Research Article

An approach for identifying historic village using deep learning

Jin Tao[1] · Geng Li[2] · Qiwei Sun[3] · Youjia Chen[2] · Dawei Xiao[1] · Huicheng Feng[2]

Received: 20 July 2021 / Accepted: 30 November 2022
Published online: 12 December 2022

Abstract

This paper aims to propose an approach to automatically identify historic villages from remote sensing images based on deep learning algorithm and accurately calculate the villages' geographical coordinates. Experimental datasets of Conghua, a typical region in fast development that retains many historic villages, are designated for training and testing. Comparison experiments of two recognition models, image classification and object detection, are designed to obtain the most suitable identification algorithm. GIS platform is adopted to visualize the distribution of the historic villages. The results show that first, the recognition accuracy of the image classification algorithm is 90.79%. However, visualization of test results shows the identified area is not a village but a surrounding. Second, the recognition accuracy of an object detection algorithm can reach 95.61%, which indicates that the algorithm is accurate and efficient. Third, by using the Historical-Modern tag as a filter, a village with a certain proportion of historic features according to specific requirements may be discriminated. Finally, 1531 historic villages in Conghua area were identified by the preferred algorithm, and their spatial locations were marked. This research will extend the detection of remote sensing image targets of deep learning algorithms from single buildings to group patterns and complex ground objects, so as to promote the integration of heritage conservation and artificial intelligence research. This time-efficiency approach can provide strong support for the discovery and field investigation of historic villages facing fast development and provide a scientific basis for the formulation of conservation policies.

Article Highlights

- Deep learning is applied to the protection of the cultural heritage of historic villages.
- Comparative experiments of different algorithms are designed to analyse their applicability in historic village recognition. A recognition rate of up to 95.61% is achieved.
- The visualization of recognition results is important for understanding the relationship between historic villages and nature, and historic village conservation.

Keywords Historic village · Deep learning · Image identification · Heritage conservation

✉ Huicheng Feng, fneocc@foxmail.com | [1]State Key Lab of Subtropical Building Science, Department of Architecture, South China University of Technology, 381 Wushan Road, Guangzhou 510641, China. [2]Department of Architecture, South China University of Technology, 381 Wushan Road, Guangzhou 510641, China. [3]School of Computer Science and Engineering, South China University of Technology, 381 Wushan Road, Guangzhou 510641, China.

SN Applied Sciences (2023) 5:13 | https://doi.org/10.1007/s42452-022-05246-y

1 Introduction

Historic villages, as a type of cultural heritage, have important historical, artistic and scientific values, and their protection is a universal concern around the world [1–3]. China maintains a large number of historic villages. Due to the influence of fast urbanization, these villages are disappearing rapidly. Thus far, China has announced 5 batches of 6799 historic villages. Meanwhile, the actual number of villages with certain historical and cultural value is estimated to be approximately 300,000, which means that only 2.3% of the villages are on the conservation list. Usually, the identification of historic villages relies on field work, which results in the lack of clues, low coverage, heavy field investigation workloads, and low time-effectiveness. Therefore, how to identify and locate these historic villages in time and correspondingly to prevent them from being destroyed in the process of rapid development is a task of great practical significance.

In addition, with the development of artificial intelligence technology, deep learning algorithms have made great progress and are becoming a trendy topic in the fields of image recognition, speech recognition, natural language processing, etc. [4–6]. In the area of intelligent image processing, the convolutional neural network (CNN) model in deep learning algorithms has achieved remarkable results in image classification, object detection, pose estimation, image segmentation and face recognition tasks. However, in the identification of built-up environments, especially in the extraction of historic villages and buildings, its applications are still in the preliminary stage.

This research attempts to combine automatic image recognition with historic village morphology research. Taking Conghua area in Zhujiang delta as an example where a large number of historic villages are preserved, a deep learning algorithm is adopted to automatically identify historic villages in high-resolution remote sensing imagery based on their morphological structures. In addition, accurate coordinates of the historic villages are extracted to obtain their geographical locations and spatial distribution. This time-efficiency approach can provide strong support for the discovery and field investigation of historic villages facing fast urbanization and provide a scientific basis for the formulation of relevant protection policies.

The paper is organized as follows. In the next section, we review papers related to general building recognition methods and raise the issue of using deep learning methods for the identification of historic villages as cultural heritage. In Sect. 3, we present the materials and deep learning algorithms used in this study. Section 4 shows the results of the different algorithms. In Sect. 5, discussion on the results and finally conclusion are given.

2 Literature review

The differences in the physical morphologies between historic villages and modern settlements are the basis for their image identification. As time has passed, the needs, technologies, materials and other aspects of the construction of villages and buildings have undergone tremendous changes, which have led to differences between the "new" and "old" in form of the villages [7, 8]. Different spatial forms will be reflected in different imaging features in remote sensing pictures. By capturing these features, a deep learning algorithm can automatically identify the target objects in a large number of samples.

The study of the target interpretation of settlements and buildings in remote sensing images is flourishing [9–12]. According to objects' appearances as planar features with certain areas, lengths and widths, the conventional approaches use region growing algorithms, such as the least square B-spline curve based method, the edge tracking-based method and edge detection [13–15]. These methods are mainly based on pixels, and the accuracies do not meet the requirements of the applications. There are also object-oriented detection methods, which mainly use the spectrum, texture, shape and background information of image objects [16, 17]. Meanwhile, object detection studies that build classifiers using machine learning are also reported [18]. However, due to the influence of natural conditions and human factors, the feature expressions that are used for the object detection methods are often designed by hand, and it is difficult to fully express the true features of the object [19].

In recent years, deep learning algorithms have provided an effective framework for automatic feature extraction and have made great progress in image and graphics recognition [20, 21]. Building recognition, as an application of deep learning in the field of Architecture [22, 23], is also gradually emerging. From the current research, higher success rates have been achieved for single building detection [24–26], and there are also acceptable discrimination rates of the geometric measurements of elements of building surfaces [27, 28].

However, different from single buildings, historic villages are formed by groups of ancient structures, and there are also various elements in the village, such as streets, squares, vegetation, water, etc. The complicated and diverse spatial forms of villages result in uncertainty in the identifications. Therefore, based on the physical morphology of historic villages, this research will extend the detection objects from single buildings to group patterns. By capturing the overall structure features of the settlement, further exploration will be made on the application of deep learning algorithm in the recognition of complex ground objects, so as to promote the integration of heritage conservation and artificial intelligence research.

3 Material and method

3.1 Material

Conghua, well known for its large amount of preserved historic villages in Zhujiang delta in south China, is designated as a research area to carry out image recognition of historic villages. A field survey shows that there are ancient buildings that are preserved in most of the villages in Conghua area (Fig. 1), and high heritage values are demonstrated by their beautifully decorated ancestral halls and dwellings.

In this research, 0.2-m resolution orthophoto remote sensing images covering the whole territory of Conghua area, which geographically spans from 113° 17′ to 114° 04′ east longitude to 23° 22′–23° 56′ north latitude over 2374.4 km^2, were used (Fig. 2). The images were taken at an aerial altitude of 4000 m during clear daytime. The original images consist of 2968 standard images, each of which has an actual east–west length of 1000 m, a north–south length of 800 m, an area of 80 hectares, a width of 5001 pixels and a height of 4001 pixels (Fig. 3a) C.

Considering the appropriate scale of the target object on a single image and the efficiency of computer processing, the standard single images were split. Every standard image was cut into 3 * 3 sections, resulting in a total of 26,712 images, each of which was reduced to 333.3 m by 266.7 m (Fig. 3b). In addition, the images were compressed to 280 × 224 pixels, which greatly increase the proportion of historic villages in a single image and meet the requirements of computer hardware device performance and recognition algorithms.

3.2 Method

In the field of deep learning, the convolutional neural network (CNN) is the structural basis of image recognition. The VGGNet of the CNN structure has achieved good performance in large-scale image recognition tasks [29]. Therefore, a convolutional neural network model is applied for the identification of historic villages.

The aim of this work is to identify whether a given remote sensing image contains historic villages, which essentially is a binary classification task. To provide a suitable method for historic village identification, we designed a comparative test of two types of algorithms. One method uses an image classification network to divide the input image into two categories with historic villages (historical) and without historic villages (non-historical), and the pictures containing historic villages are what need to be identified. Another method uses the object detection network to calculate the historic villages' confidence and locations in the image. If the confidence level is higher than the threshold, it is recognized as the historical class, and the coordinates of the historic village are obtained. Otherwise, the images would be recognized as belonging to the non-historical class.

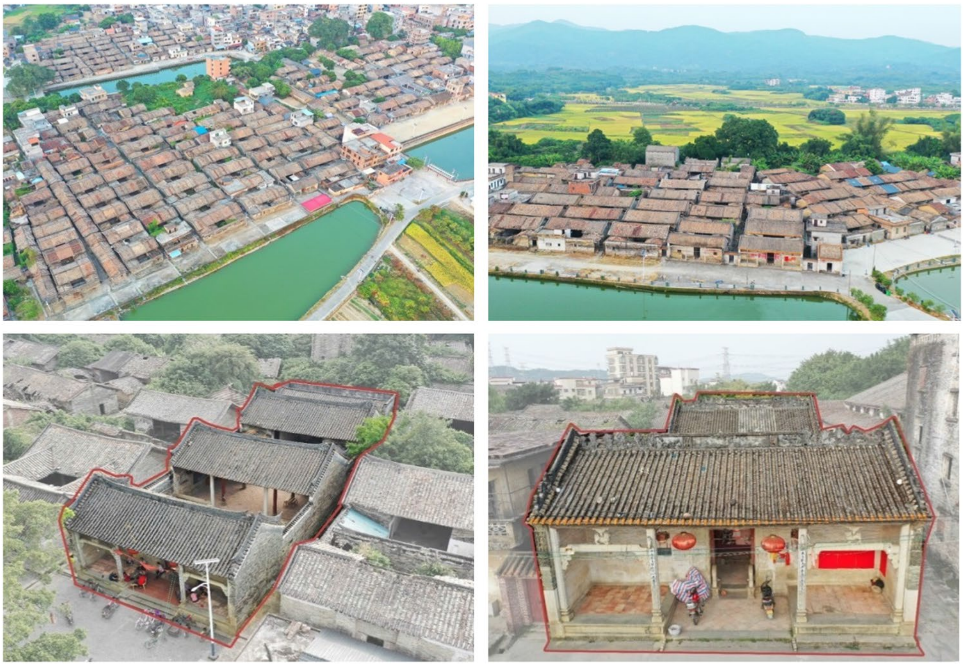


Fig.1 Traditional villages and ancient buildings in Conghua area

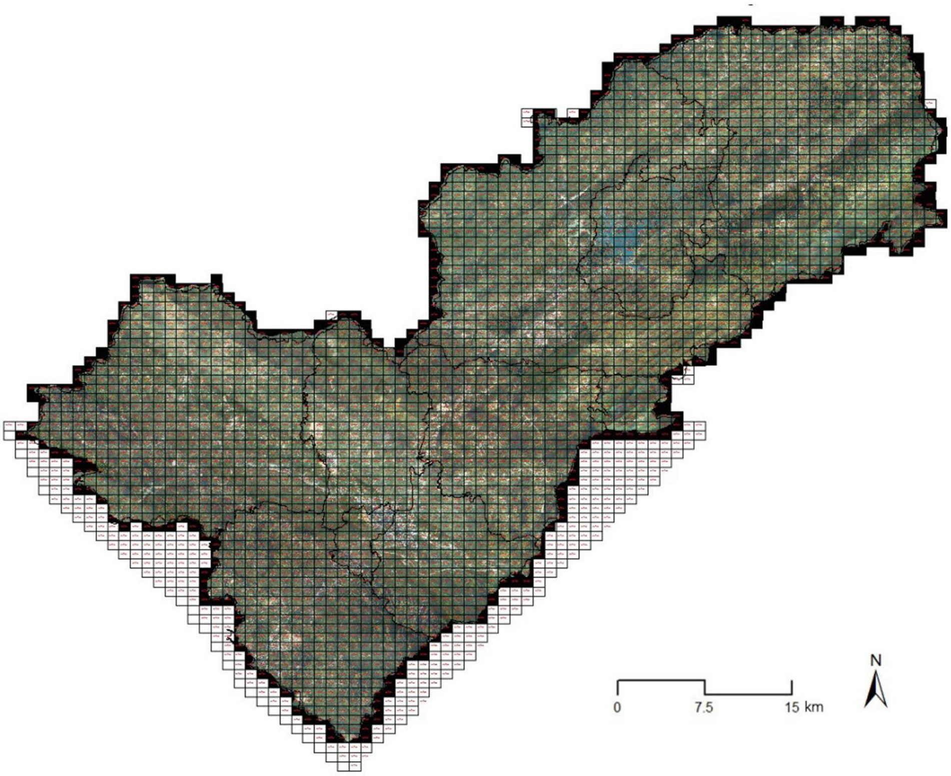
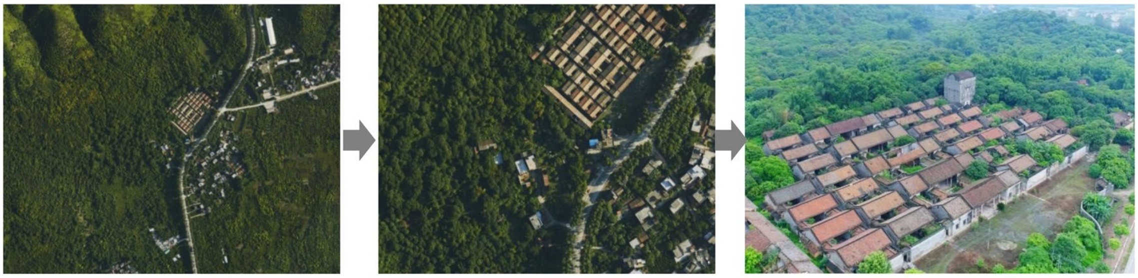


Fig. 2 Study area image and its framing



(a) Example single image

(b) Example of the image after cutting

(c) Aerial photo of the village

Fig. 3 Example of splitting single image

3.2.1 Image classification algorithm

(1) Classification task dataset

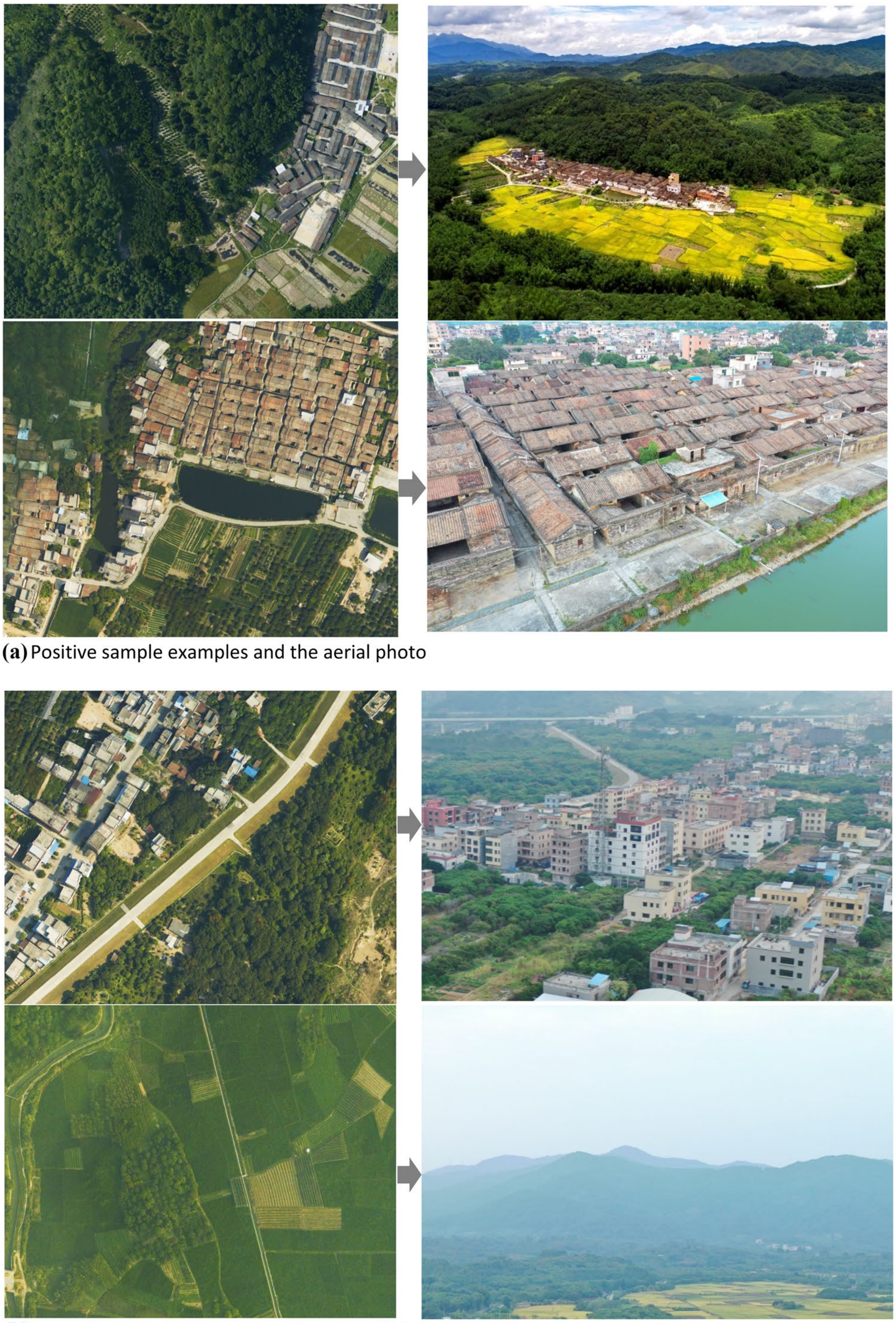
A total of 1111 samples were randomly selected from all samples, including 529 samples that included historical villages as positive samples and 582 samples that did not include historical villages as negative samples. The images were divided into the training set and the test set at a ratio of 4:1, resulting in 417 positive training samples, 466 negative training samples, 112 positive testing samples and 116 negative testing samples (Fig. 4).

(2) Image classification algorithm training

The image classification algorithm is trained using the sample training data set with the binary data annotation to form the target recognition function of the historic village, and it then calculates the classification accuracy rate using the test data set. To improve effectiveness and stability of the algorithm's training, we conduct comparative experiments in two ways using the same network structure.

Direct learning Direct learning is a common training method relative to transfer learning below, which refers to its training data is from the same data set as the test

Fig. 4 Examples of a classification task dataset

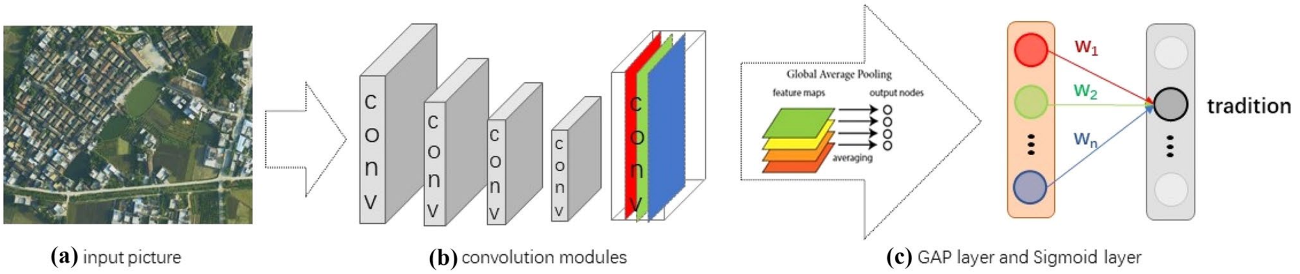


(a) Positive sample examples and the aerial photo

(b) Negative sample examples and the aerial photo

data. We use VGG16, a common VGGNet of CNN structure, for direct learning. A typical CNN classifier consists of two parts (Fig. 5). The convolution structure is used as the skeleton to extract the features of the input picture. The classifier acts as the decision-making layer, performs the weighted averaging and normalization of the output, and finally outputs the probabilities of different categories. The first part of the direct learning network structure is consistent with the structure of VGG16. The second part replaces the traditional full connection layer of VGG16 for 1000 categories into the global average pooling (GAP) layer, which can avoid this layer to overfit and be more robust to spatial translations of the input [30].

Due to the large number of convolutional neural network parameters and few classes and images, it is easy to cause network overfitting and poor robustness. The data



(a) input picture (b) convolution modules (c) GAP layer and Sigmoid layer

Fig. 5 Schematic diagram of image classification pipeline and the GAP structure. *Note* **a** The input picture size is 280 × 224. **b** After the first five convolution modules, the feature map with 512 channels is extracted. **c** After the global average pooling and Sigmoid layer, we get the final results

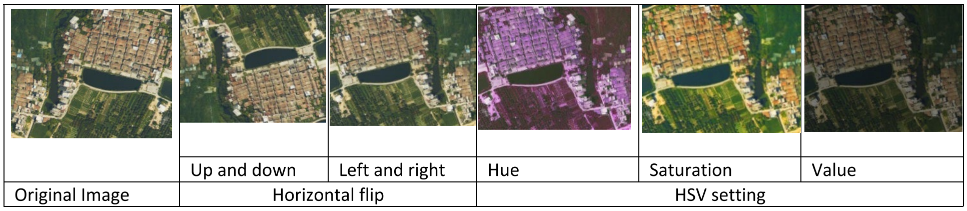


Fig. 6 Example of data augmentation

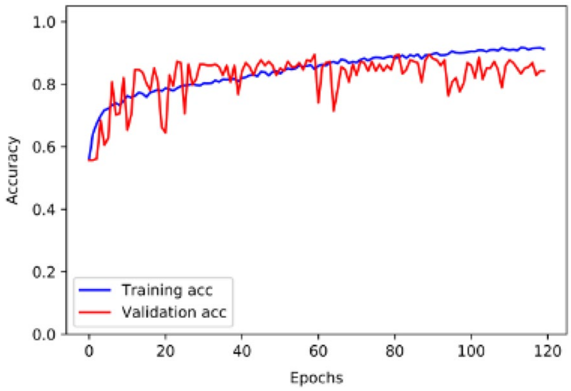


Fig. 7 Direct training process

augmentation methods of flipping images up, down, left and right and randomly setting the hue, saturation and brightness of the image in the HSV space are used to expand the number of samples (Fig. 6).

For the VGGNet, both direct training and transfer learning use same learning schedule. We trained the VGGNet using stochastic gradient descent (SGD) with a batch size of 64, a momentum of 0.9, a weight decay of 0.0005, and a learning rate of 0.01. In addition, we trained each of them for roughly 118 epochs with the whole training set. All of these experiments were run on a personal computer (PC) with a single Intel core i7 central processing unit, an NVIDIA 1080 Ti graphics processing unit with 11-GB memory. The training accuracy and validation accuracy as the number of iterations increases are shown in Fig. 7.

Transfer learning In the hierarchical structure of the CNN, the features that are learned in each layer have different meanings. In general, the lower-level network learns the lower-level features, and the higher-level network learns more advanced semantic information. Due to the small size of the training set, direct training may result in the network not being able to fully learn the characteristics of each level. Therefore, using the transfer learning method, the native VGG16 network is first trained on the ImageNet dataset consisting of 1.28 million images and 1000 categories. In this way, the network can fully learn the visual characteristics of each level. When the network is used to train a new category, the previously learned knowledge can be migrated to the new task (Fig. 8). As such, we only need to save the weight of the first 5 convolution modules of VGG16 and retrain the parameters of the entire network.

The experimental process of transfer learning is shown in Fig. 9. It can be seen that the experimental process converges quickly, and the accuracy is much higher than that of direct training. The transfer learning method can use additional dataset training to make up for the shortcomings related to insufficient training data and greatly save training time.

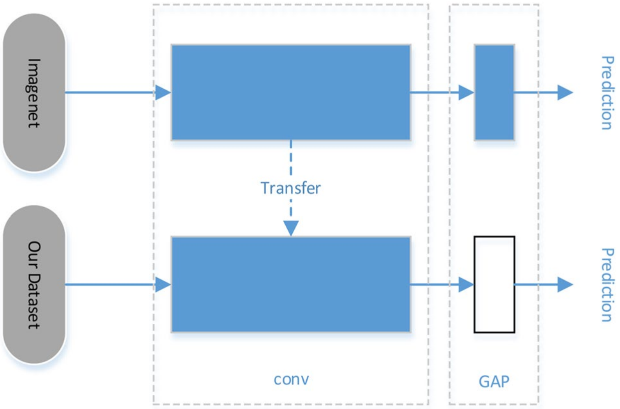


Fig. 8 Schematic diagram of the transfer learning process

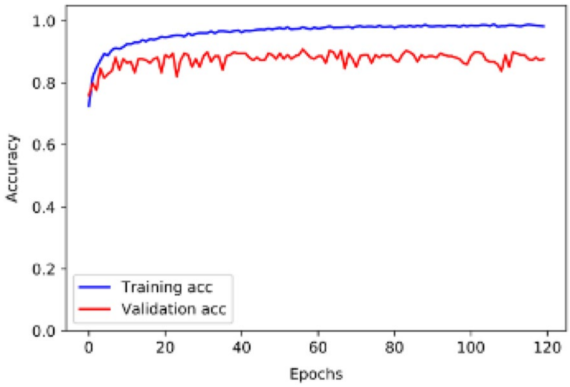


Fig. 9 Transfer learning experiment process

3.2.2 Object detection algorithm

(1) Object detection task dataset:

In fact, most of the historic villages have suffered various degrees of damage, and only a few villages are preserved intactly. Many modern buildings have been built with new technologies and materials. Since the spatial forms of the modern buildings are significantly different from those of traditional buildings, the new and old interlaced village textures exist in remote sensing images. According to this situation, on the basis of the classification task dataset, the villages' morphological features in the target detection positive samples are marked differently (Fig. 10).

- The area where the traditional features of the village are relatively intact will be marked with a Historical tag.
- The area where the historic features of the village are intertwined with modern features will be marked with a Historical-Modern tag.

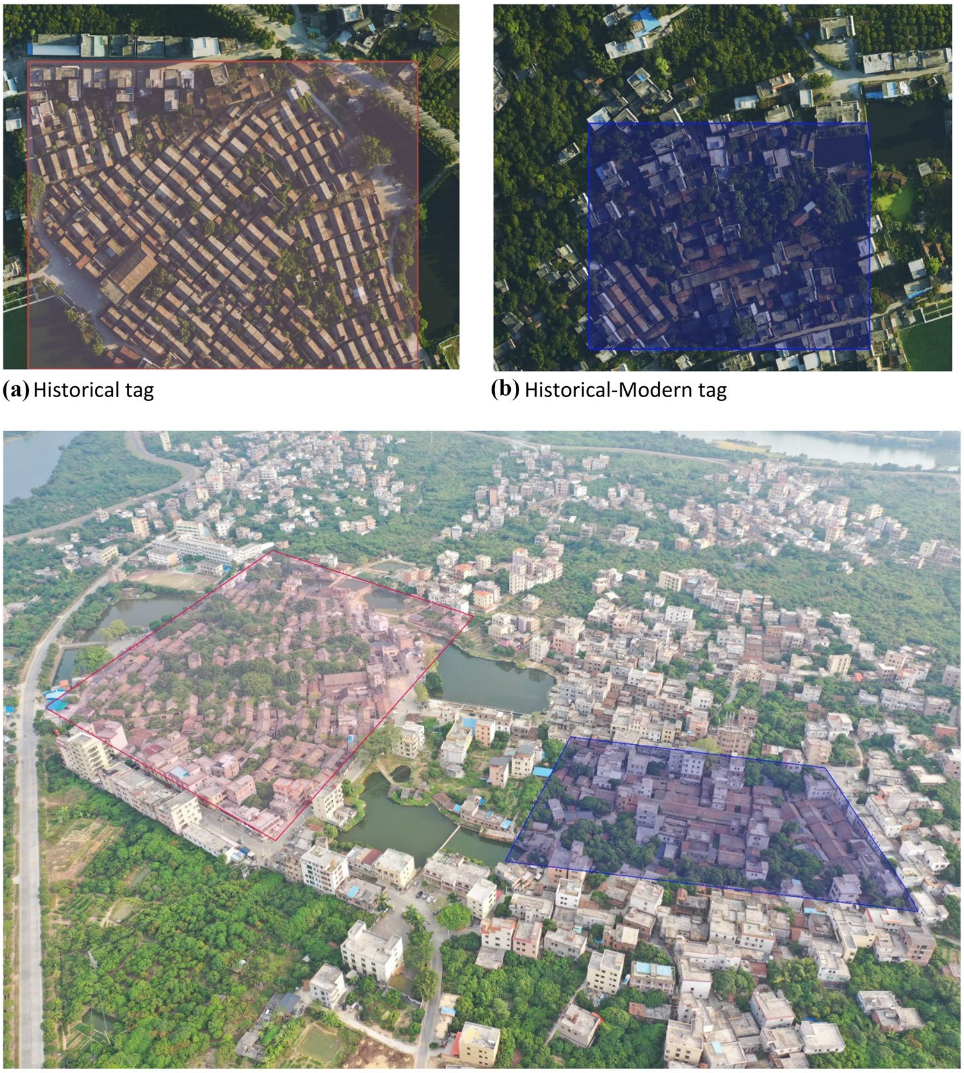
(2) Object detection algorithm training

The object detection method uses the classic two-stage object detection network Retinanet [31], which improves the recognition accuracy by training the object detection network using structured data annotations. The Retinanet network introduces a focus loss function that resolves the sample imbalance problems that are caused by too few positive samples in the target object training set or too few target instances in the positive sample. According to the pipeline, first, the training set images are input into Retinanet. Next, the network extracts the features through the convolution skeleton, and it then connects the two sub-networks of the regression and classification. Finally, it outputs the coordinates of the target (the coordinates of the upper left corner and the lower right corner) and the probability that the target belongs to a certain category (Fig. 11).

Since the project essentially classifies images, post-processing of the object detection results of the previous step is needed. As shown in Fig. 12, according to the Non-maximum suppression method [33], the 80 bounding boxes that are most likely to contain targets are extracted from an input image and the probabilities of these 80 boxes are listed. If there is a box with a probability of being Historical or Historical-Modern greater than the empirical threshold of 0.2, then the image is considered to contain historic villages. Otherwise, if the 80 boxes do not contain a village, the detection result of this picture is considered non-Historical.

For the Retinanet Object detection algorithm, both experiments use the same settings. Following original Retinanet, we trained them using SGD with a batch size of 2, a momentum of 0.9, a weight decay of 0.0001, and a learning rate of 0.01. In addition, we trained each Retinanet for roughly 118 epochs with the whole training set. The training process of the object detection algorithm network is shown in Fig. 13. Figure 13a shows the training process of combining the Historical and the Historical-Modern villages into one class. Figure 13b shows the training process divided into two categories. Both methods converge faster, and in the first epoch the accuracy of the validation set is more than 80%. Since the Adam optimization method automatically decreases the learning rate, the gradient and accuracy at the end of the training hardly change.

Fig. 10 Example images for the object detection task



(a) Historical tag

(b) Historical-Modern tag

(c) Aerial photo of the historic village and historical-Modern village

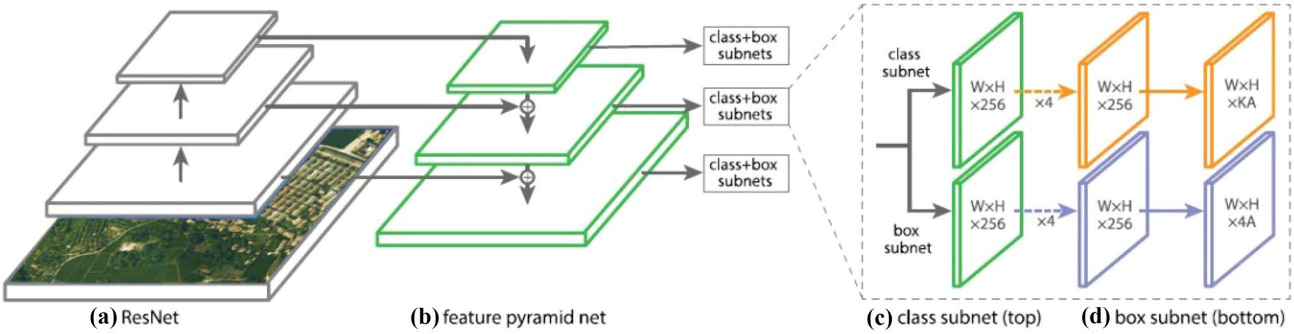
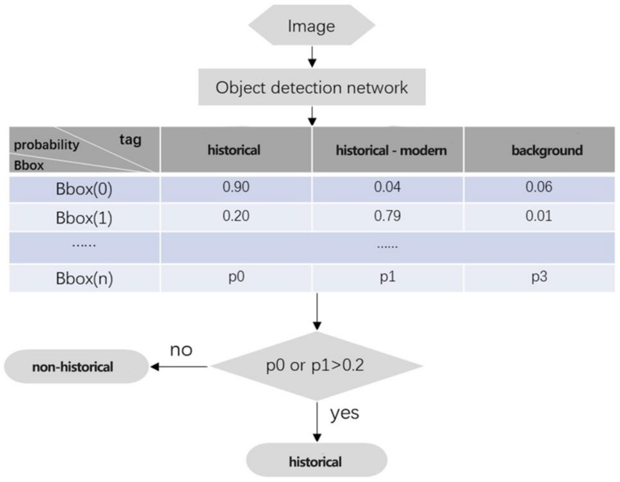


Fig. 11 Schematic diagram of Retinanet structure. *Note* The RetinaNet network uses a Feature Pyramid Network (FPN) as backbone (**a**) to generate a multi-scale convolutional feature pyramid (**b**). Two sub-networks are attached to this backbone, one classifies the objects in the box (**c**), and one regressing to the more accurate position of the object from the box (**d**) [32]



probability / tag / Bbox	historical	historical - modern	background
Bbox(0)	0.90	0.04	0.06
Bbox(1)	0.20	0.79	0.01
……		……	
Bbox(n)	p0	p1	p3

Fig. 12 Process after object detection

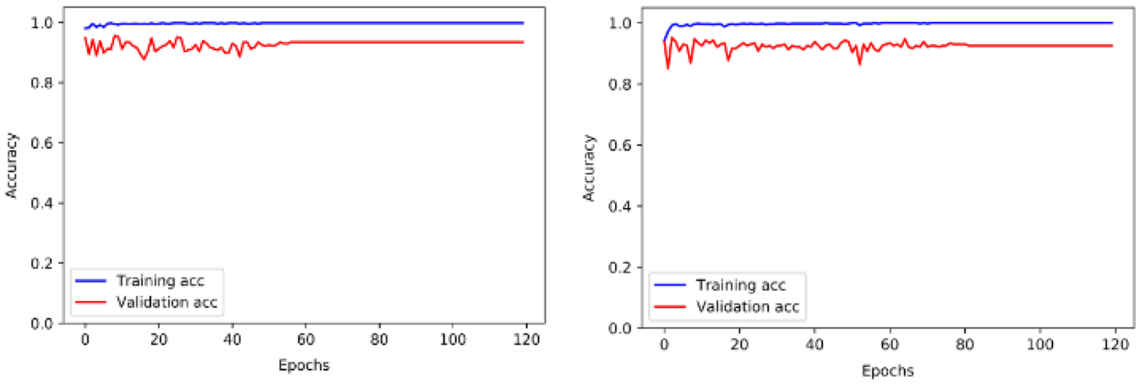
4 Results

4.1 Image classification algorithm recognition results

The image classification algorithm network with the best results (shown in Figs. 7, 8) is saved, and the input test set is tested to obtain Table 1. This experiment encodes the traditional category as 1 and the non-traditional category as 0.

From Table 1, it can be calculated that the accuracy of the direct training method is 89.47%, and the accuracy of the transfer learning method is 90.79%. Although there is not much improvement, from their training curves in Figs. 7 and 8, it suggests that the starting point of direct training is relatively low, the oscillation is more severe, and the training process is not stable. In addition, the direct training method converges slowly until the 28th epoch, and the accuracy rate is more than 80%. For the transfer learning in the 2nd epoch, the accuracy rate is more than 80%. It can be concluded that the transfer learning method has greater advantages than the direct learning method with fast convergence, stable training, and high accuracy.

To analyse the identification algorithm's recognition of the specific structures in an image, the recognition results are visualized [34].to determine whether the function identifies a historic village as the target object. As shown in Fig. 14, the first two columns are the pictures with accurate recognition results, the last two columns are pictures with incorrect recognition results, the first line is the pictures from the test set, and the second line is the corresponding visual result pictures. Among them, red indicates that the activation response is strong, and, ideally, it should overlap with the position



(a) Combining the Historical and Historical -Modern tags

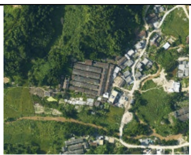
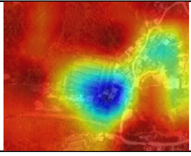
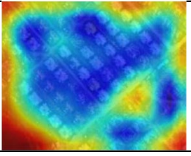
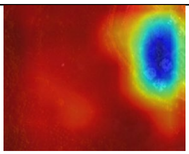
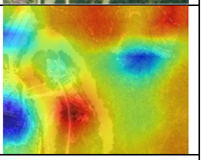
(b) Distinguishing between Historical and Historical-Modern tags

Fig. 13 Object detection network experiment process

Table 1 Test results of image classification network algorithms

(a) Test results of direct training methods				(b) Test results of transfer learning method			
	Actual: 0	Actual: 1	Total		Actual: 0	Actual: 1	Total
Predicted: 0	99	7	116	Predicted: 0	105	10	115
Predicted: 1	17	105	122	Predicted: 1	11	102	113
Total	116	112		Total	116	112	

Fig. 14 Examples of the image classification algorithm's recognition results visualization

Test images				
Visualization of results				
Test Results	Positive==>positive	Negative==> negative	Negative==>positive	Positive==>negative

of a historic village. We can see that, actually, the neural network does not truly grasp the characteristic representation of the historic village, and it only learns the correlation between the traditional village and the surrounding vegetation, although the recognition result of the whole test set is in an acceptable range. Therefore, this recognition result lacks credibility.

4.2 Object detection algorithm recognition results

4.2.1 Combining the Historical and Historical-Modern tags

In the first case, the Historical tag and the Historical-Modern tag are combined into the Historical-A1 tag for the historic village for identification. After assessing the test data set, the results indicate that the recognition accuracy of the historic villages reached 95.61% (Table 2a).

4.2.2 Distinguishing between Historical and Historical-Modern tags

In the second case, the Historical tag and the Historical-Modern tag are distinguished and identified as two different types of historic villages. In this case, both the Historical tag and Historical-Modern tag can be considered as recognition targets, and the average recognition accuracy is 95.17% (Table 2b).

Through the results of visual analysis (Fig. 15, comparison of the visual results of object detection network algorithms), it can be found that the Historical recognition area is indeed a well-preserved village, and the Historical-Modern recognition area is located above the old and new village. The results prove that the recognition function is effective, and the algorithm can more accurately identify different types of historic villages. In addition, the recognition accuracy of the two tags is slightly lower because of the more extensive classification, which increases the learning burden of the network.

4.3 Overall identification results of Conghua area

Because of the high accuracy rate and the effectiveness, the object detection algorithm is used to identify the remote sensing images of Conghua area. Then, the coordinates of the identified traditional villages are recorded and converted into the world coordinate system; therefore, the actual locations of the villages can be displayed and visualized through a GIS system. According to the identification results, there are a large number of villages (1531) with certain historic characteristics scattered in Conghua area (Fig. 16). Their spatial distribution is extensive, covering the entire region, and the number of historic villages in the western river valley is much larger than that in the eastern mountainous areas. The government of Conghua has announced 31 historic villages. Their spatial distribution is shown in Fig. 16, but this only accounts for 2% of the identified villages. This suggests, to some extent, that the government needs to strengthen their investigation of historic villages and more villages with valuable heritages should be included in the protection list. The villages that are identified in this study can be used as the basis

Table 2 Test results of the object detection algorithms

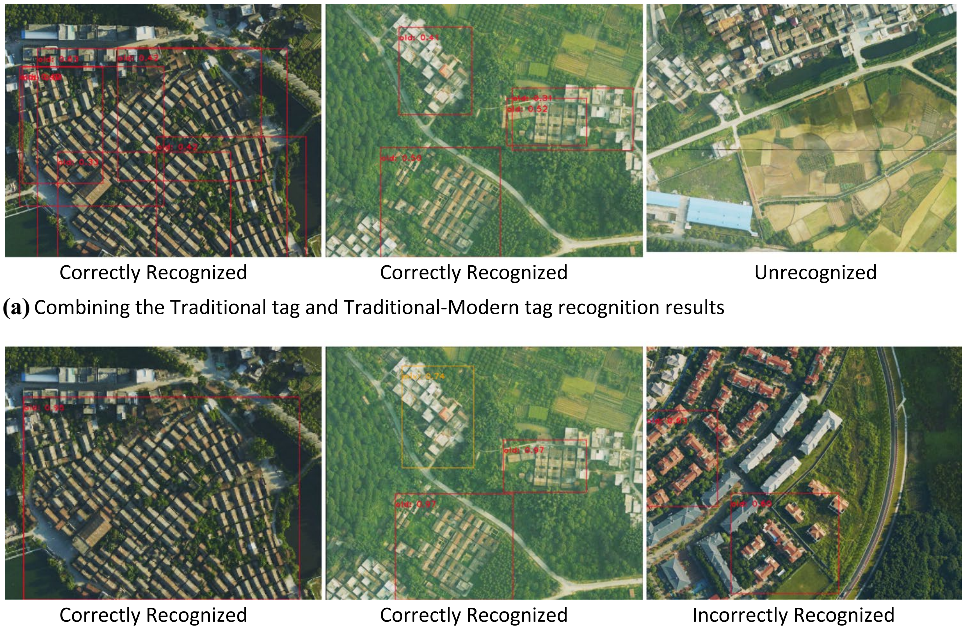
(a) Combining the Historical tag and the Historical-Modern tag recognition results

	Actual: 0	Actual: 1	Total
Predicted: 0	112	6	118
Predicted: 1	4	106	111
Total	116	112	

(b) Distinguishing between the Historical tag and Historical-Modern tag recognition results

	Actual: 0	Actual: 1	Total
Predicted: 0	109	4	113
Predicted: 1	7	108	115
Total	116	112	

Fig. 15 Comparison of the visual results of object detection network algorithms



(a) Combining the Traditional tag and Traditional-Modern tag recognition results

(b) Distinguishing between traditional labels and Traditional-Modern tag recognition results

for a field investigation, which will make the work more targeted.

5 Discussion and conclusion

In this paper, we trained a convolutional neural network model to identify historic villages. It is applied in the identification task of Conghua area. A total of 1531 villages are automatically recognized, and their distribution map is drawn. In the course of the experiment, a variety of recognition models and methods were tested, and the most suitable identification algorithm for historic villages was obtained through comparative experiments. The recognition rate is as high as 95.61%, and the selected sample only accounts for less than 5% of the whole sample, which indicates that the algorithm is accurate and efficient. This provides a new possibility for the application of deep learning technology in the field of architectural heritage conservation.

In terms of the testing data sets, since the standard frames of remote sensing images generally cover larger areas and the proportion of villages in the images is too small, the identification accuracy might be affected. Therefore, the pictures need to be cropped to an appropriate size so that the villages can occupy larger proportions in the images. Furthermore, compared to the general village building identification task where a large sample of images can be obtained from Google Earth [26], the sample of historical villages as a special kind of architectural heritage is limited in reality; thus, it is necessary to ensure that the proportion of training samples with respect to the whole is not too high. Otherwise, the recognition results will be meaningless. The transfer learning that is used in this experiment, as well as the data augmentation methods such as transformation of direction and colouration of the original image, can effectively expand the data set, and thus may improve the identification of small samples such as historic villages.

With respect to selecting the recognition algorithm, transfer learning has high accuracy, and the training process rapidly converges, which has certain advantages compared with direct training. However, according to the visual analysis, the recognition objects of the image classification algorithm are not historic villages, but rather are the mountains and green environments around the village; therefore, its identification results lack credibility for historic villages. Coincidentally, however, the recognition results that are obtained from the mountains and green environments do include historic villages, and the accuracy rate is as high as 90.79%. This outcome has forced us to re-examine the relationship between the historic villages and the surrounding landscape environment. In fact, the location of ancient Chinese villages is influenced by the local natural and human environment and will reflect a relatively fixed spatial pattern. Therefore, compared to previous studies that identified ancient villages by a single dimension of architecture, such as the identification of architectural textures [27] and edges [15], the results inspire us to recognize historic villages in terms of the relationship between architecture and nature.

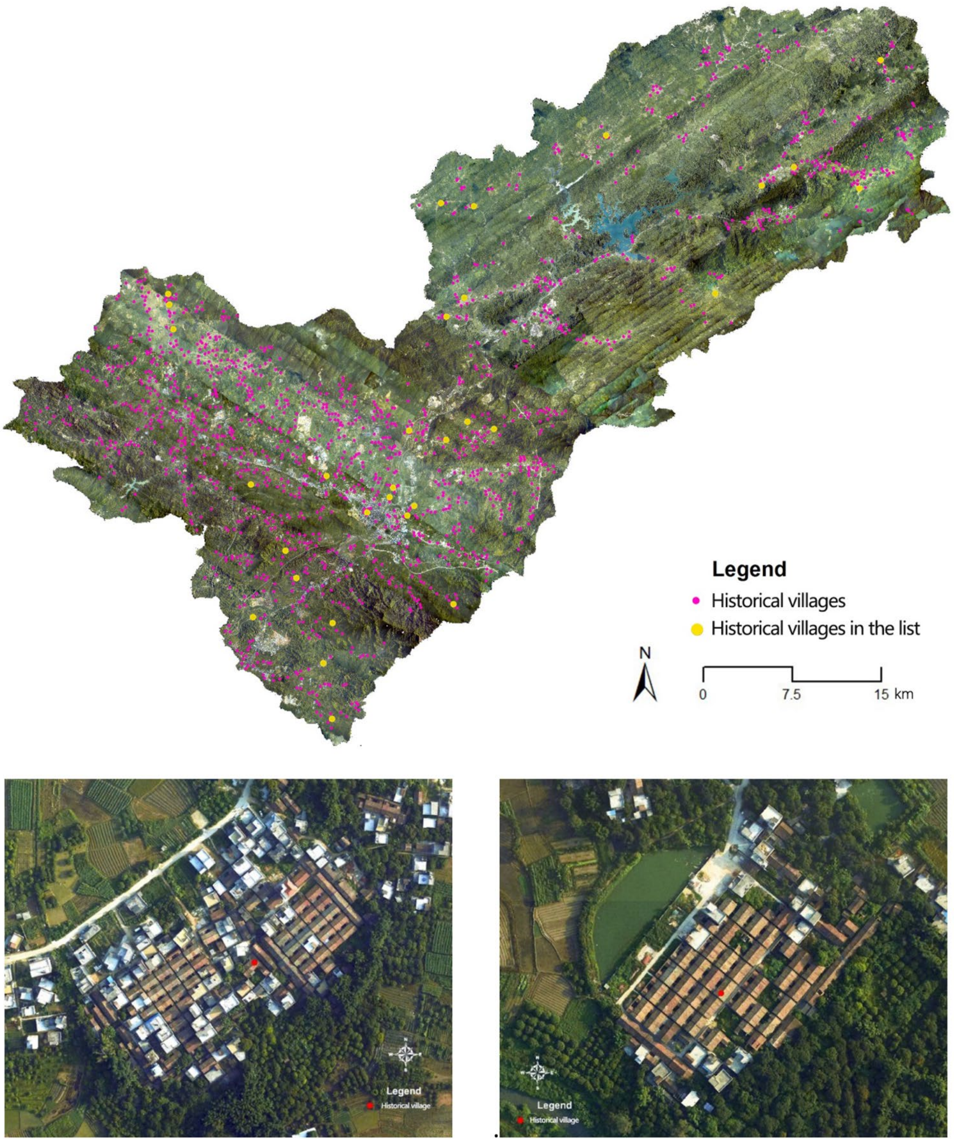


Fig. 16 The results of the identification of traditional villages in Conghua area

When using the object detection algorithm, identification errors occur mainly because the villages with weak historic features are not identified when distinguishing between Historical and Historical-Modern tags. Since the villages in the Conghua area are mainly dominated by new and old ones, but the ratio between the new and the old is different, therefore, it is worth considering what kind of village should be regarded as a historic village. The option of the Historical-Modern tag can bring adaptive, resilient answers to this issue. By using the Historical-Modern tag as a filter, it is possible to select a village with a certain proportion of historic features according to specific requirements, thus providing a scientific basis for the determination of protection objects and the formulation of protection policies.

Authors' contribution All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Jin Tao, Huicheng Feng, Dawei Xiao and Qiwei Sun. The first draft of the manuscript was written by Jin Tao, Huicheng Feng, Geng Li, Youjia Chen and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This work was supported by the National natural science foundation of china, Grant number 51878283; Guangdong basic and applied basic research foundation, Grant number 2021B1515020013 and the fundamental research funds for the central universities, Grant number x2jz/D2190720.

Availability of data and materials The remote sensing images that support the findings of this study are available in the Conghua Municipal Planning and Natural Resources Bureau. Access to these data can be allowed based on reasonable request submitted to the organization.

Code availability Arcgis software was used to calculate the villages' geographical coordinates and to display and visualize their distribution.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.The authors declare that they have no conflict of interest.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

1. Brunskill RW (2000) Vernacular architecture: an illustrated handbook. Faber and Faber, London
2. Kowalewski SA (2008) Regional settlement pattern studies. J Archaeol Res 16:225–285. https://doi.org/10.1007/s10814-008-9020-8
3. Arnold III PJ, Stark BL (1997) Gulf lowland settlement in perspective (1997), Olmec to Aztec: settlement patterns in the ancient Gulf Lowlands, University of Arizona Press, Tucson, pp 310–329
4. Liu W, Wang Z, Liu X, Zeng N, Liu Y, Alsaadi FE (2017) A survey of deep neural network architectures and their applications. Neurocomputing 234:11–26. https://doi.org/10.1016/j.neucom.2016.12.038
5. Long Y, Liu L (2016) Transformations of urban studies and planning in the big/open data era: a review. Int J Image Data Fusion 7:295–308. https://doi.org/10.1080/19479832.2016.1215355
6. Peng Y, Nijhuis S (2021) A GIS-based algorithm for visual exposure computation: the west lake in Hangzhou (China) as example. J Dig Landsc Archit 6:424–435
7. Van Eetvelde V, Antrop M (2004) Analyzing structural and functional changes of traditional landscapes—two examples from Southern France. Landsc Urban Plan 67:79–95. https://doi.org/10.1016/S0169-2046(03)00030-6
8. Günçe K, Ertürk Z, Ertürk S (2008) Questioning the "prototype dwellings" in the framework of Cyprus traditional architecture. Build Environ 43:823–833. https://doi.org/10.1016/j.buildenv.2007.01.032
9. Patino JE, Duque JC (2013) A review of regional science applications of satellite remote sensing in urban settings. Comput Environ Urban Syst 37:1–17. https://doi.org/10.1016/j.compenvurbsys.2012.06.003
10. Longbotham N, Chaapel C, Bleiler L, Padwick C, Emery WJ, Pacifici F (2011) Very high resolution multiangle urban classification analysis. IEEE Trans Geosci Remote Sens 50:1155–1170. https://doi.org/10.1109/TGRS.2011.2165548
11. Moser G, Serpico SB, Benediktsson JA (2012) Land-cover mapping by Markov modeling of spatial–contextual information in very-high-resolution remote sensing images. Proc IEEE 101:631–651. https://doi.org/10.1109/JPROC.2012.2211551
12. San Emeterio JL, Mering C (2021) Mapping of African urban settlements using Google Earth images. Int J Remote Sens 42:4882–4897. https://doi.org/10.1080/01431161.2021.1903613
13. Vaduva C, Gavat I, Datcu M (2012) Deep learning in very high resolution remote sensing image information mining communication concept. IEEE, pp 2506–2510
14. Miyazaki S, Fujii A (2011) Identification of buildings in different GIS data map using the Boolean operation method. J Asian Archit Build Eng 10:125–131. https://doi.org/10.3130/jaabe.10.125
15. Gao C, Sang N, Gao J, Tang Q (2010) Cascade of hierarchical context and appearance for object detection. Opt Eng 49:037003
16. Hofmann P (2001) Detecting informal settlements from Ikonos image data using methods of object oriented image analysis-an example from Cape Town (South Africa). Remote Sens Urban Areas/Fernerkundung in urbanen Räumen 35:107–118
17. Pesaresi M, Bianchin A (2003) Recognizing settlement structure using mathematical morphology and image texture. Remote Sens Urban Anal GISDATA 9:46–60
18. Sun X, Fu K, Long H, Hu Y, Cai L, Wang H (2008) Contextual models for automatic building extraction in high resolution remote sensing image using object-based boosting method. IEEE, pp II-437–II-440
19. Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. Science 313:504–507. https://doi.org/10.1126/science.1127647
20. Sharif Razavian A, Azizpour H, Sullivan J, Carlsson S (2014) CNN features off-the-shelf: an astounding baseline for recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 806–813
21. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3431–3440
22. Cha YJ, Choi W, Büyüköztürk O (2017) Deep learning-based crack damage detection using convolutional neural networks. Comput-Aided Civ Infrastruct Eng 32:361–378. https://doi.org/10.1111/mice.12263
23. Gonzalez D, Rueda-Plata D, Acevedo AB, Duque JC, Ramos-Pollan R, Betancourt A, Garcia S (2020) Automatic detection of building typology using deep learning methods on street level images. Build Environ 177:106805. https://doi.org/10.1016/j.buildenv.2020.106805
24. Rueda-Plata D, González D, Acevedo AB, Duque JC, Ramos-Pollán R (2021) Use of deep learning models in street-level images to classify one-story unreinforced masonry buildings based on roof diaphragms. Build Environ 189:107517. https://doi.org/10.1016/j.buildenv.2020.107517
25. Vakalopoulou M, Karantzalos K, Komodakis N, Paragios N (2015) Building detection in very high resolution multispectral data with deep learning features. In: International geoscience and remote sensing symposium (IGARSS) 2015-Novem:1873–1876. https://doi.org/10.1109/IGARSS.2015.7326158
26. Guo Z, Shao X, Xu Y, Miyazaki H, Ohira W, Shibasaki R (2016) Identification of village building via Google Earth images and supervised machine learning methods. Remote Sens 8:271. https://doi.org/10.3390/rs8040271
27. Lei MA, Haowen YAN, Zhonghui W (2017) Geometry shape measurement of building surface elements based on self-supervised machine learning. Sci Survey Mapp 42:171–177

28. Shirowzhan S, Lim S, Trinder J, Li H, Sepasgozar SME (2020) Data mining for recognition of spatial distribution patterns of building heights using airborne lidar data. Adv Eng Inform 43:101033. https://doi.org/10.1016/j.aei.2020.101033
29. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:14091556
30. Lin M, Chen Q, Yan S (2013) Network in network. arXiv:13124400
31. Lin T-Y, Goyal P, Girshick R, He K, Dollár P (2017) Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision, pp 2980–2988
32. Lin T-Y, Dollár P, Girshick R, He K, Hariharan B, Belongie S (2017) Feature pyramid networks for object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2117–2125
33. Rothe R, Guillaumin M, Gool LV (2014) Non-maximum suppression for object detection by passing messages between windows. In: Asian conference on computer vision. Springer, pp 290–306
34. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A (2016) Learning deep features for discriminative localization. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2921–2929