Evaluating Ancient Sundanese Glyph Recognition Using Convolutional Neural Network

E Paulus, S Hadi, M Suryani, I Suryana, YD Simanjuntak
Department of Computer Science, University of Padjadjaran, Jalan Raya Bandung Sumedang
KM21 Jatinangor 45363, INDONESIA
E-mail: erick.paulus@unpad.ac.id, mira.suryani@unpad.ac.id, setiawan.hadi@unpad.ac.id, inosuryana@unpad.ac.id, simanjuntak.y@gmail.com

Abstract. Handwriting recognition is still a real challenge in classification tasks. Not as in modern documents, the isolated glyph images in ancient documents have various random noises, non uniform background color, and smudge. Convolutional neural network (CNN) is one of successful method in pattern recognition and machine learning to classify the objects. The evaluation of some CNN architectures with several different convolution layers classifying the isolated glyph image are presented in this paper. The experimental study is tested on 60 classes of glyph from the ancient Sundanese dataset that published in ICDAR 2017. Beside, the batch normalization is also investigated to measure the performance of the learning process. The results shown that the recognition rate was affected by multi convolutional layers, multi fully connected layer and batch normalization. Based on the experimental study, model 8C2F could achieve 86.15% of recognition rate.

1. Introduction
In traditional document image analysis (DIA) system, recognition task usually implemented after segmentation process. The isolated glyph images resulted from segmentation process can be used to recognize the text and retrieve the meaningful information from the images of documents. Thus, the complete and good quality of glyph images can ease the content based image retrieval. But for ancient documents that written on palm leaf, there are remaining many challenges such as smudge, fraction, shadow, and touching component[1]. So that, some character recognition methods have not achieved optimal performance[2].

In the past, handcrafted feature extraction method become a good mainstay in the segmentation-free and segmentation-based techniques. For instance, word spotting as one of the trending research topics in segmentation-free recognition can be solved by the combination of template matching and blocked-based feature extraction[3] and the combination of feature extraction and SVM classifier [4]. Not only in the segmentation-free recognition process, but handcrafted feature extraction method also plays an important role in segmentation-based recognition. Our previous works showed that using k-Nearest Neighbor with the combination of several feature extraction techniques was able to increase recognition rate up to 72.91% [2]. In other words, the selection of feature extraction and classification method can improve the recognition accuracy [5].

On the other hand, the rapid evolution of neural network research has addressed huge impact in machine learning. Convolutional neural network (CNN) is one of the type of neural network that commonly use for the image. CNN has an advantage in classifying objects well though objects are in different positions, because CNN can facilitate spatial invariance in recognition. For MNIST dataset [6], CNN can led to the recognition rate to 99.28% [7]. The success story of CNN in handwriting digit
recognition (MNIST) has elevated many researchers to classify other scripts such as Javanese [8], Chinese [9], Bangla [10], and Khmer[2]. Actually, they proposed deep CNN architectures with the combination of multiple convolution layer, pooling, dropout and fully connected layer. In fact, there are no certain CNN architecture that recognize well for all scripts. Because every scripts have different challenges depend on their characteristics.

However, the interesting thing in CNN that it applied the similarity approach based on the inner product concept in multiple convolutional layer. If two vector x and y are in the similar direction, their inner product becomes larger. So that, the largest similarity offers the nearest neighbor. The motivation of this research is to evaluate the influence of multiple convolution layer against the performance of recognition rate. We proposed two CNN model that embedded 5 and 7 convolutional layers. Then, the results of both models will compare to our previous work that using 3 convolutional layer. Furthermore, we consider the prevention of overfitting by applying dropout and batch normalization.

In this paper, the evaluating of multiple convolutional layer for isolated glyph recognition are explained in detail. This article is systematically divided into sections. The collection of ancient Sundanese glyphs and their characteristics is described in the second section. Then, the forth section present the result of experiments of both CNN model are described in the fourth section. At the end, the conclusions and next research topic are conveyed in the last section.

2. Sundanese Dataset

Indonesia is rich with many cultural and historical heritage. Sundanese, which is the second largest tribe in Indonesia, has many historical documents. Those documents are stored in libraries or private collections. Preservation and conservation needs to be strived to keep the sustainability of local wisdom and national identity for the next generation. Therefore, digitization and the recognition of ancient script should be held for preservation and conservation.

Based on historical evidence, ancient Sundanese scripts can be grouped according to the variety of writing materials used (such as metals, stones, palm leaves, paper, and bamboo). The period of using that script was long enough, which was about 400 years since XV century. In this research, all of scripts are derived from palm leaf manuscripts (lontar). Almost all of West Java used those scripts at that time. In general, the symbols of the ancient Sundanese script can be arranged into groups of swara scripts, ngalagena scripts, special scripts, rarangkén scripts, and spouses’ scripts[11], as shown on Figure 1.

The isolated glyph images are resulted from annotating by philologist and some students from Informatics Department and Sundanese Study using Alethea[12] tools. Based on 66 manuscript pages, we obtained 60 classes of the glyph and the number of isolated glyph images are 7371 in total[13]. This dataset is consist of 4555 images for training set and 2816 images of testing set.

Figure 2 describes the construction of Sundanese syllables from the combination some scripts. It is not like Latin scripts that written parallel from left to right. In Sundanese writing, some rarangkén scripts are positioned either on the above, below, right, or left the main consonant to change the vocalization. The main consonant is used as the base location.
3. Methodology

3.1. Similarity-based classification

The similarity-based classification that applied in CNN architecture have a same concept with inner product. If two vectors are in the similar direction, their inner product becomes larger. Then, the nearest neighbor is determined by the largest similarity. Figure 3a shows that the glyph “pa” will be measured the similarity level to others glyphs. By doing the inner product between fully connected layers, every glyph will gather together with other glyphs if the similarity level is largest.

Figure 3b presents the characteristics of inner product. An inner product is a generalization form of the dot product. In a vector space, the result of inner product is a scalar. An inner product of two vectors is usually expressed in ordered pairs. In algebra, multiplication of vector scalar is calculated by multiplying all the elements in the vector with their pairs.

Let $a$ and $b$ be vectors and $\theta$ is the angle between the vectors then:

1. Multiplication of two-vector scalar in two dimensional spaces

\[
\langle \vec{a} \cdot \vec{b} \rangle = \langle x_{\{1\}} \cdot y_{\{2\}} + y_{\{1\}} \cdot x_{\{2\}} \rangle
\]

2. Multiplication of two-vector scalar in three dimensional spaces

\[
\langle \vec{a} \cdot \vec{b} \rangle = \langle x_{\{1\}} \cdot z_{\{2\}} + y_{\{1\}} \cdot y_{\{2\}} + z_{\{1\}} \cdot x_{\{2\}} \rangle
\]
3.2. The proposed convolutional neural network model

Convolutional Neural Network is proven able to classify objects well. This experimental study evaluated seven CNN architecture models consist of two, four, six and eight convolutional layers and two, three, and four fully connected layers. The input image is converted into grayscale and reshape into 30x30 pixels in size. All convolution layers implemented kernel size $k=3*3$, stride 2 pixel, using same padding and batch normalization. Pooling is used to reduce feature map results from convolution. The pooling type used in every two convolutional layer is average pooling with filter size 2 * 2, stride 2 pixel, and using same padding. But the result is directly used as input on the fully connected layer with 1024 neurons and followed by the Softmax activation as the last output layer composed of 60 neurons that correspond to all glyph classes. The ReLu activation function and dropout 50% is also applied after convolution two, four, five six, and eight. Figure 4 and figure 5 shown the parameters of training model in detail. The training process is iterated for 100 epochs and optimized by Adam Optimizer.

Figure 3. The illustration of similarity-based classification (left) and the properties of inner product (right)

Figure 4. The four models of CNN Architecture with different the number of convolution layers
3.3. Batch Normalization and Image Generator

Training in deep neural network is complicated. The distribution of each input on the layer changes during training. Because the parameters on the previous layer also change. It causes the training time become longer, because it requires a lower learning rate, complicated parameter initialization, and makes the model difficult to train under certain conditions. Batch Normalization allows training using a much higher learning rate and simpler parameter initialization[14].

4. Result and Discussion

The training process of seven models have been experimented on 4555 isolated glyph images. First, the model is built without implementing the batch normalization. The comparison study which illustrated in table 1 indicated that the accuracy value of training step could achieve above 98 % for two or more convolutional layers. Meanwhile, it is resulted under 95.08% of accuracy for one convolution layer. After that, the highest accuracy value for every model is used to evaluated 2816 images that different from training set. The testing process showed that the highest accuracy value was achieved by Model 8C2F about 81.50% and the lowest is obtained by Model 2C2F. It means that the multi convolutional layers was able to provide better feature map in which it could increase the performance of glyph recognition. In addition, the more number of fully connected layers also gave positive contribution to the accuracy value. It was represented by Model 1C4F that reached 77.38% as highest result. It meant that the similarity-based classification could be represented by fully connected layer or multi perceptron.

As mentioned in the previous literature, if two vectors are in the similar direction, their inner product becomes larger. So that, every glyph will gather together with other glyphs if the similarity level is largest.

The second experiment are to evaluate the performance of batch normalization on complex layers. Model 4C2F, 6C2F, and 8C2F was successful to gain 80.33%, 86.04% and 86.15% for testing set, respectively. The CNN architecture model that implementing the batch normalization was proven to get better result. For example, The accuracy value of Model 6C2F acquired 86.045% which embedded batch normalization. On the other hand, it received 79.40% for CNN architecture with no batch normalization.

Table 1. The comparison of accuracy values on various multilayer CNN architectures

| Model     | Training without BN | Training with BN | Testing without BN | Testing with BN |
|-----------|---------------------|-----------------|--------------------|-----------------|
| Model 8C2F | 98.79               | 98.62           | 81.50              | 86.15           |
| Model 6C2F | 99.12               | 98.16           | 79.40              | 86.04           |
| Model 4C2F | 99.36               | 98.99           | 81.39              | 80.33           |
| Model 2C2F | 99.99               | 97.68           | -                  | -               |
| Model 1C2F | 95.21               | 99.14           | 75.42              | -               |
| Model 1C3F | 90.98               | -               | 76.28              | -               |
| Model 1C4F | 93.98               | -               | 77.38              | -               |
| Our previous model [2] | - | - | 79.50 | - |
In the current study, the experimental simulation have indicated that our proposed CNN architecture on Model 8C2F could improve the recognition rate become 86.15%. This result was better than our previous model that achieved about 79.50%.

5. Conclusion and Suggestion
In this paper, we have evaluated two set of multilayer CNN architectures that run in Tensorflow Framework. Our experimental study show that our various CNN architecture model can recognize 60 classes of glyph and got better result rather than the previous model. The more the convolutional layers indicated the better the glyph recognition results. In addition, the performance of multi convolution layer is affected by the batch normalization to obtain better parameter initialization and learning rate.

Acknowledgments
This work is funded from the Internal Fundamental Research Grant, Universitas Padjadjaran. Beside, the early studies are supported by the International Research Collaboration and Scientific Publication Program implemented by the Ministry of Research, Technology and Higher Education Indonesia. We also thank to philologist of Sundanese Culture Study in Universitas Padjadjaran for annotation process.

References
[1] Paulus E, Suryani M, Hadi S. Improved Line Segmentation Framework for Sundanese Old Manuscripts. In: the 2nd International Conference on Computing and Applied Informatics. 2017.
[2] Kesiman MWA, Valy D, Burie J-C, et al. Benchmarking of Document Image Analysis Tasks for Palm Leaf Manuscripts from Southeast Asia. J Imaging; 4.
[3] Gatos B, Pratikakis I. Segmentation-free Word Spotting in Historical Printed Documents. 2009 10th Int Conf Doc Anal Recognit 2009; 271–275.
[4] Almazán J, Gordo A, Fornés A, et al. Segmentation-free word spotting with exemplar SVMs. Pattern Recognit 2013; 47: 3967–3978.
[5] Hossain MZ, Amin MA, Yan H. Rapid Feature Extraction for Optical Character Recognition. 2012; 1–5.
[6] Lecun Y, Cortes C. The MNIST database of handwritten digits.
[7] Meiyin Wu, Li Chen. Image recognition based on deep learning. In: 2015 Chinese Automation Congress (CAC), pp. 542–546.
[8] Wibowo GH, Sigit R, Barakah A. Javanese Character Feature Extraction Based on Shape Energy. Emit Int J Eng Technol 2017; 5: 154–170.
[9] Zeng X, Xiang D, Peng L, et al. Local Discriminant Training and Global Optimization for Convolutional Neural Network based Handwritten Chinese Character Recognition. In: The 14th IAPR International Conference on Document Analysis and Recognition Local. 2017. Epub ahead of print 2017. DOI: 10.1109/ICDAR.2017.70.
[10] Roy S, Das N, Kundu M, et al. Handwritten Isolated Bangla Compound Character Recognition: a new benchmark using a novel deep learning approach. Pattern Recognit Lett; abs/1802.0, http://arxiv.org/abs/1802.00671 (2018).
[11] Baidillah I, Darsa UA, Abdurahman O, et al. Direktorat Aksara Sunda untuk Unicode. 1st ed. Bandung: Dinas Pendidikan Provinsi Jawa Barat, 2008.
[12] Clauser C, Pletschacher S, Antonacopoulos A. Aletheia - An Advanced Document Layout and Text Ground-Truthing System for Production Environments. In: 2011 International Conference on Document Analysis and Recognition. 2011, pp. 48–52.
[13] Suryani M, Paulus E, Hadi S, et al. The Handwritten Sundanese Palm Leaf Manuscript Dataset From 15th Century. In: The 14th IAPR International Conference on Document Analysis and Recognition ICDAR 2017. 2017. Epub ahead of print 2017. DOI: 10.1109/ICDAR.2017.135.
[14] Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In: the 32nd International Conference on Machine Learning. Lille, France. Epub ahead of print 2015. DOI: 10.1007/s13398-014-0173-7.2.