A Recognition Method of the Similarity Character for Uchen Script Tibetan Historical Document Based on DNN

Xiaojuan Wang[1], Weilan Wang[1(✉)], Zhenjiang Li[1], Yiqun Wang[1], Yuehui Han[2], and Zhanjun Hao[3]

[1] Key Laboratory of China's Ethnic Languages and Information Technology of Ministry of Education, Northwest Minzu University, Lanzhou, Gansu 730000, China
wangweilan@xbmu.edu.cn
[2] College of Mathematics and Computer Science, Northwest Minzu University, Lanzhou, Gansu 730000, China
[3] College of Computer Science and Engineering, Northwest Normal University, Lanzhou, Gansu, China
[4] Library of Northwest Minzu University, Lanzhou, Gansu 730000, China

Abstract. In order to improve the similarity character recognition of Tibetan historical document, this paper applied the Depth Neural Network (DNN) to similar characters recognition of Tibetan historical document, and proposed a recognition method of the similarity character for Uchen Script Tibetan based on deep learning. The effective feature learning and recognition are automatically carried out by DNN. We also introduced a sample labeling method of Tibetan historical document of Uchen Script using unsupervised clustering and constructing sample sets of the similar characters. Compared with the traditional methods such as Support Vector Machine (SVM) and Naive Bayes Classifier (NBC) based on gradient features through simulation experiment, our method can achieve better performance. The proposed method can learn feature effectively and avoid the disadvantages of manual feature selection and extraction, and it can improve recognition rate greatly. With the increasing of training samples, the recognition rate was improved more significantly. The experimental results show that the proposed method used for similar characters of Tibetan historical document Uchen Script recognition, higher recognition rate can be obtained.

Keywords: Deep neural network (DNN) · Deep learning
Convolutional neural network (CNN) · Tibetan
Similar character of Uchen script

1 Introduction

The characters of Tibetan historical document cover modern Tibetan and Sanskrit Tibetan, so the number of characters is more than 7,000. The similarity between characters of Tibetan historical document is high and there are a lot of similar characters, such as "ཎ", "ཀ", "ཀ", "ཥ", "ཤ", "ད", "ཎ", etc., which bring a larger technical

difficulty to character recognition. In addition, many Tibetan historical documents are carved on the woodblock, which was engraved by hands, so the nicks are usually uneven. Therefore, the late manual inkiness is uneven, for example, the deep groove has less ink, leading to a loss part strokes of character of historical documents; Or a loss of strokes caused by the Image preprocessing of the ancient books, for example, "ད" "ས" "ཅ" are changed into "▰" "▰" "◨", which undoubtedly increases the difficulty of character recognition of Tibetan ancient books. At present, there is a lack of researches on the image and character recognition of Tibetan ancient books.

SVM method [1], hidden Markov model [2] and so on are more widely used in character recognition. Convolution neural network is a deep neural network which has a local connection between layers and which was put forward by American scholar LeCun. After the appearance of convolution neural network (CNN), using a variety of types of deep neural network models to analyze and recognize documents has become a research hotspot in this field. CNN has been successfully used in many areas, such as the recognition of handwritten digits, English characters, Chinese character and so on. Among 107 papers collected in ICFHR meeting held in late October 2016, whose image analysis and retrieval [3], text line segmentation [4], feature extraction [5], classification recognition processing [6] and other links involved in Chinese, English, Japanese, Mongolian, Arab, Bangladesh, etc., and more than half of the papers applied the deep learning technology. The Tibetan language includes modern Tibetan language (also known as Tibetan language or local Tibetan language) and Sanskrit Tibetan language (the Tibetan transferring form of Sanskrit). The print form of modern Tibetan characters has been studied a lot, such as professor Ou Zhu at Tibet University, professor Huang Heming at Qinghai Normal University, professor Li Yongzhong at Jiangsu University of Science and Technology, etc. And the team of professor Ding Xiaoqing at Tsinghua University studied, researched and developed the Tibetan character recognition system of practical multifont printing of more than 592 characters [7, 8], which has been well applied. The literature [9–13] shows that, for handwritten character recognition, the statistical characteristics of characters are the best, and for the off-line handwritten Chinese character recognition, gradient feature has a high recognition rate [14–16]. The researchers successfully applied the convolution neural network to digit recognition [17, 18] and character recognition [19, 20] in the natural scene, and pointed out that the convolution neural network could learn the characteristics which are better than artificial design [21, 22]. The literature [23] applied the deep convolution neural network to the recognition of offline handwritten similar characters, and the recognition rate is more significantly improved than traditional method. Therefore, this thesis proposes to use the deep convolution neural network to conduct the recognition of similar Tibetan characters. In contrast, there is no report about the application of deep convolutional neural network in the character recognition research of Tibetan ancient books.

Due to the irreproducibility of Tibetan ancient books, sample extraction of Tibetan characters can only be extracted from the document and image itself of Tibetan ancient books, and the project team has realized the preprocessing, binarization and layout analysis of document and image of Tibetan ancient books, and completed the document character segmentation. Due to the printing requirement of "soft character fine

alignment and fine carving" in the Phyi dar of Tibetan Buddhism, most of the Buddhist texts adopted Uchen Script. The striking feature of Uchen Script is that the top stroke of each letter is horizontal and straight, and the base line of the character arrangement is on a straight line. See Fig. 1. The baseline (baseline 1, baseline 2, etc. expressed by the dotted line in Fig. 1) is adopted to further segment into the vowel part above the baseline. For example, baseline 1 is adopted to express the character "▨", "◩" and so on above the baseline; The part under the baseline, such as "◪", "◪", "◪", etc. There are fewer types of characters above the baseline, about a dozen types, and there are also fewer types of similar characters. This thesis mainly studies the similar characters of the characters under the baseline.
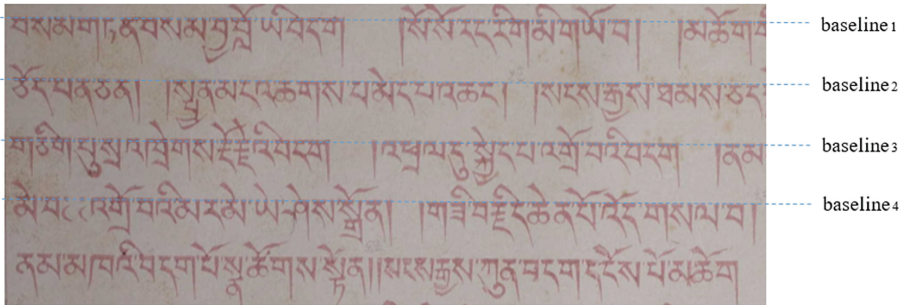

Fig. 1. Document image of Tibetan ancient books (a part)

2 Construct Sample Set of Similar Characters

In view of the current situation that there is no character sample of Tibetan ancient books, the following methods are proposed to classify and label the similar character sets.

In view of the Tibetan characters which have been segmented early, first of all, their characteristics are extracted, and three features about extraction in this paper are:

(1) Gradient 8 direction characteristics (64 D)

First of all, the character image of Tibetan ancient books is normalized to 136×50, and in order to ensure the less distortion of the image, bicubic interpolation is adopted for the deformation process. Then the uniform grid of 4×2 is used to evenly divide the original image into 8 small grids according to the size, and then the gradient feature of character pixels in each small grid is calculated. Then, the gradient is decomposed into 8 directions in accordance with the method of Bai to form 8 D gradient direction characteristics [24], and then 8 small grids features are combined to get 64-dimensional gradient direction characteristics.

(2) Features of 8×8 grid (64 D)

In the first place, the character image of Tibetan ancient books is transformed into 64×64, and in order to ensure a less distortion of the image, the deformation process adopts bicubic interpolation. Then, the original image is evenly divided

into 64 small grids by using the even grid of 8×8, and later, the percentage of the characters in each small grid in the total pixel is calculated, and the characteristics of 64-dimension are obtained.

(3) Peripheral features of characters (64 D)

The grids which are divided and extracted by using feature (2) to continue to extract the pixel periphery features from top to bottom, from bottom to top, from left to right and from right to left. The features of four directions are combined into one-dimensional features, and 64 small grids have a total of 64-D features.

After integrating the above three characteristics, there are a total of 192 D feature dimensions. Through principal component analysis, the dimension is reduced to 80 D features. k-means clustering is used to record the filename of each character and the corresponding relationship of the distance of each centroid. According to the sorting characters in the class, the former k characters which are divided into the same class and which are in a close range are divided into similar characters, constituting a set of similar characters. MATLAB is used to copy the image of similar characters in the same file, and the distance information is added before the image's original file name. Then, according to the sort of file name, the image of the same category of characters can be gathered as far as possible (Fig. 2).
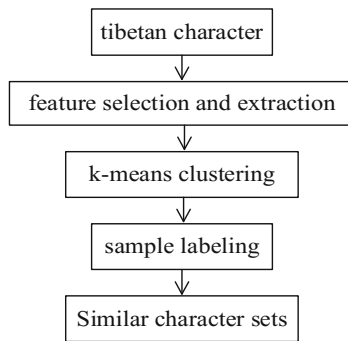
tibetan character

↓

feature selection and extraction

↓

k-means clustering

↓

sample labeling

↓

Similar character sets

Fig. 2. Construction process of similar character set of Tibetan Uchen script

3   Convolution Neutral Network (CNN)

Convolution neural network (CNN) is a neural network which is specially used to deal with similar network structure data, such as image data which can be considered as a two-dimensional pixel grid. CNN shows a high recognition rate in 2 D image recognition application, and its network structure is highly invariant to translation, scaling, tilting or other forms of deformation. CNN directly conducts the learning and character classification for the characteristics of original image, and it doesn't need too much pre-processing and feature extraction of the original character image, so it is an end-to-end recognition system, which effectively avoid the defects of losing the details of similar characters caused by artificial feature extraction and feature selection in advance. This thesis adopts the following CNN network structure, as shown in Fig. 3.
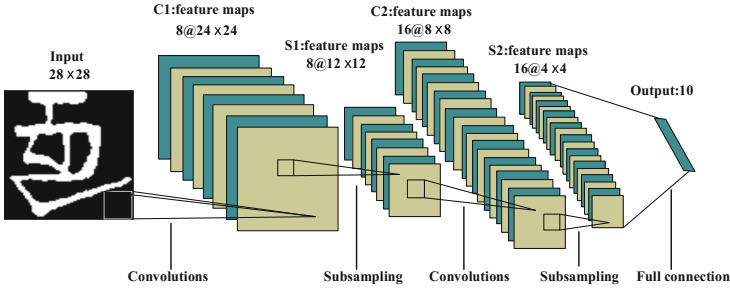
Fig. 3. CNN network structure

Convolution neural network is composed of the convolution layer and the sampling layer, and each layer is composed of multiple feature maps. Each pixel (neuron) of convolution layer is connected with a local area of the upper layer, and it can be viewed as a local feature detector. Each neuron can extract primary visual features such as direction line segments, angular point, etc. At the same time, this local connection makes the network have fewer parameters, which is beneficial to training. There is usually a sampling layer behind the convolution layer, in order to reduce the resolution of the image, and the network have a certain displacement, scaling and distortion invariance. For the convolution layer, the feature graph of the previous layer is conducted with a convolution operation with multiple group of convolution masks and then the feature graph of the layer is obtained through the activation function. The calculation form of the convolution layer is as follows:

$$a_j^l = \sigma\left(\sum\nolimits_{i \in M_j} a_i^{l-1} * w_{ij}^l + b_j^l\right) \tag{1}$$

In Eq. (1), l is the number of layers where the convolution layer is; w is convolution kernel, which is a template of 5×5. b is setover, and σ is activation function, that is $1/(1 + e^{-x})$. M_j represents an input feature graph of the upper layer.

The sampling layer is to sample the characteristics of the upper convolution layer and get the same number of feature graphs. The training of convolution neural network is the same as that of traditional neural network, and it adopts stochastic gradient descent. The input layer is a character image of Tibetan ancient books, whose size is 28×28. C1 layer is the first convolution layer, which has eight feature graphs of 24×24, and one pixel (node or neuron) in each feature graph is interconnected with a region of 5×5 corresponding to the input layer. S1 layer is a lower sampling layer containing 8 feature graphs of 12×12, and each node in the feature graph is interconnected with a region of 2×2 corresponding to the feature graph in the C1 layer. C2 is the second convolution layer with 16 feature graphs, and the size of each feature graph is 8×8. The connection between S1 and C2 plays an important role in feature extraction. S2 is the second sampling layer with 16 feature graphs, and the size of each feature graph is 4×4. The last layer is the output layer with 10 nodes, corresponding to the output category, and it has a full connection with S2 layer.

4   Experiment and Result Analysis

4.1   Experiment Data

In this paper, the experimental data is the two groups of similar characters under the baseline of Tibetan characters, and each group contains 10 Tibetan character categories. The first group is a set of similar characters formed by Tibetan vertical stacks, and it is composed of "▨", "▨", "▨", "▨", "▨", "▨", "▨", "▨", "▨" and "▨". It is represented by G1, and there are a total of 5215 experimental samples.

The second group is a set of similar characters which are composed of complete consonant characters, and it is composed of "▨", "▨", "▨", "▨", "▨", "▨", "▨", "▨", "▨", "▨". It is represented by G2, and there are a total of 24,700 experimental samples.

In order to compare the performance of CNN in the recognition of Tibetan similar characters, CNN is compared with Naive Bayes Discriminant classifier and support vector machine classifier. For Naive Bayes discriminant and SVM classification, first of all, gradient 8 direction features described in Sect. 2 are extracted to get 64 D feature vector of each sample, and then the feature vector is used to discriminate and classify. For CNN, the image of the Tibetan characters is directly compressed to the image with a resolution of 28×28, so as to reduce the parameters of CNN, and thus improve the training speed of the network.

4.2   Experiment Process

In the network training process shown in Fig. 3, the error reverse transform and the gradient random descent method are adopted to update the parameter w and b.

J(w, b) is used to express the error function, and the expression of updating parameters with the gradient descent method is as follows:

$$w := w - \alpha \frac{\partial J(w,b)}{\partial w} \tag{2}$$

$$b := b - \alpha \frac{\partial J(w,b)}{\partial b} \tag{3}$$

α is the descent rate control parameter, and the selection of α in the experiment is determined by adopting the test method. Finally, selecting $\alpha = 1.5$ as the descent rate parameter of the system.

In order to observe the influence of different α on recognition rate, first of all, other parameters are fixed, for example, the times of circuit training are 30, because smaller number of circuit training times can save the training time, but it is enough to reflect the impact of α on the recognition rate. Different α and corresponding identification error rate are shown in Table 1.

The value of α during the experimental process is conducted according to the order from top to bottom in Table 1. The error rate in Table 1 shows that the error rate is the smallest when $\alpha = 1.5$, and it is 0.2339.

Table 1. Different α and corresponding recognition error rate

α	Error rate
0.01	0.7440
0.25	0.7440
0.6	0.3706
0.9	0.2817
1.5	**0.2339**
2	0.2798
1.8	0.2716
1.6	0.2651
1.4	0.2679

4.3   Experimental Results and Analysis

The experiment adopts CNN network structure shown in Fig. 3 and uses 64 D gradient feature to conduct Naive Bayes and SVM classification. In this paper, G1 and G2 sets are conducted with K-fold cross validation (K = 10), namely, each similar set is evenly divided into 10 parts: T1, T2, T3…… T10. Each part is taken as a test set each time, and the other 9 parts are regarded as the training set. The error rate results of G1 and G2 sets are shown in Tables 2 and 3 respectively. The experimental results show that, compared with Naive Bayes and SVM recognition method, the method based on deep neural network has a lower error rate. The reason for the poor performance of SVM and Naive Bayes is that the identification information of similar Tibetan characters is lost in the process of feature extraction.

Table 2. A comparison of error rate of 10-fold cross-validation on G1 set

Classifier	NBC	SVM	CNN
Error rate of T1	0.1288	0.0250	**0.0212**
Error rate of T2	0.1288	0.0327	**0.0192**
Error rate of T3	0.1308	0.0308	**0.0269**
Error rate of T4	0.1288	0.0327	**0.0173**
Error rate of T5	0.1385	0.0423	**0.0154**
Error rate of T6	0.1115	0.0365	**0.0231**
Error rate of T7	0.1385	0.0212	**0.0154**
Error rate of T8	0.1212	0.0250	**0.0154**
Error rate of T9	0.1231	0.0346	**0.0192**
Error rate of T10	0.1654	0.0404	**0.0231**
Average error rate	0.1315	0.0321	**0.0196**

The experimental results show that, compared with Naive Bayes and SVM recognition method, the method based on deep neural network has a lower error rate.

Table 3. A comparison of error rate of 10-fold cross-validation on G2 set

Classifier	NBC	SVM	CNN
Error rate of T1	0.0526	0.0158	**0.0117**
Error rate of T2	0.0530	0.0154	**0.0134**
Error rate of T3	0.0453	0.0109	**0.0097**
Error rate of T4	0.0555	0.0170	**0.0134**
Error rate of T5	0.0951	0.0146	**0.0121**
Error rate of T6	0.0632	0.0166	**0.0162**
Error rate of T7	0.0567	0.0142	**0.0105**
Error rate of T8	0.0551	0.0153	**0.0109**
Error rate of T9	0.0579	**0.0117**	0.0130
Error rate of T10	0.0587	0.0178	**0.0117**
Average error rate	0.0593	0.0149	**0.0123**

The reason for the poor performance of SVM and Naive Bayes is that the identification information of similar Tibetan characters is lost in the process of feature extraction.

In order to illustrate the recognition performance of this paper method, The average error rate comparison of different classifiers on G1 and G2 sets is shown in Fig. 4.
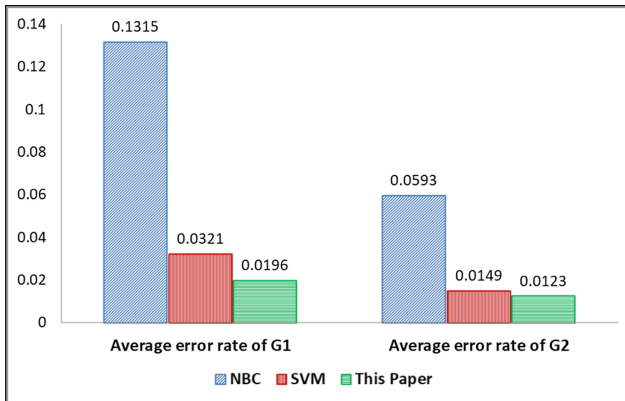

Fig. 4. Error rate of different classifiers on G1 and G2 set

Figure 4 shows this paper's method does not need human intervention in the process of training and recognition, is a kind of end-to-end approach, as well as under the condition of less training samples to achieve ideal effect.

Figures 5 and 6 shows the error curve of T10 of G1 and T10 of G2. It can be seen that CNN has smaller error in similar character recognition with the increase of the iterations.
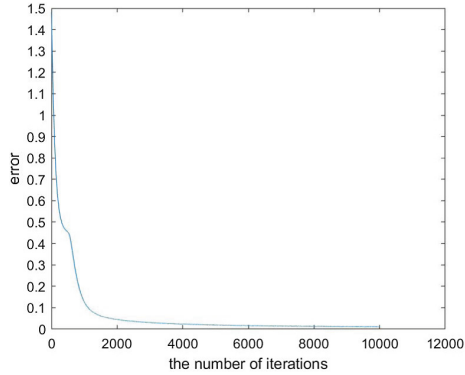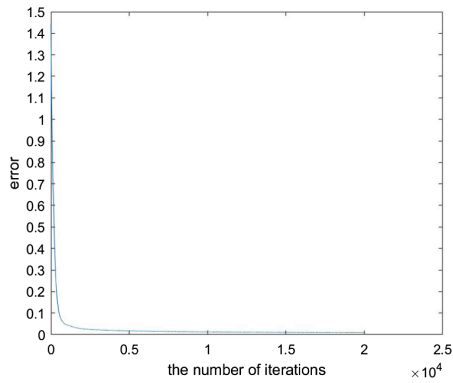
Fig. 5. T10 of G1 error curve

Fig. 6. T10 of G2 error curve

To further the robustness and stability of the network, In this paper randomly selects 1/10 of the sample from category of G2 set to form the test sample set (Te), and the number of test set is 2,470. In addition, it randomly selects five training sample sets (Tr1, Tr2, Tr3, Tr4 and Tr5) which doesn't include the test sample of Te, and the size are 1, 2 times, 3 times, 5 times and 9 times of test sample respectively, and The number of training sample sets is 2470, 4940, 7410, 12350, and 22230. The recognition error rate of these five sets of data is shown in Table 4.

Table 4. A comparison of error rate of different training samples in G2 set

Classifier	NBC	SVM	CNN
Error rate of Tr1-Te (2470-2470)	0.0628	0.0190	0.0510
Error rate of Tr2-Te (4940-2470)	0.0789	0.0202	0.0255
Error rate of Tr3-Te (7410-2470)	0.0846	0.0182	0.0227
Error rate of Tr4-Te (12350-2470)	0.0494	0.0153	0.0166
Error rate of Tr5-Te (22230-2470)	0.0526	0.0158	**0.0146**

Table 4 shows that with the increase of the sample size, the error rate of the recognition method based on the deep neural network gradually decreases, but the error rate of NBC and SVM method fluctuates up and down. It's clear that the network is more stable for the different sample collection, and the system has more robust robustness.

5   Conclusion

This thesis proposes that using convolution neural network to automatically learn and recognize the characteristics of similar characters of Uchen Script in Tibetan ancient books. At the same time, the similar characters of Tibetan ancient books constructed in this paper are adopted to train the model parameters, and the experimental results show that, compared with the traditional methods: (1) Deep convolution neural network can automatically learn the effective features and identify them from the pixel level, which avoids losing details caused by artificial selection and extraction of features and improves the recognition rate; (2) With the increase of the number of training samples, deep convolution neural network has a remarkable performance in reducing the error recognition rate, and the increase of training samples has an obvious effect on enhancing the recognition rate of deep neural network.

References

1. Gaur, A., Yadav, S.: Handwritten Hindi character recognition using k-means clustering and SVM. In: International Symposium on Emerging Trends and Technologies in Libraries and Information Services, pp. 65–70. IEEE (2015)
2. Sharma, A., Kumar, R., Sharma, R.K.: HMM based online handwritten Gurmukhi character recognition. Mach. Graph. Vis. **19**(4), 439–449 (2010)
3. Sudholt, S., Fink, G.A.: PHOCNet: a deep convolutional neural network for word spotting in handwritten documents. In: 15th ICFHR, pp. 277–282 (2016)
4. Moysset, B., Louradour, J., Kermorvant, C., Wolf, C.: Learning text-line localization with shared and local regression neural networks. In: 15th ICFHR, pp. 1–6 (2016)
5. Krishnan, P., Dutta, K., Jawahar, C.V.: Deep feature embedding for accurate recognition and retrieval of handwritten text. In: 15th ICFHR, pp. 289–294 (2016)
6. Sun, Z., Jin, L., Xie, Z., Feng, Z., Zhang, S.: Convolutional multi-directional recurrent network for offline handwritten text recognition. In: 15th ICFHR, pp. 240–245 (2016)
7. Wang, W., Ding, X., Chen, L., Wang, H.: Research on modern Tibetan language recognition in print. Comput. Eng. **29**(3), 37–39 (2003)
8. Pan, W.S., Jin, L.W., Feng, Z.Y.: Recognition of Chinese characters based on multiscale gradient and deep neural network. J. Beijing Univ. Aeronaut. Astronaut. **41**(4), 751–756 (2015)

9. Chen, K., Seuret, M., Wei, H., Liwicki, M., Hennebert, J., et al.: Ground truth model, tool, and dataset for layout analysis of historical documents. In: Proceedings of SPIE-IS&T, vol. 9402 940204-2. http://proceedings.spiedigitallibrary.org. Accessed 19 May 2015

10. Wei, H., Chen, K., Ingold, R., et al.: Hybrid feature selection for historical document layout analysis. In: 14th International Conference on Frontiers in Handwriting Recognition (ICFHR), pp. 87–92. IEEE (2015)

11. Likforman-Sulem, L., Zahour, A., Taconet, B.: Text line segmentation of historical documents: a survey. Int. J. Doc. Anal. Recognit. 9(2), 123–138 (2007)

12. Kesiman, M.W.A., Valy, D., Burie, J.C., Paulus, E., Sunarya, I.M.G.: Southeast Asian palm leaf manuscript images: a review of handwritten text line segmentation methods and new challenges. J. Electron. Imaging 26(1), 1–15 (2017)

13. Xiao, X., Yang, Y., Ahmad, T., Jin, L., Chang, T.: Design of a very compact CNN classifier for online handwritten Chinese character recognition using DropWeight and global pooling. In: ICDAR (2017)

14. Le, A.D., Nakagawa, M.: Training an end-to-end system for handwritten mathematical expression recognition by generated patterns. In: ICDAR (2017)

15. Wu, Y.-C., Yin, F., Chen, Z., Liu, C.-L.: Handwritten Chinese text recognition using separable multi-dimensional recurrent neural network. In: ICDAR (2017)

16. LeCun, Y., Boser, B., Denker, J.S., et al.: Handwritten digit recognition with a back-propagation network. In: Advances in Neural Information Processing Systems, Denver, United States, pp. 396–404 (1990)

17. Netzer, Y., Wang, T., Coates, A., et al.: Reading digits in natural images with unsupervised feature learning. In: NIPS Workshop on Deep Learning and Unsupervised Feature Learning, Granada, Spain (2011)

18. Sermanet, P., Chintala, S., LeCun, Y.: Convolutional neural networks applied to house numbers digit classification. In: Proceedings of IEEE International Conference on Pattern Recognition, Tsukuba, Japan, pp. 3288–3291 (2012)

19. Coates, A., Carpenter, B., Case, C., et al.: Text detection and character recognition in scene images with unsupervised feature learning. In: Proceedings of IEEE International Conference on Document Analysis and Recognition, Beijing, China, pp. 440–445 (2011)

20. Wang, T., Wu, D.J., Coates, A., et al.: End-to-end text recognition with convolutional neural networks. In: Proceedings of IEEE International Conference on Pattern Recognition, Tsukuba, Japan, pp. 3304–3308 (2012)

21. Jin, L., Zhong, Z., Yang, Z., et al.: Application of deep learning in handwritten Chinese character recognition. J. Automat. 42(8), 1125–1141 (2016)

22. Liu, C.L.: Normalization-cooperated gradient feature extraction for handwritten character recognition. IEEE Trans. Pattern Anal. Mach. Intell. 29(8), 1465–1469 (2007)

23. Zhao, Y., Tao, D., Zhang, S., et al.: Similar Chinese character recognition based on deep neural network under big data. J. Commun. 321(9), 184–189 (2014)

24. Bai, Z.L., Huo, Q.: A study on the use of 8-directional features for online handwritten Chinese character recognition. In: Proceedings of the 8th International Conference on Document Analysis and Recognition, pp. 262–266. IEEE, Seoul (2005)