ICIC Express Letters ICIC International ©2022 ISSN 1881-803X
Volume **16**, Number **2**, February **2022** pp. **137–144**

TOWARDS CULTURAL HERITAGE CONTENT RETRIEVAL BY CONVOLUTION NEURAL NETWORK

Sathit Prasomphan

Department of Computer and Information Science
Faculty of Applied Science
King Mongkut's University of Technology North Bangkok
1518 Pracharat 1 Road Wongsawang, Bangsue, Bangkok 10800, Thailand
sathit.p@sci.kmutnb.ac.th

Received June 2021; accepted August 2021

Abstract. *This research proposed a cultural heritage content retrieval by convolution neural network. The cultural heritage content in case of Thai's architecture was retrieved. In this research, the main contribution was to develop a classification algorithm for retrieving information from a cultural heritage image for telling a story inside that image. The cultural heritage in the case of Thai's archaeological site architecture was created, including the story of the archaeological site through the learning of machine learning and image processing. The interesting information inside the cultural heritage image was retrieved to present to the interested people who are interested in its contents. The appearance of the shape inside image can be used to distinct the characteristics of image, for example, the era, architecture and style of images. We analyze the experimental results of cultural heritage content retrieval by using the classification result by convolution neural network algorithm. In this research, the images of Thai's archaeological site architecture from world heritage provinces in Thailand were classified, for example, images from Phra Nakhon Si Ayutta province, Sukhothai province and Bangkok, which represent Ayutthaya era, Sukhothai era and Rattanakosin era in ordering. The proposed convolution neural network shows an accuracy of 80.83% in average.*

Keywords: Cultural heritage, Content retrieval, Convolution neural network, Machine learning, Image processing

1. **Introduction.** Cultural heritage is composed of tangible cultural heritage and intangible cultural heritage. Tangible cultural heritage is composed of the following categories: paintings, sculptures, monuments, and archaeological sites. The intangible cultural heritage is composed of the following categories: performing arts, rural dancing, and folklore. Cultural heritage tourism in the temple is now very popular with both Thai and foreign tourists. The tourism industry is another industry that will affect the income of the country. In addition, a tourism industry in the places where cultural heritage occurred will be required for the new generation to learn. Cultural heritage means buildings, monuments, and places that have historical value, aesthetics, archeology, or anthropology. Cultural heritage includes architecture, sculpture, painting, or natural archeology, such as caves or important places that may be works by human beings or because of natural and human works. The history of each country shows the ancestors greatness of those times through ancient monuments or antiquities. The passing of time causes those ancient ruins to deteriorate over time, leaving some parts of prosperity. As a result, understanding the history of cultural heritage or archaeological sites will have an impact on the pride of long history in that location for future generations in the country.

In this research, we aim to develop an effective cultural heritage content retrieval that enables the retrieval of stories that appear in those cultural heritages for the knowledge of

DOI: 10.24507/icicel.16.02.137

future generation's interests. For this reason, we realized that creating a cultural heritage information management system with convolution neural network, especially for Thai architecture by showing interesting information within images to present to interested people would be benefit. The development consists of generating stories from the cultural heritage images. We used the appearance of the shape for showing characteristics of the archaeological site and link to the era of the ancient monuments architecture that was created, including explaining the story of the archaeological site through machine learning and image processing. This research is a study and development of knowledge, which relies on specific data sets within Thailand.

We organize the rest of each section as follows. Theory of cultural heritage in case of architecture and convolution neural networks was introduced in Section 2. The classification algorithm for retrieving information from the cultural heritage images is detailed in Section 3. In Section 4, the results and discussions are explained; after that, we conclude the paper in Section 5.

2. **Related Work.**

2.1. **Cultural heritage.** The pagoda or stupa is one of Thailand's tangible cultural treasures that is highly known both within the country and abroad. The fundamental goal of constructing a stupa or pagoda is to collect religious symbols. Several stupa shapes or architectures exist, each of which is tied to the building's age. We can divide the main architecture of pagoda into three categories, which are Sukhothai architecture, Ayutthaya architecture, and Rattanakosin architecture. Sukhothai architecture can be divided into these styles: the bell-shaped style, the Prang style, etc. The Prang style is very common in Ayutthaya architecture. Finally, Rattanakosin architecture can be classified into these categories: the square wooden pagoda style, the Prang style, etc. The cultural heritage of each architecture can be shown in Figures 1-3.

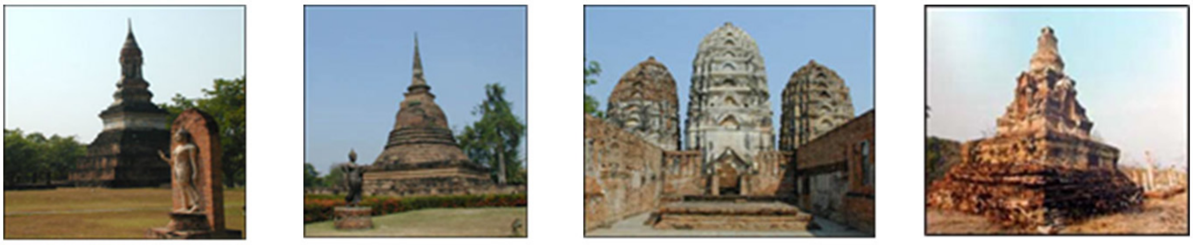

FIGURE 1. Examples of cultural heritage in Sukhothai architecture: the lotus blossom style, the bell-shaped style, the Prang style, and the Chomhar style [1]

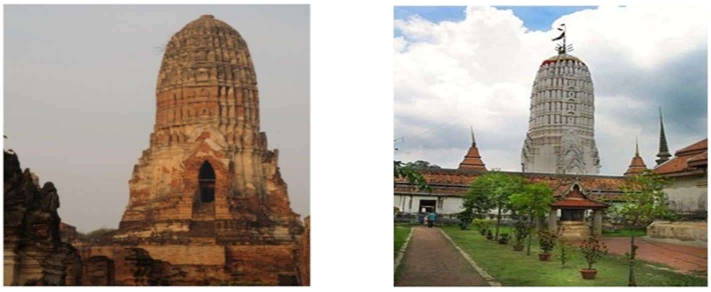

FIGURE 2. Example of cultural heritage in Ayutthaya architecture: the Prang style [1]

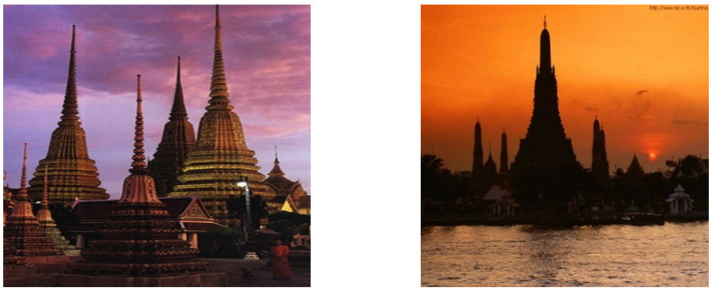

Figure 3. Examples of cultural heritage in Rattanakosin architecture: the square wooden pagoda style, the Prang style, etc. [1]

2.2. **Research on image classification and information retrieval.** In this section, we briefly explain the image classification and information retrieval. Nagano and Fukami [3] discussed the learning of skin characteristics by a convolution neural network. They used visual evaluation scores as training data and photos acquired with a microscope as input data. Saragih et al. [4] proposed research using machine learning to classify the freshness of the Ambarella fruit based on its color attribute. Liu [5] offered a CNN and wireless network-based image recognition model for intangible cultural heritage. Belhi et al. [6] introduced the system focusing on using new deep learning classification and annotation methods to automate annotation and metadata fulfillment. It also addresses concerns with physically damaged heritage artefacts using a new image reconstruction approach based on supervised and unsupervised learning to handle these issues. Ćosović and Janković [7] proposed a research in which deep learning neural network was used to classify photographs of architectural heritage into ten categories. For picture classification, the convolutional neural network was utilized, and the same architecture was applied to two sets of data. Kulkarni et al. [8] proposed platform to use Indian Digital Heritage Space (IHDS) monument data for their research. They introduced image classification and query-based retrieval of image labels by using transfer-learning algorithms. Obeso et al. [9] made use of prehispanic, colonial, and modern architectural heritage photographs created from video content in Mexico. Images have unwanted aspects added to them because of the image derivation technique. As a result, in order to increase classification rates, training of classification models should focus on relevant architectural content inside the image. Karpathy and Li [10] explained an algorithm for generating natural language descriptions of image regions. Farhadi et al. [11] developed a new system that can compute a score linking an image to a sentence. The value used the score linking an image to attach a descriptive sentence to a given image. They suggested the model for simulating the relationship between image space, mean space, and sentence space. Socher et al. [12] developed the DT-RNN model which is a method for finding the relationship between an image and sentences by using recursive neural network. The DT-RNN model uses dependency trees to process sentences into a vector space for retrieving images. Other technique that uses for searching a large image database is based on content-based image retrieval, which is based on color, texture, or shapes. However, using low-level picture features and high-level image semantics has an impact on the algorithms' performance. [13] used color histograms to image index. The weight distribution of the Gaussian was used in the experiments. [14] used image retrieval algorithm with color histogram using vector modeling to convert the image into a histogram and store it in the database. To retrieve the image from database, the image histogram search was performed by comparing the histogram between images and histogram of the image in the database using the similarity in the vector model. To measure the similarity between searching image and reference image, the similarity value was used, in which the similarity score closest to 1 means these two images are similar. From the above research, several issues were difficult to perform that affect the accuracy of algorithms, and the current problem is how to add the caption of image which has its ability as same as the aspect of human ability. According to these several issues, in this research, we proposed an algorithm for retrieving contents in the images. A cultural heritage content retrieval by convolution neural network was introduced. The cultural heritage content in case of Thai's architecture was retrieved. In this research, the main contribution was to develop a classification algorithm for retrieving information from a cultural heritage image for telling the story inside that image.

2.3. Convolution neural network.

Convolution neural network (CNN) is another type of feed-forward artificial neural network. The main objective of CNN is to develop a machine that can learn and predict the interested object. CNN differs from the original artificial neural network in that it includes a hidden layer in addition to the artificial neuron network [2]. The architecture of CNN consists of multiple layers which are convolutional layers and subsampling layers and one or more fully connected layers and Softmax layer [2]. One example of using CNN is face recognition learning. This algorithm is able to learn large face dataset, which can be applied in many different tasks, such as identifying individual faces, friend tagging on Facebook, face authentication in mobile, or used in the unmanned vehicle system [2].

3. Cultural Heritage Content Retrieval.

To perform the process of cultural heritage content retrieval, the following algorithms were executed.

```
Algorithm: Cultural Heritage Content Retrieval
1. Image input: Take the image of Thai cultural heritage that you wish to know more
   about. The database contains the training Thai cultural heritage image, which will
   be compared to the input Thai cultural heritage image.
2. Pre-processing: Use the picture improvement algorithm to improve the image quality
   after converting the RGB Thai cultural heritage image to grayscale.
3. Edge detection: The following step is used for checking the edges that passes
   through or near to the interested point. It is calculated by measuring the different
   intensity of the nearest points or finding the line surrounding the object inside
   image. One problem that will occur in the edge detection algorithms is to find the
   edge in the low quality image that has the low difference between foreground and
   background or the brightness does not cover to all of image. We used Laplacian
   method to find the edge of image.
4. Feature extraction: To retrieve the identity of each image to be a vector for using
   in the training and testing processes, feature extraction was used. In this step, we
   used convolution neural network algorithm for the process to get key points inside
   image.
5. Train image with convolution neural network.
6. Match the most similar image between the input image and the reference image in
   the database.
7. Generate cultural heritage descriptions.
End
```

The above algorithms are steps for getting the description of cultural heritage. In the feature extraction step, the convolution neural network was applied. The CNN architecture used in our research was set up with the following details: the convolution neural network architecture consists of an input, an output layer, and multiple hidden layers. Details of the hidden layers of a CNN consist of a series of convolutional layers that convolve with a multiplication or other dot product. The activation function is commonly a RELU layer and is subsequently followed by additional convolutions such as pooling layer, fully connected layers, and normalization layer, referred to as hidden layers. The final convolution often involves backpropagation to more accurately weigh the product. The process inside the convolutional layers convolves the input and passes its result to the next layer. Each convolutional neuron processes data only for its receptive field. Convolutional networks are composed of local or global pooling layers to streamline the underlying computation. The dimension of the data was reduced by pooling layers by combining the outputs of neuron clusters at one layer into a single neuron in the next layer. Local pooling combines small clusters, typically 2 × 2. Global pooling acts on the neurons of the convolutional layer. In addition, pooling may compute a max or an average. Max pooling uses the maximum value from each of a cluster of neurons at the prior layer. Fully connected layers connect every neuron in one layer to every neuron in another layer. It is in principle the same as the traditional multilayer perceptron neural network (MLP). Each neuron in a neural network computes an output value by applying a specific function to the input values coming from the receptive field in the previous layer. A vector of weights and bias determines the function that is applied to the input values. Learning in a neural network progresses by making iterative adjustments to these biases and weights. The last step for getting the description of cultural heritage is the matching process, which is the process for matching the most similar image between the input image and the reference image. In this process, a group of data, which have similar characteristics, were retrieved; several images in the group will occur. After that, the process of generating the description after the matching process will be performed. Finally, the algorithm will show and set description to the input image.

4. Results and Discussion.

4.1. **Dataset collection and description.** To verify the experimental results of cultural heritage content retrieval, the classification result by using the proposed algorithm to classify the architecture of Thai cultural heritage image was performed. We collect the dataset from the cultural heritage in Bangkok, Sukhothai province, and Phra Nakhon Si Ayutta Province, which is the UNESCO cultural heritage in Thailand. For the better accuracy of the algorithms, the image which is used in the experiments is the cultural heritage image in Ayutthaya era, Sukhothai era, and Rattanakosin era. The number of images used in the research was shown in Table 1.

Table 1. Number of cultural heritage architectures in the experiments

Cultural heritage architecture	Number of images
Ayutthaya architecture	1670
Sukhothai architecture	1440
Rattanakosin architecture	1460
All	**4570**

4.2. **Performance indexed.** To reflect the performance of classification results, one of the most powerful techniques is using confusion matrix in which each row shows the predicted class and each column shows the original class. This technique can be used to visualize the classification error. The experimental results of our proposed algorithm are shown in Tables 2-4.

4.3. **Comparing algorithms.** The following algorithms were used for comparing the performance of the proposed algorithms: convolution neural network, KNN algorithm, neural network, and SIFT with Euclidean distance algorithm. We compare the efficiency of generating descriptions which use only the classification results.

Table 2. Confusion matrix of the proposed algorithm by using convolution neural network

Predicted class	True class (Architecture)		
(Architecture)	Ayutthaya	Sukhothai	Rattanakosin
Ayutthaya	1250	260	160
Sukhothai	105	1320	15
Rattanakosin	217	93	1150

Table 3. Confusion matrix of the KNN algorithm

Predicted class	True class (Architecture)		
(Architecture)	Ayutthaya	Sukhothai	Rattanakosin
Ayutthaya	987	408	275
Sukhothai	225	1052	163
Rattanakosin	307	101	1052

Table 4. Confusion matrix of neural network and SIFT with Euclidean distance algorithm

Predicted class	True class (Architecture)		
(Architecture)	Ayutthaya	Sukhothai	Rattanakosin
Ayutthaya	1025	385	260
Sukhothai	369	962	109
Rattanakosin	350	125	985

Table 5. The precision, recall, F1-score of the proposed algorithms: convolution neural network, KNN algorithm, neural network, and SIFT with Euclidean distance algorithm

Class	Convolution neural network			KNN			Neural network and SIFT with Euclidean distance algorithm		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
Ayutthaya	0.7485	0.7952	0.7711	0.5910	0.6498	0.6190	0.6137	0.5877	0.6004
Sukhothai	0.9167	0.7890	0.8481	0.7306	0.6739	0.7011	0.6680	0.6535	0.6607
Rattanakosin	0.9167	0.8679	0.8916	0.7306	0.7060	0.7180	0.6680	0.7274	0.6965

4.4. **Experimental results.** In this research, we show the cultural heritage content retrieval performance by using the classification results by using the proposed algorithm to classify the architecture of Thai cultural heritage image. The accuracy of our image description generator depends on the classification accuracy. The confusion matrix of the tested data being classified to the original cultural heritage image was shown inside Tables 2-4. The precision, recall, F1-score of the proposed algorithms, convolution neural network, KNN algorithm, neural network, and SIFT with Euclidean distance algorithm, were shown inside Table 5. The F1-score of the proposed algorithm which is the convolution neural network gives the accuracy 0.7711, 0.8481, and 0.8916 in Ayutthaya architecture, Sukhothai architecture, and Rattanakosin architecture. The F1-score of using KNN algorithms gives the accuracy 0.6190, 0.7011, and 0.7180 in Ayutthaya architecture, Sukhothai architecture, and Rattanakosin architecture. The F1-score of using neural network and SIFT with Euclidean distance algorithm gives the accuracy 0.6004, 0.6607, and 0.6965 in Ayutthaya architecture, Sukhothai architecture, and Rattanakosin architecture. From the experimental results in this research, the proposed algorithm for retrieving contents in the images can provide the accuracy 80.83%. A cultural heritage content retrieval by convolution neural network was introduced and outperformed other methods. The cultural heritage content in case of Thai's architecture was retrieved. Accordingly, to develop a classification algorithm for retrieving information from a cultural heritage image for telling the story inside that image is benefit for the person who is interested in this cultural heritage image.

5. **Conclusions.** We have developed a cultural heritage content retrieval by convolution neural network. The research will use the cultural heritage information: case study, the architecture of Thai architecture. The interesting information inside the photo will be retrieved to present to the interested. The development consists of telling stories from photos. It can bring the appearance of the shape, the unique character of the site to link the era of the historic site. The architecture was created including the story of the archaeological site through the learning of machine learning and image processing. In the future, the goal is to categorize cultural heritage in Southeast Asian countries that have an abundance of architecture using deep learning methods such as CNN. In addition, if we can show the complete shape of the past from that remains by analyzing the similarity of the image in the database, the image can tell the story of the past.

Acknowledgment. This research was funded by King Mongkut's University of Technology North Bangkok [Contract No. KMUTNB-62-KNOW-026]. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] C. Chareonla, *Buddhist Arts of Thailand*, Buddha Dharma Education Association Inc., 1981.

[2] L. Deng and D. Yu, Deep learning: Methods and applications, *Foundations and Trends® in Signal Processing*, vol.7, nos.3-4, pp.197-387, DOI: 10.1561/2000000039, 2013.

[3] M. Nagano and T. Fukami, Development of a skin texture evaluation system using a convolutional neural network, *International Journal of Innovative Computing, Information and Control*, vol.16, no.5, pp.1821-1827, 2020.

[4] R. E. Saragih, D. Gloria and A. J. Santoso, Classification of Ambarella fruit ripeness based on color feature extraction, *ICIC Express Letters*, vol.15, no.9, pp.1013-1020, 2021.

[5] E. Liu, Research on image recognition of intangible cultural heritage based on CNN and wireless network, *J. Wireless Com. Network*, DOI: 10.1186/s13638-020-01859-2, 2020.

[6] A. Belhi, A. Bouras, A. K. Al-Ali and S. Foufou, A machine learning framework for enhancing digital experiences in cultural heritage, *Journal of Enterprise Information Management*, DOI: 10.1108/JEIM-02-2020-0059, 2020.

[7] M. Ćosović and R. Janković, CNN classification of the cultural heritage images, *The 19th International Symposium INFOTEH-JAHORINA*, Jahorina, Bosnia and Herzegovina, 2020.

[8] U. Kulkarni, S. M. Meena, S. V. Gurlahosur and U. Mudengudi, Classification of cultural heritage sites using transfer learning, *2019 IEEE 5th International Conference on Multimedia Big Data (BigMM)*, pp.391-397, 2019.

[9] A. M. Obeso, M. S. G. Vazquez, A. A. R. Acosta and J. Benois-Pineau, Connoisseur: Classification of styles of Mexican architectural heritage with deep learning and visual attention prediction, *Proc. of the 15th Int. Workshop on Content-Based Multimedia Indexing*, Florence, Italy, pp.1-7, 2017.

[10] A. Karpathy and F. Li, Deep visual-semantic alignments for generating image descriptions, *Proc. of the 2015 IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, pp.3128-3137, DOI: 10.1109/CVPR201572989328, 2015.

[11] A. Farhadi, S. M. M. Hejrati, M. A. Sadeghi, P. Young, C. Rashtchian, J. Hockenmaier and D. A. Forsyth, Every picture tells a story: Generating sentences from images, in *Computer Vision – ECCV 2010. ECCV 2010. Lecture Notes in Computer Science*, K. Daniilidis, P. Maragos and N. Paragios (eds.), Heraklion, Crete, Greece, Springer, DOI: 10.1007/978-3-642-15561-1_2, 2010.

[12] R. Socher, A. Karpathy, Q. V. Le, C. D. Manning and A. Y. Ng, Grounded compositional semantics for finding and describing images with sentences, *Transactions of the Association for Computational Linguistics*, vol.2, pp.207-218, 2014.

[13] P. Kulikarnratchai and O. Chitsoput, Image retrieval using color histogram in HSV color sampler, *The 29th Electrical Engineering Symposium (EECON-29)*, pp.1029-1032, 2006.

[14] A. Sangswang, Image search by histogram comparison using vector models, *The 2nd National Conferences Benjamit Academic*, 2012.