Batik Classification using Deep Convolutional Network Transfer Learning

Yohanes Gultom, Aniati Murni Arymurthy

Faculty of Computer Science, Universitas Indonesia

Email:yohanes.gultom@ui.ac.id

Abstract

Batik fabric is one of the most profound cultural heritage in Indonesia. Hence, continuous research on understanding it is necessary to preserve it. Despite of being one of the most common research task, Batik's pattern automatic classification still requires some improvement especially in regards to invariance dilemma. Convolutional neural network (ConvNet) is one of deep learning architecture which able to learn data representation by combining local receptive inputs, weight sharing and convolutions in order to solve invariance dilemma in image classification. Our experiments show that the proposed model, which used deep ConvNet VGG16 for transfer learning, outperformed SIFT-based and SURF-based classification models in both accuracy and speed. Our model achieved 74% accuracy and required 64s processing time (with GPU) while SIFT and SURF achieved 25% (in 492s) and 40% (in 416s) respectively.

Keywords: *Batik, classification, deep learning, transfer learning*

Abstrak

Kain Batik adalah salah satu warisan kebudayaan Indonesia yang sangat berharga. Oleh karena itu, penelitian yang berkesinambungan perlu dilakukan untuk melestarikannya. Sekalipun telah menjadi topik penelitian yang umum, klasifikasi pola Batik secara otomatis masih memiliki beberapa tantangan yang perlu diselesaikan. Salah satu tantangan tersebut adalah masalah *invariance dilemma*. *Convolutional neural network* (ConvNet) adalah salah satu arsitektur *deep learning* yang mampu mempelajari representasi data dengan mengkombinasikan teknik *local receptive inputs*, *weight sharing* dan *convolutions* untuk mengatasi masalah *invariance dilemma* pada klasifikasi citra seperti pola Batik. Melalui eksperimen ditemukan bahwa model yang diajukan, yang menggunakan *transfer learning* dari ConvNet VGG16, mencapai akurasi dan kecepatan klasifikasi pola batik yang lebih baik dari model berbasis SIFT dan SURF. Model yang diajukan mencapai akurasi 74% dan waktu proses 64 detik (karena dapat menggunakan GPU). Sedangkan model SIFT mencapai akurasi 25% dalam 492 detik dan dan SURF mencapai 40% dalam 416 detik untuk *dataset* yang sama.

Kata Kunci: *Batik, klasifikasi, deep learning, transfer learning*

1. Introduction

Batik fabric is one of the most profound cultural heritage in Indonesia. Hence, continuous research on understanding it is necessary to preserve it. One of the most popular research topic is batik classification.

Since the most prominent feature of Batik is its uniquely recurring pattern (motifs), it's natural to consider it as a key to classification. To be more specific, recognition of Batik's motifs has been considered as one of the most successful technique in Batik classification especially using Scale-Invariant Feature Transform (SIFT) [1] [2] and Speeded up robust features (SURF) [3]. Classifications using other features such as color and contrast are showing potentials but need to be researched further [4].

Deep learning based models have outperformed state-of-the-art methods in many domains including image classification and object recognition [5]. One of the deep learning models, convolutional neural network (convnet) [6], is currently considered as the state-of-the-art of image classification model as it was used as the base structure by ILSVRC-2014 top achievers [7]. Therefore convnet may also be used to improve result on other image classification problems such as Batik classification.

In this paper, a deep learning model is proposed as a better alternative to classify most common Batif motifs: Parang, Lereng, Kawung, Ceplok and Nitik. VGG16 deep convnet pretrained model [7] is used as first part of the model to automatically extract features from images and

a fully-connected neural network is trained and used as classifier. The proposed model is compared with two SVM classifiers with different hand-crafted feature extractor, SIFT and SURF.

2. Related Works

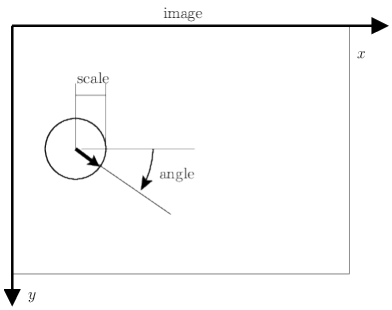

Fig. 1. SIFT Keypoint

Fig. 2. SIFT keypoints in Batik Parang

Recent researches in Batik classification can be divided into two groups: (1) Researches on classification using handcrafted features (eg. SIFT and SURF), and (2) researches on classification using automatically extracted features using deep learning.

2.1. Classification using Handcrafted Features

Since Batik classification has been researched for quite some time, current available methods are robust enough to noise addition, compression, and retouching of the input images. However most of them are still having difficulties with variance in transformations which involve either translation, rotation, scaling or combinations of them [2]. Recent improvements on Batik classification were motivated by the emergence of Scale-Invariant Feature Transform (SIFT) [8] and Speeded up robust features (SURF) [9]. Both of these keypoint-based feature extraction methods are proposed to solve the transformation invariance dilemma.

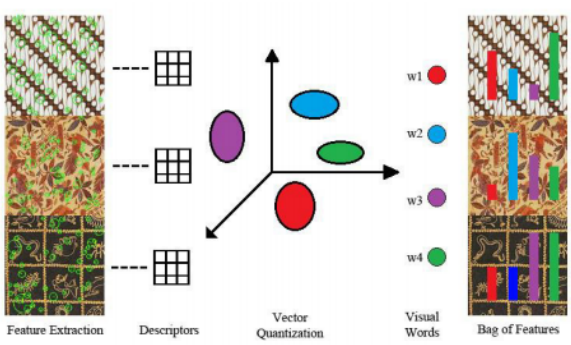

Fig. 3. SIFT for building bag of words visual vocabularies

SIFT keypoint is a circular image region with an orientation which can be obtained by detecting extrema of Difference of Gaussian (DoG) pyramid [8]. It's defined by four parameters: center coordinates x and y, scale and its orientation (an angle expressed in radians) as shown in Figure 1. An image, for example Batik image, may contains multiple keypoints as shown in Figure 1. In order to be efficiently and effectively used as a feature for classification, the keypoint need to be represented as SIFT descriptor. By definition it is a 3-dimensional spatial histogram of the image gradients characterizing a SIFT keypoint.

Recent research [2] proved that using SIFT descriptors to calculate similarity between Batik images can give 91.53% accuracy. Voting Hough Transform was applied to the descriptors to eliminate mismatched keypoint candidates. This research suggested that the original SIFT descriptor matching shouldn't be directly used to calculate similarity of Batik images due to many numbers of mismatched keypoints. The method used in this research is described by diagram in Figure 4.

Another research [1] proposed a classification method using support vector machine (SVM) fed by bag of words (BOF) features extracted using SIFT descriptors as described by Figure 5. In this research, SIFT descriptors also weren't used directly as features for SVM but were clustered using k-means vector quantization algorithm to build vocabularies. These visual vocabularies then used to describe each images and fed to SVM classifier. The experiment results showed very good average accuracy of 97.67% for normal images, 95.47% for rotated images and 79% for scaled images. Besides that SIFT and bag of words made a good feature extractor, this research also concluded that further works need to handle scaled Batik image cases.

An earlier research [3] proved that SURF can extract transformation invariant features faster than SIFT for classification of Songket, another Indonesian traditional fabric with motifs just like Batik. Unlike the others, this research used SIFT and SURF features directly to compute the matching scores between Songket images. The scores are calculated by (1) the number of matched keypoints and (2) the average total distance of the n-nearest keypoints as shown in Figure 5. The result of experiments showed that the

matching accuracy with SIFT features was 92-100% and 65-97% with SURF. With SURF features, the accuracy dropped quite significant if salt and pepper noises were added while SIFT was more stable. Apparently, this one wasn't paying much attention to transformation variance as it didn't apply transformation noise as in other research [1].

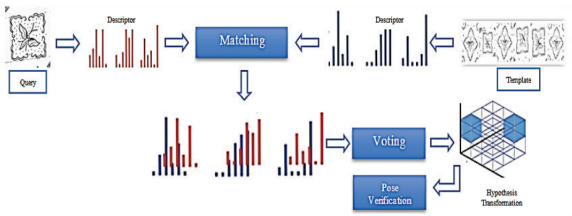

Fig. 4. SIFT with Hough voting method for Batik classification

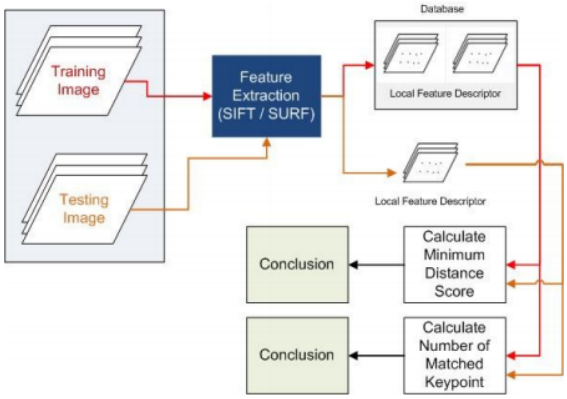

Fig. 5. SIFT vs SURF method for Songket classification

2.2. Classification using Deep Learning

Deep learning is a multilayer representation learning in artificial neural network [5]. While representation learning itself is a method in machine learning to automatically extract/learn representation (features) from raw data. The representation of the raw data then can be used for recognition or classification task. Some fundamental deep learning architectures for instances are convolutional neural network (ConvNet), deep belief network (DBN), autoencoder (AE) and recurrent neural network (RNN). Despite of being an old idea, it was recently emerged due to the several factors: (1) discovery of new techniques (eg. pretraining & dropout) and new activation functions (eg. ReLU), (2) enormous supply of data (big data), and (3) rapid improvement in computational hardware, especially GPU.

Although not yet many, the advent of deep learning also motivated a research on Batik classification using convolutional stacked autoencoder [10]. This research proposed the usage of convolutional transformations to reduce the input nodes of stacked autoencoder. The experiment showed that this deep architecture was able to achieve 81,73% accuracy by using small patches of Batik for training. When noises were added its accuracy dropped to 49% for gaussian noises, 61% for rotations, 70% for scalings and 75% for illumination noises. Another research have shown that deep architecture such as convolutional neural network should be able to outperform handcrafted features such as SIFT [11]. Therefore further research on Batik classification using deep learning architectures is encouraged.

3. Methodology

We propose a deep convolutional neural network composed by a pre-trained VGG16 (without its top layer) as automatic feature extractor and a fully-connected feed-forward neural network as classifier. The method of using pre-trained deep network as part of another neural network to solve different (but related) task can be considered as transfer learning or self-taught learning [12].

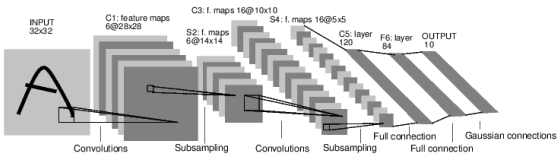

Fig. 6. LeNet5 convolutional network

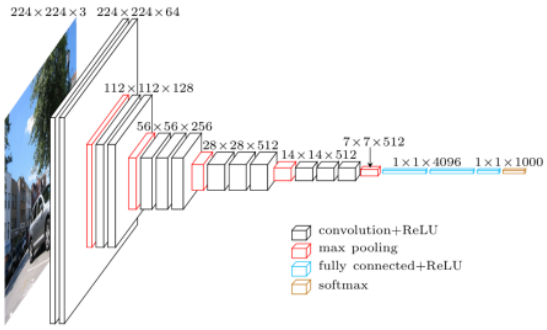

Fig. 7. VGG16 deep convolutional network model of Visual Geometry Group, Oxford

3.1. Convolutional Neural Network

Convolutional network is a special kind of neural network optimized to learn representation of an image [5]. It introduces 2 new types of hidden layers: convolutional and subsampling/pooling layers. Each layer in convnet connects neurons (pixels) from their input layer in form of local receptives (square patches) through a shared weights to a feature map [6]. On top of a set of convolutional and pooling layers, some fully-connected layers are added as classifier as described by Figure 6.

Our proposed model uses 5 set of convolutional and pooling layers using rectified unit (ReLu) activation function following example of VGG16 as shown in Figure 7. The differences are our model uses 2 fully-connected hyperbolic tangent (tanh) (Equation 1) activated layers as classifier (instead of ReLu) and a SoftMax (Equation 2) layer as an output. We also uses Dropout regularization after each tanh fully-connected layers to avoid overfitting by randomly drop/turn off (set value to zero) hidden nodes (Equation 3) [13].

$$y_i = \frac{2}{1 + e^{-2x_i}} - 1 \tag{1}$$

$$y_i = \frac{e^{x_i}}{\sum_{k=1}^{K} e^{x_k}} \quad \text{, for i=1..K} \tag{2}$$

$$r_j^x \sim Bernoulli(p), \quad \tilde{y}_i = r_i * y_i \tag{3}$$

3.2. Transfer Learning

Deep neural networks usually requires a lot of training data in order to learn the representation of the data. In case there is not enough training data, there are several techniques to help neural networks model learns data representation using small training data. One of the technique is transferring knowledge of other pre-trained neural network model to our model. This technique is known as transfer learning or self-taught learning [12].

Our proposed model uses transferred knowledge (layer weights) from pre-trained VGG16 model provided by deep learning framework Keras [1] which was pre-trained using 1,000,000 images dataset from ImageNet. We use VGG16 bottom layers weights (excluding the fully-connected layers) to initialize our model weights and train the classifier part only. This method allows us to shorten the time needed to train our model. Moreover, training the model using GPU improve the speed even more.

To improve comprehension and reproducibility, we publish our model Python code in public online code repository [2]. We also use opensource Theano-backed Keras as deep learning framework and Scikit-Learn[3] as model evaluation framework to improve reusability.

4. Experiments and Results

In order to measure the performance of our model, we trained our model and compared it with SIFT and SURF based models.

1. https://keras.io/applications/#vgg16
2. https://github.com/yohanesgultom/deep-learning-batik-classification
3. http://scikit-learn.org/

4.1. Experiments

The dataset was used in this research is a Batik dataset compiled by Machine Learning and Computer Vision (MLCV) Lab, Faculty of Computer Science, University of Indonesia. This dataset consists of 603 Batik photos ($\pm$ 78.3 MB) gathered from various sources thus having different size, quality and view angle.

We also tried to tune the dataset by deleting some duplicated (similar photos under a class) and conflicting photos (same photos exist under different classes). The tuned dataset has 523 photos left ($\pm$ 68.7 MB) and considered as separate dataset in this experiment.

All classifiers in this experiment were trained using 553 Batik photos (around 52-169 per class) and tested using 50 photos (10 classes per class) from regular dataset (9:1 ratio). The classifiers were also trained using 476 photos and tested with 47 photos from tuned dataset in order to observe the difference.

Our neural network classifier was trained for 50 epoch using Cross Entropy function to calculate loss (Equation 4) and optimized using Stochastic Gradient Descent (SGD) (Equation 5) to update weights.

$$V(f(\vec{x}), t) = -t \ln(f(\vec{x})) - (1 - t) \ln(1 - f(\vec{x})) \tag{4}$$

$$w := w - \eta \nabla Q_i(w) + \alpha \Delta w \tag{5}$$

The SIFT and SURF models were trained in similar manner using similar methods described in related research [1] and also illustrated in Figure 3:

1) Image descriptors were extracted according to their feature extractor (SIFT or SURF)
2) Descriptors were clustered to 5 clusters using K-Means to get visual vocabularies for Bag of Words (BoW)
3) Those 5 visual vocabularies then used to compute BoW features from SIFT/SURF image descriptors
4) Finally a multi-class SVM classifier were trained using the BoW features

All experiments were conducted using Intel Core i7-5960X CPU, 32 GB RAM, NVIDIA GTX 1080 8GB GPU, 240GB SSD, Debian 8 OS. The proposed model ran on GPU while SIFT/SURF SVM models ran on CPU as there were yet mature GPU implementation of the algorithms.

4.2. Results

As shown in Figure 8, the proposed model outperformed SIFT and SURF based classifier by 24% and 34% respectively when using regular dataset. Moreover by using tuned dataset, our model outperformed other models by larger margin 34% and 49%.

Parallel execution of the proposed model in GPU made the total processing time (feature extraction, training and testing) much shorter than other models which ran on single CPU. As described in Figure 9, our model only requires 10-12% of SIFT/SURF processing time.

TABLE 1
Experiment results show that the proposed model outperformed SIFT and SURF based models in term of accuracy and processing time

Method	Accuracy	Accuracy (Tuned Dataset)	Time	Time (Tuned Dataset)
SIFT + Bag of Words + SVM	0.22	0.25	554	492
SURF + Bag of Words + SVM	0.32	0.40	696	616
VGG16 + Neural Network	0.56	0.74	72	64

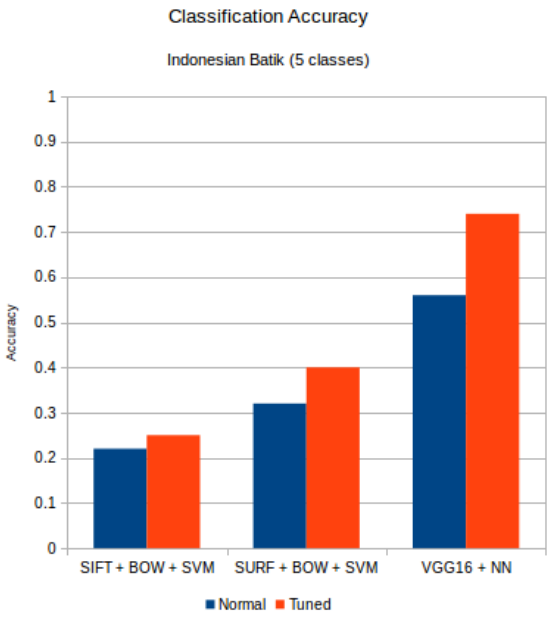

Fig. 8. Models accuracy comparison

5. Conclusion and Future Works

As shown by the experiments, our model, which is based on deep convolutional network, outperformed SIFT and SURF based models in term of accuracy as well as processing time. This confirms that automatic feature extraction using pre-trained convolutional are able to handle transformation invariant features such as Batik motifs better than SIFT and SURF as also concluded by related research [11].

Moreover, deep neural networks is more scalable as its training computation is composed of matrix multiplication which can be easily parallelized on GPU. Supported by keep-growing GPU hardwares, deep neural networks can always be improved in term of speed and capacity.

We also found out that current Batik dataset has a lot of room for improvements:

1) Multi-labeled data. As majority of the data are mixed-motif Batik, the dataset must provide more than one labels for each applicable sample.
2) Clearly distinguished samples between classes. For instance, Parang and Lereng motifs data often overlaps each other. This condition often confuses classifier during training and causes less accurate generalization.
3) Homogeneous quality of data. Due to the various sources of data, the quality (resolution, size, view angle) of the data are also various. Removing low quality data and preprocessing high quality ones may produce homogeneous data and improve classifier training process.

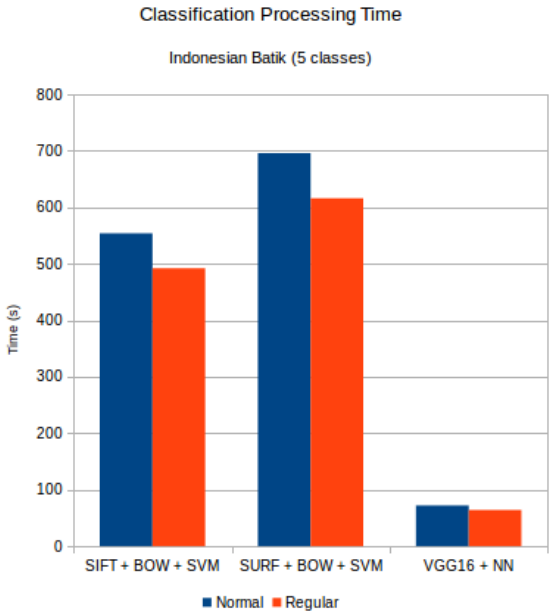

Fig. 9. Execution (from preprocessing to classification) time comparison

References

[1] R. Azhar, D. Tuwohingide, D. Kamudi, N. Suciati *et al.*, "Batik image classification using sift feature extraction, bag of features and support vector machine," *Procedia Computer Science*, vol. 72, pp. 24–30, 2015.

[2] I. Nurhaida, A. Noviyanto, R. Manurung, and A. M. Arymurthy, "Automatic indonesian's batik pattern recognition using sift approach," *Procedia Computer Science*, vol. 59, pp. 567–576, 2015.

[3] D. Willy, A. Noviyanto, and A. M. Arymurthy, "Evaluation of sift and surf features in the songket recognition," in *Advanced Computer Science and Information Systems (ICACSIS), 2013 International Conference on*. IEEE, 2013, pp. 393–396.

[4] V. S. Moertini and B. Sitohang, "Algorithms of clustering and classifying batik images based on color, contrast and motif," *Journal of Engineering and Technological Sciences*, vol. 37, no. 2, pp. 141–160, 2005.

[5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[6] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[7] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[8] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[9] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *European conference on computer vision*. Springer, 2006, pp. 404–417.

[10] R. A. Menzata, "Sistem perolehan citra berbasis konten dan klasifikasi citra batik dengan convolutional stacked autoencoder," Universitas Indonesia, 2014.

[11] P. Fischer, A. Dosovitskiy, and T. Brox, "Descriptor matching with convolutional neural networks: a comparison to sift," *arXiv preprint arXiv:1405.5769*, 2014.

[12] R. Raina, A. Battle, H. Lee, B. Packer, and A. Y. Ng, "Self-taught learning: transfer learning from unlabeled data," in *Proceedings of the 24th international conference on Machine learning*. ACM, 2007, pp. 759–766.

[13] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting." *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.