

Article

Research on Image Classification and Retrieval Using Deep Learning with Attention Mechanism on Diaspora Chinese Architectural Heritage in Jiangmen, China

Le Gao ¹, Yanqing Wu ¹, Tian Yang ^{2,*}, Xin Zhang ¹, Zhiqiang Zeng ^{1,*} , Chak Kwan Dickson Chan ³ and Weihui Chen ¹

¹ Faculty of Intelligent Manufacturing, Wuyi University, Jiangmen 529000, China

² Institute for Guangdong Qiaoxiang Studies, Wuyi University, Jiangmen 529000, China

³ Faculty of Social Sciences, Lingnan University, Hongkong 999077, China

* Correspondence: wlyangtian@163.com (T.Y.); zhiqiang.zeng@outlook.com (Z.Z.)

Abstract: The study of the architectural heritage of the Chinese diaspora has an important role and significance in China's historical and cultural background in the preservation of cultural data, the restoration of images, and in the analysis of human social and ideological conditions. The images from the architectural heritage of the Chinese diaspora usually include frescos, decorative patterns, chandelier base patterns, various architectural styles and other major types of architecture. Images of the architectural heritage of the Chinese diaspora in Jiangmen City, Guangdong Province, China are the research object of this study. A total of 5073 images of diaspora Chinese buildings in 64 villages and 16 towns were collected. In view of the fact that different types of image vary greatly in features while there are only small differences among the features of the same type of image, this study uses the depth learning method to design the Convolutional Neural Network Attention Retrieval Framework (CNNAR Framework). This approach can be divided into two stages. In the first stage, the transfer learning method is used to classify the image in question by transferring the trained parameters of the Paris500K datasets image source network to the target network for training, and thus the classified image is obtained. The advantage of this method is that it narrows the retrieval range of the target image. In the second stage, the fusion attention mechanism is used to extract the features of the images that have been classified, and the distance between similar images of the same type is reduced by loss of contrast. When we retrieve images, we can use the features extracted in the second stage to measure the similarities among them and return the retrieval results. The results show that the classification accuracy of the proposed method reaches 98.3% in the heritage image datasets of the JMI Chinese diaspora architectures. The mean Average Precision (mAP) of the proposed algorithm can reach 76.6%, which is better than several mainstream model algorithms. At the same time, the image results retrieved by the algorithm in this paper are very similar to those of the query image. In addition, the CNNAR retrieval framework proposed in this paper achieves accuracies of 71.8% and 72.5% on the public data sets Paris500K and Corel5K, respectively, which can be greatly generalized and can, therefore, also be effectively applied to other topics datasets. The JMI architectural heritage image database constructed in this study, which is rich in cultural connotations of diaspora Chinese homeland life, can provide strong and reliable data support for the follow-up study of the zeitgeist of the culture reflected in architecture and the integration of Chinese and Western aesthetics. At the same time, through the rapid identification, classification, and retrieval of precious architectural images stored in the database, similar target images can be retrieved reasonably and accurately; then, accurate techniques can be provided to restore old and damaged products of an architectural heritage.

Keywords: deep learning; attention mechanism; image classification; image retrieval; architectural heritage



Citation: Gao, L.; Wu, Y.; Yang, T.; Zhang, X.; Zeng, Z.; Chan, C.K.D.; Chen, W. Research on Image Classification and Retrieval Using Deep Learning with Attention Mechanism on Diaspora Chinese Architectural Heritage in Jiangmen, China. *Buildings* **2023**, *13*, 275. <https://doi.org/10.3390/buildings13020275>

Academic Editors: Dirk H.R. Spennemann and Humberto Varum

Received: 4 December 2022

Revised: 20 December 2022

Accepted: 14 January 2023

Published: 17 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Images from the architectural heritage of the Chinese diaspora are a powerful basis for studying modern culture, human life and history, as well as ideologies prevalent in homelands of diaspora Chinese. The image of diaspora Chinese architecture usually includes several types, such as frescos, decorative patterns, chandelier base patterns, and architectural styles. In recent years, many scholars have carried out research on and analysis of architectural images and accumulated much valuable experience and data [1–4]. However, current research in this area is mostly scattered, lacking a systematic and integrated approach to the architectural heritage and culture of the Chinese diaspora as well as an image database of diaspora Chinese architecture. This paper aims to resolve the above problems, using computer technology and artificial intelligence depth learning methods to design a set of effective images for fast recognition, classification, and building of retrieval models. The field research location selected for this study is in Jiangmen, Guangdong Province, China—hometown of the Chinese diaspora. Our research scope include frescos, decorative patterns, chandelier base patterns, architectural styles, and other images collected of diaspora Chinese culture and architecture. On the one hand, this study effectively stores the existing diaspora Chinese architectural images in Jiangmen; on the other hand, based on the constructed database, it quickly retrieves images similar to those architectural images, thus providing strong technical support for the restoration of architectural designs and the interpretation of diaspora Chinese culture.

In recent years, deep learning technology has gradually penetrated into various research fields. Experts and scholars have carried out a large number of scientific studies using deep learning technology and achieved certain research results in various fields [5–13]. Image retrieval based on deep learning is currently a research hot spot in computer vision, which can effectively search, query, and match images similar to the target task in the image database [14]. Image feature extraction technology is the core problem to improve the accuracy of image retrieval. Traditional image feature extraction methods mainly design manual descriptors for low-level features such as image color, texture and edge [15]. However, the traditional extraction methods have great defects, such as poor retrieval effects with varied lighting, background, occlusion, and other factors. With in-depth research of artificial intelligence and deep learning technology, we can use convolutional neural networks to extract high-level semantic features of images, which greatly improves the accuracy of image retrieval [16]. The classical Network models include ResNet (Residual Neural Network) and GoogLeNet, VGGNet (Visual Geometry Group Net) [17–19], which mainly extract image features using the overall characteristics of images. However, if only a single network such as ResNet, GoogLeNet, and VGGNet is used to extract image features, there will also be some disadvantages for image recognition, such as local features inside images becoming easy to ignore, or common interference by occlusions. The attention mechanism can effectively solve the above problems. This method is often used in image recognition and classification [20], object detection [21], semantic segmentation [22], multi-modal tasks [23], self-supervised learning [24] and other undertakings. Based on research on depth learning technology, this paper introduces depth learning into the recognition and retrieval of diaspora Chinese architectural images, which provides a new research idea for the preservation, retrieval, and restoration of architectural heritage digital images.

At present, some scholars have done similar research on the recognition and retrieval of architectural images. For example, the authors of [25] constructed a dataset for application to examples of the traditional Chinese architectural heritage and proposed a deep retrieval method for the recommended task. In [26], The author puts forward the application of rural architectural design in view of architectural image recognition based on machine learning and other technologies. This research uses the Histogram of Oriented Gradient (HOG) algorithm to extract the contour features of the building image, uses the trained SVM classifier to classify the image features, and finally completes the recognition of the building. The authors of [27] combine 3D image technology with Internet of Things technology to study the development of the classical architectural art style. This paper

studies the collection and data preprocessing of 3D architectural images and in 3D displays the images through simulation experiments, providing data support for policy formulation and technical intervention into the architectural heritage. In [28], the author used depth learning technology to classify architectural heritage images, especially by using convolutional neural networks. Their research results can correctly classify the existing images, thus contributing to the research and interpretation of relevant aspects of the architectural heritage. The authors of [29] proposed a method to identify the roof type of rural buildings based on high resolution remote sensing images of UAVs. This method can obtain clear and accurate proofs of roof profiles and of types of rural buildings; it can be used for green roof evaluation of potential construction of roof solar energy. In addition, Hong et al. [30] applied building image recognition and classification technology to post disaster building damage assessment, which can effectively classify buildings with various damage levels. Taoufig et al. [31] applied the automatic classification technology of architectural images to smart city construction and cultural heritage projects. The architectural image database they built provides strong support for urban planning, virtual city tours and digital archiving of cultural relics. To sum up, most of the above studies only extract the overall characteristics of the building facade or focus on some areas of the building but fail to extract the spatial characteristics of separate components. In retrieval research with a variety of different building image datasets, it is difficult to achieve both rapid recognition and high accuracy. Based on the above depth learning technology in image building research, this paper adopts the depth learning architectural image hierarchical retrieval strategy for research. We propose an architectural style classification and retrieval method based on Convolutional Neural Network (CNN) and channel spatial attention. First, the CNN feature extractor is used to extract the depth feature. Secondly, the channel space attention module is introduced to generate attention maps, which can not only enhance the texture of feature representation of architectural images but can also focus on the spatial features of various architectural elements. Then, the Softmax classifier is used to predict the score of the target class. Finally, based on the image classification, the similarity measure is used to retrieve the results. This paper adopts the method of the Convergent Block Attention Module (CBAM) to extract the features of diaspora Chinese architectural images in Jiangmen city. By fusing this Module, the features of similar patterns in architectural images can become more prominent. CBAM is a simple and effective attention module, which is mainly used in the feed forward convolutional neural network. Because it is a general module with broad compatibility, it can be integrated into any CNN model architectures.

Our contributions in this paper mainly address the following four areas: (1) Aiming at the current situation of the world's cultural heritage of Chinese diaspora architecture in Jiangmen, China, which has been in disrepair and damaged for many years, this study enables a complete image database of the Chinese diaspora architectural heritage to be built. Through the rapid identification, classification, and retrieval of the precious architectural images and patterns stored in this database, the targeted similar images can be retrieved reasonably and accurately, thus providing an accurate restoration plan for the architectural heritage. (2) The world cultural heritage of diaspora Chinese architecture in Jiangmen City has a great many frescos and patterns rich in the cultural connotations of the homeland of diaspora Chinese. The construction of the image database of architectural heritage in this study provides rich and reliable data support for a follow-up study of architectural trends of the times and the integration of Chinese and Western aesthetics. (3) We used the data enhancement method to expand the samples of diaspora Chinese architectural images, and at the same time, we used migration learning to move the parameters trained by the image source network of the Paris500K datasets to the image datasets to train the diaspora Chinese architectural image network. Through this method, we classified architectural images into four types: frescos, decorative patterns, chandelier base patterns, and architectural styles. (4) Given that distinct types of architectural images are largely different from one another and the same types have small differences, we propose a two-stage training image retrieval network, CNNAR Framework, with deep learning and

attention fusion mechanisms. This network framework measures the distances between classified diaspora Chinese architectural images by using the fine features extracted from similar images and then returns the retrieved images, greatly improving the accuracy of architectural image retrieval.

2. Methodology

2.1. Study Area and Datasets

Jiangmen City, located in Guangdong Province, China, is a famous homeland of diaspora Chinese in South China (Figure 1). From 1869 to 1939, a large number of Jiangmen people traveled to North America, Southeast Asia and other countries, expanding their population, materials, culture, information, and ideas into an endless network through human exchanges and north-south trade via the Pacific Ocean route. Under the influence of nationalist sentiment, democratic thought, and gradually open vision, the diaspora Chinese returned home and integrated the modern industrial civilization and fashion into the Chinese cultural context and architectural spirit. The values of a new society, new traditions, and new culture are embodied in the rural society of South China. The Jiangmen diaspora Chinese were the first Chinese people to fix their eyes on the Western world.

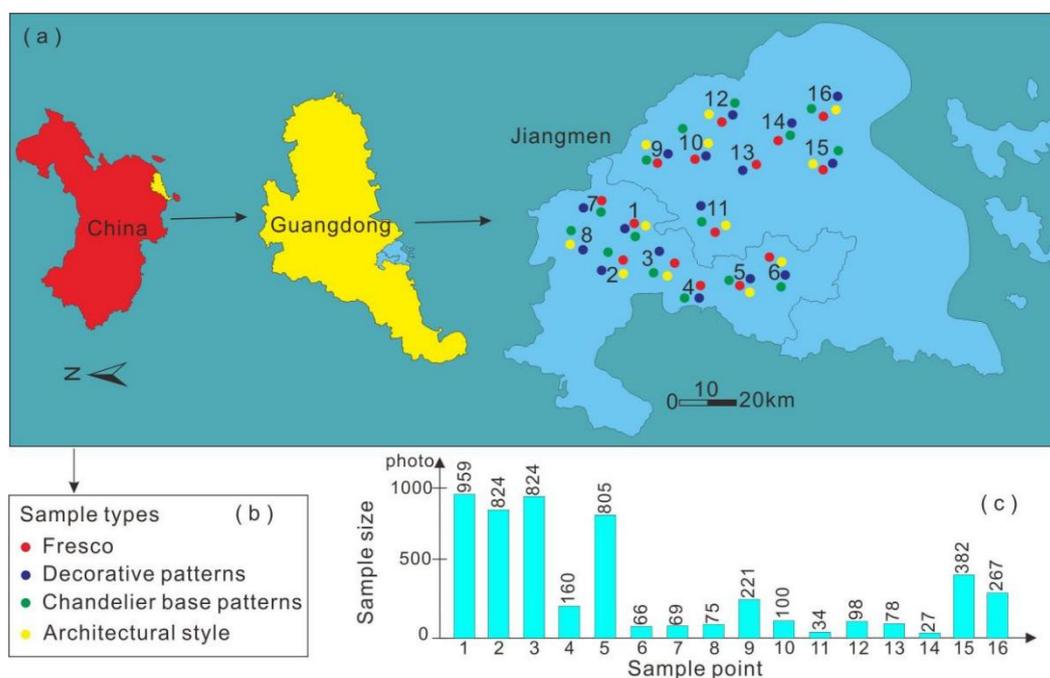


Figure 1. Sampling point map of the research area (a) geographical location; (b) sampling type; (c) number of samples.

According to the research data, there are 2019 barbicans (Diaolou), 2000 arcade buildings and 2800 villas in Jiangmen, and the buildings with various Western architectural elements account for more than 95% of buildings in the village. The architectural images of overseas Chinese contain important cultural information and characteristics of the times. For example, the four-petal lotus is the most frequent feature in images, and it has been regarded as an auspicious pattern since the Han Dynasty in China. In addition, the rooster image is a symbol of the light guarding the portal. It is a psychological continuation of Chinese traditional culture that drives away ghosts and has multiple meanings of exorcising evil spirits. It has a strong stability in the traditional root culture of diaspora Chinese. The Syncretic of Chinese and Western cultures is vividly reflected in the pattern of “lion rolling hydrangea ball”. The traditional pattern of “lion rolling silk ball” is displayed in the decoration of diaspora Chinese architecture, which turns into a lion standing in the middle of the earth.

This paper takes the diaspora Chinese architectural images in Jiangmen City, Guangdong Province, China as the research object, and collects 5073 architectural images in 16 towns and 64 villages (Figure 2). Based on the collected data, we constructed an architectural Heritage Image Dataset JMI (Jiangmen Images). The experiments in this paper were carried out on JMI datasets containing several types of images, such as frescos, decorative patterns, chandelier base patterns, and architectural styles. This dataset contained a large number of frescos and patterns rich in the cultural connotations of diaspora Chinese. The construction of this architectural heritage image database provided rich and reliable data support for follow-up research on architectural trends of the times and the integration of Chinese and Western aesthetics.

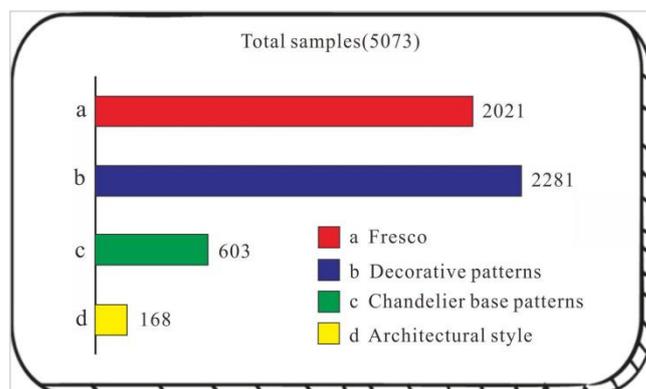


Figure 2. The type and quantity of samples.

We take 80% of each type of image for experimental training and 20% for experimental testing. In our experiment, two auxiliary datasets are also used, namely Paris500K datasets [32] and Corel5K datasets [33]. The Paris500K datasets is a collection of landmark images from photos taken by various researchers and online albums. Images have a “natural” distribution; there is approximate duplication and a large number of unrelated images, including pictures of parties, pets, etc. It contains 94,303 images of 79 landmarks. We selected 80,000 images for training and the remaining 14,303 images for testing. The Corel5K datasets contain 5000 images collected and curated by the company Corel. The image database covers multiple topics and consists of 50 CDS, each representing a semantic topic and each containing 100 equal-sized images that can be converted to different formats. We selected 4000 images for training and the remaining 1000 images for testing.

2.2. Classification and Retrieval

According to the differences among image features of different types of the JMI architectural heritage, we designed a two-stage training image retrieval network framework, CNNAR Framework, with in-depth learning and an attention fusion mechanism, as shown in Figure 3. The network framework is mainly divided into building image classification modules and image retrieval modules, which greatly improves the accuracy of diaspora Chinese architectural image retrieval. First, the task of the image classification module mainly includes a target network of Transfer Learning ResNet50 (TL ResNet50). We use the migration learning method to migrate, or move, the parameters trained by the image source network of the Paris500K datasets to the architectural heritage image dataset to train the image network. Diaspora Chinese architectural images are classified into four types: frescos, decorative patterns, chandelier base patterns and architectural styles. Then, the image retrieval module task includes a Convolutional Block Attention Module Net (CBAM Net) module, which can be used to extract the feature vector of distance measurement. During the classification and retrieval of architectural images, the architectural image database determines the image type through the TLResnet50 network in the offline phase, and then obtains the feature vector through the CBAM Net network and constructs the feature vector library. First, we get the feature vector of the input image through online

searching of the same network as the building image database, and then calculate the distance with the feature vector of the same category of building images. Finally, according to the calculated value, we derive the image retrieval result with the closest similarity.

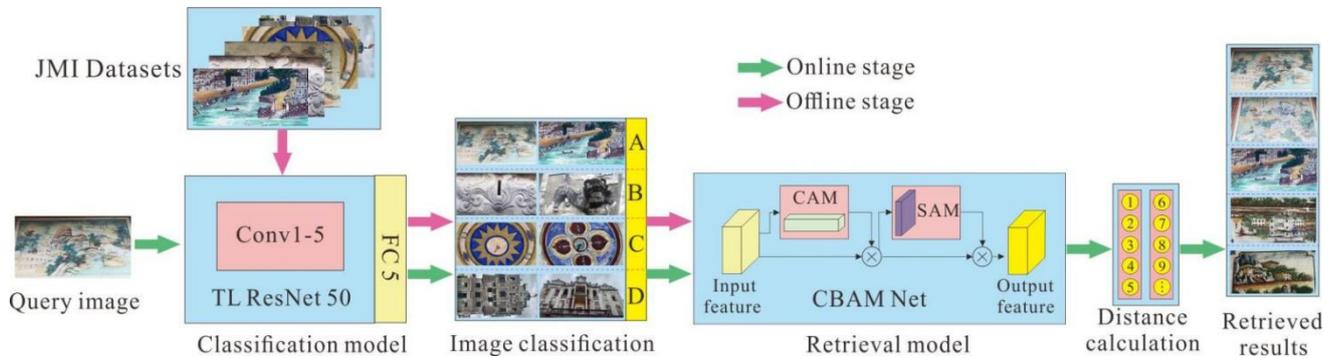


Figure 3. Block diagram of Diaspora Chinese architectural heritage image classification and retrieval.

2.2.1. Classification Module Task (Phase I)

Distinct types of image data in the image database based on diaspora Chinese architectural heritage vary largely in characteristics. The task of our classification module network is to classify diverse types of images in the building image data set. In order to improve the accuracy of building image classification, we combine the migration learning method with the residual network to classify building images and use ResNet50 as the classification network.

With the increase in number of deep learning layers, the problem of gradient disappearance becomes more serious, which makes the network parameters of the output layer impossible to learn effectively. In order to solve this problem, residual networks appear. A residual network is an effective network to alleviate the problem of gradient disappearance, which greatly improves the depth of the network that can be effectively trained [34]. Figure 4 is the residual structure diagram, in which the residual block is divided into two parts: direct mapping x and residual $F(x)$. We added the input of the previous layer without any other calculation to the result of convolution calculation and passed it into the network structure of the next layer. In the process of forward transmission, as the number of layers deepens, the image information contained in the Feature Map will decrease layer by layer, while the direct mapping of the residual network ensures that the network of layer $n + 1$ must contain more feature information than that of layer n , thus effectively alleviating the problem of vanishing gradient without adding additional parameters.

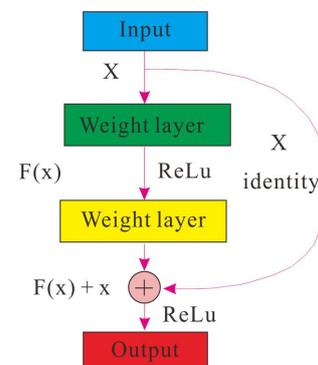


Figure 4. Residual block structure.

In recent years, transfer learning methods have been widely used in different fields of research, such as building image classification [35], industrial machinery fault detection [36], water quality prediction [37], material performance prediction [38], medical image analysis [39], etc. However, such research on architectural heritage image datasets

is relatively sparse. If we directly use the depth learning model for classification training, it is easy for the phenomenon of over fitting to appear, so we use the transfer learning method to avoid that. Therefore, the parameters trained in the Paris500K public datasets are migrated to the JMI architectural heritage datasets for training by using the migration learning method. Figure 5 is a schematic diagram of the transfer learning process. Figure 6 is a flow diagram of the image classification stage. First, input Paris500K dataset is applied to a designed model to train the given task. After the model training is completed, we fix the parameter weights of the model, replace the input data source with JMI data set, and change the final classification of the model Task to obtain a new model for re training. After the Model training is completed, we fix the parameter weight of the model, change the input data source to the JMI data set, and change the final classification task of the model to get a new model for re-training.

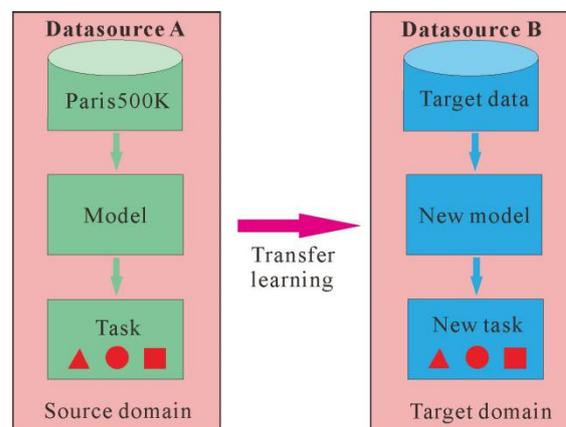


Figure 5. Schematic diagram of transfer learning process.

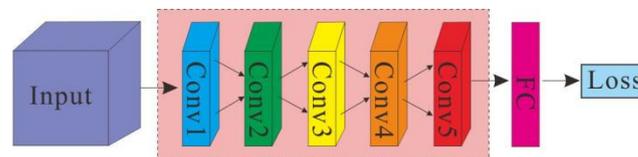


Figure 6. Classification phase model diagram.

This paper use ResNet50 as an architectural image classification network. ResNet is proposed by He et al. [40] of Microsoft Research. The main ResNet50 model structure is shown in Figure 7. A direct channel is established through the ResNet neural network, which combines input and output, effectively alleviates the loss of image feature information caused by convolution operation, and plays a positive role in solving the problem of gradient disappearance or explosion in the deep network. As shown in the network structure in Figure 7, identity block defines three convolution operations and protects the integrity of the image features information through direct channels. After convolution, the feature map is normalized in batches and processed by activation function, so as to enhance the ability of the model to extract image features. Specifically, we first use 7×7 and the maximum pooling layer to extract sample features and reduce dimensions. Then, the image features of the samples are further extracted through four convolution layers containing 3, 4, 6 and 3 identity blocks, respectively. Through the learning of the classification module task, we expanded the distribution space of architectural image features and classified the architectural heritage image features. The classification feature results are shown in Figure 8. The image feature space is expanded from one quadrant to the whole space, which is more conducive to improving the discrimination performance of image features.

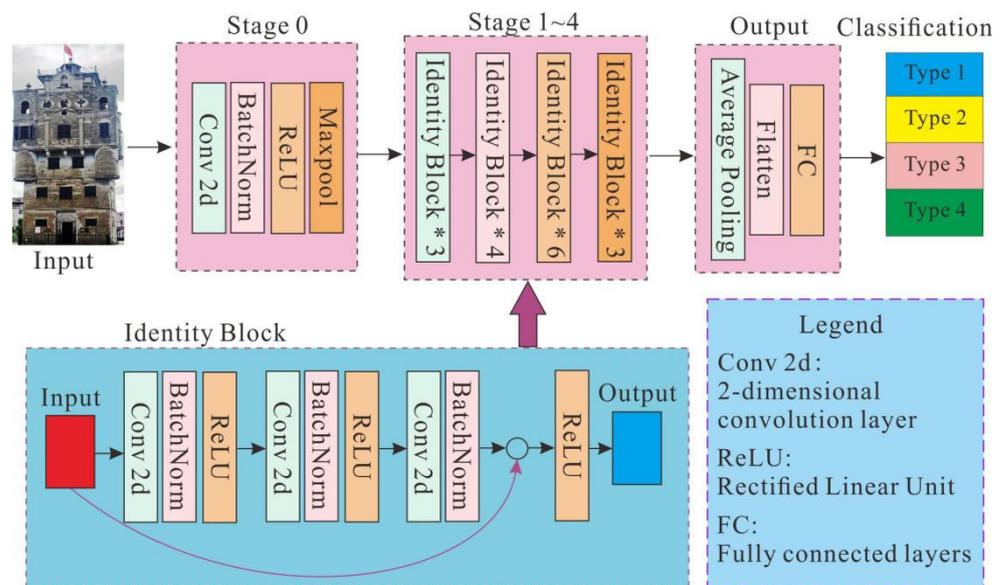


Figure 7. The structure chart of Resnet50.

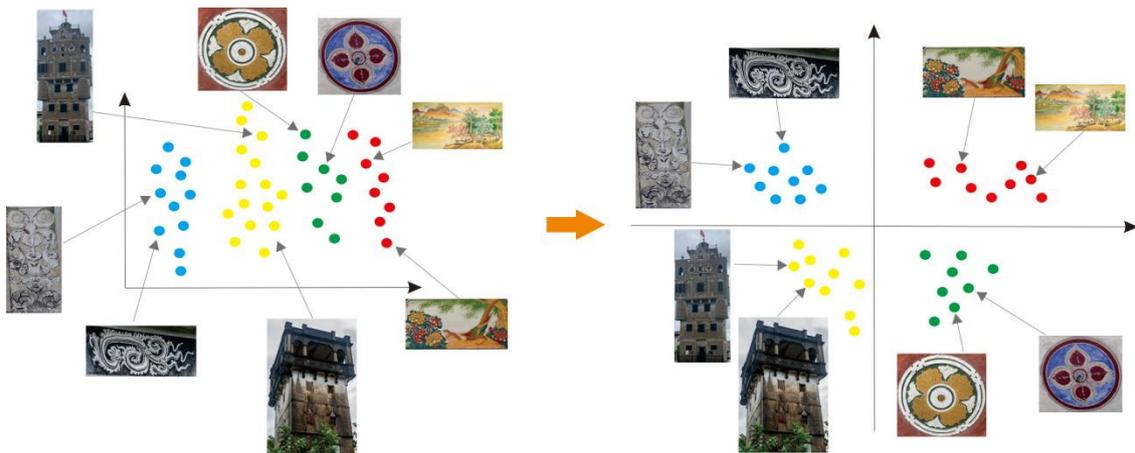


Figure 8. Schematic diagram of image feature classification space.

2.2.2. Retrieval Module Tasks (Phase II)

There are many types of architectural images of diaspora Chinese architectural heritage, including frescos, decorative patterns, chandelier base patterns, and architectural styles, and each type of image contains different features. Because the differences between each image feature are small, using the features trained in the first stage directly for metric retrieval will lead to low accuracy. Therefore, in this study, the attention mechanism is integrated on the basis of the first stage of image classification, which can highlight the feature differences between the same type of images. We construct the CBAM Net network training for retrieving image features.

The attention mechanism in neural networks is a resource allocation scheme that allocates computing resources to more important tasks while solving the problem of information overload when computing power is limited. In recent years, the attention mechanism has made important breakthroughs in image, natural language processing, and other fields, which has been proved to effectively improve the performance of the model [41,42]. In this paper, the features of the same type of images of diaspora Chinese architectures have few differences, and various features are easily disturbed by background and other factors, so the network model needs to have a strong ability to learn detailed features. We use VGG16 network as the backbone network and integrate the attention

mechanism in the feature extraction module to make the model adaptively focus on the features themselves and highlight the features of the image. After extracting features from architectural images through the convolutional layer, the image first passes through Channel Attention Module (CAM) to obtain channel attention features by multiplying the weight of the channel by the input feature layer. Then, through the Spatial Attention Module (SAM), the weight of spatial information is multiplied by the input feature layer to get the fusion attention feature. Finally, the network model is trained by distance ranking and contrast loss function. The model diagram is shown in Figure 9.

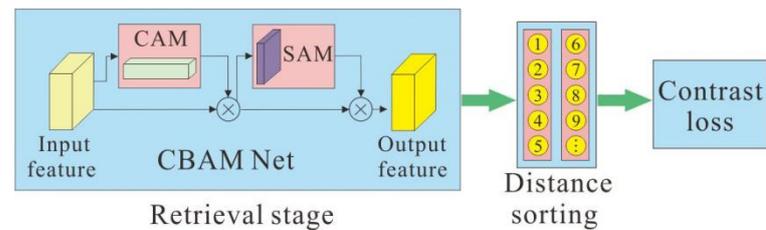


Figure 9. Integrated attention mechanism model.

Each channel of the feature map is a feature detector, and the channel attention mechanism pays a different amount of attention to different image channels [43]. The channel attention module is shown in Figure 10. For the image input features, we first use the maximum pooling and mean pooling algorithms simultaneously, then get the transformation results through several Multilayer perceptron and MLP layers. Finally, we apply the sigmoid function to the two channels respectively, so that the attention result of the channel can be obtained. The calculation procedure is shown in Equation (1).

$$M_c = \text{sigma} (\text{MLP} (\text{AvgPool} (F) + \text{MaxPool} (F))) \quad (1)$$

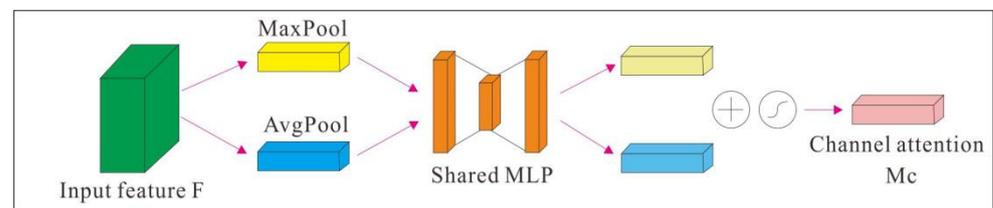


Figure 10. Channel attention module.

In Equation (1), the input data source is the feature F extracted through the convolution layer, and M_c is the channel attention result, σ It refers to sigmoid function, MLP refers to multilayer perceptron, AvgPool refers to average pooling, and MaxPool refers to maximum pooling. Because we are dealing with building image pixels, each pixel is a number of three channels, so the unit is 1.

Spatial attention can be seen as an adaptive spatial region selection mechanism: where to focus [44]? In the spatial attention module in Figure 11, we first reduce the dimension of the channel itself, obtain the results of maximum pooling and mean pooling, respectively, and then concatenate them into a feature map; following that, we use a convolutional layer for learning. The calculation procedure is shown in Equation (2).

$$M_s = \text{sigma} (\text{Conv} (7 \times 7 (\text{AvgPool} (F); \text{MaxPool} (F)))) \quad (2)$$

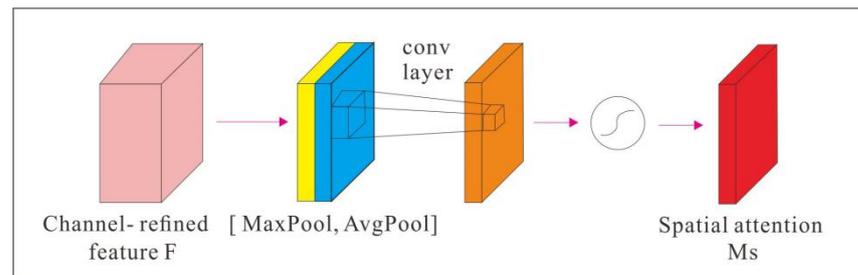


Figure 11. Spatial attention module.

In Equation (2), M_s is spatial attention as a result, F for the input characteristics of sigma as sigmoid function, f 7×7 to 7×7 size of convolution kernels, AvgPool average pooling, MaxPool for maximum pool.

Loss of contrast is a dimension reduction method of study that can take into account a mapping relationship, which can be made in a high dimensional space, the same category but with far distant points. Through function is mapped to a low dimensional space after getting close, various categories where distance is relatively close, by mapping after the low dimensional space becomes broader [45]. In this paper, the contrast loss function is used to process the image of diaspora Chinese architecture. In order to make architectural images of the same type highly similar, each image of the same type is regarded as a small class, and the image after data augmentation forms a positive sample pair with the original image, while other images form a negative sample pair with the original image. After fusing the feature map extracted from the attention mechanism network, we calculate the Euclidean distance between images and train the network using the contrast loss function.

3. Experiment and Discussion

3.1. Data Preprocessing

Neural networks usually need a lot of data to complete effective training. In the case of less data, the training effect and generalization ability of the network model are poor. Zhang et al. [46] proposed a data enhancement model based on the antagonism of feature reconstruction and deformation information for the lack of data samples. The loss of image features can be improved by using feature reconstruction methods, and the data samples can be effectively expanded by using deformation information. Based on this, this study preprocesses the image data to expand the sample number of JMI datasets and achieve better data training effect. In addition, Yu et al. [47] proposed a Hermite interpolation method for the problem of the number of samples in the fault data set. According to the interpolation curve of sample characteristics, a synchronous sampling method was used to build virtual samples. The features of virtual samples are mapped to three-dimensional space to expand the number of enhanced data samples. In order to facilitate the training of the in-depth learning network model, this study first normalizes the size of all images in the JMI building data set to keep the size of $224 \times 224 \times 3$. To solve the problem that the number of images in the building image data set is small and the training can easily be over fit, we use four data enhancement methods to expand the data set. Our goal is to keep the augmented image as similar as possible to the original image, which is helpful for the second stage of loss measurement training. In this study, the methods of Gaussian noise [48], salt and pepper noise [49], histogram equalization [50], and adaptive histogram equalization with limited contrast [51] are used for data expansion, and the number of database images is expanded to 25,365. The image enhancement effect is shown in Figure 12, which shows the data enhancement effect of four types of images, including mural, decorative pattern, lantern flower, and architectural style.

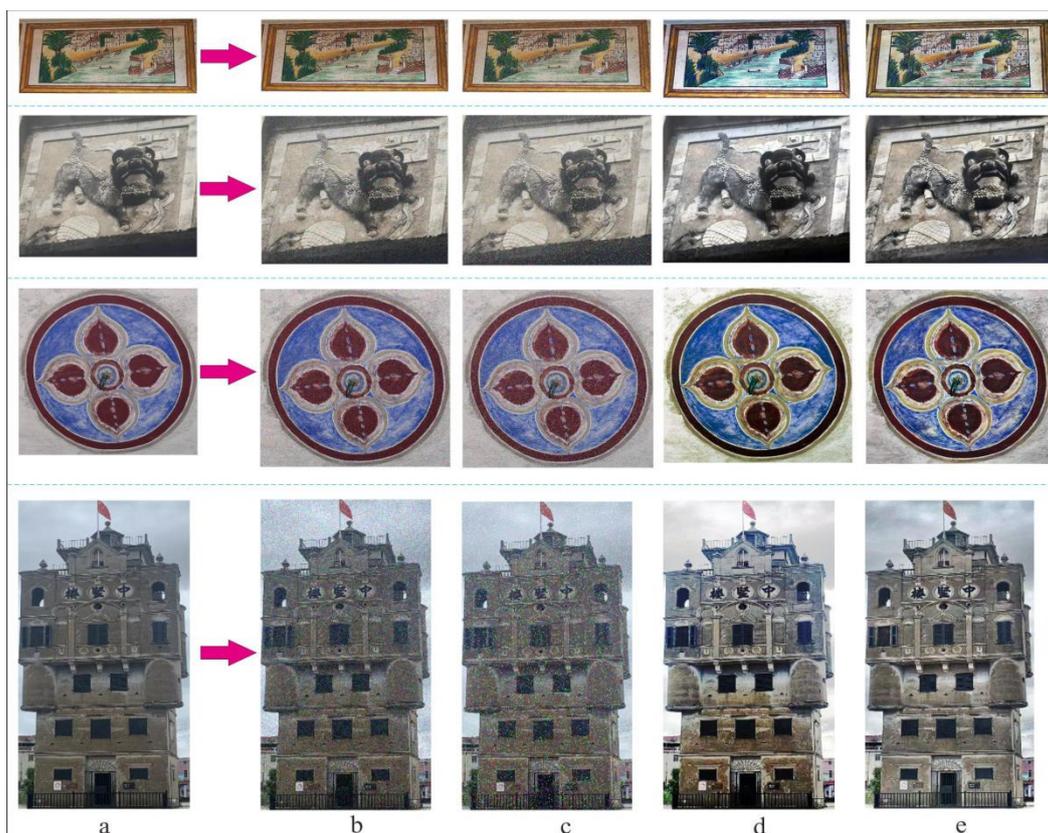


Figure 12. Image enhancement effect (a) original image, (b) gaussian noise, (c) salt and pepper noise, (d) histogram equalization, (e) restricted contrast adaptive histogram equalization.

3.2. Image Classification Experiment

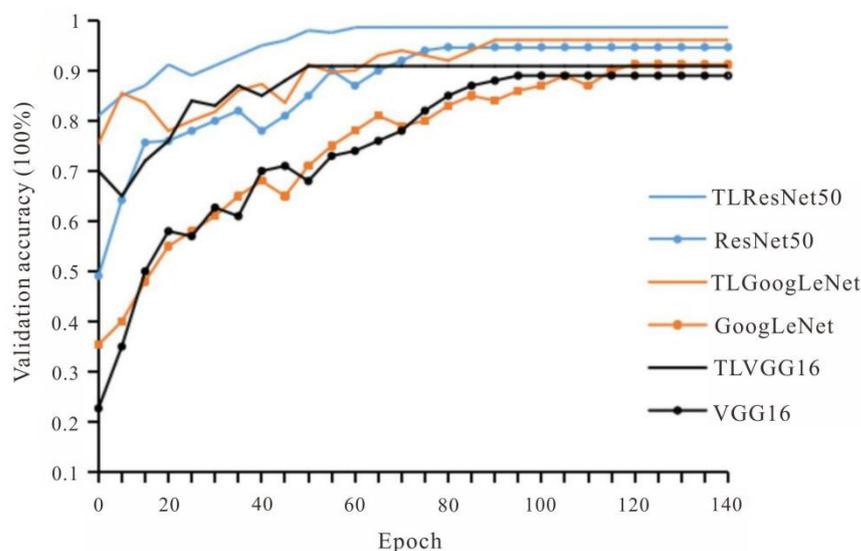
The objects that recognize images and distinguish different categories are called image classification. We train the model of image classification to recognize different types of images. Wang et al. [26] used machine learning technology to identify and extract architectural image features and the SVM classifier to classify image features, finally completing the identification and classification of buildings, providing technical support for the design and planning of rural buildings. Zhao et al. [52] designed a voting mechanism for segmented images using deep learning technology, established a VGG Vote network model, and applied this model to the recognition and classification of remote sensing images. The experimental results show that this vote mechanism significantly improves the classification accuracy, robustness, and anti-interference qualities of the VGG Vote.

In the image classification experimental model training, the source domain datasets are Paris500K, and the target domain datasets are JMI building heritage image datasets. The trained Paris500K datasets image source network parameters are migrated to the JMI architectural heritage image datasets through the migration learning method. The task of the image classification stage is to classify architectural images with high accuracy and narrow the scope of image retrieval, so accuracy is used as the evaluation index of test results. The loss function of network training uses cross entropy loss. In this experiment, the batch size was set as 16, the initial learning rate was 0.0001, and it began to decay 10 times after 10 epochs. The Adam optimizer was used for optimization. Three network models, ResNet50, GoogLeNet, and VGG16, were used to carry out experiments in the image datasets of JMI expatriate architectures. Each model was tested with and without transfer pre-training weights. The accuracy results of the six experiments are shown in Table 1.

Table 1. Accuracy of several network models.

Network Model	Classification Accuracy (%)
TL ResNet50	98.3
ResNet50	94.5
TL GoogLeNet	95.2
GoogLeNet	91.2
TL VGG16	91.4
VGG16	89.3

The experimental results in Table 1 show that the experimental accuracy of the three network models with migration pre-training weights is higher than that of the original network. After GoogLeNet network integration and migration learning, the accuracy rate was improved to 4.0% at most, indicating that its adaptability was higher. But the best performance result is the Resnet50 with migration pre-training weight. Its accuracy rate reaches 98.3% at the highest level, and its original network accuracy rate is 94.5%. It can be seen from the model training accuracy chart in Figure 13 that the network model has a similar change trend in the training process. The migration pre-training weight training network not only improves the experimental accuracy, but also accelerates the convergence speed of model training. The TL ResNet50 starts to converge at about 60 epochs, while the Resnet50 starts to converge at 80 epochs. TL GoogLeNet starts to converge at about 90 epochs, while GoogLeNet starts to converge at 120 epochs. TL VGG16 starts to converge at about 50 epochs, while VGG16 starts to converge at 90 epochs. Based on the above experimental analysis, Resnet50 network with migration weight is the best network for JMI building image datasets classification.

**Figure 13.** Training accuracy of several models.

3.3. Contrast Experiment of Mainstream Image Retrieval Methods

Image retrieval can quickly meet the needs of users to retrieve files in the building image database. On the one hand, image retrieval is used to collect and process architectural image resources, extract features, analyze and index, and establish an image index database. On the other hand, the similarity between the user query image and the database image is calculated according to the similarity algorithm, and the images meeting the threshold are extracted and output in descending order of similarity. Ma et al. [53] present a method for organizing and retrieving photos from massive facility management photo databases using photo-metadata: photographed location, camera perspective, and image semantic content information. The method is applied to 21 building image datasets. The research shows that their metadata-based image retrieval system can achieve fast image retrieval

according to the needs of users. Sun et al. [54] proposed an improved EAST detector algorithm to identify and retrieve images. The algorithm uses a full convolution neural network structure to extract multi-scale features of image text and ensures the balance between positive and negative samples by adjusting the loss function. The experimental results show that the algorithm can effectively solve the feature difference between images of the same category and improve the low recall problem of detection. We use our CNNAR Framework network model to do image retrieval research on JMI architectural heritage datasets. In order to prove the superiority of our model on JMI datasets, this paper conducts comparative experimental analysis with several mainstream network model methods. In this paper, the image datasets of JMI diaspora Chinese architectural heritage reflects the CTSL (CNN + Transfer learning + Scenario3 + LBP) method in literature [55] and the FLM(Four convolution Layer) in literature [56] Model, direct retrieval of features extracted from the first stage image classification network TL ResNet50 and other methods were used to compare and analyze the experimental results. We hope that the returned image after image retrieval is as similar as possible to the original image and retain as many similar images as possible, so the average retrieval accuracy mAP and recall R@10 are used as evaluation indexes in this study. The experimental results are shown in Table 2.

Table 2. Retrieval accuracy of different methods on the image datasets of diaspora Chinese architectural heritage.

Network	Map (%)	R@10 (%)
CNNAR	76.6	19.7
CTSL [55]	70.4	18.5
FLM [56]	73.2	19.0
TL ResNet50	68.8	14.6

As can be seen from Table 2, the retrieval effect of the four models on the JMI building image datasets is relatively good. The lowest average retrieval accuracy mAP is 68.8% of TL ResNet50 network, and the highest is 76.6% of our CNNAR network. Therefore, it can be seen from the comparative experimental results that in the image retrieval experiment of JMI building image datasets, the method CNNAR proposed in this paper highlights the features of similar images between the same type of images through the attention mechanism on the basis of the first stage of image classification, and the retrieval effect is optimal, with the average retrieval accuracy of 76.6% and a recall rate of 19.7%. However, using the features of TL ResNet50 training directly for metric retrieval, the mAP is only 68.8%. The reason is that the trained network is only for four types of architectural images, and the extracted image features cannot distinguish between similar images, so the recall rate is the lowest. The mAP of the other two mainstream methods CTSL and FLM network models are 70.4% and 73.2%, respectively, which is lower than the retrieval accuracy of the network model proposed by us.

3.4. Attention Mechanism Ablation Experiment

In order to verify the impact on the accuracy of the network model in this paper after removing the attention mechanism, we test the attention mechanism ablation experiment in this section. After building image recognition and classification in JMI datasets through the image classification model task network, feature and metric retrieval between images of the same type in the image retrieval model task network are retrained. Then we take each image of the same type and the augmented four images as a small class and use the contrast loss function for training; the threshold is set to 1.5. The main purpose of feature extraction network using attention mechanism is to train the features extracted by the network to better recognize the differences between the same type of images. We use the image retrieval accuracy AP as the evaluation index of network effects in this stage. By testing each type of image data separately, the test experimental results are shown in Table 3.

Table 3. Ablation experiment accuracy.

Network	AP (%)			
	Fresco	Decorative Pattern	Chandelier Base Pattern	Architectural Style
No attention mechanism	73.0	75.6	80.5	74.7
channel attention mechanism	78.9	77.4	82.2	75.3
Spatial attention mechanism	76.5	76.7	82.0	75.0
Fusion attention mechanism	80.5	79.9	85.3	75.8

It can be seen from the ablation experiment results in Table 3 that the image retrieval accuracy without the attention mechanism is at least 73~80.5%, and the image retrieval accuracy with the attention mechanism is at most 75.8~85.3%. The attention fusion mechanism has greatly improved the retrieval accuracy of different types of architectural images, among which the accuracy of fresco has increased by up to 7.5%, the accuracy of decorative patterns has increased by 4.3%, the accuracy of chandelier base patterns has increased by 4.8%, and the accuracy of architectural style images has increased by only 1.1%. The retrieval accuracy of channel attention mechanism is slightly higher than that of spatial attention mechanism. Analysis of the above results shows that the background of architectural style images is relatively complex, and that there are many features and similarities in the images. The interference in extracting image features is large, so the accuracy improvement is minimal. The feature gap between mural images is obvious and adding the attention mechanism can effectively enhance feature representation. Therefore, in the image retrieval stage of this paper, it is best to select the fusion attention mechanism model network to extract the architectural image features of chandelier base patterns.

3.5. Top 10 Retrieved Images

After the experimental test of the JMI architectural image in the JMI datasets by the above different network models, this paper randomly extracts one image from the JMI architectural image datasets as the query image. We use our two-stage training image retrieval network framework CNNAR Framework, which combines deep learning with attention mechanism, to perform image retrieval. Figure 14 shows the top 10 results of similarity ranking of four overseas Chinese architectural image data retrieval. From the retrieval results, we can see that the retrieved image results are very similar to the query image. For example, in the Fresco image, the query image contains tall buildings, cruise ships, ports, trees, and other information, and the detection result image also contains the above information. In the decorative patterns image, the query image is a standard four petal lotus, and the retrieved results are all four petal lotus images. The above proves that the research method in this paper has excellent performance whether considering the classification effect or retrieval accuracy of architectural images. In view of the current state of disrepair and damage of the world cultural heritage of overseas Chinese architecture in Jiangmen City, through our architectural image retrieval research, we can provide an accurate repair plan for overseas Chinese architectural heritage images.

Image retrieval	Top 10 of retrieved result											
 Fresco						1-5						6-10
 Decorative patterns											1-10	
 Lamp base patterns											1-10	
 Architectural style											1-10	

Figure 14. Image retrieval results of top 10 architectural heritage.

3.6. Experiments on Public Datasets

In order to verify the generalization performance of the proposed CNNAR model, we used CNNAR, CTSL and FLM methods to conduct comparative experimental studies on Paris500K and Corel5K public datasets, respectively. The experimental results are shown in Tables 4 and 5.

Table 4. Retrieval accuracy of different models in Paris500K.

Network	mAP (%)	R@10 (%)
CNNAR	71.8	6.4
CTSL	71.2	6.1
FLM	73.4	7.5

Table 5. Retrieval accuracy of different models in Corel5K.

Network	mAP (%)	R@10 (%)
CNNAR	72.5	7.3
CTSL	70.2	6.6
FLM	73.9	8.0

Through experimental analysis, the retrieval accuracy of the CNNAR method proposed in this paper on the Paris500K and Corel5K, the two public datasets are 71.8% and 72.5%, respectively, which is slightly lower than the experimental performance of the JMI image datasets. According to the analysis of the reasons, there are few types of building images of diaspora Chinese architectural heritage, and they are all building types and images, while Paris500K has 79 different landmark images and Corel5K has 50 different types of images. When we perform image retrieval experiments, the accuracy of image classification

in the first stage decreases. Moreover, the images in the two public datasets contain various interference factors such as background and occlusion, which makes it difficult to ensure the accuracy of the research method in this paper. However, our image classification and retrieval method still have high retrieval accuracy and strong generalization performances, which can be further improved for model generalization ability in subsequent studies. The comparative experimental results on the Paris500K and Corel5K public datasets show that our CNNAR model has strong generalization ability and can also be effectively applied to other topics datasets. At present, the retrieval accuracy of our model in the above two public datasets is 71.8~72.5%. Therefore, in the subsequent research, we can enhance and supplement the building image data in the JMI datasets with other methods to further improve the generalization ability of our model.

4. Conclusions

In this paper, taking the Jiangmen diaspora Chinese architectural heritage image in South China as the research object, the architectural image recognition and retrieval based on deep learning and attention fusion mechanism are studied. The images of the diaspora Chinese architectural heritage are of great significance to history and humanity, to understanding the trends of thought of the times, data preservation, and image restoration, etc. Especially in the face of the serious situation of disrepair and damage faced by the world cultural heritage of diaspora Chinese architecture, it is of great significance to build an image database to effectively identify, classify and retrieve various architectural images. In order to effectively preserve diaspora Chinese architectural heritage images and improve the accuracy of model classification and retrieval, the following research has been done in this paper.

- (1) We have built a JMI architectural heritage image database containing architectural images such as frescos, decorative patterns, chandelier base patterns, and architectural styles. The method of adding Gaussian noise, salt and pepper noise, histogram equalization, and adaptive histogram equalization with limited contrast is used to expand the data, and the number of images in JMI database is expanded to 25,365. This database contains a large number of frescos and patterns with rich cultural connotations of diaspora Chinese, and the constructed architectural heritage image database can provide rich and reliable data support for the follow-up study of architectural trends of the times and the integration of Chinese and Western aesthetics.
- (2) In this paper, the parameters trained by Paris500K datasets image source network are migrated to JMI architectural heritage image dataset for image network training through the migration learning method. We used the ResNet50, the GoogLeNet and the VGG16, three excellent convolutional neural network models, to conduct migration training experiments in the JMI image dataset. The results show that the Resnet50 network with migration weight not only has the fastest convergence speed, but also the highest accuracy, of 98.3%. It is the best network for JMI building image datasets classification.
- (3) To solve the problem of small difference in image features of the same type of buildings, we propose a two-stage training image retrieval network framework CNNAR Framework network model based on deep learning and the attention mechanism. The CNNAR network is used to conduct image retrieval research on the JMI diaspora Chinese architectural heritage datasets, and at the same time, it is compared with several mainstream network model methods for experimental analysis. The analysis results show that the CNNAR retrieval method proposed in this paper has the best retrieval effect, with an average retrieval accuracy of 76.6% and a recall rate of 19.7%. The architectural image results retrieved by this method are highly similar to the query image. In view of the current state of disrepair and damage of the world cultural heritage of diaspora Chinese architecture in Jiangmen City, through our architectural image retrieval research, we can provide an accurate repair plan for diaspora Chinese architectural heritage images.

- (4) The experimental results of image retrieval on the Paris500K and the Corel5K public datasets show that our CNNAR model has a strong generalization ability and can be effectively applied to other topics datasets. In subsequent research, we can enhance and improve the building image data in the JMI datasets, so as to further improve the generalization ability of our model.

Author Contributions: Conceptualization, L.G. and T.Y.; methodology, L.G.; software, Y.W.; validation, L.G., Z.Z. and X.Z.; investigation, T.Y. and W.C.; resources, T.Y.; data curation, Y.W.; writing—original draft preparation, L.G.; writing—review and editing, T.Y.; visualization, Y.W. and X.Z.; supervision, T.Y. and Z.Z.; project administration, T.Y. and C.K.D.C.; funding acquisition, L.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by the Wuyi University-Hong Kong-Macao Unite Research Funds (Grant nos. 2019WGALH23) and the Wuyi University Youth Team Funds (Grant nos. 2019td10) and the Guangdong Province Philosophy and Social Science Planning Discipline Joint Project (Grant nos. GD20XSH06) and National I & E Program for College Student (Grant nos. 202211349032).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: In this study, the authors used two publicly available datasets for analysis: (1) Paris500K which has been deposited on the website <https://www.vision.rwth-aachen.de/page/paris500k>, accessed on 5 October 2022. (2) Corel5K which has been deposited on the website <https://github.com/watersink/Corel5K>, accessed on 5 October 2022.

Conflicts of Interest: The authors declare that there are no conflict of interest regarding the publication of this paper.

References

- Caciora, T.; Herman, G.V.; Ilies, A.; Baias, S.; Ilies, D.C.; Josan, I.; Hodor, N. The use of virtual reality to promote sustainable tourism: A case study of wooden churches historical monuments from Romania. *Remote Sens.* **2021**, *13*, 1758. [CrossRef]
- Li, J.Y.; Huang, X.; Tu, L.L.; Zhang, T.; Wang, L.G. A review of building detecting from very high resolution optical remote sensing images. *Giscience Remote Sens.* **2022**, *59*, 1199–1225. [CrossRef]
- Cai, Y.M.; Ding, Y.L.; Zhang, H.W.; Xiu, J.H.; Liu, Z.M. Geo-Location algorithm for building targets in oblique remote sensing images based on deep learning and height estimation. *Remote Sens.* **2020**, *12*, 2427. [CrossRef]
- Munawar, H.S.; Aggarwal, R.; Qadir, Z.; Khan, S.I.; Kouzani, A.Z.; Mahmud, M.A.P. A gabor filter-based protocol for automated image-based building detection. *Buildings* **2021**, *11*, 302. [CrossRef]
- Cao, D.G.; Xing, H.F.; Wong, M.S.; Kwan, M.P.; Xing, H.Q.; Meng, Y. A stacking ensemble deep learning model for building extraction from remote sensing images. *Remote Sens.* **2021**, *13*, 3898. [CrossRef]
- Khoshboresh-Masouleh, M.; Alidoost, F.; Arefi, H. Multiscale building segmentation based on deep learning for remote sensing RGB images from different sensors. *J. Appl. Remote Sens.* **2020**, *14*, 034503. [CrossRef]
- Kwak, Y.; Yun, W.; Kim, J.; Cho, H.; Park, J.; Choi, M.; Jung, S.; Kim, J. Quantum distributed deep learning architectures: Models, discussions, and applications. *ICT Express* **2022**, *in press*. [CrossRef]
- Coulibaly, S.; Foguem, B.K.; Kamissoko, D.; Traore, D. Deep learning for precision agriculture: A bibliometric analysis. *Intell. Syst. Appl.* **2022**, *16*, 200102. [CrossRef]
- Arora, T.K.; Chaubey, P.K.; Raman, M.S.; Kumar, B.; Nagesh, Y.; Anjani, P.K.; Ahmed, H.M.; Hashmi, A. Optimal facial feature based emotional recognition using deep learning algorithm. *Comput. Intell. Neurosci.* **2022**, *2022*, 8379202. [CrossRef]
- Balogh, Z.A.; Kis, B.J. Comparison of cT noise reduction performances with deep learning-based, conventional, and combined denoising algorithms. *Med. Eng. Phys.* **2022**, *109*, 103897. [CrossRef]
- Gao, L.; Huang, Y.; Zhang, X.; Liu, Q.; Chen, Z. Prediction of Prospectin Target Based on ResNet Convolutional Neural Network. *Appl. Sci.* **2022**, *12*, 11433. [CrossRef]
- Jackulin, C.; Murugavalli, S. A comprehensive review on detection of plant disease using machine learning and deep learning approaches. *Meas. Sens.* **2022**, *24*, 100441. [CrossRef]
- Huang, Y.; Feng, Q.; Zhang, W.; Zhang, L.; Gao, L. Prediction of prospecting target based on selective transfer network. *Minerals* **2022**, *12*, 1112. [CrossRef]
- Hameed, I.M.; Abdhussain, S.H.; Mahmmud, B.M. Content-based image retrieval: A review of recent trends. *Cogent Eng.* **2021**, *8*, 1927469. [CrossRef]

15. Aziz, M.A.; Ewees, A.A.; Hassanien, A.E. Multi-objective whale optimization algorithm for content-based image retrieval. *Multimed. Tools Appl.* **2018**, *77*, 26135–26172. [[CrossRef](#)]
16. Fu, R.; Li, B.; Gao, Y.; Wang, P. Content-based image retrieval based on CNN and SVM. In Proceedings of the 2016 2nd IEEE International conference on computer and communications (ICCC), Chengdu, China, 14–17 October 2016; pp. 638–642. [[CrossRef](#)]
17. Kilic, S.; Askerzade, I.; Kaya, Y. Using ResNet transfer deep learning methods in person identification according to physical actions. *IEEE Access* **2020**, *8*, 220364–220373. [[CrossRef](#)]
18. Hua, C.; Chen, S.; Xu, G.; Lu, Y.; Du, B. Defect identification method of carbon fiber sucker rod based on GoogLeNet-based deep learning model and transfer learning. *Mater. Commun.* **2022**, *33*, 104228. [[CrossRef](#)]
19. Prasetyo, E.; Suciati, N.; Fatichah, C. Multi-level residual network VGGNet for fish species classification. *J. King Saud Univ.-Comput. Inf. Sci.* **2022**, *34*, 5286–5295. [[CrossRef](#)]
20. Shi, Z.; Wang, F.; Wang, Y.; Jia, H. Image emotion recognition research based on separable convolution attention mechanism (SCAM) neural network. *Laser J.* **2022**, *43*, 88–93. [[CrossRef](#)]
21. Lan, Y.; Peng, B.; Wu, X.; Teng, F. Infrared dim and small targets detection via self-attention mechanism and pipeline correlator. *Digit. Signal Process.* **2022**, *130*, 103733. [[CrossRef](#)]
22. Vanian, V.; Zamanakos, G.; Pratikakis, I. Improving performance of deep learning model for 3d point cloud semantic segmentation via attention mechanisms. *Comput. Graph.* **2022**, *106*, 277–287. [[CrossRef](#)]
23. Wang, Y.; Feng, Y.; Zhang, L.; Zhou, J.; Yong, L.; Goh, S.M.R.; Zhen, L. Adversarial multimodal fusion with attention mechanism for skin lesion classification using clinical and dermoscopic images. *Med. Image Anal.* **2022**, *81*, 102535. [[CrossRef](#)]
24. Wang, X.; Yuan, Y.; Guo, D.; Huang, X.; Cui, Y.; Xia, M.; Wang, Z.; Bai, C.; Chen, S. SSA-Net: Spatial self-attention network for COVID-19 pneumonia infection segmentation with semi-supervised few-shot learning. *Med. Image Anal.* **2022**, *79*, 102459. [[CrossRef](#)] [[PubMed](#)]
25. Ma, K.; Wang, B.W.; Li, Y.Q.; Zhang, J.X. Image retrieval for local architectural heritage recommendation based on deep hashing. *Buildings* **2022**, *12*, 809. [[CrossRef](#)]
26. Wang, Y.S.; Hu, X. Machine learning-base image recognition for rural architectural planning and design. *Neural Comput. Appl.* **2022**, 1–10. [[CrossRef](#)]
27. Xie, X.S.; Wen, X.; Deng, F.F. Applications of 3D image using internet of things in the exhibition of classical architecture art style. *Mob. Inf. Syst.* **2021**, *2021*, 2283354. [[CrossRef](#)]
28. Llamas, J.; Leronés, P.M.; Medina, R.; Zalama, E.; Gomez-Garcia-Bermejo, J. Classification of architectural heritage images using deep learning techniques. *Appl. Sci.* **2017**, *7*, 992. [[CrossRef](#)]
29. Wang, Y.J.; Li, S.C.; Teng, F.; Lin, Y.H.; Wang, M.J.; Cai, H.F. Improved mask R-CNN for rural building roof type recognition from UAV high-resolution images: A case study in hunan province, China. *Remote Sens.* **2022**, *14*, 265. [[CrossRef](#)]
30. Hong, Z.H.; Zhong, H.Z.; Pan, H.Y.; Liu, J.; Zhou, R.Y.; Zhang, Y.; Han, Y.L.; Wang, J.; Yang, S.H.; Zhong, C.Y. Classification of building damage using a novel convolutional neural network based on post-disaster aerial images. *Sensors* **2022**, *22*, 5920. [[CrossRef](#)]
31. Taoufiq, S.; Nagy, B.; Benedek, C. HierarchyNet: Hierarchical CNN-based urban building classification. *Remote Sens.* **2020**, *12*, 3794. [[CrossRef](#)]
32. Weyand, T.; Leibe, B.T. Visual landmark recognition from internet photo collections: A large-scale evaluation. *Comput. Vis. Image Underst.* **2015**, *135*, 1–15. [[CrossRef](#)]
33. Jiu, M.; Sahbi, H. Context-aware deep kernel networks for image annotation. *Neurocomputing* **2022**, *474*, 154–167. [[CrossRef](#)]
34. Gupta, A.; Pawade, P.; Balakrishnan, R. Deep residual network and transfer learning-based person re-identification. *Intell. Syst. Appl.* **2022**, *10*, 200137. [[CrossRef](#)]
35. Li, Z.C.; Dong, J.W. A framework integrating deeplabV3+, transfer learning, active learning, and incremental learning for mapping building footprints. *Remote Sens.* **2022**, *14*, 4738. [[CrossRef](#)]
36. Huang, M.; Yin, J.; Yan, S.; Xue, P. A fault diagnosis method of bearings based on deep transfer learning. *Simul. Model. Pract. Theory* **2023**, *122*, 102659. [[CrossRef](#)]
37. Peng, L.; Wu, H.; Gao, M.; Yi, H.; Xiong, Q.; Yang, L.; Cheng, S. TLT: Recurrent fine-tuning transfer learning for water quality long-term prediction. *Water Res.* **2022**, *225*, 119171. [[CrossRef](#)]
38. Zhu, C.; Ni, J.; Yang, Z.; Sheng, Y.; Yang, J.; Zhang, W. Bandgap prediction on small thermoelectric material dataset via instance-based transfer learning. *Comput. Theor. Chem.* **2022**, *1217*, 113872. [[CrossRef](#)]
39. Yu, X.; Wang, J.; Hong, Q.; Teku, R.; Wang, S.; Zhang, Y. Transfer learning for medical images analyses: A survey. *Neurocomputing* **2022**, *489*, 230–254. [[CrossRef](#)]
40. He, K.M.; Zhang, X.; Ren, S. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [[CrossRef](#)]
41. Wang, Y.; Xie, Y.; Zeng, J.; Wang, H.; Fan, L.; Song, Y. Cross-modal fusion for multi-label image classification with attention mechanism. *Comput. Electr. Eng.* **2022**, *101*, 108002. [[CrossRef](#)]
42. Chen, Q.; Fan, J.; Chen, W. An improved image enhancement framework based on multiple attention mechanism. *Displays* **2021**, *70*, 102091. [[CrossRef](#)]
43. Zhang, S.; Deng, X.; Lu, Y.; Hong, S.; Kong, Z.; Peng, Y.; Luo, Y. A channel attention based deep neural network for automatic metallic corrosion detection. *J. Build. Eng.* **2021**, *42*, 103046. [[CrossRef](#)]

44. Qiu, Z.; Becker, S.I.; Pegna, A.J. Spatial attention shifting to fearful faces depends on visual awareness in attentional blink: An ERP study. *Neuropsychologia* **2022**, *172*, 108283. [[CrossRef](#)] [[PubMed](#)]
45. Cheng, Y.; Wang, H. A modified contrastive loss method for face recognition. *Pattern Recognit. Lett.* **2019**, *125*, 785–790. [[CrossRef](#)]
46. Zhang, Z.C.; Wang, H.B.; Wang, N.B. Sample extraction and method with feature reconstruction and deformation information. *Appl. Intell.* **2022**, *52*, 15916–15928. [[CrossRef](#)]
47. Yu, W.X.; Lu, Y.; Wang, J.N. Application of small sample virtual expansion and spherical mapping model in wind turbine fault diagnosis. *Expert Syst. Appl.* **2021**, *183*, 115397. [[CrossRef](#)]
48. Koyuncu, H.; Ceylan, R. Elimination of white gaussian noise in arterial phase CT images to bring adrenal tumours into the forefront. *Comput. Med. Imaging Graph.* **2018**, *65*, 46–57. [[CrossRef](#)]
49. Piroozmandan, M.M.; Farokhi, F.; Kangarloo, K.; Jahanshahi, M. Removing the impulse noise from images based on fuzzy cellular automata by using a two-phase innovative method. *Optik* **2022**, *255*, 168713. [[CrossRef](#)]
50. Vijayalakshmi, D.; Nath, M.K. A novel multilevel framework based contrast enhancement for uniform and non-uniform background images using a suitable histogram equalization. *Digit. Signal Process.* **2022**, *127*, 103532. [[CrossRef](#)]
51. Ullah, Z.; Farooq, M.U.; Lee, S.; An, D. A hybrid image enhancement based brain MRI images classification technique. *Med. Hypotheses* **2020**, *143*, 109922. [[CrossRef](#)]
52. Zhao, J.H.; Wang, X.; Dou, X.T.; Zhao, Y.X.; Fu, Z.X.; Guo, M.; Zhang, R.J. A high-precision image classification network model based on a voting mechanism. *Int. J. Digit. Earth* **2022**, *15*, 2168–2183. [[CrossRef](#)]
53. Ma, J.W.; Czerniawski, T.; Lite, F. An application of metadata-based image retrieval system for facility management. *Adv. Eng. Inform.* **2021**, *50*, 101417. [[CrossRef](#)]
54. Sun, W.W.; Wang, H.Q.; Lu, Y.; Luo, J.S.; Liu, T.; Lin, J.Z.; Pang, Y.; Zhang, G. Deep-learning-based complex scene text detection algorithm for architectural images. *Mathematics* **2022**, *10*, 3914. [[CrossRef](#)]
55. Khatami, A.; Babaie, M.; Tizhoosh, H.R.; Khosravi, A.; Nguyen, T.; Nahavandi, S. A sequential search-space shrinking using CNN transfer learning and a radon projection pool for medical image retrieval. *Expert Syst. Appl.* **2018**, *100*, 224–233. [[CrossRef](#)]
56. Singh, P.; Hrisheekesha, P.N.; Singh, V.K. CBIR-CNN: Content-based image retrieval on celebrity data using deep convolution neural network. *Recent Adv. Comput. Sci. Commun.* **2021**, *14*, 257–272. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.