Classification of Pandavas Figure in Shadow Puppet Images using Convolutional Neural Networks

Wiwit Supriyanti[1]*, Dimas Aryo Anggoro[2]
[1]Politeknik Indonusa Surakarta
Surakarta, Indonesia
[2]Informatics Department
Universitas Muhammadiyah Surakarta
Surakarta, Indonesia
*wiwitsupriyanti@poltekindonusa.ac.id

Abstract-Indonesia is a nation with various ethnicities and rich cultural backgrounds that span from Sabang to Merauke. One of the cultural products of Indonesian society is shadow puppet. Shadow puppet has been internationally renowned as a masterpiece of cultural art and recognized by UNESCO. The development of Indonesian society is very dependent on technological sophistication and it may shift the existing traditional culture out from the memory of the nation. Practices of modern life and the busy activities of the people exacerbate the condition and may make the society to ignore traditional culture. This study seeks to preserve traditional Indonesian culture by making shadow puppets as the object of classification. We use a deep learning algorithm called convolutional neural network (CNN) to classify 430 puppet images into 4 classes. The proportion of training, validation and test data is 70 by 20 by 10. The experiments show that the most efficient model is obtained with 3 convolution layer. It reaches an accuracy rate of 0.93 and a drop out rate of 0.2

Keywords: shadow puppet, pandavas, image classification, deep learning; convolutional neural network

1. Introduction

The Indonesian nation is a nation that has various ethnic and cultural backgrounds from Sabang to Merauke. One of the cultural products of Indonesian society is shadow puppet. This art developed on the islands of Java and Bali. There are two versions of shadow puppet, namely wayang orang and wayang kulit. Wayang orang is a puppet that is played directly by people using costumes as their trademark, while the puppet in the form of a puppet is a puppet played by the puppeteer. Some of these puppet-shaped puppets include shadow puppets, puppets and grass puppets. Wayang kulit is a shadow puppet performance art made from dried animal skins originating from Central Java and Yogyakarta. The stories in puppet shows usually come from the Mahabharata and Ramayana which have been changed by poets and masters in the archipelago [1].

In the international world, shadow puppet has now been recorded as a masterpiece of cultural art, namely by UNESCO, an institution under the United Nations that deals with issues of education, science and culture. On 7th November 2003 the Indonesian shadow puppet was announced by UNESCO as a world masterpiece in Paris. This demonstrates that shadow puppetry, as a traditional

cultural heritage, has been recognised internationally as a cultural heritage rich in values that contributes significantly to the creation and growth of national identity [2]. Seeing this award, as citizens with integrity, they should protect and preserve this culture. However, along with the times with increasingly sophisticated technological advances, shadow puppet shows are now increasingly being marginalized from the arena of the entertainment scene. The knowledge of this shadow puppet art among teenagers is decreasing, one of which is the ignorance of the puppet characters. This is because many puppet characters have different shapes and characters.

The many types of shadow puppet in Indonesia make researchers interested in creating a program to recognize the type of wayang kulit, especially for Pandavas figure based on a dataset of photos or shadow puppet images. The introduction of the shadow puppet image will later be classified according to the five types of pandavas figure that exist. Seeing the development of an era in which Indonesian society is very dependent on technological sophistication, it is possible if the existing culture begins to be forgotten. Habits of modern life and the daily activities of the people also make it possible for traditional culture to be forgotten. Therefore, the selection of the shadow puppet object in this

study is expected to be able to contribute to the preservation of Indonesian culture. In this case, the data collection of shadow puppet objects will be classified according to their type using a computer program.

Researchers are currently focusing on the development of deep learning, especially the neural network process, since its output is superior to that of other methods. Several studies on convolutional neural network (CNN) have been conducted by previous researchers, with a test error of 17%, the ImageNet LSVRC-2010 dataset was divided into 1000 classes, yielding very important results in testing. [3]. Comparing the convolutional neural network (CNN) method with several other classification methods implemented for animal recognition, the result is that the CNN method gives the best results with an accuracy rate of up to 98% [4]. When the CNN approach is used to introduce road traffic signs, an accuracy rate of about 85 percent to 90 percent is achieved [5]. Based on the exposure of several studies above, it can be concluded that the use of the CNN method in classifying images has better advantages over other classification methods. The researcher chose the convolutional neural network (CNN) algorithm to identify the shadow puppet images for this purpose.

2.   Method

a.   Pandavas Figure

Figure 1. Pandavas Figure

One of the shadow puppet which is well known as shadow puppet culture in Indonesia is pandavas figure. Pandavas figure is one of the puppets originating from the Mahabharata story. There are five characters in the Pandava puppet, namely Yudistira, Bima, Arjuna, Nakula and Sadewa. Each pandavas figure in the shadow puppet, especially the Javanese version of the pandava, has a distinctive and meaningful character, the following are the characters in the Pandava puppet quoted from the internet [6]:

1)   Yudistira

Puntadewa is the small word for Yudistira. He is the oldest of Pandu and Dewi Kunti's five sons, the Pandavas.

He is Lord Yama's embodiment. The Kingdom of Amarta was ruled by Yudistira. Yudistira is a wise man who has no rivals and has rarely lied in his life. Has a high morale, enjoys forgiving and forgiving enemies who have given up. Fairness, tolerance, integrity, religious observance, self-assurance, and the willingness to speculate are some of the other qualities that stand out.

2)   Bima

Sena is Bima's first name. Pandu and Dewi Kunti have a son named Bima. Since he is an embodiment of Dewa Bayu, he is known as Bayusutha. Bima is the most strong of his brothers, with long arms, a tall stature, and the most terrifying face. Despite this, he has a good spirit. When it comes to mace guns, you're a natural. Rujakpala is the name of the club's arms. Werkudara is another nickname for Bima. Bima has three children in Javanese puppetry: Gatotkaca, Antareja, and Antasena. Bima is a brave, steadfast, strong, steadfast, obedient, and honest character. Bima is also cruel and frightening to his rivals, despite the fact that his heart is tender. He has a single-minded attitude, dislikes small talk, is never ambivalent, and never licks his own saliva.

3)   Arjuna

Permadi is Arjuna's first name. He is the youngest son of Dewi Kunti and Pandu. Lord Indra, the god of war, has come to life in him. He is a wise knight who enjoys wandering, meditation, and learning new things. Arjuna was a skilled archer and was regarded as a knight. His military prowess made him the Pandavas' rock in order to win the major fight against the Kauravas. Janaka is another name for Arjuna. In Madukara, he was the ruler. Arjuna is a clever, quiet, gentle, cautious, respectful, courageous warrior who enjoys protecting the helpless.

4)   Nakula

Pinten is Nakula's nickname. Nakula is one of Dewi Madrim and Pandu's twin sons. He is the embodiment of Aswin, the god of medicine, who is one of the twin gods. Nakula is a master of the sword weapon. Nakula is the most attractive man on the planet and a formidable swordsman. Nakula's characters are trustworthy, faithful, and obedient to their parents. They also know how to repay favors and keep secrets.

5)   Sadewa

Tangsen is Sadewa's nickname. Sadewa is one of Dewi Madri and Pandu's twin sons. He is the embodiment of Aswin, the god of medicine, who is one of the twin gods. Sadewa is a hardworking and wise person. Sadewa is an astronomy specialist as well. Sadewa's characters are trustworthy, faithful, and obedient to his parents. They also know how to repay favors and keep secrets.

The puppet dataset used in this study was taken from open sources via the website http://tokohwayangpurwa. blogspot.com/ [7] with a total of 418 images classified into 4 classes: Yudistira, Bima, Arjuna, Nakula Sadewa (second these figures are physically identical twins).

b. Deep Learning

Deep learning has recently become a hot topic in machine learning research. The explanation for this is that deep learning has shown incredible results in the field of computer vision. Deep learning is a branch of machine learning that makes use of neural networks to solve problems involving large datasets. For supervised learning, deep learning techniques have a very powerful architecture. The learning model will better reflect labeled image data by adding more layers.

There are strategies for extracting features from training data and special learning algorithms for classifying images and detecting sounds in machine learning. However, there are several disadvantages to this approach in terms of speed and accuracy. Convolutional neural network (CNN) is a deep learning system that addresses the shortcomings of the previous method. With this model, a number of independent parameters can be decreased, and input image deformations including translation, rotation, and scale can be treated [8].

The application of deep (multi-layered) artificial neural network concepts can be suspended on existing machine learning algorithms so that today's computers can learn at speed, accuracy, and at large scale. Deep learning was widely used in the science community and industry to help solve many big data issues, such as computer vision, speech recognition, and natural language processing, as this theory evolved. One of the key features of deep learning is feature engineering, which extracts useful patterns from data to make it easier for models to distinguish between classes. The most critical technique for achieving good results on predictive tasks is feature engineering. However, different data sets and data types necessitate different engineering approaches, making it difficult to understand and master.

CNN methods or convolutional neural networks are very good at identifying good features in the image to the next layer to shape non-linear hypotheses that can increase the complexity of a model in deep learning, CNN methods or convolutional neural networks are very good at finding good features in the image to the next layer to form non-linear hypotheses that can increase the complexity of a model [9].

c. Convolutional Neural Network

Convolutional neural networks, or ConvNets, are a type of neural network that processes data in the form of multiple arrays, such as a color image made up of three 2D arrays containing pixel intensities in three different colors. Convolutional neural networks (ConvNets) are a subset of artificial neural networks (ANNs) that are widely regarded as the most effective model for solving object recognition problems.

A convolutional neural network, in technical terms, has a trainable architecture that consists of several steps. Each stage's input and output are feature maps, which are a collection of arrays. A two-dimensional matrix, for example, is the input for a greyscale image. Each stage's output is a feature map containing the processing results from all points in the input image. Convolution, activation, and pooling are the three layers that make up each level.

Figure 2 depicts the design of a convolutional neural network in general, as used by LeCun [10]. In this example, CNN's input is a picture of a specific size. The convolutional stage is the first in CNN. Convolution is done with the aid of a kernel of a certain scale. The number of kernels used is determined by the number of features that are generated. This stage's output is then passed through an activation function, such as a tanh function or a Linear Unit Rectifier (ReLU). The output of the activation function then goes through a sampling or pooling operation. Depending on the pooling mask used, the pooling process produces an image that has been reduced in size.
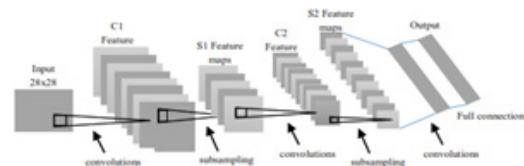

Figure 2. Convolutional Neural Network Architecture

1) Convolutional Layer

Neuron to (i, j) in the hidden layer, has a value of activity y which is calculated according to Equation (1), where the value (m, n) in the equation shows the size of the local receptive fields / kernel.

$$y_{ij} = \sigma(b + (w * x)) \qquad (1)$$

$$(W * x)_{ij} = \sum_{k=0}^{m}\sum_{l=0}^{n} w_{k,l}\, x_{i+j,k+l} \qquad (2)$$

The multiplication between the input and the kernel above (Equation (2)) is usually called a convolution. However, convolution is carried out on an inverted kernel [11], as in Equation (3). Meanwhile, if the kernel is not reversed, the function is called cross-correlation. Even so, many machine learning libraries use the cross-correlation formula and call it the convolutional formula.

$$(W * x)_{ij} = \sum_{k=0}^{m}\sum_{l=0}^{n} w_{k,l}\, x_{i-j,k-l} \qquad (3)$$

The size of the convoluted image is reduced compared to the initial image and can be expressed by Equation (4). In this case, if an image with a size of 28x28 is subjected to convolution with a kernel size of 3x3 then the final size becomes 28-3 + 1 x 28-3 + 1 = 26x26.

$$ukuran_{hasilkonvolusi} = ukuran_{awal} - filtersize + 1 \qquad (4)$$

Figure 3 shows an illustration of the convolution process in the image, which is a two-dimensional array I, with a weight of K (two dimensions) [12]. In this figure, a 4x3 image is convoluted using a 2x2 kernel. The resulting image is 3x2 in size. The first element in the convoluted

image is the sum of the multiplication of the kernel weight and the corresponding image value.
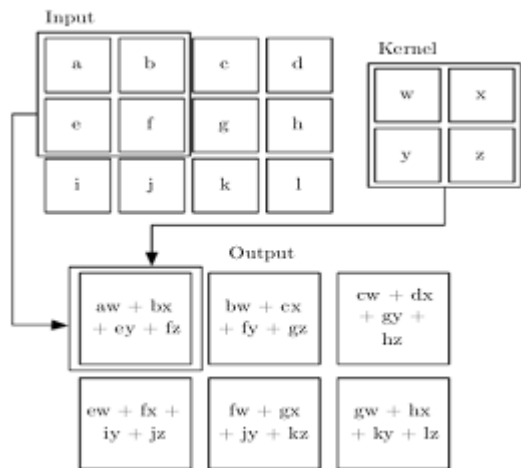

Figure 3. Convolution process on 2D array input with 2D weights

Via performance optimization, the convolutional layer is exposed to the model's complexity in a significant way. Three parameters, width, stride, and zero padding settings, are used to optimize this [13].

2) Pooling Layer

As convolution is performed, the pooling layer keeps the size of the data constant, i.e. by reducing the size of the matrix / sample reduction (downsampling) [14]. After the Convolutional Layer, the Pooling Layer is typically applied. The pooling layer is essentially a filter with a specific size and stride that alternately shifts the entire feature map region. In general, max pooling or average pooling are used in the pooling process. When using max pooling, the largest value is used, while when using average pooling, the average value is used. Of the two ways the pooling process is most often encountered is to use max pooling, for average pooling it is very rarely used but in several network architectures it can be found [15]. To monitor overfitting, a pooling layer inserted between successive convolutional layers in the CNN model architecture will gradually reduce the size of the output volume on the Feature Map, lowering the number of parameters and calculations on the network. Each feature map stack's size is reduced by the pooling layer, which runs on top of it. In most cases, the pooling layer employs a 2x2 filter that operates on each slice of the input in two stages. A max-pooling process is depicted in Figure 4.
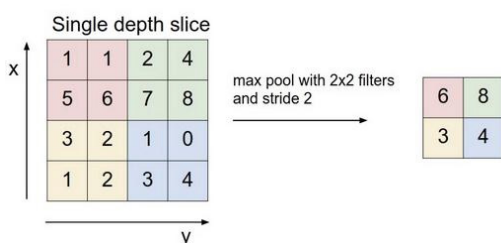

Figure 4. Max-Pooling Operation

The method of max-pooling is depicted in the diagram above. The pooling process produces a matrix with smaller dimensions than the original image. Each slice of the input volume depth will be processed by the pooling layer above. The max-pooling operation uses a 2x2 filter scale, as seen in the image above. The process's input is 4x4 in size; the maximum value is taken from each of the four numbers in the procedure input, and then a new output size of 2x2 is generated.

3) Normalization Layer

Normalization layer is useful for overcoming significant differences in value ranges. However, currently the normalization layer is still not widely used because the effect on this layer is not that big [14].

4) Fully Connected Layer

As with ordinary neural networks, the Fully Connected Layer is a layer in which all activation neurons from the previous layer are connected to all neurons in the next layer. This layer is most commonly used in MLP (Multi Layer Perceptron), which aims to convert data dimensions so that data can be categorized linearly.

The distinction between a fully connected layer and a regular convolutional layer is that the convolutional layer's neurons are only connected to a small portion of the input, while the fully connected layer's neurons are all connected. The two layers, however, still control the product dot, so their functions are similar.
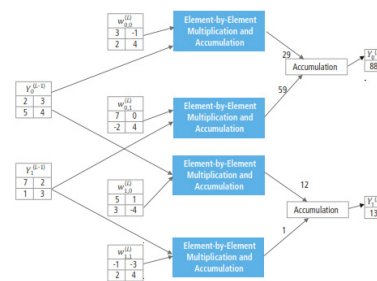

Figure 5. Processing of a Fully Connected Layer

5) The Activation Function

The activation function is a linear or non-linear function that defines the relationship between levels of internal activity (summation function). This role determines whether or not neurons are active. The ReLU (Rectified Linear Unit) activation function is one of the most widely used activation functions in CNN. The ReLU (Rectified Linear Unit) function basically performs a threshold operation from 0 to infinity. A graph of the ReLU activation function is shown below:

If the input from the neurons is a negative number, the function will convert that value to a value of 0, and if the input is positive, the neuron's output will be the activation value itself.

6) Loss Layer

Loss layer is the last layer in CNN where in this process it will show the predicted results and loss function values during the training process.

d. Dropout Regularization

Dropout is a regularization strategy for neural networks that prevents overfitting while also speeding up the learning process [16]. Overfitting is a situation in which almost all of the data that has gone through the training phase has reached a good percentage, but the prediction process has a gap [17].
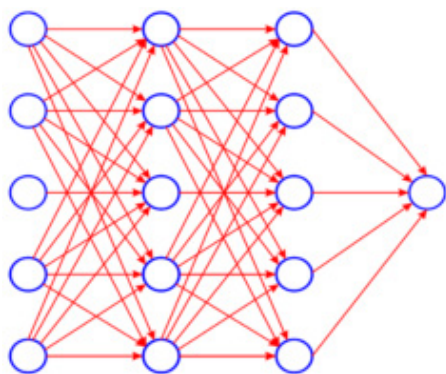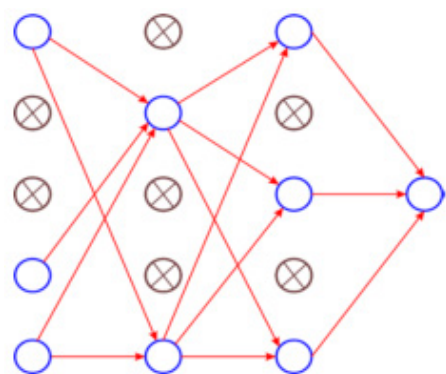

Figure 7. Standard Neural Net

Figure 8. After Applying Dropout

Dropout refers to removing neurons that are either hidden or visible layers in the network. By removing a neuron, means removing it temporarily from the existing network. The neurons to be removed will be randomly selected. Each neuron will be assigned a probability p that is between 0.0 and 1.0 [18]. The dropout temporarily eliminates a neuron from the network in the form of a hidden layer or a visible layer in the working system. Figure 7 (a neural network with two hidden layers) and Figure 8 (a dropout process has been carried out) describes the dropout regularization process.

e. Adam Optimizer

Adam Optimizer is a way to optimize a parameter, this optimization can make the parameter to be a maximum or a minimum. Adam Optimizer is one of the optimizations that combines the AdaGrad and RMSProp methods [19].

f. Confusion Matrix

The accuracy matrix was used to test the model. To comprehend the matrix, true positive (TP), false positive (FP), false negative (FN), and true negative (TN) were previously specified as shown in Table 1's confusion matrix. Predicted positive data is classified as TP. True is characterized as negative data that is predicted to be true, while TN is defined as negative data that is predicted to be true. FN is the polar opposite of TP, which is negative data predicted to be incorrect, and FP is the polar opposite of TN, which is positive data predicted to be incorrect.

Table 1. Confusion Matrix

Predicted Class	True Class	
	Positive	Negative
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

The ratio of all correctly categorized data (both positive and negative) divided by the total number of data is known as accuracy. The formula for accuracy is Equation (5).

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (5)$$

g. Research Flow

The flowchart in Figure 9 shows the procedure for classifying shadow puppet images using a convolutional neural network algorithm.
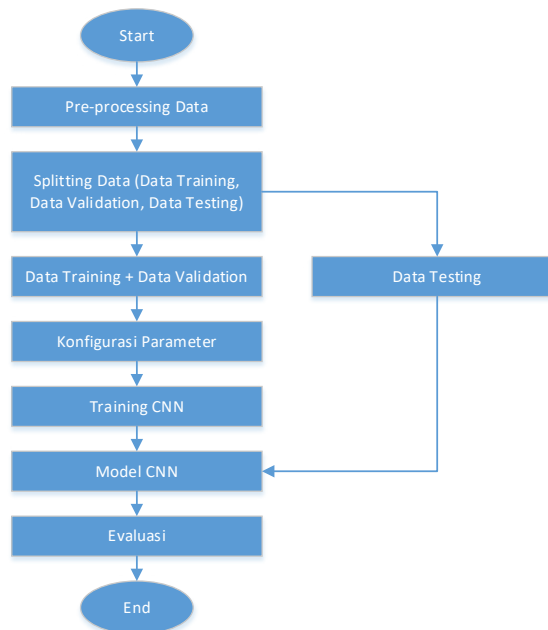

Figure 9. Research Flowchart

3. Result and Discussion

Using the Convolutional Neural Network (CNN) algorithm, the researcher categorized it into four image classes: yudistira (0), bima (1), arjuna (2), and nakula-sadewa (3). The training and data validation processes are

the most important steps in creating this model. The aim of this method is to construct a model that can detect the desired object with a high degree of accuracy.

Tests are carried out to see the effect of the depth of the convolution layer used on system performance by activating dropout regularization so that there is no overfitting due to too high an accuracy level or underfitting if the accuracy level is too low during the training process. The best-fit model would be used by the dropout regularization method, which has an effect on reducing noise data during the testing process, resulting in a high degree of accuracy. On a 46x46 pixel puppet image with 1-layer to 4-layer depth of convolution layers, this scenario is tested using 70% training data, 20% validation data, and 10% testing data.

The test results with the dropout regularization technique for each number of convolutional layers can be seen in Figures 10 to 13.
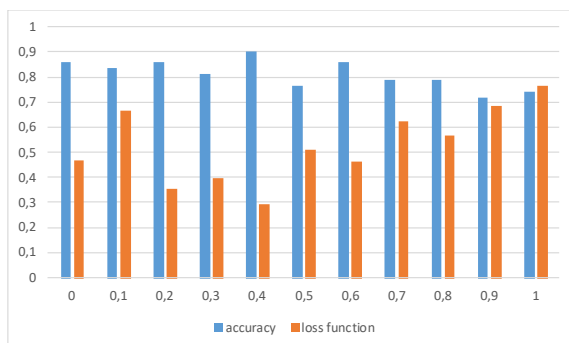

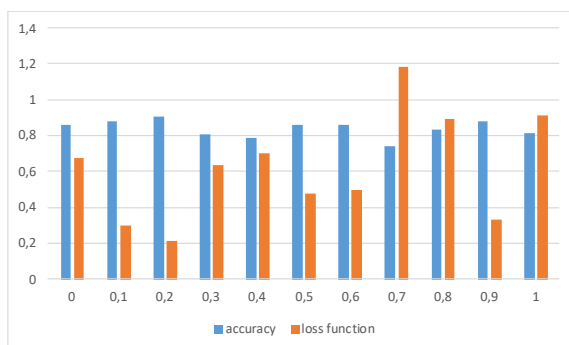**Figure 10. Adding a Dropout to a 1-Layer Convolution Layer**

Figure 11. Adding a Dropout to a 2-Layer Convolution Layer
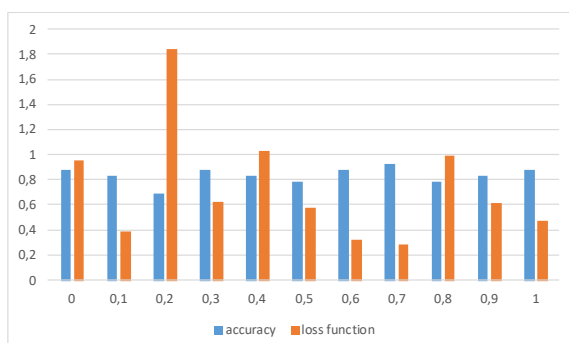

Figure 13. Adding a Dropout to a 4-Layer Convolution Layer

Based on the picture above, it can be analyzed that by using the dropout technique, the optimal accuracy rate is 0.93 with a loss function value of 0.197 in the 3-layer convolution layer, the dropout value is 0.2. Whereas in the 4-layer convolution layer the highest accuracy was also obtained at 0.93 with a loss function value of 0.291 dropout value of 0.7. Even though the accuracy value obtained is the same, the loss function value obtained in the number of 3-layer convolutions is smaller than the number of 4-layer convolutions, so the number of 3-layer convolutions was chosen as the best result in this study. Comparison of the level of accuracy obtained for each number of convolutional layers without adding a dropout (dropout = 0) with a dropout added can be seen in Figure 14. -layer) up to 0.88 (4-layer convolution) increases to 0.93 with a dropout value of 0.2.

4. Conclusion

With the addition of a 0.2 dropout in the number of 3-layer convolution layers, this study successfully implemented the convolutional neural network approach for the classification of shadow puppet images with the best accuracy percentage of 0.93 and loss function of 0.197. Of the 10% testing data, almost all have a match with existing training data. In the training process, the position and size of the images affect the accuracy and time of training data. The larger the size of the trained image, the longer the learning process, the inverted image position will affect the validity of the test results. The use of the number of layers in the training process also affects the level of accuracy in testing data. The more layers that are used, the better the results will be obtained, although it will have an impact on the length of the training process.

In the future, further research will combine convolutional neural network algorithms with several other methods to get better results. It is hoped that the existence of shadow puppets will be maintained by the classification of this shadow puppet image, namely by implementing it in an intelligent system application on a computer.

Acknowledgement

Reference

[1] Lukman Pasha, *Buku Pintar wayang*. Yogyakarta: In AzNa Books, 2011.

[2] H. S. Sudjarwo, Sumari, and U. Wiyono, *Rupa & karakter wayang purwa : Dewa-Ramayana-Mahabharata*. Jakarta: Kaki Langit Kencana, 2010.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Image Net Classification with Deep Covolutional

Neural Networks," *Commun. ACM*, pp. 1097–1105, 2012.

[4] T. Trnovszky, P. Kamencay, R. Orjesek, M. Benco, and P. Sykora, "Animal Recognition System Based on Convolutional Neural Network," *Digit. IMAGE Process. Comput. Graph.*, vol. 15, no. 3, pp. 517–525, 2017.

[5] S. Visalini, "Traffic Sign Recognition Using Convolutional Neural Networks," *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 5, no. 6, pp. 11259–11266, 2017.

[6] "MENGENAL KARAKTER TOKOH PANDAWA LIMA." [Online]. Available: http://yokimirantiyo.blogspot.com/2013/01/mengenal-karakter-tokoh-pandawa-lima.html?m=1. [Accessed: 10-Oct-2020].

[7] M. H. Purwa, "Album Wayang Indonesia," 2020. [Online]. Available: http://tokohwayangpurwa.blogspot.com/. [Accessed: 12-May-2020].

[8] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, 1998.

[9] L. Deng and D. Yu, "Deep Learning: Methods and Applications," *Found. Trends Signal Process.*, vol. 7, no. 3–4, pp. 197–387, 2014.

[10] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *ISCAS 2010 - 2010 IEEE International Symposium on Circuits and Systems: Nano-Bio Circuit Fabrics and Systems*, 2010, pp. 253–256.

[11] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. 2016.

[12] Rismiyati, "Implementasi Convolutional Neural Network Untuk Sortasi Mutu Salak Ekspor Berbasis Citra Digital," Universitas Gajah Mada, Yogyakarta, 2016.

[13] K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks," 2015.

[14] Suyanto, K. N. Ramadhani, and S. Mandala, *Deep Learning Modernisasi Machine Learning Untuk Big Data*. INFORMATIKA, 2019.

[15] U. Michelucci, *Advanced Applied Deep Learning Convolutional Neural Networks and Object Detection*. SSBM Finance Inc., 2019.

[16] H. Abhirawa, Jondri, and A. Arifianto, "Pengenalan Wajah Menggunakan Convolutional Neural Network," in *eProceedings of Engineering*, 2017, pp. 4907–4916.

[17] A. Santoso and G. Ariyanto, "Implementasi Deep Learning Berbasis Keras untuk Pengenalan Wajah," *J. Emit.*, vol. 18, no. 1, pp. 15–21, 2018.

[18] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout : A Simple Way to Prevent Neural Network from Overfitting," *J. Mach. Learn. Res.*, 2014.

[19] D. P. Kingma and J. L. Ba, "ADAM: A Method For Stochastic Optimization," *Energy Educ. Sci. Technol. Part B Soc. Educ. Stud.*, vol. 5, no. 2, pp. 365–380, 2013.