Indonesian ID Card Recognition using Convolutional Neural Networks

M Octaviano Pratama, Wira Satyawan, Bagus Fajar, Rusnandi Fikri, Haris Hamzah
Artificial Intelligence Department
PT Premier Optima Sattiga
Jakarta, Indonesia
{ octav, wrsatya, bagus, rusnandi, haris } @premieroptima.com

***Abstract*— Indonesian ID Card can be used to recognize citizen of Indonesia identity in several requirements like for sales and purchasing recording, admission and other transaction processing systems (TPS). Current TPS system used citizen ID Card by entering the data manually that means time consuming, prone to error and not efficient. In this research, we propose a model of citizen id card detection using state-of-the-art Deep Learning models: Convolutional Neural Networks (CNN). The result, we can obtain possitive accuracy citizen id card recognition using deep learning. We also compare the result of CNN with traditional computer vision techniques.**

***Keywords*— Indonesia Citizen Id Card, Deep learning, Convolutional Neural Networks**

I. INTRODUCTION

Indonesian id card is used by various companies with several purposes like for sales and purchasing recording, campus admission system, online booking and other purposes. Almost current input process of id card was done by using manually input from clerk or admin. The lack of this system is not efficient, time consuming and prone to error. There are several techniques involving automatically process of citizen id card recognition like Optical Character Recognition (OCR) [1] to recognize fields value in citizen id card image. Currently the growth of deep learning has become state-of-the-art of various fields like text processing [2], music classification [3] until image recognition [4]. The last aforemention is used not only to classify or recognize given image but also used to recognize character in image. One of deep learning model that can be used as text recognition in image is Convolutional Neural Networks (CNN) [5]. The performance of CNN to classify or recognize text in image has been proved in various experiments like from Lai [2], The process of recognition character in image starting from extracting text features in images using two layers: Convolutional layers and Sub sampling layer, then it will be forwarded into classification layer: softmax to determine the final class or text in image.

Based on above explanation, this study aim is to create system or model in order to recognize Indonesian citizen id card using one of deep learning technique: Convolutional Neural Networks (CNN). The contributions of this research are: (1) the propose experiments to recognize citizen id card fields like number, name and address using CNN, (2) preprocessing techniques of recognition process, (3) dataset contains alphanumeric character from thousand Indonesian citizen id card. this research is expected can be used by industry to input the fields in Indonesian id card by using automatically input process.

II. RELATED WORK

TABLE I. RELATED WORK

Ref	Related Work	
	Model	***Tasks***
[6]	Bileving, RLSA	Indonesian Id Card Detection
[7]	Template Matching	Indonesian Id Card Detection
[5]	Convolutional Neural Networks	Image Recognition
[8]	Histogram	document segmentation using histogram analysis
[9]	Morphological Transformation	Morphological Transformation in Image Prcoessing

There are several empirical researches related to image recognition. Most of them used image processing and optical character recognition (OCR) models. Less id card data is required by using OCR and image processing techniques, reversely when using machine learning or deep learning, much data is required in order to be recognized.

III. METHOD

Fig. 1. Workflow of Research

The workflow of method in this research can be seen in figure 1. The process starting from data collection. We collected citizen id card, then thousand citizen id cards are preprocessed using various image processing techniques. After having appropriate pre-processed image, text area in citizen id card is extracted and separate per characters. The next step classification model using Convolutional Neural Networks (CNN) is perfromed to recognize each character in citizen id card. The last step, all of recognized character will be merged to produce complete recognized fields in citizen id card.

For data collection, we collected almost a thousand citizen id card data from various companies. The dataset then pre-processed by using Grayscale techniques to convert RGB layer into grayscale, then thresholding (Eq. 1) is perfromed to convert grayscale image $G(x, y)$ into binary by selecting appropriate threshold.

$$I(x,y)\begin{cases}1, if\ G(x,y) > T\\ 0, otherwise\end{cases} \quad (1)$$

After having binary image representation of id card, morphological transformation is performed by using dilation, erotion, opening and closing to produce image without noise. The result of pre-processing then forwarded into text area extraction and separation. We used automatically text area separation by forming kernel (the area or box contain character). The kernel is fixed sized move from left to right (Figure 2). We determine the coordinate of fields: id card number, id card name and id card address as follow:

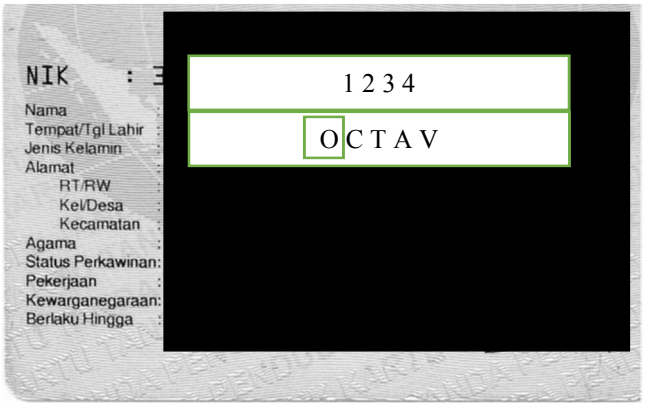

Fig. 2. Citizen Id Card Text Area Extraction and Separation. Box inside name in id card is kernel that move from left to right

When the kernel find character inside of identity card, then it will be crop and save into new file. Let say the name "OCTAV" will be saved into 5 different file JPEG type in the same folder. On the other side, when having name "WIRA" it will be saved into 4 different file JPEG in the same folder but different with the first one. All of the file is scaling into the same size 32 x 32 pixels size. The file is then given a label by annotators in order to be recognized by CNN in the next step. The class label contains 26 label (A – Z) and 10 label (0 – 9) and 1 label whitespace. Total 10.000 character and 10.000 labels { x_i, y_i } is obtained from this process.

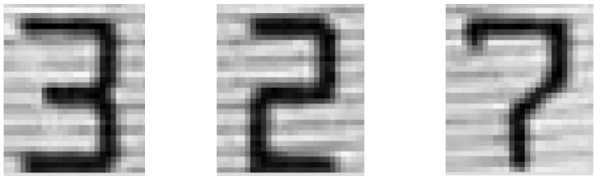

Fig. 3. Image File Containing Number and Character. Left image is annotated with 3, center image is annotated with 2, and the last image is annotated with 7

After obtaining character for each identity card, the data is separated into training data and validation data with weight of separation 70% for training data and remaining fro testing data. Training data then forwarded into Convolutional Neural Networks (CNN) architecture (figure. 4) to produce model. The training process using CNN starting from each digits or alphanumeric characters put in input layer, then Convolutional process with 5 x 5 kernels scan through image overlapping from left to right. The result of this process $S(x,W)$ then put into sub sampling layer which scan by kernels 2 x 2 non-overlapping through image. The result of sub sampling or max pooling $Z(x,W)$ then put into fully connected layer which contains Softmax layer the classified into class label prediction.

$$S(x,W) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij} W_{(i-m,j-n)} \quad (2)$$

$$Z(S,U) = \sum_{i=1}^{n}\sum_{j=1}^{m} S_{ij} U_{(i-m,j-n)} \quad (3)$$

$$Softmax(z_i) = \frac{\exp(Z_i)}{\sum_{i=1}^{n}\exp(Z_j)} \quad (4)$$

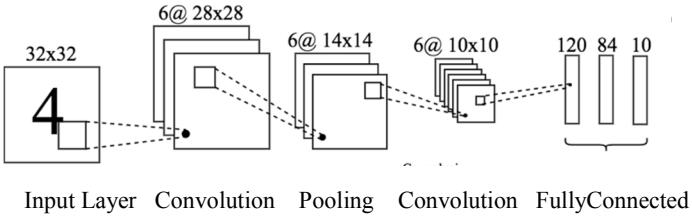

Fig. 4. Convolutional Neural Networks (CNN) architecture which consists of 2 convolutional layers, 2 sub sampling layers and fully connected layer.

The process of training is repetitive until convergence or none of the weight or kernel change. After having appropriate model, measurement evaluation then performed using Precision (Eq. 5), Recall (Eq. 6), and F-Score (Eq.7). Precision score indicated that item retrieved relevant to query, meanwhile Recall score indicated that item relevant retrieved and F-score is the average of Precision and Recall. After measuring a model, the model can be used to predict unknown or existing id card. The prediction result then combining each other to form the final field recognition of Indonesian citizen id card.

$$Precision = \frac{|\text{Relevant} \cap \text{Retrieved}|}{Relevant} \quad (5)$$

$$Recall = \frac{|\text{Relevant} \cap \text{Retrieved}|}{Retrieved} \quad (6)$$

$$F - Score = 2\,\frac{Precision.Recall}{Precision+Recall} \quad (7)$$

TABLE II. DESIGN EXPERIMENTS

Model	***Feature***	***Pre-Processing***
Convolutional Neural Networks	10.000 Alpha-numeric image	Yes
Convolutional Neural Networks	10.000 Alpha-numeric image	No
Support Vector Machine	Sobel, Gabor Filter	Yes
Support Vector Machine	Sobel, Gabor Filter	No

We used 10.000 alphanumeric character extracted from Indonesian citizen id card with 37 labels (A – Z, 0 – 9, and whitespace). The experiment environment is Python with various libraries like Numpy, Tensorflow, and others. The experiment is tried in GPU computer with CUDA support. Research experiment of this research can be seen in table 2. There are four experiments which involving CNN and SVM models as benchmarking. For CNN model, we used 10.000 alpha numeric image and pre-processing techniques. For benchmarking SVM model, we used 10.000 image extracted using Sobel and Gabor filter features.

IV. Result and Discussion

We created 10.000 alphanumeric character dataset derived from Indonesian citizen id card. The dataset is encoded in HDF5 file. The dataset was preprocessed and ready to be used in machine learning or deep learning model. After performing 100 epoch training, we obtained 91% and 90% for training accuracy and validation accuracy using CNN with pre-processing (Figure 5). On the other side we also obtain 64% and 62% for training accuracy and validation accuracy using SVM with pre-processing. Without pre-processing the result of evaluation is below than with pre-processing. The beneficial of Pre-processing to citizen id card image is can remove noise, filter merely appropriate dataset and other.

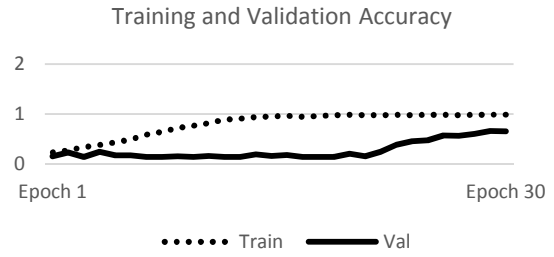

Fig. 5. Training and Validation Accuracy

For measurmenet evaluation using Precsion, Recall and F-score, the complete result can be seen in table 3. CNN result is better than SVM because it consists automatically feature extraction from pre-processed images, reversely in SVM we used manually feature extraction Gabor Feature and Sobel.

TABLE III. Measurement Evaluation

Model	*Precision*	*Recall*	*F-Score*	*Support*
Convolutional Neural Networks	0.84	0.85	0.84	1000
Convolutional Neural Networks (Pre-Processing)	0.89	0.88	0.88	1000
Support Vector Machine	0.62	0.64	0.63	1000
Support Vector Machine (Pre-Processing)	0.71	0.74	0.73	1000

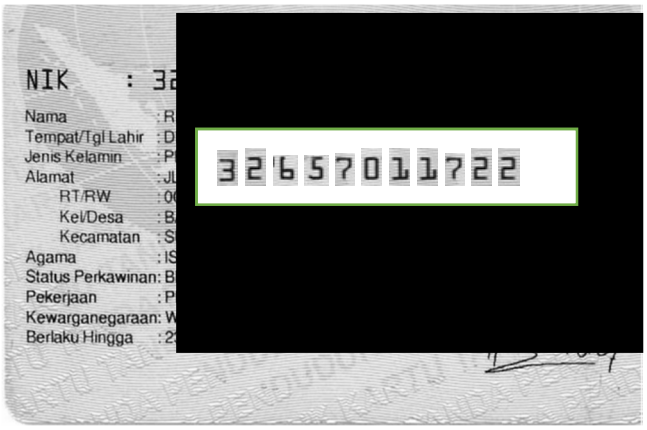

Fig. 6. Recognition result. The field in id card number is recognize one by one using CNN. After obtaining character prediction, then the result is combined to form field result.

We tried to predict complete citizen id card number (16 digits) using CNN. After trying several sampling of Indonesian citizen id card, from a hundred complete id card, the number of fully correct prediction is 76 meanwhile 24 id card incorrect predicted. The factors affects to incorrectly prediction are: (1) citizen id card is not readable (noise, vanish, or scratch), and (2) citizen id card image is cut off.

TABLE IV. Prediction for Each Id Card

Ground Truth	Prediction	Success
3204090311910003	3204090311910003	Yes
3204092505600001	3204092505600001	Yes
3204091812630002	3204091812630002	Yes
3216037009810006	3216037009810006	Yes

V. Conclusion

We created dataset of alphanumeric character derived from Indonesian citizen id card which consists of 10.000 images. The dataset was pre-processed using several image processing techniques: Grayscale, binary, morphological transform. the dataset then separated into training, validation and testing which training used for forming a model, validation to validate each epoch training and testing data is used for measurement evaluation. Each result of classification then combining form id card field like name, number or address. After peforming 100 epoch training, we obtained 91% testing accuracy. For measurement evaluation, we obtained 0.89, 0.88, and 0.88 using precision, recall and F-score. As benchmarking, we used other model: Support vector machine (SVM). The performance of CNN is better than SVM.

For the future experiments, we plan to use other deep learning model like Recurrent Neural Networks (RNN), Residual Neural Networks (ResNet) [4] or other models. We also to plan augment more and more dataset in order to improve the performance.

References

[1] R. Mithe, S. Indalkar, and N. Divekar, "Optical Character Recognition," *Int. J. Recent Technol. Eng.*, vol. 2, no. 1, pp. 72–75, 2013.

[2] S. Lai, L. Xu, K. Liu, and J. Zhao, "Recurrent Convolutional Neural Networks for Text Classification," *Twenty-Ninth AAAI Conf. Artif. Intell.*, pp. 2267–2273, 2015.

[3] M. O. Pratama and M. Adriani, "Music Era Classification using Hierarchical-level Fusion," in *ICACSIS 2018*, 2018.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Dec. 2015.

[5] A. Krizhevsky, I. Sutskever, and H. Geoffrey E., "ImageNet Classification with Deep Convolutional Neural Networks," *Adv. Neural Inf. Process. Syst. 25*, pp. 1–9, 2012.

[6] S. Widodo, "Optical Character Recognition for Indonesian Electronic Id-Card Image," pp. 225–232, 2014.

[7] M. Ryan and N. Hanafiah, "An Examination of

Character Recognition on ID card using Template Matching Approach," in *Procedia Computer Science*, 2015, vol. 59, pp. 520–529.

[8] V. J. Dongre and V. H. Mankar, "DEVNAGARI DOCUMENT SEGMENTATION USING HISTOGRAM APPROACH," *Int. J. Comput. Sci. Eng. Inf. Technol.*, vol. 1, no. 3, 2011.

[9] K. Sreedhar and B. Panlal, "ENHANCEMENT OF IMAGES USING MORPHOLOGICAL TRANSFORMATIONS," *Int. J. Comput. Sci. Inf. Technol.*, vol. 4, no. 1, 2012.