International Journal of Advanced Smart Convergence Vol.8 No.3 95-101 (2019)
http://dx.doi.org/10.7236/IJASC.2019.8.3.95

IJASC 19-3-13

A Comparative Study on OCR using Super-Resolution for Small Fonts

Wooyeong Cho[1], Juwon Kwon[1], Soonchu Kwon[2], Jisang Yoo[1]

[1]Department of Electronics Engineering, Kwangwoon University, Korea
[2]Graduate School of Smart Convergence, Kwangwoon University, Korea
[1]chowy1628@gmail.com, [1]02kjw0203@gmail.com, [2]ksc0226@kw.ac.kr, [1†]jsyoo@kw.ac.kr

Abstract

Recently, there have been many issues related to text recognition using Tesseract. One of these issues is that the text recognition accuracy is significantly lower for smaller fonts. Tesseract extracts text by creating an outline with direction in the image. By searching the Tesseract database, template matching with characters with similar feature points is used to select the character with the lowest error. Because of the poor text extraction, the recognition accuracy is lowerd. In this paper, we compared text recognition accuracy after applying various super-resolution methods to smaller text images and experimented with how the recognition accuracy varies for various image size. In order to recognize small Korean text images, we have used super-resolution algorithms based on deep learning models such as SRCNN, ESRCNN, DSRCNN, and DCSCN. The dataset for training and testing consisted of Korean-based scanned images. The images was resized from 0.5 times to 0.8 times with 12pt font size. The experiment was performed on x0.5 resized images, and the experimental result showed that DCSCN super-resolution is the most efficient method to reduce precision error rate by 7.8%, and reduce the recall error rate by 8.4%. The experimental results have demonstrated that the accuracy of text recognition for smaller Korean fonts can be improved by adding super-resolution methods to the OCR preprocessing module.

Keywords: *Korean OCR, Tesseract, Super-resolution, Text-recognition, and Deep-learning*

1. INTRODUCTION

The text recognition field was developed by G. Tauschek in 1928 with a patent application for the world's first printed number reading and character recognition methods. Since the 1950s, computers have been produced and are used for various automated tasks. OCR (Optical Character Recognition) [1] is conversions from captured images written by humans or printed by a machine with an image scanner to machine-readable characters. Early systems can only read a sample of the typeface beforehand to read that particular typeface. But with the development of deep learning [2] technology, it has a good prospect as the development of artificial intelligence or machine vision research field. The recognition accuracy for English characters is good, but the performance is low for Hangul, so Korean researchers are trying to improve the recognition accuracy of

Manuscript Received: July. 16, 2019 / Revised: July. 20, 2019 / Accepted: July. 27, 2019
***Corresponding Author: jsyoo@kw.ac.kr
Tel: +82-2-940-5112, Fax: +82-2-912-7332
Department of Electronics Eng. Kwangwoon University, Korea

OCR. To improve the recognition accuracy, preprocessing is an important process. Preprocessing is an image refinement that adjusts margins, layout, brightness, contrast, and resolution to make it easier for the computer to recognize characters in images. Images are preprocessed in various and complex ways that vary from case to case. In this paper, we raise the problem of low recognition accuracy when applying Tesseract OCR for small letters, set up an experimental procedure, experiment with various super-resolution methods as a solution for low recognition accuracy, and compare the results.

2. TESSERACT-OCR

Tesseract-OCR is an open-source OCR engine based on a command-line interface that has been evolved with Google's support since 2006. It can be combined with 'Leptonica' image processing library to process images of various formats. For example, By including layout analysis, the computer can detect whether the text is monospaced or proportionally spaced. In the early stages of development, only English was recognized, but more than 60 languages were supported and new languages could be made recognizable by machine learning. In 2007, it was reported that text recognition performed much better than other open-source tools and currently, it adds an LSTM [3] based engine and provides a total of 116 languages on Unicode (UTF-8) support. The train data file for each language is an archive file in a Tesseract specific format. It contains several uncompressed component files that are needed by the tesseract-OCR process. The recognition accuracy of Tesseract-OCR is greatly influenced by image characteristics such as brightness, tilt, etc. Therefore, it is important to preprocess images before OCR. If the angle of the image is distorted, characters may not be found, and the black borders of the general document may also be mistaken for text and should be erased.

3. SUPER-RESOLUTION

Super-resolution is a method of upscaling video images or a single image by enhancing the resolution. This topic was first published in 1984 and the term of super-resolution is originated in the 1990s. This technique is used to extract high quality upsized images from low-quality images. It is mainly used in image processing fields such as radar or sonar imaging applications. The basic concept of HR (High-Resolution) image reconstruction is to obtain several LR (Low-Resolution) images by moving sub-pixel [4] units form HR images and to reconstruct the HR images by estimating sub-pixel movements from the LR images [5]. Therefore, in order to reconstruct the HR image, information of other images is needed to upsize one image and a precise matching process between LR images is also important. A lot of research has been done on the new method that combines super-resolution based on CNN (Convolutional Neural Network) [6] and this will be a new base technology that can watch the image with the HR. In this part, various super-resolution methods of SRCNN [7], ESRCNN, DSRCNN [8], DCSCN [9] are applied to find out how they affect recognition accuracy.

3.1 SRCNN (SUPER-RESOLUTION CNN)

SRCNN is meaningful as the first method to apply deep learning to SISR (Single Image Super-Resolution). Since then, many papers have been used as a basis and the simple structure has surpassed the existing SR methods such as Bilinear [10], Bicubic [11] method.

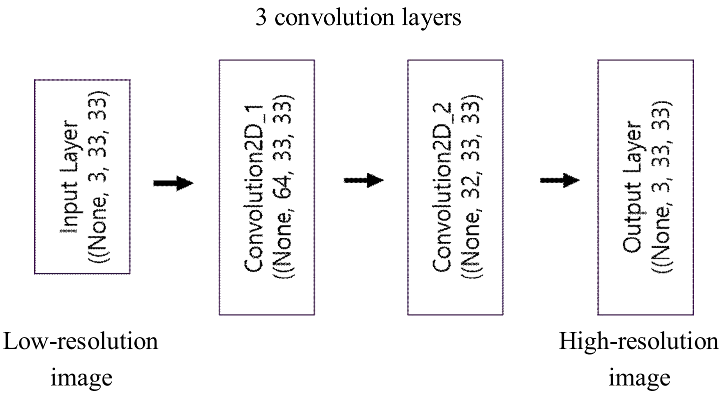


Figure 1. SRCNN structure

This model consists of three stages as shown in Figure 1. It takes an LR image as input and goes through a series of processes to output the HR image. It has higher PSNR (Peak Signal to Noise Ratio) value when more images for training are used while using larger filter size because if the model has a larger filter size, feature extraction performance is better than using small filter size. This method was only applied to the Luminance channel at that time, but further studies showed better results when applied to the RGB channel.

3.2 ESRCNN (Expanded Super Resolution CNN)

This method is slightly improved in SRCNN. The model is shown in Figure 2. below. The difference from SRCNN is that the layer is added in parallel in the middle of the layer. Neural networks connected in parallel consist of the added feature of merging them into the next step. The kernel size was 1x1 for non-linear mapping in SRCNN, while ESRCNN adjusted kernel size 5x5 to maximize the amount of information from each layer. Finally, the structure is extracted from the more upscaled image by taking the average at the output side.

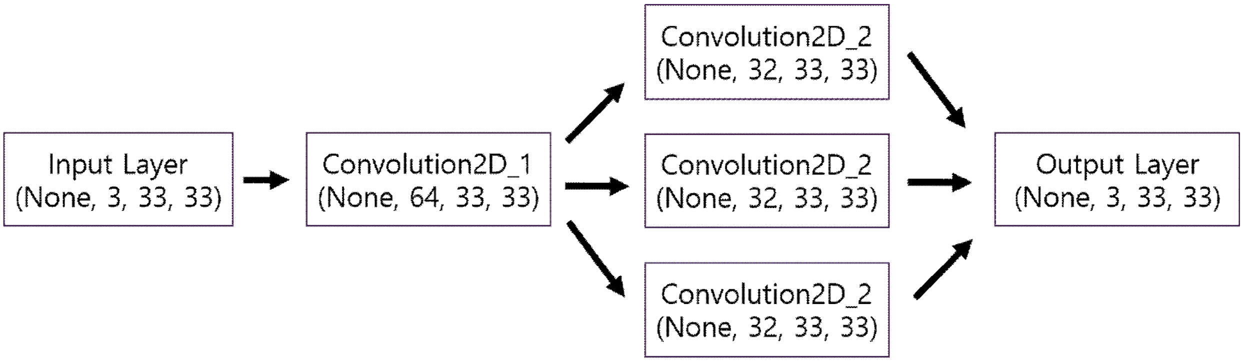


Figure 2. ESRCNN structure

3.3 DSRCNN (Denoising Super Resolution CNN)

DSRCNN is an evolved model against SRCNN's noisy output with higher performance on Set5 [12]. It consists of several convolutional and deconvolutional layers [13] which are used for upsampling and downsampling respectively. Also, the skip connection of RESNET [14] was used in the middle part of the model. The reason for this is that if the model network is deeper, it will not recover well whenever the weight updates pass the deconvolutional layer. The operation of this model is to downscale the training image by passing a Gaussian filter with a sigma of 0.5 to 1/3 and then upscaling it to a size of 33x33.

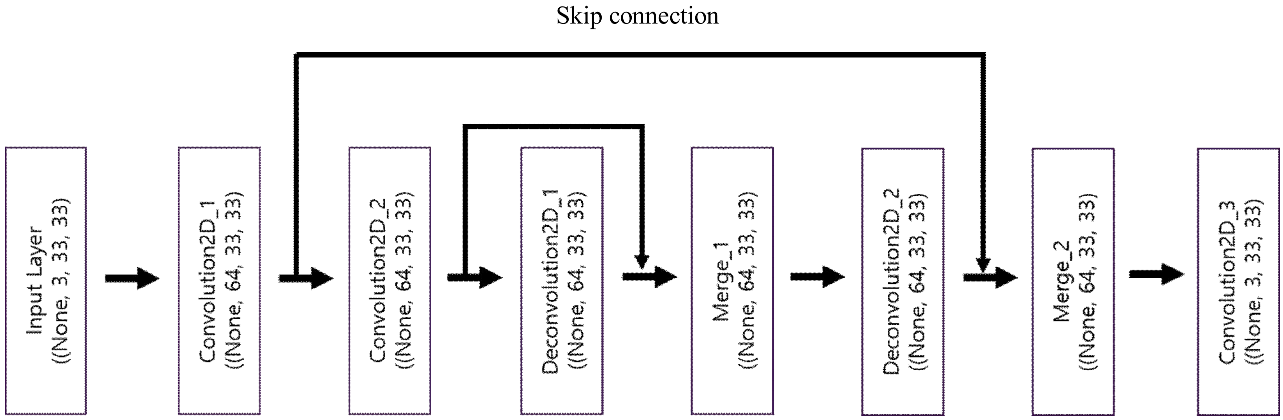

Skip connection
Input Layer ((None, 3, 33, 33))
Convolution2D_1 ((None, 64, 33, 33))
Convolution2D_2 ((None, 64, 33, 33))
Deconvolution2D_1 ((None, 64, 33, 33))
Merge_1 ((None, 64, 33, 33))
Deconvolution2D_2 ((None, 64, 33, 33))
Merge_2 ((None, 64, 33, 33))
Convolution2D_3 ((None, 3, 33, 33))

Figure 3. DSRCNN structure

3.4 DCSCN (Deep CNN with Skip Connection and Network in Network)

This model combines a deep neural network and skip connection techniques that is known as a high efficiency deep learning-based super-resolution model. Deeply stack layers are used to improve feature extraction performance and to solve large computational and finite computational resource problems. SISR's latest model trend consists of 20 to 30 CNN layers, but this model reduces the total computation of CNN filters by 10 to 100 times by using 11 layers. In addition, the SRCNN and other models have redundant pixels due to the operation of upscaling the image by the product of the scale factor when inputting the super-resolution. There is an advantage that can get a better understanding of the feature and by configuring the 1x1 size CNN layer. This method calls depthwise convolution, meaning each input channel is convolved in isolation of other channels. It is possible to reduce the dimension of the previous layer to minimize the loss of information and to reduce the amount of computation.

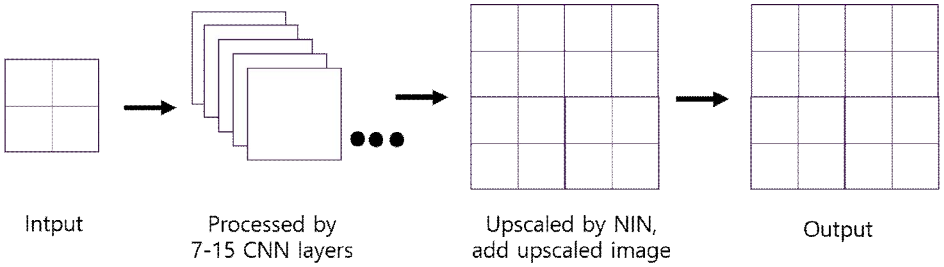

Intput
Processed by 7-15 CNN layers
Upscaled by NIN, add upscaled image
Output

Figure 4. DCSCN structure

4. Experiments and results

4.1 Dataset

Dataset is made up of the scanned image of 1315 Hangul letters and 26 English capital letters at 300 DPI based on A4 paper size. First, the image was converted to an 8-bit grayscale image and binarized to a value of 189 to increase the text recognition accuracy. Then, the image was cropped to create random letters, and the sample size was extracted at a constant pixel size of 153x23. The characters are 12pt, the letter-spacing is 1, and the font style is *Gulim*. To apply the super-resolution of the extracted character sample, the images are downsampled by the size of 0.1 times per step. In addition, there are 137 character samples for each step, and there are 6 steps, so the dataset consists of 822 images.

4.2 Experiment Method

The experimental environment includes Tesseract-OCR ver. 4.0, Tensorflow 1.8.0, CUDA 9.0, cuDNN 7.4, and open-cv ver 4.0 for image preprocessing and resizing. First, some basic preprocessing to recognize the text in a scanned document is conducted. During the preprocessing process, the threshold was set to configure the image to binarize the image and after that skew correction for slightly rotated parts and set up the ROI (Region Of Interest) to reduce false positives for non-text elements.

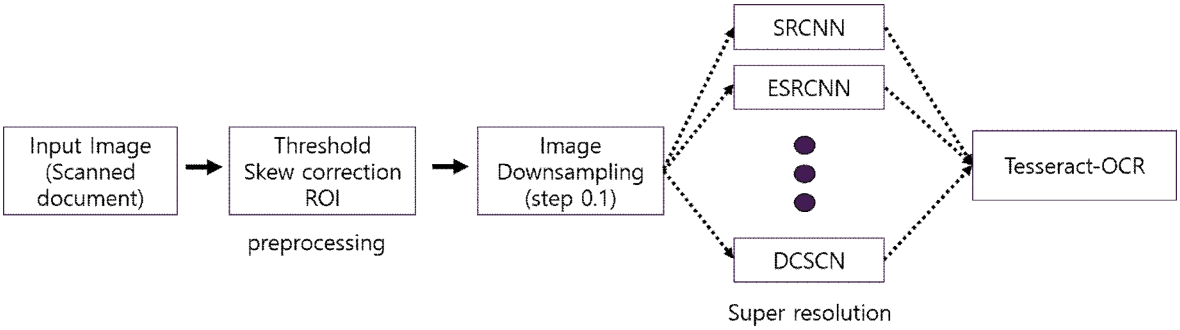
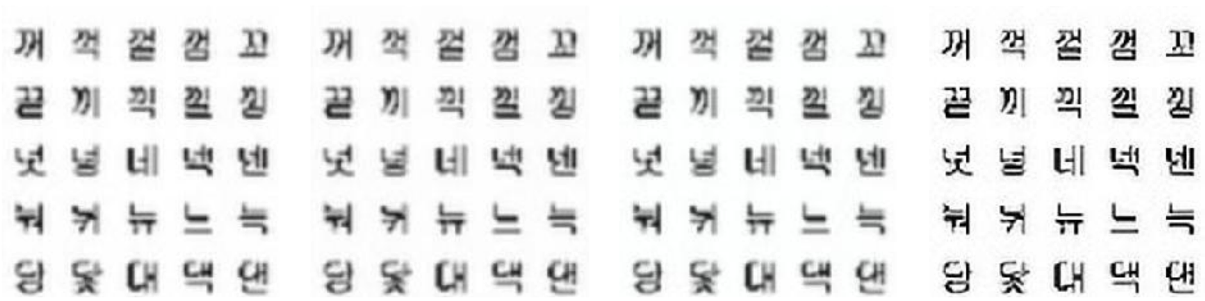

Input Image (Scanned document)
Threshold Skew correction ROI
preprocessing
Image Downsampling (step 0.1)
SRCNN
ESRCNN
DCSCN
Super resolution
Tesseract-OCR

Figure 5. Experiment procedure

Afterward, 5 steps of downsampling were performed at intervals of 0.1 times in order to check the recognition accuracy and obtain super-resolution results. First, the text recognition accuracy when the font size is small is displayed as the index of the recognition accuracy using precision and recall. When comparing the OCR accuracy of images resized from 0.5 to 0.9 times and that of applying super-resolution, the latter accuracy is increased compared to the former.

4.3 Results

Table 1 shows the precision error and recall error based on the text data obtained by tesseract-OCR after applying several super-resolution to our dataset. The error rates of the samples of the dataset are 34.3% and 33.4%, respectively. However, the error rate when samples are downsampled by x0.5 times shows a high error rate value of 90.4%. It means that the error rate significantly increases as the font size is reduced.



(a) **(b)** **(c)** **(d)**

Figure 6. Super-resolution results. (a) SRCNN, (b) ESRCNN, (c) DSRCNN, (d) DCSCN

Table 1. Experiment results of precision error and recall error (scale = x2)

Ratio	x 0.5		x 0.6		x 0.7		x 0.8	
Method	Precision error	Recall error	Precision error	Recall error	Precision error	Recall error	Precision error	Recall error
SRCNN	88.9	88.6	81.6	81.5	66.3	66.3	56.6	56.6
ESRCNN	86.1	85.8	82.3	82.2	68.3	68.3	54.5	54.3
DSRCNN	86.1	85.9	75.1	74.8	63	63	55.7	55.6
DCSCN	82.6	82	73.9	73.9	62	63	52	53

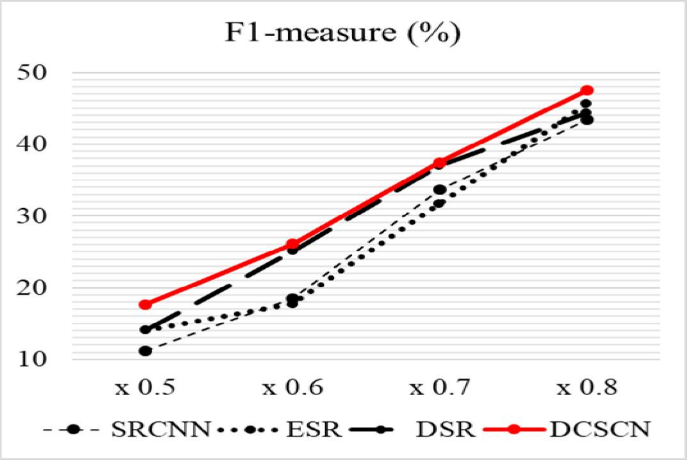


Figure 7. The result of f1-measure of SRCNN and other methods on our dataset

Figure 7 shows the result of calculating the f1-measure value based on the result of table 1. It shows the f1-measure values by size, the DCSCN model has the upper hand in each case. Table 2 shows the operating time for each super-resolution for each character sample sized 0.5 times. The DSRCNN model was the longest with 3.77 seconds per sample and the SRCNN model was the shortest with 1.44 seconds. The DCSCN model is 0.07 seconds slower than the SRCNN model but can be said to be an efficient model for reducing precision error and recall error.

Table 2. The experiment result of operating time (scale=x2)

Method	SRCNN	ESRCNN	DSRCNN	DCSCN
operating time(s)	2.26	3.10	3.77	2.33

5. Conclusion

In this paper, when OCR was applied with Tesseract in a scanned document, super-resolution was performed on small text parts. The precision error and recall error rates of cropped samples are 34.4% and 33.4%, respectively. However, if the text image is downsampled by 0.5 times, the error rate will increase to 90.4%. To solve this problem, we compared the existing deep learning-based super-resolution methods based on text recognition accuracy and operating time. The DSRCNN model has the longest operating time of 3.77s. In terms of operating time, the SRCNN model has the best performance, but the precision error and recall error is 88.9%, 88.6% respectively, which is lower than the other models. As a result, the DCSCN method with the highest recognition accuracy and smallest operating time was applied, and the error rate was reduced by 7.8% and 8.4% based on 0.5 times resized the image, respectively. The text recognition accuracy by tesseract-OCR resulted in better results, and our experiment found that the DCSCN method is the most efficient. Adding it to the OCR preprocessing module can further improve the text recognition accuracy.

ACKNOWLDEGEMENT

The present Research has been conducted by the Research Grant of Kwangwoon University in 2017.

REFERENCES

[1] S. Mori, C.Y. Suen, and K. Yamamoto, “Historical Review of OCR Research and Development,” *Proceedings of the IEEE*, Vol. 80, No. 7, pp. 1029-1058, July 1992.

[2] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*, MIT press, 2016.

[3] K. Greff, R.K. Srivastava, J. Koutnik, B.R. Steunebrink, and J. Schmidhuber, “LSTM: A search space odyssey.” *IEEE transactions on neural networks and learning systems 28*, No. 10, July 2016.

[4] W. shi, J. Caballero, F. Huszar, J. Totz, A.P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, “Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. *IEEE Conference on Computer Vision and Pattern Recognition,* pp. 1874-1883, 2016.

[5] C. Dong, C.C. Loy, K. He, and X. Tang, “Learning a deep convolutional network for image super-resolution., “ *European conference on computer vision”*. Springer, Cham, 2014.

[6] C. Dong, C.C. Loy, and X. Tang, “Accelerating the super-resolution convolutional neural network.,” *European conference on computer vision*. Springer, Cham, 2016.

[7] C. Dong, C.C. Loy, and X. Tang, “Image super-resolution using deep convolutional networks.,” *IEEE transactions on pattern analysis and machine intelligence 38*., No. 2, pp.295-307, June 2015.

[8] L. faning, Z. Xiaoshu, and H. Chunjiao, “Single Image Super-Resolution Restoration Model Using Deep Convolutional Networks.” *Guangxi Sciences*, pp.231-235, 2017.

[9] J. Yamanaka, S. Kuwashima, and T. Kurita, “Fast and accurate image super resolution by deep CNN with skip connection and network in network.” *International Conference on Neural Information Processing*., pp. 217-225, Springer, Cham, November 2017.

[10] W. Sen, and Y. Kejian, “An image scaling algorithm based on bilinear interpolation with CV++” *Journal of Techniques of Automation & Applications*, pp.44-45, 2008.

[11] C. De Boor, “Bicubic spline interpolation.” *Journal of mathematics and physics 41*.” No. 1-4, pp. 212-218, Apr 1962 .

[12] M. Bevilacqua, A. Roumy, C. Guillemot, and M.L. Alberi-Morel, “Low-Complexity Single-Image Super-Resolution based on Nonnegative Neighbor Embedding.” *British Machine Vision Conference*, 2012.

[13] L. Xu, JS. Ren, C. Liu, J. Jia, “Deep convolutional neural network for image deconvolution.”, *Advances in neural information processing systems*., pp. 1790-1798, 2014.

[14] K. He, X. Zhang, S. Ren, J. Sun, “Deep residual learning for image recognition.” *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778, 2016.