Word Recognition by Combining Outline Emphasis and Synthesize Background

Yukihiro Achiha, Takayoshi Yamashita, Mitsuru Nakazawa, Soh Masuko, Yuji Yamauchi, Hironobu Fujiyoshi

▶ **To cite this version:**

Yukihiro Achiha, Takayoshi Yamashita, Mitsuru Nakazawa, Soh Masuko, Yuji Yamauchi, et al.. Word Recognition by Combining Outline Emphasis and Synthesize Background. 16th International Conference on Entertainment Computing (ICEC), Sep 2017, Tsukuba City, Japan. pp.492-496, 10.1007/978-3-319-66715-7_70 . hal-01771238

HAL Id: hal-01771238

https://inria.hal.science/hal-01771238

Submitted on 19 Apr 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Word Recognition by Combining Outline Emphasis and Synthesize Background

Yukihiro Achiha[1], Takayoshi Yamashita[1], Mitsuru Nakazawa[2], Soh Masuko[2], Yuji Yamauchi[1], and Hironobu Fujiyoshi[1]

[1] Chubu University
[2] Rakuten Institute of Technology, Rakuten, Inc.

Abstract. Character recognition collects item keywords from images from e-commerce websites; however, it requires a huge amount of training data. In this paper, we propose an efficient method to collect the training data by generating synthesis images and emphasizing outlines to obtain realistic images. The proposed method improves recognition accuracy on both generated images and real images from e-commerce websites.

keywords : character recognition, synthesis image, CNN

1 Introduction

Deep Convolutional Neural Network (DCNN) is a common approach in character recognition of handwritten characters and signs in scenes. Character recognition in scene images consists of two parts: character detection and recognition [4]. This technique can be applied to item images from e-commerce websites to collect the item's information. DCNN requires a huge amount of training data in order to obtain high accuracy [3][5]. Although public datasets are available, the range of fonts in these datasets is too small. In order to address this problem, a character synthesis method was proposed to reduce the image collection cost [1], which generates synthesis images by using font data and background images. However, it assumes English characters on a simple background. In this paper, we propose a method to generate synthesis characters and word images using Japanese font data and complex background images for item images from e-commerce websites. In addition, we introduce an approach to emphasize the outline of characters. Our method trained with synthesized images, which include both with and without outline emphasis, improves recognition accuracy.

2 Proposed method

The proposed method consists of three steps: generation of character images, addition of margins and emphasis of outlines, and synthesis of the characters with a complex background. Figure 1 shows the flow of character image generation. First, character images are generated from character lists with font data; these

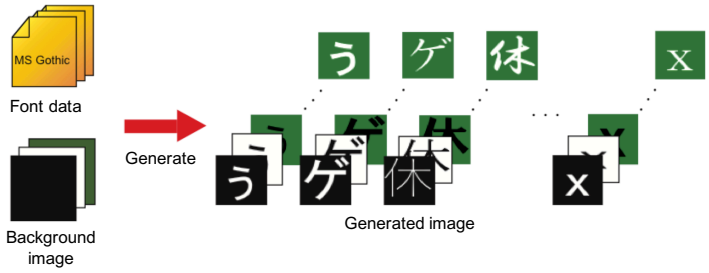


Fig. 1. The flow of character image generation

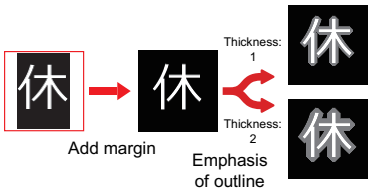


Fig. 2. Addition of margins and emphasis of outline

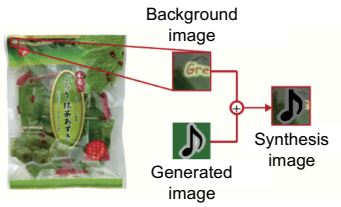


Fig. 3. Synthesis of background image

images are then synthesized with a background image. We prepared 22 fonts commonly used in e-commerce websites such as MS Mincho and Yuri Gothic.

Character images on complex backgrounds of e-commerce images are sometimes decorated, for example with borders. In order to improve recognition accuracy of such characters, outline emphasis of characters is introduced during character generation. First, a margin is added to the generated image. Then, an outline is added to emphasize characters in the image. At that time, two types of images are generated with different outline thicknesses. Figure 2 shows the flow of the addition of margins to the character image, followed by outline emphasis.

The background region is replaced with a complex background such as an item image. As shown in Fig. 3, the color of the background is green, and the color of the character and the outline are different. The background image is cropped randomly from a banner image of an e-commerce store.

The DCNN is trained by using the synthesis character images. To test the effectiveness of the method, we also applied the proposed method to word synthesis. The word images are generated based on a word list and synthesized with complex backgrounds.

2.1 Structure of DCNN

Figure 4 shows the DCNN network structure. The network consists of 4 layers: 3 convolution layers and 1 fully connected layer. The filter size of each layer is 5×5. Max pooling is employed for the pooling layer. The fully connected layer has 4,096 units and it employs Dropout[7] during the learning phase. The activation function of each layer is ReLU[6]. The output units have 1,253 classes for

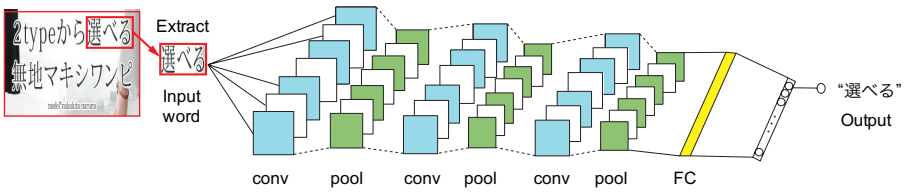


Fig. 4. The structure of DCNN

character recognition and 241 classes for word recognition. The input image size is 32×32 for character recognition and 96×96 for word recognition, respectively. AdaGrad [8] is used for the optimization method. The mini batch size is 32 and the epoch number is 50.

3 Evaluation

First, we evaluate the effectiveness of outline emphasis and synthesis with background images using top 5 accuracy. We trained 1,253 characters using 145 images for each (Fig. 5). For evaluation, 227 images collected from e-commerce websites were used. Table 1 shows the result for synthesis images. The method with emphasis and synthesis achieves best performance on top 1 accuracy.

The evaluation results of character recognition and word recognition on real images of e-commerce are shown in Table 2 and Table 3, respectively. From Table 2, the method with emphasis performs 12.4% better than baseline, which is without emphasis and synthesis, on top 1 accuracy. On the other hand, the method with synthesis also improves 13.8% than baseline on top 1 accuracy. The combination of emphasis and synthesis achieves best performance with an improvement of 14.0%. The results of word recognition on real images are shown in Table 3. The method with emphasis improves about 20.4% and 19.8% than baseline on top 1 and top 5, respectively. Synthesis is also effective for real images; it improves accuracy by 24.4% and 22.3% than baseline on top 1 and top 5, respectively. The combination of emphasis and synthesis achieves best

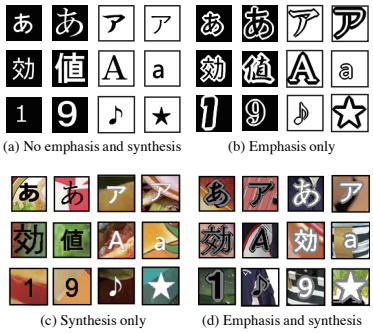


Fig. 5. Example of synthesis images

Table 1. The comparison of character recognition on synthesis images[%]

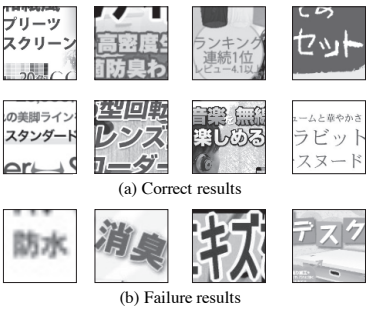
outline	background	top1	top2	top3	top4	top5
–	–	99.4	99.8	99.8	99.8	99.9
x	–	96.0	97.7	98.2	98.6	98.8
–	x	99.1	99.6	99.8	99.9	99.9
x	x	99.6	99.9	99.9	99.9	99.9

Table 2. Comparison of character recognition on real images[%]

outline	background	top1	top2	top3	top4	top5
–	–	50.3	62.0	66.2	69.6	71.5
x	–	62.7	72.5	75.8	77.5	78.7
–	x	64.1	73.1	76.9	79.4	80.9
x	x	64.3	74.4	78.2	80.2	81.4

Table 3. Comparison of word recognition on real images[%]

outline	background	top1	top2	top3	top4	top5
–	–	27.6	34.1	36.8	40.2	42.1
x	–	48.0	56.0	58.8	60.7	61.9
–	x	52.0	58.5	60.7	63.8	64.4
x	x	62.8	69.3	70.9	73.4	76.2



(a) Correct results

(b) Failure results

Fig. 6. Examples of recognition results

performance. Figure 6 shows recognition results. It recognizes words correctly even when the number of characters is different. However, recognition fails when the characters are blurred or rotated.

4 Conclusion

In this paper, we proposed a method to generate outline emphasis of word images and synthesize them with complex background images. The DCNN trained with generated images obtained high accuracy on both synthesized images and real images from e-commerce websites.

References

1. M. Jaderberg, K. Simonyan, A. Vedaldi, A. Zisserman, "Synthetic data and artificial neural networks for natural scene text recognition", arXiv 2014, NIPS Deep Learning Workshop, 2014
2. T. Kobayashi, M. Nakagawa, "A Pattern Classification Method of Linear-Time Learning and Constant-Time Classification", IEICE, 89(11):981-992, Nov. 2006.
3. Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-Based Learning Applied to Document Recognition", Proceedings of the IEEE, 86(11):2278-2324, Nov. 1998.
4. C.-L. Liu, M. Koga, H. Fujisawa, "Lexicon-driven segmentation and recognition of handwritten character strings for Japanese address reading", TPAMI, 24(11),1425-1437, Nov. 2002.
5. T. Wang, D. J. Wu, A. Coates, A. Y. Ng, "End-to-End Text Recognition with Convolutional Neural Networks", ICPR, 2012.
6. V. Nair, G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines", ICML, pp.807-814, 2010.
7. G. E. Hinton, N.Srivastava, A.Krizhevsky, I.Sutskever, and R.R.Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors", Clinical Orthopaedics and Related Research, vol.abs/1207.0850, 2012.
8. D. John, E. Hazan, Y. Singer,"Adaptive subgradient methods for online learning and stochastic optimization." Journal of Machine Learning Research 12.Jul (2011): 2121-2159.