Computer Systems Science & Engineering
DOI: 10.32604/csse.2023.025293
Article

Tech Science Press

Toward Fine-grained Image Retrieval with Adaptive Deep Learning for Cultural Heritage Image

Sathit Prasomphan*

Department of Computer and Information Science, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, 1518 Pracharat 1 Rd., Wongsawang, Bangsue, Bangkok, 10800, Thailand
*Corresponding Author: Sathit Prasomphan. Email: sathit.p@sci.kmutnb.ac.th

Received: 19 November 2021; Accepted: 09 February 2022

Abstract: Fine-grained image classification is a challenging research topic because of the high degree of similarity among categories and the high degree of dissimilarity for a specific category caused by different poses and scales. A cultural heritage image is one of the fine-grained images because each image has the same similarity in most cases. Using the classification technique, distinguishing cultural heritage architecture may be difficult. This study proposes a cultural heritage content retrieval method using adaptive deep learning for fine-grained image retrieval. The key contribution of this research was the creation of a retrieval model that could handle incremental streams of new categories while maintaining its past performance in old categories and not losing the old categorization of a cultural heritage image. The goal of the proposed method is to perform a retrieval task for classes. Incremental learning for new classes was conducted to reduce the re-training process. In this step, the original class is not necessary for re-training which we call an adaptive deep learning technique. Cultural heritage in the case of Thai archaeological site architecture was retrieved through machine learning and image processing. We analyze the experimental results of incremental learning for fine-grained images with images of Thai archaeological site architecture from world heritage provinces in Thailand, which have a similar architecture. Using a fine-grained image retrieval technique for this group of cultural heritage images in a database can solve the problem of a high degree of similarity among categories and a high degree of dissimilarity for a specific category. The proposed method for retrieving the correct image from a database can deliver an average accuracy of 85 percent. Adaptive deep learning for fine-grained image retrieval was used to retrieve cultural heritage content, and it outperformed state-of-the-art methods in fine-grained image retrieval.

Keywords: Fine-grained image; adaptive deep learning; cultural heritage; image retrieval

1 Introduction

Cultural heritage refers to buildings, monuments, and places that have historical value, aesthetics, archeology, and anthropology. Cultural heritage includes architecture, sculpture, painting, or natural archeology, such as caves or important places that may be works by humans or because of natural and human work. Cultural heritage can be divided into two groups: tangible and intangible. Examples of tangible cultural heritage sites include murals, statues, temples, and ancient monuments. Performing arts, rural dancing, and folklore are examples of intangible cultural heritage [1–4]. Tourists are increasingly interested in cultural heritage tourism at various locations [4]. Through ancient monuments and antiquities, each country's history demonstrates the majority of its ancestors at all times. In addition, with time, the ancient ruins deteriorated, leaving some places of prosperity. Consequently, knowing the history of cultural heritage or archaeological sites gives an impression of the pound of strong history in that location for the country's next generation [4].

The following ten main industries, which can be grouped into two categories, have the potential to drive the country's economic growth engine and increase competitiveness. The first category includes modern transportation, smart electronics, tourism, agriculture, biotechnology, and food. The second group includes five future industries: industrial automation, aviation and logistics, biochemical industry, digital industry, and comprehensive medical industry, all of which are emerging businesses in which Thailand has competitive potential and investors are interested (Medical Hub). Tourist business is another target industry that will have an impact on the country's economy. In addition, for the younger generation to learn, a tourism sector in the regions where cultural heritage occurred will be required [2–4].

The motivation of this research is, at present, fine-grained image classification is a challenging research topic due to the high degree of similarity among categories and the high degree of dissimilarity for a specific category caused by different poses and scales [5–8]. A cultural heritage image is one of the fine-grained images because each image has the same similarity in most cases. Using the classification technique, distinguishing cultural heritage architecture may be difficult. Accordingly, this study proposes a cultural heritage content retrieval method using adaptive deep learning for fine-grained image retrieval (ADLFIG). The key contribution was to create an up-to-date retrieval model that could handle incremental streams of new categories while maintaining its past performance on old categories without losing sight of a cultural heritage image. Cultural heritage in the case of Thai archaeological site architecture was retrieved through machine learning and image processing.

The remainder of each section is organized as follows: The theory of cultural heritage in the case of architecture, image retrieval techniques, and fine-grained image retrieval is introduced in Section 2. The retrieval of cultural heritage content using adaptive deep learning for fine-grained image retrieval (ADLFIG) is detailed in Section 3. The experimental setup is detailed in Section 4. The results and discussion are explained in Section 5. Finally, in Section 6, we summarize the paper.

2 Related Theory

2.1 Cultural Heritage

The pagoda or stupa is one of Thailand's tangible cultural heritage sites that is widely famous both in the country and abroad. The primary objective of building a stupa or pagoda is to gather religious iconographies. Several pagoda shapes and architectures exist, each of which can be related to the era of the building. The major architecture of pagodas can be divided into three categories: Sukhothai architecture (Buddhist Era 1792–2006), Ayutthaya architecture (Buddhist Era 1893–2310), and Rattanakosin architecture (Buddhist Era 2325–present). In each era, the architecture of the pagoda has its sub-categories for example, in Sukhothai, architecture can be divided into these styles: the Prang style, the Haripunchai style, and the others present Sukhothai and Lankan art styles, the bell-shaped pagoda style, etc. The Prang style is common in Ayutthaya architecture.

Finally, the architecture of Rattanakosin can be divided into the following categories: Prang style, square wooden pagoda style, and so on [1–4,9]. Figs. 1–3 depict the cultural heritage of each architectural style. From the main categories and sub-categories of the pagoda which mostly occur in a similar style, it is difficult to retrieve the correct image in the database. Using a fine-grained image retrieval technique for this group of cultural heritage images in a database can solve this problem.

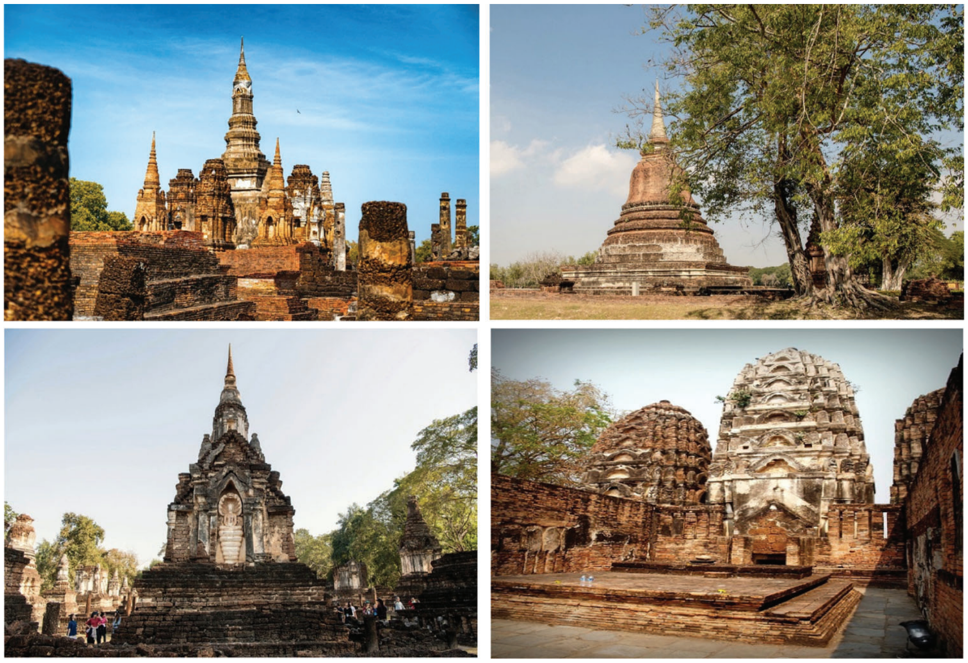

Figure 1: Sukhothai architectural example [1–4]

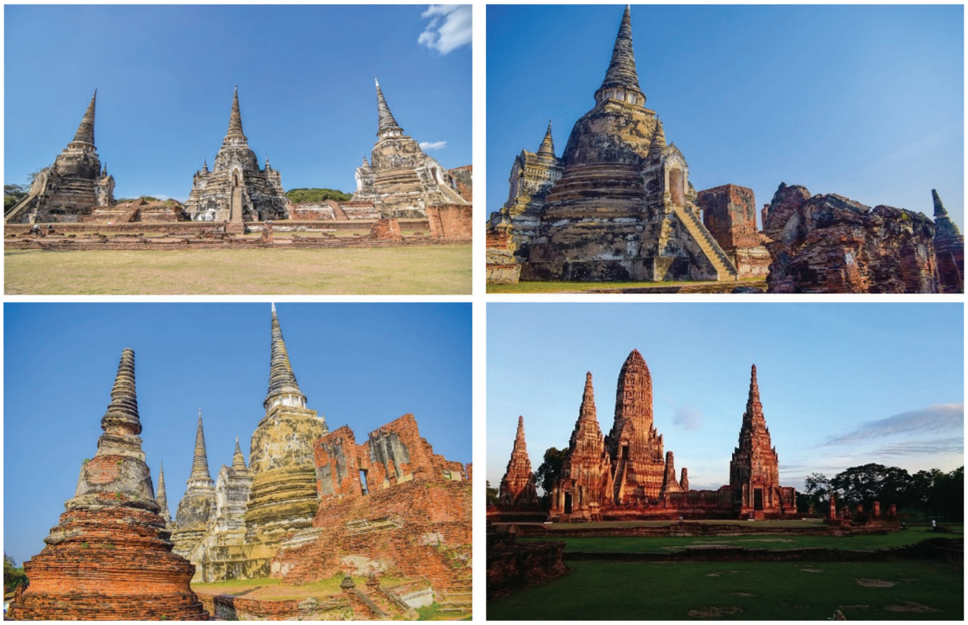

Figure 2: Ayutthaya architectural example [1–4]

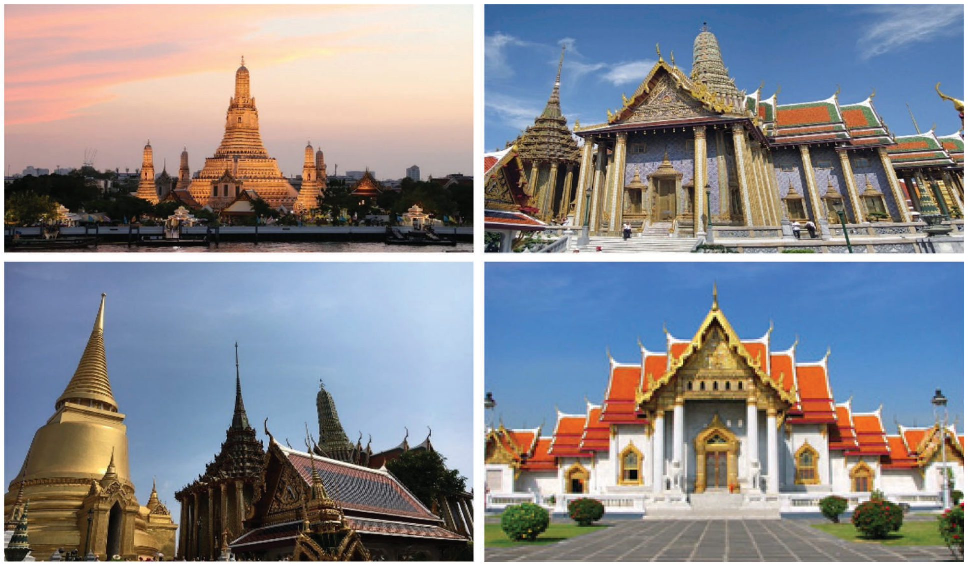

Figure 3: Rattanakosin architecture example [1–4]

2.2 Image Retrieval Techniques

In this section, we briefly discuss relevant image retrieval studies. Previously, image retrieval strategies relied on obtaining picture file names or using an image caption, which is not a component of the image itself [1–4]. Consequently, it is not possible with the existing massive database because each user's description of a photo is unique and may not always be relatable [1–4]. Another problem that can occur in a search image in a huge dataset is the use of a text search, which is more difficult if a few words of an image are given in that lage database. Moreover, seaching an image from a large number of images that contain the same class, such as a car, bird, building, or tree, can yield incorrect information about the results. When working with large databases, it is also necessary to rely on humans to manually classify images, which takes time.

Another technique for image retrieval uses the color, shape, and texture of an image for a query inside an image database. We refer to the features of an image as image characteristics. Image indexing is a technique that is used for image retrieval. This is a technique for storing and creating a unique vector within each image to be stored in a database. Each image's characteristics, such as color histogram, color, texture, and shape can be stored in a database and used for the query process. Most content-based picture retrieval relies on feature extraction of the characteristics of these image [10,11].

Several researchers have proposed techniques for image retrieval from a large database. Uday et al. [12] proposed a platform for image classification and retrieval using transfer-learning algorithms. Indian Digital Heritage Space (IHDS) monument data were used for their research. Prasomphan et al. [2] proposed a methodology for synthesizing stupa image descriptions. Key points obtained by scale-invariant feature transform (SIFT) algorithms were used to determine the stupa age, architecture, and other descriptions, and artificial neural networks were used to learn stupa descriptions from the generated key points. Relevant information from the query results is displayed on a mobile application.

Latif et al. [13] reviewed content-based image retrieval (CBIR) and image representation enhancements in recent years. Belhi et al. [14] introduced an image reconstruction technique and image annotation using a combination of deep learning classification techniques. Using different distance metrics, a hybrid feature-based efficient content-based image retrieval system was proposed in [15]. Color auto-correlograms, color moments, and hue, saturation, value (HSV) histogram features are utilized in the spatial domain, whereas frequency-domain features, such as stationary wavelet transform (SWT) moments and Gabor wavelet transform features, are employed in the frequency domain. Color and edge directivity descriptor features are also used to improve the statistical image features of precision binaries to construct an effective content-based image retrieval system. M. K. Bashir [16] used a semantic labels and binary hash codes from convolutional autoencoders for image retrieval. Y. Rao and W. Liu [17] reported a region-division-based image retrieval approach that integrates a low-level local color histogram and texture features. In [18], the authors developed a pairwise constraint propagation-based interactive medical image search system. The core strategy is to take pairwise constraints from the user feedback and transmit them across the full image set to reconstruct the similarity matrix. They then ranked the queried images using this matrix. In [19], the choice of the best features to be utilized in image categorization was introduced. The fuzzy artmap (FAM) classifier was adapted, called the non-proliferation fuzzy artmap (NPFAM). In [20], a content-based picture retrieval system was built using an approach that combines several shape, color, and relevance feedback elements. M. Nagano and T. Fukami [21] discussed how a convolutional neural network can learn skin features using pictures taken with a microscope as input data and visual evaluation scores as training data. R. E. Saragih [22] proposed utilizing machine learning to classify the freshness of Ambarella fruits based on their color. For intangible cultural assets, E. Liu [23] presented a convolution neural network and wireless network based image recognition approach.

Content-based image retrieval methods were proposed in [24]. The goal of this method is to learn all features of an image without using the label value. Using a modified U-net model with no label information, the proposed framework generates hash codes of appropriate length. These results are extremely promising and competitive with those obtained using other methods. In [25], the medical image retrieval was introduced. Images of the body parts were used to analyze the disease. Using image features, the proposed method aims to generate the most effective and least parameterized hash codes. The convolutional neural network architecture was used to extract deep characteristics from the medical pictures. In [26], content-based image retrieval methods were introduced by searching for points with the most similar content to query features from within a large dataset. In their study, they proposed an approximate nearest neighbor (ANN) search. The low-dimension feature representation was focused. Deep learning based approaches are used for feature extraction in hashing methods.

From the above methods, using content-based image retrieval for the high degree of similarity among categories and a high degree of dissimilarity for a specific category may be difficult.

2.3 Fine-grained Image Retrieval

In computer vision, fine-grained image retrieval is considered the most significant issue because there is less variation among classes but a great variation within the same class. Therefore, it is more difficult than content-based image retrieval [27–31]. Recently, several studies on fine-grained image retrieval techniques have been proposed. A convolution neural network was applied in the retrieval process. The loss function was applied and modified to improve the network performance. In [27], a piecewise cross entropy loss function was introduced. A convolution neural network model was introduced for retrieval performance in fine-grained images. In this study, the model is trained with the decorrelated global piecewise centralized ranking loss (DGPCRL) and can be regarded as a general classification model with a normalize-scale layer. H. Xu [28] provided a revolutionary approach to fine-grained image classification that can employ useful information from either organized knowledge sources or unstructured text. They suggested a visual-semantic embedding model that investigates semantic embedding using knowledge bases and text, then trains a novel end-to-end convolution neural network framework to map picture features linearly to a rich semantic embedding space. The experimental results on a large-scale dataset were reported. A significant improvement was observed in these studies. F. Li [29] presented sketch-based image retrieval (SBIR) for fine-grained image retrieval problem. The described the process of using a freehand scenario sketch to retrieve scene photos from a large database that meets a user's particular needs. In their study, a graph-based method was used to learn the similarity between images in a database and scene sketches. W. Chen [30] examined the problem of fine-grained picture retrieval in an incremental scenario, where new categories were introduced over time. The process of transferring gained knowledge from an original model to an incremental one is known as incremental learning. Image classification, image synthesis, object identification, hashing image retrieval, and semantic segmentation are some applications that have been studied [30]. D. Wu [31] introduced a deep incremental hashing network (DIHN), a new deep hashing system for incrementally learning hash codes. The hash codes of the new incoming images were generated directly using the deep incremental hashing network (DIHN). Subsequently, the similarity between the training images and query image was calculated. The results were experimented using a benchmark database. The efficiency of their algorithms showed that they outperformed the traditional method. However, old groups of images were not considered in the hash code learning process. In [32], a framework for content-based fine-grained image retrieval (CB-FGIR) by using a convolutiion neural network was introduced. They experimented the proposed framework with Oxford flower-17 dataset. They reported that the CB-FGIR framework achieved high accurate retrieval results. To produce compact binary codes for fine-grained images, Q. Cui [33] proposed a fine-grained hashing topic. ExchNet is a single end-to-end trainable network that they proposed. In [7,8,30,31,33–35], a fine-grained image retrieval technique using a convolution neural network was introduced. They applied algorithms with several domains, such as resnet architectures for fine-grained vehicle classification [8] and fine-grained fashion similarity [34].

In [34] fine-grained fashion similarity was introduced. Similarity searching between images was performed using an attribute-specific embedding network (ASEN). Multiple attribute-specific embeddings were used. More fine-grained features can be extracted by using the attributes provided.

Based on the above research, it was difficult to determine the effect of several issues on the accuracy of the algorithms. Based on these issues, the key contribution of this research was to create a retrieval model that could handle incremental streams of new categories while maintaining its past performance on old categories without losing the old categorization of a cultural heritage image. In this research, we compare the performance of the proposed algorithms with attribute-specific embedding network (ASEN) technique and decorrelated global piecewise centralized ranking loss (DGPCRL) technique. The reason for using these two algorithms is that they provide high accuracy and are state-of-the-art in this fine-grained image retrieval technique.

3 Cultural Heritage Content Retrieval by Adaptive Deep Learning for Fine-Grained Image Retrieval

To perform the process of cultural heritage content retrieval using adaptive deep learning for fine-grained image retrieval (ADLFIG), the following algorithms were executed.

3.1 Problem Statement

Let fine-grained image that we used to train for adaptive deep learning for fine-grained image retrieval (ADLFIG), which have m class of images denoted with $\boldsymbol{X} = \{(\boldsymbol{xi}, \boldsymbol{yi}\ )\}$ where $\boldsymbol{i} = \boldsymbol{1}, \ldots, \boldsymbol{m}$ and each image is annotated with one of the C fine-grained class labels denoted with $\boldsymbol{Y} = \{\boldsymbol{y_1}, \boldsymbol{y_2}, \boldsymbol{y_3}, \ldots,\ \boldsymbol{yC}\ \}$. The goal of our proposed method is to perform a retrieval task for m classes. Incremental learning was performed for n new classes to reduce the re-training process. In this step, the original class is not necessary for re-training which we call the adaptive deep learning technique proposed in this research.

3.2 Cultural Heritage Content Retrieval by Adaptive Deep Learning for Fine-Grained Image Retrieval

The input in this framework is composed of three inputs, which is the original database image (D_1), the incremental database image (D_2) and the query images(D_3). At first, the original database image (D_1) was train using convolution neural network to determine the mapping function $f1 : X \to Y$. Next, the incremental database image (D_2) was trained using the pre-trained network from the original database image. In this step, the new fine-grain image class was trained to calculate the new mapping function by including the increment database image in the model denoted as $f2 : X' \to Y$'. Finally, a query image was used to extract the feature and retrieve its class from the $f2$ function. The following algorithms were used to retrieve the images from the database.

ALGORITHM: *Adaptive Deep Learning for Fine-Grained Image Retrieval (ADLFIG)*

```
Input:
    Fine-grain images with $\boldsymbol{X} = \{(\boldsymbol{xi}, \boldsymbol{yi})\}$ where i = 1,…, m and each image is labeled with one of C fine-grained class labels denoted by $\boldsymbol{Y} = \{\boldsymbol{y_1}, \boldsymbol{y_2}, \boldsymbol{y_3}, \ldots, \boldsymbol{yC}\}$.
    The original database image $(D_1)$
    The incremental database image $(D_2)$
    The query images $(D_3)$
Output:
    Fine-Grain Image class
Process:
for $i = 1 \to$ number of data in database $(D_1)$ do
    1. Extract image feature with convolution neural network algorithm.
    2. Train image in the original database image $(D_1)$ with convolution neural network to find the mapping function $f1 : X \to Y$.
    3. Store the parameter from the original database image training.
end
The incremental database image $(D_2)$ was trained using a pretrained network from the original database image
The new fine-grain image class was trained to calculating the new mapping function by including the increment database image in the model denoted as $f2 : X' \to Y$'.
Calculate the similarity score between the query image and reference images in the original database images $(D_1)$ and incremental database images$(D_2)$.
```

The above algorithms are steps for obtaining an image from the database using *adaptive deep learning for fine-grained image retrieval (ADLFIG)*.

The process of the feature extraction steps has the following details. We used a convolution neural network to extract the important features in the fine-grained images. The architecture of the convolution neural network was set up as follows: The Restnet50 architecture was performed. The activation function in each layer is Rectified Linear Unit (ReLU). We set the size of the image as the input to the network to 224 × 224 × 3. The RGB image was used. The hidden layer is composed of convolutional and pooling layers. The final layer is the output layer used for image classification.

Within the convolutional layers, the input is convolved and the result is passed on to the next layer.

In our proposed algorithms, the four blocks of convolutional layers were set up. Subsequently, filters with a size of 3 × 3 were used. We used 6 layers, 8 layers, 12 layers and 6 layers in each block in ordering.

4 Experimental Setup

4.1 Dataset Collection and Description

To verify the experimental results of cultural heritage content retrieval by adaptive deep learning for fine-grained image retrieval, the suggested technique was used to categorize the style of Thai cultural heritage images, and the classification results in terms of accuracy, precision, and recall were obtained. In this study, pagodas were obtained in several styles. In each era, the architecture of the pagoda has its subcategories; for example, in Sukhothai, architecture can be divided into these styles: the Prang style, Haripunchai style, and the others present Sukhothai and Lankan art styles, bell-shaped pagoda styles, etc. The Prang style is common in Ayutthaya architecture. Finally, architecture in Rattanakosin can be divided into the following categories: Prang style, square wooden pagoda style, etc. We collected the dataset from the cultural heritage or archaeological site in Bangkok, Sukhothai Province, and Phra Nakhon Si Ayutthaya Province, which is the United Nations Educational, Scientific and Cultural Organization (UNESCO) cultural heritage in Thailand. Tab. 1 shows the total number of images used throughout the study.

Table 1: Total number of images used throughout the study

Cultural heritage architecture	Number of image
Prang style	1670
Haripunchai style	1440
Lankan art style	1460
Bell-shaped pagoda style	1480
Square wooden pagoda style	1500
All	**7550**

4.2 Performance Indexed

One of the most powerful techniques for reflecting the performance of classification results is the use of a confusion matrix. In this technique, each row shows the predicted class and each column shows the original class. This technique can be used to visualize classification errors. We calculated the precision, recall, f-1 score and accuracy using the following equation:

$$\boldsymbol{Precision} = \frac{TP}{TP + FP} \tag{1}$$

where TP is the true positive value and FP is the false positive value. Precision is the value that measures the number of correct predicting answers by dividing by the total number of images in the dataset.

$$\boldsymbol{Recall} = \frac{TP}{TP + FN} \tag{2}$$

where TP is the true positive value and FN is the false negative value. Recall is the value that measures the number of correct predicting answers for each class and is divided by the total number of that class which is the ground truth.

$$\boldsymbol{F1 - Score} = 2\frac{PrecisionxRecall}{Precision + Recall} \tag{3}$$

where Precision is the precision value calculated from (1) and Recall is the recall value calculated from (2).

$$\boldsymbol{Accuracy} = \frac{TP + TN}{N} \tag{4}$$

where TP is the true positive value, TN is the true negative value, and N is total number of images.

Recall@K is another performance metric that we used to assess retrieval performance. The average recall scores across all query photos in the test set were calculated as Recall@K. We specifically returned the top K-related photos for each query. If there is at least one positive image among the top K returning photos, the score is 1: otherwise, it is 0.

4.3 Comparing Algorithms

To compare the performance of the proposed methods, the following algorithm was employed by adaptive deep learning for fine-grained image retrieval (ADLFG), decorrelated global piecewise centralized ranking loss (DGPCR), attribute-specific embedding network (ASEN). We compare the efficiency of algorithms by using the classification results as shown in Tab. 2.

Table 2: Precision, recall, f-1 score of the proposed algorithm by using adaptive deep learning for fine-grained image retrieval (ADLFG), decorrelated global piecewise centralized ranking loss (DGPCR), attribute-specific embedding network (ASEN)

Class	Accuracy			Recall			Precision			F1-Score		
	ADLFG	DGPCRL	ASEN	ADLFG	DGPCRL	ASEN	ADLFG	DGPCRL	ASEN	ADLFG	DGPCRL	ASEN
Prang style	**0.89**	0.81	0.79	**0.89**	0.82	0.79	**0.86**	0.80	0.86	0.88	0.81	0.82
Haripunchai style	0.73	**0.80**	0.69	**0.73**	0.65	0.68	0.78	**0.85**	0.67	**0.75**	0.74	0.67
Lankan art style	0.75	0.70	**0.80**	0.75	0.69	**0.80**	0.70	0.76	**0.79**	0.72	0.72	**0.79**
Bell-shaped pagoda style	0.60	0.70	**0.80**	0.61	0.68	**0.70**	0.70	0.80	**0.85**	0.65	0.74	**0.77**
Square wooden pagoda style	**0.80**	0.78	0.75	**0.81**	0.75	0.69	**0.75**	0.71	0.73	**0.77**	0.73	0.71

5 Result and Discussion

5.1 Experimental Results

In this research, we have developed a cultural heritage content retrieval by adaptive deep learning for fine-grained image retrieval. The classification results by using precision, recall and F1-score is presented. The cultural heritage content retrieval performance is shown by using the classification results. To compare the performance of the proposed methods, the following algorithm was employed by adaptive deep learning for fine-grained image retrieval (ADLFG), decorrelated global piecewise centralized ranking loss (DGPCR), and an attribute-specific embedding network (ASEN).

The accuracy, precision, recall, and F1-score of the adaptive deep learning for fine-grained image retrieval (ADLFG) is shown in Tab. 2. The accuracy of the adaptive deep learning for fine-grained image retrieval (ADLFG) of Prang style, Haripunchai style, Lankan art style, Bell-shaped pagoda style, and square wooden pagoda style are 0.89, 0.73, 0.75, 0.60, and 0.80, respectively. The accuracy of the decorrelated global piecewise centralized ranking loss (DGPCR) is shown in Tab. 2. The accuracy of Prang style, Haripunchai style, Lankan art style, Bell-shaped pagoda style, and square wooden pagoda style are 0.81, 0.80, 0.70, 0.70, and 0.78, respectively. The accuracy of the attribute-specific embedding network (ASEN) is shown in Tab. 2. The accuracy of Prang style, Haripunchai style, Lankan art style, Bell-shaped pagoda style, and square wooden pagoda style are 0.79, 0.69, 0.80, 0.80, and 0.75, respectively.

The F1-score of the adaptive deep learning for fine-grained image retrieval (ADLFG) of Prang style, Haripunchai style, Lankan art style, Bell-shaped pagoda style, and square wooden pagoda style are 0.88, 0.75, 0.72, 0.65, and 0.77, respectively. The F1-score of the decorrelated global piecewise centralized ranking loss (DGPCR) is shown in Tab. 2. The F1-score of Prang style, Haripunchai style, Lankan art style, Bell-shaped pagoda style, and square wooden pagoda style are 0.81, 0.74, 0.72, 0.74, and 0.73, respectively. The F1-score of the attribute-specific embedding network (ASEN) is shown in Tab. 2. The F1-score of Prang style, Haripunchai style, Lankan art style, Bell-shaped pagoda style, and square wooden pagoda style are 0.82, 0.67, 0.79, 0.77, and 0.71, respectively.

The Recall@K of the proposed algorithm by using adaptive deep learning for fine-grained image retrieval (ADLFG), decorrelated global piecewise centralized ranking loss (DGPCR), and an attribute-specific embedding network (ASEN) is shown in Tab. 3. The number k is ranked as 1, 2, 5, and 10. Recall@5 of the proposed algorithm using adaptive deep learning for fine-grained image retrieval (ADLFG) of Prang style, Haripunchai style, Lankan art style, Bell-shaped pagoda style, and square wooden pagoda style are 0.95, 0.90, 0.93, 0.92, and 0.93, respectively.

Table 3: Recall@K of the proposed algorithm by using adaptive deep learning for fine-grained image retrieval (ADLFG), decorrelated global piecewise centralized ranking loss (DGPCR), attribute-specific embedding network (ASEN)

Class	K = 1			K = 2			K = 5			K = 10		
	ADLFG	DGPCRL	ASEN	ADLFG	DGPCRL	ASEN	ADLFG	DGPCRL	ASEN	ADLFG	DGPCRL	ASEN
Prang style	**0.89**	0.81	0.79	**0.94**	0.90	0.80	**0.95**	0.91	0.82	**0.95**	0.91	0.82
Haripunchai style	0.73	**0.80**	0.69	**0.80**	0.75	0.70	**0.90**	0.80	0.75	**0.90**	0.80	0.75
Lankan art style	0.75	0.70	**0.80**	0.80	0.75	**0.85**	**0.93**	0.80	0.85	**0.93**	0.80	0.85
Bell-shaped pagoda style	0.60	0.70	**0.80**	0.78	0.80	**0.85**	**0.92**	0.80	0.85	**0.92**	0.80	0.85
Square wooden pagoda style	**0.80**	0.78	0.75	**0.85**	0.80	0.75	**0.93**	0.80	0.80	**0.93**	0.80	0.80

5.2 Discussion

From the main categories and sub-categories of the pagoda, which mostly occur in a similar style, it is difficult to retrieve the correct image in the database. Using a fine-grained image retrieval technique in this group of cultural heritage images in a database can solve this problem, as shown in Tabs. 2 and 3. According to the findings of this study's experiments, the proposed method for retrieving the correct image from a database can deliver an average accuracy of 85 percent. Adaptive deep learning for fine-grained image retrieval was used to retrieve the cultural heritage content, and it outperformed the previous methods.

The experimental results show significant improvement in the Prang style and square wooden pagoda style. The accuracies of the Prang style and square wooden pagoda style were 0.89 and 0.80, respectively. However, the accuracy and precision of the Haripunchai and Bell-shaped pagoda styles were not significantly improved. The reason for this is that the shape of this style is similar to that of the Prang style; therefore, the accuracy is lower than that of the state-of-the-art method. The case of the Lankan art style and Bell-shaped pagoda style in terms of recall and F1-score are lower than those of the state-of-the-art method. In this case, we used the first rank for each retrieval to show the results. However, in the recall Recall@5 rank, the results are better than those of the other methods, as shown in Tab. 3.

Additionally, if the time performance in each case is faster than in other methods, it is not necessary to re-train the training dataset.

6 Conclusion

We developed a cultural heritage content retrieval method that uses adaptive deep learning for fine-grained image retrieval. This research used the cultural heritage image: a case study and the architecture of Thai architecture. The key contribution of this research was the creation of an up-to-date retrieval model that could handle incremental streams of new categories while maintaining its past performance on old categories and not losing the old categorization of a cultural heritage image. We analyzed the experimental results of cultural heritage content retrieval using cultural heritage content retrieval efficiency by incremental learning for fine-grained images. In this research, images of Thai archaeological site architecture from world heritage provinces in Thailand were retrieved, for example, images from Sukhothai Province, Ayutthaya Province, and Bangkok, which are mostly similar in their architecture. The proposed adaptive deep learning for Thai archaeological site architecture retrieval outperformed the state-of-the-art method in fine-grained image retrieval. In the future, the proposed fine-grained image retrieval technique will be applied to multiclass classification to improve the retrieval accuracy rate with optimized feature selection strategies.

Funding Statement: This research was funded by King Mongkut's University of Technology North Bangkok (Contract no. KMUTNB-62-KNOW-026).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] C. Chareonla, "Buddhist arts of Thailand," Master Thesis. Magadh University, India, 1981.

[2] S. Prasomphan and J. E. Jung, "Mobile application for archaeological site image content retrieval and automated generating image descriptions with neural network," *Mobile Networks and Applications*, vol. 22, no. 4, pp. 642–649, 2017.

[3] S. Prasomphan, "Cultural heritage content management system by deep learning," in *Proc. the 2020 Asia Service Sciences and Software Engineering Conf. (ASSE'20), Association for Computing Machinery*, New York, NY, USA, pp. 21–26, 2020.

[4] S. Prasomphan and N. Pinngoen, "Feature extraction for image matching in Wat Phra Chetuphon Wimonmangklararam balcony painting with SIFT algorithms," in *Proc. 2021 IEEE 4th Int. Conf. on Computer and Communication Engineering Technology (CCET)*, Beijing, China, pp. 79–84, 2021.

[5] W. Chen, Y. Liu, W. Wang, T. Tuytelaars, E. M. Bakker *et al.*, "On the exploration of incremental learning for fine-grained image retrieval," *ArXiv, abs/2010.08020*, pp. 1–15, 2020.

[6] D. Wu, Q. Dai, J. Liu, B. Li and W. Wang, "Deep incremental hashing network for efficient image retrieval," in *Proc. 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, California, USA, pp. 9061–9069, 2019.

[7] J. Yu, M. Tan, H. Zhang, D. Tao and Y. Rui, "Hierarchical deep click feature prediction for fine-grained image recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 2, pp. 1–10, 2019.

[8] R. Watkins, N. Pears and S. Manandhar, "Vehicle classification using ResNets localisation and spatially-weighted pooling," *ArXiv, abs/1810.10329*, pp. 1–15, 2018.

[9] S. Prasomphan, P. Nomrubporn and P. Pathanarat, "Automated generating Thai stupa image descriptions with grid pattern and decision tree," in *Proc SCDS 2016*. Singapore, vol. 652, 1–7, 2016.

[10] M. A. Nielsen, "Deep learning," in *Neural Networks and Deep Learning*, 1st ed., vol. 1, USA: Determination Press, pp. 167–205, 2018.

[11] L. Deng and D. Yu, “Deep learning: Methods and applications,” *Foundations and Trends R in Signal Processing*, vol. 7, no. 3–4, pp. 197–387, 2013.

[12] U. Kulkarni, S. M. Meena, S. V. Gurlahosur and U. Mudengudi, “Classification of cultural heritage sites using transfer learning,” in *Proc. 2019 IEEE Fifth Int. Conf. on Multimedia Big Data (BigMM)*, Singapore, pp. 391–397, 2019.

[13] A. Latif, A. Rasheed, U. Sajid, J. Ahmed, N. Ali *et al.,* “Content-based image retrieval and feature extraction: A comprehensive review,” *Mathematical Problems in Engineering*, vol. 2019, no. 4, pp. 1–21, 2019.

[14] A. Belhi, A. Bouras, A. K. Ali and S. A. Foufou, “Machine learning framework for enhancing digital experiences in cultural heritage,” *Journal of Enterprise Information Management*, vol. 33, no. 1, pp. 1–13, 2020.

[15] Y. Mistry, D. Ingole and M. Ingole, “Content based image retrieval using hybrid features and various distance metric,” *Journal of Electrical Systems and Information Technology*, vol. 5, no. 3, pp. 878–888, 2017.

[16] M. K. Bashir and Y. Saleem, “Deep hashing for semi-supervised content based image retrieval,” *KSII Transactions on Internet and Information Systems*, vol. 12, no. 8, pp. 3790–3803, 2018.

[17] Y. Rao and W. Liu, “Region division for large-scale image retrieval,” *KSII Transactions on Internet and Information Systems*, vol. 13, no. 10, pp. 5197–5218, 2019.

[18] M. Wu, Q. Chen and Q. Sun, “Medical image retrieval with relevance feedback via pairwise constraint propagation,” *KSII Transactions on Internet and Information Systems*, vol. 8, no. 1, pp. 249–268, 2014.

[19] K. Anitha and A. Chilambuchelvan, “NPFAM: Non-proliferation fuzzy ARTMAP for image classification in content-based image retrieval,” *KSII Transactions on Internet and Information Systems*, vol. 9, no. 7, pp. 2683–2702, 2015.

[20] Y. Mussarat, S. Muhammad, M. Sajjad and I. Isma, “Content based image retrieval using combined features of shape, color and relevance feedback,” *KSII Transactions on Internet and Information Systems*, vol. 7, no. 12, pp. 3149–3165, 2013.

[21] 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.

[22] 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.

[23] E. Liu, “Research on image recognition of intangible cultural heritage based on CNN and wireless network,” *EURASIP Journal on Wireless Communications and Networking*, vol. 240, no. 1, pp. 1–12, 2020.

[24] Ş. Öztürk, “Image inpainting based compact hash code learning using modified U-Net,” in *Proc. 2020 4th Int. Sym. on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, Istanbul, Turkey, pp. 1–5, 2020.

[25] Ş. Öztürk, “Class-driven content-based medical image retrieval using hash codes of deep features,” *Biomedical Signal Processing and Control*, vol. 68, pp. 1–9, 2021.

[26] Ş. Öztürk, “Comparison of pairwise similarity distance methods for effective hashing,” *Int. Conf. on Applied Scientific Computational Intelligence using Data Science (ASCI 2020)*, vol. 1099, no. 1, pp. 1–13, 2021.

[27] X. Zeng, Y. Zhang, X. Wang, K. Chen, D. Li *et al.,* “Fine-grained image retrieval via piecewise cross entropy loss,” *Image and Vision Computing*, vol. 93, no. 5, pp. 1–6, 2020.

[28] H. Xu, G. Qi, J. Li, M. Wang, K. Xu *et al.,* “Fine-grained image classification by visual-semantic embedding,” in *Proc. the Twenty-Seventh Int. Joint Conf. on Artificial Intelligence, IJCAI-18*, Stockholm, Sweden, pp. 1043–1049, 2018.

[29] F. Liu, C. Zou, X. Deng, R. Zuo, Y. Lai *et al.,* “SceneSketcher: Fine-grained image retrieval with scene sketches,” in *Proc. 16th European Conference on Computer Vision–ECCV 2020*, Glasgow, UK, pp. 718–734, 2020.

[30] X. Zhang, H. Xiong, W. Zhou, W. Lin and Q. Tian, “Picking deep filter responses for fine-grained image recognition,” in *Proc. 2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 1134–1142, 2016.

[31] L. Xie, J. Wang, B. Zhang and Q. Tian, “Fine-grained image search,” *IEEE Transactions on Multimedia*, vol. 17, no. 5, pp. 636–647, 2015.

[32] V. Kumar, V. Tripathi and B. Pant, “Content based fine-grained image retrieval using convolutional neural network,” in *Proc. 2020 7th Int. Conf. on Signal Processing and Integrated Networks (SPIN)*, Noida, India, pp. 1120–1125, 2020.

[33] Q. Cui, Q. Y. Jiang, X. S. Wei, W. J. Li and O. Yoshie, “ExchNet: A unified hashing network for large-scale fine-grained image retrieval,” in *Proc. 16th European Conference on Computer Vision–ECCV 2020*, Glasgow, UK, pp. 189–205, 2020.

[34] J. Dong, Z. Ma, X. Mao, X. Yang, Y. He *et al.,* “Fine-grained fashion similarity prediction by attribute-specific embedding learning,” *IEEE Transactions on Image Processing*, vol. 30, pp. 1–15, 2021.

[35] X. Wei, J. Luo, J. Wu and Z. Zhou, “Selective convolutional descriptor aggregation for fine-grained image retrieval,” *IEEE Transactions on Image Processing*, vol. 26, no. 6, pp. 2868–2881, 2017.