Heritage Science

RESEARCH Open Access

Deep learning based identification and interpretability research of traditional village heritage value elements: a case study in Hubei Province

Gangyi Tan[1,2,3,4], Jiangkun Zhu[1,2,3,4] and Zhanxiang Chen[1,2,3,4*]

Abstract

The preservation and transmission of traditional villages is crucial to the prosperity and development of ethnic cultures. However, current traditional village surveys usually require a large number of experts and scholars to conduct field research, which is an expensive and time-consuming method, especially for large-scale tasks. Therefore, this study proposes an automatic classification method based on deep learning (DL) for the identification of traditional village heritage value elements (TVHVE). The study evaluates four selected convolutional neural network (CNN) frames using traditional villages in Hubei Province as a sample dataset. The results show that Residual Network152 (ResNet152) is the most suitable CNN frame for identifying TVHVE in Hubei. The stability and consistency of various TVHVE present in the ResNet152 model were evaluated using Area Under Curve (AUC) and Precision Recall Curve (PRC), which indicated satisfactory prediction performance for most elements, except for specific elements such as tombstones and stone carvings, which showed lower accuracy. In addition, the study sheds light on the areas of concern of the model with respect to different TVHVE images and elucidates the reasons behind the confusion between elements through semantic clustering based on image classification and interpretability analysis using the Gradient-Weighted Class Activation Mapping (Grad-CAM) heat map. By using an automated classification method based on DL, this study significantly reduces the cost and effort associated with traditional surveys. At the same time, insight into areas of concern and confusion in the model improves guidance for conservation efforts and provides valuable references for subsequent research.

Keywords Convolutional neural network, Traditional village, Heritage value elements, Element identification, Interpretability analysis

Introduction

Traditional villages are important repositories of tangible and intangible cultural heritage, serving as the foundation of a nation's culture and playing a crucial role in promoting the prosperity and development of its cultural identity [1]. However, rapid urbanization has led to a significant decline in the rural population. Since 1980, more than 600 million people have migrated from rural to urban areas [2]. As a result, many rural dwellings have been abandoned, leading to a decline in the number of villages from 3.63 million to 2.71 million between 2000

*Correspondence:
Zhanxiang Chen
chenzx_0107@hust.edu.cn
[1] School of Architecture and Urban Planning, Huazhong University of Science and Technology, No. 1037 Luoyu Road, Wuhan 430074, China
[2] Hubei Engineering and Technology Research Center of Urbanization, Wuhan 430074, China
[3] Hubei Rural Construction Center, Wuhan 430074, China
[4] HUST Built Heritage Research Center, Wuhan 430074, China

and 2010 [3]. Moreover, this rural exodus has put many traditional villages at risk of demolition or destruction [4]. The gradual disappearance of the physical space that supports the unique elements of rural and historical culture poses a significant challenge to the preservation and transmission of vernacular culture.

In 2012, China introduced the concept of "Traditional Villages" and established the "List of Traditional Villages in China" to identify and recognize villages with significant cultural heritage value [5]. These villages include both tangible elements, such as architectural structures and landscapes, and intangible elements, such as folklore and cultural traditions. The Ministry of Housing and Urban–Rural Development of the People's Republic of China, in cooperation with experts in the fields of architecture, folklore, art, aesthetics, and economics, has developed the Traditional Village Evaluation and Recognition Index System (for Trial Implementation) to select traditional villages in China. This evaluation system focuses on three key aspects: traditional architecture, village location and layout, and the presence of intangible cultural heritage [6]. In the selection process, specific external manifestations of TVHVE, including the unique village environment, streets and alleys, architecture, intricate decorations and folk culture, play an important role.

Currently, research and work on traditional villages mainly revolves around empirical studies conducted through fieldwork, photography, and mapping [7–10]. However, the collection of TVHVE through field methods is often constrained by various factors such as transportation limitations, climatic conditions, and complex topography, resulting in significant human and material costs [11]. Moreover, the subsequent data fusion is a labor-intensive task, as the efficiency of human visual perception in identifying and classifying a large amount of image information is relatively low. Consequently, the involvement of domain experts becomes necessary to ensure accurate classification, increasing both time and labor costs [12].

In the field of architecture, machine learning (ML) algorithms have made rapid progress and are widely used for pattern recognition and solving complex engineering problems. Traditional ML methods such as Random Forest, Artificial Neural Networks, Bayesian Learning, Decision Tree, Support Vector Machine, and K-Nearest Neighbor algorithm have been used. However, these methods have certain limitations. They often require manual design and extraction of architectural features, leading to limitations in model accuracy. In addition, traditional ML algorithms require complex image preprocessing, making it difficult to achieve end-to-end automatic recognition of building features. In recent years, with the increasing complexity of the task, traditional ML methods have proven inadequate to effectively identify architectural features [13, 14].

DL methods, especially those based on image classification, have made significant progress in various domains, including image classification, target detection, and semantic segmentation [15–17]. In the field of architecture, DL methods have also produced notable research results [18, 19]. Therefore, the use of DL techniques to develop an innovative technical approach for collecting data on TVHVE for the purpose of identification and modeling has great potential for widespread application in rural areas. This approach effectively overcomes the limitations associated with traditional ML algorithms, providing improved accuracy and efficiency. As shown in Table 1, DL methods based on image classification in the construction field have been applied to various research areas, such as building structure damage assessment, building classification and extraction, building material classification, urban street and building group pattern classification, land use classification, building quality complaint text analysis, building energy prediction and classification, and others.

In summary, there is a limited amount of research on the application of DL in rural areas, with most studies focusing on urban settings. Furthermore, existing research often tends to classify rural architecture or specific individual elements [35, 36], while a comprehensive classification of the tangible and intangible characteristics of traditional villages remains relatively scarce. Therefore, there is a need to establish a model that can effectively and automatically identify and classify the TVHVE.

Research aim

This study aims to develop an automatic classification model using DL technology to identify the TVHVE in Hubei. The main contributions of this study are as follows: (1) establishing a classification framework for TVHVE in Hubei; (2) comparing the performance of four commonly used CNN models for detecting TVHVE in Hubei; (3) developing a training model specific to TVHVE in Hubei; (4) conducting interpretability analysis of the classification results for TVHVE in the test set. This study demonstrates the practical application of image recognition technology in architectural recognition and preservation, and provides theoretical and technical support for the preservation of traditional villages.

Materials and methods

Study area area

Hubei Province, located in central China at 108° 21′–116° 07′ east longitude and 29° 05′–33° 20′ north

Table 1 Research on DL methods based on image classification in the architecture field

Research category	Author	State-of-the-art review findings
Research on building structure damage	Pathirage, C.S.N. et al. [20]	This article presents a deep sparse autoencoder framework for effectively addressing complex pattern recognition problems, particularly those involving highly nonlinear relationships
	Alcantara, E. A. M. et al. [21]	The study introduces a building damage identification and structural response prediction method based on wavelet spectrogram data in CNN
Research on building classification and extraction	Zhenyu Lu et al. [22]	The study classifies three types of buildings, single-family houses, multi-family houses, and non-residential buildings, using spatial and landscape attributes from laser scanning remote sensing data
	Jianfeng Huang et al. [23]	The study presents a building extraction method based on a gated residual refinement network (GRRNet) using high-resolution aerial images and LiDAR data
	Joachim Höhle et al. [24]	The study investigates automatic building extraction and image enhancement methods based on DSM orthoimages
	Qintao Hu et al. [25]	The study proposes a novel DL model, DABE-Net, for automated building extraction
	E. J. Hoffmann et al. [26]	The study addresses the classification of commercial, residential, public, and industrial buildings using DL techniques and aerial and street view images
	Jian Kang et al. [27]	The study proposes a research framework based on CNN for functional classification of individual buildings
Research on building material classification	Andrei Kliuev et al. [28]	This paper examines the use of artificial neural networks and deep machine learning to predict physicomechanical properties of functional materials
Research on urban street and building group pattern classification	Chuan-Bo Hu et al. [29]	The study investigates the classification of Hong Kong urban street geometries using Google Street View images and multi-task DL
	Xiongfeng Yan et al. [30]	The study focuses on the classification of building group patterns at a macro level using a graph GCNN model framework
Research on land use classification	Victor Alhassan et al. [31]	This article introduces a deep learning-based method for mapping land-use and land-cover (LULC) from multispectral satellite imagery
Research on building quality complaint text classification	Botao Zhong et al. [32]	The study employs DL methods for the classification of building quality complaint text data
Research on building energy prediction and classification	Hansaem Park et al. [33]	This paper proposes deep migration learning integrated with stacked integration integration (SDTL) to predict optimal operational strategies for future energy distribution in buildings
	Kadir Amasyali et al. [34]	This paper presents a deep learning approach for predicting building energy consumption in an occupant behavior-sensitive manner

latitude, has a profound historical heritage. It served as the capital of Chu State during the Warring States period and established itself as the cultural center of Chu in Hubei. Throughout its history, from the Qin, Han and Six Dynasties to the Tang, Song, Ming and Qing Dynasties, Hubei's culture has preserved the essence of Chu culture while assimilating the characteristics of other regions. In particular, Hubei is the province through which the Yangtze River flows the longest, positioning it as a convergence point for China's four cardinal directions: east, west, south and north. In addition, Hubei's geographic location intersects with major ancient transportation routes. These include the Sino-Russian Ten Thousand Mile Tea Road, the Ancient Tea and Horse Trade Road, the Ancient Salt Road connecting Sichuan and Hubei, and the Huguang to Sichuan Immigrant Passage. These ancient routes crisscross Hubei, bringing a wealth of historical and cultural elements to its traditional villages.

Hubei is famous for its diverse and culturally rich traditional villages, which exhibit the unique socio-cultural characteristics of "blending northern and southern influences and incorporating elements from east and west".

These villages represent typical settlement development in the Yangtze River Basin. As of May 2023, a total of 270 traditional villages in Hubei have been recognized and included in the prestigious list of Chinese traditional villages [37]. Figure 1 shows that the terrain of Hubei is mainly characterized by mountainous regions in the east, west, and north, while the central area consists of low-lying areas and a partially open basin in the south. Traditional villages are mainly concentrated in the hilly and mountainous areas in the western and eastern parts.

Data collection and processing

In 2022, our research team, supported by the Department of Housing and Urban–Rural Development of Hubei, embarked on the comprehensive "Survey and Archiving of Traditional Villages". From June to July, we meticulously conducted extensive field research on traditional villages in various regions of Hubei. The survey covered more than 700 villages, 270 of which were designated as national-level traditional villages. The images covered in this article are from these 270 traditional villages (Fig. 1). Our research encompassed field photography, questionnaire distribution, and interviews with local villagers, utilizing a range of devices including drones, cameras, and smartphones. However, given that images captured by these devices may possess varying pixel sizes, we conducted preprocessing during data collection. This preprocessing involved data normalization and image resizing to ensure uniform size and feature representation across all images.

In this study, we focus on traditional villages in Hubei Province as our research subject, and we collected over 12,000 images from 270 villages through field surveys. During the selection process, we identified 3805 images that represent the characteristics of traditional villages across various regions of Hubei Province. These images serve as the foundation for constructing our dataset. This serves as a solid foundation for our study, facilitating an in-depth analysis of the characteristics, cultural heritage, and challenges facing traditional villages in Hubei. Through careful organization and analysis of these data, we aim to improve our understanding of the TVHVE. In addition, we plan to apply advanced DL techniques, such as image classification, to automatically identify and classify relevant features.

Establishing a classification framework for the TVHVE is important because it allows for the systematic identification of criteria and the accurate categorization of these elements. The rational design and application of classification rules not only contributes to an in-depth understanding of village characteristics and facilitates further analysis, but also serves as a foundation for subsequent research and conservation efforts. Figure 2 shows the classification and label settings of TVHVE. The TVHVE are classified into three categories: environmental elements, architectural elements, and cultural

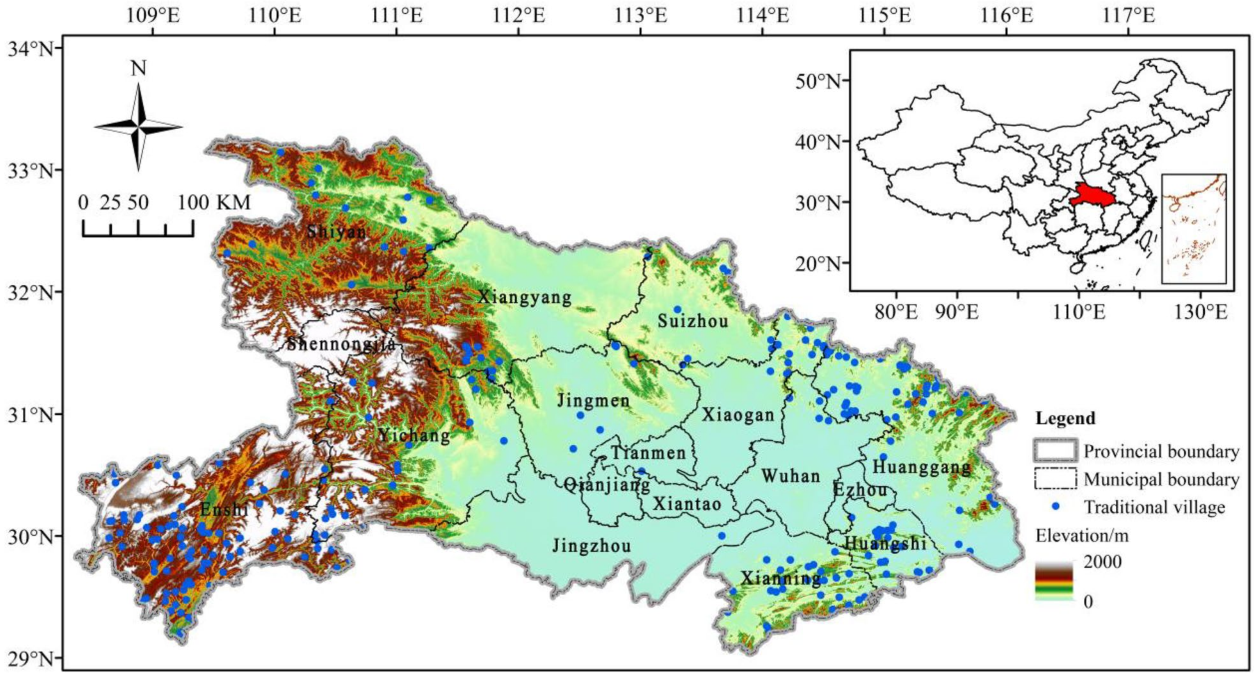


Fig. 1 Distribution of traditional villages in Hubei Province

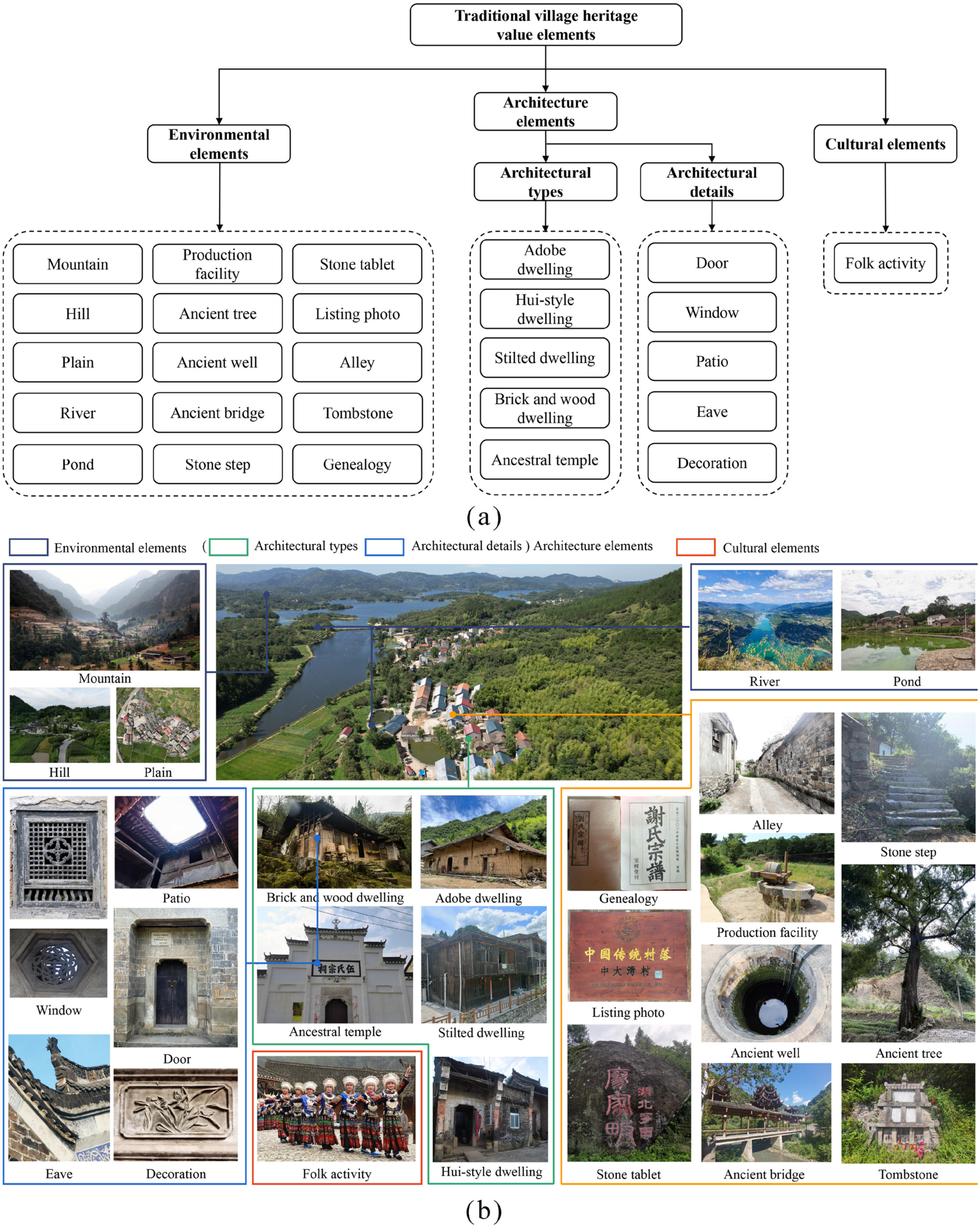


Fig. 2 **a** Framework of traditional village heritage value elements; **b** Label classification

elements. Meanwhile, the indicators for TVHVE include 26 detailed elements that meticulously characterize the essence of traditional villages. This scientifically rigorous categorization framework provides a robust tool and methodology for advancing research on TVHVE [38].

Compared to studies that rely on internet-sourced image data of traditional villages, the database utilized in this study is built upon a substantial collection of field-photographed images specific to Hubei. This aspect lends a higher degree of reliability and relevance to our research. To ensure the integrity of the image data, researchers meticulously screened and cleaned tens of thousands of traditional village images, adhering to the principles outlined for categorizing the elements of heritage value as described earlier. Furthermore, this screening method took into consideration the regional characteristics specific to traditional villages in Hubei. The image dataset was categorized and extracted based on these characteristics, effectively reducing the duplication rate of similar images and minimizing computational time required for modeling.

During dataset construction, we meticulously selected training and test sets at a 4:1 ratio, ensuring label consistency [22, 32]. We also considered the quantity of data for each sample type, meeting model computational requirements (Table 2). To enhance recognition accuracy, we took the following steps:

1. Diverse image types: on-site photography captured heritage elements from various angles for comprehensive coverage.
2. Varied image backgrounds: the dataset included images with diverse lighting and weather conditions, improving adaptability.
3. Diverse target scenes: within the same classification, targets of varying sizes were included, enhancing scene recognition.
4. Data enhancements: we use two data enhancement techniques during model training: random rotation and cropping. These techniques can increase the diversity of training data and improve the generalisation ability of the model.

Table 2 Model database of heritage value elements

Parent classes	Child classes	Heritage value elements	Train-set	Test-set	Total
Environmental elements	–	Mountain	96	32	128
		Hill	110	36	146
		Plain	126	42	168
		River	75	25	100
		Pond	102	34	136
		Production facility	75	25	100
		Ancient tree	202	67	269
		Ancient well	90	30	120
		Ancient bridge	99	32	131
		Stone step	94	31	125
		Stone tablet	102	33	135
		Listing photo	90	29	119
		Alley	102	34	136
		Tombstone	87	28	115
		Genealogy	83	27	110
Architecture elements	Architectural types	Adobe dwelling	127	42	169
		Hui-style dwelling	108	35	143
		Stilted dwelling	155	51	206
		Brick and wood dwelling	186	62	248
		Ancestral temple	75	25	100
	Architectural details	Door	188	62	250
		Window	83	27	110
		Patio	107	35	142
		Eave	120	40	160
		Decoration	88	29	117
Cultural elements	–	Folk activity	92	30	122

The overall workflow of this study, as shown in Fig. 3, consists of four main steps: data collection and classification, data processing, model comparison and selection, and analysis of identification results including data interpretation.

In the data collection phase, extensive research and photography of traditional villages was conducted to establish a comprehensive sample database. A detailed classification of the TVHVE was carried out during this phase. Next, in the data pre-processing step, the collected data were manually screened and organized according to the 26 categories of TVHVE. This process involved careful data selection and preparation to construct the corresponding data set. In the model comparison and selection step, the training effects of four CNN models (ResNet18, Visual Geometry Group Network19 (VGG19), ResNet152, and Dense Convolutional Network121 (DenseNet121)) were examined. The purpose of this step was to identify the most appropriate model for TVHVE detection. Finally, in the data and interpretability analysis step, the recognition results of the test

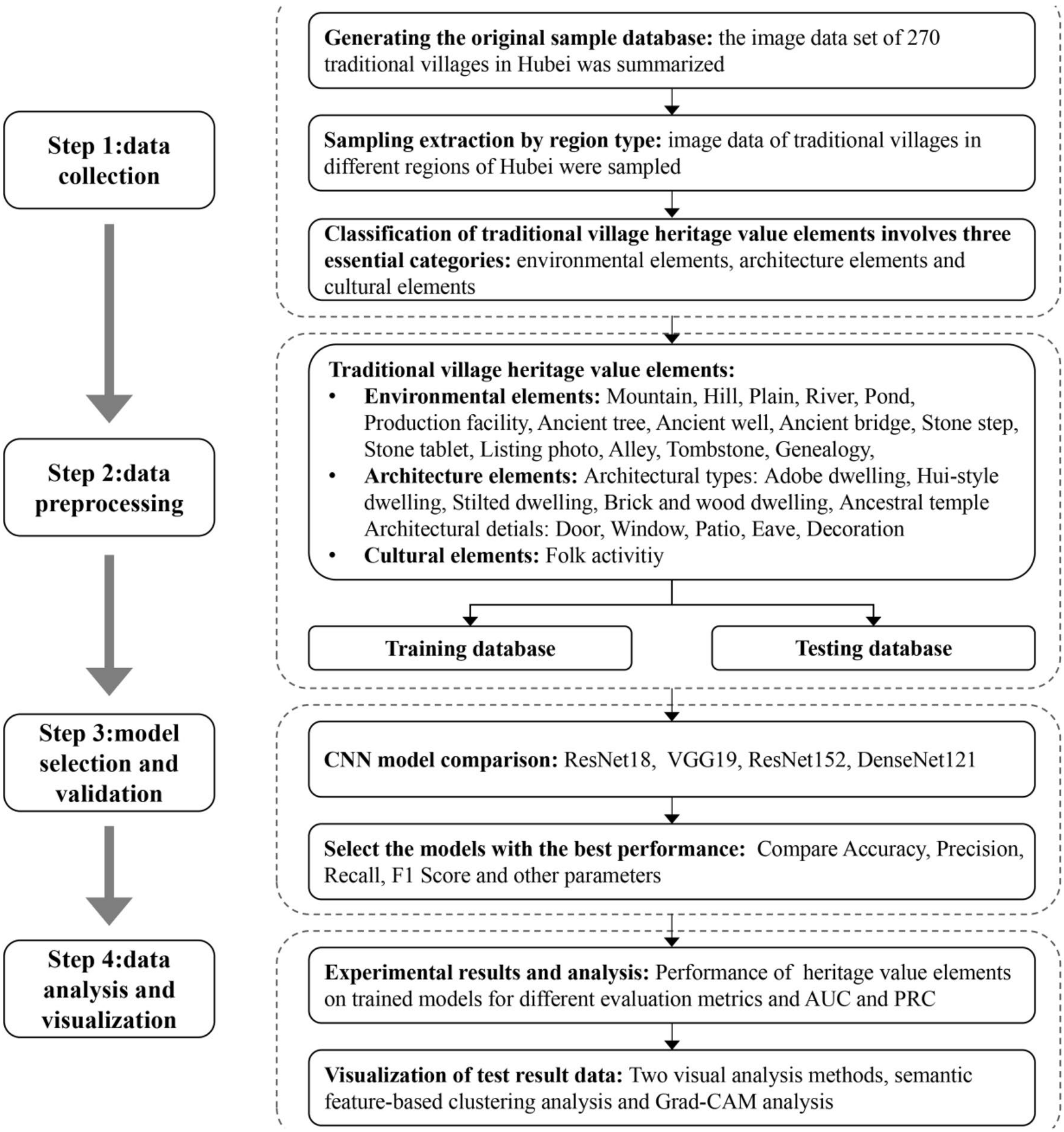

Step 1:data collection
Step 2:data preprocessing
Step 3:model selection and validation
Step 4:data analysis and visualization
Generating the original sample database: the image data set of 270 traditional villages in Hubei was summarized
Sampling extraction by region type: image data of traditional villages in different regions of Hubei were sampled
Classification of traditional village heritage value elements involves three essential categories: environmental elements, architecture elements and cultural elements
Traditional village heritage value elements:
• Environmental elements: Mountain, Hill, Plain, River, Pond, Production facility, Ancient tree, Ancient well, Ancient bridge, Stone step, Stone tablet, Listing photo, Alley, Tombstone, Genealogy,
• Architecture elements: Architectural types: Adobe dwelling, Hui-style dwelling, Stilted dwelling, Brick and wood dwelling, Ancestral temple Architectural detials: Door, Window, Patio, Eave, Decoration
• Cultural elements: Folk activitiy
Training database
Testing database
CNN model comparison: ResNet18, VGG19, ResNet152, DenseNet121
Select the models with the best performance: Compare Accuracy, Precision, Recall, F1 Score and other parameters
Experimental results and analysis: Performance of heritage value elements on trained models for different evaluation metrics and AUC and PRC
Visualization of test result data: Two visual analysis methods, semantic feature-based clustering analysis and Grad-CAM analysis

Fig. 3 Research flow chart

set data on the trained models were evaluated. In addition, interpretability analysis techniques such as semantic clustering and Grad-CAM heat map were used to gain insights and interpret the results in a meaningful way.

Model selecting

The CNN is a widely used DL framework specifically designed for image classification tasks. It consists of a feature extraction layer, which performs convolutional computations, and multiple hidden layers. The CNN is capable of automatically extracting low-level features from the original input and integrating them into high-level features that serve as the basis for target recognition. This network framework exhibits powerful recognition performance. In the field of image classification, several classical CNN models have gained significant popularity. These models include ResNet, VGG, DenseNet, and others. These models have demonstrated their effectiveness in various image recognition tasks.

The VGG model utilizes 3×3 convolution kernels and successive 3×3 convolutions to maintain a consistent receptive field while increasing network depth, improving feature capture efficiency. In contrast, ResNet introduces skip connections and residual learning to tackle deep neural network optimization challenges, alleviating the vanishing gradient problem. DenseNet enhances model performance and reduces parameters through feature reuse and dense connections, maximizing information flow between layers, representing significant advancements in deep learning techniques.

In this study, we have selected four CNN image recognition models commonly used in the architectural field, ResNet152, VGG19, ResNet18, and DenseNet121. They represent the classical and commonly used model architectures in deep learning. resNet152 and VGG19 are relatively deep networks with a large number of layers and parameters, whereas ResNet18 and DenseNet121 are relatively shallow with fewer parameters. By comparing these models of different depth and complexity, their trade-off between performance and resource consumption can be evaluated.

Model training

All model training and testing procedures in this study were performed on a cloud computing platform provided by FEATURIZE [39]. The rented computer used in this study had an Intel Xeon Gold Xeon Gold 6142 CPU model and a GeForce RTX 3080 GPU model. The available video memory of the GPU was 10.5 GB. TensorFlow and PyTorch, which are popular DL programming frameworks, were used for the experiments.

To ensure a fair comparison between different models, the hyperparameters for model training were standardized in the experiments. The Adam optimization algorithm was used as the gradient optimization algorithm for training all models, with a learning rate of 0.001. The number of training iterations was set to 100, and the loss function chosen was Cross Entropy.

Evaluation criterion

Accuracy is a commonly used metric to evaluate the correctness of a model. However, when dealing with unbalanced data sets, accuracy alone may not be an appropriate metric to evaluate the results. Therefore, in this study, we used four evaluation metrics: Accuracy, Precision, Recall, and F1 Score. Accuracy represents the proportion of correctly predicted positive samples out of all samples. Precision measures the proportion of correctly predicted positive samples out of all samples predicted to be positive. Recall quantifies the proportion of correctly predicted positive samples out of all actual positive samples. The F1 score combines precision and recall, seeking a balance between the two to achieve the optimal trade-off.

By calculating the average of these evaluation metrics, we can select the best performing model based on its overall performance. This approach provides a comprehensive evaluation of model effectiveness and takes into account the impact of sample imbalance.

$$Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{1}$$

$$Precision = \frac{T_P}{T_P + F_P} \tag{2}$$

$$Recall = \frac{T_P}{T_P + F_N} \tag{3}$$

$$F1score = 2 \times \frac{precision \times recall}{precision + recall} \tag{4}$$

where $\mathrm{T_P}$ (True Positive) represents positive samples are rated as positive by the model, $\mathrm{T_N}$ (True Negative) represents negative samples are rated as negative samples, $\mathrm{F_P}$ (False Positive) represents negative samples are rated as positive samples and $\mathrm{F_N}$ (False Negative) represents positive samples are rated as negative samples.

AUC and PRC are key metrics used to assess the classification performance of a model. AUC describes the relationship between true positive and false positive rates, while PRC demonstrates the trade-off between precision and recall. Both provide a comprehensive assessment of model performance and are particularly suitable for unbalanced datasets. The AUC, which ranges from 0 to 1, with 0.5 representing random guessing, 0 indicating poor performance, and 1 denoting excellent performance, AUC offers a comprehensive evaluation of a model's overall proficiency, independent of threshold choices. PRC also ranges from 0 to 1, with higher values indicating better model performance. Unlike AUC, PRC focuses more on positive classes and is particularly useful for evaluating unbalanced datasets.

Interpretability analysis based on image classification

Image classification interpretability analysis is a valuable approach utilizing visualization techniques to gain deeper insights into model classification outcomes. It aids in understanding the relationships between categories, pinpointing misclassifications, and exploring image features. In this study, we employed two common methods, semantic feature visualization and Grad-CAM heat maps, to demystify the inner workings of our deep learning model, shedding light on the "black box" and offering intuitive insights into the TVHVE classification and recognition task's similarities and distinctions among various elements.

Dimensionality reduction visualization of semantic features in image classification involves reducing high-dimensional image feature vectors to a lower-dimensional space (e.g., 2D or 3D) and visualizing them. This method aims to provide visual insights into the clustering, distribution, and distinctions among image data categories. It encompasses four key steps: feature extraction using a trained CNN model to obtain high-dimensional feature vectors, applying a dimensionality reduction algorithm (e.g., t-SNE) to condense these vectors while preserving essential information, using visualization techniques like scatter plots or 3D graphs to display feature distributions, and analyzing the results to interpret classification outcomes, feature representations, category relationships, and potential outliers.

In addition, Grad-CAM heat map is an interpretable method used to interpret the prediction results of CNN in image classification tasks. It generates a heat map that visualizes the attention paid by the model to different regions of the image during the classification process. Higher heat values are typically associated with regions that have a strong influence on the predicted categories. Therefore, heat maps help to understand how the model makes classification decisions and serve as a visual and explanatory tool for the model's prediction process. It is important to note that Grad-CAM is an interpretability technique that explains the prediction results of a trained CNN model for a specific image. It does not modify or tune the model itself, but provides an interpretation of the model's predictions.

Results and discussions

Model performance comparison

In order to select a suitable CNN model to detect the TVHVE in Hubei, we conducted a comparative experiment with four prominent CNN models: ResNet152, VGG19, ResNet18, and DenseNet121. In the experiments, we ensured consistency by using the same training and test datasets for all models. We also maintained consistency in the optimization algorithm, learning rate, number of training iterations, and cross-entropy parameters across all models. The experiments were conducted in the same hardware and software environments, ensuring fairness and reliability in the comparison.

Figure 4 show the variation of the evaluation metrics of the test set for the four models during the iteration process. These curves show that the evaluation metrics and cross-entropy loss trends of the four models follow similar patterns, indicating that all models have achieved convergence without overfitting or underfitting problems. During the first 30 iterations, the models undergo the learning phase, with all evaluation metrics showing an upward trend, while the cross-entropy loss gradually decreases. Among the four models, ResNet152, VGG19, ResNet18, and DenseNet121 reach their optimal performance at the 49th, 31st, 34th, and 63rd iterations, respectively. Thereafter, the evaluation metrics and cross-entropy loss curves maintain a stable fluctuation within a certain range, indicating that the models have converged on the network parameters that yield the best classification results.

By comparing the training curves of each model, it is evident that the ResNet152 model consistently outperforms the VGG19, ResNet18, and DenseNet121 models on the test set. In addition, the cross-entropy loss curve of the ResNet152 model remains consistently lower than the other three models. Taken together, these observations lead to the conclusion that the ResNet152 model is the most suitable choice for the task of classifying and detecting the TVHVE in Hubei. It demonstrates superior classification performance and lower loss function values, allowing for more accurate classification of images related to TVHVE.

Table 3 presents the evaluation metrics and their averages on the test dataset for the best performing model obtained after training each model. The test results indicate that ResNet152 achieves the highest performance in five metrics: accuracy (0.851), precision (0.852), recall (0.838), F1 score (0.842), and Loss (0.491). As a result, ResNet152 is identified as the most effective DL CNN model for the task of identifying the TVHVE in Hubei.

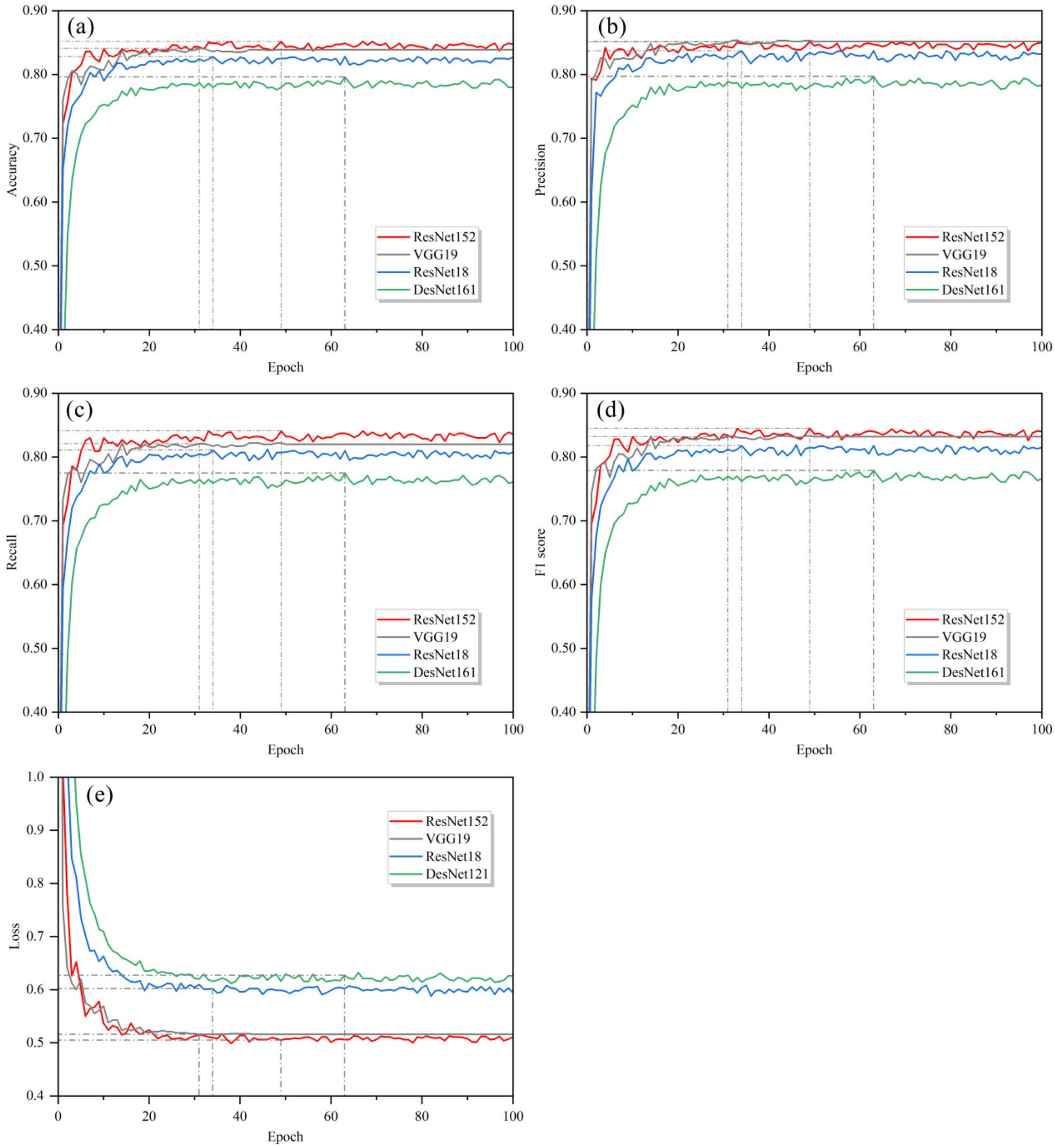


Fig. 4 **a** Accuracy; **b** Precision; **c** Recall; **d** F1 Score; **e** Loss

Based on ResNet152 model training results

Result of recognition

Table 4 presents the specific values of various evaluation metrics, including Accuracy, Precision, Recall, and F1 Scores, for the trained ResNet152 model in recognizing the TVHVE within the test set. In order to provide a more intuitive understanding of the effectiveness of the various categories of heritage value elements identified in the model, Table 4 has been added to highlight the effectiveness of each category using the average effectiveness ratio as an example. The Table clearly shows that the model can effectively recognize most of the TVHVEs in the test set. In particular, nine heritage value elements, such as ancient bridges, ancient trees, and hanging

Table 3 Identification validity parameters of the four models

Models	Loss	Accuracy	Precision	Recall	F1-score	Mean
ResNet18	0.602	0.828	0.837	0.811	0.818	0.824
VGG19	0.516	0.841	0.851	0.821	0.832	0.836
ResNet152	**0.505**	**0.852**	**0.852**	**0.841**	**0.845**	**0.84**
DenseNet121	0.627	0.796	0.797	0.775	0.779	0.829

Bolded values in Table 3 indicate that this model outperforms the other three models on all metrics, and is key evidence for the selection of this model

Table 4 Precision of identification of heritage value elements

Heritage value elements	Accuracy	Precision	Recall	F1-score	Average effectiveness
Alley	0.813	0.929	0.813	0.867	0.856
Ancestral temple	0.850	0.810	0.850	0.829	0.835
Ancient bridges	0.875	1.000	0.875	0.933	0.921
Ancient trees	0.962	0.944	0.962	0.953	0.956
Ancient well	0.833	0.909	0.833	0.870	0.833
Brick and wood dwelling	0.837	0.788	0.837	0.812	0.810
Adobe dwelling	1.000	0.804	1.000	0.891	0.937
Decoration	0.737	0.824	0.737	0.778	0.769
Door	0.960	0.873	0.960	0.914	0.940
Eaves	0.875	1.000	0.875	0.933	0.897
Folk activity	0.958	0.920	0.958	0.939	0.944
Genealogy	0.955	0.955	0.955	0.955	0.955
Hills	0.621	0.750	0.621	0.679	0.668
Hui-style dwelling	0.714	0.714	0.714	0.714	0.714
Listing photo	0.870	0.833	0.870	0.851	0.870
Mountains	0.842	0.640	0.842	0.727	0.725
Patio	0.947	0.947	0.947	0.947	0.947
Plains	0.824	0.778	0.824	0.800	0.806
Pond	0.963	0.897	0.963	0.929	0.974
Production facility	0.850	0.810	0.850	0.829	0.850
River	0.950	0.950	0.950	0.950	0.933
Stilted dwelling	0.697	0.852	0.697	0.767	0.732
Stone step	0.737	0.875	0.737	0.800	0.750
Stone tablets	0.667	0.750	0.667	0.706	0.748
Tombstone	0.600	0.818	0.600	0.692	0.704
Window	0.864	0.905	0.864	0.884	0.879

footstools, have an average effectiveness ratio greater than 90%, indicating high recognition effectiveness for these elements. However, seven heritage value elements, including tombstones, stone carvings, and mountains, have an average effectiveness ratio of less than 80%. This discrepancy can be attributed to the presence of more synonyms and confusion in the semantic categorization of these particular heritage value elements.

The model's ability to accurately recognize and classify key heritage elements is crucial for the documentation and preservation of cultural heritage sites. For elements with high recognition rates, this technology can facilitate detailed archival records and aid in maintenance planning by providing reliable identification at scale.

AUC and PRC analysis

In this study, the AUC is used to evaluate the stability and consistency of the trained model in recognizing elements of traditional village heritage values. Figure 5a shows that the AUC values for all categories of the TVHVE

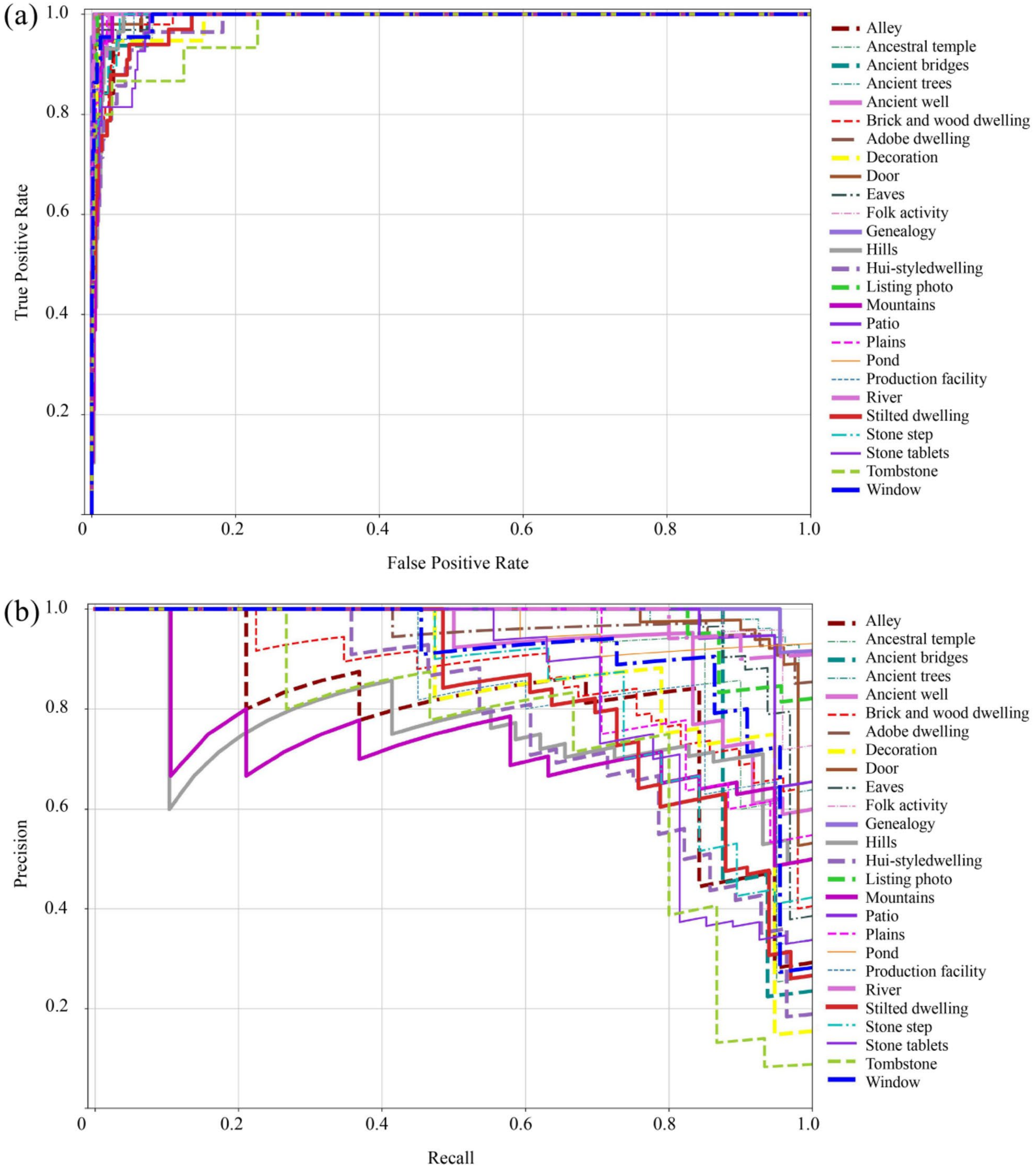


Fig. 5 **a** AUC plot of heritage value elements; **b** PRC plot of heritage value elements

are above the random guess level, indicating a relatively accurate recognition rate for each element type in the model. However, certain heritage value elements, such as gravestones and stone carvings, show slightly lower recognition performance in the AUC.

Additionally, PRC is employed in this study to compare the recognition performance of TVHVE in the trained model. Figure 5b shows that the PRC for all categories of TVHVE are above the random guess level, indicating a relatively accurate recognition rate for each element category in the model. Compared to the AUC, the PRC is better at illustrating the differences between different categories of TVHVE and can more clearly depict the variations in recognition between elements. It should be noted that heritage value elements such as tombstones and stone carvings, which show poor performance in the AUC, show a similar trend in the PRC. In addition, mountains and hills in topographic environments show lower recognition performance in the PRC.

It is worth noting that the TVHVE that performed poorly in the AUC and PRC, such as tombstones, stone carvings, mountains and hills, were not among the least numerous types of elements in the dataset. Therefore, it can be further illustrated that the use of AUC and PRC to evaluate the recognition results of TVHVE is able to cope with the problems posed by unbalanced datasets. In practical applications, various representation methods can enrich the dataset to identify underperforming elements. Advanced analytical techniques like deep feature extraction or ensemble learning are then applied to improve precision and recall.

Interpretability analysis

Semantic cluster analysis

In Sect. "Data collection and processing", we present the results of evaluation metrics to assess the recognition performance of different TVHVEs in the trained model. However, these metrics alone do not provide direct explanations or insights into the underlying reasons for the performance. To address this limitation, this study employs visualization techniques to delve into the inner workings of the DL network. Through visualization, we aim to understand and analyze the differences and confounds between different TVHVE, as well as identify the critical aspects that the model focuses on during the detection process. This approach improves the accuracy of element categorization and identification, and provides guidance for further improvements in model performance. By revealing the "black box" of the DL networks, we gain valuable insights into the model's operation, leading to a better understanding of its performance in recognizing TVHVE.

Figure 6 shows the results of semantic clustering analysis for image recognition of TVHVE based on the ResNet152 training model:

Among the environmental elements, the semantics of mountains, hills, and plains can be divided into similar categories. However, the model shows confusion between hilly terrain and mountains/plains, leading to lower accuracy in recognizing hills as a heritage value element. This confusion explains the relatively lower performance of the model in recognizing hilly terrain compared to other types of terrain. For the others, rivers and ponds are recognized with high accuracy and there is essentially no semantic confusion between them. However, there is some semantic confusion between rivers and mountains due to their close association, as rivers often flow through hilly terrain. This semantic confusion may explain the model's difficulty in accurately recognizing rivers and mountains.

Environmental elements like production facilities, ancient bridges, stone steps, and genealogy exhibit high recognition accuracy and minimal semantic confusion. Ancient trees, while having relatively compact semantic clustering, show some confusion with traditional village images. Semantic confusion also arises between ancient wells and ponds, as well as between production facilities and stone steps, likely due to shared materials or elements. In contrast, stone carvings and tombstones, with lower recognition accuracy, display dispersed semantic clustering, indicating higher confusion with other environmental elements. This is primarily due to their similarity, requiring more sophisticated categorization methods for distinction. Streets, on the other hand, exhibit greater decentralization but tend to be confused with architectural elements in terms of semantics.

Within the category of architectural elements, the heritage value elements of the five architectural types show high recognition accuracy in the model and similar semantic categorization in the semantic clustering. In particular, there is a significant degree of semantic confusion between the stilted dwellings and the brick dwellings. This confusion may be due to the similarities in architectural features or materials used in these two types of dwellings. In addition, there is a significant semantic similarity between Hui-style dwellings and ancestral temples, which is consistent with the common sense in the field of architecture that the architectural elements of southern ancestral temples have many similarities to Hui-style dwellings.

Within the category of architectural detail elements, the semantic clustering of elements such as doors, windows, terraces, eaves, and decorations shows a higher degree of decentralization. The level of confusion between these elements is minimal, indicating that their semantics are relatively independent. However, there is some semantic confusion between certain elements, such as eaves, and ancestral temples, which can be attributed to the inclusion of certain architectural elements in the semantics of ancestral temples.

Within the category of cultural elements, there is currently only one heritage value element, so its semantic categorization is highly independent and does not overlap with other elements.

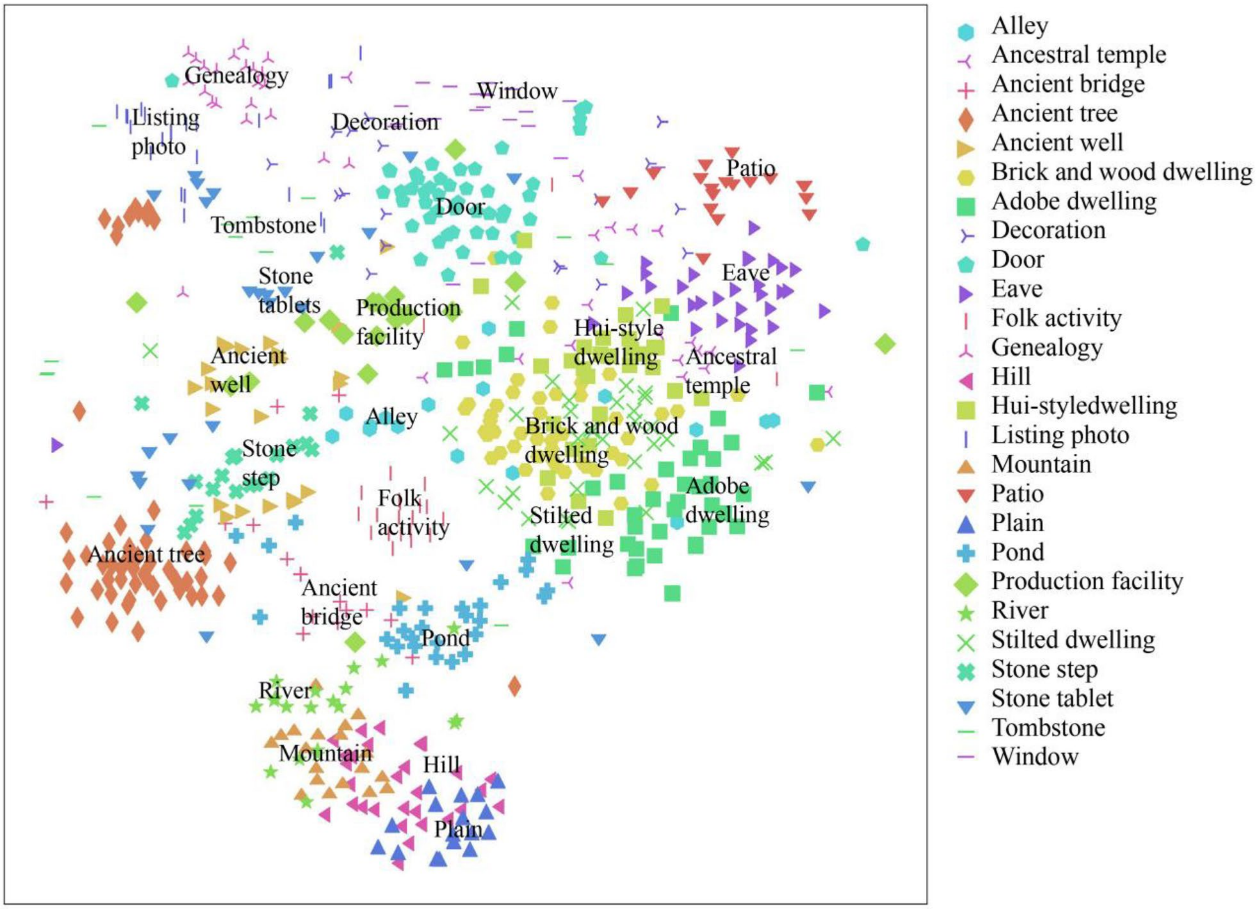


Fig. 6 Semantic clustering map of heritage value elements

Grad-CAM analysis

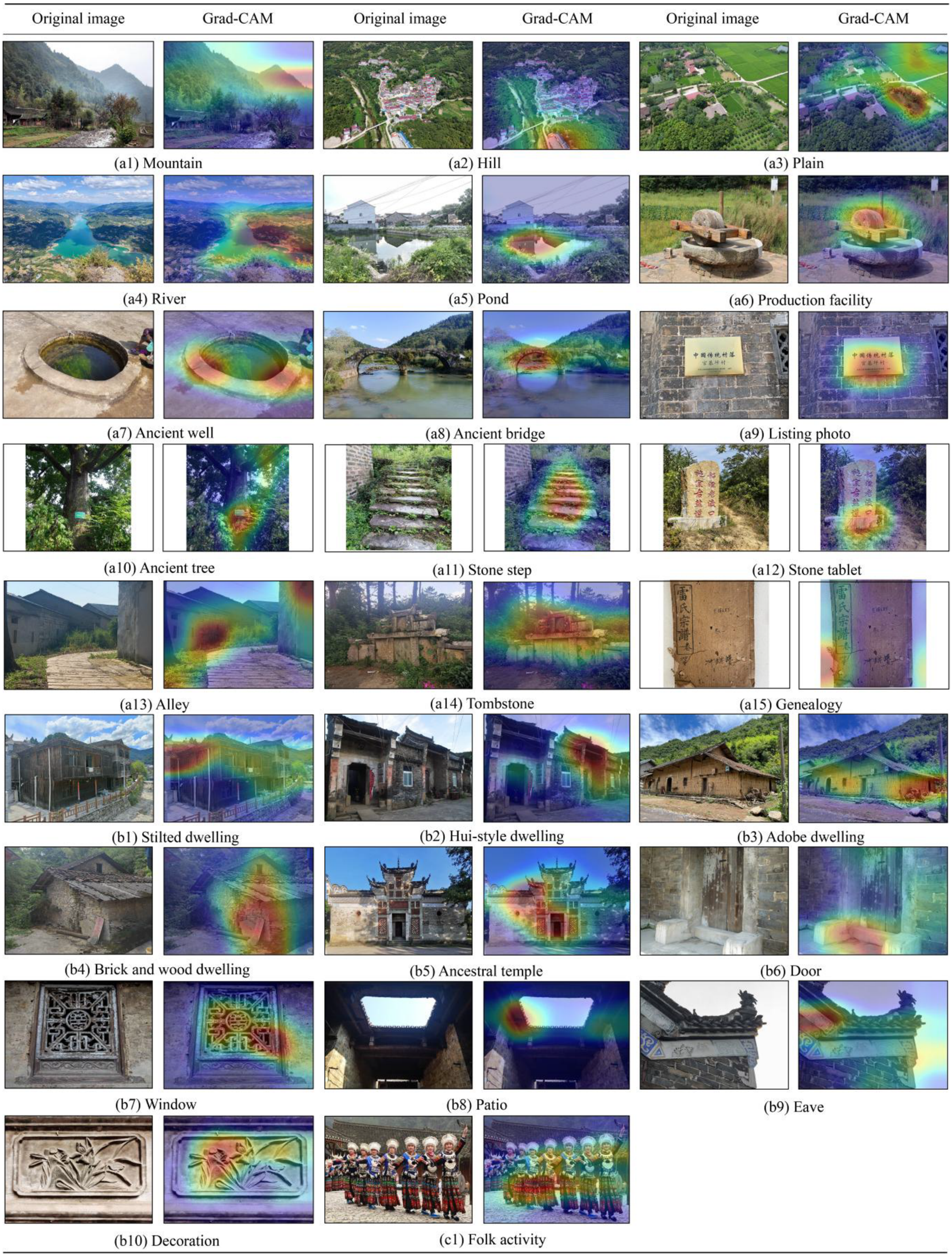
Figure 7 shows the original images of 26 TVHVE along with their corresponding class activation heat maps. The visualization results highlight the activated regions for each element that serve as key features in the images of the TVHVE. The darker regions in the heat map indicate higher attention from the model, helping researchers make accurate classification decisions. For example, in the case of stone carvings and tombstones, which show a high degree of confusion, the heat map focuses on areas with similar elements, resulting in potential confusion and requiring more detailed classification decisions. Similarly, the heat map for the river element focuses not only on the river itself, but also on the adjacent mountains, leading to some confusion between the river and mountain elements.

For instance, the heat maps for stone carvings and tombstones, which are frequently confused by the model, reveal a focus on overlapping features that contribute to this confusion. This observation suggests a need for more distinct feature extraction in training the model to better differentiate between these similar categories. Similarly, the heat map for rivers shows significant attention not only to the water bodies but also to the adjacent mountainous regions, which could be misleading the model into confusing these two separate elements.

The detailed visualisations provided by Grad-CAM allow researchers to identify the features prioritised by the model. This can guide further feature engineering and data preprocessing to emphasise or de-emphasise certain features to reduce confusion. Examples include distinguishing between stone carvings and tombstones or rivers and mountains. Adding images that provide clearer, more characteristic features of the elements at the time of dataset acquisition will help the model learn to recognise and differentiate between these elements more effectively.

Conclusions

In this study, we propose a DL-based approach to recognize heritage value elements in traditional villages in Hubei Province. We compared the performance of



(a1) Mountain (a2) Hill (a3) Plain

(a4) River (a5) Pond (a6) Production facility

(a7) Ancient well (a8) Ancient bridge (a9) Listing photo

(a10) Ancient tree (a11) Stone step (a12) Stone tablet

(a13) Alley (a14) Tombstone (a15) Genealogy

(b1) Stilted dwelling (b2) Hui-style dwelling (b3) Adobe dwelling

(b4) Brick and wood dwelling (b5) Ancestral temple (b6) Door

(b7) Window (b8) Patio (b9) Eave

(b10) Decoration (c1) Folk activity

Fig. 7 Grad-CAM heat map of each traditional village heritage value element. **a1** Mountain; **a2** Hill; **a3** Plain; **a4** River; **a5** Pond; **a6** Production facility; **a7** Ancient well; **a8** Ancient bridge; **a9** Listing photo; **a10** Ancient tree; **a11** Stone step; **a12** Stone tablet; **a13** Alley; **a14** Tombstone; **a15** Genealogy; **b1** Stilted dwelling; **b2** Hui-style dwelling, **b3** Adobe dwelling; **b4** Brick and wood dwelling; **b5** Ancestral temple; **b6** Door; **b7** Window; **b8** Patio; **b9** Eave; **b10** Decoration; **c1** Folk activity

four CNN models (ResNet152, VGG19, ResNet18, and DenseNet121) using evaluation metrics such as accuracy, precision, recall, F1 score, and loss rate. The results show that the ResNet152 model achieves the highest performance on the test set with an Accuracy of 0.851, Precision of 0.852, Recall of 0.838, F1 Score of 0.842, and Loss rate of 0.491. The ResNet152 model effectively recognizes most of the TVHVE, with average recognition rates exceeding 90% for elements such as ancient bridges, ancient trees, and stilted dwellings. However, elements such as gravestones and stone tablets have lower average recognition rates (below 80%), which may be due to semantic categorization challenges. This suggests the need for more accurate classification methods and model enhancements for these specific elements.

This study conducted a comprehensive assessment of TVHVE using AUC and PRC, and the results demonstrate the overall good performance of the model. The AUC values for all element categories exceeded random guesses, indicating high recognition accuracy of the model. However, certain elements such as tombstones and stone tablets exhibited slightly lower recognition performance. Similarly, the PRC values for all element categories surpassed random guesses, indicating elevated identification accuracy of the model. Compared to the AUC, the PRC provided clearer differentiation between the element categories, revealing distinct recognition patterns. Notably, elements like tombstones and stone tablets displayed subpar performance in both AUC and PRC. Additionally, elements associated with mountainous and hilly terrain showed lower identification performance. These results highlight the significance of considering both AUC and PRC for a comprehensive assessment of model performance in the context of traditional village.

Semantic clustering analysis revealed variations and confusions among different heritage value elements. Environmental elements such as production facilities, ancient bridges, stone steps, and genealogy showed higher accuracy and less semantic confusion in their classification. Conversely, certain elements showed poorer results. For example, confusion between hilly terrain, mountains, and plains led to lower recognition accuracy for hilly elements. Some semantic confusion was observed between river and mountain elements. Stone tablets and tombstones had low recognition accuracy in the model and showed more confusion with other elements in the living environment category. In addition, there was increased semantic confusion between road elements and building elements.

Grad-CAM's heat map analysis reveals the regions of interest in the model for different images of heritage elements. These heat maps highlight the key features that influence the model's classification decisions. In the case of stone tablets and tombstones, the heat maps show similar regions of interest, leading to potential confusion. Similarly, the heat map for the river element focuses on both the river and the adjacent mountains, which can contribute to confusion between river and mountain elements.

The interpretability analysis provides valuable insights into the performance and decision-making process of the DL model in recognizing TVHVE. These findings provide guidance for improving the model's classification and recognition capabilities, and shed light on the relationships and semantic distinctions between different heritage value elements. Ultimately, this knowledge contributes to the preservation and transmission of the cultural heritage of traditional villages.

Acknowledgements
Not applicable.

Author contributions
GT: Reviewing and Editing. JZ: Investigation, Experiment, Methodology, Data analysis, Writing-Original draft preparation. ZC: Conceptualization, Writing.

Funding
This research was supported by a grant from the National Natural Science Foundation of China (No. 52278018).

Availability of data and materials
Materials used for the study are available upon reasonable request.

Declarations

Competing interests
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Received: 5 February 2024 Accepted: 8 June 2024
Published online: 18 June 2024

References

1. Hu Y, Chen S, Cao W, Cao C-Z. The concept and cultural connotation of traditional villages. Urban Dev Stud. 2014;21:10–3.
2. Doe J. Xinhua China's Rural Population Percentage Cut 50% over 70 Years. 2019. http://english.www.gov.cn/archive/statistics/201909/03/content_WS5d6e6561c6d0c6695ff7fbab.html. Accessed 3 Sep 2019.
3. Liu Y-S, Liu Y, Chen Y-F, Long H-L. The process and driving forces of rural hollowing in China under rapid urbanization. J Geogr Sci. 2010;20:876–88. https://doi.org/10.1007/s11442-010-0817-2.
4. Gao J-L. Study on the strategy of controlling and guiding to the protection and development planning of traditional villages in Chongqing, Chongqing University, 2017.
5. Cao Y-C, Zhang Y-K. Appraisal and selection of "Chinese traditional village" and study on the village distribution. Arch J. 2013;12:44–9.
6. Jiancun DJ. No. 125 on the issuance of the "Traditional Village Evaluation and Identification Index System (Trial)" notice. 2012. https://www.mohurd.gov.cn/gongkai/zhengce/zhengcefilelib/201208/20120831_211267.html. Accessed 23 Sept 2023.

7. Altassan AI. Sustainability of heritage villages through eco-tourism investment (case study: Al-Khabra Village, Saudi Arabia). Sustain. 2023;15(9):7172. https://doi.org/10.3390/su15097172.
8. Wang C, Zhong H, Su W-B. Gene recognition and genealogy construction of settlement cultural landscape: a case study of Dong traditional village in northern Guangxi. Soc Sci. 2020;2:50–5.
9. Li B-H, Li Z, Liu P-L, Dou Y-D. Landscape gene variation and differentiation law of traditional villages in Xiangjiang River Basin. 2022;37:362-377.
10. Volovyk V, Lavryk O, Yatsentyuk Y, Maksiytov A. Polish ethnocultural landscape of Podillya: structure, use, protection of cultural heritage. Geol Geogr Ecol. 2022;57:68–80. https://doi.org/10.26565/2410-7360-2022-57-06.
11. Liu P-L, Zeng C, Liu R-R. Environmental adaptation of traditional Chinese settlement patterns and its landscape gene mapping. Habitat Int. 2023;135: 102808. https://doi.org/10.1016/j.habitatint.2023.102808.
12. Yin L, Wang Z-X. Measuring visual enclosure for street walkability: using machine learning algorithms and Google Street View imagery. Appl Geogr. 2016;76:147–53. https://doi.org/10.1016/j.apgeog.2016.09.024.
13. Kamath CN, Bukhari SS, Dengel A. Comparative study between traditional machine learning and deep learning approaches for text classification. In Proceedings of the ACM Symposium on Document Engineering. Association for Computing Machinery, New York, NY, USA, 2018;1–11. https://doi.org/10.1145/3209280.3209526.
14. Dong B, Wang X. Comparison deep learning method to traditional methods using for network intrusion detection, Proc. 2016 8th IEEE Int. Conf. Commun. Softw. Networks, ICCSN. 2016;581–585. https://doi.org/10.1109/ICCSN.2016.7586590.
15. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–44. https://doi.org/10.1038/nature14539.
16. Mathew A, Amudha P, Sivakumari S. Deep learning techniques: an overview. Adv Mach Learn Technol Appl. 2021. https://doi.org/10.1007/978-981-15-3383-9_54.
17. Shin HC, Roth HR, Gao M-C, Lu L, Xu Z-Y, Nogues I, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging. 2016;35(5):1285–98. https://doi.org/10.1109/TMI.2016.2528162.
18. Damodaran BB, Höhle J, Lefèvre S. Attribute profiles on derived features for urban land cover classification. Photogramm Eng Remote Sensing. 2017;83(3):183–93. https://doi.org/10.14358/PERS.83.3.183.
19. Taylor ME, Stone P. Transfer learning for reinforcement learning domains: a survey. J Mach Learn Res. 2009;10:1633–85.
20. Pathirage CSN, Li J, Li L, Hao H, Liu W-Q, Wang R-H. Development and application of a deep learning–based sparse autoencoder framework for structural damage identification. Struct Health Monit. 2019;18(1):103–22. https://doi.org/10.1177/1475921718800363.
21. Alcantara EAM, Bong MD, Saito T. Structural response prediction for damage identification using wavelet spectra in convolutional neural network. Sensors. 2021;21(20):6795. https://doi.org/10.3390/s21206795.
22. Lu Z-Y, Im J, Rhee J, Hodgson M. Building type classification using spatial and landscape attributes derived from LiDAR remote sensing data. Landsc Urban Plan. 2014;130:134–48. https://doi.org/10.1016/j.landurbplan.2014.07.005.
23. Huang J-F, Zhang X-C, Xin Q-C, Sun Y, Zhang P-C. Automatic building extraction from high-resolution aerial images and LiDAR data using gated residual refinement network. ISPRS J Photogramm Remote Sens. 2019;151:91–105. https://doi.org/10.1016/j.isprsjprs.2019.02.019.
24. Höhle J. Automated mapping of buildings through classification of DSM-based ortho-images and cartographic enhancement. Int J Appl Earth Obs Geoinf. 2021;95: 102237. https://doi.org/10.1016/j.jag.2020.102237.
25. Hu Q-T, Zhen L-L, Mao Y, Zhou X, Zhou G-Z. Automated building extraction using satellite remote sensing imagery. Autom Constr. 2021;123: 103509. https://doi.org/10.1016/j.autcon.2020.103509.
26. Hoffmann EJ, Wang Y-Y, Werner M, Kang J, Zhu XX. Model fusion for building type classification from aerial and street view images. Remote Sens. 2019;11(11):1259. https://doi.org/10.3390/rs11111259.
27. Kang J, Körner M, Wang Y-Y, Taubenböck H, Zhu XX. Building instance classification using street view images. ISPRS J Photogramm Remote Sens. 2018;145:44–59. https://doi.org/10.1016/j.isprsjprs.2018.02.006.
28. Kliuev A, Klestov R, Bartolomey M, Rogozhnikov A. Recommendation system for material scientists based on deep learn neural network. In: Antipova T, Rocha A, editors. Digital science. Cham: Springer International Publishing; 2019. p. 216–23. https://doi.org/10.1007/978-3-030-02351-5_26.
29. Hu C-B, Zhang F, Gong F-Y, Ratti C, Li X. Classification and mapping of urban canyon geometry using Google Street View images and deep multitask learning. Build Environ. 2020;167: 106424. https://doi.org/10.1016/j.buildenv.2019.106424.
30. Yan X-F, Ai T-H, Yang M, Yin H-M. A graph convolutional neural network for classification of building patterns using spatial vector data. ISPRS J Photogramm Remote Sens. 2019;150:259–73. https://doi.org/10.1016/j.isprsjprs.2019.02.010.
31. Alhassan V, Henry C, Ramanna S, Storie C. A deep learning framework for land-use/land-cover mapping and analysis using multispectral satellite imagery. Neural Comput Appl. 2020;32:8529–44. https://doi.org/10.1007/s00521-019-04349-9.
32. Zhong B-T, Xing X-J, Love P, Wang X, Luo H-B. Advanced engineering informatics convolutional neural network: deep learning-based classification of building quality problems. Adv Eng Informatics. 2019;40:46–57. https://doi.org/10.1016/j.aei.2019.02.009.
33. Park H, Park DY, Noh B, Chang S. Stacking deep transfer learning for short-term cross building energy prediction with different seasonality and occupant schedule. Build Environ. 2022;218: 109060. https://doi.org/10.1016/j.buildenv.2022.109060.
34. Amasyali K, El-Gohary N. Machine learning for occupant-behavior-sensitive cooling energy consumption prediction in office buildings. Renew Sustain Energy Rev. 2021;142: 110714. https://doi.org/10.1016/j.rser.2021.110714.
35. Meng C-M, Song Y-S, Ji J-Q, Jia Z-Y, Zhou Z-X, Gao P, et al. Automatic classification of rural building characteristics using deep learning methods on oblique photography. Build Simul. 2022. https://doi.org/10.1007/s12273-021-0872-x.
36. Wang X, Jin X-L, Feng Y-T. Landscape reconstruction of traditional village couplets based on image recognition algorithm. J Opt. 2023;52(1):224–32. https://doi.org/10.1007/s12596-022-00843-x.
37. Doe J. The prestigious list of Chinese traditional villages. http://www.mohurd.gov.cn. Accessed 30 Jan 2024.
38. Tan G-Y, Yi L-W. Critical thinking on the protection of traditional villages and rural construction based on the concept of built heritage. New Arch. 2023;2:4–10.
39. Doe J. Machine learning online laboratory. https://featurize.cn/me/billing. 2023. Accessed 30 Jan 2024.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.