

Development of OCR-based Applications to Improve Inventory Management

Min-Seok Seo¹, Jae-Min Lee¹, So-Yeol Lee¹, Dong-Geol Choi^{*1}

¹Undergraduate Student, Department of information and Communication Engineering, Hanbat National University, Daejeon, 34152, Republic of Korea

^{*1}Professor, Department of information and Communication Engineering, Hanbat National University, Daejeon, 34152, Republic of Korea

king_v_@naver.com¹, sy320329@gmail.com¹, jaemin.lee@gmail.com¹, dgchoi@hanbat.ac.kr^{*1}

Article Info

Volume 81

Page Number: 236 - 244

Publication Issue:

November-December 2019

Abstract

Background/Objectives: Inventory management is essential in today's business. Accurate inventory tracking allows company save their time and money for quick and accurate inventory scanning, we suggest OCR-based inventory management software.

Methods/Statistical analysis: The process go through 5 steps: pre-processing, text detection, text classification, text recognition, post-processing. We adopted character-level text detection method instead of word-level's. The annotation of character-level is extracted from word-level's through weakly supervised learning.

Findings: In this way, we experimented with various previous character recognition methods with the same data set to compare performance. Furthermore, we have combined each method, and tested it with various data sets. The results showed generally higher performance than conventional methods. Unlike other image processing tasks that require RGB channels, we have found out gray scaling image in OCR process brings high accuracy of character recognition. You can check this on Table1. We have made the network spatially invariant without data augmentation by using STN(Spatial Transformer Network).

Improvements/Applications: The use of OCR technology has been confined to limited fields such as document translation or license plate recognition due to it's performance. We have improved the accuracy of text recognition by combining the existing OCR methods. Based on our results, we try to increase the use of OCR technology to inventory management system.

Keywords: OCR, Attention, Inventory Management, STN, Text Detection, Text Recognition

Article History

Article Received: 3 January 2019

Revised: 25 March 2019

Accepted: 28 July 2019

Publication: 22 November 2019

I. Introduction

A Deep Learning-based OCR technology has drawn great attention in the academic world in recent years. Despite the highly mature scene text detection technique[1][2][3], the field of scene text recognition technique remains at the level of document translation and copy. We compare the methodologies in the existing scene text recognition technique field to find the methodologies with the highest accuracy and apply them to the inventory management application. The technique is largely divided into pre-processing, recognition, and post-processing step. In the pre-processing step, text is found with a segmentation-based algorithm[4]. In the recognition step, STN(Spatial Transformer Network)[5] converts irregular images filtered from the preprocess to regular images that are easy to recognize. Regular images pass through SRN(Sequence Recognition Network) and output predicted results. In the post-processing process, the recognized characters are refined to suit each situation and the results are return. As a result, we can expect three effects. First, we solved the traditional problem that inventory management efficiency is highly dependent on employees. Next, we improve the poor working conditions in which employees read, write and manage goods with very high concentration. Finally, we have created consistent accuracy and inventory control efficiency that is independent of the environment and staff. By creating an easy working environment, we are looking forward to the effect of developing jobs for the disabled.

II. Related Work

OCR is an image scanner that captures images of human-written or machine-printed characters and converts them into machine-readable characters. Software that converts textual images of documents that can be obtained by image scanning into a format that can be edited by a computer is commonly

called OCR. OCR began as a field of research in artificial intelligence and machine vision. Optical character recognition using optical technologies such as mirrors and lenses and digital character recognition by scanners and algorithms were considered different domains, but the term optical character recognition is now considered to include digital character recognition. Early systems required "training," which meant reading a sample of a font in advance to read that particular font, but now most fonts can be converted with high probability. However, the recognition rate is lower in the environment where various noises exist (irregular background), not in a specific environment, and the application remains at the same level as the document conversion and license plate recognition application. in this section examines various traditional methods of using OCR application. In addition, we search for various optimization methods that are briefly mentioned in existing OCR studies or exist only in code, and find the optimal method by combining them. Section 2.3 compares the accuracy of existing studies with various test data sets.

License Plate Recognition

The license plate recognition system will be recognized after the license plate has been detected on the vehicle. Car imaging, license plate extraction, pre-processing process according to shading or tilting are performed, and license plate recognition is performed according to predetermined reference letters. In particular, the recognition of license plates can be simple. Since the aspect ratio of the license plate is constant at 2:1, it is easy to utilize the license plate by using the aspect ratio. The recognized license plate characters are output

via the user interface. In figure 1 is the pipe line of the algorithm described earlier. Until then, however, OCR technology did not develop much, so it was used only in a given

static environment. Even in the past when there was no development of CNN, OCR was used in many applications, especially license plate recognition.

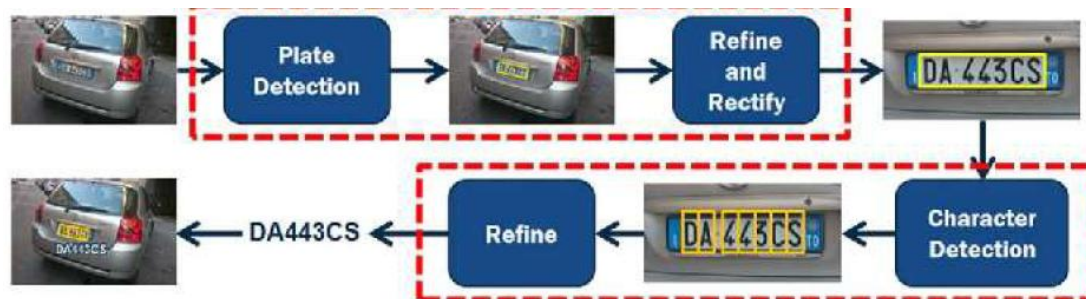


Figure 1. License Plate Recognition pipeline

Document Character Recognition

With the development of CNN, a variety of characters are also recognized. Typical applications are OCR-based document character recognition programs. Document character recognition is a technique that extracts letters from documents and stores them on a computer. Once the image is

entered, remove the background, locate the character area, extract the character, insert it into a specific network, and predict the character. However, still sensitive to noise, there were many differences in accuracy depending on the variety of viewpoints and background. The pipe lines described earlier are described in figure 2.

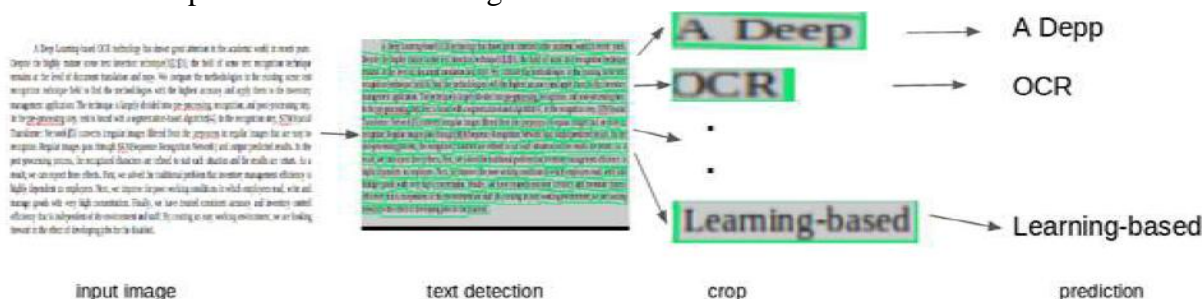


Figure 2. Document Character Recognition pipeline

Compare text recognition model

In this section, we compare various character recognition methods in various public databases to create OCR-based inventory management applications. We also compare our method with the existing methods. The training datasets are MJSynth (MJ) and SynthText (ST). The MJ dataset is a data set for character recognition, which is relatively noise-free and is labeled in word units. The SJ dataset is a text detection-only dataset that is labeled with a bounding box of text and words.

But in our experiment, we wanted to create a recognition model that was robust with background and variety of noise, so we learned the character part of the SJ dataset by cropping it up and learning about the character recognition model. In Figure 1, (A) is MJ data set and (b) is SJ dataset. Test datasets include IIIT, SVT, IC03, IC13, IC15, SP, and CT. In figure 2, (c) shows regular images (IIIT5K, SVT, IC03, IC13) and (d) shows Irregular images (IC15, SVTP, CUTE) real-world datasets. Table 1 is a quantitative comparison

of existing methods and Table 2 is a qualitative comparison. In this Table 2, the poor quality of the image tends to be poorly predicted by any

network. If the image is clear and regular, it tends to predict well.




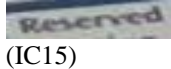




Figure 3. training dataset and test dataset

Table 1. Compare existing methods

Model	Train data	IIIT 300 0	SVT 647	IC03 860 867		IC13 857 1015		IC15 1811 2077		SP 645	CT 288
CRNN[12]	MJ	78.2	80.8	89.4	-	-	86.7	-	-	-	-
RARE[13]	MJ	81.9	81.9	90.1	-	88.6	-	-	-	71.8	59.2
R2AM[15]	MJ	78.4	80.7	88.7	-	-	90.0	-	-	-	-
STAR-Net[14]	MJ+PRI	83.3	83.6	89.9	-	-	89.1	-	-	73.5	-
GRCNN[16]	MJ	80.8	81.5	91.2	-	-	-	-	-	-	-
ATR[17]	PRI+C	-	-	-	-	-	-	-	-	75.8	69.3
FAN[18]	MJ+ST+C	87.4	85.9	-	94.2	-	93.3	70.6	-	-	-
Char-Net[19]	MJ	83.6	84.4	91.5	-	90.8	-	-	60.0	73.5	-
AON[20]	MJ+ST	87.0	82.8	-	91.5	-	-	-	68.2	73.0	76.8
EP[21]	MJ+ST	88.3	87.5	-	94.6	-	94.4	73.9	-	-	-
SSFL[22]	MJ	89.4	87.1	-	84.7	84.0	--	-	-	73.9	62.5
Ours model, RGB	MJ+ST	87.9	87.5	94.9	94.4	93.6	92.3	78.0	74.3	80.2	74.4
Ours model, gray-scale	MJ+ST	89.5	88.4	95.0	95.5	94.0	93.0	77.0	75.2	81.1	77.2

Table 2. Qualitative comparison

Image	CRNN	RARE	STAR-Net	Char-Net	EP	SSFL
 (IIIT5K)	LOCAOOW	LOCACOLA	COCACOLA	COCACOLA	COCAOOLA	COCA6OLA
 (IC03)	PEACOCKS	PEACOCKS	PEACOCKS	PEACOCKS	PEACOCKS	PEACOCKS
 (IC13)	SUWWORS	SUMMERS	SUMMER'S	SUMMERS	SUMMER'S	SUMMCR'S
 (IC15)	ACIRVE	REVENI	REIENT	REIENED	REVENVE	REVERED

 (CUTE)	ACM	ACM	ACM	ACM	ACM	ACM
 (SVT)	RIRS	RIPST	FRST	FRSTI	FWRSI	FWRSI

III. OCR-based Applications to Inventory Management

Many distribution companies, such as Amazon and Coupang, still employ a lot of manpower to check Inventory. Also, many employees complain about poor working conditions and difficulties. We propose OCR-based Inventory Management applications to solve these problems. As shown in Figure 3, The pre-processing process is to re-size the image to fit

our model in the image containing letters, make the letters clearer, or change the three channels to one channel. Detection is divided into CNN-based segmentation and regressions, and the application we proposed used segmentation. The classifier is what detects and predicts the language of the cropped image. The Recognition used an Attention, which recognized letters in cropped images. Finally, in post-processing, the extracted text is modified to suit the situation.

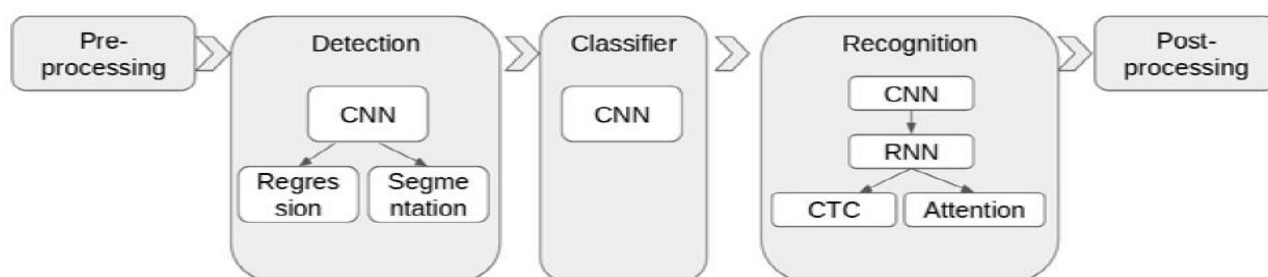


Figure 4. Pipelines for OCR-based inventory management applications.

STN + SRN Modeling

First Inventory management application is adapted in input image into text detection network. Next, the image is resized by the appropriate interpolation method and is

adapted to STN+SRN. STN converts irregular images to regular images. The post-processing process returns the results of inventory management. The figure1 below gives a general overview. In Figure 4 outlines the overall system.

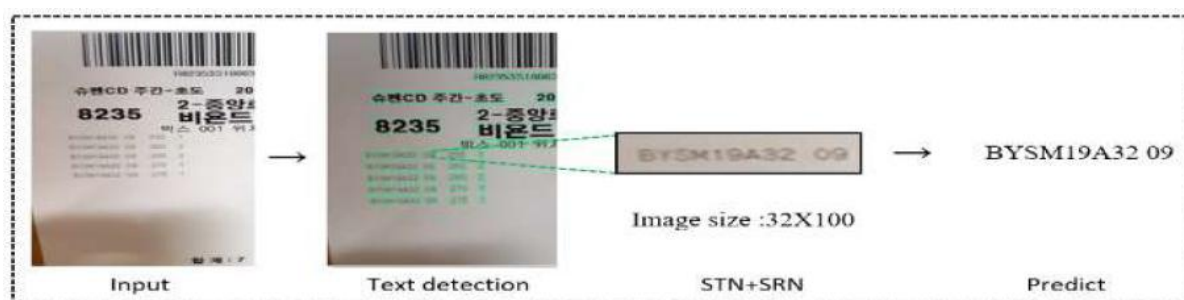


Figure 5. System Outline

Pre-processing

Color information can be noise in OCR because color information is not relevant to character characteristics.

Therefore, the image of three channels is changed to gray scale. For example, green "hello" and yellow "hello" are the same, but the addition of color information can act like noise and interfere with learning. In addition, if the characters in the image are densely packed, the images can be binaryized to increase learning efficiency. Finally, it makes a strong model for variety of viewpoints by making data comments. However, it was not chosen to reverse the word or rotate it to more than 90 degrees. For example, word 'b' and 'd' are different words, but flipping them over can

confuse learning because they look perfectly the same.

Text Detection

A method of detecting scene text in word units has recently emerged, showing promising results. Previous methods, trained with strict word-level bounding boxes, have limitations in expressing text areas in arbitrary form. In this paper, a method is adopted to effectively detect text areas by linking all the associations between each character and the character. Inputs an image and finds text through word level segmentation in Figure 5. As shown, if the letter density is low and the letter size is relatively small, it tends to be difficult to find. It does not distinguish well between the hole and the number zero.



Figure 6. Experience in the real inventory management environment.

Classifier

It is a network that classifies the language that makes up the words that are detected. The reason for classifying languages is that each character of the language has different post-processing methods to consider. For example, if you are in Japanese, you should post-processing it by considering vertical writing. Since the performance of language classification networks is sufficiently high, we adopted MobileNetv2 which has less computing resources. As show in table 1 classifier structure was used.

Improve Image Quality

To pass STN and SRN at once, all images must be resized equally. The feature of the characters included in the image is more in the width component than in the height. Therefore images are resized to a height of 32 width 100. When images are resized, many important features disappear. Therefore, there should be

appropriate interpolation. BICUBIC is the most popular interpolation used in text recognition. However, the LANCZOS interpolation was adopted because accuracy was important in the area of inventory management.

STN

Depending on the angle at which the picture is taken and the lighting, the text may appear bent or tilted. CNNs are not invariant to rotation and scale and more general affine transformations. In preparation for affine transform, you need to perform data aggregation and learn more learning data. STN was adopted to reduce this burden. We also conducted an experiment at MNIST to understand STN intuitively. You can see how MNIST data changes in the direction of learning well without any special customizations. Figure 6 shows MNIST converting affine through learning.

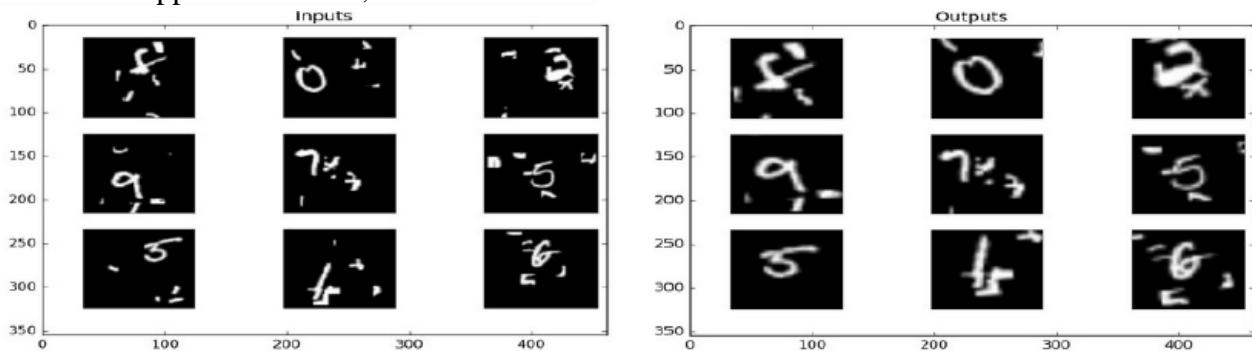


Figure 7. System Outline

SRN

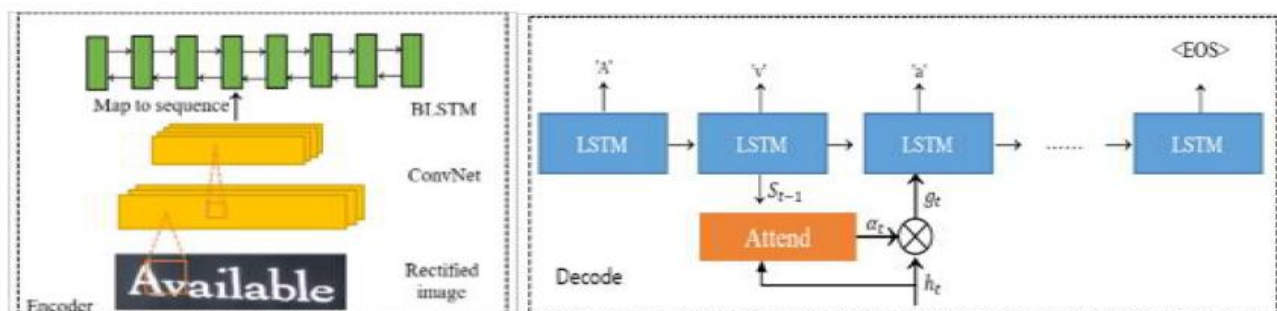


Figure 8. SRN Architecture

The height and width of the image passed through RESNET are reduced by 1/32 and 1/8 ratios respectively. Because all images are 32*100 in size, the height component is 1. Removing the height component will result in a sequence feature. Sequence feature is entered in BiLSTM and the feature passed through BiLSTM predicts characters in the Decoder.

Conclusion

In this paper, we have shown good results in most test datasets by briefly mentioning in existing research or combining various optimization methods that exist only in code. We also found that in Table 2, learning with gray-scale is better than learning with RGB, and we found that LANCZOS has better accuracy than BICUBIC when resizing images. Also, as mentioned above, OCR does not perform image rotation augmentation. Because of this problem, the OCR network is vulnerable to affine transformation. To solve this problem, the affine transformation problem is solved by attaching STN network to the front. By producing the inventory management application using the proposed technology, it is easy for people with insufficient expertise or physical discomfort to manage logistics. It is also expected to reduce the intensity of the work of many salespeople who work at department stores and road shops.

References

- [1] Shi, Baoguang, et al. "Aster: An attentional scene text recognizer with flexible rectification." *IEEE transactions on pattern analysis and machine intelligence* (2018).
- [2] Shi, Baoguang, et al. "Robust scene text recognition with automatic rectification." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016.
- [3] Jaderberg,Max, Karen Simonyan, and Andrew Zisserman. "Spatial transformer networks." *Advances in neural information processing systems*. 2015.
- [4] Deng, Dan, et al. "Pixellink: Detecting scene text via instance segmentation." *Thirty-Second AAAI Conference on Artificial Intelligence*. 2018.
- [5] He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
- [6] Sandler, Mark, et al. "Mobilenetv2: Inverted residuals and linear bottlenecks." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.
- [7] Baek, Youngmin, et al. "Character Region Awareness for Text Detection." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019.
- [8] Baek, Jeonghun, et al. "What is wrong with scene text recognition model comparisons? dataset and model analysis." *arXiv preprint arXiv:1904.01906* (2019).
- [9] Liao, Minghui, et al. "Textboxes: A fast text detector with a single deep neural network." *Thirty-First AAAI Conference on Artificial Intelligence*. 2017.
- [10] Liu, Xuebo, et al. "Fots: Fast oriented text spotting with a unified network." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

- [11] Liu, Wei, et al. "STAR-Net: a spatial attention residue network for scene text recognition." *BMVC*. Vol. 2. 2016.
- [12] B. Shi, X. Bai, and C. Yao. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. In *TPAMI*, volume 39, pages 2298–2304. IEEE, 2017. 1, 2, 4, 5, 10, 11, 12, 13
- [13] B. Shi, X. Wang, P. Lyu, C. Yao, and X. Bai. Robust scenetext recognition with automatic rectification. In *CVPR*, pages 4168–4176, 2016. 1, 2, 3, 4, 5, 10, 11, 12
- [14] W. Liu, C. Chen, K.-Y. K. Wong, Z. Su, and J. Han. Star-net: A spatial attention residue network for scene text recognition. In *BMVC*, volume 2, 2016. 1, 2, 4, 11
- [15] C.-Y. Lee and S. Osindero. Recursive recurrent nets with attention modeling for ocr in the wild. In *CVPR*, pages 2231–2239, 2016. 1, 2, 4
- [16] J. Wang and X. Hu. Gated recurrent convolution neural net-work for ocr. In *NIPS*, pages 334–343, 2017. 1, 2, 4, 10, 11, 12
- [17] X. Yang, D. He, Z. Zhou, D. Kifer, and C. L. Giles. Learning to read irregular text with attention mechanisms. In *IJCAI*, 2017. 1, 2, 3,
- [18] Z. Cheng, F. Bai, Y. Xu, G. Zheng, S. Pu, and S. Zhou. Fo-cusing attention: Towards accurate text recognition in natural images. In *ICCV*, pages 5086–5094, 2017. 1, 2, 3, 4, 5, 10, 11, 13
- [19] S. M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong, and R. Young. Icdar 2003 robust reading competitions. In *ICDAR*, pages 682–687, 2003. 3
- [20] Z. Cheng, Y. Xu, F. Bai, Y. Niu, S. Pu, and S. Zhou. Aon: To-wards arbitrarily-oriented text recognition. In *CVPR*, pages 5571–5579, 2018. 1, 2, 3, 8, 13
- [21] F. Bai, Z. Cheng, Y. Niu, S. Pu, and S. Zhou. Edit probability for scene text recognition. In *CVPR*, 2018. 1, 2, 3, 10, 13
- [22] Y. Liu, Z. Wang, H. Jin, and I. Wassell. Synthetically super-vised feature learning for scene text recognition. In *ECCV*, 2018. 1, 2