Hindawi
Mobile Information Systems
Volume 2022, Article ID 8952381, 11 pages
https://doi.org/10.1155/2022/8952381

*Research Article*

# Classification of Ancient Buddhist Architecture in Multi-Cultural Context Based on Local Feature Learning

**Yali Wu**

*Sichuan University, Chengdu, Sichuan 610041, China*

Correspondence should be addressed to Yali Wu; wuyali@stu.scu.edu.cn

Received 7 April 2022; Accepted 6 May 2022; Published 23 May 2022

Academic Editor: Chia-Huei Wu



Ancient Buddhist architecture plays an important role in the development of Chinese architectural culture. Under the background of multiculturalism, the ancient Buddhist architectural style has also been influenced to varying degrees. In order to realize automatic classification of ancient Buddhist architecture under multi-cultural background, this paper proposes an automatic classification algorithm based on local feature learning. Firstly, the ancient Buddhist architecture images are gridded, so that the backbone network can obtain relatively flat ancient Buddhist architecture image blocks. At the same time, the backbone network can learn more local details. Then, the grid reconstruction module is designed to strengthen the connection between the features of each block and highlight the distinguishing detail features. The accuracy of ancient Buddhist architecture classification can be effectively improved through image meshing and mesh reconstruction. Experiment and analysis are carried out by using the dataset of ancient Buddhist architecture images on the Internet. Experimental results show that the proposed algorithm has better recognition accuracy and robustness than other comparison algorithms.

## 1. Introduction

Buddhism, as a foreign culture, originated from India and flourished in the Han and Tang dynasties of China. In the course of the development of Chinese history, Buddhism was closely combined with Confucianism, Taoism, and other diversified cultures and gradually formed the Buddhist culture with Chinese characteristics [1]. At the same time, the ancient Buddhist architecture also continuously permeates and influences the whole Chinese society and traditional culture. Therefore, ancient Buddhist architecture is an indispensable part of the study of China's traditional architectural style and culture [2]. On the material level, ancient Buddhist architecture is the material carrier of Buddhist culture. From the dominant level, ancient Buddhist architecture is the architectural expression of Buddhist culture. The change of ancient Buddhist architecture undoubtedly reflects the relationship between Buddhism and Chinese traditional culture, that is, the process from confrontation and conflict to gradual integration. In terms of culture, Buddhist architecture plays an important role in the development of Chinese architectural culture. At the same time, Buddhist culture and architecture also have an important impact on people's lives [3]. Nowadays, with the rapid development of social economy and culture, all kinds of traditional culture have been greatly impacted. Compared with ancient Chinese temples, modern temples are difficult to match in both quantity and architectural art. Therefore, the study of the style of ancient Buddhist architecture can complement and perfect the content of ancient Buddhist architecture system more comprehensively.

With the increasing frequency of religion, Confucianism and Taoism, the phenomenon of multi-culture arises at the historic moment [4]. Multiculturalism is a compound word that consists of "pluralism" and "culture." Pluralism is different from diversity. Diversity usually describes different states or forms of existence of things. Pluralism mainly reflects the differences in the nature of things. Culture is one of the most frequently used concepts in many subjects. From a philosophical point of view, the essence of culture is "humanization." Man changes nature according to his own needs and makes it suitable for human survival and development. It has traces of human existence and development, human nature, history, and subjectivity. The heterogeneity of the value orientation of cultural subjects often makes different cultures different. However, different cultures are not completely mutually exclusive, and there are some commonalities between them. In dealing with different cultures, we should adopt a dialectical and unified attitude and seek common ground on the premise of recognizing different cultural differences, so as to achieve harmony and difference. Different forms of culture have different values, and different cultures are not simply against each other. They are interrelated and influence each other to meet the cultural needs of different groups. It is of great significance to study the style of ancient Buddhist architecture under the multi-cultural background. How to identify the style of ancient Buddhist architecture through image classification will be the main content of this paper.

Image classification refers to the classification of images into a certain category according to the information in the image. Therefore, the extraction of image feature information is an important research content of image classification. Traditional image classification mainly uses machine learning to extract features. With the continuous development of deep learning, various deep learning algorithms are gradually applied to image classification. In 2012, AlexNet neural network [5] surpassed traditional methods in image classification effect. After AlexNet, a series of improved convolutional neural network (CNN) models [6] emerged to continuously improve the classification accuracy. However, there are some defects in CNN's model [7]. Firstly, the pooling layer of CNN will lead to the loss of a large number of valuable feature information, thus affecting the classification accuracy. Secondly, CNN is insensitive to location information, which leads to its weak ability to recognize spatial relations between objects. The capsule network [8] proposed subsequently can better deal with the above problems. The capsule network abandons the pooling layer of CNN and retains a large amount of picture information, which makes the capsule network useless training data to achieve the ideal effect. In addition, the capsule network is a partial prediction of the whole. In the prediction process, it can better retain the attitude of features, such as location, size, direction, and other information. This enables the capsule network not only to carry out more accurate classification but also to effectively identify the image after a series of spatial transformation such as affine transformation. However, due to the high cost of computation and memory load, the capsule network has a relatively shallow structure and is mainly suitable for simple datasets, but does not perform well in processing complex data.

In recent years, it has been found that comprehensive use of features at different levels extracted from depth models can improve the classification effect compared with only using the features at the highest level. CNN is used in literature [9] to extract template features and apply them to classification tasks. This method splices the features extracted from the last pooling layer and the full connection layer in CNN as the final feature representation. Compared with some representative methods, it has a higher classification accuracy. An idea of cross-layer connection on the basis of traditional CNN was introduced in literature [10]. It directly connects the features of the second pooling layer in CNN with the full connection layer across the hidden layer in the middle of the model and finally uses this feature to predict the sample category. This method can effectively combine high-level features and low-level features, and achieve higher accuracy than traditional CNN in classification tasks. Literature [11] proposed a recognition method based on cross-connected LeNet-5 network to solve the problem of low recognition rate of LeNet-5. This method can fuse the low-level features and high-level features extracted by neural network and improve the recognition rate. Recently, some studies have shown that the combination of prototype learning and deep learning can extract the discriminant features with small differences within classes and large differences between classes. Literature [12] combines prototype learning with CNN and proposes convolutional prototype learning. This model can significantly reduce the intra-class differences of features extracted by CNN and improve the robustness of CNN. A prototype integration method based on prototype learning was proposed in literature [13]. This method can not only narrow the intra-class differences of depth features but also enlarge the inter-class differences, thus improving the robustness of the detection of new categories in incremental learning. Literature [14] proposed a deep coding and classification model in the case of modal loss. This model makes full use of the idea of complete information coding and prototype classification to improve the performance of multi-modal data classification in the case of modal loss. It can be seen that prototype learning and deep learning are combined for classification tasks, and features of different levels extracted from the deep model are comprehensively utilized in the classification stage. It not only obtains and discriminates features but also further improves classification accuracy.

However, the above classification methods have some limitations in local feature extraction ability, comprehensive utilization degree of local features, amount of calculation, and final accuracy. In this paper, a learning model for grid reconstruction is established based on differential theory to extract fine-grained local features. This algorithm is applied to the classification and recognition of ancient Buddhist architectural styles. The results show that the proposed algorithm can effectively improve the classification and recognition accuracy of ancient Buddhist architectural styles.

The innovations and contributions of this paper are listed as follows:

(1) The ancient Buddhist architecture images are gridded, so that the backbone network can obtain relatively flat ancient Buddhist architecture image blocks, and it can learn more local details.

(2) The grid reconstruction module is designed to strengthen the connection between the features of each block and highlight the distinguishing detail features.

(3) The accuracy of ancient Buddhist architecture classification can be effectively improved through image meshing and mesh reconstruction.

This paper consists of five main parts: the first part is the introduction, and the second part is the state of the art, which is classification of ancient Buddhist architectural styles. The third part is the recognition methodology of ancient Buddhist architectural style based on local feature learning. The fourth part is the experiment and analysis, and the fifth part is the conclusion.

## 2. State of the Art

After Buddhism was introduced to China since the Western Han dynasty, it was originally only revered by the court. Later, with the increase of monks, the government and the people gradually formed a custom of building temples, pagodas, and grottoes. According to the spreading routes and schools of Buddhism, Chinese Buddhist architecture can be divided into three categories. The first category is the Han Buddhist temples, which are more numerous and widely distributed. The second type is Tibetan monasteries, mainly distributed in Tibet, Inner Mongolia, Qinghai, Gansu, Sichuan, Yunnan, and other places. The third category is the Southern Theravada temples, mainly distributed in the southwest of Yunnan Province. This paper focuses on the analysis of Chinese Buddhist architecture and divides ancient Buddhist architectural styles into three categories, namely, official architectural style, minority architectural style, and integrated architectural style [15]. Examples of these three types of buildings are shown in Figure 1.

The influence of multi-culture on the main architectural style of the temple is very great. This is mainly because under the influence of Confucianism, Buddhism, Taoism, and worship, the temple presents different styles and characteristics. At the same time, combined with the aesthetic characteristics of the local ethnic group, it makes full use of local building materials and decorative techniques. This prevents the overall architectural style from being generalized as plain or gorgeous. Table 1 shows the specific situation of ancient temples investigated in a certain province of China.

As can be seen from Figure 1 and Table 2, among the 32 typical monasteries surveyed, official style accounts for 25%, integrated style for 60%, and minority style for 15%. Most monasteries adopt a comprehensive architectural style. Its building roof is mostly grey small green tile, color to white, red, and dark grey. The larger, more important monasteries were built in the official style. It is decorated with red or yellow glazed tiles, and the walls are mostly red or yellow, imposing, and resplendent. Temples in Western Hunan usually adopt the architectural style of ethnic minorities, with exaggerated roof warping and far-reaching eaves.

During the historical development of Buddhism, many sects spread and developed in China. There is no doubt that the architectural style of the temple has a great relationship with the religious sect of the temple. And the orientation of the monastery sect practice is related to the high monks living in tin. For example, if a abbot practices Zen, the temple may develop a Zen monastic style. In addition, the architectural style of the temple will change with the development of history. For example, a temple in the pure land style may gradually evolve into a Zen temple if it is managed by a Zen monk in later generations. On the basis of the original hall, it added space for cultivation such as Zen Hall and Dharma Hall. In this case, posterity can only determine its specific type according to the existing main style of the temple.

Among the 32 temples surveyed, there are 17 Zen temples, accounting for 53.12% of the total. There are 8 monasteries with double cultivation of Zen and Jing, accounting for 25%. There are 3 pure land temples, accounting for 9.38%. There were 4 temples of the rooftop sect or other types, accounting for 12.5%. These data are basically consistent with the view that Buddhism is dominated by Zen and pure land monasteries. In ancient temples, the architecture of Zen temples mostly follows the idea of "emptiness, nothingness, and harmony" of Zen. In terms of architectural color and material use, it generally adopts a comprehensive architectural style. Pure land temple also basically follows the architectural style of Han temple. However, the pure land school of Buddhism takes reciting Buddha as the main practice mode, specially called "Amitabha Buddha" in order to live in the Western paradise. The pure land sect believes that the Western paradise is full of precious jewels and the architecture is golden, so the pure land sect temples are relatively ornate. The temple of Chanjing Shuangxiu combines the architectural styles of both. It has both the simplicity of Zen temple and the solemnity of pure land temple in the main hall. Such as Shimen Jiashan Temple for Zen double repair temple, Nanyue Temple is a Zen temple. Even one architectural style is relatively simple, and the use of architectural styles is integrated. It can be seen that different schools of Buddhism have little influence on architectural styles.

## 3. Methodology

This paper proposes a network model as shown in Figure 2 to realize automatic classification of ancient Buddhist architectural styles. Firstly, the ancient Buddhist architecture images are gridded to obtain relatively flat ancient Buddhist architecture image blocks in backbone network. This solves the problem of large space between classes and small space between classes in ancient Buddhist architecture. At the same time, the backbone network can learn more local details. However, grid will destroy the integrity of ancient Buddhist architectural structure, so the grid reconstruction module is redesigned to strengthen the connection between the features of each block and highlight the distinguishing details. Through the above two parts of image grid of ancient Buddhist architecture and grid reconstruction, the accuracy of ancient Buddhist architecture classification can be effectively improved.

*3.1. Grid Processing Method.* The complex characteristics of ancient Buddhist architectural styles bring great challenges to its classification. To solve this problem, based on the idea

(a) (b) (c)

Figure 1: Three styles of ancient Buddhist architecture. (a) Official style. (b) Minority style. (c) Integrated style.

Table 1: Typical architectural style of the ancient temple-type table.

| No. | Name of the temple | Architectural style types |
|---|---|---|
| 1 | South YueMiao | Official style |
| 2 | Nanyue Zhusheng Temple | Official style |
| 3 | Lushan Temple in Changsha | Official style |
| 4 | Kaifu Temple in Changsha | Official style |
| 5 | Liuyang Shishuang Temple | Integrated style |
| 6 | Longxing Temple in Yuanling | Minority style |
| 7 | Dayong Puguang Temple | Minority style |
| 8 | Nanyue Nantai Temple | Integrated style |
| 9 | Fuyan Temple of Nanyue | Integrated style |
| 10 | Shimen Jiashan Temple | Official style |
| 11 | Youxian Baoning Temple | Integrated style |
| 12 | Mount Nanyue Seal Temple | Integrated style |
| 13 | Nanyue Sutras Hall | Official style |
| 14 | Nanyue Fangguang Temple | Integrated style |
| 15 | Nanyue Gaotai Temple | Integrated style |
| 16 | Iron Buddha Temple of Nanyue | Integrated style |
| 17 | Nanyue Wuyue Temple | Integrated style |
| 18 | Xiangnan Temple of Nanyue | Integrated style |
| 19 | Zhurong Temple of Nanyue | Integrated style |
| 20 | Guangji Temple, Nanyue | Integrated style |
| 21 | Baiyuan Temple in Yuanling | Minority style |
| 22 | Phoenix Temple in Yuanling | Minority style |
| 23 | Longquan Ancient Temple in Yuanling | Minority style |
| 24 | Xiangtan Zhaoshan Temple | Integrated style |
| 25 | Nanyue Dashan Temple | Integrated style |
| 26 | Yongzhou Blue Mountain Pagoda Temple | Integrated style |
| 27 | Nanyue Shoudian Buddha | Integrated style |
| 28 | Liuyang Baogai Temple | Integrated style |
| 29 | Ningxiang Miyin Temple | Integrated style |
| 30 | Xiangxiang Yunmen Temple | Integrated style |
| 31 | Tielu Temple in Changsha | Official style |
| 32 | Changsha Xixin Zen Temple | Official style |

Table 2: Experimental results under different segmentation parameters (%).

| $T$ | ACC | SEN | SPE | PPV | NPV |
|---|---|---|---|---|---|
| 1 | 80.24 | 83.18 | 77.21 | 79.9 | 80.63 |
| 2 | 80.96 | 86.21 | 75.25 | 80.72 | 81.26 |
| 3 | 82.41 | 85.81 | 77.51 | 81.18 | 82.11 |
| 4 | 80.63 | 83.64 | 77.43 | 80.84 | 80.39 |
| 5 | 80.44 | 84.94 | 75.62 | 80.29 | 80.63 |

of differential approximation, the ancient Buddhist architectural images to be classified are grid-processed. The image is divided into several relatively flat local areas to reduce the interference of feature extraction caused by the complex appearance of ancient Buddhist architecture.

Taking the segmentation process shown in Figure 3 as an example, the input image $X$ is firstly divided into $T \times T$ grids. If $T$ is too small and the mesh is too thick, the obvious area will be retained, as shown in Figure 3(a). With the increase of $T$, the local features of ancient Buddhist architecture gradually approach a clear state in the grid. However, if $T$ is too large, too many blank grids will be generated and too many invalid features will be introduced, as shown in Figure 3(c). Considering the general distribution of ancient Buddhist architectural images, $T = 3$ is set in this paper, and the 9 grids obtained are represented by $R_z$ and $k \in [1, 9]$, respectively, where, starting from $R_1$ in the upper left corner, it is numbered clockwise to $R_8$, and the middle grid is numbered $R_9$. However, simple grid segmentation can lead to the loss of the associated information between grids. Therefore, the algorithm in this paper needs to set the overlapping area at the neighboring meshes when cutting. Thus, it can retain the structural features between the meshes, as shown in Figure 3(d).

*3.2. Network Structure Based on Local Feature Learning.* Different from natural images, the distinct characteristics of small differences between different categories (especially local areas) of ancient Buddhist architectural images result in that no local grid can independently cover the subject semantic information of original ancient Buddhist architectural images. Therefore, while extracting independent grid features, the algorithm in this paper also needs to have the ability to perceive the membership relationship between the local and the whole of ancient Buddhist architecture. The global structure can be preserved to some extent by the overlapping redundant information between adjacent meshes that needs to be specially preserved during segmentation. However, if the feature extraction is biased to the overlapping region, the global optimal solution cannot be obtained, which affects the accuracy of subsequent classification. Therefore, a grid reconstruction module is specially designed to compensate for the global feature loss caused by meshing.

In this algorithm, $j = F(.)$, a residual network with shared weight $\Theta$ is adopted. After feature maps were extracted from

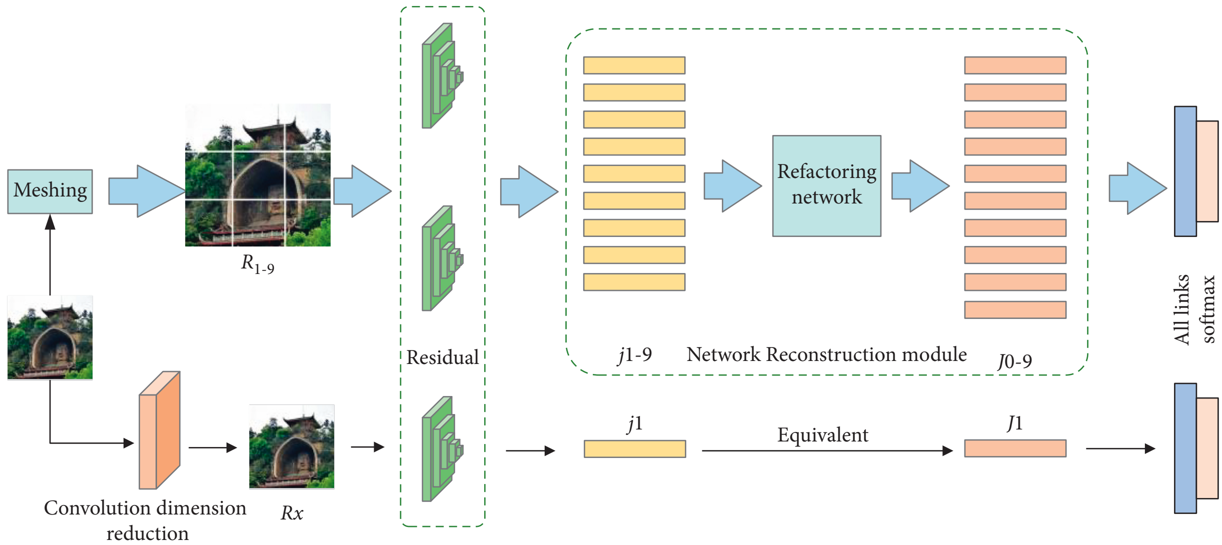



Figure 2: The deep learning model proposed in this paper.

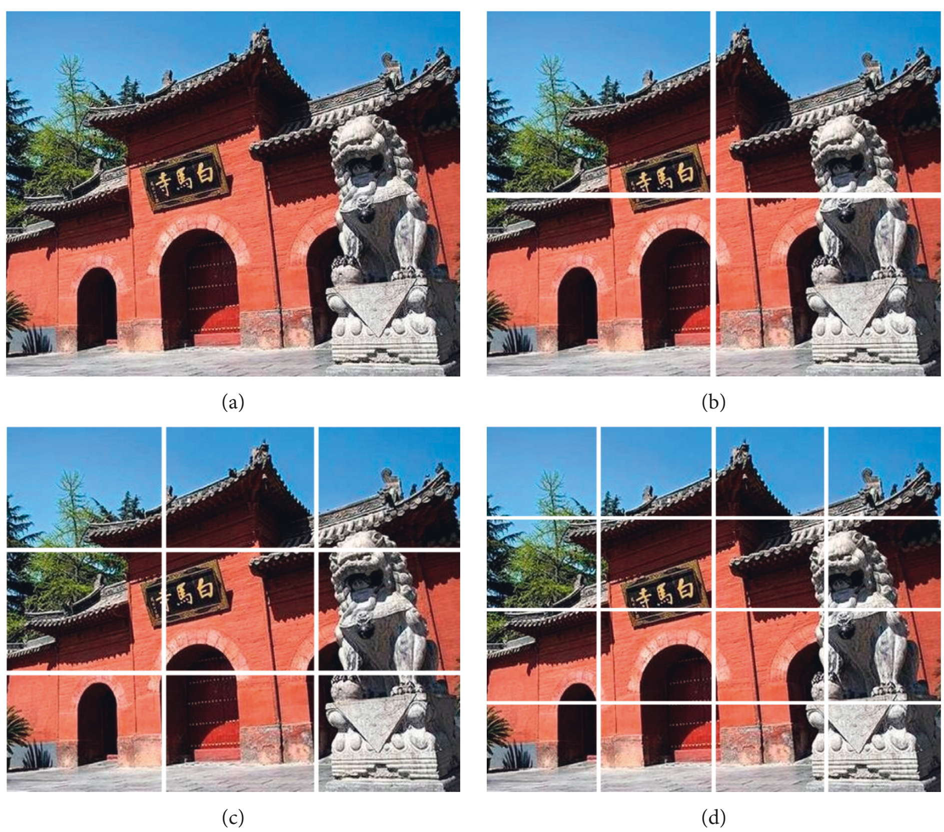



Figure 3: Processing effect of Buddhist architecture image grid. (a) $T = 2$. (b) $T = 3$. (c) $T = 4$. (d) $T = 5$.

the two branches, local feature $j_{1-9}$ and global feature $j_x$ were obtained by global leveling.

$$j_z = F(\Theta, R_z), z \in \{1 \sim 9, X\}, \quad (1)$$

$R_X$ is the global image after dimensionality reduction. $J_z$ is the feature vector extracted from the residual network, and the number of channels $C$ is determined by the number of convolution kernels at the output layer of the residual network.

Subsequently, the proposed algorithm reconstructs the network through the attention mechanism shown in Figure 4. The weight of each grid's contribution to the final classification was obtained by the training of formula (2), and the feature vector $q_z$ was constructed.where $M_z$ and $h_{\mathbf{z}}$ are the model weights and bias parameters of grid $z$. $\times$ means Hadamard multiplication.

$$q_z = \frac{1}{1 + \exp(m_z \times j_z + h_z)} \times j_z, \tag{2}$$

The recombination of local features is achieved by two branches. On the one hand, the features screened by adjacent grids are fused with formula (3) to strengthen the association and further feature screening is carried out.

$$J_z = m_z^* (j_z + q_{z-1}) + h_z'. \tag{3}$$

where $q_{z-1}$ is the feature filtering result of $j_{z-1}$. $m_z^{'}$ and $h_z^{'}$ are parameters of the full connection layer.$^*$is matrix multiplication.

On the other hand, $T \times T$ $C$ channel feature vector $q_z$ is reassembled into $T \times T \times C$ feature graph according to the segmentation sequence, and then, the global feature $J_0$ is obtained through maximum pooling.

$$J_0 = \max(q_z^w), w \in [1, C]. \tag{4}$$

*3.3. Loss Function Design.* Among the images to be classified, ancient Buddhist buildings are mostly in the middle of the image. However, the information content of ancient Buddhist architecture will be different according to the different areas covered by each grid, and the contribution to classification will also be different. Therefore, in the training process, the algorithm in this paper adopts formula (5) to set the loss weight $m_z^{''}$ of each feature of the upper branch.

$$m_z^{''} = \begin{cases} z, & m_z, \\ 10, & z = \{1 \sim 9\}. \end{cases} \tag{5}$$

Among them, $J_0$ as the global feature has the largest weight. In addition, $q_z$ always contains $q_{z-1}$ information. Therefore, the loss weight of features $J_{1\sim g}$ increases in a clockwise direction, while the lower branch learns the whole image and gives a large weight at the beginning of training. Therefore, the scale factor $\beta$ is used to adjust the weight proportion of upper and lower branches. At the beginning of the training, the lower branches lost $L_x$ weight, and when the training cycle $e$ reached $\lambda$, the upper and lower branches gradually increased to equal weight. The total loss function is shown in

$$L_S = \sum_{z=0}^{9} \min\left(1, eiu(e - \lambda) + m_z^{''*} \beta\right)^* L_z + L_X. \tag{6}$$

Among them, $L_z$ and $L_X$ were obtained by calculating the cross-entropy of

$$L = \frac{1}{H} \sum_{h=1}^{H} n_h \log(u_h). \tag{7}$$

*3.4. Implementation Steps of the Algorithm in This Paper.* The model training sample is composed of the original input image $X$ and its corresponding ancient Buddhist building number $n$, denoted as $<x, n>$, $n \in \{1, \ldots, 22, I, J\}$. The algorithm model in this paper maps the input image to the probability vector $u$ that it belongs to various ancient Buddhist buildings through the function $u = \Phi(X, \Omega)$. $\Omega$ (including $\Theta$) represents all learnable parameters in the classification model.

*Step 1.* Cut the complete image $X$ into 9 meshes of the same size $\{R_z\}$, $z \in [1, 9]$. Meanwhile, the complete image is downsampled to the same size of the grid through convolution, which is denoted as $R_X$.

*Step 2.* Send $\{R_X, R_z\}$ into residual convolution units (all units share weights), respectively, to obtain global feature $j_X$ and grid feature vector $j_z$.

*Step 3.* Feature vector $j_z$ is reconstructed through the network to obtain the reconstructed feature $J_z$, $z \in [0, 9]$.

*Step 4.* Use linear classifier to classify reconstructed feature $J_z$ and complete image feature $J_X = j_z$, and obtain 11 classification probability vectors $u$.

*Step 5.* Complete the classification based on the classification probability vector $u$ and calculate the cross-entropy loss.

*Step 6.* An adaptive momentum estimation optimizer is used to optimize network parameters.

*Step 7.* When reasoning, the images $X$ to be classified need not be segmented, and the classification and recognition of ancient Buddhist buildings can be completed directly through the sub-branch of the model.

## 4. Result Analysis and Discussion

In this study, a dataset from the Internet was used to verify the validity of the proposed method. The dataset, which shares images of ancient Buddhist architecture from Chinese provinces, aims to accelerate the understanding of ancient Buddhist architectural styles through large-scale data. It contains 541 temple image data from 22 provinces. Due to the fuzzy images of 12 temples, the remaining 529 ancient Buddhist buildings were used as experimental data after screening. An example of these data images is shown in Figure 5.

*4.1. Influence of the Number of Hidden Layers on Classification Performance.* In this section, in order to determine the impact of grid segmentation parameter $T$ on performance,

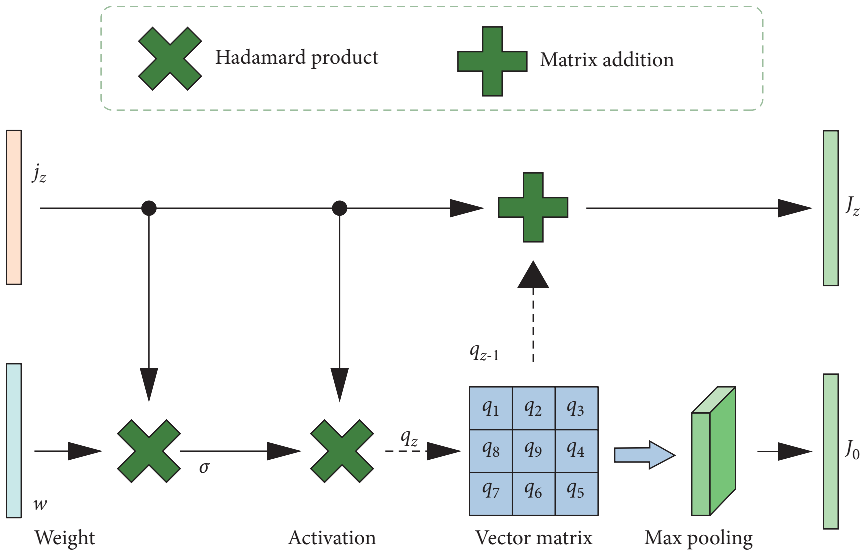
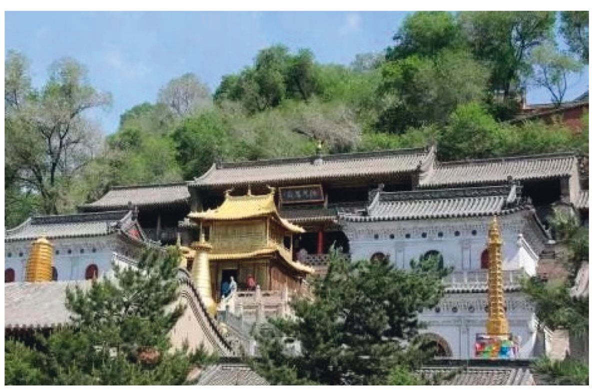



Figure 4: Network structure based on local feature learning.



(a)

(b)


(c)


(d)

Figure 5: Examples of ancient Buddhist architecture from Chinese provinces. (a) Xiantong Temple, Mont Wutai, Shanxi Province. (b) Wannian Temple, Mount Emei, Sichuan Province. (c) Huacheng Temple, Jiuhua Mountain, Anhui Province. (d) Lingyin Temple in Hangzhou, Zhejiang Province.

we constructed multiple networks with different grid segmentation parameters $T$. To facilitate comparison, the loss function training model proposed in this paper is used. The other hyperparameters are set the same except for the grid partitioning parameter $T$. Table 2 shows the experimental results under different partitioning parameters $T$.

As can be seen from Table 2, when $T = 1$, the classification accuracy is the lowest. The reason is that it cannot

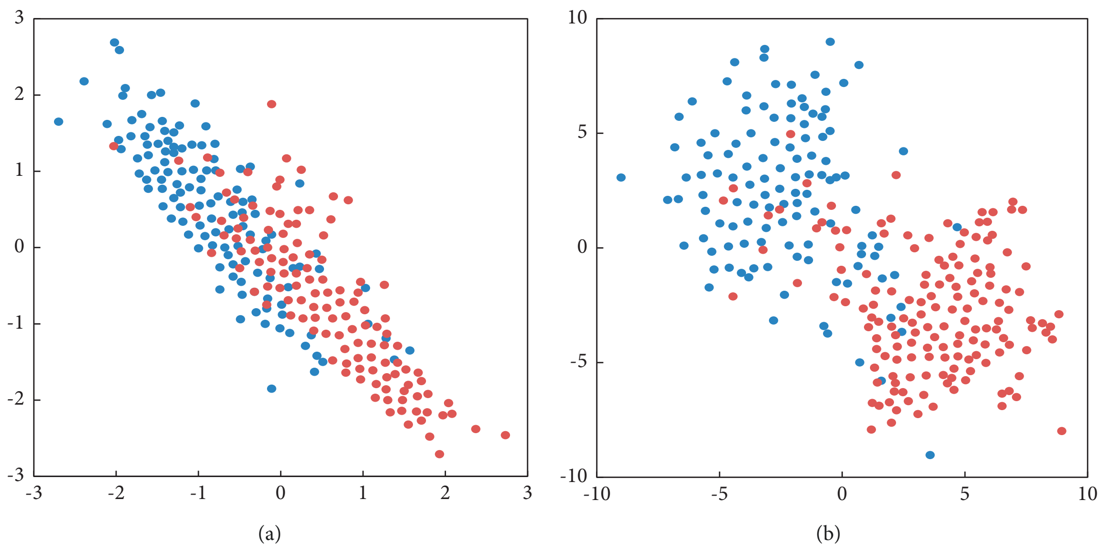



Figure 6: Comparison of feature distribution between the two methods. (a) Traditional algorithm. (b) Proposed algorithm.

adequately extract the deep features of functional connections in the brain. As the number of hidden layers increases, the classification accuracy gradually increases. When $T = 3$, the classification accuracy is the highest, reaching 82.41%. And in SPE, PPV, and NPV, three indicators have reached the highest value. However, as the number of hidden layers continues to increase, the performance of classification declines. The reason is that when the network structure is complex, the over-fitting phenomenon will occur, thus affecting the classification effect. Therefore, the grid segmentation parameter with $T = 3$ was selected as the basic structure in subsequent experiments.

*4.2. Influence of Prototype Learning on Feature Distribution.* For the effect of network learning on feature distribution, we also train a traditional algorithm without network learning based on the most commonly used softmax loss function. Then, the feature extracted from the last hidden layer in the algorithm is projected into the two-dimensional space for visualization, and the result is shown in Figure 6. Figure 6(a) shows the feature distribution extracted by the traditional algorithm. Figure 6(b) shows the distribution of features extracted by the method proposed in this paper. It can be seen from the figure that the feature extracted by the traditional algorithm has great differences within classes and small differences between classes. However, the feature differences extracted by the algorithm in this paper are significantly smaller within classes and larger between classes. Therefore, the algorithm in this paper can extract discriminative features.

*4.3. Convergence of Algorithm.* In order to intuitively demonstrate the convergence of the algorithm proposed in this paper, loss curves of unsupervised pretraining and supervised fine-tuning are drawn, as shown in Figure 7.

In the pretraining stage, it can be seen from the loss curve shown in Figure 7 that the overall convergence effect is good. When the verification set loss is the lowest, the network parameters are saved for subsequent training. In the fine-tuning stage, the loss of the training set showed a slight wobble, but the overall trend was downward. The loss of verification set and test set also decreases as the number of iterations increases. Finally, the network parameters at the lowest loss of verification set are saved for testing. In summary, the algorithm in this paper is convergent.

*4.4. Robustness Comparison Experiment.* In order to verify the robustness of the model, the dataset of this paper was randomly rotated between angles of [−25, −15, 0, 15, and 25] to obtain the rotated-rotation of the dataset. Then, the trained model is validated on the rotated test set. At the same time, this paper also conducts a robustness comparison experiment with CNN with the same number of layers as the model in this paper, and the comparison results are shown in Table 3 and Figure 8. It can be seen from Table 3 that the classification accuracy of CNN in the processing of rotating images is reduced by 4.9%, while the accuracy of the model proposed in this paper is only reduced by 0.51% in the rotating dataset. The experimental results not only prove that the proposed network is more robust to affine transformation images than CNN. At the same time, it verifies that the network proposed in this paper is a further improvement and improvement of traditional CNN in terms of robustness.

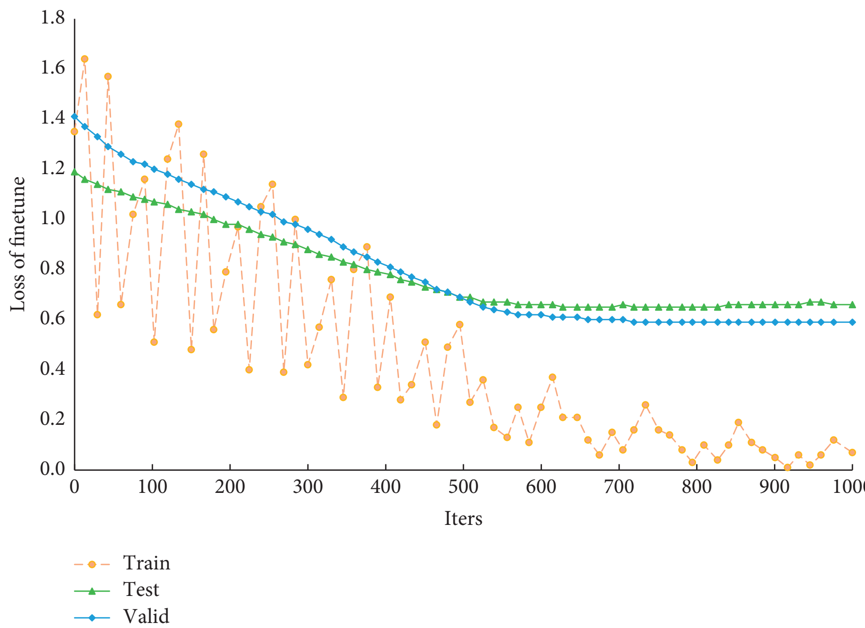



FIGURE 7: Convergence curve of the proposed algorithm.

TABLE 3: Robustness comparison results.

| Model | Dataset (%) | Dataset rotation (%) |
|---|---|---|
| CNN | 0.85 | 5.75 |
| Proposed | 0.33 | 0.84 |

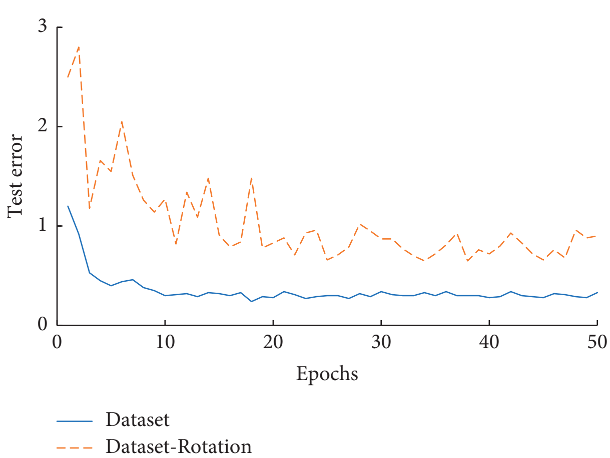



FIGURE 8: Robustness comparison curve.

*4.5. Comparative Experiments with Various Classical Algorithms.* In order to verify the advantages and disadvantages of the proposed method compared with other methods, we selected six classical algorithms from the traditional machine learning method and deep learning method, respectively, for comparative experiments. Methods based on traditional machine learning include literature [16] and literature [17]. Methods based on deep learning include literature [18], literature [19], literature [20], and literature [17]. The results are shown in Table 4 and Figure 9.

TABLE 4: Classification performance data of seven methods.

| Methods | ACC (%) | SEN (%) | SPE (%) | PPV (%) | NPY (%) |
|---|---|---|---|---|---|
| Literature [16] | 74.99 | 76.98 | 72.08 | 75.52 | 74.46 |
| Literature [17] | 76.18 | 81.61 | 70.22 | 76.71 | 75.26 |
| Literature [18] | 77.91 | 84.79 | 70.75 | 77.38 | 78.96 |
| Literature [19] | 78.44 | 81.75 | 75.12 | 78.30 | 78.96 |
| Literature [20] | 78.70 | 85.46 | 72 | 76.85 | 81.35 |
| Literature [17] | 79.36 | 84.79 | 73.80 | 78.17 | 80.55 |
| Proposed | 82.41 | 86.78 | 77.77 | 81.88 | 83.87 |

As can be seen from the above experimental results, the proposed method achieves the optimal performance in all indicators, and the classification effect is significantly better than other algorithms. Among the seven methods, the method based on deep learning is significantly better than the method based on traditional machine learning. ACC indexes of algorithms based on traditional machine learning are all lower than 77%. This is because the traditional machine learning model is a shallow model, unable to fully excavate valuable deep information. However, ACC indexes of deep learning-based methods are all above 77%, reflecting the advantages of deep learning methods in processing high-dimensional information. Compared with other four methods based on deep learning, this algorithm is significantly better than the other algorithms, with ACC index reaching 82.41%. The advantage of the proposed algorithm is also quite obvious in other indexes. The reasons are mainly in two aspects: (1) the network in this paper can extract the discriminant features with small differences within classes and large differences between classes, which improves the discriminant ability of the model. (2) The method proposed in this paper integrates low-level features

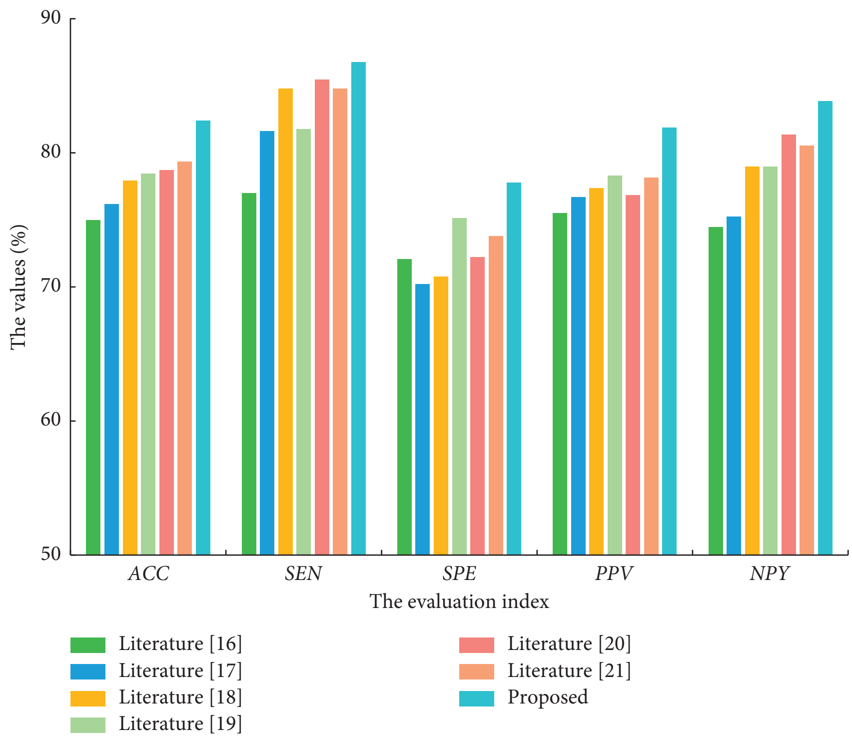



Figure 9: Result diagram of classification performance.

and high-level features extracted by depth model and can extract local information and global information, further improving the classification accuracy. The above results verify the effectiveness of the proposed method.

## 5. Conclusion

Ancient Buddhist architecture is the material carrier of Buddhist culture. The architectural style of ancient Buddhism reflects the influence of multi-culture on Buddhism to some extent. Buddhist architecture plays an important role in the development of Chinese architectural culture. At present, it is difficult to classify ancient Buddhist architecture automatically because of its complex structure and changing styles. In order to solve this problem, this paper uses the differential idea to segment the ancient Buddhist architectural images reasonably. After extracting features through residual network, reconstruction network, and other modules, the classification model is constructed. The experimental results show that the algorithm proposed in this paper can effectively improve the classification accuracy, and provide new ideas for the follow-up study of ancient Buddhist architecture. In the future work, we will focus on improving image gridding and grid reconstruction to obtain more and more accurate detail features, so as to more effectively improve the accuracy of ancient Buddhist architecture classification.

## Data Availability

The labeled dataset used to support the findings of this study is available from the corresponding author upon request.

## Conflicts of Interest

The author declares that there are no conflicts of interest.

## Acknowledgments

This study was supported by Sichuan University.

## References

[1] C. Zhai, "Cultural diversity in the world and socialist culture with Chinese characteristics-review of the second world cultural forum," *International Critical Thought*, vol. 8, no. 2, pp. 326–338, 2018.

[2] Q. Liu, Z. Liao, Y. Wu, and D. Y. Mulugeta Degefu, "Cultural sustainability and vitality of Chinese vernacular architecture: a pedigree for the spatial art of traditional villages in Jiangnan region," *Sustainability*, vol. 11, no. 24, p. 6898, 2019.

[3] V. H. Van, "Comparative buddhism in India, China, vietnam and the spirit of localization in Vietnamese buddhism[J]," *International Journal of Recent Scientific Research*, vol. 10, no. 6, pp. 1–7, 2019.

[4] N. Dreamson, "Culturally inclusive global citizenship education: metaphysical and non-western approaches," *Multicultural Education Review*, vol. 10, no. 2, pp. 75–93, 2018.

[5] R. A. Minhas, A. Javed, A. Irtaza, M. Mahmood, and Y. B. Joo, "Shot classification of field sports videos using AlexNet convolutional neural network," *Applied Sciences*, vol. 9, no. 3, p. 483, 2019.

[6] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into Imaging*, vol. 9, no. 4, pp. 611–629, 2018.

[7] B. Yildiz, A. Durdu, A. Kayabaşi, and M. Duramaz, "CNN based sensor fusion method for real-time autonomous

robotics systems," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 30, no. 1, pp. 79–93, 2022.

[8] C. Xiang, L. Zhang, Y. Tang, and W. C. Zou, "MS-CapsNet: a novel multi-scale capsule network," *IEEE Signal Processing Letters*, vol. 25, no. 12, pp. 1850–1854, 2018.

[9] P. Perera and V. M. Patel, "Learning deep features for one-class classification," *IEEE Transactions on Image Processing*, vol. 28, no. 11, pp. 5450–5463, 2019.

[10] Y. Liu, B. Yin, J. Yu, and Z. Wang, "Image classification based on convolutional neural networks with cross-level strategy," *Multimedia Tools and Applications*, vol. 76, no. 8, pp. 11065–11079, 2017.

[11] S. M. S. Abdullah and A. M. Abdulazeez, "Facial expression recognition based on deep learning convolution neural network: a review[J]," *Journal of Soft Computing and Data Mining*, vol. 2, no. 1, pp. 53–65, 2021.

[12] Y. Zhao, S. Guo, Y. Wang, J. Cui, and N. Xiao, "A CNN-based prototype method of unstructured surgical state perception and navigation for an endovascular surgery robot," *Medical, & Biological Engineering & Computing*, vol. 57, no. 9, pp. 1875–1887, 2019.

[13] K. Moran, C. Bernal-Cardenas, M. Curcio, and R. D. Bonett, "Machine learning-based prototyping of graphical user interfaces for mobile apps," *IEEE Transactions on Software Engineering*, vol. 46, no. 2, pp. 196–221, 2020.

[14] L. Wu, Y. Wang, and L. Shao, "Cycle-consistent deep generative hashing for cross-modal retrieval," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 1602–1612, 2019.

[15] J. Liu, C. Cao, X. Liu et al., "The evolution of the timber structure system of the buddhist buildings in the regions south of the yangtze river from 10th-14th century based on the main Hall of baoguo temple," *International Journal of Architectural Heritage*, vol. 13, no. 1, pp. 114–127, 2019.

[16] N. Prakash, A. Manconi, and S. Loew, "Mapping landslides on EO data: performance of deep learning models vs. Traditional machine learning models," *Remote Sensing*, vol. 12, no. 3, p. 346, 2020.

[17] H. Liu and B. Lang, "Machine learning and deep learning methods for intrusion detection systems: a survey," *Applied Sciences*, vol. 9, no. 20, p. 4396, 2019.

[18] C. Zhao, Q. Sun, C. Zhang, and Y. F. Tang, "Monocular depth estimation based on deep learning: an overview," *Science China Technological Sciences*, vol. 63, no. 9, pp. 1612–1627, 2020.

[19] Z. Cui, F. Xue, X. Cai, Y. Cao, G. G. Wang, and J. Chen, "Detection of malicious code variants based on deep learning," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 7, pp. 3187–3196, 2018.

[20] K. Wang, H. Wang, M. Liu, and X. T. Xing, "Survey on person re-identification based on deep learning," *CAAI Transactions on Intelligence Technology*, vol. 3, no. 4, pp. 219–227, 2018.