■ 3010

Batik image retrieval using convolutional neural network

Heri Prasetyo*[1], Berton Arie Putra Akardihas[2]
Department of Informatics, Universitas Sebelas Maret (UNS), Surakarta, Indonesia
*Corresponding author, e-mail: heri.prasetyo@staff.uns.ac.id[1], bertonarie123@student.uns.ac.id[2]

Abstract
This paper presents a simple technique for performing Batik image retrieval using the Convolutional Neural Network (CNN) approach. Two CNN models, i.e. supervised and unsupervised learning approach, are considered to perform end-to-end feature extraction in order to describe the content of Batik image. The distance metrics measure the similarity between the query and target images in database based on the feature generated from CNN architecture. As reported in the experimental section, the proposed supervised CNN model achieves better performance compared to unsupervised CNN in the Batik image retrieval system. In addition, image feature composed from the proposed CNN model yields better performance compared to that of the handcrafted feature descriptor. Yet, it demonstrates the superiority performance of deep learning-based approach in the Batik image retrieval system.

Keywords: autoencoder, CNN, deep learning, feature extraction, image retrieval

1. Introduction

The brain is an amazing organ in the human body. With our brains, we can understand what we see, smell, taste, hear and touch. The infant brain weight is only about half a kilogram but can solve a big problem, and even supercomputers cannot. After several months of birth, the baby can recognize the face of his parents, discern discrete objects from the background, and begin to speak. Within one year the baby has an intuition about natural objects, can follow objects and understand the meaning of a sound. When they are children, they can understand grammar and have thousands of words in their vocabulary.

Building machines that have intelligence like our brains are not easy, to make machines with artificial intelligence we have to solve very complex computing problems that we have even struggled with, problems that our brains can solve in a matter of seconds. To overcome this problem, we have to develop other ways to program computers that have been used in this decade. Therefore there arises an active field of artificial computer intelligence and also commonly called deep learning [1].

Nowadays Artificial intelligence has undergone very rapid development. Ai has been used in many fields of research, in the field of computer vision Content-Based Image Retrieval (CBIR) has been developed in multi-level schemes with low-level features to high-level features. Convolutional Neural Network (CNN) has been successfully used to be an effective descriptor feature and gain accurate results. In general, the features gain by the deep learning method are trained by mimic human perceptions through various operations such as convolution and pooling. Deep learning has become a descriptor feature that is better than low-level features. Although now the CNN module has become state of the art in computer vision this does not guarantee the features obtained from the highest level always get the best performance [2].

In the Content-Based Image Retrieval system aims to provide the right way to do the browsing, retrieving and searching some desired images that have been stored in the image database. The image database contains many images that have been stored and arranged in a storage device. Usually, the size of the image database is very large so that the process of searching for specific images manually requires a lot of time, and causes conditions that are uncomfortable for the user. For example, Batik is a cultural heritage of the archipelago Indonesia that has a high value and blend of art, laden with philosophical meanings and meaningful symbols that show the way of thinking of the people making it. Batik is a craft that has been a part of Indonesian culture especially Javanese for a long time, batik have

a lot of motives, pattern and color so to take specific batik picture from the database very challenging [3].

This paper offers a solution to use convolutional neural networks to carry out CBIR tasks to solve problems that occur in taking batik images. The method intended is to produce effective image descriptors from the CNN architecture. Descriptors of this feature are very important for content-based shooting systems. The Image feature is used to improve the performance and to solve problems in existing batik shooting systems.

2. Content-based Image Retrieval System

Image retrieval is a computer system for searching and retrieving a specific image in large or big size of image databases. The classical approach appends on the metadata such as texts, keywords, or descriptions embedded in an image. Thus, the image retrieval can be performed with the search key as aforementioned text, keywords, etc. This technique is inefficient since the manual image annotations are time-consuming and exhausting process. Even though, large amounts of automatic images annotations have been proposed in literature [4], an image retrieval system with content annotation still cannot deliver satisfactory result.

CBIR is computer application dealing with the searching problems over large-scale image database. CBIR, also recognized as Query-Based Image Content (QBIC) and Content-Based Visual Information Search (CBVIR), differs with the content-based approach. The CBIR analyzes the image content rather than metadata information such image keywords, tags, or image descriptions [5].

In this paper, the usability of CNN model is extended to the CBIR task. The main reason is the superiority performance offered by CNN model compared to the handcrafted feature in the computer vision and recognition tasks. The CNN or Deep Learn network achieves the outstanding retrieval performance in the ImageNet challenge [6]. The CNN model inspires the other deep learning-based approaches, such as AlexNet [7], VGGNet [8], GoogLeNet [9], Microsoft ResNet [10], etc., to tackle the obsolete of handcrafted feature in the image retrieval domain.

The CNN model receives a three-dimensional image of size $h \times w \times d$, where h and w are spatial dimensions and d is the number of channels. This image is further processed thorough the CNN architecture consisting several convolutions, max-poolings, and activation functions to perform end-to-end image feature generation. Let X_{ij} be a vector data located at spatial position (i, j) in specific layers. The CNN computes a new data Y_{ij} as follow:

$$Y_{ij} = f_{ks}\left(\left\{X_{si+\delta i, sj+\delta j}\right\}_{0 \leq \delta i, \delta j < k}\right) \tag{1}$$

where k and s denote kernel size and stride, respectively. The function f_{ks} is the layer type used such as matrix dot multiplication for convolutional layers, max spatial for max pooling layers, nonlinear functions for activation functions, and other types of layers. This form of functionality is maintained using kernel size and step composition while still using the transformation rules.

$$f_{ks} \circ g_{k's'} = (f \circ g)_{k'+(k-1)s', ss'} \tag{2}$$

While a general network computes general nonlinear functions, a network with only layers of this form computes a nonlinear filter, which we call a deep filter or fully convolutional network. FCN naturally operates at any size input and produces the appropriate spatial dimensions. The loss function is valued composed with the FCN defines task. If the loss function is a sum over the spatial dimensions of the final layer $l(x; \theta) = \sum_{ij} l'(x_{ij}; \theta)$, the parameter gradient will be a sum over the parameter gradients of each of its spatial components. Thus stochastic gradient on l computed on whole images will be the same as the stochastic gradient on l', taking all the final receptive fields as minibatch. When calculating this receptive field is done repeatedly with forward and backward propagation operations feedback will be more effective if the calculation is done layer by layer in all images compared to computing patch by patch to the part of the image. An illustration of a CNN operation can be seen in Figure 1.

The proposed CNN model constructs the feature descriptor from Batik image. This feature descriptor is to measure the similarity between query and target images in database under the K-Nearest Neighbors (KNN) [11] strategy. This KNN technique performs similarity matching with the distance score criterion. This paper investigates two CNN models in the training stage, i.e. with supervised and unsupervised learning approaches. Figure 1 illustrates an example of proposed supervised CNN architecture for Batik image retrieval. The supervised terminology refers to the utilization of class label, whereas unsupervised disobeys the image label in the training process. Autoencoder is simple example of unsupervised CNN method which compresses the data features into smaller size and recovers back to the original data [12].
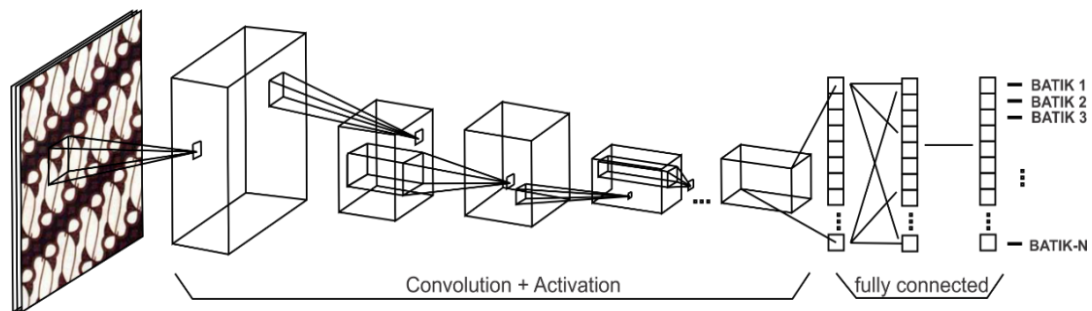

Figure 1. Ilustration operation using CNN

3. Method

This section presents two methods for generating the feature descriptor in the Batik image retrieval system. We firstly explain the supervised CNN model. Then, the unsupervised CAE model [13] is subsequently described in this section.

3.1. Supervised Learning

The CNN model is the supervised deep learning-based approach commonly used in the image classification [14], prediction [15], segmentation, analysis [16], etc. The supervised CNN model consists of several layers such as convolutional layer, max pooling layer, etc. These layers are repeated over several times and fed into the fully connected layer at the end of CNN layer [17]. Our proposed image retrieval system employs the CNN architecture with six convolutional layers and two fully connected layers to generate Batik feature descriptor. Table 1 summarizes the CNN architecture used in our proposed method.

Table 1. The Supervised CNN Architecture for Batik Image Retrieval

Layer Type	Size	Output Shape
Input	(128,128,3)	-
Convolutional + Relu	8 (3x3) filters, 1 stride, 2 padding	(128,128,8)
Max Pooling	8 (2x2) filters, 2 stride, 0 padding	(64,64,8)
Convolutional + Relu	16 (3x3) filters, 1 stride, 2 padding	(64,64,16)
Max Pooling	16 (2x2) filters, 2 stride, 0 padding	(32,32,16)
Convolutional + Relu	32 (3x3) filters, 1 stride, 2 padding	(32,32,32)
Max Pooling	32 (2x2) filters, 2 stride, 0 padding	(16,16,32)
Convolutional + Relu	64 (3x3) filters, 1 stride, 2 padding	(16,16,64)
Max Pooling	64 (2x2) filters, 2 stride, 0 padding	(8,8,64)
Convolutional + Relu	128 (3x3) filters, 1 stride, 2 padding	(8,8,128)
Max Pooling	128 (2x2) filters, 2 stride, 0 padding	(4,4,128)
Convolutional + Relu	256 (3x3) filters, 1 stride, 2 padding	(4,4,256)
Max Pooling	256 (2x2) filters, 2 stride, 0 padding	(2,2,256)
Flatern + Dropout (30%)	(1,1,1024) 1024 neurons	256
Dense	256 neurons	97
Softmax	97 way	97

After performing six convolution and max-pooling operations, an input image of size $128 \times 128 \times 3$ is converted into new representation with dimensionality $2 \times 2 \times 256$. This new data representation is then flatten to become one dimensional data of size $1 \times 1 \times 1024$. This flatten data is subsequently processed and trained with the Multi-Layer Perceptron (MLP). Herein, the MLP receives 1024 input feature and feeds into 1024 input neurons. The hidden and output layers are set as 256 and 97, respectively. The value of 97 in output layers is equivalent to that of the desired class target, i.e. the number of Batik image classes used in the proposed image retrieval system.

3.2. CAE Unsupervised Learning

This paper also considers the other CNN model, namely Convolutional Auto-Encoder (CAE), for generating image feature. The CAE is an unsupervised deep learning-based method, i.e. the image label is not required in the training process. In order to generate image feature, this technique learns and captures the information from input data directly without the availability of class label.

The CAE involves two parts, i.e. encoder and decoder blocks. The encoder block processes the sample data X consisting n samples and m features to yield the output Y. In the opposite side, the decoder aims to reconstruct the original sample data X from the Y. Let X' be the reconstructed data produced at the decoder side. The main goal of CAE is to minimize the difference between the original data X and reconstructed version X'. Specifically, the encoder simply maps the input X into new representation Y with the help of function f. This process can be formulated as follow:

$$Y = f(X) = s_f (WX + b_X) \tag{3}$$

where s_f denotes the nonlinear activation function in encoder side. CAE simply performs a linear operation if one simply uses identity function for s_f. The W and $b_X \in R^n$ are encoder parameters, respectively, referring as weight matrix and bias vector. In contrast, the decoder reconstructs X' from Y representation by means of function g. This process can be simply illustrated as:

$$X' = g(Y) = s_g (W' Y + b_Y) \tag{4}$$

where s_g represents the activation function in decoder side. The b_Y and W are the bias vector and weight matrix, respectively, denotingas decoder parameter.

Strictly speaking, the CAE model searches the global or near optimum parameter $= (W, b_X, b_Y)$ in the training process. This task is equivalent to the minimization process of loss function over all dataset X under the following objective function:

$$\theta = \min_\theta L(X, X') = \min_\theta L(X, g(f(X))) \tag{5}$$

where $L(\cdot, \cdot)$ denotes the auto-encoder loss function. In this paper, we simply use linear reconstruction L_2 for loss function, or commonly referred as Mean Squared Error (MSE) [18]. This loss function is formally defined as:

$$L_2(\theta) = \sum_{i=1}^{n} \|x_i - x'_i\|^2 = \sum_{i=1}^{n} \|x_i - g(f(x_i))\|^2 \tag{6}$$

where $x_i \in X, x'_i \in X'$ and $y_i \in Y$, respectively denote the original input data, reconstructed data, and new compact representation of input data.

In this paper, the CAE architecture was built with four encoding blocks and four decoding stages. This architecture includes a stacked Convolutional Auto-Encoder. The summary of CAE architecture used in this paper can be seen in Table 2. Suppose that an input image is of size $128 \times 128 \times 3$. As it can be inferred from Table 2, this image is convolved four times to obtain new simpler and compact representation. This process can be also considered as repetitive encoding. Herein, the new representation is regarded as neural code

with dimensionality $4 \times 4 \times 128$. By using the backward approach and decoding process, this neural code can be recovered back to yield the reconstructed image of original size $128 \times 128 \times 3$. This reverse process performs the deconvolution and unpooling operations. The CAE neural code can be further utilized as the feature descriptor in the proposed Batik image retrieval system.

Table 2. The CAE Architecture for Batik Image Retrieval System

Layer Type	Size	Output Shape
Input	(128,128,3)	-
Convolutional + Relu	32 (3x3) filters, 1 stride, 2 padding	(128,128,32)
Max Pooling + Dropout	32 (2x2) filters, 2 stride, 0 padding	(64,64,32)
Convolutional + Relu	64 (3x3) filters, 1 stride, 2 padding	(64,64,64)
Max Pooling + Dropout	64 (2x2) filters, 2 stride, 0 padding	(32,32,64)
Convolutional + Relu	64 (3x3) filters, 1 stride, 2 padding	(32,32,64)
Max Pooling + Dropout	64 (2x2) filters, 2 stride, 0 padding	(16,16,64)
Convolutional + Relu	128 (3x3) filters, 1 stride, 2 padding	(16,16,128)
Max Pooling + Dropout	128 (2x2) filters, 2 stride, 0 padding	(8,8,128)
Max Pooling (Neural Code)	128 (2x2) filters, 1 stride, 2 padding	(4,4,128)
Unpoling	128 (2x2) filters, 2 stride, 0 padding	(8,8,128)
Deconvovution + Relu	128 (3x3) filters, 1 stride, 2 padding	(8,8,128)
Unpooling + Dropout	128 (2x2) filters, 2 stride, 0 padding	(16,16,128)
Deconvovution + Relu	64 (3x3) filters, 1 stride, 2 padding	(16,16,64)
Unpooling + Dropout	64 (2x2) filters, 2 stride, 0 padding	(32,32,64)
Deconvovution + Relu	64 (3x3) filters, 1 stride, 2 padding	(32,32,64)
Unpooling + Dropout	64 (2x2) filters, 2 stride, 0 padding	(64,64,64)
Deconvovution + Relu	32 (3x3) filters, 1 stride, 2 padding	(64,64,32)
Unpooling + Dropout	32 (2x2) filters, 2 stride, 0 padding	(128,128,32)
Deconvovution + Sigmoid	32 (3x3) filters, 1 stride, 2 padding	(128,128,3)

3.3. Learning process and Hyperparemeter Tuning

The CNN model is very sensitive to hyperparameter changes in the learning process, since it utilizes the Restructured Linear Unit (ReLu) $f(x) = max(0, x)$ for its activation function. This function is with the gradient descent making it very unstable in comparison with the tanh and sigmoid activation functions. Compared to the aforementioned activation functions, ReLu yields an identical error with 25% less iteration in learning stage [7].

In the training process of our proposed image retrieval system, we simply split the image dataset as two folds, i.e. 75% and 25% for training and testing purpose, respectively. The Adaptive Moment Estimation (Adam) [19] is exploited for CNN optimizer with learning rate 0.0001. We simply employ the Mean Square Error (MSE) [20] for calculating the loss function. For avoiding the overfitting problem and dealing with small size of dataset, the proposed system uses data augmentation technique to improve the data variation. The training and testing processes are conducted under the Intel Core i5 2010 processor. From our experiment, the supervised CNN and CAE models require around 10 hours and 3 days, respectively, for the training process. At the end of training process, two deep learning based models produce a set of image features which can be used for the descriptor in the Batik image retrieval. These image features are simply obtained from the last layer and neural code layer of supervised CNN and CAE models, respectively.

4. Experimental Study

Extensive experiments were carried out to investigate and examine the proposed method performance in the Batik image retrieval system. Firstly, we give a brief description about the image dataset used in the experiment. The effectiveness of the proposed method is subsequently observed under visual investigation. Then, the objective performance comparisons are further evaluated to overlook the effect of different distance metrics and superiority of the proposed method in comparison with the former competing schemes.

4.1. Dataset

This experiment utilizes a set of Batik images, refered as Batik image dataset, over various patterns, colors, and motifs. This image database consists of 1552 image. This database is further divided into 97 image classes. Each class contains a set of similar images

regarding to their motifs and content appearance. Each image class owns 16 similar images, in which all images belonging to the same class are considered as similar images. Figure 2 gives several examples of Batik images from the dataset.

4.2. Practical Application on Batik Image Retrieval

This sub-section evaluates the performance of the proposed method under visual investigation. The proposed method utilizes the image feature obtained from CNN and CAE approach for performing Batik image retrieval system. The correctness of the proposed method is determined whether the system returns a set of retrieved images correctly or not.

Figure 3 displays the retrieved images returned by the proposed image retrieval system using the CNN and CAE image features. We only show six-teen retrieved images arranged in ascending manner based on their similarity score. The similarity criterion is measured using the distance score and given at the top of each image. Smaller distance value indicates more similar between the query and target image in database. As shown in this figure, the proposed method with CNN feature returns all retrieved images correctly. It is little regrettable that the proposed method with CAE feature only produces six retrieved images correctly.

Figure 2. Some image samples in the Batik dataset
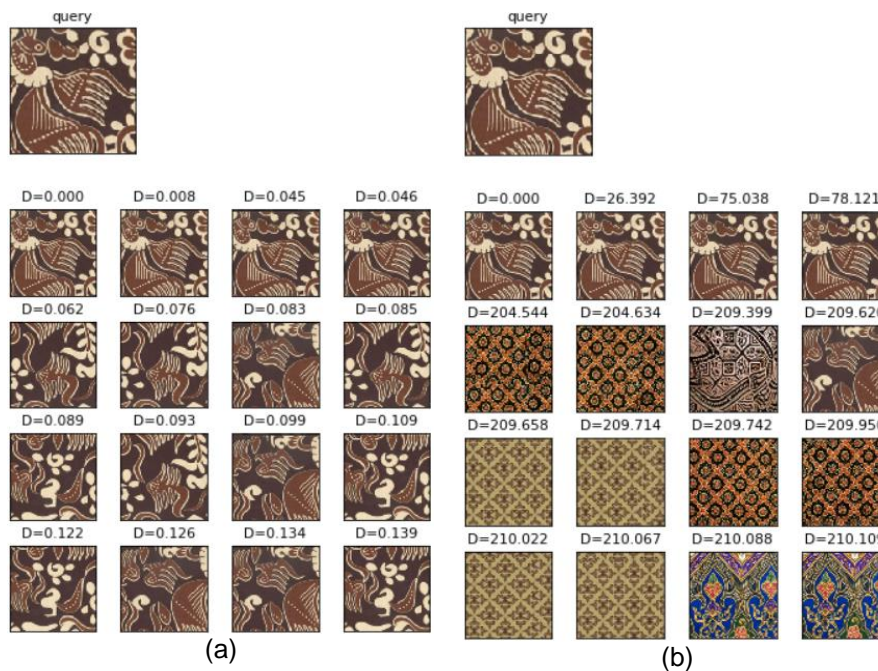

(a)   (b)

Figure 3. Performance evaluation in terms of visual investigation
for the proposed method with: (a) CNN, and (b) CAE image feature

4.3. Comparison of Porposed Methods with Direfferent Distance Metrics

This sub-section reports the effect of different distance metrics on the proposed method. In this experiment, three distance metrics, namely Euclidean [21], Manhattan [22], and Bray-Curtis distance [23], are extensively examined over two performance criterion, i.e. precision and recall rate. These two scores are formally defined as:

$$p_i(n) = \frac{R_V}{n} \qquad (7)$$

$$r_i(n) = \frac{R_V}{M} \qquad (8)$$

where $p_i(n)$ and $r_i(n)$ denotes the precision and recall rate, respectively, if image i is turned as query image. The symbols n and M represent the number of retrieved images and total images in database which is relevant to image i, respectively. R_V is the number of images which are relevant to query image i obtained at n retrieved images.

Figure 4 shows the performance comparison over various distance metrics in terms of Precision and Recall scores. All images in database are chosen as query image. The number of retrieved images are set as $n = \{1,2,...,16\}$. In most cases, Bray-Curtis distance yields the best retrieval performance compared to that of the other distance metrics for both CNN and CAE image feature. In the Batik image retrieval system, the Bray-Curtis distance becomes a good candidate for measuring the similarity between the query and target images in database.

Table 3 tabulates more complete comparsions for the proposed image retrieval system using CNN and CAE features over various distance. This comparison is evaluated in terms of average recall rate with the number of retrieved images as $n = 16$. Herein, all images in database are turned as query image. As reported in this table, the proposed method with supervised CNN delivers better performance compared to that of CAE technique. The image feature obtained from proposed supervised CNN method is more suitable for Batik image retrieval task.
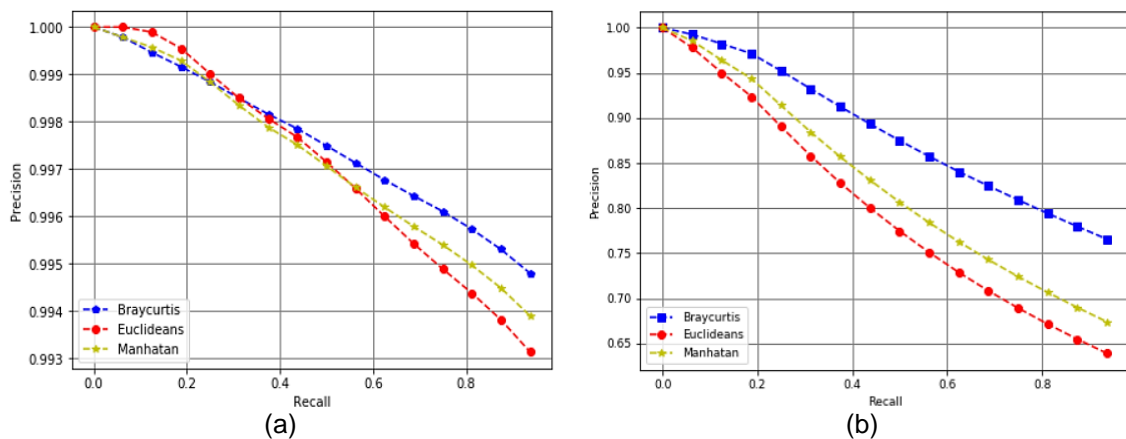

Figure 4. Performance comparisons in terms of precision and recall rates
over various distance metrics with the image features from: (a) CNN, and (b) CAE method

4.4. Comparison against Former Methods

This sub-section summarizes the performance comparison between the proposed supervised CNN method and former existing schemes on Batik image retrieval system. This comparison is conducted in terms of Average Precision Recall (APR) score. The APR is formally defined as:

$$APR = \frac{1}{N}\sum_{i=1}^{N} r_i(n) \tag{9}$$

where $r_i(n)$ and N are the recall rate for query image i and the total number of images in database, respectively. Herein, all images in database are turned as query image indicating that $N = 1552$. Thus, the APR value is averaging over all query images. The number of retrieved images is set as 16 yielding $n = 16$. To make a fair comparison, this experiment also investigates the dimensionality of image feature.

Table 4 reports the performance comparison in terms of feature dimensionality and APR value. As shown in this table, the proposed supervised CNN yields the best performance in comparison with the other competing schemes. It is noteworthy that the proposed method requires lowest feature dimensionality (with exceptional on comparison to LBP [20] scheme). This lower dimensionality indicates the faster process on KNN searching for effective Batik image retrieval system. Thus, the proposed method can be considered on implementing the Batik image retrieval and classification system.

Table 3. APR CNN and CAE

Method	Euclidean	Manhattan	Bray-curtis
CNN	0.9938	0.9931	0.9947
CAE	0.6737	0.6387	0.7654

Table 4. APR Comparison with Former Method

Method	Feature Size	APR (%)
LBP [24]	59	92.57
LTP [25]	118	95.65
CLBP [26]	118	95.17
LDP [27]	236	93.52
Gabor Filter [28]	144	96.55
ODBTC+PSO [3]	384	97.68
Proposed Supervised CNN	97	99.47

5. Conclusions

A new content-based image retrieval system has been presented in this paper. This system achieves the retrieval accuracies 99.47% and 76.54%, respectively, while the image feature is constructed from CNN and CAE deep learning-based architecture on Batik image database. The CNN outperforms the former existing schemes in terms of retrieval accuracy. In addition, it requires the lowest image features, i.e. 97 feature dimensionality, compared to other methods. For future work, a slight modification can be carried out for CAE model by adding fully-connected layers before and after the neural code section. This scenario may reduce the dimensionality of image feature, at the same time, it improves the performance for Batik image retrieval.

References

[1] Johnson MH. The neural basis of cognitive development. In: Damon W. *Editor*. Handbook of child psychology: Cognition, perception, and language. Hoboken: John Wiley & Sons Inc. 1998: 1-49.

[2] Liu P, et al. Fusion of deep learning and compressed domain features for content-based image retrieval. *IEEE Transactions on Image Processing*. 2017. 26(12): 5706-5717.

[3] Prasetyo, H, et al. Batik Image Retrieval Using ODBTC Feature and Particle Swarm Optimization. *Journal of Telecommunication, Electronic Computer Engineering*. 2018. 10(2-4): 71-74.

[4] Datta R, Li J, Wang JZ. *Content-based image retrieval: approaches and trends of the new age*. Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval. 2005.

[5] Eakins JP, Graham ME. Content based image retrieval: A report to the JISC technology applications programme. 1999.

[6] Russakovsky O, et al. Imagenet large scale visual recognition challenge. International Journal of *Computer Vision*. 2015; 115(3): 211-252.

[7] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*. 2012: 1097-1105.

[8] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv*. 2014.

[9] Szegedy C, et al. *Going deeper with convolutions*. Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.

[10] He K, et al. *Deep residual learning for image recognition*. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[11] Cover T, Hart P. Nearest neighbor pattern classification. *IEEE transactions on information theory*. 1967; 13(1): 21-27.

[12] Petscharnig S, Lux M, Chatzichristofis S. *Dimensionality reduction for image features using deep learning and autoencoders*. Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing, ACM. 2017.

[13] Masci J, et al. *Stacked convolutional auto-encoders for hierarchical feature extraction*. International Conference on Artificial Neural Networks. 2011: 52-59.

[14] Wang R, et al. A Crop Pests Image Classification Algorithm Based on Deep Convolutional Neural Network. *TELKOMNIKA Telecommunication Computing Electronics and Control*. 2017; 15(3): 1239-1246.

[15] Baharin A, Abdullah A, Yousoff SNM. Prediction of Bioprocess Production Using Deep Neural Network Method. *TELKOMNIKA Telecommunication Computing Electronics and Control*. 2017; 15(2): 805-813.

[16] Sudiatmika IBK, Rahman F, Trisno T, Suyoto S. Image forgery detection using error level analysis and deep learning. *TELKOMNIKA Telecommunication Computing Electronics and Control*. 2019; 17(2): 653-659.

[17] Setiawan W, Utoyo MI, Rulaningtyas R. Classification of neovascularization using convolutional neural network model. *TELKOMNIKA Telecommunication Computing Electronics and Control*. 2019; 17(1): 463-472.

[18] Meng Q, et al. *Relational autoencoder for feature extraction*. 2017 International Joint Conference on Neural Networks (IJCNN). 2017: 364-371.

[19] Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv. 2014.

[20] Hagan MT, Menhaj MB. Training feedforward networks with the Marquardt algorithm. *IEEE transactions on Neural Networks*. 1994; 5(6): 989-993.

[21] Danielsson PE. Euclidean distance mapping. *Computer Graphics image processing*. 1980; 1 4(3): 227-248.

[22] Craw S. Manhattan distance. In: Sammut C, Webb GI. Encyclopedia of Machine Learning and Data Mining. Springer. 2017: 790-791.

[23] Kokare M, Chatterji B, Biswas P. *Comparison of similarity metrics for texture image retrieval*. *TENCON 2003*. IEEE, Conference on Convergent Technologies for the Asia-Pacific Region. 2003; 2: 571-575.

[24] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis machine intelligence*. 2002; 24(7): 971-987.

[25] Tan X, Triggs B. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE transactions on image processing*. 2010; 19(6): 1635-1650.

[26] Guo Z, Zhang L, Zhang D. A completed modeling of local binary pattern operator for texture classification. *IEEE Transactions on Image Processing*. 2010; 19(6): 1657-1663.

[27] Zhang B, et al. Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor. *IEEE transactions on image processing*. 2010; 19(2): 533-544.

[28] Prasetyo H, Wiranto W, Winarno W. Statistical Modeling of Gabor Filtered Magnitude for Batik Image Retrieval. *Journal of Telecommunication, Electronic Computer Engineering*. 2018; 10(2-4): 85-89.