International Journal of Advanced Smart Convergence Vol.8 No.3 131-137 (2019)
http://dx.doi.org/10.7236/IJASC.2019.8.3.131

IJASC 19-3-17

Atypical Character Recognition Based on Mask R-CNN for Hangul Signboard

Sooyeon Lim*

Department of Game, Dongyang University, Korea
sylim@dyu.ac.kr

Abstract

This study proposes a method of learning and recognizing the characteristics that are the classification criteria of Hangul using Mask R-CNN, one of the deep learning techniques, to recognize and classify atypical Hangul characters. The atypical characters on the Hangul signboard have a lot of deformed and colorful shapes beyond the general characters. Therefore, in order to recognize the Hangul signboard character, it is necessary to learn a separate atypical Hangul character rather than the existing formulaic one. We selected the Hangul character '닭' as sample data and constructed 5,383 Hangul image data sets and used them for learning and verifying the deep learning model. The accuracy of the results of analyzing the performance of the learning model using the test set constructed to verify the reliability of the learning model was about 92.65% (the area detection rate). Therefore we confirmed that the proposed method is very useful for Hangul signboard character recognition, and we plan to extend it to various Hangul data.

Keywords: *Atypical Character Recognition, Hangul Signboard, Mask R-CNN, Hangul signboard character*

1. INTRODUCTION

For character recognition, there is a need for an effective and flexible method of extracting character features of images in various and complex environments. Recently, global IT companies such as Google, Facebook, etc have been actively researching hierarchical and automated feature learning methods through multilayer neural networks [1-3]. Deep learning, which is the hottest topic of machine learning, learns the features that best represent characters in various environments through unsupervised learning from the learning images, and converts them into higher-level features using multi-layered neural networks such as CNN. These features can be used as inputs to the character detection model for character recognition, resulting in a more accurate and high performance character detection and recognition model.

This study uses Mask R-CNN, one of the deep learning techniques, to automatically recognize and classify atypical Hangul characters. Hangul character recognition is more difficult than English character recognition because the number of Hangul characters is much larger than that of English characters. In particular, the Hangul signboard, which is commonly seen on the street, is composed of characters that have been transformed

Manuscript Received: July. 31, 2019 / Revised: Aug. 1, 2019 / Accepted: Aug. 12, 2019
Corresponding Author: sylim@dyu.ac.kr
Tel: +82-31-839-9065, Fax: +82-31-851-3625
Department of Game, Dongyang University, Korea

and colorful to attract the attention of people passing through the street or in the form of a brand logo. We selected '닭', one of the most frequently seen Hangul characters in real life, and constructed 5,383 Hangul character image data sets and used them for learning and verifying deep learning models.

2. RELATED WORKS

Recently, research on image recognition and analysis based on deep learning technology CNN (Convolution Neural Network) is active. Deep learning has supervised and unsupervised learning. This study classifies and predicts data based on supervised learning. CNN is one of the most influential neural networks in computer vision, including image processing and speech processing [4-5]. CNN constructs a reasonable network using a combination of specific layers based on the fact that the input consists of images. It generally consists of three core structures: the convolutional layer, the pooling layer, and the fully-connected layer [6-8].

CNN image classification has been improved to R-CNN, Fast R-CNN, Faster R-CNN, Mask R-CNN, etc. R-CNN (Region-based CNN) uses the selective search algorithm [9] to find region proposals, which are areas where objects in an image exist. This algorithm merges adjacent pixels with similar patterns such as color distribution and intensity. About 2,000 of the extracted bounding boxes are used as input data of CNN. At the end of the CNN, the image is classified using SVM (Support Vector Machine) [10]. Finally, a linear regression model is used to more accurately match the bounding box coordinates of the classified objects.

Fast R-CNN [11] introduced the RolPool concept to solve the cost problem of using all three models of selective search, SVM, and linear regression for many regions when classifying images. This is how to apply CNN to the input image first and find the object area in the CNN feature map.

Faster R-CNN [12] solved the bottleneck of the region proposal stage in Fast R-CNN. The Selective search algorithm, which extracts the object region in Fast R-CNN, is changed to RPN (Region Proposal Network). This speeds up the Fast R-CNN by putting the region proposal step into the CNN.

Unlike other methods, Mask R-CNN [13] does not search the image area by the box but searches the image area by the pixel. This improves the accuracy and speed of recognition by adjusting RoIPool through a method called RoIAlign to achieve the correct alignment of the feature map (Figure 1).

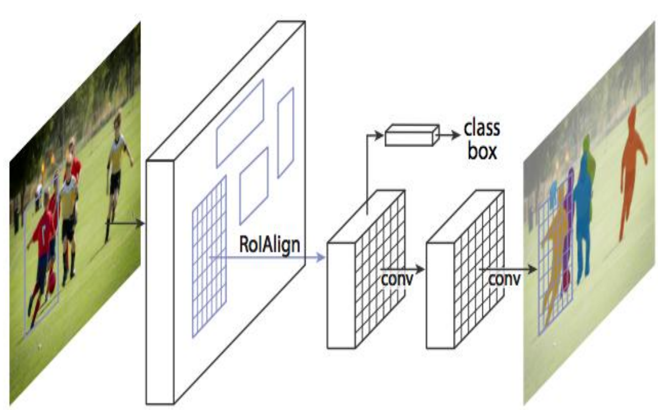

Figure 1. The Mask R-CNN framework for image segmentation

3. HANGUL CHARACTER RECOGNITION BASED ON MASK R-CNN

There are many deep learning examples for character classification. Deep Font [14] is an English font classification model based on AlexNet. There are also many deep learning examples for classifying Hangul characters and fonts [15-17]. However, these studies have the disadvantages that the characters and fonts that can be classified using a model composed of simple layers are limited and that it is difficult to recognize a large number of Hangul fonts.

In character recognition process in image, when it is determined that a character exists in a given image, the result is indicated by a bounding box. Therefore, various techniques have been developed for extracting the area of individual characters in a box, recognizing them as individual characters, and connecting them and recognizing them as characters. Most of these techniques are based on some assumptions, such as the spacing of characters, the position of characters on a straight line, or the same size of characters.

However, the characters on the signboard seen in the street are not affected by such assumptions because the shape, size, etc. of characters are atypical. Therefore, a recognition process that is completely different from general printed characters or handwriting is required. This study aims to learn Hangul characters with various visual forms, not to obtain Hangul characters from the characters of a specific font, and to extract and classify each feature. For this purpose, the atypical visualized characters on the street signboard were chosen as the test subjects.

3.1 Data set

In this study, we constructed a new Hangul database for character recognition in Hangul signboard. First of all, 5,383 original images were collected by selecting '닭', one of the most frequently used characters in Hangul signboards. These images were obtained from relevant searches from Google and Naver.

We divided 5,383 images into training data and test data, and each data set was randomly divided into 80% and 20% of the total data. Figure 2 shows part of the dataset we built.

Figure 2. Some images of data set for Hangul signboard recognition

3.2 Atypical Hangul Recognition

Recognition and classification of Hangul characters based on Mask R-CNN proposed in this study is done as shown in the figure 3. In order to recognize a specific Hangul character in an atypical character image, first, a task of quickly and accurately separating and acquiring each character from an image composed of several characters should be performed. After that, each extracted character image is stored in DB. We performed annotation manually on the assumption that the preprocessing process described above was successfully completed for the images in the collected data set. Annotated Hangul character images are input to the Mask R-CNN model and used for training. Figure 3 shows how the character '닭' used in the Hangul signboard is recognized by the Mask R-CNN model.

Original image after Annotation Result of Mask R-CNN

Figure 3. Mask R-CNN learning process on original image

4. IMPLEMENTATION

We set Hangul character '닭' as a related search term and collected Hangul signboard data through web crawling and actual photography to build a data set. The built dataset was used to implement a supervised learning system that applied a mask R-CNN to detect and recognize '닭' character. The total 5,383 data collected were divided into a train set of 4,306 data and a test set of 1,077 data. Images used in all experiments were upsampled to 1024x1024 with preprocessing. Training data begins to be trained after initializing the model using weights obtained through pretraining on the MSCOCO data set [18]. Such a method has an advantage that the learning speed can be increased and the accuracy can be increased compared to learning using only user data. In this learning system, the batch size is 4 and the learning rate starts with 0.001 and ends with 0.0001. Training is performed by setting up a network model of a total of 200 epochs using stochastic gradient descent of Momentum 0.9. We experimented with a combination of different learning times and different learning coefficients, but there was no noticeable improvement by additional learning.

The implementation of this study was performed on a system using an Intel (R) i9-9900KF CPU 3.60GHz processor, 3,072 CUDA cores, and two NVIDIA GeForce RTX2080 GPUs with 1,845GHz and 8GB of memory. We also implemented experiments using Tensorflow, a representative AI open source library used in Python, and Object Detection API, an open source framework that can be used in conjunction with it.

We analyzed the performance of the learning model using the test set constructed to verify the reliability of the learning model. As a result of applying the method proposed in this study, the area detection rate was 92.65% for '닭' character on the Hangul signboard. Through the validation and test results of the learning model, we confirmed that the proposed method is very useful for Hangul signboard character recognition.

The resulting image can be applied to interactive art using the motion of the viewer. Fig. 4 is a flow chart showing the process of visualizing the recognized Hangul character image as 3D character interacting with the viewer by being processed as input of 3D interactive typography [19-20].

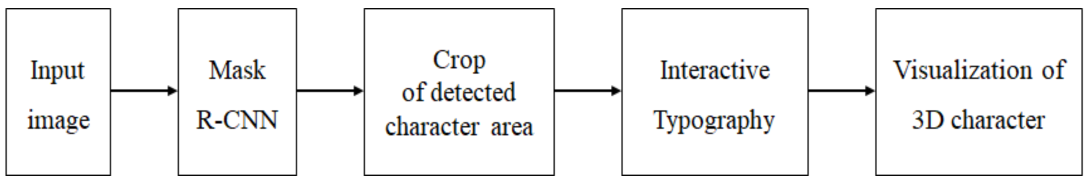

Figure 4. A flow chart of Application to 3D Interactive Typography

Figure 5 illustrates the process of deconstructing the components and connecting them to the viewer's skeleton in order to connect the Hangul recognized by Mask R-CNN with the viewer's motion.

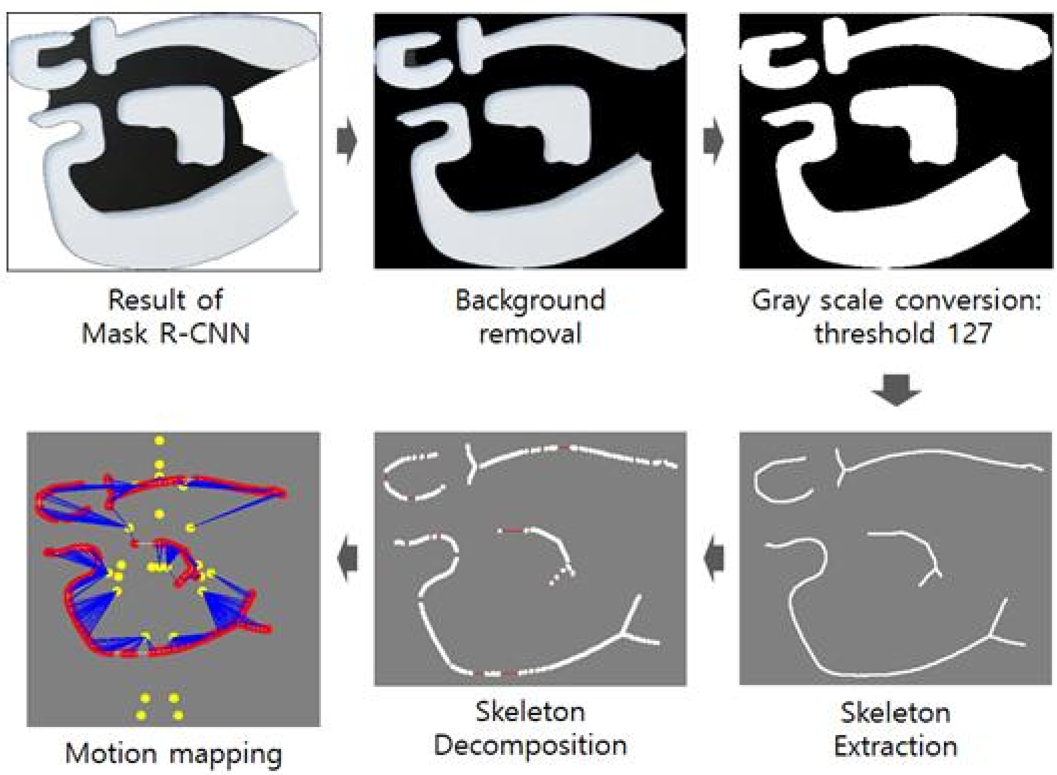

Figure 5. The process that Hangul character '닭' recognized through Mask R-CNN are connected to viewer's skeleton

5. CONCLUSION

This study is about deep learning based character recognition for extracting character from Hangul signboard images with many atypical characters. As a result of training by Mask R-CNN model, we have detected the position of characters added up to semantic segmentation and extracted more meaningful results. The proposed Hangul character recognition method consists of three steps. First, data for learning Hangul characters existing in signboards of various standards is collected and annotated. This process is the most important process in deep learning based recognition method. This is because errors in annotations result in less accurate learning. The second step is to learn the collected data using a deep learning algorithm based on Mask R-CNN. Finally, the region forming the stroke of the character is detected through the post-processing of the character candidate region derived by the learned weights. The experiment was carried out on a signboard containing the character '닭' one of the most commonly used characters in daily life. As a result, the area detection rate was 92.65% for the character '닭' on the Hangul signboard.

The model proposed in this study is sufficiently recognizable, but it is necessary to secure the data set and to further improve the learning model in order to become an optimal learning model for recognizing and classifying many Hangul signboards currently in circulation in Korea. If the proposed method is extended to more optimized network structure and sufficient dataset for atypical Hangul character recognition in the future, real-time character recognition and extraction can be performed in more diverse and complex environments. It can be widely used in various commercial and public service developments and public arts.

ACKNOWLEDGEMENT

This study was supported by grant from Dong Yang University in 2018.

REFERENCES

[1] Y. Sun, Y. Chen, X. Wang and X. Tang, "Deep learning face representation by joint identification-verification," *Advances in neural information processing systems*, pp. 1988-1996, 2014.
https://arxiv.org/abs/1406.4773

[2] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint*, 2016.
https://arxiv.org/abs/1409.1556

[3] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee and A. Y. Ng, "Multimodal deep learning," in *Proc. 28th international conference on machine learning(ICML-11)*, pp. 689-696, 2011.
http://hdl.handle.net/10203/198493

[4] X. B. Zhang, F. C. Chen and R. Y. Huaug, "A Combination of RNN and CNN for Attention-based Relation Classification," *Procedia Computer Science*, Vol. 131, pp. 911-917, 2018.
DOI: https://doi.org/10.1016/j.procs.2018.04.221

[5] H. C. Moon, A. N. Yang and J. G. Kim, "CNN-Based Hand Gesture Recognition for Wearable Applications," *The Korean Society of Broad Engineers*, Vol. 23, No. 2, pp. 246-252, 2018.
DOI: https://doi.org/10.5909/JBE.2018.23.2.246

[6] A. Krizhevsky, I. Sutskever and GE. Hinton, "Imagenet classification with deep convolution neural networks," *Advances in Neural Information Processing Systems*, pp.1106-1114, 2012.

[7] K, Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint*, 2014.
https://arxiv.org/abs/1409.1556

[8] C. Szegedy, W. Liu, Y. Jia and P. Sermanet, "Going deeper with convolutions," in *Proc. IEEE conference on computer vision and pattern recognition*, pp. 1-9, 2015.

[9] J. R. Uijlings, K. E. Van De Sande, T. Gevers and A. W. Smeulders, "Selective search for object recognition," *International journal of computer vision*, Vol. 104, No. 2, pp. 154-171, 2013.

[10] H. Drucker, C. J. Burges, L. Kaufman, A. J. Smola and V. Vapnik, "Support vector regression machines," *Advances in neural information processing systems*, pp. 155-161. 1997.

[11] R. Girshick, "Fast R-CNN," in *Proc. IEEE international conference on computer vision*, pp. 1440-1448, 2015.

[12] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, pp. 91-99, 2015.

[13] K. He, G. Gkioxari, P. Dollár and R. Girshick, "Mask r-cnn," in *Proc. IEEE international conference on computer vision*, pp. 2961-2969, 2017.

[14] Z. Wang, J. Yang, H. Jin, E. Shechtman, A. Agarwala, J. Brandt and T. S. Huang, "Deepfont: Identify your font from an image," in *Proc. 23rd ACM international conference on Multimedia*, pp. 451-459, 2015.
DOI: https://doi.org/10.1145/2733373.2806219

[15] I. K Hwang, *Study on Hangul font characteristics using CNN*, Doctoral dissertation, Seoul National Univ., 2017.
http://hdl.handle.net/10371/131338

[16] I. J. Kim, C. Choi and S. H. Lee, "Improving discrimination ability of convolutional neural networks by hybrid learning," *International Journal on Document Analysis and Recognition*, Vol. 19, No. 1, pp. 1-9, 2016.
DOI: https://doi.org/10.1007/s10032-015-0256-9

[17] J. H. Yang, H. B. Kwak and I. J. Kim, "Large-Scale Hangul Font Recognition Using Deep Learning," in *Proc. Annual Conference on Human and Language Technology*, pp. 8-12. 2017.

[18] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Proc. European conference on computer vision*, pp. 740-755, 2014.
DOI: https://doi.org/10.1007/978-3-319-10602-1_48

[19] S. Lim, "Emotional Communication on Interactive Typography System," *International Journal of Contents*, Vol. 14, No. 2, pp. 41-44, 2018.
DOI: http://doi.org/10.5392/IJoC.2018.14.2.041

[20] S. Lim, "3D Spatial Interaction Method using Visual Dynamics and Meaning Production of Character," *International journal of advanced smart convergence*, Vol. 7, No. 3, pp. 130-139, 2018.
DOI: https://doi.org/10.7236/IJASC.2018.7.3.130