International Journal of Informatics and Computation (IJICOM)
Vol. 3, No.2, December 2021
ISSN: 2685-8711, E-ISSN: 2714-5263 DOI: 10.35842/ijicom

Pooling Comparison in CNN Architecture for Javanese Script Classification

Mohammad Diqi[1], R. Nurhadi Wijaya[2], Mujastia Feliati M[3].

Abstract

Javanese script is evidence of the past culture, which contains various current language learning, including script recognition. However, learning traditional scripts becomes less attractive to the students. Thus, we propose a learning method to enable character recognition among students to deal with the issues. We offer a novel CNN architecture and compare different pooling layers for Javanese script classification. We calculate the separate pooling layer to reduce extensive feature extraction of the image. We present the model comparison results in Javanese character classification to convince our development.

Keywords

Keywords**:** Classification, CNN, Javanese Script, Deep Learning

1. Introduction.

The Sonobudoyo Museum is one of the state museums in Indonesia where stores historical objects like the Javanese script. This script presents street names, especially Yogyakarta, to represent local elementary, middle, and high schools. To read and understand Javanese script, people must learn it within a specified period, whereas Latin letters are easier and faster to understand. Javanese script is one of the historical relics and is clear proof of an earlier era before the existence of the Indonesian nation. Javanese characters are also an icon of Yogyakarta included in Javanese language learning formulated in essential competencies such as fairy tales, song, and Javanese script [1].

Nowadays, learning Javanese script is one of the essential competencies of students. However, it is less attractive because students have difficulty reading and writing Javanese characters. As well as learning that has not actively involved students, especially in fun learning, it is one factor that influences students' interest in learning Javanese script. Thus, various techniques have been proposed to attract students to learn Javanese script [2]. To enable Character classification, several papers presented teaching methods to deal with Javanese script classification [3].

Deep learning is a growing research area to address pattern recognition and image classification. Deep learning can be considered as a subset of machine learning. It is a field based on learning and improving by examining computer algorithms. While machine learning uses more straightforward concepts, deep learning works with artificial neural networks designed to imitate humans' thinking and learning [4]. CNN method is an effective

1 Mohammad Diqi, Universitas Respati Yogyakarta (diqi@respati.ac.id)
2 R. Nurhadi Wijaya, Universitas Respati Yogyakarta (nurhadi@respati.ac.id)
3 Mujastia Feliati Muhdalifah, Universitas Respati Yogyakarta, mujastia@gmail.com

deep learning architecture to deal with various computer vision issues in many areas by calculating image datasets. The CNN utilizes pixel neighbor information in the feature extraction process with convolution and pooling operation between inputs and kernel. The data is then classified using Softmax to determine its class based on its features. The experimental results show that the discriminative model of deep learning has confirmed recognizing 20 essential Javanese characters with an accuracy of 94.57 % [5].

This paper presents novel deep learning with CNN architecture to construct a Javanese script classification model. We propose a novel technique to deal with character recognition using CNN as a new model to enable character recognition based on the review paper.

2. Related Works

A study proposed a CNN architecture to deal with image processing issues. The testing phase of the study uses comparison parameters such as filter size, amount of data, and data training scenario to get the best model. The filter size in this comparison is 3x3 and 5x5, for the amount of data compared to 90, 150, 300, and the data training scenario in question is the sharing of train and test data starting from 60%:40%, 70%:30: and 80%:20%. The comparison produced a model with an accuracy rate of 95% [6]. A CNN also proposed to establish a classification model with 2000 pieces. This test divides the dataset into ten folders containing 2000 images. Based on the result, the model can identify the type of plant genus with accuracy = 90.8% [7].

An article explored average pooling as a pooling layer and a convolutional neural network algorithm. The data will be divided into left-eye data, correct eye data, second eye data. Then the results of the three CNN will be averaged with the highest value taken as a result of classification. System testing uses stages, first stage classification with three classes, and second stage classification with nine types. Based on the research that has been done, it can be concluded that the test results using three categories are better than using nine classes. The 3-class classification pays only attention to the angle horizontally and ignores the vertical angle so that the neural network can predict easily [8].

In classifying moon orchids images, a paper proposed CNN to deal with dendrobium orchids images and squirrel tail orchids mages classification. The study gathered 90 ideas for data training and 30 shots for data testing. They applied CNN with feedforward and backpropagation methods. The test results of the classification of image objects on the image of white moon orchid plants, dendrobium, and squirrel tails were able to achieve an accuracy = 83%, recall = 80%, and precision = 89% [9].

An article about research to combine domain knowledge in machine learning for football result prediction. The 2017 Football Prediction Challenge task is to use machine learning to predict the outcome of future football matches based on a data set explaining the results of 216,743 past football matches. The research presented two new ideas for integrating football domain knowledge into the modeling process, a method for predicting match results, which we describe as the extraction of reviewer features and the learning of ranking features. Using this method, two learning sets will be built from challenge data [10]. Another CNN implementation introduces objects of footballs, birds, and basketballs with a dataset of up to 450 images. Based on the experimental result, the model can produce training accuracy = 98% and testing accuracy = 81.3% [11].

A study discusses hand gesture imagery taken using the camera will then be done hand signal recognition and, in the process, using a single-board computer to be recognized. The introduction results were forwarded to Leonardo Arduino and DC motors to move twelve-wheeled robot movements. The methods used in this study are contrasted stretching of preprocessing and CNN for hand signal recognition. This method is tested with a slight variation of 26-140 lux. The distance from hand to the camera is 120-200 cm. Hand signal recognition system using the method produces accuracy of 97.5%, precision

97.57%, sensitivity 97.5%, specificity 99.77% and f1 score of 97.45% [12].

An article discussed the comparative architecture of CNN for the classification of fundus imagery. The test results in both methods showed the best architectures, namely VGG19 and VGG16. The first phase of trials resulted in sensitivity, specificity, and accuracy of 87.8%, 90.7%, and 89.3%. For the second stage of the sensitivity trial, specificity and accuracy are 94.2%, 90.4%, and 92.31% [13]. Another implementation of CNN is infant facial expression videos with the CNN, Autoencoder, and LSTM Network to detect crying and pain levels in a baby's facial videos [11].

One way to classify characters is by examining how they change throughout a story. Grouped in this way by character development, character types include the dynamic character, the round character, the static character, the stock character, and the symbolic character. The script is primarily used to write the Javanese language and several other regional languages. The current paper explored Javanese script classification by collecting various character datasets. The report constructed the Javanese script classification dataset to help users detect Javanese characters. The results of this training the application of Javanese script classification can produce a certain level of recognition of Javanese script patterns in an actual application [17]

3. Our Method

CNN is a development of artificial neural networks inspired by human neural networks and is commonly used in image data to detect and recognize an object [14]. It has several steps to compute and classify the image. In this study, we focus on pooling layer comparison, which reduces the size of the matrix used to reduce the dimensions of the Feature Map (downsampling). In principle, the Pooling layer consists of a filter with size and stride sure that will shift to the entire area feature map. We compare the classification model's average and max pooling performance to get the best model. CNN is one of the developments of artificial neural networks inspired by human neural networks and is commonly used in image data to detect and recognize an object in an image [14].

In the CNN architecture, the input layer is the image that will be convoluted in the form of a Carakan script image. The second layer is feature extraction, which contains the convolution, pooling, and activation functions. Convolutional Layer or the part that performs convolution operation that combines linear filters against local areas. This layer is the first to receive an image inputted into the architecture. The shape of this layer is a filter with a length (pixel), width (pixel), and thickness according to the channel image data input. Fig. 1 depicts CNN filter with a 5x5 filter.

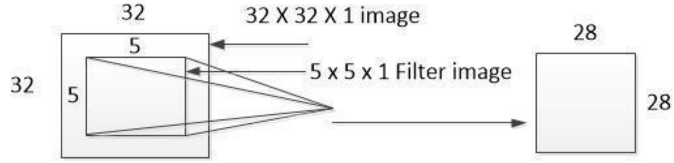


Fig. 1 Illustration of a Convolution process of CNN

Figure 1 shows a 32x32 size image inserted into the convolution layer with a filter 5x5. The next layer is the pooling layer. It reduces the matrix size used to reduce the dimensions of the Feature Map (downsampling), thereby accelerating computing because the parameters that have to be updated are fewer and overcome overfitting. The usual pooling layers are Max Pooling and Average Pooling. Max pooling is done by selecting the maximum value in a specific area in Fig. 2

Fig. 2 Max Pooling

While average pooling is done by selecting the average value of a particular area. Can be seen in Fig. 3

Fig. 3 Average Pooling

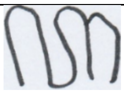
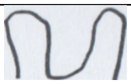
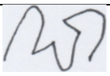
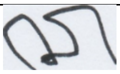
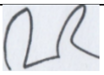
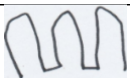
ReLu is an activation function that will produce zero value if x < 0 and then linear with slope 1 when x > 0 [14] [8]. The third layer is Classification has two layers, namely Flatten and Fully Connected layer. Flatten is reshaping the feature map into a vector to be input from a fully-connected layer.

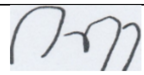
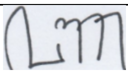
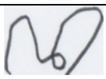
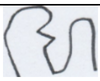
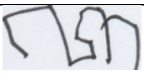
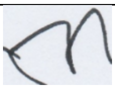
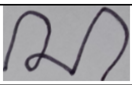
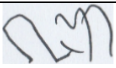
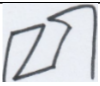
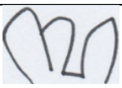
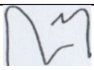
While the Fully connected layer is a collection of convolution processes, this layer gets input from previous approaches to determine which features are most correlated with a particular class. The function of this layer is to unite all nodes into one dimension [15]. The last layer is the loss function to determine how the training provides a penalty for storage between the predicted results and the label [16].

4. Experimental Setup

A fundamental letter in the Javanese script is called an aksara which represents a syllable that contains around 45 letters. However, some notes became obsolete and some are only used in certain contexts. It is common to divide the letters into several groups based on their function. In this experiment, we utilized 2000 images of Carakan script as our dataset that consisting of 20 letters namely Ha, Na, Ca, Ra, Ka, Da, Ta, Sa, Wa, La, Pa, Dha, Ja, Ya, Nya, Ma, Ga, Ba, Tha, and Nga. Table 1 describes our dataset to conduct our experiment.

Table 1 Dataset in this study

Number	Letter Name	Pictures	Number	Letter Name	Pictures
1	Ha		11	Pa	
2	Na		12	Dha	
3	Ca		13	Ja	
4	Ra		14	Ya	

5	Ka		15	Nya	
6	Da		16	Ma	
7	Ta		17	Ga	
8	Sa		18	Ba	
9	Wa		19	Tha	
10	La		20	Nga	

The design of the model using a flowchart can be seen in figure 4

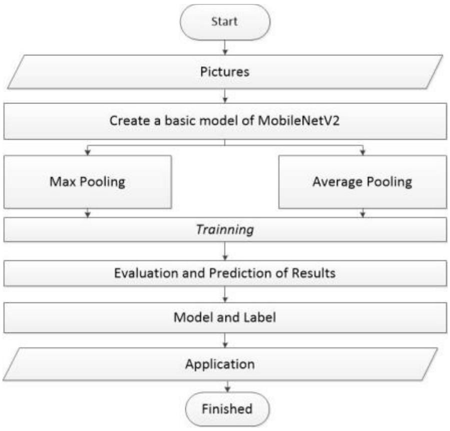


Fig. 4 Model development of Javanese character recognition

Fig. 4 depicts our CNN architecture consisting of several layers, including pooling, average pooling, and max pooling. The first step is to collect image data, then create a basic model with average pooling and max pooling. Each model will be trained to produce an evaluation and prediction value. The results of the two models will be compared to find the best model. This paper constructs a learning model with average pooling and max pooling. The model will be divided into four parts: the average model pooling with a 5x5 and filter 3x3, and a max-pooling with a 5x5 and filter 3x3.

5. Results and Discussion

a. Training and Testing Result

This study conducts the training process to construct our classification model using the training dataset. Then, we perform the validation process to evaluate the model performance. This study trained until 1600 dataset and tested \ 400 Java character as the testing dataset. Table 2 describes the training and validation results in the CNN architecture.

Table 2 Training Results

Pooling Layer	Epoch	Filter	Accuracy	Accuracy Validation	Loss	Loss Validation
Average Pooling	100	5x5	1.0000	0.9250	0.0025	0.3779
		3x3	1.0000	0.9125	0.0025	0.4344
Max Pooling	100	5x5	1.0000	0.9075	0.0026	0.4351
		3x3	1.0000	0.8950	0.0028	0.4974

b. Accuracy Comparison

This experiment also calculates the accuracy metric to measure the algorithm's performance interpretably. The accuracy of a model is usually determined after the model parameters and is calculated in the form of a percentage.

Table 3 Accuracy Comparison

Pooling Layer	Filter	Accuracy
Average Pooling	5x5	93%
	3x3	91%
Max Pooling	5x5	92%
	3x3	90%

Table 3 depicts that the average pooling with filter 5x5 has a higher value than the average pooling filter 3x3. The pooling operation involves sliding a two-dimensional filter over each channel of the feature map and summarizing the features lying within the region covered by the filter.

c. Evaluation Metrics

An evaluation metric quantifies the performance of a predictive model. This typically involves training a model on a dataset, using the model to make predictions on a holdout dataset not used during training, then comparing the projections to the expected values in the holdout dataset. Fig. 5 depicts the evaluation metrics of the proposed model.

Classification Report

	precision	recall	f1-score	support
BA	0.69	0.90	0.78	20
CA	1.00	0.90	0.95	20
DA	0.95	1.00	0.98	20
DHA	1.00	1.00	1.00	20
GA	0.90	0.95	0.93	20
HA	1.00	0.75	0.86	20
JA	1.00	0.95	0.97	20
KA	1.00	0.90	0.95	20
LA	0.82	0.90	0.86	20
MA	1.00	0.80	0.89	20
NA	0.74	1.00	0.85	20
NGA	1.00	0.90	0.95	20
NYA	0.88	0.70	0.78	20
PA	0.91	1.00	0.95	20
RA	1.00	0.90	0.95	20
SA	1.00	1.00	1.00	20
TA	1.00	1.00	1.00	20
THA	1.00	1.00	1.00	20
WA	1.00	1.00	1.00	20
YA	0.87	1.00	0.93	20
accuracy			0.93	400

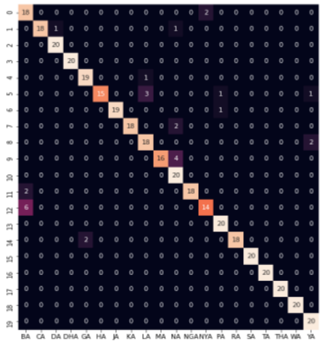


Fig. 5 Classification Matrix and Confusion Matrix Average Pooling Filter 5x5.

The Classification Matrix and Confusion Matrix Max Pooling Filter 5x5 can be seen in Fig. 6

Classification Report

	precision	recall	f1-score	support
BA	0.85	0.85	0.85	20
CA	0.90	0.95	0.93	20
DA	1.00	1.00	1.00	20
DHA	1.00	1.00	1.00	20
GA	0.89	0.80	0.84	20
HA	1.00	0.70	0.82	20
JA	1.00	0.85	0.92	20
KA	0.90	0.95	0.93	20
LA	0.61	0.95	0.75	20
MA	1.00	0.90	0.95	20
NA	0.90	0.95	0.93	20
NGA	1.00	0.90	0.95	20
NYA	0.88	0.70	0.78	20
PA	0.83	1.00	0.91	20
RA	0.95	0.90	0.92	20
SA	1.00	1.00	1.00	20
TA	1.00	1.00	1.00	20
THA	1.00	1.00	1.00	20
WA	0.95	0.95	0.95	20
YA	0.91	1.00	0.95	20
accuracy			0.92	400

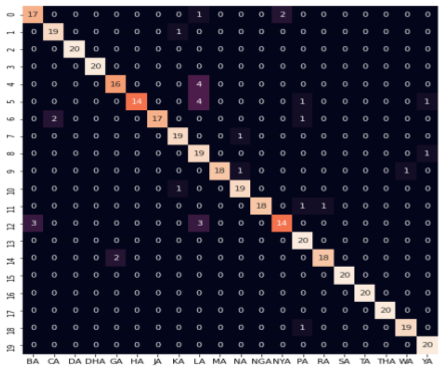


Fig. 6 Classification Matrix and Confusion Matrix Max Pooling Filter 5x5.

Based on the evaluation metrics, the proposed model can obtain a better result using the average pooling layer using the evaluation matrix. The pooling layer can increase the accuracy, precision, and recall value compared to the max-pooling layer.

6. Conclusion

Javanese script is a critical element of cultural heritage that becomes a basic competency in current learning in the students. However, it is less attractive to students to learn the character. Thus, it requires a new method to retain Javanese script classification and attract children's interest in learning Javanese characters. Therefore, we propose a novel learning technique using CNN to construct the classification model of the Javanese character by training Javanese script images. This study conducts a training and testing phase to measure evaluation and prediction results. We calculate and compare the pooling layer performance in the script classification to increase the CNN performance. Based on the experimental result, the average pooling can produce a higher accuracy than the max-pooling layer. Moreover, the average pooling can also gain a faster training time. It can be concluded that the average pooling can obtain the best performance for Javanese script classification

7. References

[1] Rismiyati, Khadijah and A. Nurhadiyatna, "Deep learning for handwriting Java character recognition," 2017 1st International Conference on Informatics and Computational Sciences (ICICoS) IEEE , pp. 59-64, 2017.

[2] G. Abdul Robby., et al, “Implementation of Optical Character Recognition using Tesseract with the Javanese Script Target in Android Application,” Procedia Computer Science, Vol. 157, pp 499-505, ISSN 1877-0509, 2019

[3] D.K. Ratih, et al., "Usability Evaluation of Mobile-Based Application for Javanese Script Learning Media.", Journal of Information Technology and Computer Science, Vol. 3, No. 1, pp. 88-93, 2018.

[4] Hidayatullah, A. F., Cahyaningtyas, S., & Pamungkas, R. D., “Attention-based cnn-bilstm for dialect identification on javanese text.”, Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, pp. 317-324, 2020

[5] M. A. Wibowo, M. Soleh, W. Pradani, A. N. Hidayanto and A. M. Arymurthy, "Handwritten javanese character recognition using descriminative deep learning technique," 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), 2017, pp. 325-330

[6] Jimmy Pujoseno, "Implementation of Deep Learning Using Convolutional Neural Networks for Classification of Stationery", Final Project: Islamic University of Indonesia, 2018.
[7] Sarirotul Ilahiyah and Agung Nilogiri, Implementation o++f Deep Learning in Plant Identification Based on Leaf Image Using Convolutional Neural Network, JUSTINDO, Vol 3 No 2, 2018.
[8] William Sugiarto, Yosi Kristian and Eka Setyaningsih, Estimation of Eye Gaze Using Average Pooling Convolutional Neural Network, Technology Dynamics,, Vol 9 No 2 Page 62-68, 2017
[9] Boki Laptupono, Implementation of Deep Learning Using Convolutional Neural Network for Image Classification (case study: Sports Images of Soccer, bird, and Basketball), Final Task, Yogyakarta: Universitas Islam Indonesia, 2018.
[10] Rahayu Kia Sandi Cahaya Putri, Implementation of Deep Learning Using Convolutional Neural Network Method for Image Klasifikais (Case Study: White Moon Orchid, Dendrobium Orchid and Squirrel Tail Angrek), Final Task, Yogyakarta: Islamic University of Indonesia, 2018.
[11] Kristian, Yosi, et al. Classification of Pain in Infant Facial Expression Videos Using DCNN Autoencoder and LSTM. National Journal of Electrical Engineering and Information Technology (JNTETI), , 7.3: 308-316. 2018
[12] Rika Rokhana, Joko Priambodo, Tita Karlita, Made Gede Sunarya, Eko Mulyanto Yuniarno, I Ketut Eddy Purnama and Mauridhi Hery Purnama, Convolutional Neural Network for Femoral Bone Detection in B-Mode Ultrasonic Image, JNTETI, Vol 8 No1, 2019
[13] Setiawan, Wahyudi. Comparison of Convolutional Neural Network Architecture for Fundus Classification. *Jurnal Simantec*, 7.2: 48-53. 2019.
[14] Royani Darma, and Gunawan Ariyanto, Implementation of Deep Learning Tensorflow-Based for Fingerprint Recognition, Journal of Emitter, Surakarta: Muhammadiyah University of Surakarta (UMS), Vol. 18 No. 01, 2018
[15] Erlyna Nour Arrofiqoh and Harintaka, Implementation of the Convolutional Neural Network Method for Classification of Plants in High-Resolution Images, Geomatics, 2018, Volume 24 No.2, Yogyakarta: Gadjah Mada University.
[16] Suryanto. Kurniawan N.R and Satria M. Deep Learning Modernizing Machine Learning for Big Data. Bandung: Informatics Publisher. 2019.
[17] Diqi, Mohammad; MUHDALIFAH, Mujastia. Design and Building Javanese Script Classification in The State Museum of Sonobudoyo Yogyakarta. International Journal of Informatics and Computation, [S.l.], v. 1, n. 2, p. 35-45. 2020.