Journal of Theoretical and Applied Information Technology
30th April 2022. Vol.100. No 8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

IMPROVING OPTICAL CHARACTER RECOGNITION ACCURACY FOR INDONESIA IDENTIFICATION CARD USING GENERATIVE ADVERSARIAL NETWORK

EMILDA ZHANG[*,1], VINCENT ARDYAN PUTRA[*,2], GEDE PUTRA KUSUMA[*,3]

[*]Computer Science Department, BINUS Graduate Program - Master of Computer Science,

Bina Nusantara University, Jakarta, Indonesia, 11480

E-mail: [1]emilda@binus.ac.id, [2]vincent.putra001@binus.ac.id, [3]inegara@binus.edu

ABSTRACT

e-KTP (Kartu Tanda Penduduk Elektronik) is a national identity card for Indonesian has been widely used not only as an identification but also used in various aspects of life such as for administration, finance-related matters, etc. Detecting information from the ID card field can be done using previous methods with segmentation and text extraction. However, these methods need a proper high-quality captured image. In fact, most of the ID card images are taken using mobile phones' cameras, which do not really produce good quality images all the time. Image enhancement using GAN (Generative Adversarial Network) was also proposed as a method of increasing the accuracy of OCR (Optical Character Recognition). But in previous studies, this method was only limited to document text images with a white background. To overcome these problems, we propose a new pre-processing approach consists of DeblurGAN (Generative Adversarial Network for deblurring image), shadow removal, and binarization to pre-process the image for Tesseract-OCR. We also propose a simple post-processing method for Tesseract-OCR output to extract key-value pairs for each field in e-KTP. By using our approach, we have achieved an average Character Error Rate of 18.82% which is better compared to without pre-processing which is 38.13%.

Keywords: *Optical Character Recognition, Tesseract Model, Generative Adversarial Network, Pre-processing, Character Error Rate*

1. INTRODUCTION

e-KTP (*Kartu Tanda Penduduk Elektronik*) is not only used as an ID card by Indonesians. There are many life's aspects that use data from it, especially in administration, document creation, the process of borrowing money, buying a house, creating a bank account, etc. Along with technological development especially in computer vision, data from e-KTP can be extracted from the captured image from a mobile phone's camera.

There are several obstacles in the recognition process using current technology from dealing with low-light, hazy, rainy images captured in outdoor environments, image orientation, hand's reflection when capturing the image, etc. To optimize this process, two processes can be done, applying a good image pre-processing [1] for the input and post-processing [2] for the text recognition output.

Text recognition process can be done using Optical Character Recognition (OCR) algorithm with Tesseract-OCR [3]. Tesseract-OCR algorithm has a high level of accuracy for scanned documents and images with good quality and clear as the input. Therefore, the right image pre-processing method is so needed to generate a high-quality image that the OCR process [1] can give an expected output which is also supported by a good post-processing algorithm.

A lot of research has been done to increase OCR accuracy by enhancing the input image through pre-processing. Bieniecki, Grabowski, & Rozenberg [1] proposed a method focusing on preprocessing the image using images from a digital camera, this method adjusts the orientation of text areas of the image as the pre-processing technique for OCR. In 2019, Satyawan et al. [4] also applied a combination of image processing for the OCR process. These processes consist of pre-processing by changing the image into grayscale, binarize the image using thresholding, use Sobel as the edge detection, applying a morphological transformation (consists of dilation, erosion, opening on, closing a by b, and Tophat), applying Otsu technique to represent the image into 2D gray-level intensity function, text area detection, and segmentation. Akhter, Bhuiyan, & Uddin [5] also using image enhancement and background removal for the OCR process. This method used a very simple spot removal to enhance the image that has already binarized using the Otsu technique. So, it will not output optimal accuracy of OCR for noisy images. Improving OCR accuracy for ID card images based on the segmentation method was proposed by Soeseno & Liliana [6] and Thanh & Trong [7], but this method is limited to images with high-quality images taken at good angles.

Ian et al. introduced a framework called Generative Adversarial Networks (GAN) [8] which a rising issue in deep learning for image enhancement recently. Lat & Jawahar [9], Su et al. [10], and Kong & Wang [11] proposed a method for increasing the accuracy of Tesseract OCR by applying GAN-based image pre-processing to perform a resolution enhancement on document text images which has a white background. However, this method is not suitable for a more complex dataset of text images such as ID cards or images taken from mobile phones' cameras. In addition, pre-processing using GAN to increase image resolution for OCR tasks was too limited to a specific dataset in previous studies.

We believe that the use of pre-processing using a combination of GAN as well as other pre-processing methods can cover a wider dataset and applicable not only for document scanned text images but also raw image which has complex background. Therefore, the contribution of this paper includes the following contents. A method for enhancing image quality in pre-processing based on machine learning using combinations of techniques which are: Binarization, Deblurring, and shadow removal. We use DeblurGAN for deblurring the image and used Wasserstein-loss as the loss function. A new post-processing algorithm to improve the output of Tesseract-OCR for Indonesian ID cards. Our dataset of Indonesian ID cards which are captured using a mobile phone’s camera with different lighting situations and different environments.

2. RELATED WORKS

Text image recognition using OCR has been a great area for researchers these couple of years. A lot of methods have been proposed to enhance the accuracy of OCR whether it is using Tesseract, ABBYY FineReader, or HANWANG. A & N [12] proposed a method for preprocessing the image using local brightness and contrast adjustment method to effectively handle lighting variations and the irregular distribution of image illumination, and use an optimized grayscale conversion algorithm to transform the document image to greyscale level. At last, this method sharpens the useful information in the resulting grayscale image using the Un-sharp Masking method. An optimal global binarization approach was also used to prepare the final document image for OCR recognition. Bieniecki, Grabowski, & Rozenberg [1] also proposed a method focusing on preprocessing the image using images from a digital camera, this method adjusts the orientation of text areas of the image as the pre-processing technique for OCR.

Akhter, Bhuiyan, & Uddin [5] also used image enhancement and background removal for the OCR process. This method used a very simple spot removal to enhance the image that has already binarized using the Otsu technique. In 2019, Satyawan et al. [4] also applied a combination of image processing for the OCR process. These processes consist of pre-processing by changing the image into grayscale, binarize the image using thresholding, use Sobel as the edge detection, applying a morphological transformation (consists of dilation, erosion, opening on, closing a by b, and Tophat), applying Otsu technique to represent the image into 2D gray-level intensity function, text area detection, and segmentation. Thanh & Trong [7] proposed a method by focusing on OCR for Vietnamese ID cards using Tesseract. This method analyzes image structure and applying image pre-processing. The pre-processing consists of tilt adjusting, noise filtering, background removal, color channel analyzing, connected component analyzing, mask line creating, table structure analyzing, and binary image.

Background elimination approach also proposed by Shen & Lei [13] which enhances document images by utilizing the brightness and chromaticity as contrast parameters. Then convert color images to gray and threshold it to eliminate the background. This method also uses Tesseract, ABBYY FineReader, and HANWANG for the OCR. Soeseno & Liliana [6] proposed a segmentation method for OCR process for Indonesian National ID Cards. The area of the National ID Card is determined using the blueness level of every pixel. Then Canny Edge Detection is used to mark the edges of the National ID Card, which is later processed by using dilation to thicken the edges. The next step involves converting the image into a binary image. By dividing each edge into twelve partitions and using lines to mark each edge. Vamvakas, Gatos, Stamatopoulos, & Perantonis [14] proposed a complete OCR methodology for historical documents, either printed or handwritten without any knowledge of the font. This method applies 3 steps: the pre-processing step that includes image binarization and enhancement, top-down segmentation approach to detect text lines, words, and characters that will be saved to a database for the recognition process, and segmentation approach is used as part of the recognition process. In 2015, Hartanto, Sugiharto [15] proposed a method using Template Matching Correlation Algorithm for OCR. Before applying the algorithm, this method pre-processed the input image by applying binarization, segmentation, and normalization.

An approach for post-processing was also proposed by Llobet, Navarro-Cerdan, Perez-Cortes, & Arlandis [2]. Instead of focusing on the pre-processing of the image before the OCR process, Llobet et al. [2] focused on post-processing of the OCR result by using a sequence of vectors of a posteriori class probability to build a WFST (Weighted Finite-State Transducers) that is then composed with independent WFSTs for the error and language models. A method based on Generative Adversarial Network (GAN) was also proposed by Lat & Jawahar [9] which by focusing on document images, Lat & Jawahar enhance the OCR accuracy with super-resolution image generated based on Super-Resolution Generative Adversarial Network (SRGAN) as the pre-processing. The same as Lat & Jawahar [9], Su et al. [10] also proposed a method to improve text image resolution based on Generative Adversarial Network (GAN) using Conditional Generative Adversarial Network (cGAN) for OCR process, and Kong & Wang [11] also proposed a method for increasing the accuracy of Tesseract OCR by applying GAN based image pre-processing using SRGAN.

In this paper, we focus on improving Tesseract-OCR accuracy for Indonesian ID cards by pre-processing and post-processing the input image. Previously proposed methods based on Generative Adversarial Network (GAN) [9][10][11] seems very convincing for the pre-processing image to generate a clear image for Tesseract-OCR. Hence, we want to improve the quality of images using combinations of pre-processing techniques which include GAN method. And finally, we use simple post-processing techniques using regex to enhance Tesseract-OCR accuracy for analyzing Indonesian ID cards.

3. PROPOSED METHOD

Our approach consists of 3 general stages, namely pre-processing, character recognition using tesseract, and post-processing. In pre-processing stages, we apply different combinations of pre-processing techniques for enhancing image quality as the input. There are 3 main techniques that applied: Deblurring, shadow removal, and Binarization.

These couple of years, a machine learning technique for enhancing image quality called generative adversarial network (GAN) [8] grab our attention, so a method by Kupyn, Budzan, Mykhailych, Mishkin, & Matas [16] based on GAN is adopted for deblurring images, called DeblurGAN. DeblurGAN is an end-to-end learning method for motion deblurring. The learning is based on a conditional GAN (cGAN) and the content loss. This method used WGAN-GP [17] as the critic (discriminator in WGAN [17]) function with gradient penalty and perceptual loss. By using WGAN-GP the training process will be more stable. The generator used in DeblurGAN consists of two convolution blocks with stride ½, nine residual blocks (ResBlocks), and two transpose convolution blocks. Each ResBlock consists of a convolution layer, instance normalization layer, and ReLU activation. DeblurGAN will be used as the deblurring technique in this paper which described in the Experiment Section. We used the pre-trained model provided by the author and using Wasserstein-loss from WGAN [17] as the loss function. Therefore, we applied DeblurGAN to the dataset for deblurring them with result shown in Figure 1.

Because the dataset used in this paper was taken from a mobile phone's camera with different lighting, the image contains a lot of reflections. To improve the quality of such an image, we applied shadow removal after deblurring the image. First, we split channels of the image into their individual planes, then for each plane we dilate them and blur them to make it smoother using median blur. After that, we calculate the deficiency of the planes by subtracting the absolute value of the plane and the blurred image by 255. In the last step, we normalize it and append it into an array before merging back all the planes as a single image (Figure 2).

At last, for binarization, we used 3 different thresholding methods to binarize the image, which is TRUNC thresholding, TOZERO thresholding, and we also tried Otsu thresholding which shown by A & N [12], Akhter et al. [5], and Satyawan et al. [4] will output a better result for binarizing images. In

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

TRUNC thresholding methods, values lower than the threshold value are unchanged, but any value higher is set to the threshold value which represents by the following:

$$\mathrm{dst}(x,y)=\begin{cases} \text{maxval} & \text{if } \mathrm{src}(x,y)>\text{thresh} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

TOZERO thresholding adjusts pixel values lower than the threshold value to set to the 0 and value above are unchanged which represent by the following:

$$\mathrm{dst}(x,y)=\begin{cases} \mathrm{src}(x,y) & \text{if } \mathrm{src}(x,y)>\text{thresh} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Otsu thresholding is an adaptive thresholding for image binarization. From pixel value 0 to 255, Otsu threshold find the optimal threshold values by calculating and evaluating their between-class variance (or within-class variance). Result after binarizing the images are shown in Figure 3.

Then, the image is then processed using Pytesseract (Python Tesseract version 5.0.0 alpha) to extract text which can be seen in Table 1. Pytesseract (Python Tesseract) is a tool that enables us to perform OCR using python. It is a wrapper for the open-source Tesseract-OCR engine which was originally developed by Hewlett Packard in 1985 [3]. Tesseract-OCR is currently maintained by Google under the Apache 2.0 license [3] and is now considered as one of the most accurate and widely used OCR-engine [18].

After we get the result from Pytesseract, we split the result into lines and perform key-value pairs extraction using regex. We used Regex.search for fields that contain limited types of content, such as “Agama”, “Jenis Kelamin”, “Golongan Darah”, “Kewarganegaraan”, “Status”, and “Berlaku Hingga”. If the regex.search failed to find the matched value, then we will split the string using a colon. We also use this method to find other field key values detected by Tesseract and then split them by colon. If the string doesn’t contain a colon, we will split it with regex matched value and take the last array. If the regex fails and the string doesn’t contain a colon, we will return the initially detected string based on the array index. After that, we will process each field's value by removing extra spaces, lowercase characters, symbols, and punctuations, but we don’t remove ”:.,-” punctuation, because that is part of the content in the ID card. We filter each field’s value according to its own characteristic. For example, we filter the “NIK” field’s value so that it only contains numbers, while the “Tempat/ Tgl Lahir” field’s value we will filter out extra spaces, lowercase characters, symbol, and punctuations, but still keeping ”:.,-” because they are parts of the formats in “Tempat/ Tgl Lahir” field. The result example can be seen in Table 1.

4. EXPERIMENTS

4.1 Dataset

The dataset used in this paper contains 100 Indonesian ID cards which were taken by mobile phone’s camera and used for testing purposes only. Because the dataset images contain not only the ID card (there are also faces, hand’s reflection, blobs, different lighting conditions, etc.) as shown in Figure 4, so we rectified the images into 600x400 so the images will only contain the ID card, then we masked the area of the right side of Indonesian ID card which contains profile picture, city name, date, and signature into dominant colors detected from the images (Figure 5b). The masking process is used to improve the Tesseract-OCR result so that it only detects text from the needed information.

4.2 Experimental Design

We applied 5 different combinations of different techniques of pre-processing to the input images before doing the OCR process to evaluate each combination. The result of the pre-processing is shown in Figure 5. All of the binarization steps used 3 different thresholding, which are TRUNC, TOZERO, and Otsu thresholding. For TRUNC thresholding we applied thresholding within the range of 100 – 250. The result showed that using 200 and 215 fits best for each image. TOZERO showed the best result using 127 threshold value. The Detail process of our pre-processing technique can be seen in Table 2.

Then we applied a post-processing algorithm after the Tesseract-OCR process the image. The post-processing result will be export to a text file. Lastly, we measure the CER (Characters Error Rate) for each image by comparing extracted output values to its ground truth. CER is computed with the minimum number of operations required to transform the ground truth into the output. The CER is defined as:

$$\mathrm{CER} = \frac{(i+s+d)}{n} \times 100\% \quad (3)$$

where n is the number of characters in the reference text, i is the number of characters inserted, s is the number of characters substituted, and d the number of characters deleted [19][20].

During CER calculation, we will ignore the keys and only compute the values of each field. We provide one example of how we compute CER on one image which can be seen in Table 3. After we

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

calculate all CER of each image, we will then calculate the average CER by dividing the sum of all CER by the number of images in the dataset.

4.3 Experimental Results

Using combination of shadow removal and binarization, we effectively remove the background image and thus the readability of e-KTP improves accordingly. Figure 5 shows example of original images and processed images. The background in Figure 5a is effectively removed by using our method which can be seen in Figure 5d. Table 4 shows the result of combinations of techniques that we used to improve Tesseract-OCR accuracy. The results are divided into 4 columns based on different thresholding method.

Although at first glance it doesn't seem to be any different between original image in Figure 5c and deblurred image in Figure 5e, Table 4 shows that DeblurGAN improve Tesseract-OCR accuracy by comparing Binarization and Deblurring + Binarization or Shadow removal + Binarization and Deblurring + Shadow removal + Binarization.

We also found in our study that binarized shadow removal results in better accuracy than performing either "shadow removal" or "binarized" alone. This is the reason why we choose to do shadow removal first, and then continue with binarization. By using combination of shadow removal and binarization, we reduce the average CER up to 18.93%. This result show that combination of shadow removal and binarization is very impactful for increasing OCR accuracy.

As the comparison with A & N [12], Akhter et al. [5], and Satyawan et al. [4], we can see from Table 4 shows Otsu thresholding does not perform well in our dataset. We assume that is because there are some datasets containing reflections. If there are a lot of reflections, the image will contain a lot of noise and the text area cannot be detected because of the reflection overlap with it as shown in Figure 6. Our research shows that TRUNC threshold works well in Tesseract-OCR. Combining deblurring, shadow removal, and binarization using TRUNC thresholding, we managed to get the lowest CER of 18.82%. techniques using regex to enhance Tesseract-OCR accuracy for analyzing Indonesian ID cards.

Based on these results, GAN is confirmed to be capable of improving the accuracy of Tesseract-OCR's OCR. Utilizing GAN in the pre-processing stage will yield the best results when combined with shadow removal and binarization with TRUNC Thresholding.

5. CONCLUSIONS

Most of the ID card images are taken using mobile phones' cameras, which do not really produce good quality images all the time, thus affecting accuracy in character recognition. To solve this problem, we try to find the best pre-processing technique to enhance image quality.

In this experiment we use GAN as the image enhancer in the pre-processing process which can effectively increase Tesseract-OCR performance, specifically DeblurGAN. We also use shadow removal and binarizing the image as the additional pre-processing methods.

Our method showed that using TRUNC threshold in the binarization process works well with Tesseract-OCR. The CER improved also significantly by applying shadow removal after DeblurGAN. Combining those 3 pre-processing methods, we managed to get the lowest CER of 18.82%.

There are many rooms to improve the OCR result, researchers may try to apply different types of GAN, combine multiple GAN, or find another pre-processing technique to reduce noise, dots, and blobs of the image to improve the dataset quality. Researchers may also try to use different post-processing method or OCR method to achieve higher accuracy.

REFERENCES:

[1] W. Bieniecki, S. Grabowski, and W. Rozenberg, "Image preprocessing for improving OCR accuracy", *Proceeding of the 3rd International Conference of Young Scientists 'Perspective Technologies and Methods in MEMS Design', MEMSTECH 2007*, 2007, pp. 75–80.

[2] R. Llobet, J. R. Navarro-Cerdan, J. C. Perez-Cortes, and J. Arlandis, "OCR post-processing using weighted finite-state transducers", *Proceedings - International Conference on Pattern Recognition*, 2010, pp. 2021–2024.

[3] S. Bhaskar, N. Lavassar, and S. Green, "Implementing Optical Character Recognition on the Android Operating System for Business Cards", *CogITo Smart Journal*, Vol. 2, No. 2, 2016, pp. 194.

[4] W. Satyawan, M. Octaviano Pratama, R. Jannati, G. Muhammad, B. Fajar , H. Hamzah, … K. Kristian, "Citizen Id Card Detection using Image Processing and Optical Character Recognition", *Journal of Physics: Conference*

Series, 1235(1), 2019.

[5] M. R. Akhter, M. H. Bhuiyan, and M. S. Uddin, "Extraction of words from the national ID cards for automated recognition", *International Conference on Graphic and Image Processing (ICGIP 2011)*, 8285(October), 828521, October, 2011.

[6] J. H. Soeseno and Liliana, "Segmentasi Area KTP dari Image untuk Otomatisasi Pembacaan Data", *Jurnal Infra Petra*, Vol 5, No. 1, 2017, pp. 1–5.

[7] T. N. T. Thanh and K. N. Trong, "A method for segmentation of Vietnamese identification card text fields", *International Journal of Advanced Computer Science and Applications*, Vol. 10, No. 10, 2019, pp. 415–421.

[8] Ian J. Goodfellow, J. Pouget-Abadie, M. Mirza, , B. Xu, D. Warde-Farley, S. Ozair, … Y. Bengio, "Generative Adversarial Nets", *Advances in Neural Information Processing Systems*, December, 2017, pp. 4089–4099.

[9] A. Lat and C. V. Jawahar, "Enhancing OCR Accuracy with Super Resolution", *Proceedings - International Conference on Pattern Recognition*, August, 2018, pp. 3162–3167.

[10] X. Su, H. Xu, Y. Kang, X. Hao, G. Gao, and Y. Zhang, "Improving text image resolution using a deep generative adversarial network for optical character recognition", *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, 2019, pp. 1193–1199.

[11] J. Kong and C. Wang, "Resolution Enhancement for Low-resolution Text Images Using Generative Adversarial Network", *MATEC Web of Conferences*, 2018, pp. 246.

[12] A. El Harraj and N. Raissouni, "OCR Accuracy Improvement on Document Images Through a Novel Pre-Processing Approach", *Signal & Image Processing : An International Journal*, Vol.6, No.4, 2015, pp. 01–18.

[13] M. Shen and H. Lei, "Improving OCR performance with background image elimination", *2015 12th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2015*, 2015, pp. 1566–1570.

[14] G. Vamvakas, B. Gatos, N. Stamatopoulos and S. J. Perantonis, "A Complete Optical Character Recognition Methodology for Historical Documents," *2008 The Eighth IAPR International Workshop on Document Analysis Systems*, 2008, pp. 525-532.

[15] S. Hartanto, A. Sugiharto, and S. N. Endah, "Optical Character Recognition Menggunakan Algoritma Template Matching Correlation", *Jurnal Masyarakat Informatika*, Vol. 5, No. 9, 2015, pp. 1-12.

[16] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8183–8192.

[17] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GaN". *ArXiv*, 2017.

[18] M. G. Marne, P. R. Futane, S. B. Kolekar, A. D. Lakhadive, and S. K. Marathe, "Identification of optimal Optical Character Recognition (OCR) engine for proposed system", *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, 2018, pp. 1–4.

[19] I. S. MacKenzie amd R.W. Soukoreff, "A character-level error analysis technique for evaluating text entry methods", *ACM International Conference Proceeding Series*, Vol. 31, No. 1, 2002, pp. 243–246.

[20] J. Read, E. Mazzone, and M. Horton, "Recognition errors and recognizing errors - Children writing on the tablet PC", *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 3585 LNCS(I), 2005, pp. 1096–1099.

(a) (b)

Figure 1: Image Before And After Deblurred: (a) Before Deblurred (b) After Deblurred

(a) (b)

Figure 2: Deblurred Images Before And After Shadow Removal: (a) Before Shadow Removal (b) After Shadow Removal

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

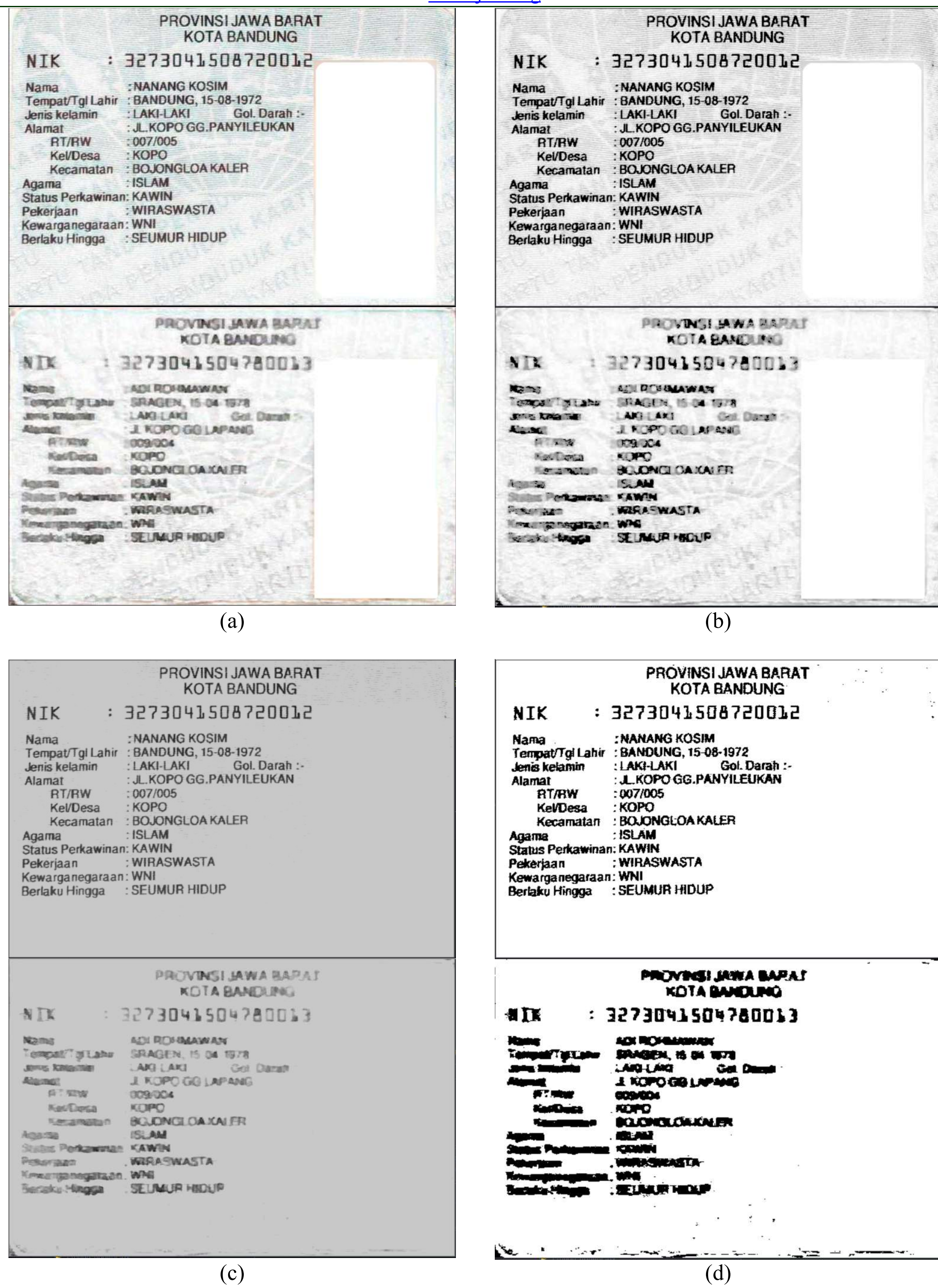


Figure 3: Binarize Image With Different Thresholding Methods: (a) Affter Shadow Removal (b) TOZERO Thresholding (c) Otsu Thresholding

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 1: Post Processing Results

	Dataset 1	Dataset 2

Ground Truth	{
	"0": "PROVINSI JAWA BARAT",
	"1": "KOTA BANDUNG",
	"NIK": "3273041508720012",
	"nama": "NANANG KOSIM",
	"tempat_tanggal_lahir": "BANDUNG, 15-08-1972",
	"jenis_kelamin": "LAKI-LAKI",
	"gol_darah": "-",
	"alamat": "JL.KOPO GG.PANYILEUKAN",
	"rt_rw": "007/005",
	"kel_desa": "KOPO",
	"kecamatan": "BOJONGLOA KALER",
	"agama": "ISLAM",
	"status_perkawinan": "KAWIN",
	"pekerjaan": "WIRASWASTA",
	"kewarganegaraan": "WNI",
	"berlaku_hingga": "SEUMUR HIDUP"
	}
		{
		"0": "PROVINSI JAWA BARAT",
		"1": "KOTA BANDUNG",
		"NIK": "3273041504780013",
		"nama": "ADI ROHMAWAN",
		"tempat_tanggal_lahir": "SRAGEN, 15-04-1978",
		"jenis_kelamin": "LAKI-LAKI",
		"gol_darah": "-",
		"alamat": "JL.KOPO GG.LAPANG",
		"rt_rw": "009/004",
		"kel_desa": "KOPO",
		"kecamatan": "BOJONGLOA KALER",
		"agama": "ISLAM",
		"status_perkawinan": "KAWIN",
		"pekerjaan": "WIRASWASTA",
		"kewarganegaraan": "WNI",
		"berlaku_hingga": "SEUMUR HIDUP"
		}

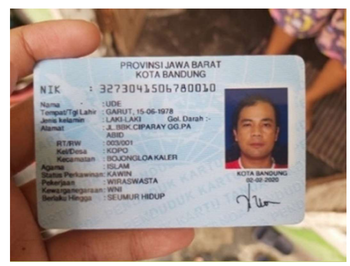
OCR Output using Pytesseract	PROVINSI JAWA BARAT KOTA BANDUNG NIK : 32730415087?20012 Nama INANANG KOSIM TempatTgi Lahir : BANDUNG, 15-08-1972 Jenis kelamin : LAKI-LAKI Gol. Darah :- Alamat : JL.KOPO GG.PANYILEUKAN RTRW :007/005 KeVDesa :KOPO Kecamatan : BOJONGLOA KALER Agama : ISLAM Status Perkawinan: KAWIN Pekerjaan :WIRASWASTA Kewarganegaraan: WNI Berlaku Hingga: SEUMUR HIDUP	PROVINSI JAWA RAPAT KOTA BANDUNG NIK : 3073041508 ?80053 Nam3 ADI ROHMAWAN Tempat yisne SRAGEN, 1S Oa 1573 us Kemasan AKG LAKI Gor Daan Mam 1 KOPO GG LAPANG 87 02006 KecTarsa OPO meramatn — BOJONGIOAXAIFR Dalan ISLAM Sisam Penaunnan KAWIN Senarman .WARASWASTA Kema ganegaraso. WNI Berak Hingga SEUMUR HIDUP be .
Key-Value Pairs Extracted	{ "0": "PROVINSI JAWA BARAT", "1": "KOTA BANDUNG", "NIK": "3273041508720012", "nama": "INANANG KOSIM", "tempat_tanggal_lahir": "BANDUNG, 15-08-1972", "jenis_kelamin": "LAKI-LAKI", "gol_darah": "-",	{ "0": "PROVINSI JAWA RAPAT", "1": "KOTA BANDUNG", "NIK": "307304150880053", "nama": "N ADI ROHMAWAN", "tempat_tanggal_lahir": "Tempat yisne SRAGEN, 1S Oa 1573", "jenis_kelamin": "LAKI-LAKI", "gol_darah": "-", "alamat": "M 1 KOPO GG LAPANG",

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

	"alamat": "JL.KOPO GG.PANYILEUKAN", "rt_rw": "007/005", "kel_desa": "KOPO", "kecamatan": "BOJONGLOA KALER", "agama": "ISLAM", "status_perkawinan": "KAWIN", "pekerjaan": "WIRASWASTA", "kewarganegaraan": "WNI", "berlaku_hingga": "SEUMUR HIDUP" }	"rt_rw": "87 02006", "kel_desa": "KT OPO", "kecamatan": "BOJONGIOAXAIFR", "agama": "ISLAM", "status_perkawinan": "KAWIN", "pekerjaan": "S WARASWASTA", "kewarganegaraan": "WNI", "berlaku_hingga": "SEUMUR HIDUP" }
CER	**0.58%**	**29.70%**

(a)

(b)



(c)

(d)

(e)

(f)

Figure 4: Datasets with Different Conditions: (a) Non Frontal View (b) With Hand's Reflection (c) Contains Hand (d) Contains Lighting On Top Of The Text Area (e) Reflections (f) Blurry Image

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 2: Combination Of Pre-processing Techniques

Technique	Description
Original	We rectified the input image into 600x400 and masked it in the area of the right side of Indonesian ID card which contains profile picture, city name, date, and signature into blue color.
Binarization	We rectified and masked the input image using the same technique as the original, and then we binarized it. We tried 3 different thresholding, which are TRUNC, TOZERO, and OTSU.
Shadow removal + Binarization	We rectified and masked the input image using the same technique as the original, and then we applied shadow removal and binarized it using 3 different approaches (TRUNC, TOZERO, and OTSU).
Deblurring + Binarization	We rectified and masked the input image using the same technique as the original, and then we applied DeblurGAN [16] and binarized it using 3 different approaches (TRUNC, TOZERO, and OTSU).
Deblurring + Shadow removal + Binarization	We rectified and masked the input image using the same technique as the original, and then we applied DeblurGAN [16], shadow removal, and binarized it using 3 different approaches (TRUNC, TOZERO, and OTSU).

Figure 5: Image Pre-processing Results: (a) Original (b) Original With Image Masking (c) Binarization (d) Shadow Removal + Binarization (e) Deblurring (DeblurGAN) + Binarization (f) Deblurring (DeblurGAN) + Shadow Removal + Binarization

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 3: CER Calculation On One Dataset

Ground truth	Output

{
"0": "**PROVINSI JAWA BARAT**",
"1": "**KOTA BANDUNG**",
"NIK": "**3273041508720012**",
"nama": "**NANANG KOSIM**",
"tempat_tanggal_lahir": "**BANDUNG, 15-08-1972**",
"jenis_kelamin": "**LAKI-LAKI**",
"gol_darah": "-",
"alamat": "**JL.KOPO GG.PANYILEUKAN**",
"rt_rw": "**007/005**",
"kel_desa": "**KOPO**",
"kecamatan": "**BOJONGLOA KALER**",
"agama": "**ISLAM**",
"status_perkawinan": "**KAWIN**",
"pekerjaan": "**WIRASWASTA**",
"kewarganegaraan": "**WNI**",
"berlaku_hingga": "**SEUMUR HIDUP**"
 	
}
	{
	"0": "**PROVINSI JAWA BARAT**",
	"1": "**KOTA BANDUNG**",
	"NIK": "**3273041508720012**",
	"nama": "**INANANG KOSIM**",
	"tempat_tanggal_lahir": "**BANDUNG, 15-08-1972**",
	"jenis_kelamin": "**LAKI-LAKI**",
	"gol_darah": "-",
	"alamat": "**JL.KOPO GG.PANYILEUKAN**",
	"rt_rw": "**007/005**",
	"kel_desa": "**KOPO**",
	"kecamatan": "**BOJONGLOA KALER**",
	"agama": "**ISLAM**",
	"status_perkawinan": "**KAWIN**",
	"pekerjaan": "**WIRASWASTA**",
	"kewarganegaraan": "**WNI**",
	"berlaku_hingga": "**SEUMUR HIDUP**"
	
	}

Calculation Insert (i) : 0 Detete (d) : 1 Substitude (s) : 0 Number of characters in the ground truth (n) : 171 CER = $\frac{0+1+0}{171} x\ 100\% = 0.58\%$	

Table 4: Average CER Results Of Tesseract-OCR

	Average CER (%)			
	TRUNC Thresholding	**TOZERO Thresholding**	**OTSU Thresholding**	**Non-Thresholding**
Original	**-**	**-**	**-**	**38.13**
Binarization	**30.56**	51.64	41.27	**-**
Shadow Removal	48.95	48.95	48.95	-
Shadow removal + Binarization	**19.20**	20.81	33.43	**-**

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Deblurring + Binarization	**29.16**	52.88	40.92	**-**
Deblurring + Shadow removal + Binarization **(Proposed method)**	**18.82**	19.65	33.06	**-**

(a) (b)

Figure 6: Otsu Thresholding Result For An Image Dataset That Has Reflection: (a) Before Otsu Thresholding and (b) After Otsu Thresholding