



## Ensemble Convolutional Neural Network for Robust Batik Classification

---

Yufis Azhar, Moch Chamdani Mustaqim and Agus Eko Minarno

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

March 9, 2021

# Ensemble convolutional neural network for robust batik classification

Yufis Azhar, Moch. Chamdani Mustaqim, Agus Eko Minarno

Universitas Muhammadiyah Malang

Jl. Raya Tlogomas 246, Malang, East Java, Indonesia

danicham14@gmail.com

**Abstract.** Some researchers propose using the Convolutional Neural Network (CNN) method to classified batik images. It can extract features automatically without the need to define feature manually from the image. However, CNN's weakness is that its accuracy is quite low, especially for small-sized datasets, compared to machine learning methods that use hand-crafted feature extraction. In this research, an ensemble CNN method is proposed to improve the accuracy of the CNN method in classifying batik images. This method will train several CNN models at once, and then by voting and averaging techniques, the output label will be determined. Test results for two different datasets show this method can improve the accuracy of the CNN method and get an accuracy value of 100%. This method is also proven to extract features faster than the state-of-the-art method like MTCD+SVM, which is included in the hand-crafted feature extraction category.

## 1. Introduction

Batik is one of Indonesia's cultural heritage. Research on batik has started in 1974 [1], and since then, the development of research on batik is multiplying. In computer vision, batik is often used as a research object about pattern recognition, image retrieval, object detection, and so on.

Machine learning is an approach that is often chosen for the classification of batik images. Handayani et al. managed to get a maximum accuracy value of 99% after using the voting features interval method to classify 163 batik images into five categories [2]. In machine learning, feature extraction plays a significant role. As evidenced by Suciati et al. when using a fast discrete curvelet transform on HSV color features. The accuracy results showed that the use of HSV is better than RGB color features [3]. Rangkuti also mentioned color features as a critical feature to recognize batik patterns in his research. In the study, Rangkuti used treeval and treefit as decision tree function in optimizing content based batik image retrieval. The results showed that the decision tree approaches properly-recognized batik patterns with 90% accuracy [4]. Batik pattern recognition was also done by Nurhaida et al. in 2015. The SIFT method that used in this research was able to recognize a batik pattern with an error rate of 0.08 [5]. In 2018, Caraka et al. used geometric invariant moments to detect batik parang rusak patterns. This method is proven to perform exceptionally well when used on binary classification datasets with 92% accuracy [6]. While in 2015, Aditya et al. proposed Grey Level Co-occurrence Descriptor (GLCM) and Neural Network to recognize batik patterns. The accuracy obtained by this research was 91% [7]. Both studies prove that in addition to color features, edge detection and texture detection can also be used to recognize batik patterns. GLCM's success in

improving models' accuracy to recognize batik patterns makes this method widely used in other studies [7–11]. In 2020, Azhar et al. also proposed a combined approach of the Multi Texton Co-occurrence Descriptor, which is based on GLCM, with the TF-IDF weighting method commonly used in text retrieval. This method is called Texton Frequency - Inverse Image Frequency (TF-IIF). The results obtained show an increase in 7% precision compared to the previous process [12]. In addition to GLCM, the SIFT approach is also widely used in research on batik pattern recognition. Azhar et al. and Setyawan combine SIFT and KNN [13,14]. Pradnyana et al. also used KNN as a Balinese fabric classifier because of its simplicity [15].

Besides of machine learning approach, deep learning is also often used for the classification of batik images. Among the many proposed methods, the Convolutional Neural Network (CNN) is one of the most widely used methods [16–20]. In 2017, Handayani et al. used CNN and the Deep Belief Network (DBN) to classify batik datasets. However, both methods provide poor accuracy results, with 63% on CNN and 44% on DBN [21]. In 2018, Agastya et al. used augmentation data techniques to increase the number of datasets before they were classified using CNN and its derivative variants (VGG16 and VGG19). Although the accuracy results reached 90%, its accuracy dropped to 56% when the augmentation data was applied [22]. Based on research that has been proposed by Arsa et al., classification using deep learning also has not been able to provide maximum results [23]. It depends on the dataset used. The amount of data in each class dramatically affects the learning process in the deep learning algorithm. Also, the determination of convolution layer models, hidden layers, and other CNN parameters significantly affect the accuracy results.

In the previous research, Minarno et al. classified batik images by utilizing the sub-band image [11]. The wavelet and GLCM methods are combined with the Probabilistic Neural Network (PNN) to obtain 72% accuracy. In the next research, Minarno et al. proposed Multi Texton Histogram (MTH) method with six textons as a feature extractor for batik combined with PNN, KNN, and SVM methods solve the same problem. The results obtained are accuracy scores of 92%, 82%, and 76% respectively [24,25]. In 2019, Minarno et al. using the GLCM method combined with KNN and SVM. The results obtained are 78% for KNN and 92% for SVM [26]. Based on the previous research, both GLCM and MTH could be used to classify batik images quite well. In 2020, Minarno et al. combined MTH and GLCM in a new method called Multi Texton Co-occurrence Descriptor (MTCD). This method used six filters called textons to generate a color histogram and edge on an image by convolution. The two histograms are then combined with the entropy, correlation, energy, and contrast values obtained through the GLCM method. In this research, MTCD was paired with KNN and SVM for the classification of batik images. The results obtained were proven to be able to outperform the previous studies. Both KNN and SVM were able to get an accuracy value of 96% [27]. From these studies, it can be seen that the accuracy value of the classification model is largely determined by the features selected to be extracted. And these features must be determined manually or better known as hand-crafted feature extraction. This is one of the weaknesses of the machine learning method.

In terms of practicality of use, deep learning is still better than machine learning because it did not need to define features that must be extracted first, as it does in machine learning methods. Deep learning was used to cover up the deep learning method's weaknesses that are still getting low accuracy on small datasets. An ensemble CNN method was proposed to improve the accuracy of the CNN method in classifying batik images. This method will train several CNN models at once, and then by voting and averaging techniques, the output label will be determined. Test results for two different datasets show this method can improve the accuracy of the CNN method and get an accuracy value of 100%. This method is also proven to extract features faster than the MTCD + SVM method, which is included in the category of hand-crafted feature extraction.

## **2. Research method**

This research's primary focus is to create a deep learning architecture that can classify batik images well. The primary method used is the Convolutional Neural Network (CNN). A dropout layer was

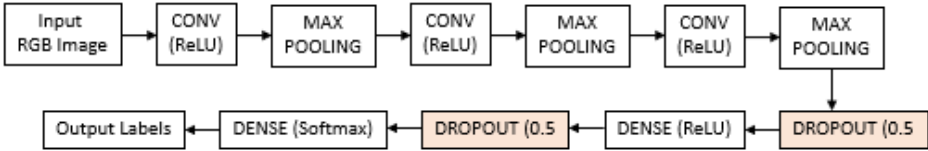
added to the CNN architecture to produce better accuracy. Figure 1 shows the architecture used in this research.

As we know, overfit is one of the problems that often arise in implementing deep learning in image classification. And the way to handle it is by using the concept of regularization. Adding layer dropout to the network is one form of regularization that can be applied. How it works from the dropout layer was to break the input relationship from the previous layer to the next layer with a certain probability. In this research, the probability value was set to 0.5, so it was expected that there would be half of the interrupted relationship between the extraction and the fully connected layers. Thus, overfit is expected to be prevented. From Figure 1, it can be seen that there are two dropout layers installed on the network. The first layer is before the Fully Connected (FC) layer (after the feature was extracted), and the second is between the dense layer on the FC Layer.

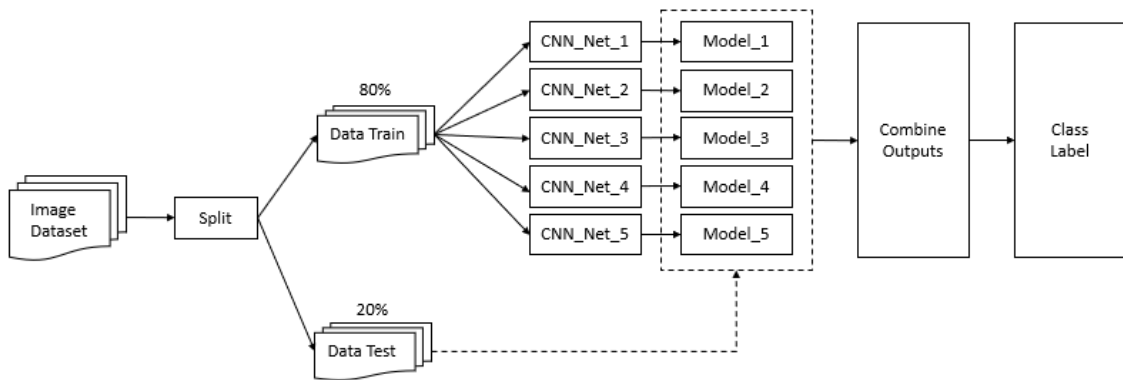
In detail, there are a total of 10 layers used on the CNN architecture in this research. Consists of three pairs of Convolution and Max-Pool layers, two dense layers, and two dropout layers. Each layer uses ReLU as its activation function, except for the last dense layer that uses the softmax function. Each convolution layer has 32, 50, and 80 filters (kernels), respectively. Each convolution and pool layer have filters of the same size, namely 2x2, and stride = 1. Each input image will be resized to a size of 32x32 with a depth of 3 (RGB) before being incorporated into the CNN architecture.

Apart from adding a dropout layer, the CNN architecture was also combined with the ensemble method. Figure 2 shows the design using the ensemble and CNN methods in classifying batik images. There are two main phases in the design method proposed. The first phase is the training phase. In this phase, the data train will be entered into CNN with the same architecture five times. The result is five different models, with varying values of accuracy. This has happened because, on CNN, the weight of each kernel was initially set at random. As the training process progresses, the filter's weight will become more convergent towards the best value. However, because the initial value is different, each model's best value in the last epoch will also be different. Although the difference is not very significant, it will also affect the classification process of data. And then the second phase is the testing phase. In this phase, the test data will be predicted using the five models that have been created previously. Then, the label of the test data will be determined using voting techniques. All these stages are referred to as the ensemble CNN method.

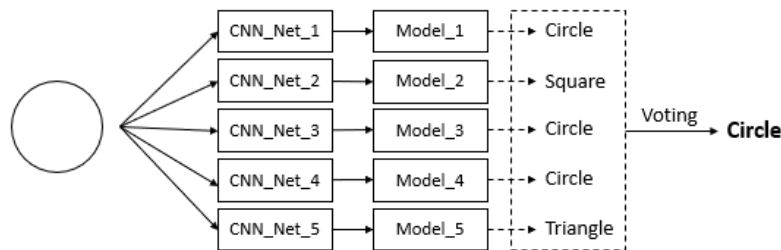
The ensemble method itself refers to a technique that uses a large number of models (where the exact number of models used is very dependent on the case of classification), then combines the results of the prediction through voting or averaging to improve the accuracy of the classification results. Figure 3 exemplifies how the ensemble method is paired on CNN to combine the output produced by each model. Using the voting technique, the label that is guessed by many models is the label that will be selected as the output label.



**Figure 1.** Network architecture with the addition of a dropout layer after the feature extraction layer and before the output layer



**Figure 2.** The proposed ensemble CNN method



**Figure 3.** The ensemble (combine output) process on CNN to determine the output label of a data

### 3. Result and discussion

There are two types of datasets used in this research. The first dataset is Batik300 (Figure 4). This dataset consists of 300 batik images, which are divided into 50 different classes. Each class consists of 6 images. These six images are the results of the augmentation of an image with a large dimension. All images in this dataset have the same dimensions, i.e., 128 x 128 pixels. The second dataset is Batik41k (Figure 5). This dataset has characteristics similar to Batik300, but with a higher number of classes, namely 351 classes. Another thing that also distinguishes this dataset from Batik300 is the imbalance dataset condition. Batik351 has a different amount of data in each category. This dataset's total batik image is 41,621 data, with the dimensions of each image being the same, namely 400 x 600 pixels. The two datasets were split into two parts, 20% as the test data, and the rest as the data train.



**Figure 4.** Examples of Image Contained in Batik300 Dataset



**Figure 5.** Examples of Image Contained in Batik41k Dataset

Three test scenarios will be carried out in this research. The first scenario is to look at adding layer dropouts on CNN architecture in classifying batik images. The purpose of this test is to see how much influence the dropout layer has on handling overfit.

There are two architectures to be tested—one architecture without a layer dropout and another architecture with a layer dropout (Figure 1). Both architectures used the same hyperparameter.

Hyperparameter tuning is performed using the Keras Tuner library. The led hyperparameter is the learning rate and optimization function. Of the 5 learning rate scores tested (1e-1, 1e-2, 1e-3, 1e-4, 1e-5), the most optimal results for learning rate is 1e-3. As for optimization function, testing of Adam, SGD, and RMSprop function. Adam gives the best results. The results obtained when testing the two architectures in the Batik300 dataset are shown in Figure 6. Although both architectures can produce reasonably good accuracy values (above 0.8), it is seen that architectures that do not use dropout layers tend to overfit. This is concluded from the observation that the value of loss tends to increase, and the value of accuracy tends to decrease from the 86<sup>th</sup> epoch. Different things are shown in the results of architectural training using layer dropouts. It is seen that the value of loss continues to fall close to 0 and does not indicate an increase. The same thing happened in the accuracy value plot, which continuously increased to close to 1. From this test, it can be concluded that the addition of the dropout layer proved to be sufficient enough to suppress the tendency to overfit the classification results.



**Figure 6.** Comparison of loss and accuracy values obtained by models

The second testing scenario conducted in this research is the use of the ensemble method on CNN. The purpose of using this method is to increase the accuracy value of CNN. In this research, the ensemble method was carried out by combining the results of the classification of 5 CNN models trained using the same architecture and dataset. Comparing the proposed method and the previous method is also applied using the same hardware and the same dataset. The results can be seen in Table 1 and Table 2.

**Table 1.** The accuracy values of the proposed method compared with previous studies for the Batik300 dataset

Method	Accuracy
MTH+SVM [26]	0.92
MTCN+SVM [27]	0.96
CNN	0.98
Ensemble CNN	1

**Table 2.** Comparison of result before and after the ensemble method was applied to the CNN architecture

Dataset	Parameter	CNN	Ensemble CNN
Batik300	Accuracy	0.98	1
	Precision	0.95	1
	Recall	0.98	1
	F1-Score	0.96	1
Batik41k	Accuracy	0.99	1
	Precision	0.97	1
	Recall	0.99	1
	F1-Score	0.98	1

**Table 3.** Comparison of the time needed to extract features in batik41k test data between the previous and proposed method

Method	Time (s)
MTCD+SVM [27]	235.23
Ensemble CNN	9.54

In Table 1, the proposed method (ensemble CNN) can produce an excellent accuracy value (100%) for the Batik300 dataset. It outperformed the results obtained by previous studies in [26] and [27]. Even when compared to traditional CNN methods, the ensemble CNN is still superior. Meanwhile, in Table 2, if we focus on the comparison between the traditional CNN method and the ensemble CNN for the two types of datasets (Batik300 and Batik41k), it can be seen that the ensemble CNN can outperform almost all measurement parameters. Especially for the batik41k dataset, the Ensemble CNN's accuracy value is the same as [27]. But, the use of the Ensemble CNN method was considered to have advantages. For example, at the speed of the extraction of its features, this hypothesis was proven using the third testing scenario.

In the third test scenario, predictions were made using models that have been formed by the ensemble CNN and MTCD + SVM methods. Because both methods have been tested using the same dataset, and both can get perfect accuracy values (i.e., 100%). Both methods were considered to have the same ability in classifying batik images. Variables that will be compared in this test are the time needed by the two methods in extracting features for testing data (new data). Both methods will be given the same dataset, namely Batik41k. Then, both methods were trained using the same number and variety of train data. After the model was formed, the two methods were asked to predict as many as 8,094 batik images, which were still in the form of raw images. The results obtained in Table 3 show that the ensemble CNN method is superior to the MTCD + SVM method in terms of its feature extraction speed. In the MTCD method, each feature (color, edge, and texture) must be extracted separately. For color and edge features, extraction is done by converting six filters one by one. This process needs to be done for each batik image (be it a training image or test image). Unlike MTCD, the extraction of features on CNN is done by creating a feature map in each layer. The process of creating this feature map can be done very quickly at the testing stage because the kernel weight in each layer has been obtained at the time of training. On the MTCD+SVM method, feature extracted is performed by MTCD. While on CNN, the feature extraction process is done on the input layer until the flatten layer. The total time to classify an image is calculated from the image is inserted into the model until all features have been extracted (histogram feature vector for MTCD and flatten feature map for CNN). Since each kernel's best weight value has been obtained during the training process, then at the time of testing, CNN only needs to perform one convoluted process in each convolutional

layer to get the best feature map (only one feed-forward step from the input layer to output layer). This is what causes the testing process of the CNN ensemble to be faster than MTCN+SVM.

#### 4. Conclusion

This research has presented a method for classifying two batik datasets using dropout, ensemble testing scenarios, and comparing testing times between Ensemble CNN vs. previous research method. Based on the test results, it can be concluded that the addition of the dropout layer was able to overcome overfitting, and the ensemble technique can improve the accuracy of the previous method, like MTH+SVM or MTCN+SVM. Ensemble CNN can also increase the accuracy, precision, recall, and f-measure value of the CNN method, both for batik300 and batik41k dataset. Although ensemble CNN and MTCN+SVM reach accuracy 1, in the testing process, Ensemble CNN has a faster computing process than MTCN + SVM with a time difference of 225.69 seconds.

#### References

- [1] Bartleson FE. Totem Name Batik Kit. United State; U.S. Patent No. US3840113A, 1973.
- [2] Handhayani T. Batik Lasem images classification using voting feature intervals 5 and statistical features selection approach. Proceeding - 2016 Int Semin Intell Technol Its Appl ISITIA 2016 Recent Trends Intell Comput Technol Sustain Energy. 2016;13–6.
- [3] Suciati N, Kridanto A, Naufal MF, Machmud M, Wicaksono AY. Fast discrete curvelet transform and HSV color features for batik image classification. In: 2015 International Conference on Information & Communication Technology and Systems (ICTS). Surabaya; 2015. p. 99–104.
- [4] Rangkuti AH, Rasjid ZE, Santoso DJ. Batik Image Classification Using Treeval and Treefit as Decision Tree Function in Optimizing Content Based Batik Image Retrieval. In: Procedia Computer Science. 2015. p. 577–83.
- [5] Nurhaida I, Noviyanto A, Manurung R, Arymurthy AM. Automatic Indonesian's Batik Pattern Recognition Using SIFT Approach. In: Procedia Computer Science [Internet]. Elsevier Masson SAS; 2015. p. 567–76. Available from: <http://dx.doi.org/10.1016/j.procs.2015.07.547>
- [6] Caraka RE, Pardamean TWCB. Batik Parang Rusak Detection Using Geometric Invariant Moment. In: 2018 Indonesian Association for Pattern Recognition International Conference (INAPR). 2018. p. 71–4.
- [7] Aditya CSK, Hani'ah M, Bintana RR, Suciati N. Batik classification using neural network with gray level co-occurrence matrix and statistical color feature extraction. In: 2015 International Conference on Information & Communication Technology and Systems (ICTS). Surabaya; 2015. p. 163–8.
- [8] Sulistianingsih N, Soesanti I, Hartanto R. Classification of Batik Image using Grey Level Co-occurrence Matrix Feature Extraction and Correlation Based Feature Selection. In: 2018 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI). Yogyakarta, Indonesia; 2018. p. 492–7.
- [9] Juwita AR, Solichin A. Batik Pattern Identification using GLCM and Artificial Neural Network Backpropagation. In: 2018 Third International Conference on Informatics and Computing (ICIC). Palembang, Indonesia; 2018. p. 1–6.
- [10] Irawan C, Ardyastiti EN, Setiadi DRIM, Rachmawanto EH, Sari CA. A survey: Effect of the number of GLCM features on classification accuracy of lasem batik images using K-nearest neighbor. In: 2018 International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2018. 2018. p. 33–8.
- [11] Minarno AE, Munarko Y, Kurniawardhani A, Bimantoro F. Texture Feature Extraction Using Co-Occurrence Matrices of Sub-Band Image For Batik Image Classification. In: Information and Communication Technology (ICoICT). IEEE; 2014. p. 249–54.
- [12] Azhar Y, Minarno AE, Munarko Y, Ibrahim Z. Image Retrieval Based on Texton Frequency-



- Inverse Image Frequency. *Kinet Game Technol Inf Syst Comput Network, Comput Electron Control*. 2020;5(2):87–94.
- [13] Azhar R, Tuwohingide D, Kamudi D, Sarimuddin, Suciati N. Batik Image Classification Using SIFT Feature Extraction, Bag of Features and Support Vector Machine. In: *Procedia Computer Science* [Internet]. Elsevier Masson SAS; 2015. p. 24–30. Available from: <http://dx.doi.org/10.1016/j.procs.2015.12.101>
- [14] Setyawan I, Timotius IK, Kalvin M. Automatic batik motifs classification using various combinations of SIFT features moments and k-Nearest Neighbor. In: *2015 7th International Conference on Information Technology and Electrical Engineering (ICITEE)*. Chiang Mai; 2015. p. 269–74.
- [15] Pradnyana GA, Suryantara IKA, Darmawiguna IGM. Impression Classification of Endek (Balinese Fabric) Image Using K-Nearest Neighbors Method. *Kinet Game Technol Inf Syst Comput Network, Comput Electron Control*. 2018;3(3):213–20.
- [16] Rasyidi MA, Bariyah T. Batik pattern recognition using convolutional neural network. *Bull Electr Eng Informatics*. 2020;9(4):1430–7.
- [17] Wicaksono AY, Suciati N, Fatichah C, Uchimura K, Koutaki G. Modified Convolutional Neural Network Architecture for Batik Motif Image Classification. *IPTEK J Sci*. 2017;2(2):26–30.
- [18] Prasetyo H, Putra Akardihas BA. Batik image retrieval using convolutional neural network. *Telkomnika (Telecommunication Comput Electron Control)*. 2019;17(6):3010–8.
- [19] Tristanto J, Hendryli J, Erny Herwindiati D. Classification of Batik Motifs Using Convolutional Neural Networks. *SSRN Electron J*. 2018;1–5.
- [20] Puarungroj W, Boonsirisumpun N. Convolutional Neural Network Models for HandWoven Fabric Motif Recognition. In: *2019 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT-NCON)*. 2019. p. 300–3.
- [21] Handhayani T, Hendryli J, Hiryanto L. Comparison of shallow and deep learning models for classification of Lasem batik patterns. In: *Proceedings - 2017 1st International Conference on Informatics and Computational Sciences, ICICoS 2017*. 2017. p. 11–6.
- [22] Agastya IMA, Setyanto A. Classification of Indonesian Batik Using Deep Learning Techniques and Data Augmentation. In: *2018 3rd International Conference on Information Technology, Information System and Electrical Engineering (ICITISEE)*. Yogyakarta, Indonesia; 2018. p. 27–31.
- [23] Arsa DMS, Susila AANH. VGG16 in Batik Classification based on Random Forest. In: *2019 International Conference on Information Management and Technology (ICIMTech)*. Jakarta/Bali, Indonesia; 2019. p. 295–9.
- [24] Minarno AE, Munarko Y, Kurniawardhani A, Bimantoro F. Classification of Texture Using Multi Texton Histogram and Probabilistic Neural Network. In: *International Conference on Engineering and Technology for Sustainable Development*. 2015. p. 12–22.
- [25] Minarno AE, Maulani AS, Kurniawardhani A, Bimantoro F, Suciati N. Comparison of methods for Batik classification using multi texton histogram. *Telkomnika (Telecommunication Comput Electron Control)*. 2018;16(3):1358–66.
- [26] Minarno AE, Sumadi FDS, Wibowo H, Munarko Y. Classification of batik patterns using K-Nearest neighbor and support vector machine. *Bull Electr Eng Informatics*. 2020;9(3):1260–7.
- [27] Minarno AE, Azhar Y, Sumadi FDS, Munarko Y. A Robust Batik Image Classification using Multi Texton Co-Occurrence Descriptor and Support Vector Machine. In: *2020 3rd International Conference on Intelligent Autonomous Systems (ICoIAS)*. Singapore; 2020. p. 51–5.