Image Retrieval of Indonesian Batik Clothing Based on Convolutional Neural Network

Mutia Fadhilla
Teknik Informatika
Universitas Islam Riau
Indonesia
tiafadhilla@eng.uir.ac.id

Des Suryani
Teknik Informatika
Universitas Islam Riau
Indonesia
des.suryani@eng.uir.ac.id

Nesi Syafitri
Teknik Informatika
Universitas Islam Riau
Indonesia
nesisyafitri@eng.uir.ac.id

Hendra Gunawan
Teknik Informatika
Universitas Islam Riau
Indonesia
hendra@eng.uir.ac.id

***Abstract*—Indonesian Batik is best-known for unique and distinct pattern. Searching Indonesian Batik clothing images is a challenging problem due to its wide pattern variations. In this paper, proposed image retrieval model of Indonesian Batik clothing image searching based on Convolutional Neural Network (CNN). Autoencoder proposed as CNN model that trained to reconstructed original input batik clothing image. So, the visual features can be extracted from CNN Autoencoder. Based on the experimental results, the proposed method can reach 90.8% in retrieval accuracy, 58.8% in mean average precision, and 61.9% in average recall.**

Keywords—Image Retrieval, Indonesian Batik Clothing, Convolutional Neural Network, Computer Vision

I. INTRODUCTION

Indonesia is known as a country that has cultural diversity. One of the cultural heritages in Indonesia is batik. It is a traditional cloth or fabric which is made using wax-resist dyeing technique. Similar fabric could be found in various countries, but Indonesian Batik is probably the well-known and being designated as a Masterpiece of the Oral and Intangible Heritage of Humanity by UNESCO. Where batik has symbolic meaning patterns that represent the identity of the Indonesian people which expresses their creativity and spirituality. In addition, batik cloth is also widely used as the main material in fashion products in Indonesia, such as clothing that is used for daily activities or certain formal occasions.

Nowadays, batik clothing can be purchased easily through e-commerce. But the customers commonly need to input some keywords related to products and it will be time consuming. Furthermore, it is not easy to represent batik clothing which has wide various styles and patterns using only some keywords or captions. Because of these problems, clothing image retrieval recently has become interesting topic on research.

Clothing image retrieval is a technique that used for searching and retrieving relevant images of clothing product based on image visual features. However, the most challenging part of clothing image retrieval is to extract useful visual features from the image. The main reason is batik clothing image has big variations and uniqueness in attributes such as color, pattern, style, and image viewpoint. Besides that, the clothing images also can be differentiated by fine and small detail probably can be seen in different viewpoint.

The computer vision is now widely used in several fields such as robotics, medicine, fashion, etc. Convolutional Neural Networks (CNN) is one of computer vision model that have been succeeded in several vision-related task in artificial intelligence, such as feature extraction, image classification, and image detection such as proposed in [1], [2]. So far, there are some research that are related to batik pattern retrieval. [3] proposed a method for retrieving batik pattern image based on combination of local and global features that were extracted according to the Zernike moments (ZMs). [4] developed content-based batik retrieval using Maximum Run Length (MRL) from Local Binary Pattern. Other existing methods used CNN-based model for recognize batik pattern. [5], [6] proposed CNN model of batik pattern classification. Meanwhile, [7] proposed supervised and unsupervised CNN model to extract image features then KNN searching for batik retrieval. However, most of the research only discuss about image retrieval of batik pattern fabric, almost no research focused on batik clothing image retrieval. Meanwhile, there are many studies related to fashion image retrieval. Some research proposed cross-domain clothing retrieval model, [8] proposed novel model used feature fusion and quadruplet, [9] used triplet embedding. Meanwhile [10] conducted survey on proposed cross-domain model. Others implemented capsule network [11] and multi-task learning CNN model [12] to extract features for retrieval multi-view images. But most of studies focused on general fashion product images.

Based on the problems mentioned above, motivating us to develop a batik clothing image retrieval model. It implements CNN model for extract visual features of images. The visual features of images will be used for measuring similarity between query and images collection in retrieval model. In this study, conducted some experiments that uses different visual features extracted and similarity measurements.

II. PROPOSED METHOD

A. CNN Auto Encoder

The proposed method in this study is illustrated in fig. 1 and fig 2. Fig. 1 shows training step of CNN Networks and fig. 2 shows testing step in batik clothing image retrieval. In this study, the dataset used is a collection of batik clothing images obtained from several shops that sell batik clothes online in e-commerce such as Shopee or Tokopedia. Since the dataset has no label, CNN-based network used for training phase is Autoencoder network as shown in fig 1. This network model will be trained to reconstruct the images in dataset which consists of two parts i.e., Convolutional Part (Encoder) and Deconvolutional Part (Decoder).

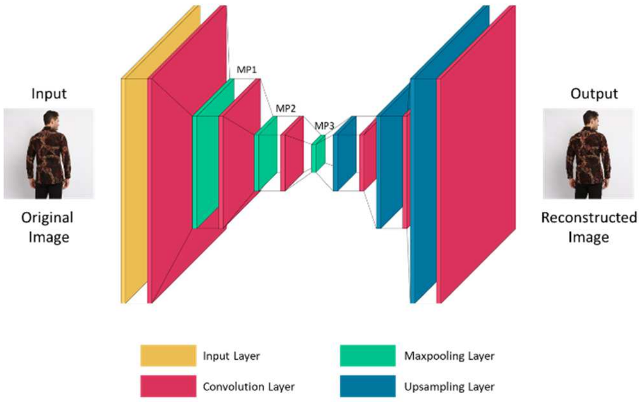

Fig. 1. Autoencoder CNN Network of Proposed Method

Encoder part consist of three convolution layers where each layer uses 32 filters with kernel size 3x3 followed by max-pooling layer. It is trained for obtaining feature maps from convolution layer as representation of input image. Meanwhile, Decoder part will be trained to reconstruct the original input image based on the feature maps that obtain from encoder part. It consists of three up-sampling layers followed by convolution layer that has same filter configuration in encoder part. Autoencoder network is implemented to minimize loss value between input and reconstructed images so the image representative features can be extracted from the model and used in retrieval testing phase.

B. Batik Clothing Image Retrieval

The second part of proposed method is measuring similarity score to get matching or similar batik clothing images as shown in fig 2. Visual features of batik clothing images will be extracted from trained autoencoder network in first phase. In this study, visual features are extracted from the output of convolution and max-pooling layer, namely MP1, MP2, and MP3 as shown in fig 1. The visual features of query image will be compared with visual features of images collection to find relevant batik clothing image. The similarity measurement for comparison uses three functions i.e., Cosine Similarity, Euclidean Distance and Manhattan Distance. After measuring all of similarity between query image and each image in collection of batik clothing images, the similarity or distance score will be sorted or ranked. Score from cosine similarity will be sorted from largest value to smallest. While score from Euclidean and Manhattan Distance will be sorted from the smallest value to biggest. Then, the top-k matching images can be collected from index information in ranked similarity or distance score.

Cosine Similarity can be calculated using equation (1) as shown below:

$$Cos(X_q, X_i) = \frac{X_q \cdot X_i}{\|X_q\| \times \|X_i\|} \quad (1)$$

where X_q is the visual feature of query image and X_i is visual features of image and Xi is visual features in images collection. The larger value of the cosine similarity, the more similar batik clothing images will be collected. Meanwhile, Euclidean Distance and Manhattan Distance can be calculated using equation (2) and (3) below:

$$Edist(X_q, X_i) = \sqrt{\sum_{i=1}^{n} (X_q - X_i)^2} \quad (2)$$

$$Mdist(X_q, X_i) = \sum_{i=1}^{n} |X_q - X_i| \quad (3)$$

where n is vector size on visual features, X_q and X_i are the visual features of query image and each image in batik clothing images collection. The smaller value of Euclidean or Manhattan Distance, the more similar or matching images will be found.

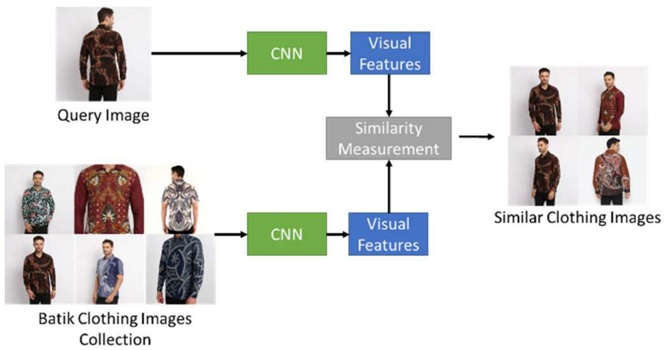

Fig. 2. Image Retrieval Testing of Proposed Method

C. Performance Measurements

The performance of proposed method is measured with three measurements that related to image retrieval. The measurements are image retrieval accuracy, Mean Average Precision (MAP) and Average Recall (AR). Retrieval accuracy is used to calculate accuracy rate for top-k result in batik clothing image retrieval. It measures how accurate the proposed method can retrieve relevant images at least one image at top-k. Image retrieval accuracy is calculated by equation below:

$$Top\_k\_acc = \frac{\sum_{n=1}^{N} R_n}{N} \quad (4)$$

where N is number of query images and R_n is relevant value for image retrieval of each query image which is equaling to 1 if one of the images is in the top-k retrieval is the same batik clothing image as the query image, otherwise is equaling 0.

MAP for image retrieval is the mean of the average precision score of each query:

$$MAP@K = \frac{\sum_{n=1}^{N} AP@K(n)}{N} \quad (5)$$

Where N is total of query images. AP@K(n) is the average precision at top-k retrieval images for each query image. Average precision is calculated by formulation below:

$$AP@K = \frac{\sum_{k=1}^{K} P(k) \times rel(k)}{M} \quad (6)$$

Where M is number of matching or relevant batik clothing images in top-k retrieval images, P(k) is the precision score at top-k retrieval images, and rel(k) is an indicator that equaling 1 if product at rank k is a matching batik clothing product, 0 otherwise.

AR is used to calculate the probability that all relevant images are retrieved in top-k retrieval images using the proposed method. AR can be calculated by equation below:

$$AR@K = \frac{\sum_{n=1}^{N} R(k)}{N} \quad (7)$$

where N is the total of query images and R(k) is recall score at top-k retrieval images.

III. EXPERIMENTAL RESULTS

To evaluate the proposed method, image collection in dataset is separated into training and testing images. Testing images consist of 65 batik clothing images which each image has different batik pattern. This data partition is used as query images in proposed method. Meanwhile training images are batik clothing images with different viewpoint corresponding with each testing image as shown in fig 3. It is used as training data in Autoencoder CNN training phase and collection images in image retrieval in proposed method. Autoencoder training phase is conducted with several hyperparameters which are 50 epochs, Adam Optimizer, and Mean Absolute Error loss function.

Fig. 3. Examples of Batik Clothing Image Dataset

The results of performance measurement are shown in table I – IV. Table I is result of retrieval accuracy at top-k in each combination of visual feature and similarity function. As shown in table I, the best score is top-50 accuracy with visual feature MP1 and MP3 using Cosine Similarity which score is 90.8%. In addition, the best scores of each top-k are 81.5% accuracy at top-10 using MP2 visual features and Manhattan Distance, 84.6% accuracy at top-20 using MP1 visual feature and Manhattan Distance and MP3 visual feature and Manhattan Distance, 87.7% accuracy at top-30 using MP3 visual feature and Cosine Similarity and MP3 visual feature and Manhattan Distance, then 89.2% accuracy at top-40 using MP1 visual feature and Cosine Similarity then MP3 visual feature and Cosine Similarity.

TABLE I. RETRIVEL ACCURACY OF BATIK CLOTHING IMAGE

Visual Feature	Similarity Function	Top-k Retrieval Accuracy				
		Top-10	*Top-20*	*Top-30*	*Top-40*	*Top-50*
MP1	Cosine	70.8%	80.0%	83.1%	89.2%	90.8%
	Euclidean	70.8%	76.9%	83.1%	83.1%	84.6%
	Manhattan	80.0%	84.6%	86.2%	86.2%	89.2%
MP2	Cosine	73.8%	83.1%	86.2%	87.7%	89.2%
	Euclidean	70.8%	78.5%	83.1%	83.1%	84.6%
	Manhattan	81.5%	83.1%	84.6%	86.2%	89.2%
MP3	Cosine	72.3%	81.5%	87.7%	89.2%	90.8%
	Euclidean	69.2%	81.5%	83.1%	83.1%	86.2%
	Manhattan	80.0%	84.6%	87.7%	87.7%	89.2%

Compared with [7], the proposed method provides better performance at top-16 retrieval accuracy than the model used visual features extracted from unsupervised CNN or auto encoder model (CAE) as shown in table II. But the proposed method can't provide better performance at top-16 retrieval accuracy score comparing with visual features extracted from supervised CNN model. The main reason is that the dataset used in this study does not have label or attributes that can help CNN model learn more information about the image dataset and get better visual features.

TABLE II. COMPARISON OF RETRIEVAL ACCURACY

Method / Visual Features	Top-16 Retrieval Accuracy	
	Euclidean	Manhattan
CNN [7]	99.38%	99.31%
CAE [7]	67.37%	63.87%
Proposed Method		
▪ MP1	75.38%	80.00%
▪ MP2	76.92%	81.54%
▪ MP3	76.92%	81.54%

Table III shows Mean Average Precision results of each visual feature and similarity function combination. As shown in table III, the best score of each MAP at top-k is visual features MP2 with Manhattan Similarity. The best score is 58.8% at top-10 retrieval images, following by top-20 with MAP score 55.5%, top-30 with MAP score 54.7%, top-40 with MAP score 54.6%, and top-50 with MAP score 54.7%.

TABLE III. MAP OF BATIK CLOTHING IMAGE RETRIEVAL

Visual Feature	Similarity Function	Mean Average Precision				
		Top-10	*Top-20*	*Top-30*	*Top-40*	*Top-50*
MP1	Cosine	35.0%	34.6%	33.5%	33.1%	32.8%
	Euclidean	35.7%	34.6%	33.5%	32.8%	32.7%
	Manhattan	55.8%	53.1%	51.8%	51.8%	51.9%
MP2	Cosine	39.4%	37.2%	37.3%	35.1%	35%
	Euclidean	37.5%	36.5%	35.7%	34.6%	33.8%
	Manhattan	58.8%	55.5%	54.7%	54.6%	54.7%
MP3	Cosine	37.3%	35.8%	35.5%	33.8%	33.0%
	Euclidean	36.1%	35.4%	34.9%	34.2%	32.8%
	Manhattan	54.6%	52.4%	52.2%	50.2%	50.3%

Table IV shows average recall of top-k result of each visual feature and similarity function combination. As shown in table IV, the best score of AR is visual features MP1 with Manhattan Distance and visual features MP2 with Manhattan Distance at top-50 images retrieval. Its score is 61.9%. AR at top-40 has best score that is using MP1 visual features MP1 and Manhattan Distance which has 61% AR score. Meanwhile, AR at top-10, top-20, and top-3 has best score with MP2 visual features and Manhattan Distance as shown in table IV.

TABLE IV. AR OF BATIK CLOTHING IMAGE RETRIEVAL

Visual Feature	Similarity Function	Average Recall				
		Top-10	*Top-20*	*Top-30*	*Top-40*	*Top-50*
MP1	Cosine	36.8%	46%	49.5%	55.9%	59.9%
	Euclidean	35.8%	44.6%	50.8%	54.5%	56.5%
	Manhattan	48.1%	54.9%	60.0%	61.0%	61.9%
MP2	Cosine	37.6%	48.6%	51.7%	55.4%	58.8%
	Euclidean	37.6%	46.7%	51.3%	54.6%	57.1%
	Manhattan	49.4%	56.9%	59.0%	60.0%	61.9%
MP3	Cosine	37.1%	47.7%	52.2%	55.9%	58.8%
	Euclidean	36.5%	47.2%	51.8%	55.1%	58.6%
	Manhattan	47.6%	55.5%	57.5%	60.4%	61.8%

In addition, fig 4 illustrates some results of batik clothing image retrieval using proposed method. The left column is query Image and other columns shows top-10 ranking image retrieval results based on correspondence query image. The red rectangle means the relevant or matching batik clothing product of query image retrieved by proposed method. As shown in figure 4, the second query image is the example of the best result on image retrieval using proposed method. The matching images occurred on the 1st and 2nd images in retrieval results. Meanwhile, the matching images of first query image example are occurred on 6th and 7th image in retrieval results and the matching image of the last example is occurred on 10th image in results.

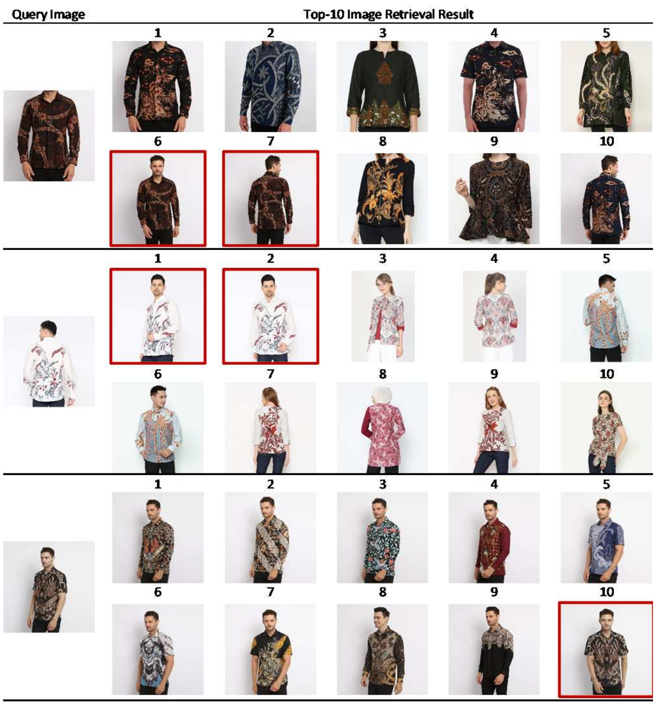

Fig. 4. Examples of Batik Clothing Image Retrieval Result

IV. CONCLUSIONS

In this paper, image retrieval of Indonesia batik clothing based on convolutional neural network was proposed. Autoencoder model was proposed as convolutional neural network that is trained for extract visual features of batik clothing image used for comparing similarity between query image and dataset collection.

Some experiments by combining different visual feature and similarity function were conducted to measure the performance of proposed method. Based on the results, for batik image retrieval model of proposed method can reach 90.8% in retrieval accuracy at top-50 using MP1+Cosine Similarity, 58.8% in mean average precision at top-10 using MP2+Manhattan Distance, and 61.9% in average recall at top-50 using MP1+Manhattan Distance and MP2+Manhattan Distance. The proposed method needs some improvement for the next study, such as providing more label information of the dataset and implementing other deep neural networks for extracting useful image features.

REFERENCES

[1] Y. Zhu and S. Newsam, "DenseNet for dense flow," in Proceedings - International Conference on Image Processing, ICIP, 2018, vol. 2017-September. doi: 10.1109/ICIP.2017.8296389.

[2] M. Tan and Q. v. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in 36th International Conference on Machine Learning, ICML 2019, 2019, vol. 2019-June.

[3] Q. Yuan, S. Xu, and L. Jian, "A new method for retrieving batik shape patterns," J Assoc Inf Sci Technol, vol. 69, no. 4, 2018, doi: 10.1002/asi.23977.

[4] H. Prasetyo and J. W. Simatupang, "Batik Image Retrieval Using Maximum Run Length LBP and Sine-Cosine Optimizer," 2019. doi: 10.1109/ICSECC.2019.8907190.

[5] Y. Gultom, A. M. Arymurthy, and R. J. Masikome, "Batik Classification using Deep Convolutional Network Transfer Learning," Jurnal Ilmu Komputer dan Informasi, vol. 11, no. 2, 2018, doi: 10.21609/jiki.v11i2.507.

[6] Y. Azhar, Moch. C. Mustaqim, and A. E. Minarno, "Ensemble convolutional neural network for robust batik classification," IOP Conf Ser Mater Sci Eng, vol. 1077, no. 1, 2021, doi: 10.1088/1757-899x/1077/1/012053.

[7] H. Prasetyo and B. A. Putra Akardihas, "Batik image retrieval using convolutional neural network," Telkomnika (Telecommunication Computing Electronics and Control), vol. 17, no. 6, 2019, doi: 10.12928/TELKOMNIKA.v17i6.12701.

[8] Y. Miao, G. Li, C. Bao, J. Zhang, and J. Wang, "ClothingNet: Cross-Domain Clothing Retrieval with Feature Fusion and Quadruplet Loss," IEEE Access, vol. 8, 2020, doi: 10.1109/ACCESS.2020.3013631.

[9] S. Jiang, Y. Wu, and Y. Fu, "Deep bidirectional cross-Triplet embedding for online clothing shopping," ACM Transactions on Multimedia Computing, Communications and Applications, vol. 14, no. 1, 2018, doi: 10.1145/3152114.

[10] C. Ning, Y. Di, and L. Menglu, "Survey on clothing image retrieval with cross-domain," Complex and Intelligent Systems, 2022, doi: 10.1007/s40747-022-00750-5.

[11] F. Kinli, B. Ozcan, and F. Kirac, "Fashion image retrieval with capsule networks," 2019. doi: 10.1109/ICCVW.2019.00376.

[12] M. Fadhilla, J. Y. Lin, W. J. Chen, and G. S. Lin, "Multi-View Clothing Image Searching Based on Deep Neural Network," 2020. doi: 10.1109/ICCE-Taiwan49838.2020.9258235.