Study on Text Irregular Image Algorithm Based on Convolutional Neural Network

Aisi Luo, Zeyu Jia, Hongyu Fu, Sheng Han

Hebei University of Science & Technology, Yuxiang Street, Shijiazhuang, 050024, China

Abstract: In view of character defects due to irregular interference of text images, this paper proposed a restoration algorithm model based on convolutional neural network and key point detection, implements restoration training on the defective character areas. By studying and analyzing the characteristics of ancient text images of different styles, a digital text image database was established. Based on the Chinese character regional positioning technology and key point detection technology, the restoration effect was evaluated and the training test was conducted after restoration of defective strokes.

1. Introduction

With the development of imaging equipment and science and technology, cameras now possess extremely high image acquisition capability. If the document image contains only text image, the recognition efficiency will be greatly improved, especially with the application of OCR (Optical Character Recognition) technology in printed text. Nevertheless, with the passage of time, some inscriptions are seriously damaged in circulation due to improper preservation, natural weathering, rain erosion and other factors, making it difficult to recognize text and image. Traditional manual restoration method has complicated and time-consuming process, which causes problems in archeologists' information investigation. With the development of science and technology, science and technology are applied in document recovery. However, the current technology has difficulty in distinguishing the annotated information and ancient documents marked by ancient people, so recognition often encounters interferences, with recognition success rate greatly declined. In order to solve the problem of character defects caused by irregular interference from text images [1], this paper proposed a restoration algorithm model based on convolutional neural network and key point detection, and carried out restoration training on the defective character areas to further deepen the theoretical and practical research on the restoration of inscriptions and other text works.

2. Convolutional Neural Network and Three-Dimensional Point Cloud

2.1. Convolutional Neural Network

Convolutional network, also known as convolutional neural network (CNN), is a neural network specially designed to process data with a grid-like structure. Usually there are two parameters, x and

w. The input is x and the output is a specific mapping. In the specific operation, some high dimension arrays are generally input. It is possible to sum up a finite number of arrays to replace summation of infinite number of arrays. The correspondingly required two-dimensional kernel K is:

$$s(i, j) = (I * k)(i, j) = \sum_m \sum_n I(m, n)K(i - m, j - n) \tag{1}$$

Convolution is commutative and can be written equivalently as:

$$S(i, j) = (K * I)(i, j) = \sum_m \sum_n I(i - m, j - n)K(m, n) \tag{2}$$

2.2. Three-Dimensional Point Cloud Technology

The two key steps in 3D point cloud target extraction are feature extraction and selection, classification. Is the whole process similar to the target recognition of image? Essentially, the method flow is roughly the same when it comes to target recognition. The target for recognition is generally in a large scenario with a variety of mixed targets. Since there is need to recognize a target, certainly an index or value is needed to maximize the difference between different targets. This index or value is the so-called target features. Hence, when recognizing the target, we often need to adopt the features appropriate for the target. CNN features are mastered by learning, and feature representation and classifier are jointly optimized. Classifiers include SVM, boosting, decision tree and so on. Where, 2D three-dimensional point clouds have such geometric features as area, radius and point density. Compared with 2D, the geometric features of 3D point cloud have two more items: elevation difference and elevation standard deviation. The statistical graph can be divided into point feature histogram (PFH), fast point feature histogram (FPFH) and viewpoint feature histogram (VFH).

2.3. Stone Flower Restoration Technology

Stroke extraction of Chinese characters based on triangular network [2] aims to split the Chinese characters extracted from inscription images into strokes and connecting components used to restore damaged Chinese characters later. The purpose of stroke extraction is to eliminate the loss of stroke continuity information caused by cross connection of strokes in Chinese characters. Among the stroke extraction methods in this paper, we mainly analyze "X-shaped" crossing and "T-shaped" crossing of strokes.

In this paper, we adopt a stone flower restoration technique based on deep learning and stroke decomposition, in which the recovery of Chinese characters was focused on the stroke level. First, strokes of complete Chinese characters extracted from the image were extracted to form a set of stroke templates for restoration. Then, similar strokes of the damaged strokes were searched from the template set by matching to restore the damaged strokes and remove the stone flowers.

2.4. Font Generation Technology

Firstly, the collected inscriptions were preprocessed, converted to point cloud data, and denoised by binarization [3] to form clear images. Secondly, the image of inscriptions was segmented based on the regional positioning technology of Chinese characters in AreaVoronoi diagram, the fuzzy part of the text image was restored using stone flower restoration technology, and the stroke extraction of Chinese characters was performed based on triangular grid [4] .In addition, different calligraphy font types were classified and processed to establish a characteristic font library meeting

the needs of different population. Inscriptions are of great practical value, especially in today with advanced science and technology. Its practical value is reflected in the fact that, after outline refinement, it can be edited into bitmap or vector map, and at the same time, it can be used in different forms in various works. Pictures can be reprocessed to form character inheritance of character calligraphy characteristics. Also, the two-dimensional outline of stone inscriptions can be made into TTF font file, which can be applied in different works of computer like other fonts, thus demonstrating high artistic value.

3. Text Image Restoration Model

3.1. Convolutional Neural Network

Two-dimensional convolution was carried out on the convolution layer to extract features from the local neighborhood on the feature map of the previous layer. Additive bias was then applied and the result was conveyed through an S-type function. Formally, it is the value of the unit at position ðx. In the j th feature map of the i th layer, it is the value of a unit at the position (x, y) of the first feature map on the second layer, which is the deviation of this feature map. m indexes were implemented on the feature map set of the (i-1) layer connected to the current feature map, which was the value of the kernel connected to the k th layer at the position (p, q), the height and width of the kernel, respectively. In order to strengthen the invariance of the distortion on the input and weaken the resolution of the feature map, the feature map of the previous layer was concentrated into the local area in the subsampling layer of the experiment.

input -- >H1: The input of the neural network was 7 consecutive frames with the size of 60*40, and 7 frames were fed with hard wired kernels to yield 5 different features: Gray level, gradient in x direction, gradient in y direction, optical flow in x direction and optical flow in y direction. Information of the first three channels can be directly obtained by operation on each frame respectively, while the extraction of the latter optical flow (x, y) requires information of two frames. Thus, the number of feature maps in layer H1 is as follows: (7+7+7+6+6=33), the size of feature maps is still 60*40. H1 -- >C2: Two 7*7*3 3D convolution kernels were used for convolution of 5 channels respectively to yield two series. Each series has 5 channels (7*7 means space dimension, 3 means time dimension. Namely, 3 frames of image are processed each time). C2 layer of two different convolution kernels are used.

3.2. Matching Algorithm

Obviously, the main diagonal in matrix A has zero elements. This description contains both local information and global information. In addition, when the sampling point representing the contour changes, that is, when the starting point changes as in $C_i=\{p_k, p_{k+1}, \cdots, p_n, p_1, \cdots, p_{k-1}\}$, the corresponding matrix becomes:

$$A = \begin{bmatrix} a_{kk} & \cdots & a_{k1} & \cdots & a_{k(k-1)} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{1k} & \cdots & a_{11} & \cdots & a_{1(k-1)} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{(k-1)k} & \cdots & a_{(k-1)1} & \cdots & a_{(k-1)(k-1)} \end{bmatrix}$$

$$(3)$$

The kth point is the starting point of the contour sequence. In order to find the partial matching segments of the two outlines R1 and R2, their description matrices A1 and A2 should be compared,

which have dimensions M×M and N×N, respectively. Suppose M≤N. The main purpose of partial matching is to recognize similar parts of two shapes. Comparing two description matrices equals to finding a block of elements of size r×r starting with the elements A1(s, s) and A2(m, m) on the main diagonal of the two matrices. The two element blocks should make the difference Dα(s, m, r) minimal. Where, Dα(s, m, r) is defined as:

$$D_a(s,m,r) = \frac{1}{r^2} \sum_{i=0}^{r-1} \sum_{j=0}^{r-1} \left[A_1(s+i, s+j) - A_2(m+i, m+j) \right]^2 \qquad (4)$$

In order to calculate all possible similar values {s, m, r}, integral image is adopted for acceleration, with N M×M integral images Int1···IntN established to represent N different description matrices. MnD is defined as:

$$M_D^n = A_1(1:M, 1:M) - A_2(n:n+M-1, n:n+M-1) \qquad (5)$$

Different MND represent different possible matches in N. Using the integral image, the difference of all blocks, Dα(s, m, r), can be calculated within constant time. The final result of the calculation will yield the difference of all the ternary {s, m, r} groups.

4. Experimental Results and Analysis

The techniques used in this paper mainly include convolutional neural network, stone flower restoration algorithm and font generation technology.

4.1. Convolutional Neural Network

In 2D central neural network, 2D convolution was performed on the convolutional layer to extract features from the local neighborhood on the feature map of the previous layer. Additive bias was then applied, with the result passed through an S-type function. Formally, it is the value of the unit at position ðx; In the j th feature mapping of the i layer, it is formally a unit value at the position (x, y) of the first feature map in the second layer, which is denoted as i, j, x and y, respectively. Different values are compared with different convolutions to yield result after a total of 50 rounds, with the recognition accuracy by convolution increased by about 20.13%.

4.2. Stone Flower Restoration Algorithm

The contour-based stroke extraction algorithm presents a good effect on Song typeface and black typeface, but with poor effect on regular script and official script. Nonetheless, Chinese character stroke extraction based on triangular grid has the advantages of stronger anti-noise, recovery of missing contour, better prediction of Chinese character stroke, etc. Compared with the above three methods, there are more advantages and less disadvantages. As a result, we adopted the Chinese character stroke extraction method based on triangular grid, which can improve the document recognition accuracy.

4.3. Font Generation Technology

During key point detection experiment, fuzzy areas of strokes would appear, which would greatly affect stroke extraction. Hence, we adopted the connected triangle recognition technology for deletion to improve data availability. Six cases were selected: 1:1, 1:1.5, 1:2, 1:2.5,1:3 and 1:3.5. The test results showed that the effect was better under the length-width ratio of 1:3.

A single gray image [5] was trained, and the data resulting from the training was taken as the initial value in subsequent experiment. The training lasted 3000 times with a total of 15 rounds, LOSS=0.846, PSNR=21.1982dB. Then, the text subject to different degrees of interference was mixed for 60 rounds of experiment, and the normalized data through BN layer was processed 9000 times per round, with a total of 50 rounds [6]. Without BN, the training was 6000 times per round, with a total of 15 rounds. After the algorithm processing, the final correct recognition rate of restoration was up by 29.99% [7]. The experiment showed that the text components restored by the model herein were effective in improving the recognition rate in image text restoration. In the stroke extraction method, the triangular grid was used to represent Chinese characters, and the fuzzy region of the stroke was detected according to the triangle features in the triangular grid, with good experimental results achieved. The undamaged Chinese characters from the same inscription image or the same author's inscriptions were split into strokes, and the resulting strokes and connecting components contained in the Chinese characters together form a stroke component set M for restoring the damaged Chinese characters. When restoring damaged Chinese characters, similar strokes or connecting components of damaged Chinese characters were found from the stroke set [8].

In this paper, the result of stroke extraction was used as the sample template needed for Chinese character restoration to replace the damaged stroke in the damaged Chinese characters, so as to achieve the purpose of restoring Chinese characters, and good experimental results were also achieved.

5. Conclusion

Based on convolutional neural network, this paper employed the regional positioning technology of Chinese characters based on AreaVoronoi diagramand the stroke extraction method of Chinese characters based on triangular grid to solve the problem of character mutilation caused by irregular interference which greatly reduced the accuracy of document character recognition. Meanwhile, a large number of image data of inscriptions were trained and good results were achieved, which provided a feasible way for document restoration. In this paper, the contaminated or defective part of the text was predicted based on the region segmentation of the text image, stroke direction and shape of the font. The highest PSNR was 30.23dB, the highest SSIM was 0.875, and the best LOSS was 0.016. The trained interference image and restored image test results were loaded into Baidu OCR for testing, finding that the correct recognition rate of restored images was increased by 31.21%. When different kinds of Chinese characters were used for training, PSNR reached 28.92dB and SSIM reached 0.901. The experiment showed that the model had good effect on the damaged images with different damage degrees and different styles of ancient Chinese characters.

References

[1] Xue Wenjuan. The significance of stone tablets to the study of local history and culture. World of Antiquity, 2014 (01): 64-65+49.

[2] Ilya Blayvas, Alfred M. Bruckstein, and Ron Kimmel. Efficient computation of adaptivethreshold surfaces forimage binarization In 2001 IEEEComputerSociety Conference on Computer Vision andPattern Recognition (CVPR 2001) 2001: 737-742.

[3] Yibing Yang and Hong Yan. An adaptive logical method for binarization of degradeddocument images Pattern Recognition, 2000, 33 (5): 787–807.

[4] Moon-Soo Chang, Sun-Mee Kang, Woo-Sik Rho, Heok-Gu Kim, and Duck-Jin Kim. Im-proved binarization algorithm for document image by histogram and edge detection.In ICDAR. 1995: 636-639.

[5] Xiangyun Ye and Mohamed Cheriet and Ching Y. Suen. Stroke-model-based characterextraction from gray-level document images. IEEETransactions on Image Processing, 2001, 10 (8): 1152–1161.

[6] Junsong Zhang, Guohong Mao, Hongwei Lin, Jinhui Yu, and Changle Zhou. Outlinefont generating from images of ancient chinese calligraphy. T. Edutainment, 2011, 5: 122-131.

[7] Zhang Galun. Let cultural heritage be "immortalized" under digital technology. Science & Technology Daily, 2021-09-17 (008).

[8] SONG Junhua. Some thoughts on the digital protection of intangible cultural heritage. Cultural Heritage, 2015 (02): 1-8+157.