

Document image segmentation into text, continuous-tone and screened-halftone region by the neural networks

Kazuaki Nakamura and Shinji Yamamoto
Toyohashi Univ. of Tech.

Tetsuya Itoh
Minolta Co., Ltd.

Abstract

In this paper we propose the image segmentation method by the neural networks (NN) to extract text, continuous-tone and screened-halftone region in the document image. Each feature extractor for text, continuous-tone and screened-halftone is composed of the 3-layer NN which is upper part of 5-layer hourglass model NN; each hourglass model NN is trained by different learning samples which specify its own characteristics. We call these extractors as the Feature Neural Networks (FNNs). The classification stage is composed of another 3-layer NN whose inputs are applied from feature extractors. We call this as the Classification Neural Network (CNN). Experimental results show this method successful to segment text, continuous-tone and screened-halftone region.

1 Introduction

Common documents are composed of some segments of images that have different attributes; e.g., characters and symbols generally referred as texts, continuous-tone as photographs and screened-halftone with various screen rulings. Recently, digital copiers appeared in the market and it became possible to process images to improve the image quality. The best way to get high-fidelity copies of such composite documents is to process an image adequately to the attribute for the region in a page, which is automatically segmented by the pre-processing. However, it is common to apply the specified image processing to the whole page by a user-selected mode (i.e., texts, photographs and texts/photos) in practical digital copiers.

Several methods have been investigated, which could discriminate the three different regions (i.e., texts, continuous-tone and screened-halftone) from the scanned images by image scanner using the known statistical characteristics (feature parameters) generally related to each image attributes [1][2]. But, finding the sensitive parameters to the three regions should be done through the experience or trial and error (heuristic approach). It must be a time-consuming task. Moreover, the threshold of the discriminator based on such parameters should be also determined by the experience or trial and error.

The purpose of this paper is to propose a new image segmentation method using an artificial neural network (NN) to solve this conventional problem, which will discriminate the image attributes such as text, continuous-tone and screened-halftone and segment the image regions automatically. This method is characterized by two processing stage, feature extraction stage and classification stage, as follows;

a) Feature extraction stage:
A hourglass model NN (5 layers) is known to concentrate the input information at the 3rd layer, if the learning is made to get the output equivalent to the input (identity mapping)[4][8]. So we used the first part (1 to 3 layer) of the hourglass model NN as the feature extraction stage.

b) Classification stage:
At the classification stage, another NN is used which discriminates the image attributes from the outputs of feature extraction NNs. As a post-processing, we applied Hough-transformation to the result map to get the appropriate image-region map.

In the following second chapter, details of our segmentation algorithm are explained. In the third chapter, experimental results are shown for the gray-scale images scanned by a 400 dpi / 8 bits scanner.

2 Segmentation Algorithm

2.1 Outline of Segmentation algorithm

A flow chart employed in this segmentation system is shown in Figure 1. In this figure, "Feature Extraction Learning stage", "Segmentation Learning stage", and final "Segmentation stage" are shown in every vertical group respectively. An arrow with the dotted line means to use same NN after a training is finished with fixed weight coefficient.

First, in Feature Extraction Learning stage, five-layered hourglass model NNs are prepared for five kinds of images. They are learned the identity mapping using different training samples that represent one of the five kinds of images. After

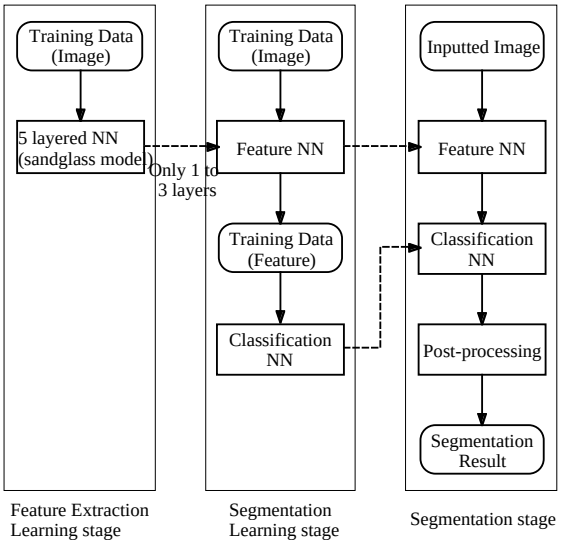

Figure 1: Flow diagram of segmentation using NN.

this training is finished, hourglass model NNs are divided into two parts: NNs of the first half parts in five layers (from first to third layer) are used as Feature Extraction NN (FNNs).

Next, in Segmentation Learning stage, Classification NN (CNN) is connected to FNNs. CNN is learned the segmentation, after training image data are changed into feature data by FNNs. However, the weight coefficients of FNNs are fixed at this time. Finally, in Segmentation stage, a segmentation result for an input image is obtained by those NNs.

With the sequence shown in the above, the segmentation algorithm is carried out separately at every pixel, so it does not consider how the segmentation results of the neighboring pixels are. Therefore, a post-processing that checks the interrelation of the neighboring pixels is added to the system, and it modifies segmentation results.

2.2 Feature Extraction NNs

We used the Multiple Layered Perceptron[7], which is well known that it can compose various mappings with the optional precision with Error Back-propagation algorithm[3]. Especially, when the number of neurons in an input-layer are same as in an output-layer and the number of neurons in a hidden-layer are smaller than in a input- / output-layer, the NN is called "hourglass model NN". It can learn an identity mapping by giving the same data as the input for training data.

It is considered that a mapping to the low dimension space (in other word, the compression of the information) is carried out in the first half part of the NN (from the input-layer to the hidden-layer located the center of this NN's architecture). In the same way, it is considered that the reconstruction is carried out in the latter half part (from the hidden-layer located the center to the output-layer)[4][5]. If such a NN can be trained well, the feature parameter is considered to appear in the hidden-layer located the center, which shows the characteristic of the input data suitably. It means that the feature extraction is carried out automatically without depending on the experience.

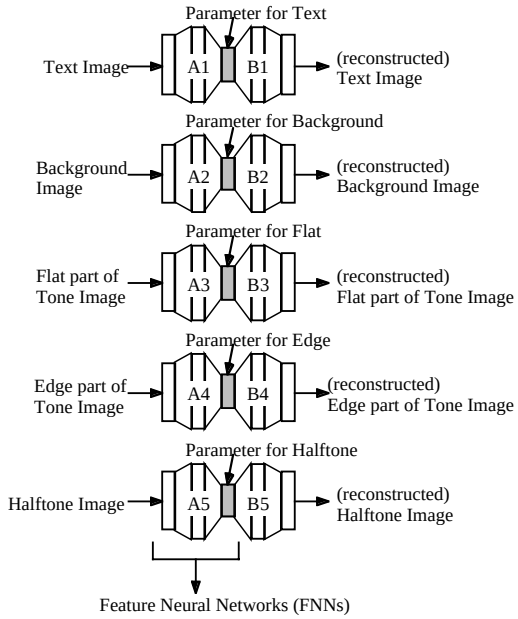

Figure 2: NN structure for the feature extraction (FNNs).

Figure 2 illustrates the network layout. First, five kinds of five layered hourglass model NNs, which are composed with network-As and -Bs, are prepared separately for five image attributes: texts, background, the flat and edge parts in the continuous-tone, and screened-halftone images. Then, training samples are also prepared respectively for those NNs. For example, only typical text images are prepared for one NN, and this NN learns the identity mapping. When this training is finished the feature parameters that correspond to text image are collected intensively into the third layer's neurons. Other NNs are the same too: the training and extraction of the feature parameters of each image are carried out automatically. From now on, NNs that are from the first to third layers in those are used for the feature extraction; they are called Feature Extraction NNs (FNNs).

By the way, we segregated between the flat and edge parts in the same continuous-tone image, and between the text and background parts in the same text image. Because it is considered that their characters were very different even if they were same image region.

2.3 Classification NN

The classification stage is composed of another 3-layer NN (CNN) whose inputs are applied from feature extractors as shown in figure 3. In other words, feature parameters that are extracted by each FNN, which corresponds to feature extractor of text, background, flat or edge part of continuous-tone, or screened-halftone, are applied to the CNN. Output of CNN is composed of five outputs corresponding to the five regions.

In this stage, all kinds of input images including five regions are used as the training sets of CNN. For instance, in the case of text input image, output of A1 will be strongly excited, but outputs of other FNNs will be weakly excited. Using such information, CNN training is performed.

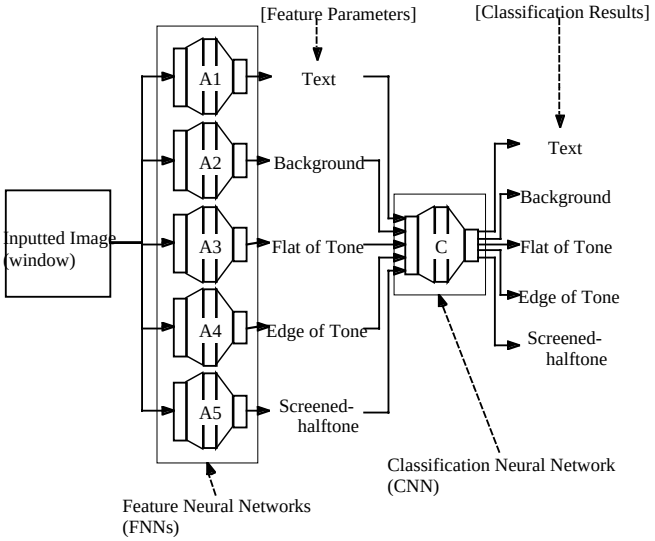

Figure 3: NN structure for the classification (CNN).

2.4 Post-processing

Classification of figure 3 is carried out separately at every pixel, so it does not consider how the segmentation results of the neighboring pixels are. And also feature extraction area are restricted to the local NxN picture elements. Because of these processing architecture, some mis-classification are observed as the results of figure 3. Therefore, a post-processing that checks the interrelation of the neighboring pixels is added as following two stages.

a) Area check:

Small area regions which are different from surrounding regions are omitted and assimilated into surrounding regions. We assumed that the effective regions less than the size of a 8 point character don't exist, and such regions are omitted.

b) Hough-transformation[9]:

Boundary lines of each region are smoothed by the Hough-transformation. In this paper we assumed the boundary lines are composed of vertical of horizontal straight lines.

3 Experiment

3.1 Experimental Condition

Experiment was carried out for gray-scale images which were digitized by an image scanner of 400 dpi resolution, 256 gray levels per pixel.

Each training set of 200 image block data (8x8 pixels) was selected randomly from one of the five regions (text, background, flat and edge part of continuous-tone and screened-halftone) for each FNN (feature extractor) respectively. In the case of continuous-tone image, edge parts were separated from flat parts by edge extraction operation using Laplacian-filter[10].

Furthermore, five kinds of data were prepared, which were changed the bias values of original image from -20 to +30 by 10 step. So total 6000 data were used for FNNs training. Gray levels of these data originally having 0-255 levels were normalized to 0-1 before starting the training.

Dimensions (number of input/output) of FNNs and CNN were selected as follows:

a) FNNs:
64(8x8) - 50 - 10
(Dimension of hourglass NN was 64-50-10-50-64),
b) CNN:
50(10x5) - 40 - 5.

We used two kinds of training algorithm. At the first stage of training (1-15000 times training), Kick-Out algorithm [6] was used to speed up the training. After that, usual Error Back-propagation algorithm[7] was used until the mean-squared error of training reached stable state.

Training coefficient was 0.001 and the momentum was 0.7.

3.2 Experimental result

3.2.1 Mean-squared Error

The mean-squared errors after the training of FNNs and CNN finished, are as follows:

- FNN
  - Text 0.1039
  - Background 0.0092
  - Flat part (Continuous-Tone) 0.0078
  - Edge part (Continuous-Tone) 0.0485
  - Screened-Halftone 0.1154
- CNN 0.0521

3.2.2 Segmentation Result

Figure 4 shows an example of test image, and figure 5 shows the classification result of figure 4. The test image of figure 4 includes three kinds of different image, i.e., a part of face image as a continuous-tone image, character image having the size of 8, 10.5, and 14 points as a text, and screened-halftone image of 65, 100, 133 and 175 lines.

In the figure 5, flat and edge part of continuous-tone image are combined to the same area, and text and background image are also combined to the text area. So the white area shows the continuous-tone region, gray area shows the text region, and black area shows the screened-halftone region.

Figure 5 shows our algorithm quite successful.

4 Conclusion

A newly developed image segmentation method is described. This method extracts text, continuous-tone and screened-halftone region in the document image by the neural networks (NN). Each feature extractor for text, continuous-tone and screened-halftone is composed of the 3-layer NN which is

Figure 4: Test image.

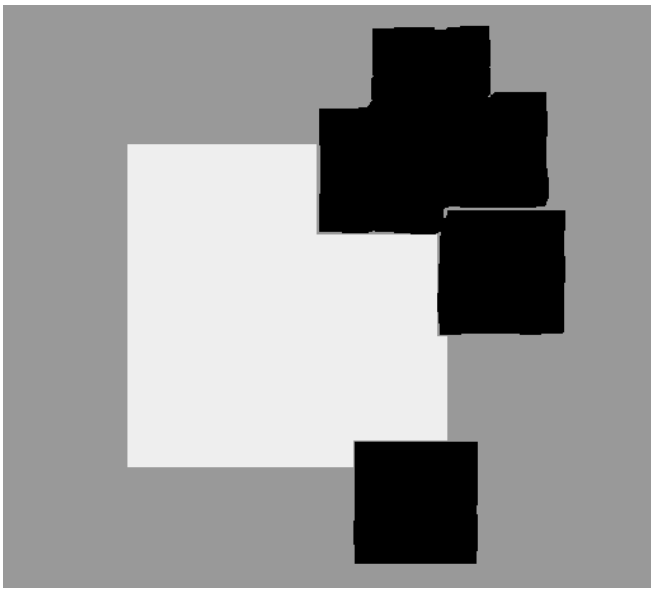

Figure 5: Segmentation result of test image.

upper part of 5-layer hourglass model NN; each hourglass model NN is trained by different learning samples which specify its own characteristics. The classification stage is composed of another 3-layer NN whose inputs are applied from feature extractor. And as a post-processing, Hough-transformation is applied to extract the boundary lines of each region, which is very effective to suppress the mis-classification in the small areas. Experimental result shows this method successful to segment text, continuous-tone and screened-halftone region.

References

1. Y. Inoue, Y. Nagata and T. Satoh, "Region Segmentation for Color Images Applying Neural Network", Conf. of Simulation Technology, Vol.13, pp.243-246 (1994)
2. S. Ohuchi, K. Imao and W. Yamada, "Segmentation Method for Documents Containing Text / Picture (Screened Halftone, Continuos Tone)", IEICE Trans. D-II, Vol. J75-D-II, No.1, pp.39-47 (1992)
3. K. Funahashi, "On the Approximate Realization of Identity Mappings by Three-Layer Neural Networks", IEICE Trans. A, Vol. J73-A, No.1, pp.139-145 (1990)
4. B. Irie and M. Kawato, "Acquisition of Internal Representation by Multi-Layered Perceptrons", IEICE Trans. D-II, Vol. J73-D-II, No. 8, pp.1173-1178 (1990)
5. T. Yonekura, S. Yokoi and J. Toriwaki, "A Method of Information Integration by Multi Layer Neural Networks and Its Theoretical Study", IEICE Trans. D-II, Vol.J73-D-II, No. 8, pp.1205-1212 (1990)
6. Y. Ochiai, N. Toda and S. Usui, "Kick-Out Learning Algorithm to Reduce the Oscillation of Weight", Neural Networks, Vol. 7, No. 5, pp.797-807 (1994)
7. D.E. Rumelhart, J.L. McClelland, "Parallel Distributed Processing", 1, MIT Press (1986)
8. G.W. Cottrell, P. Munro, D. Zipser, "Image Compression by Back Propagation: An Example of Extensional Programming", Tech. Rep. 8720, Univ. of California, San Diego, Institute for Cognitive Science (1987)
9. V.F. Leavers, "Shape Detection in Computer Vision Using the Hough Transform", Springer-Verlag London (1992)
10. T.Y. Young, "Handbook of Pattern Recognition and Image Processing", vol. 2, Computer Vision, Academic Press (1994)