Image Classification of Golek Puppet Images using Convolutional Neural Networks Algorithm

Tuti Purwaningsih, Triano Nurhikmat, Pertiwi Bekti Utami

Lecturer of Islamic University of Indonesia
tuti.purwaningsih@uii.ac.id
Student of Islamic University of Indonesia
e-mail: 14611209@students.uii.ac.id
Student of Islamic University of Indonesia
e-mail: 14611038@students.uii.ac.id

Abstract

Golek puppet is traditional art that developed in Indonesia especially in West Java province. Golek puppet has become one of entertainment for West Java people. But in this era, many people has forget about golek puppet. It causes the youth generation starting to leave about this culture. One of the reason why this can be happened is golek puppet has to many category and it makes hard to remember all of them. With the number of Golek puppet category, the classification of category Golek puppet automatically is needed through recognition of Golek puppet image. Deep learning can be used to recognize Golek puppet images. The best method to classifying image is using Convolutional Neural Network (CNN). This study resulted accuracy with CNN method in amount of 100% accuracy to classifying Golek puppet image. Which can be decided as the best method in classifying image.

Keywords : Deep Learning, Image Classification, Golek Puppet, Accuracy, Convolutional Neural Networks

1   Introduction

Indonesia is a rich country in tradition and culture. Each region in Indonesia has its own characteristics. One of the culture that existing in Indonesia is Golek Puppet. Which is from West Java region. Golek Puppet is an art performance which the

puppet is made by wood. In International side, golek puppet has record by UNESCO as world creation in Paris on November 7th, 2003. From the achievement, golek puppet should be conserve by Indoensian people. But in reality many people has forget about golek puppet. It causes the youth generation starting to leave about this culture. One of the reason why this can be happened is golek puppet has to many category and it makes hard to remember all of them.

Currently computer vision is one of the tools that have the ability like humans, which can recognize an object in an image. In Computer Vision is inseparable from an Algorithm that helps the recognition process is included in the classification of images. Some algorithms used in the process of classification of an image that is Supoort Vector Mechine, Naive Bayes, Fuzzy, and so forth. research on image classification in an image ever done by Rosli in 2012 that is about image classification with fuzy inference engine method on the classification of mango fruit. The level of accuracy produced using this method is 80% [4].

Research from Dong-Chul Park in 2016 about the classification with many categories on the Caltech dataset using Naive Bayes [3]. The accuracy level generated using this method is 77%. Later research conducted by Hosseini in 2017 used the Support vector Mechine (SVM) method to classify hyperspectral images of space dimensions. Accuracy rate obtained by 73% - 80%. But from some of these studies the accuracy level obtained is not very optimal. So it takes a better method in the process of image classification [1].

The development of deep learning, especially the conventional neural network method, is currently widely used by researchers because the performance of this method has better results from other methods. Some research on CNN has been done by Alex Krizhevsky in 2012 This study uses the ImageNet LSVRC-2010 dataset into 1000 classes, showing a very significant result in testing test with 17% test error [2]. Then research by Tibor Trnovszky 2017, et al regarding the implementation of Convolutional Neural Network (CNN) on the introduction of animals by comparing several methods of classification. The use of CNN method gives the best result among other methods that is by giving accuracy rate of 98% [5]. Then the last application of the CNN method is implemented on the introduction of road traffic signs, such as research conducted by S. Visalini in 2017 the accuracy rate is 85% - 90% [6]. Based on the comparison of the above methods, the use of CNN method has advantages over other methods. So it can be concluded CNN has better image classification capabilities. Therefore in this study aims to classify 2 kinds of puppet show using CNN method.

2   Related Work

Table 1. Table Comparison of CNN Method

No	Writers	Dataset	Convolutional Layer	Information
1	Krizhevsky et al, (2012)	1.2 million images	5	Resulted 17% error rate
2	Muhammad Zufar dan Budi Setiyono (2016)	Primay Data From WEB CAM	2	Resulted accuracy more than 89% in 2 frame/second
3	Andrian Yusuf Wicaksono, et al, (2017)	7112 Image of Batik	2	Resulted accuracy 70.84 %
4	Yiyu Hong dan Jongweon Kim (2017)	30000 Image of Painting	5	Resulted 2% *error* and error in SIFT method resulted 15,6%, where the gap is about 13,6 %
5	Tibor Trnovszky, et al, (2017)	500 Image of Animals	2	Resulted accuracy 98 %
6	S. Visalini (2017)	500 Image of *Traffic signs*	2	Resulted accuracy 85-95 %

Table 2. Comparison Table Other Methods

No	Writers	Dataset	Method	Information
1	Rosli, et, al (2012)	Data of Mango	Fuzzy Logic	Resulted accuracy 80 %
2	Dong-Chul Park, (2016)	Data *Caltech*	*Naive Bayes*	Resulted accuracy 77%
3	Lida Hosseini,(2017)	Data *Hyperspectral*	*Support Vector Machine*	Resulted accuracy 73-80 %

3　Problem Formulations or Methodology

Since Golek puppet is one of culture in Indonesia that should be conserve, one way how us solving the problem is do the classification image using Convolutional Neural Networks. Convolutional Neural Network (CNN) is the development of multilayer perceptron (MLP) designed to process two dimensional data in image form. CNN is included into the type of Deep Neural Network because of the high depth of the network and applied to many image data. Basically image classification can be used with MLP, but with MLP method is less suitable for use because it does not store spatial information from the data of ideals and assume each pixel is an independent feature to produce unfavorable results.

Convolutional Neural Networks is one of method in images classification that gives the best result than the other method because in this method there are a process called feature extraction. Where in this process the input image will be changed to a matrix with ordo size nxn. The matrix that has been formed is representative of the image and will become the input for the next process.

Another reason why we choose this method is because this network has a special coating which is named with a layer of convolution, where on this layer an image insert will be processed based on the filters that are already specified. Of each layer it will produce a pattern of some parts of the image that will be easier to be classified, making learning process more efficient to be implemented.

The image of Golek Puppet is taken from search engine named google. We choose 2 category of Puppet ie Cepot and Semar. Each of puppet image has 100 images, where 160 images is used as data training and the rest is 40 images used to testing the model so the total of input image is 200 images. These are the image of Golek Puppet that used in this study.

Fig 1. Cepot Puppet

Fig 2. Semar Puppet

4   The Proposed Method

This study proposed to classify 2 category images of Golek   Puppets using Convolutional Neural Networks algorythm, fig. 3 shows the design flow of Convolutional Neural Networks algorithm.
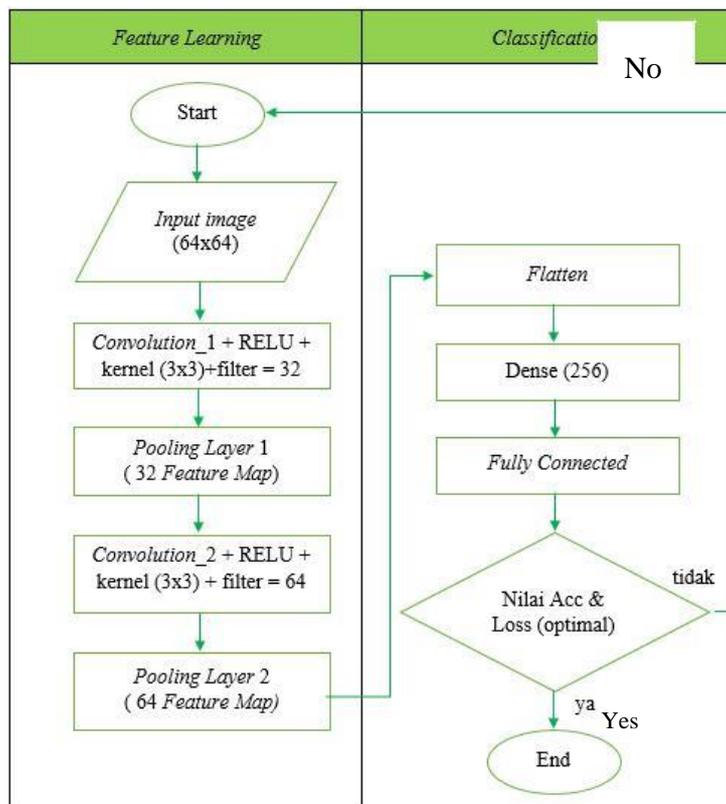
Fig 3. Design Flow Chart of Convolutional Neural Network

Design of the model that used in this study has two stages in the CNN model in general, namely the feature of extraction and classification. At the feature extraction stage, the input image size is 64x64x3. The third digit is an image that has 3 channels namely Red, Green, and Blue (RGB). This research uses 3 layer convolution and 2 layer pooling. Each of the convolution layers uses a 3x3 size filter. Then the process of pooling using max-pooling method. Then at the classification stage of artificial neural network that has a hidden layer.

4.1   Convolutional Layer

Convolution layer is part of the stage on CNN architecture. This stage performs a convolution operation on the output of the previous layer. The layers are the main process that build the CNN architecture network. Convolution operation is an operation on two functions of a real-valued argument. This operation applies the output function as Feature Map of the input image. These inputs and outputs can be seen as two real-valued arguments. The convolution operation can be written as follows:

$$S(t) = (x * t)(t) = \sum_{\alpha = -\infty}^{\infty} [\![x(\alpha) * w(t-\alpha)]\!] \tag{1}$$

Explanation:
S (t)   = Function of convolution operation result
X   = Input
W   = weight (kernel)

In addition, the determination of the output volume can also be determined from each layer with hyperparameters. The hyperparameter used in the equation below is used to calculate the number of activation neurons in a single output. Consider the following equation:

$$(W-F + 2P) / (S + 1) \tag{2}$$

Explanation :
W   = Image volume size
F   = Filter Size
P   = Padding value used
S   = Size Shift (Stride)

Here is a picture of the convolution process.
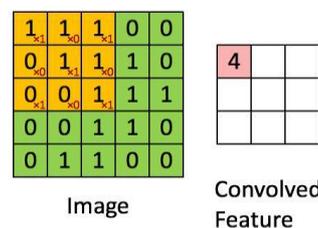
Fig 4. Convolution Process

4.2 Pooling Layer

Convolutional networks may include local or global pooling layer, which combine the outputs of neuron clusters at one layer into a single neuron in the next layer. For example, *max pooling* uses the maximum value from each of a cluster of neurons at the prior layer. Another example is *average pooling*, which uses the average value from each of a cluster of neurons at the prior layer. In this study we use max pooling. Fig. 5 shows the max pooling process.
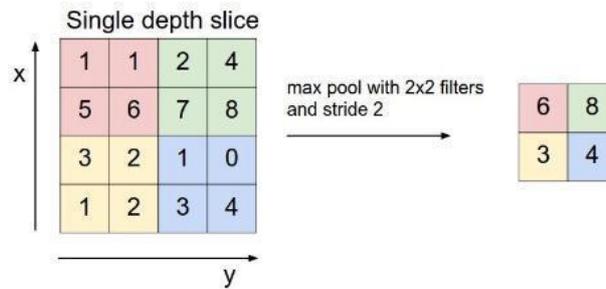
Fig 5. Max Pooling Process

4.3 Fully Connected Layer

The Fully-Connected Layer is a layer where all the activation neurons of the previous layer are all connected with neurons in the next layer just as with ordinary neural networks. Basically this layer is usually used on MLP (Multi Layer Perceptron) which has a purpose to transform the dimensions of data so that data can be classified in a linear manner. The difference between the Fully-Connected layer and convolution layer is that neurons in the convolution layer has an overall connected neuron. However, both layers still operate the dot product, so its function is not so different. Fig. 6 shows the Fully-Connected layer process.
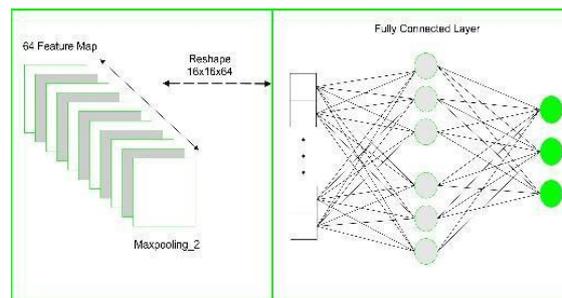
Fig 6. Fully-Connected layer process

4.4   Activation Function

The activation function is a function that measures the relationship between the internal level (sum function) which may be linear or non-linear. This function aims to determine whether neurons are activated or not. One of the functions used in CNN is the activation function of ReLU (Rectified Linear Unit). Basically the ReLU (Rectified Linear Unit) function performs a "treshold" from 0 to infinity. ReLU Activation layer and convolution layers is that neurons in the convolution layer connect only to specific regions of the input, while the Fully-Connected layer has an overall Function In this function, the input of the neurons consists of negative numbers, then this function will be used for those values, and if the positive value of the output of the neuron is the value of the activation itself. Fig. 7 shows about actiovation function of ReLU:
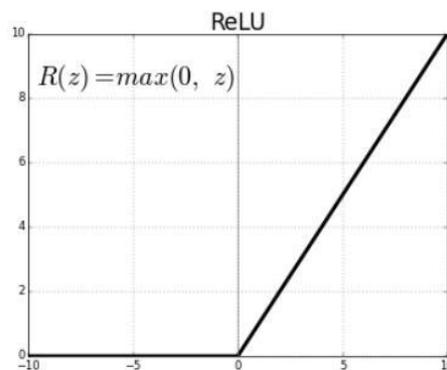
Fig 7. Activation Function of ReLU

In this function the input of the neurons is a negative number, then this function will translate that negative value into 0 (zero), and if the input is positive then the output of the neuron is the value of the activation itself.

5   Results, Analysis and Discussions

Based on the results of the training model obtained the following results:

5.1   Result of Training Model

After going through some process in Convolutional Neural Network (CNN) algorithm got the result of training and validation. This process uses the number 100 epoch, the value of learning rate 0,001. The following graph of training results using tensorboard:

Tuti Purwaningsih et. al
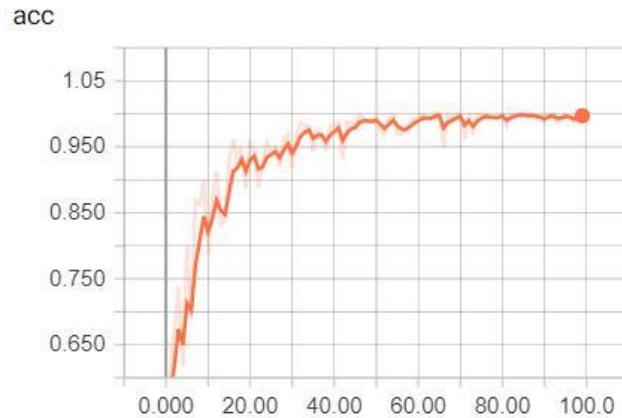
Fig 8. *Training Accuracy*
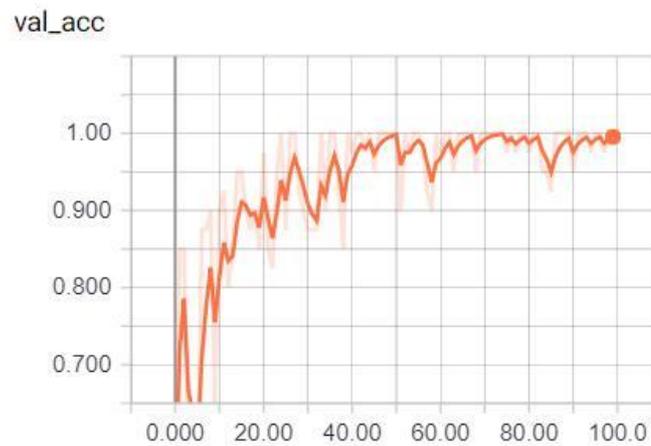
Fig 9. *Validation Accuracy*

Based on Fig 6 and 7 obtained accuracy from training model reaches 100% with loss value of 0.0066433. The training time required for 100 epoch in the training of this model is 583 seconds. Then the accuracy of the data validation reaches 100% with a loss value of 0.0320.

5.2   Result of Testing New Data

The testing process uses 40 test data, for each puppet type class of 20 images. The result of the confusion matrix is as follows:

Table 3. *Confusion Matrix*

Matrix		Pred Class	
		Cepot	Semar
Actual	Cepot	20	0
Class	Semar	0	20

Based on the above table the results of predictions from the model of data testing new data shows optimal results. The prediction on Cepot is classified into Cepot as much as 20 images, this means that the classification of the picture can be said very precisely. Then Prediction of Semar classified into Semar of 20 images, this means that the classification of the image can be said very precise matrix above is as follows:

$$\text{Overall Accuracy} = \frac{\text{TTPall}}{\text{Total Number of Testing Entries}}$$

$$\text{Overall Accuracy} = \frac{40}{40} = 100\ \%$$

The accuracy generated by the model with 64x64 pixel image input, the value of the learning rate of 0.0001 resulted in the accuracy of the new data testing of 100%.

5.3   Result of Convolution Layer

In this section we try to add another experiments with another number of convolution layer. And this is the result :

Table 4. Comparison of Convolutional Layer

Layer Convolution	Accuracy Validation	Loss Validation	Time (Seconds)
2	100%	0.0320	528
3	100%	0.0168	650
4	95%	0.2451	675
5	97%	0.1492	613

Based on table 4, the use of many convolution layers in this study, the more use of convolution layer will slow down the process of model training, this is caused by many extraction stages of the features or images performed by the computer and it takes a long time. So from the table can be seen the more the number of convolution layer, the time required in the process of training the model will be more and more. In this study only uses 2 layer convolution to minimize the time in the process of modeling.

5.4   Result of Training and Testing Data

In this section we try to add another experiments with another number of training and testing data. And this is the result :

Tuti Purwaningsih et. al

Table 4. Comparison of Testing and Training Data

(Training :Testing)	Accuracy Validation	Loss Validation	Time (Seconds)
(70 : 20)	97%	0.05682	719
(80 : 20)	100%	0.03200	528
(90 : 20)	100%	0.01545	536

Based on table 4, the experiments show the three scenarios of training and testing data, it can be assumed that the more the amount of data train, the higher the acquisition obtained by the model. This is because the more the model is trained with many pictures, the model will recognize the pattern of the image more accurately.

6   Conclusion

Based on the results of the analysis that has been done, obtained some conclusions are:

1. The CNN model in this research uses 64x64 input shape, 0.0001 learning rate, 3x3 filter size, Epoch 100 count, data training 160, and data testing 40. Produce the accuracy level of training and testing in classification of golek puppet image are 100% accuracy.

2. This study uses a new data testing of 40 to be tested into the model that has been made. The result of data test resulted in a new level of accuracy in classifying puppet images of 100%.

References

[1] Hosseini, L., & Ramin Shaghaghi Kandovan. (2017). Hyperspectral Image Classification Based on Hierarchical SVM Algorithm for Improving Overall Accuracy. *Scientific Research Publishing*, 66-75

[2] Krizhevsky, A., Ilya Sutskever, and Geoferry E. Hinton. (2012). Image Net Classification with Deep Covolutional Neural Network. *Communications of the ACM*, 1097-1105

[3] Park, D.-C. (2016). Image Classification Using Naïve Bayes Classifier. *International Journal of Computer Science and Electronics Engineering (IJCSEE)*, Vol. 4 , ISSN 2320–4028

[4] Rosli, R., et al. (2012). Mango Grading By Using Fuzzy Image Analysis. *International Conference on Agricultural, Environment and Biological*, (hal. pp.18–22).

[5] Trnovsky, T., Dkk. (2017). Animal Recognition System Base On Convolutional Neural Network . *Digital Image Processing And Computer Graphics*, Vol.15, No.3

[6] Visalini, S. (2017). Traffic Sign Recognition Using Convolutional Neural Network. *International Jurnal of Innovative Research in Computer and Communication Engineering*, Vol.5.