: The sequence in which strokes are written deviates from the standard.
- * Stroke direction: Strokes are drawn in the wrong orientation.
- * Geometric errors: Misplacement or misalignment of strokes within the character structure.

(3) Non-standardized DICCs:

- * Overall structural imbalance: Characters may appear skewed or improperly proportioned.
- * Non-standard strokes: Variations in stroke shapes that do not conform to standard forms.

The complexity and frequency of these errors increase with the number of strokes and the intricacy of stroke combinations. As a result, the accuracy and standardization of foreign students' DICCs tend to decline, particularly for more complex characters. These challenges are illustrated in Figure 1, highlighting the need for robust and adaptive recognition systems that can accommodate the diverse and nonstandard writing patterns of foreign students.

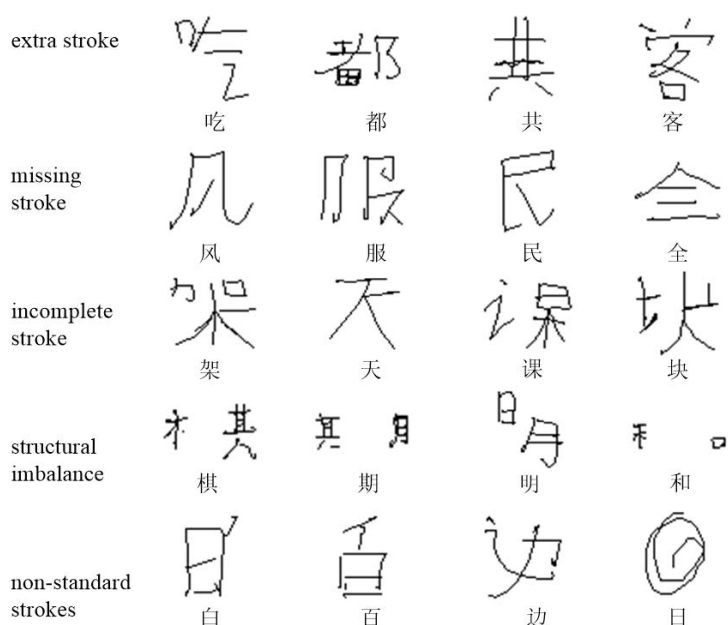


Figure 1. Examples of DICCs for foreign students

There has been some research on DICCR for foreign students, such as the Chinese character recognition method based on stroke names and whole character structures proposed by Bai and Zhang [4]. Later, they utilized Hidden Conditional Random Field (HCRF) to enhance model performance [5]. These methods all belong to structure-based recognition methods, which posit that recognizing Chinese characters should follow the process of writing them: composed of strokes to form radicals and radicals to form Chinese characters. However, there are various stroke errors and stroke order errors in the DICCs of foreign students, and many strokes are not written in a standardized manner, which makes it difficult to stably extract the stroke structure and its interrelationships of Chinese characters. Although it is possible to obtain certain experimental results and even some demonstration systems under certain conditions, recognition methods that rely on extracting stroke structures cannot solve the DICCs recognition problem for foreign students. Xu and Zhang [6] proposed the 1-D ResNetDC for classifying DICCs writing trajectories. This method is specifically designed for learning sequential data; therefore, it relies on the writing trajectory of Chinese characters and cannot achieve stroke-order freedom [7].

There have been many studies on DICCs targeting native Chinese speakers, and good recognition performance has been achieved, such as RNN-based recognition methods [8, 9] and Convolutional Neural Network (CNN)-based recognition methods [3, 10–12]. RNN-based recognition methods process raw sequence data to better utilize the rich temporal and spatial information contained in the sequence data. However, there are errors of stroke order and stroke direction in DICCs for foreign students, which will affect the accuracy of RNN models that use sequence data as input. CNN-based recognition methods convert the sequential data of DICCs into digital images or extract feature maps. These methods all study the recognition of DICCs from native language writers, while DICCs for foreign students have their own characteristics, and these methods are not applicable.

This paper converts the handwriting sequence into a 2D black-and-white image, which can avoid effects such as stroke direction, stroke order, continuous strokes, and broken strokes. The converted images have no noise from scanning or taking photos, and the data volume is small, which is beneficial for training the model. This paper proposes a residual network based on dilated convolution (ResNetDC), which can not only capture the important features of DICCs in short distances but also ensure the extraction of long-distance correlation. It has a good recognition rate for foreign students' DICCs existing writing errors, such as extra, missing, connected, broken, and incomplete strokes, as well as non-standard writing of DICCs.

2 Related Works

DICCR methods that rely on handcrafted features are limited by these low-capacity features, making it difficult to improve recognition performance [2, 13]. With the rapid development of GPU parallel computing support for deep learning, new breakthroughs have been brought to DICCR. The performance of DICCR has rapidly improved, surpassing traditional methods comprehensively, and the recognition accuracy exceeds human level. Based on the representation of input data, we classify existing deep learning methods for DICCR into three categories.

A. Recurrent Neural Network (RNN) method based on sequence data

RNN methods based on sequence data directly process the original sequence data. Zhang et al. [8] used both LSTM and GRU for RNN modeling and built a deep RNN model by stacking bidirectional RNNs to achieve end-to-end recognition of DICCs. Zhang et al. [9] proposed a trajectory-based component analysis network (TRAN). Ren et al. [14] proposed the variance constraint and attention weight vector to improve the performance of the RNN network. Due to the recurrent computing mechanism of RNN, it is not easy

to parallelize and has low computational speed; moreover, the recognition effect of these methods relies on the writing order of DICCs, making it difficult to achieve stroke-order freedom. There are many stroke-order issues with the Chinese characters written by foreign students, so this type of method is not applicable.

B. Graph Neural Network (GNN) method based on graph-structured data

GNN methods process graph-structured data and explicitly model the geometric semantics of DICCs. Specifically, Gan et al. [15] proposed SGCN for DICCR. SGCN uses spatial graph convolution to combine information about nearby neighborhoods and a hierarchical residual structure to use the global shape properties to make the final classification. Gan et al. [16] proposed PyGT to recognize graph-structured DICCs. This method requires constructing direct geometric graphs based on coordinate sequences, extracting the features of vertices in the graph, designing graph convolution kernels, etc., which is more complex to implement. Due to the confusion between similar characters, GNNs cannot accurately recognize characters with similar structures.

C. CNN method based on grid data

CNN methods process two-dimensional grid data. Cireřan and Meier [17] introduced MCDNN, which is a set model that classifying Chinese characters images. Gan [18] developed a one-dimensional CNN architecture for recognizing DICCs. Hu et al. [19] utilized a CNN-based method for IAHCCR. These methods all fall under the end-to-end approach category, where the system learns to recognize characters directly from raw input data without the need for manual feature extraction or intermediate processing steps.

Integrating domain-specific knowledge can enhance the recognition performance of traditional CNNs. For instance, Graham [12] utilized path signature to improve recognition accuracy of DICCs. Zhong et al. [20] extracted Gabor features, gradient features, and other features using traditional feature extraction methods, embedding these features into network to improve the recognition rate. Zhang et al. [3] put forward the method of DirectMap + ConvNet + Adaptation by combining the CNN with the domain knowledge of directional feature maps. These approaches integrate manually extracted feature images from Chinese characters as prior knowledge into the CNN architecture. This integration aids CNN in learning auxiliary features of Chinese characters more effectively, thereby significantly enhancing the network's recognition performance. However, this method requires a deep understanding of the complex domain knowledge involved in extracting these feature images.

These traditional convolutional networks aggregate contextual information through continuous convolution strides and pooling operations, with relatively fixed kernel sizes, and the convolutions are all common convolutions. Structures of some DICCs from foreign students are not standard, with excessively dispersed components, as shown in the fourth line of Figure 1. The traditional convolutional network lacks the ability to obtain long-distance correlation of DICCs and cannot effectively learn advanced structural knowledge of Chinese characters.

Dilated convolution, also known as atrous convolution, is an efficient technique for expanding the receptive field of a neural network without increasing the number of parameters or the computational cost. This method is particularly valuable in tasks such as semantic segmentation [21–23], depth estimation [24], and object detection [25], where capturing multi-scale context is crucial. By adjusting the dilation rate, the same convolutional kernel can capture information at varying scales: smaller dilation rates focus on local, short-range details, whereas larger dilation rates enable the kernel to encompass broader, long-range contexts.

In our model, in addition to traditional methods, the resolution of feature maps is considered. The combination of convolution kernel and different dilated rates is used to increase the receptive field and aggregate multi-scale context information while maintaining the resolution of the feature map. This article applies dilated convolution to the recognition of DICCs by foreign students. It studies the application of dilated

convolution in classification tasks, marking a novel and innovative approach in this field.

3 Deep Residual Network Based on Multi-Scale Convolution and Dilated Convolution

This section mainly elaborates on the technical details of the DICCR method proposed in this article for foreign students, including model design, multi-scale context, and visualization.

A. Design of deep networks based on residual Blocks

The increasing number of layers of CNNs brings about significant improvement in performance. Although deep architecture is beneficial for feature learning, there are still many problems in training deep neural networks, including gradient vanishing and gradient explosion. Therefore, we propose network architecture ResNetDC, which adopts residual connections [26], as shown in Figure 2.



Figure 2. Network architecture ResNetDC

In Figure 2, the notation "Conv-64, $k=7 \times 7$, $s=2$ " specifies that this convolutional layer produces 64 output channels, uses a kernel size of 7×7 , and applies a stride of 2. The term "Max pooling 2×2 , $s=2$ " signifies a max pooling operation with a 2×2 window and a stride of 2. We have designed a residual module named Block, which comprises six convolutional layers, as illustrated in Figure 3. These layers are sequentially labeled

as "1st Conv", "2nd Conv", "3rd Conv", "4th Conv", "5th Conv", and "6th Conv". For instance, "2nd Conv k=3×3, stride=1, dilate=1" describes the second convolutional layer, characterized by a 3×3 kernel, a stride of 1, and a dilation rate of 1. The symbol "⊕" represents the element-wise addition operation. Following each convolution, we apply batch normalization and activate the outputs using the ReLU function. It's worth noting that there is variability in the number of channels across the convolutional layers within the four Block modules. The specific channel counts for each convolutional layer in these four modules are detailed in Table 1.

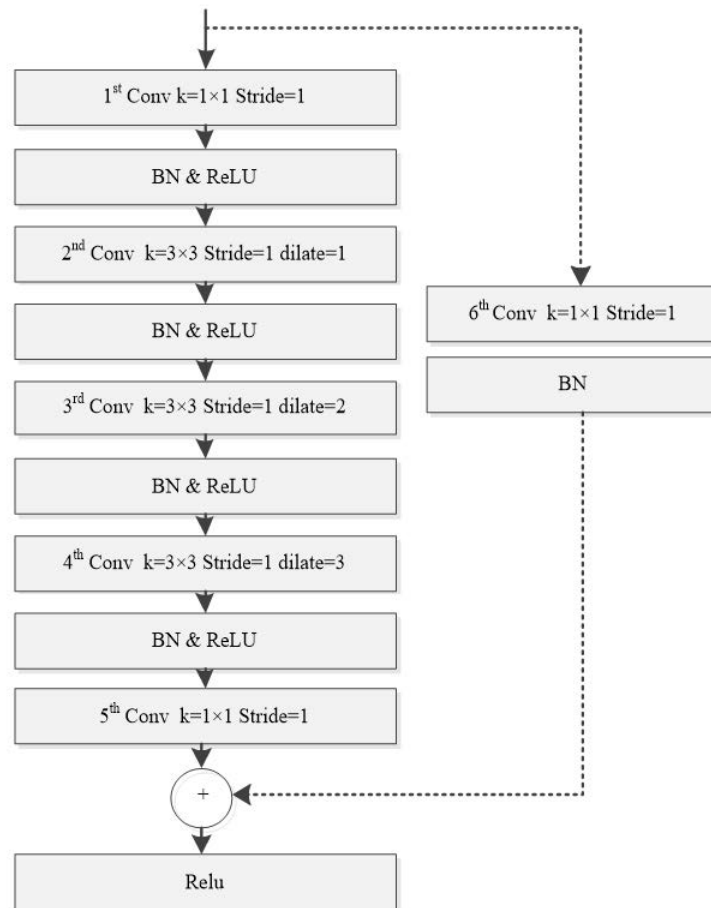


Figure 3. Residual module Block 1

Table 1. Channel counts of each convolutional layer for the Block modules

Layer Name	Block 1	Block 2	Block 3	Block 4
1 st Conv	64	128	256	512
2 nd Conv	64	128	256	512
3 rd Conv	64	128	256	512
4 th Conv	64	128	256	512
5 th Conv	256	512	1024	2048
6 th Conv	256	512	1024	2048

The input Chinese character image is a black-and-white image with a single channel of 64×64 pixels. The first three layers of the network structure down-sample the image by 8 times, and the input feature map

size of Block 1 is 8×8 . Given that the resolution of the feature maps has already been significantly reduced, all convolutional layers within the four Block modules employ a stride of 1. This means that no further down-sampling is applied. As a result, both the input and output feature maps for each of the four Block modules maintain a spatial dimension of 8×8 .

B. Multi-scale convolutional kernel

The ResNetDC network designed in this article employs a multi-scale convolutional kernel strategy. Specifically, the initial convolutional layer utilizes a larger 7×7 kernel to capture broader spatial features. This choice is motivated by the fact that larger kernels provide a wider effective receptive field and higher shape bias, aligning with how humans primarily use shape cues for object recognition [27]. Moreover, smaller (3×3) kernels facilitate the construction of such deep architectures [28]. To leverage these advantages, ResNetDC incorporates multiple Block modules, each featuring 3×3 and 1×1 convolutional kernels. The 1×1 convolutional layers are particularly useful for cross-channel information exchange, enhancing feature integration. By stacking these Block modules, ResNetDC achieves a deep network architecture that effectively captures both local and global features. This multi-scale kernel design significantly boosts the network’s expressive power, enabling it to better model the complex patterns present in Chinese character images.

C. Multi-scale dilated convolution

For handwritten Chinese character recognition tasks, low-level visual cues and advanced structural knowledge are both necessary for predicting successfully. Therefore, the convolutional layers in Block employ varying rates to learn contextual information from deferent spatial ranges.

Traditionally, each convolutional layer is followed by a pooling layer, which helps integrate multi-scale contextual information through the combination of convolutional strides and pooling operations. Besides the traditional way, in the design of this model, we also considered the resolution of the feature map, the four Block modules having a convolution step of 1, removed the pooling layer, and then aggregated multi-scale context information by combining the convolution kernel and different dilation rates.

For a $k \times k$ convolutional kernel with a dilated rate of r , the size of the dilated convolutional kernel is $k_d \times k_d$, where $k_d = k + (k - 1) \times (r - 1)$, so as to realize the purpose of expanding the receptive field. However, it’s important to note that within the expanded $k_d \times k_d$ region, only the original $k \times k$ pixels actually participate in the computation, and a significant portion of the information is lost, so it is a sparse sampling method. When multiple dilated convolutions are used continuously, it is easy to cause Grid Effect [29], losing the continuity and correlation.

To effectively address the issues caused by the grid effect, the 3×3 convolution layers of the each Block in the ResNetDC use varying dilated rates: 1, 2 and 3 respectively. Correspondingly, the dilated kernel sizes are 3×3 , 5×5 and 7×7 respectively, and the receptive fields on the Block’s input feature map are 3×3 , 7×7 and 13×13 . This approach ensures that all holes are covered, thereby significantly increasing the receptive field without sacrificing spatial resolution. By maintaining the 8×8 resolution of the feature maps, the model can capture both local and global features, aggregating context information from multiple scales. This multi-scale context learning enhances the model’s ability to understand complex patterns and improves classification accuracy. Notably, with the input feature image resolution down-sampled to 8×8 , the 7×7 dilated convolution kernels effectively approximate global convolutions, capturing long-range dependencies across the entire feature map. The detailed configuration of the convolutional layers within the Block module is summarized in Table 2.

According to ResNetDC shown in Figure 2, the network focuses on the areas shown in the right figures of Figure 4 and Figure 5, respectively, when predicting the categories of characters “棋” and “和”. If dilated

convolution is not used, that is to say, the dilated rates of all convolutional layers are set to 1, the areas of interest for the network are shown in the left figures of Figure 4 and Figure 5, respectively. They illustrate that ResNetDC we designed can not only capture low-level visual clues of Chinese characters but also acquire advanced structural knowledge. Both short-distance features and long-range correlations can be learned.

Table 2. Detailed configuration of Block

Layer	1	2	3	4	5
Convolution	1×1	3×3	3×3	3×3	1×1
Dilation rate	1	1	2	3	1
Dilated kernel size	1×1	3×3	5×5	7×7	1×1
Receptive field	1×1	3×3	7×7	13×13	13×13

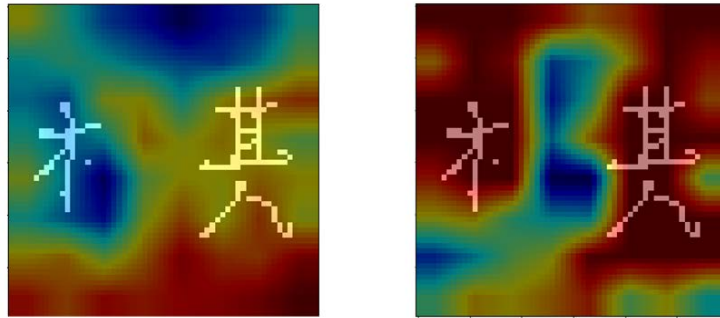


Figure 4. Comparison of regions concerned by not dilated network and ResNetDC when classifying the character “棋”

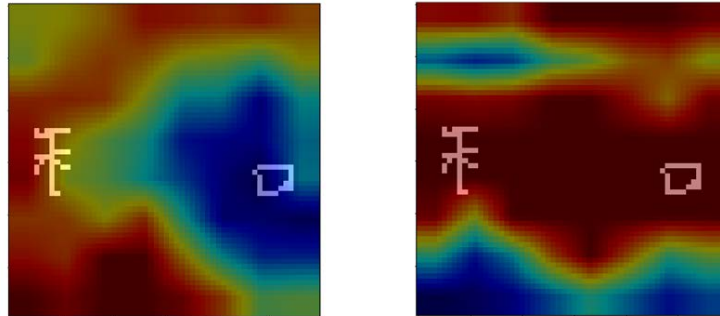


Figure 5. Comparison of regions concerned by not dilated network and ResNetDC when classifying the character “和”

D. Visualization

Figure 6 shows the visualization effect of the focus areas of each layer in the ResNetDC model, which are areas of interest for the 7×7 convolutional layer, 3×3 convolutional layers, and Block 1 – Block 4 from left to right. The input Chinese character is “报” .

Figure 6 illustrates that the low-level convolution operation learns the local and detailed information of the Chinese character image, and has a small receptive field. With the increase of the network depth, the receptive field gradually increases, and the scope of attention is also growing, which conforms to the design intention of the ResNetDC model. That is, the layer-by-layer feature extraction of the image is completed

through multi-layer convolution. The ResNetDC model can capture not only low-level visual clues but also advanced structural knowledge of Chinese characters.

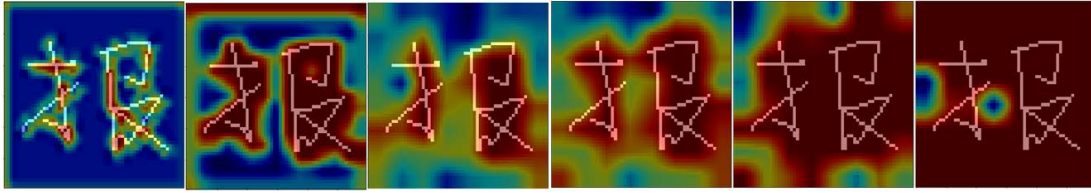


Figure 6. Visualization of regions concerned by each layer in ResNetDC

4 Experiment

A. Datasets

The experimental data of this study is based on the DICC dataset for "zero starting point" foreign students [30, 31]. The dataset uses Anoto digital paper and pen to collect data, which contains 525 categories of Chinese characters and 31,734 samples. The sample size of each category varies greatly, and the dataset is imbalanced. CASIA-OLHWDB1.0 [1] is used to pre-train the ResNetDC, which can effectively alleviate the challenges brought by imbalanced dataset. CASIA-OLHWDB1.0 is established by the Institute of Automation of the Chinese Academy of Sciences and produced by 420 authors, involving 3,866 commonly used Chinese characters, of which 3,740 categories overlap with the GB1 set. Some examples of datasets are shown in Figure 7 and Figure 8.

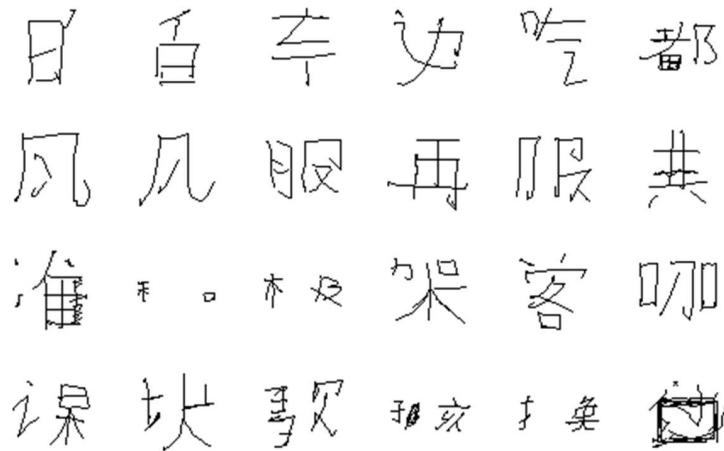


Figure 7. DICC for foreign students

B. Data preprocessing

Because the category of Chinese characters is very large, we rasterize the sampled handwriting sequence into an image, keeping the aspect ratio scaled to 56×56 pixels, and place it on a slightly larger 64×64 image center to allow for various geometric deformations during the training process. Set the foreground color to 0 and the background color to 255.

The amount of images for DICC is large. If each image is stored as a file, not only will the number of files be large, but the loading of data will also be slow. Therefore, we use Lightning Memory Mapped Database (LMDB) to store images. LMDB puts the entire dataset in one file, avoiding the overhead of file system addressing. After preprocessing, the images and labels are stored uniformly in the LMDB database.



Figure 8. CASIA-OLHWDB1.0

Because the number of Chinese character categories in the DICC's training set for foreign students is much greater than the number of samples for each category of Chinese character, the method of data augmentation is used to increase the number of samples for each category of character, making the training set much richer and the model more generalizable. Before feeding samples into the network, data augmentation is carried out through methods such as cropping, flipping, and rotation.

C. Implementation details

ResNetDC is a CNN built on the PyTorch deep learning framework, designed to achieve efficient and effective model training through a series of carefully crafted optimization strategies. The detailed parameter configurations are listed in Table 1. The training process of ResNetDC aims to minimize the cross-entropy loss. To achieve rapid convergence and adapt to the learning requirements of different parameters, ResNetDC employs the Adam optimizer. The choice of batch size is a balancing act that needs to be adjusted based on specific hardware conditions and model complexity. A batch size of 256 provides a good balance in most cases, accelerating the training process while avoiding excessive memory consumption. Initially, the learning rate is set to 0.01, which is a relatively high starting value intended to speed up the initial convergence. As training progresses, when the objective loss no longer decreases significantly, the learning rate is multiplied by 0.35. Training is terminated when the model's performance on the validation set no longer shows significant improvement.

D. Experimental results

Figure 9 shows the changes in training loss for networks configured with 2, 3, and 4 Blocks, while Figure 10 presents the corresponding validation accuracy for these three network configurations.

The results show that the ResNetDC configured with four Blocks achieves the lowest training loss and the highest validation accuracy. This performance is attributed to the increased depth provided by stacking more Block modules, which, combined with the residual structure within each Block, effectively mitigates issues of gradient vanishing and explosion. Consequently, this deeper architecture facilitates better feature learning and enhances classification accuracy. Through experiments, it was found that the area of interest of the ResNetDC configured with four Blocks has almost covered the entire Chinese character image. Increasing the number of Block will only increase the size, parameters, and training time of the network, but will not improve classification accuracy and performance. Therefore, in the ResNetDC model, setting up four Block

modules is the most reasonable solution.

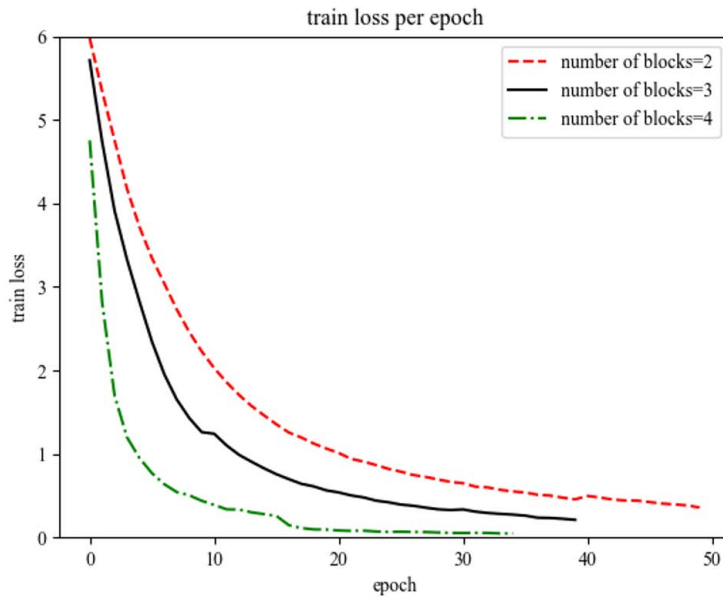


Figure 9. The changes in training loss for networks configured with 2, 3, and 4 Blocks

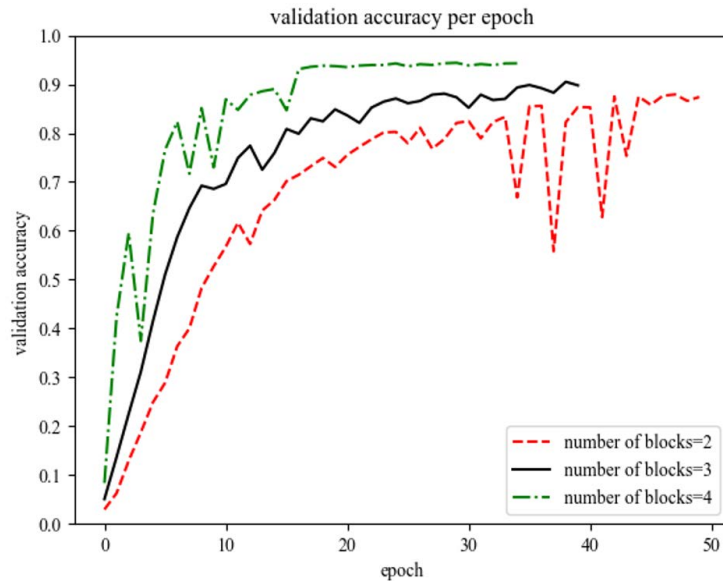


Figure 10. Accuracy comparison for the three networks

To evaluate the effectiveness of dilated convolution in the ResNetDC, a comparison of the classification performance between common convolutional networks and dilated convolutional networks is conducted. In the ResNetDC shown in Figure 2 and Figure 3, the dilation rates of 3×3 convolutional layers of every Block module in a common convolutional network are set to 1, and they are set to 1, 2 and 3 respectively in the dilated convolutional network, Figure 11 shows the training losses of these two networks, and Figure 12 shows their validation accuracy.

Figure 11 and Figure 12 clearly demonstrate the superior performance of the dilated convolutional network compared to the common convolutional network. Specifically, the training loss of the dilated convolutional network is consistently lower, and the validation accuracy stabilizes after approximately 16 epochs, reaching a

significantly higher level than that of the common convolutional network. Its Top 1 accuracy reaches 94.3%, and Top 5 accuracy reaches 98.9%. These results highlight the substantial impact of dilated convolutions on DICCR for foreign students.

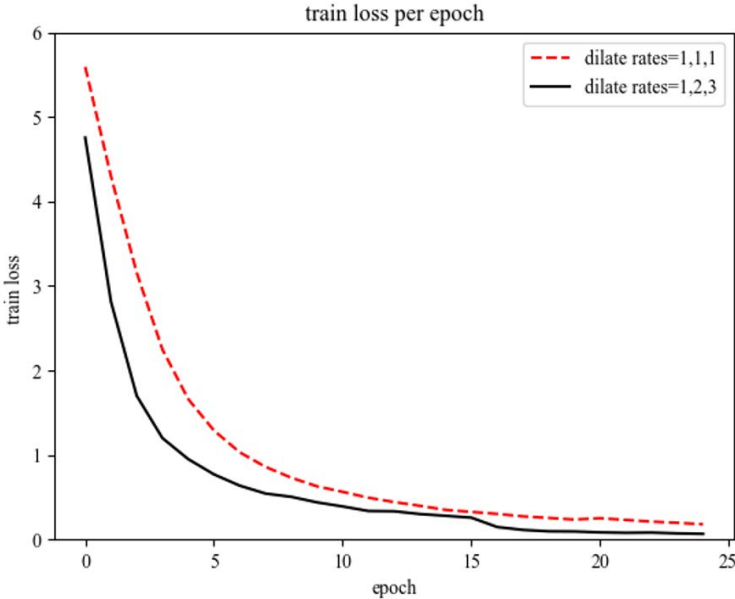


Figure 11. Changes in training loss of two networks

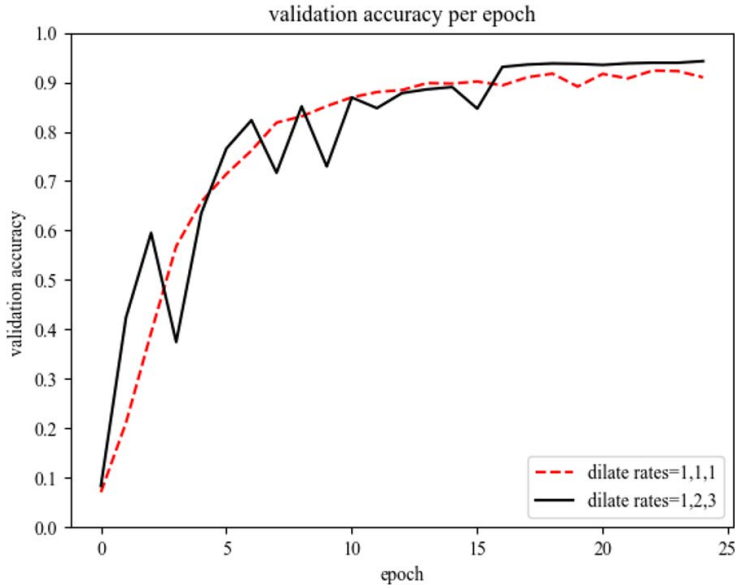


Figure 12. Changes in validation accuracy of two networks

The sample size of DICCs for foreign students is small and imbalanced. By pre-training the ResNetDC on the CASIA-OLHWDB1.0 dataset, we can transfer the knowledge learned from a large dataset to the target task, thereby improving the model’s generalization ability and performance. Additionally, this approach helps avoid overfitting, which is a common issue when training models directly on small and imbalanced datasets. The validation accuracy changes of pre-trained models compared to models without pre-training on the foreign student DICCs dataset are shown in Figure 13.

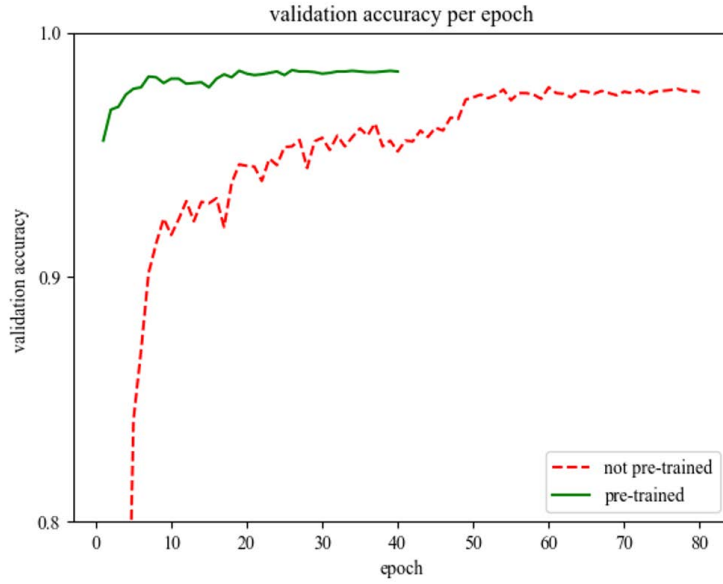


Figure 13. The validation accuracy changes

Our pre-trained model achieves an accuracy of 98.5%, which shows a significant advantage over Hao Bai’s method, as illustrated in the accuracies comparison in Table 3.

Table 3. Accuracies comparison

Approach	Accuracy
Hierarchical model [4]	92.55%
Improved hierarchical models [5]	97.91%
ResNetDC (ours)	98.5%

Table 4. Prediction effect of ResNetDC model on writing errors and non-standard writing samples

Sample	(a)	(b)	(c)	(d)	(e)
Problem	Extra stroke	Missing stroke	Incomplete stroke	Structural imbalance	Structural imbalance
Actual categories	吃	风	块	棋	在
Top 1 prediction	吃	风	块	棋	在
Top 5 prediction	吃气汽 空还	风见几 贝识	块知步 边球	棋快楼 样糕	在石礼 木本

E. Result analysis

Using the trained ResNetDC model, single-sample prediction and batch prediction are achieved for the samples in the test set. The single sample prediction interface is shown in Figure 14. Select a sample by opening a dialog box, and the prediction window displays the sample content, actual category, Top 1 prediction value, and Top 5 prediction value. Batch prediction saves the file names, actual categories, Top 1 predicted values, and Top 5 predicted values of all samples in the test set to a file.

The ResNetDC model can effectively recognize DICCs by foreign students with writing errors and non-standard problems, as shown in Table 4. The Top 1 prediction value is completely correct for samples with extra, missing, incomplete strokes, and structural imbalance. Figure 15 shows some samples correctly recognized by ResNetDC. The ability of ResNetDC to effectively recognize DICCs of foreign students highlights its adaptability, making it a powerful tool for accurate DICCR, despite the challenges posed by various imperfect input data.

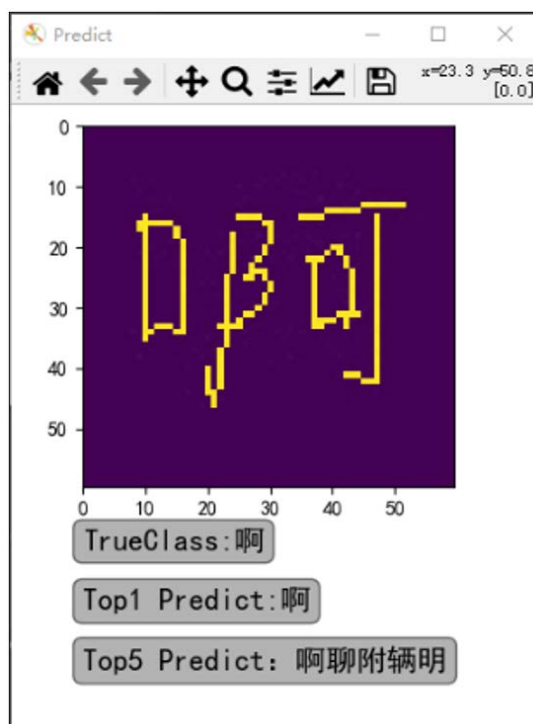


Figure 14. Single sample prediction interface



Figure 15. Samples correctly recognized by ResNetDC

5 Conclusion

For the recognition of foreign students' DICCs, we propose network architecture ResNetDC that learns multi-scale context information, which has the characteristics of easy optimization of the residual network. The ResNetDC overcomes the limitation of traditional convolutional networks, which struggle to capture long-distance correlations in DICCs. By employing varying dilation rates, the model aggregates multi-scale contextual information, ensuring a comprehensive acquisition of both low-level visual clues and high-level structural knowledge. This approach enables the model to achieve excellent classification accuracy, even for DICCs with various errors and non-standard.

The experimental results clearly demonstrate the effectiveness of the proposed ResNetDC model in DICCR for foreign students. This robust performance provides a solid technical foundation for a wide range of applications. For example, Assessing the accuracy and legibility of handwritten Chinese characters, enhancing the learning experience by providing immediate feedback on writing mistakes, assisting foreign students in practicing and improving their handwriting skills, facilitating the recognition and processing of handwritten text in various digital systems, ensuring accurate and efficient evaluation of handwritten responses in online testing environments, and so on.

The ResNetDC model's ability to handle diverse and imperfect input data makes it particularly valuable for these applications, offering significant development prospects and paving the way for advancements in international Chinese language education and technology.

Data Availability

Data used to support research findings are available from relevant authors upon request.

Conflict of Interests

The authors declare that they have no conflicts of interest.

References

- [1] C. L. Liu, F. Yin, D. H. Wang, and Q. F. Wang, "CASIA online and offline Chinese handwriting databases," in *2011 International Conference on Document Analysis and Recognition*, Beijing, China, 2011, pp. 37–41. <https://doi.org/10.1109/ICDAR.2011.17>
- [2] L. W. Jin, Y. Gao, G. Liu, Y. Y. Li, and K. Ding, "SCUT-COUCH2009 — A comprehensive online unconstrained Chinese handwriting database and benchmark evaluation," *Int. J. Doc. Anal. Recognit.*, vol. 14, pp. 53–64, 2011. <https://doi.org/10.1007/s10032-010-0116-6>
- [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, 2017. <https://doi.org/10.1016/j.patcog.2016.08.005>
- [4] H. Bai and X. W. Zhang, "Recognizing Chinese characters in digital ink from non-native language writers using hierarchical models," in *Proceedings of the Second International Workshop on Pattern Recognition*, Singapore, 2017, pp. 45–50. <https://doi.org/10.1117/12.2280237>
- [5] H. Bai and X. W. Zhang, "Improved hierarchical models for non-native Chinese handwriting recognition using hidden conditional random fields," in *Proceedings of the Fifth International Workshop on Pattern Recognition*, Chengdu, China, 2020, pp. 41–45. <https://doi.org/10.1117/12.2574420>
- [6] H. Xu and X. Zhang, "Recognizing digital ink Chinese characters written by international students using

- a residual network with 1-dimensional dilated convolution,” *Information*, vol. 15, no. 9, p. 531, 2024. <https://doi.org/10.3390/info15090531>
- [7] W. Cai, S. Uchida, and H. Sakoe, “Comparative performance analysis of stroke correspondence search methods for stroke-order free online multi-stroke character recognition,” *Front. Comput. Sci.*, vol. 8, pp. 773–784, 2014. <https://doi.org/10.1007/s11704-014-3207-6>
- [8] X. Y. Zhang, F. Yin, Y. M. Zhang, C. L. Liu, and Y. Bengio, “Drawing and recognizing Chinese characters with recurrent neural network,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 849–862, 2018. <https://doi.org/10.1109/TPAMI.2017.2695539>
- [9] J. S. Zhang, Y. X. Zhu, J. Du, and L. R. Dai, “Trajectory-based radical analysis network for online handwritten Chinese character recognition,” in *2018 24th International Conference on Pattern Recognition (ICPR)*, Beijing, China, 2018, pp. 3681–3686. <https://doi.org/10.1109/ICPR.2018.8546074>
- [10] W. X. Yang, L. W. Jin, Z. C. Xie, and Z. Y. Feng, “Improved deep convolutional neural network for online handwritten Chinese character recognition using domain-specific knowledge,” in *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, Tunis, Tunisia, 2015, pp. 551–555. <https://doi.org/10.1109/ICDAR.2015.7333822>
- [11] L. Chen, S. Wang, W. Fan, J. Sun, and S. Naoi, “Beyond human recognition: A CNN-based framework for handwritten character recognition,” in *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, Kuala Lumpur, Malaysia, 2015, pp. 695–699. <https://doi.org/10.1109/ACPR.2015.7486592>
- [12] B. Graham, “Sparse arrays of signatures for online character recognition,” *arXiv preprint arXiv:1308.0371*, 2013. <https://doi.org/10.48550/arXiv.1308.0371>
- [13] C. L. Liu, F. Yin, D. H. Wang, and Q. F. Wang, “Online and offline handwritten Chinese character recognition: Benchmarking on new databases,” *Pattern Recognit.*, vol. 46, no. 1, pp. 155–162, 2013. <https://doi.org/10.1016/j.patcog.2012.06.021>
- [14] H. Q. Ren, W. Q. Wang, and C. L. Liu, “Recognizing online handwritten Chinese characters using RNNs with new computing architectures,” *Pattern Recognit.*, vol. 93, pp. 179–192, 2019. <https://doi.org/10.1016/j.patcog.2019.04.015>
- [15] J. Gan, W. Q. Wang, and K. Lu, “Characters as graphs: Recognizing online handwritten Chinese characters via spatial graph convolutional network,” *arXiv preprint arXiv:2004.09412*, 2020. <https://doi.org/10.48550/arXiv.2004.09412>
- [16] J. Gan, Y. Y. Chen, B. Hu, J. X. Leng, W. Q. Wang, and X. B. Gao, “Characters as graphs: Interpretable handwritten Chinese character recognition via Pyramid Graph Transformer,” *Pattern Recognit.*, vol. 137, p. 109317, 2023. <https://doi.org/10.1016/j.patcog.2023.109317>
- [17] D. Cireşan and U. Meier, “Multi-column deep neural networks for offline handwritten Chinese character classification,” in *2015 International Joint Conference on Neural Networks (IJCNN)*, Killarney, Ireland, 2015, pp. 1–6. <https://doi.org/10.1109/IJCNN.2015.7280516>
- [18] J. Gan, W. Q. Wang, and K. Lu, “A new perspective: Recognizing online handwritten Chinese characters via 1-dimensional CNN,” *Inform. Sci.*, vol. 478, pp. 375–390, 2019. <https://doi.org/10.1016/j.ins.2018.11.035>
- [19] M. J. Hu, X. W. Qu, J. Huang, and X. G. Wu, “An end-to-end classifier based on CNN for in-air handwritten-Chinese-character recognition,” *Appl. Sci.*, vol. 12, no. 14, p. 6862, 2022. <https://doi.org/10.3390/app12146862>
- [20] Z. Y. Zhong, L. W. Jin, and Z. C. Xie, “High performance offline handwritten Chinese charac-

- ter recognition using GoogLeNet and directional feature maps,” in *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, Tunis, Tunisia, 2015, pp. 846–850. <https://doi.org/10.1109/ICDAR.2015.7333881>
- [21] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. Yuille, “Semantic image segmentation with deep convolutional nets and fully connected CRFs,” *arXiv preprint arXiv:1412.7062*, 2016. <https://doi.org/10.48550/arXiv.1412.7062>
- [22] F. Yu and V. Koltun, “Multi-scale context aggregation by dilated convolutions,” *arXiv preprint arXiv:1511.07122*, 2016. <https://doi.org/10.48550/arXiv.1511.07122>
- [23] L. C. Chen, G. Papandreou, F. Schroff, and H. Adam, “Rethinking atrous convolution for semantic image segmentation,” *arXiv preprint arXiv:1706.05587*, 2017. <https://doi.org/10.48550/arXiv.1706.05587>
- [24] C. Q. Zhuang, Z. D. Lu, Y. Q. Wang, J. Xiao, and Y. Wang, “ACDNet: Adaptively combined dilated convolution for monocular panorama depth estimation,” *arXiv preprint arXiv:2112.14440*, 2022. <https://doi.org/10.48550/arXiv.2112.14440>
- [25] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, “SSD: Single shot multibox detector,” in *Computer Vision – ECCV 2016*. Cham: Springer International Publishing, 2016, pp. 21–37. https://doi.org/10.1007/978-3-319-46448-0_2
- [26] K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016, pp. 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- [27] X. H. Ding, X. Y. Zhang, Y. Z. Zhou, J. G. Han, G. G. Ding, and J. Sun, “Scaling up your kernels to 31×31 : Revisiting large kernel design in CNNs,” *arXiv preprint arXiv:2203.06717*, 2022. <https://doi.org/10.48550/arXiv.2203.06717>
- [28] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2015. <https://doi.org/10.48550/arXiv.1409.1556>
- [29] P. Q. Wang, P. F. Chen, Y. Yuan, D. Liu, Z. H. Huang, X. D. Hou, and G. Cottrell, “Understanding convolution for semantic segmentation,” in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Lake Tahoe, NV, USA, 2018, pp. 1451–1460. <https://doi.org/10.1109/WACV.2018.00163>
- [30] H. Bai, “Study on digital ink characters stroke error extraction by beginning learners of Chinese as a foreign language,” Ph.D. thesis, Beijing Language and Culture University, China, 2018.
- [31] H. Bai, X. W. Zhang, Y. G. Fu, and W. H. An, “Adaptive visualization of extracted digital ink characters in Chinese,” *Comput. Eng. Appl.*, vol. 48, no. 15, pp. 153–158, 2012.