Resume Paper

[redacted]

Research Area : Computer Vision

Paper Title,Writer, Name of Journal/Conference and Publised Year :

Implementation of Data Augmentation Using Convolutional Neural Network for Batik Classification, Chan Uswatun Khasanah; Ema Utami; Suwanto Raharjo, 2020 8th International Conference on Cyber and IT Service Management (CITSM), 2020

Research Problem :

How the data Augmentations affect the classification result for batik dataset ?

Proposed Method :

A. Uswatun Khasanan et al proposed 8 data augmentations using CNN to classify batik

Table I Proposed data Augmentation Methods for batik

Augmentation Method	Result	Explanation (From me)
Brightness		Brightness was used for making image much more bright, in case there are some images were dark. When image too dark, it prevents the neural network to learn the essential feature from the corresponding image
Channel Shift		Channel Shifting was used to shift each pixel values randomly chosen from -N up to **N** pixels (but it could not below 0 or above 255) the purpose of this shifting is to generalize more better, since it

		simulate different lightning conditions or environments
Random erasing		Random Erasing was used for replacing some proportion of pixels from image with some uniform distribution. It enables the neural network to recognize objects even when they are partially occluded or missing from the image. One example when the batik input was ripped apart in the future.
Zoom		Zoom was used for taking care a problem when dealing with objects at different scales or resolutions. When this is applied through the whole neural network. The neural network will able to recognize the objects from the different resolutions, in which making the model generalization more well
Rotation		Rotation was used for taking care a problem with a different angle in the future input. Thus, it making the model much more better to generalize future inputs
Height Shift		Height shift was used for simulate variations in camera angle or subject movement, but the change only from its height. Whether the augmented image moved the image up or down.
Width Shift		Same as height shift, but the changes made from its own width . Whether the augmented image moved the image to the left left or right.
Shear		Shear was used to create a new perspective effect that simulates the appreance of the object from a different angle, so then the model was exposed to different range of viewing angles which

		can help improve its ability to generalize well to new data

The value for each data augmentations, can be shown at table II below

Table II Data Augmentation Value

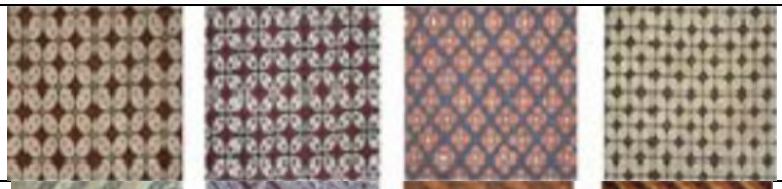
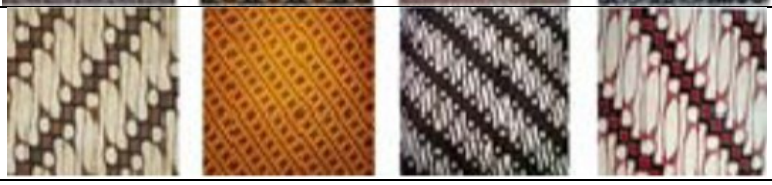
Data Augmentation	Value	Explanation(From me)
Width Shift	0.2	It means that allowed maximum horizontal shift will randomly chosen between -45 to 45 pixels from 224 width size.
Height shift	0.2	It means that allowed maximum vertical shift will randomly chosen between -45 to 45 pixels from 224 height size.
Shear	0.2	Image will randomly sheared by maximum of 20% of its width or its height
Zoom	0.2	Image will randomly zoomed in or out by up to 20% of their original size. If the original size has size 224 x 224, a zoom value of 0.2 would allow for a maximum zoom of 45 pixels in either direction from the original size. So it could be (224+45) x (224+ 45) or (224-23) x (224-23)
Random erasing	P=0.5, s_l=0.02, s_h=0.4, r_l=0.3	This was used to delete some portion of an image by the given parameters.
Brightness	[0.2,1.0]	Range for brightness

Channel shift	100	The shifthing value for each channel is -100 to 100 pixels. It is applied for each pixels in an image.
		
		Some example :
		Before channel shifting :
		`[[[200, 100, 50], [150, 75, 25], [100, 50, 0]],`
		`[[255, 255, 255], [0, 0, 0], [100, 100, 100]],`
		`[[50, 150, 200], [25, 75, 150], [0, 50, 100]]]`
		
		After channel Shifting :

		[[[160, 150, 100], [200, 125, 75], [75, 75, 50]], [[255, 255, 205], [0, 50, 50], [155, 105, 155]], [[25, 200, 150], [50, 100, 200], [150, 150, 100]]]

From table II we know that the proposed data augmentations are made for making the model generalize well in later study case of batik recognition. Here are the batik types that have been used by the researcher . Table III shows the batik types and it name

Table III Batik Types

Batik types	Type name
	Ceplok
	Kawung
	Lereng
	Nitik
	Parang

Dataset was collected from the google images which consists of 50 pictures/class. The researcher also expand the dataset to 100 by using slicing method . Table IV shows the slicing method.

Table IV Slicing Method

Batik Image	Slicing Method Result
	

The collected data will then divided by training and testing data. Figure 1 shows the division of dataset

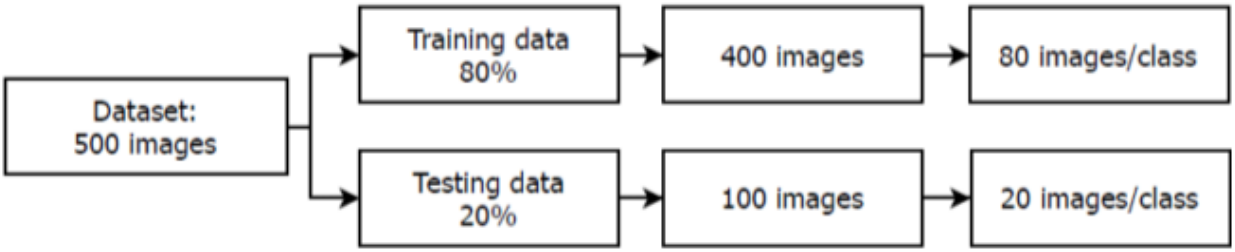


Figure 1 Division of dataset

From here we know that, the author does not used train-val-test method. Therefore the result from the testing dataset could not represent the whole batik motifs just yet.

B. VGG-16 as pretrained method to classify batik

The author was used the VGG-16 by fine tuning the model. So then x layer will learn additional features from the input. In this research block5 convolutional layer was used. Table V shows the paramaters that were used by the model and figure … show the visualization of trainable layers

Table V Model Parameters

Parameter	Value
Input_image	(224,224,3)
Optimizer	RMSprop
learning_rate	1e-4 (0.0004)
Loss	Categorical CrossEntropy

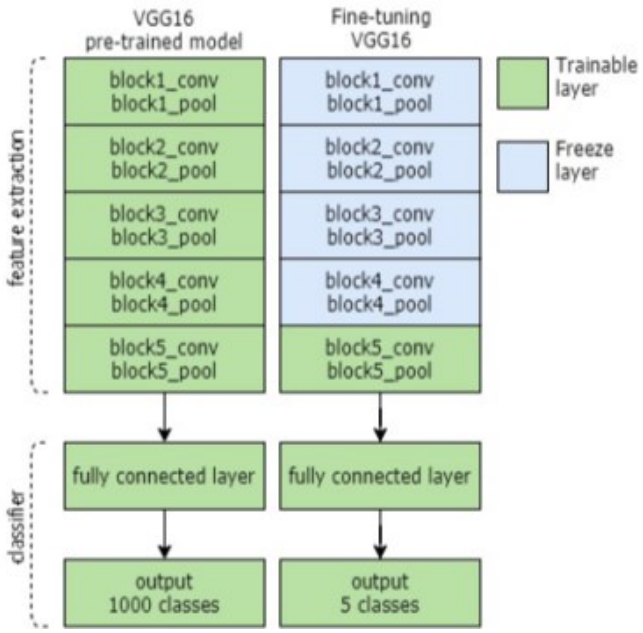


Figure 2 trainable layers

Experiment Result :

The experiment was done by dividing the augmentations method into some group. Table VI shows the group and table VII shows the result for each combined augmentations

Table VI Group Augmentation

Group	Data Augmentation	Accuracy With data Augmentation	Accuracy Without data Augmentation
I	Width shift	97.92	95.83
	Zoom	97.92	
	Rotation	96.88	
	Shear	96.88	
II	Height Shift	95.83	95.83
	Random Erasing	95.83	
	Brightness	88.54	
	Channel Shift	83.33	

From the table VI above we can see each augmentation resulted a different accuracy. The highest is width Shift and the lowest was Channel shift.

The author does not stated, whetheher the dataset was shuffled or not for each data augmentation. Therefore I could not say whether the resulted accuracy was accurate or not. But here I provide my analysis.

- Group 1 having a great improvement compared with group 2 and without data augmentation. This can happen because
    - Could be the distribution of the dataset
    - The consistency of the resulted augmentation image
    - Additional generalization from each augmentation that have a big impact compared with another augmentation methods, such as
        - Shear, Zoom, Rotation, width shift → Both have a similarity in which making the model much more better at learning different angles / scales/ resoultion.
- Without data augmentation
    - Lack of additional generalization informations, therefore in some circumstances it does not work well at predicting some input images.
- Group 2 Lower accuracy compared without data augmentation
    - This can be happen because it is too complex for the model to learn
        - Brightness -> The author does not stated wheter the brightness is applied to all dataset or not. But if so, then the input image that has a great brightness should not been increased the brightness rather it should decreased the brightness. If all is randomly increased or decreased. Therefore it could make the learning data more bias, in which leads to some model having a tendency into some brightness range
        - Channel Shift → Same as Brightness

Then the author combined each Group data augmentation into one. Table … shows the resulted accuracy
The author stated grouping was based on the highest accuracy.

TABLE VII Combined Group Augmentation

Model	Accuracy
Without data augmentation	95.83
Group II	96.88
Group I + II	97.92
Group I	98.96

From here, we can say that. In this study, the data augmentation methods from group 1 much more reliable compared with group 2 and combined. Therefore, what we can

say that the angles and resolutions are matters for model to recognize the batik. But we cannot say that data augmentation methods from group 2 does not reliable. Since there is no indepth analysis within each data augmentation methods.

Paper Contribution :

- It provides insight for determining the data augmentation methods for batik study case. From what I get they are
    - Rotation
    - Zoom
    - Horizontal Shift
    - Vertical Shift
    - Shear
    - Zoom
    - Random erasing
- The data augmentation methods for batik provide a high overview which method is suitable for later study case
    - If you need client to recognize the input image from randomly resolutions / angles then you should use
        - Zoom
        - Rotation
        - Shear

      This is helpful since it can make the time efficient and model more simpler

Paper Drawback/Limitation :

- It is only used the Accuracy metrics
- The authors not doing some kind of error analysis
- The dataset having same resolutions in the beginning
- The complex batik shape haven’t been touched by the authors
- There’s no explanation why choosing the 8 data augmentation methods and using model VGG-16

Paper Title,Writer, Name of Journal/Conference and Publised Year :
VGG16 in Batik Classification based on Random Forest, Dewa Made Sri Arsa; Anak Agung Ngurah Hary Susila, 2019 International Conference on Information Management and Technology (ICIMTech), 2019

Research Problem :

How to achieve a better model generalization by using VGG-16 as feature extractor and Random forest as the classifier ?

Proposed Method :
Here's the overview of the proposed method

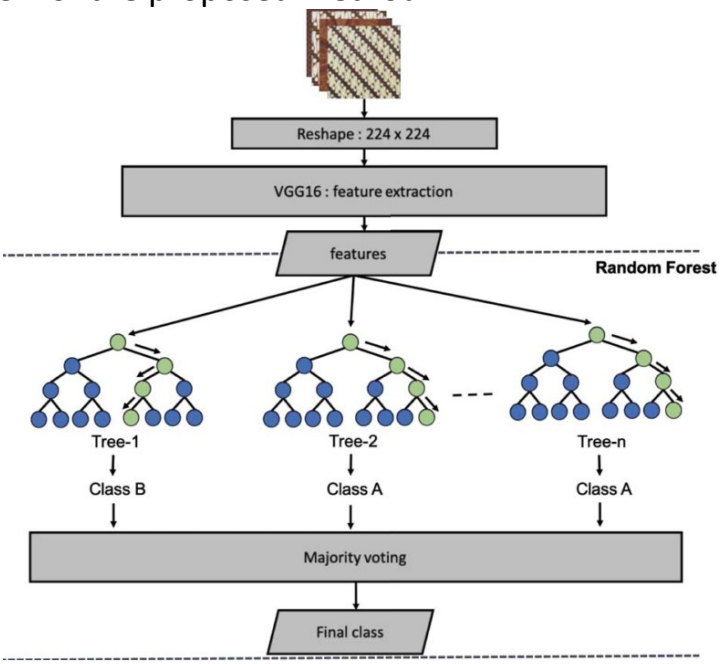


Figure 3 Proposed Method

A. Deeplearning – VGG 16 Model as Feature extractor
   Since VGG 16 is used as the feature extractor therefore all the layers will be freezed, the purpose of this step is to just extract the essential feature that are passed from layer to layer and then the outputted features will be feed into random forests.
B. Random Forest as the classifier

Random forest was used as the classifier because each class only have 6 images, which in this case very tiny to be trained on VGG-16. Therefore, Random forest model was chose. To provide a reliable output class, majority voting from the random forests were done in this research. Gini Index Minima was used as the measure of impurity (Kemurnian).

- Analogy Impurity → Say you have different types of fruits (oranges, apples, and banana). The **Gini index measures the impurity** of the Fruit basket in terms of the variety of fruits. IF the basket contains only apples, the the Gini index is 0, perfect purity. On the other hand, if the basket contains an equal number of apples, oranges, and bananas, then the gini index is 0.5, equal proportions. Range value 0 - 1

The authors did not state other reasons why using random forest. In my perspective Knn should be used too as comparison, because of how the algorithm learn the pixels through neighboring. Since each batik has kinds of repeating pattern.

C. Experiment Setup

Figure 4 shows the dataset that was used by the author

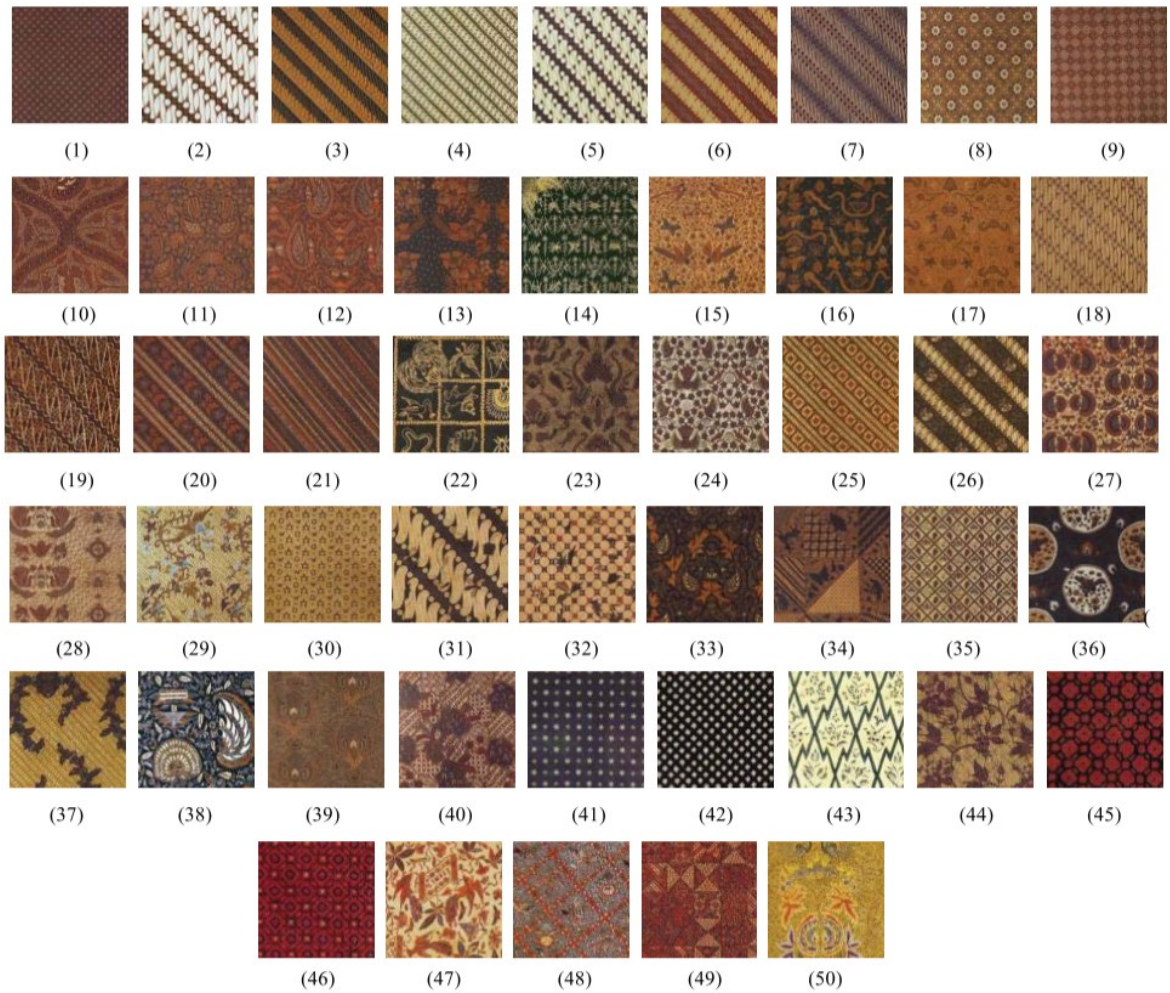


Figure 4 Dataset

The dataset consists of 50 batik images, in which 6 images per class, but the author did not tell what are their class names. The dataset was splitted into 80:20 proportion and was tested 20 times by the proposed model, table VIII shows the parameters that were used for diagnosing the model

Table VIII Classification Metrics

Parameter Name	Explanation
Precision	Measures the proportions of **true positive** among all **positive predicitions (How often hunter finds gold coins out of all the objects he digs up)**
Recall	Measures the proportion of the **true positive** among all **actual**

	positive example (How often hunter finds gold coins out of all the gold coins that are actually in the field)
Accuracy	Measure the rate
F1-Score	Combine f1 precision and recall into a signle number (**Hunter overall efficiency** in finding Gold coins while **avoiding** digging up useless objects)

Experiment Result :

TABLE IX Experiment Result

Method	Precision	Recall	F1-Score	Accuracy
Proposed Method	97.34±2.76	97.40 ± 2.50	97.03 ± 2.81	97.58±2.32
VGG-16	88.59±6.64	89.31 ± 6.64	88.48 ± 6.56	88.42 ± 6.63

From the table above, we can see that the proposed model have a great perfomance compared the VGG-16. This can be happen because of the random forest classifiers. By using gini index minima, the proposed model know in which what is the best random forest that can minimize the gini index, and also there is a majority voting, that made the output much more reliable. Unfortunately the author does not give an explanation regarding the majority voting. Therefore, we could not answer wheter or not its 50:50 / 70:30 / 60:40. Since those are important to interpret the proposed model.

Paper Contribution :

- It gave a new perspective to tackled a small dataset by using pretrained model and then use the features for simple machine learning model
- It can be use as the baseline for creating a best model to classify batik, in order to create SOTA for batik model. That can be used for Style transfer later.

Paper Drawback/Limitation :

- It does not provide some error analysis
- It is only used by a small dataset
- Classes does not defined clearly
- It does not explain the majority voting output in details