Project Report - MLBatikNet: Deep Learning Based Multilabel Batik Impression Classification

Hani Ramadhan - 201793254
Student of Big Data Department
Data Science Laboratory
Pusan National University
hani042@pusan.ac.kr

Yoga Yustiawan - 201783629
Student of Big Data Department
Data Science Laboratory
Pusan National University
yoga017@pusan.ac.kr

***Abstract*—Batik is Indonesian traditional heritage applied as patterns in some fashion items. Its patterns like any fashion item can express several impressions when worn. Thus, this problem falls into multilabel classification to predict single batik image's impressions. However, as multilabel dataset, annotated batik images are still scarce yet imbalanced. Recently, deep learning techniques indicate great technological advancement on classifying multilabel datasets, such as ImageNet. In this study, the application of transfer learning on some pretrained models on batik dataset is going to be studied. Chosen models in this study are Residual Network based model in Tensorflow such as InceptionNet and MobileNet. Experimental results proved that InceptionNet had improved the validation performance to 0.906 accuracy and 0.940 average precision better than non-deep learning techniques, yet MobileNet performed worse than non-deep learning techniques and InceptionNet in this study. Data augmentation also suggests good enhancement to both InceptionNet and MobileNet performance.**

***Keywords*—** Multilabel, batik, ResNet, Tensorflow, data augmentation

I. Introduction

Batik has been recognized as an element of global cultural heritage produced by Indonesians. The wide diversity of patterns reflects a variety of influences, one example such as human characters or so called impressions. Ordinary people even most common Indonesians have the affliction to understand what kind of characters that are represented in a pattern of Batik. Utilizing the deep learning methods for helping people to understand the pattern impression of Batik is challenging computer vision task. Since each pattern is not only belong to one impression, but can reflect two or more impressions.

Multi-label classification tasks which are a well-known challenging issue due to the complexity of data are sometimes also intricate to define information concerning the classes that are not mutually exclusive. In medical field, [1] exercises Deep Neural Networks (DNN) to identify patterns in patient data for intelligent health risk prediction. [2] identify the movie genre using Deep Convolutional Neural Networks (ConvNets). One interesting previous work to identify the pattern of Batik has been developed by exercising ensemble of multi-label classifier [3]. However, this work does not implement deep learning techniques to address the issue.

In this work, the Convolutional Neural Networks (CNN) as one prominent deep learning technique are aimed to be exercised to identify the pattern of Batik that reflects impressions. This is distinguishable issue that very limited few works that has tried to address this issue especially for multi-label classification. In order to address the above-mentioned issue, this work proposes a CNN architecture, benefiting from transfer learning of pretrained models, that is particularly devised to identify multi-label classification of Batik called MLBatikNet.

II. Related Works

Our proposed work exploits several studies in the past. The base of our neural network architecture make use of Inception Network (InceptionNet) [4] and Mobile Network (MobileNet) [5]. To minimize the computing effort, we also incorporate transfer learning [6]. The two approaches then will be modified to be adapted our multilabel batik dataset and compared to previous works on multi-label classification [3].

A. Transfer Learning

Simply said, transfer learning is one method to benefit pretrained model to perform different learning task [6]. For example, a model already trained to classify the detection of the cat in some image dataset. This model can be used to detect dogs in other image dataset with several parameter adjustments and relearning. However, the cost of the adjustments and relearning is not much as training model from zero.

The interesting part of this study is that there is currently minimum number of study of batik classification using deep learning. Thus, there is no accessible model to be used as pretrained model that corresponds to batik image features. In this study, we will discuss the performance of object detection pretrained model such as ImageNet to learn scarce multi-label batik dataset compared to training using multi-label batik dataset from zero.

B. Inception Network

Inception Network (InceptionNet) is an architecture of deep learning that employs locally optimal convolutional building block and repeat it. In some manners, it behaves like ensembles of relatively shallow networks [7]. The principle of the inception network is to consider how an optimal local sparse structure of a convolutional vision network can be approximated and covered by readily available dense components [4]. The performance of InceptionNet had the best score in ImageNet object detection and classification problem on 2015. However, it is not being tested on smaller dataset and different task.

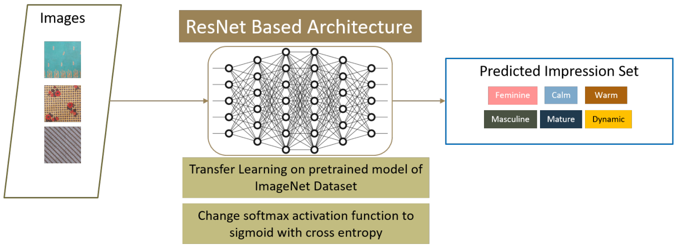

Fig. 1. Our proposed change from existing InceptionNet-MobileNet Architecture

In this study, the performance will be discussed according to multi-label batik classification task.

C. Mobile Network

Mobile Network (MobileNet) is an architecture based on depthwise separable convolutions which is a form of factorized convolutions which factorize a standard convolution and 1 x 1 convolution called a pointwise convolution [5]. The MobileNet chooses 3 x 3 depthwise separable convolutions and puts nearly all of the computation into dense 1 x 1 convolutions. By using a small, low-latency, low-power model, MobileNet can meet the resource constraint of a variety of uses cases. The performance of MobileNet on ImageNet object detection was relatively satisfying and exceeded the popular models such as VGG16 on 2017. However, since it is not being tested in smaller dataset and different task, this study will discuss the performance of MobileNet for multi-label batik classification task.

III. Methods

In MLBatikNet, we use pretrained models to overcome the work to build the model from the beginning. This work is limited to usage of transfer learning to perform the classification. The proposed pretrained models are InceptionNet [4] and MobileNet [5] for image classification and taken from official [8] site.
To adapt our dataset, the parameter of the last layer (before the output layer) of the pretrained models is defined to six output layers. For the activation function, instead of softmax or ReLU, which used for multi-class classification problem, we will use sigmoid with cross entropy [1]. Our proposed change is depicted in Figure 1.

IV. Experimental Setup

A. Dataset

The provided dataset in this project are batik cloth images which consists of 102 labeled multilabel images and 11 unlabeled image. The dimension of single image is 500 x 500 pixel. The labels of the images are defined by set of six distinct impressions: mature, calm, warm, dynamic, masculine, feminine. The distribution of labelset, which is also imbalanced, can be seen in Table I. Some of the example images can be seen in Figure 2

TABLE I
Distribution of Multilabel Batik Cloth Image Labelset

Impression labelset	Total
feminine	51
dynamic, masculine	23
mature, calm	13
mature, calm, warm	7
mature, calm, feminine	4
mature, calm, warm, feminine	2
mature, calm, warm, masculine	1
Total	102

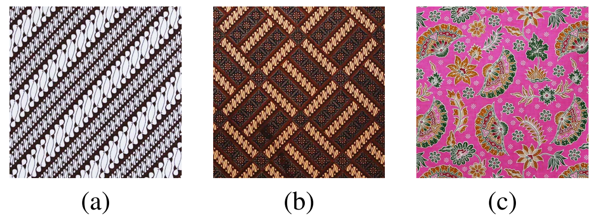

Fig. 2. The sample batik images and their labelset: (a) dynamic, masculine, (b) mature, calm, (c) feminine

Regarding the limited and imbalanced dataset, which is not in deep learning's favor, several techniques of data augmentation are configured. The data augmentation is led to slightly smaller images but still capturing the prominent patterns of batik images. The data augmentation process is led by random crops on the original image, several rotated images, and horizontal-vertical flipped images.

B. Evaluation Metric

Multilabel classification performance cannot be measured using accuracy like single label classification, thus in this project the performance is measured by two different metrics: Hamming Loss and Average Precision [9], [10].

Hamming Loss (1) of measures the misclassification of a single label for each instance y_i, where each available label of an instance is represented by y_{i_j} and the predicted label is $\hat{y}_{i_j}$. It indicates better performance when it is closer to 0 and worse when closer to 1. The Hamming Loss is equivalent to 1 subtracted by multilabel accuracy (2). Since the code of InceptionNet [11] and MobileNet [12] already provide implementation of the multilabel accuracy, we will report the performance of the approaches using multilabel accuracy instead of hamming loss.

$$hloss = \frac{1}{m}\sum_{i=1}^{m} |\hat{y}_{i_j} \neq y_{i_j}, j = 1, ..., m| \tag{1}$$

$$ml\_acc = \frac{1}{m}\sum_{i=1}^{m} |\hat{y}_{i_j} = y_{i_j}, j = 1, ..., m| \tag{2}$$

Meanwhile, the average precision (3) measures the performance of the classification based on the rank of all true labels y_i in each instance. Each predicted label of an instance $\hat{y}_{i_j}$ actually had 0 to 1 probability value, claimed to have the label as they are rounded following a threshold. Thus, all of these labels of each instance has certain order of rank from highest to the lowest defined by the ranking function $r_i(label)$. If all

TABLE II
IMAGE TRANSFORMATION FOR DATA AUGMENTATION

Transformation	Treatment
Flip:	Horizontal, Vertical
Rotation:	90°, 180°, 270°
Focus:	Sharpen, Blur
Brightness and Contrast:	increase, decrease
Original	None

TABLE III
DATASET SIZE FOR INCEPTIONNET AND MOBILENET

Classifier	Non-augmented	Augmented
InceptionNet	102	10200
MobileNet	156	15600

the true labels of the prediction has top ranks, the average precision should be closer to 1.

$$AP = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{|y_i|}\sum_{l\in y_i}\frac{|\{l'|r_i(l') \le r_i(l), l' \in y_i\}|}{r_i(l)} \quad (3)$$

C. System Setup

This work is conducted in the system with Intel Xeon processor E5-2600 v4 @ 1.70 GHz x 16, 64 GB RAM, GeForce GT1050 4 GB and OS Ubuntu 16.04.4 LTS 64 bit.

V. RESULT AND DISCUSSION

A. Preprocessing

As described in the experimental setup, we would like to see whether an augmentation of dataset can affect the performance of deep learning on the batik multilabel dataset. The preprocessing consisted of two steps: image feature transformation and random cropping. The transformation step is done using Batch Processing by [13], while the random cropping resulting 10 smaller images from a single image is done in Python. The list of the transformation included in random cropping is listed in Table II, thus yields 100 times the original instance of augmented data.

Implementation of InceptionNet and MobileNet for multilabel classification have different settings. The codes for MobileNet multilabel classification [12] requires redundant image for different labels. Which means if an image has two labels, two image instances will appear in two separate label association. Meanwhile, the code of InceptionNet [11] multilabel classification can easily capture the association of an image instance to multiple labels by a simple text file. Thus, this leads to different amount of dataset used as input for each classifier which is depicted in Table III.

B. Experimental Results

Both Deep Learning approaches performs on the same hyperparameter configuration: Learning rate 0,01; 1000 training-validation steps; sigmoid with cross entropy optimization function; and batch size of 100. We would not like to see the performance on test as our dataset is very limited. The chart is produced by the TensorBoard visualization on the training summary for each case. The training performance is illustrated by orange lines and the validation performance is illustrated by blue lines.

1) InceptionNet Approach: InceptionNet performance by means of accuracy and average precision is depicted on Figure 3 and Figure 4 for non-augmented and augmented dataset respectively. We can see that when InceptionNet is applied for both scenario, both accuracy and average precision converges nicely and yields great result, although on the non-augmented dataset, the average precision had a hard time to converge in the beginning. However, the result on non-augmented dataset tends to show that the InceptionNet still slightly overfits the data, while the result on augmented data has no sign of heavily overfitting the data.

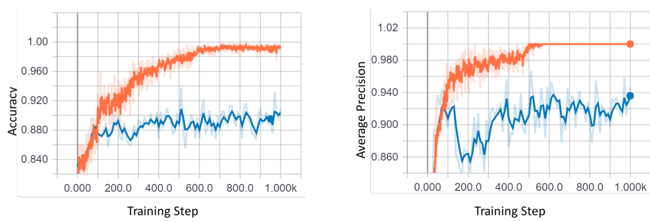

Fig. 3. Performance of InceptionNet on non-augmented dataset

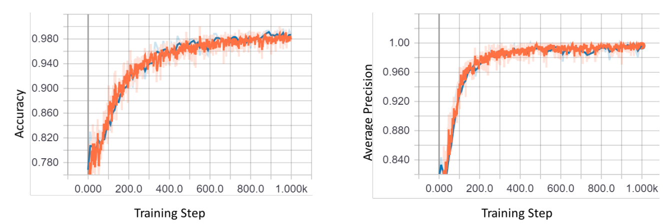

Fig. 4. Performance of InceptionNet on augmented dataset

2) MobileNet Approach: MobileNet, which has lesser parameters to train and lighter computation, visibly has worse performance than InceptionNet for both accuracy and average precision. Its results on non-augmented and augmented dataset are depicted in Figure 5 and Figure 6 respectively. For non-augmented dataset in Figure 5, we see that MobileNet achieved convergence on training set and validation set, although it declined after few steps for the validation. On augmented dataset, although it has better performance than the non-augmented dataset, both training and validation had not converged very well. Even though the training and the validation share similar performance, it is hard to determine whether MobileNet overfitted the data or not.

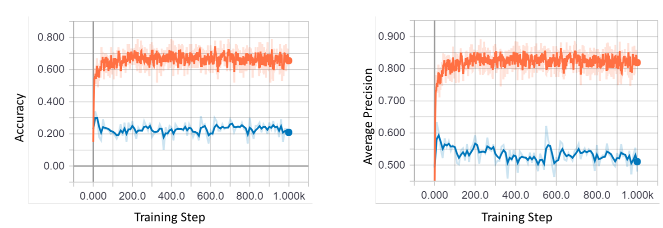

Fig. 5. Performance of MobileNet on non-augmented dataset

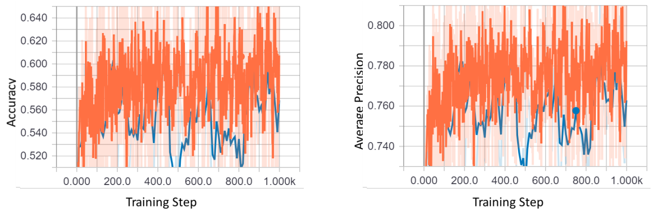

Fig. 6. Performance of MobileNet on augmented dataset

TABLE IV
PERFORMANCE COMPARISON ON NON-DEEP LEARNING APPROACH TO INCEPTIONNET AND MOBILENET

Dataset and Metrics	Methods		
	Ensemble	Inception	Mobile
Non-augmented dataset			
-Train Acc	-	0.989	0.623
-Train AP	-	1.000	0.791
-Val. Acc	0.837	0.906	0.196
-Val. AP	0.866	0.940	0.483
Augmented dataset			
-Train Acc	-	0.982	0.544
-Train AP	-	0.997	0.729
-Val. Acc	-	0.987	0.537
-Val. AP	-	0.996	0.738

C. Overall Comparison

In addition to both deep learning approaches performance comparison, a non-deep learning technique performance using ensemble of multilabel classifiers [3] is also presented as baseline approach. The ensemble is consisted of which consisted of the combination of Ensemble of Classifier Chain (ECC) and Multi-label K-Nearest Neighbor. Although only ensemble's performance on non-augmented validation data was available, we still would like to see how deep learning approach compared to non-deep learning approach.

Table IV shows the comparison of all approach performances. We can see that in all terms of training and validation evaluation MobileNet performed the worst while InceptionNet performed nearly perfect in both non-augmented and augmented dataset. On the other hand, we can see the performance of MobileNet is rather unacceptable since it is significantly lower than the baseline approach. Meanwhile, InceptionNet significantly outperformed the baseline approach. The bad performance of MobileNet on this multilabel batik dataset may be caused by its lesser parameter and lighter computation to InceptionNet. The redundant setting of input for MobileNet may also affect how it performed on imbalanced multilabel dataset. Apart from the comparison of InceptionNet and MobileNet performance, the augmentation of multilabel dataset had greatly improved the evaluation of both methods.

VI. CONCLUSION

In this study, we performed comparison on two different ResNet based deep learning architecture to multilabel batik dataset classification. The two studied approach are InceptionNet and MobileNet. The challenge of this problem resides on the dataset's imbalanced label distribution, thus it can easily misclassify few label on an instance.

The result shows that InceptionNet performed very well on the dataset while MobileNet did not. This is caused by MobileNet's characteristic that has lesser parameters to train and lighter computation, plus a different setting on multilabel input. Data augmentation was proved to improve the performance of both method in the case of imbalanced dataset.

In the future, there is still much possibility to do with the multilabel batik dataset. One of deep learning approaches called Generative Adversarial Network (GAN) is possible to generate a new example based on existing dataset. This is interesting since batik incorporates deep values of Indonesia culture and hard to devise a meaningful design. On the other hand, there is another possibility to perform style transfer which applies some pattern on an image to another image. This is a good opportunity to design a fashion item based on batik without losing its emotional impression.

REFERENCES

[1] A. Maxwell, R. Li, B. Yang, H. Weng, A. Ou, H. Hong, Z. Zhou, P. Gong, and C. Zhang, "Deep learning architectures for multi-label classification of intelligent health risk prediction," *BMC bioinformatics*, vol. 18, no. 14, p. 523, 2017.

[2] J. Wehrmann and R. C. Barros, "Movie genre classification: A multi-label approach based on convolutions through time," *Applied Soft Computing*, vol. 61, pp. 973–982, 2017.

[3] H. Ramadhan, I. Arieshanti, A. Yuniarti, and N. Suciati, "Impression determination of batik image cloth by multilabel ensemble classification using color difference histogram feature extraction," *Kursor*, vol. 7, no. 4, 2014.

[4] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, *et al.*, "Going deeper with convolutions," Cvpr, 2015.

[5] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[6] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[7] A. Veit, M. J. Wilber, and S. J. Belongie, "Residual networks are exponential ensembles of relatively shallow networks," *CoRR*, vol. abs/1605.06431, 2016.

[8] TensorFlow, "Models and examples built with TensorFlow," May 2018. original-date: 2016-02-05T01:15:20Z.

[9] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Mining multi-label data," in *Data mining and knowledge discovery handbook*, pp. 667–685, Springer, 2009.

[10] M.-L. Zhang and Z.-H. Zhou, "Ml-knn: A lazy learning approach to multi-label learning," *Pattern recognition*, vol. 40, no. 7, pp. 2038–2048, 2007.

[11] R. Bartyzal, "Multi-label-Inception-net: Multi-label image classification using pretrained Inception net," May 2018. original-date: 2017-04-02T14:01:52Z.

[12] G. Codelabs, "Tensorflow-for-poets-2 development," May 2018. original-date: 2017-05-09T22:48:44Z.

[13] GIMP, "BIMP. Batch Image Manipulation Plugin. | GIMP Plugin Registry," 2018.