SEMANTIC SEGMENTATION METHOD ACCELERATED QUANTITATIVE ANALYSIS OF THE SPATIAL CHARACTERISTICS OF TRADITIONAL VILLAGES

Zhang Mengdi [1], Li Zhe [1,2,*], Wu Xiaomin [3]

[1]School of Architecture, Tianjin University, Tianjin, 300072, (meng123, lee_uav)@tju.edu.cn
[2]Key Laboratory of Information Technology for Architectural Cultural Inheritance, Ministry of Cultural and Tourism, Tianjin, 300072
[3]School of Urban Design, Wuhan University, Wuhan, 430072

KEY WORDS: Semantic Segmentation, Convolutional Neural Network (CNN), Remote Sensing Image, Traditional Chinese village, Spatial Form.

ABSTRACT:

Rapid investigation and quantitative analysis are crucial for heritage conservation and renewal design. As an important category of architectural heritage - traditional settlements - with their large number and complex spatial characteristics, their spatial character patterns are an important support to assist settlement conservation and renewal design. However, the current means of analysis often requires manual data collection, secondary mapping of the collected data, extraction of individual elemental patterns and village boundaries. Then settlement boundary form, settlement density will be calculated by mathematical methods. The above methods are inefficient and prone to manual mapping errors, making it difficult to quantify and analyze a large number of traditional villages in a short period of time. Semantic segmentation is a computer vision technique for quickly segmenting different objects. Based on the collected remote sensing data of traditional villages, this paper established a dataset of semantic segmentation of spatial features of traditional settlements, segmenting village buildings, water systems, roads and plants. Using Transfer learning, data augmentation and other methods, a model was trained that can automatically segment elements of the villages. From the national traditional villages that have been announced so far, 60 traditional villages from different regions in the north and south were selected for analysis. The experiments show that the model established in this paper has an accuracy rate of above 86% in segmenting elements of villages, can effectively identify the location of different elements in remote sensing images, effectively improves the quantification rate of spatial features of settlements and saves the cost of mapping and data transcription. The results of the spatial characteristics of the 60 villages studied in this paper can also provide some theoretical basis and inspiration for the study, conservation, design and transformation of traditional villages.

1. INTRODUCTION

The traditional village is a representative of Chinese residential architecture, and is a settlement space with basic forms such as cluster, band and radiation, formed by people gathering and settling in accordance with the natural environment and social culture (Hao and Zao, 2019). The number of national-level traditional villages certified in China in 2020 has reached 6,819, making it one of the largest living farming settlements in the world and an important vehicle for the transmission of Chinese civilization (Zhe et al., 2019). With the rapid promotion of China's rural revitalization project, accelerated urbanization and urbanization migration of the population, traditional villages have undergone different degrees of extinction, alteration and construction of new villages, which followed with the destruction of the village style or complete disappearance of traditional village features (Ya-Juan et al., 2016). In order to record valuable regional cultural features, the existing spatial patterns urgently need to be extracted and counted, thus assisting the implementation of corresponding design decisions.

In the study of quantification of traditional village features, commonly the CAD plans of all buildings in the village are drawn manually first (Yun, 2009), buildings are abstracted as points, and then the mathematical model is established by computer and the concept of village space is analyzed, or the village morphology is quantified through fractal dimension (Weiguo and Mengjia, 2021), accordingly the quantification of the village interior spatial characteristics of the village can be realized. With the development of photogrammetry and point cloud technology, automatic extraction of point cloud data based on village features has been realized (Zhe et al., 2019), however, it still requires manual operation of UAVs to collect images for village field research and obtain point cloud data through photogrammetry. In summary, the current approaches to spatial quantification of traditional villages to some extent require manual collection, data extraction, or remapping of image data. The extraction of spatial features of all national traditional villages by the current methods is labor-intensive. Therefore, it is difficult to achieve a spatial overview of all villages in a short period of time.

In order to improve efficiency of the quantification, we propose an automatic quantification system of traditional village spatial features based on remote sensing data, which adopts deep learning techniques and computer vision technology. The contributions of this paper are summarized in the following points:

1. A semantic segmentation dataset of remote sensing images of Chinese traditional villages had been built, a model that can be used to segment remote sensing images of villages was trained by transfer learning, and then the effectiveness of this model was verified within the test dataset, finally, the automatic segmentation of the locations of buildings in

Chinese traditional villages based on satellite images was realized.

2. Based on the semantic segmentation of village architecture, automatic extraction of village boundaries and automatic quantitative calculation of village morphology were achieved using computer vision algorithms in this paper.
3. Traditional villages in different regions were also analyzed using the method in point 2, thus the spatial features of different villages were compared and the distribution characteristics in different provinces were summarized.

The remainder of the paper is organized as follows: the progress of research on semantic segmentation based on remote sensing images and deep learning is introduced in section 2. A deep learning method for semantic segmentation of traditional village architecture based on remote sensing images, and an automatic quantification method for village spatial features based on the results of semantic segmentation are proposed in section 3. The experimental results are analyzed in section 4, and section 5 is a discussion of this method. The final section is the conclusion of the paper and in which the future research situation is discussed.

2. RELATED WORK

The method of extracting buildings based on remote sensing images is widely used in urban planning, natural disaster prevention, and population area development management (Liu et al., 2017). In earlier studies, the extraction of buildings relied mainly on the spectral-structural properties of buildings (Xin et al., 2013) or semi-automatic extraction of buildings based on edge detection and linear features of buildings (Dan and Wei-Dong, 2011). As deep learning techniques continue to advance, the semantic segmentation task, whose assignment is to predict the category of each pixel in the image, is being accomplished much more efficiently. Neural networks such as Fully Convolutional Networks (FCNS) (Long et al., 2015) and U-Net (Ronneberger et al., 2015) have been proposed and gradually applied in semantic segmentation tasks, making it possible to automatically segment buildings in remote sensing images. The method has been applied in urban sprawl monitoring, urban village extent detection (Qian et al., 2019) , and automatic road extraction (Alshaikhli et al., 2019) to improve the efficiency of the corresponding studies.

In deep learning methods, it is necessary to rely on a large number of datasets for training and learning in order to obtain models for recognition and prediction, especially for semantic segmentation algorithms, but there is a lack of a suitable dataset for traditional Chinese villages. Currently, there are WHU building dataset (Dengxin and Wen, 2011), Massachusetts building dataset (Mnih, 2013) , and Inria Aerial Image dataset (Maggiori et al., 2017) , of which have thousands of high-resolution original and annotated images. However, there is a certain gap between the three and the remote sensing images of traditional villages (Figure 1). The former datasets have more regular forms and greater building spacing, while the latter has irregular forms and small building spacing. So, all the three datasets are not suitable for the traditional village building segmentation task in this paper. It is necessary to build the dataset based on the remote sensing data of the traditional villages by manually labelling the buildings.

(a) WHU building dataset (b) Massachusetts building dataset

(c) Inria Aerial Image dataset (d) China traditional village

Figure 1. Comparison of three semantic segmentation dataset images of architectural with remote sensing images of traditional Chinese villages.

3. METHOD

This paper introduces remote sensing images and deep learning methods in the spatial quantification of traditional villages to achieve semantic segmentation of buildings; and combines computer vision related algorithms to achieve automatic quantitative extraction of spatial features of traditional villages based on the results of semantic segmentation.

3.1 Date Pre-Processing

In this paper, five remote sensing images of traditional villages in different areas from the five batches of national-level lists published in China were selected as the annotation objects, and the Google remote sensing images were selected with the resolution between 0.4-0.8m. The buildings are labelled in white and the background in black.

According to the experimental results in the paper (Etten, 2018), for large scale remote sensing images, they can be partitioned into manageable cutouts first and ensure a certain overlap rate between images to avoid incomplete recognition targets due to splitting images. After that, the model is trained and runed to analyze the remote sensing images for prediction. The final step seeks to stitch together the prediction into one final image strip. The accuracy of the operation results can be guaranteed while the amount of operation does not exceed the memory requirement. Therefore, in this paper, the annotated traditional village images are partitioned into 512*1024 size images (see Figure 2). A total of 520 original images and corresponding labelled images were finally obtained. The workflow of image slicing and model training and prediction is shown in Figure 2.

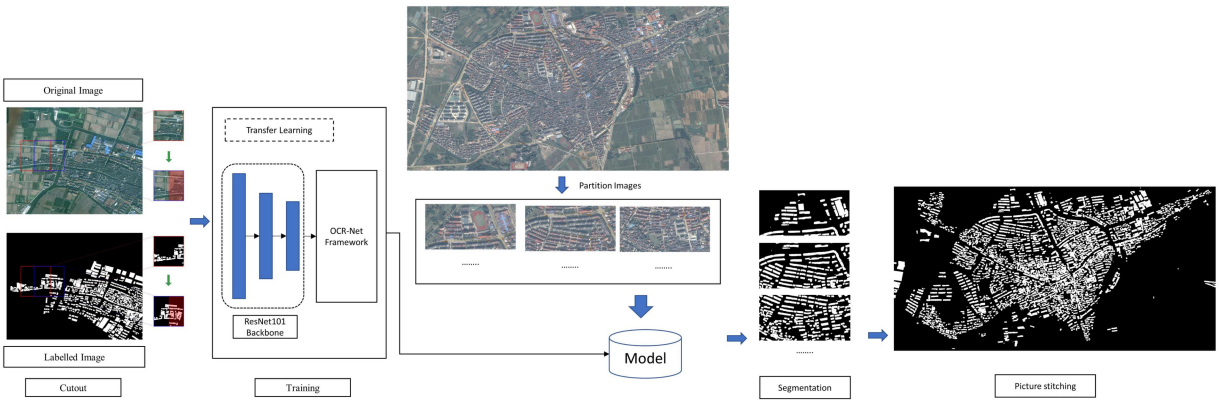


Figure 2. Pipeline of this research for traditional village semantic segmentation using deep learning.

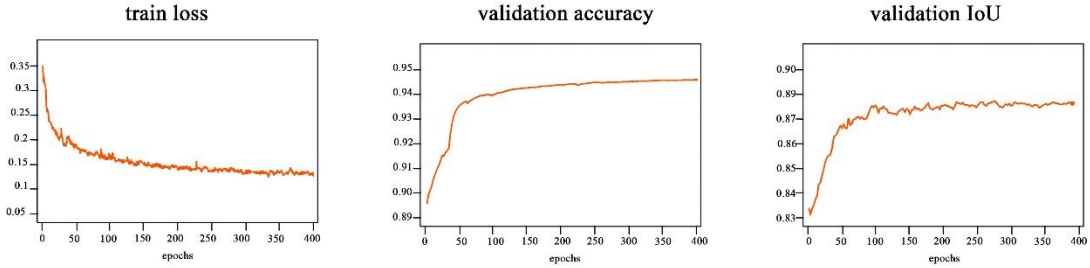


Figure 3. Training loss, validation accuracy and validation IoU by training epochs.

To cope with problems such as changes in lightness and darkness and image deformation that may occur in real remote sensing images, and to enhance the robustness and the performance of the model, image dataset augmentation was applied. A total of 2600 images were obtained using random rotation, width and height adjustment, scaling and horizontal and vertical flip, and lightness and darkness adjustment, of which 70% were used for training, 20% for validation, and 10% for testing.

3.2 Deep Learning Model and Transfer Learning

The OCR-Net model used in this study has achieved 83.6% accuracy on the Cityscapes dataset, which is a standard dataset used to measure the accuracy of semantic segmentation algorithms (Yuan et al., 2020), owns the highest accuracy on the current semantic segmentation dataset. OCR-Net is constructed with contextual information explicitly augmented with contribution values from the same class of object pixels to enhance the accuracy of the final model prediction.

Usually, a large dataset is indispensable in order to obtain a reliable and accurate model, but when it is difficult to build a large dataset, transfer learning is a better approach (Ali et al., 2014). The method is to train the model with a large dataset that is relatively similar to the target dataset, and the resulting model is then fine-tuned and trained with the target dataset.

Compared with the three image semantic segmentation datasets with thousands of very large resolution data mentioned in the previous paper, the dataset built in this paper has only five large scale images before segmentation, which is a small sample dataset and is suitable for the training method of transfer learning. In this paper, the Inria Aerial Image dataset was selected for training, because it has a certain degree of similarity to traditional villages and has more than 9,000 annotated images after partition. After obtaining the pre-trained model, transfer learning is performed with the traditional village dataset built in this paper.

3.3 Training

The specific training method is as follows. First, the Inria Aerial Image dataset was used on the OCR-Net network to obtain a model that can be used to recognize the Inria Aerial Image dataset, and then the weights were updated using transfer learning.

After 430 epochs of training, the final loss and accuracy were relatively stable (Figure 3). This study used the mmsegmentation[1] toolbox for experiments implemented on an NVIDIA 2080Ti GPU under CUDA 10.0 on an Ubuntu 16.04 system.

3.4 Automatic Extraction of Boundaries and Quantization Based on Semantic Segmentation Images

Relying only on the semantic segmentation results of remote sensing images of villages, it is not possible to directly extract information such as the boundaries of villages to achieve the quantification of village space. The currently available algorithms for automatic extraction of village boundaries require drawing CAD plans of village buildings, then manually extracting the location point of each building based on the CAD plans, thus extracting the village boundaries based on the set of location points using the Delaunay Triangulation algorithm (Xincheng et al., 2020).

The edge detection algorithm provides a method for automatically finding the boundary contours of a particular graph. In computer vision, there are several edge detection algorithms such as Canny operator, Sobel operator, etc. Among them, Canny operator has lower error rate, higher positionability

[1] https://github.com/open-mmlab/mmsegmentation.git

and minimum response (Canny, 1986), so Canny operator is chosen for this paper.

The Canny algorithm was introduced in this paper to extract the outer contour of each building in the semantic segmentation result. The four location points of each building were obtained by calculating the smallest outer rectangle of the contour. The collection of location points for all the buildings in the village can be obtained.

However, some villages or are on the same image with other villages, or around which there are new buildings. In order to extract the boundaries of buildings of the same village only as much as possible, the DBSCAN clustering algorithm was used in this paper to aggregate consecutive and close location points together as building points of a village (clustering extraction results in Figure 4).

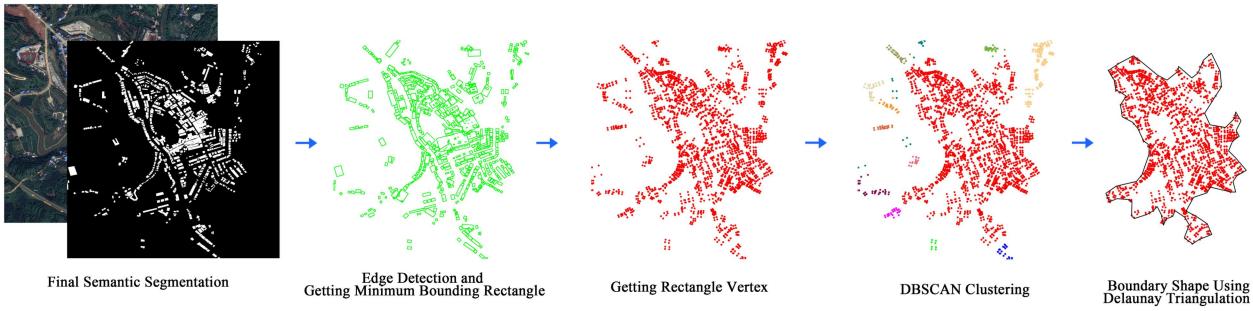


Figure 4. Methodology process diagram of extracting village boundary automatically by computer vision algorithms and semantic segmentation images.

Then the Delaunay Triangulation algorithm is used to find the boundaries of the villages based on the set of location points as shown in Figure 4. The automatic extraction of building locations reduces the need for manual spotting of building locations.

After obtaining the village boundary, the shape of the village was measured by the shape index in landscape ecology (Pu Xincheng 2013). The formula is given as follows.

$$S = \frac{P}{1.5\lambda - \sqrt{\lambda} + 1.5}\sqrt{\frac{\lambda}{A\pi}} \quad (1)$$

where S = the shape index
P = the perimeter of the village boundary
A = the area of the village
λ= the aspect ratio of the minimum external ellipse of the settlement boundary

4. RESULT

4.1 Performance of Deep Learning Models

In this study, we evaluated the models using the same test dataset. The models were trained on the Inria Aerial Image dataset or on the traditional village dataset using transfer learning. The two models were measured using the three most common evaluation metrics, including Precision, Dice, and IoU (Intersection over Union), which are defined as:

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Dice = \frac{2TP}{2TP+FP+FN} \quad (3)$$

$$IoU = \frac{TP}{FN+FP+TP} \quad (4)$$

where TP = true positive, the building pixels predicted as buildings
TN = true negative, the ground pixels predicted as ground
FP = false positive, the ground pixels predicted as buildings
FN = false negative, the building pixels predicted as ground

Name	Precision	Dice	IoU
Pre-Trained Model	77.83%	50.25%	33.56%
Transfer Learning Model	88.61%	90.65%	82.9%

Table 1. Performance of two models with the same test dataset.

Based on the evaluation of the metrics in Table 1, the result of the transfer learning model is better. The visualization shows that the transfer learning model is able to accurately segment most of the buildings (Figure 5(c)), while the original model is less effective, with only some of the buildings in the top left being partially segmented (Figure 5(b)).

(a) Remote sensing image (b) Predict image using the pre-trained model (c) Predict image using the transfer learning model

Figure 5. Visualization of recognition results of pre-trained and transfer learning models.

In addition, the model can compute an average of 2.9 images per second with a resolution of 500*1024 when measured in terms of runtime speed. 70 traditional villages were split into a total of 23,730 images and the semantic segmentation took a total of about 2.7 hours (9,880 seconds).

4.2 The Performance of Spatial Feature Extraction

The remote sensing images of 70 national traditional villages were automatically quantified and analyzed using the automatic spatial feature extraction method for traditional villages established in this paper. Based on the quantification metrics (Table 2) in the literature (Xincheng, 2013), the semantic segmented images were calculated using the statistical method

in 3.4, consequently quantification results in Figure 6 were obtained. The whole process took over three hours.

S	λ	Type of village
S≥2	λ<1.5	Cluster trending finger-like settlements
	1.5 ≤ λ < 2	No trending finger-like settlements
	λ ≥ 2	Band trending finger-like settlements
S < 2	λ<1.5	Clustered settlements
	1.5 ≤ λ < 2	Band trending cluster settlements
	λ ≥ 2	Band settlements

Table 2. Quantitative classification of settlement boundary patterns.

The 70 traditional villages counted in this paper are distributed in four provinces (Figure 6). As shown in Figure 6, it can be found that there are more clustered settlements in Hebei Province, more banded settlements in Shanxi Province, and nearly no banded settlements in Jiangxi Province, and more finger-like villages in Zhejiang Province compared to the other three provinces. As this paper only explores ways to automatically quantify village characteristics, the data collected from remote sensing images of traditional villages is still relatively small and cannot fully represent the characteristics of each province, and is only an illustrative example of the methods used in this paper.

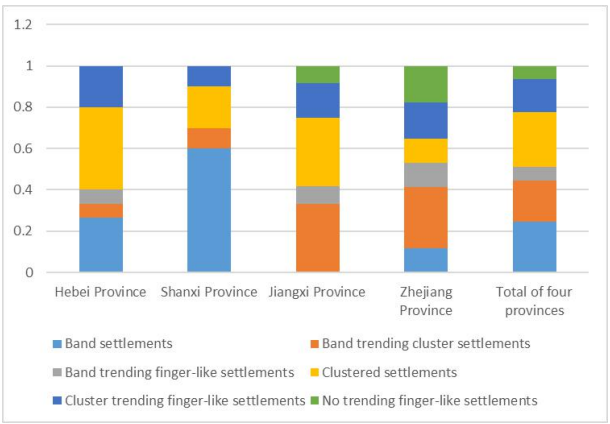


Figure 6. Statistical histogram of the traditional village boundary fractal index.

The traditional villages in Hebei Province collected in this article are mostly defensive fortresses. A fortress refers to a settlement that combines defense and housing, usually using fortress walls to enclose streets, lanes, dwellings and public buildings (Chunmei et al., 2013). Fortified for defensive purposes, the forts are mostly located on high terraces to get a better lookout, which may also cause it to gravitate towards regular shapes such as rectangles, and thus explain the statistical results on village morphology in Hebei Province in this paper to some extent.

5. DISCUSSION

The results of the experiments and index evaluation show that the semantic segmentation method of remote sensing images of traditional villages established in this paper, after transfer learning, can better cope with the needs of segmenting traditional village buildings in different regions of China, and avoids the influence of subjective factors in the process of manually extracting village buildings. The efficiency of processing an average of one image per second is also relatively high, and the task of segmenting remote sensing images of 70 traditional villages can be completed in three hours. The proposed algorithm for automatic extraction of village boundaries based on segmented images is able to automatically extract traditional village boundaries and improve efficiency.

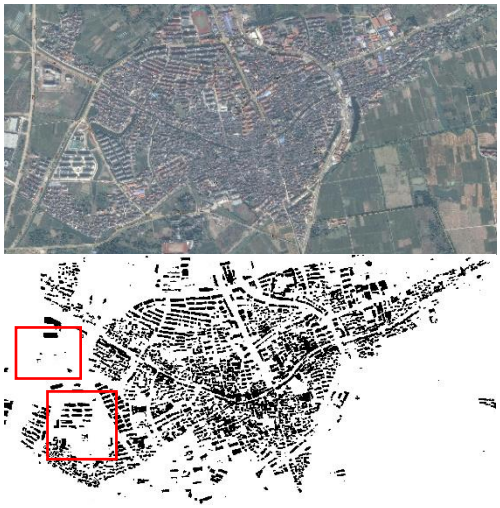
However, there is still a degree of misclassification in this model during the semantic segmentation process. As shown in Figure 8 (red box is the case of missed identification, blue box is the result of tree shading influence), factors such as tree shading may lead to the failure to identify village buildings and may also affect the quantified results.

Figure 8. Misidentification in semantic segmentation.

In addition, this paper aims to introduce advanced algorithms to improve the efficiency of village quantification by using only the smallest range in the calculation of automatic village boundaries based on semantic segmentation results. The results calculated in this step are not precise enough because the method for finding the boundary morphology at macro-, meso- and microscopic scales at 100m, 30m and 7m in the literature (Xincheng, 2013) is not used.

The results of the semantic segmentation can be used to quantify the spatial characteristics of the village, but also to count the architectural changes in the village. For example, the remote sensing images of the town of Fenghui in Guangxi province were segmented with this model for different years, and it is obvious from the segmentation results that there are changes in the two images on the left. If the difference is calculated using the two images, an estimate of the new building area can be made. It is easy to monitor the expansion of buildings in the vicinity of traditional villages using this method.

(a) Remote sensing image and semantic segmentation image of Fenghui Town in 2007



(b) Remote sensing image and semantic segmentation image of Fenghui Town in 2016

Figure 9. Remote sensing image and semantic segmentation image of Fenghui Town in different years.

The semantic segmentation model of traditional villages established in this paper can be applied to remote sensing images of traditional villages in different regions of China to automatically segment village buildings. Automatic quantitative analytical processing from remote sensing images to spatial feature indicators has been implemented. The quantitative results can be provided to government decision makers and designers for assessing and adjusting village design and plan. It can also be used for monitoring tasks based on remote sensing data in the construction of villages and towns.

6. CONCLUSION

In this paper, we proposed an automatic method for quantifying spatial features of traditional villages using deep learning and computer vision algorithms. A semantic segmentation dataset for Chinese traditional villages was built based on remote sensing images, and the OCR-Net model was used to train a model that can automatically segment remote sensing images. The experimental results demonstrate that the village semantic segmentation model trained using transfer learning can guarantee a good performance in the task of segmenting traditional village buildings from remote sensing images. Using algorithms such as edge detection, minimum outer rectangle and clustering, automatic transformation from semantically segmented images to quantitative indicators of village spatial morphology is achieved.

Taking 70 traditional villages as an example, it is proved that this workflow can automate the process of quantifying the spatial patterns of traditional villages, improve the efficiency of the analysis of the village spatial features and reduce the manpower consumption. Moreover, the automated workflow can avoid the errors that may arise in the manual work of segmenting buildings and delineating village boundaries.

The performance of the algorithm can be further improved in the future, or other data can be introduced for comprehensive analysis to enhance the accuracy of prediction of traditional village boundaries. Applying the method of this paper, the spatial quantification statistics of all national traditional villages will be carried out, and a database of the spatial characteristics of all traditional villages will be established to preserve the information on the planar forms of traditional villages.

Based on the statistical results, a quantitative comparison of the spatial characteristics of villages across geographical areas can achieved, and the wisdom of traditional village space creation will be unearthed. It not only provides data to support the research, conservation and renovation of traditional villages, but also provides a reference for the design of a livable community based on the spatial creation techniques of the villages.

REFERENCES

Ali, S., Razavian, Hossein, A., Josephine, S., and Stefan, C., CNN Features off-the-shelf: an Astounding Baseline for Recognition, *2014 IEEE conference on computer vision and pattern recognition workshops*, 10.1109/CVPRW.2014.131.

Alshaikhli, T., Liu, W., and Maruyama, Y., 2019: Automated Method of Road Extraction from Aerial Images Using a Deep Convolutional Neural Network, *Applied Sciences*, 9, 10.3390/app9224825.

Canny, J., 1986: A Computational Approach to Edge Detection, *Ieee T Pattern Anal*, PAMI-8, 679-698, 10.1109/TPAMI.1986.4767851.

Chunmei, Z., Lifeng, T., and Yan, L., 2013: Study on the Fortress Settlement by Typology Methods in Yu County of Hebei Province, *Fujian Architecture & Construction*, 33-35.

Dan, W. and Wei-Dong, S., 2011: A Method of High-resolution Remote Sensing Images Building on Edge Extraction, *IEEE Computer Society*.

Dengxin, D. and Wen, Y., 2011: Satellite Image Classification via Two-Layer Sparse Coding With Biased Image Representation, *IEEE Geoscience & Remote Sensing Letters*, 8, 173-176, 10.1109/LGRS.2010.2055033.

Etten, A. V., 2018: You Only Look Twice: Rapid Multi-Scale Object Detection In Satellite Imagery.

Hao, L. and Zao, L., 2019: Quantitative Study on the Spatial Structure of Rural Settlements in the Surrounding Areas of Hefei, *Urbanism and Architecture*, 016, 17-18.

Liu, X., Liang, X., Li, X., Xu, X., Ou, J., Chen, Y., Li, S., Wang, S., and Pei, F., 2017: A future land use simulation model (FLUS) for simulating multiple land use scenarios by coupling human and natural effects, *Landsc. Urban Plan.*, 168, 94-116, 10.1016/j.landurbplan.2017.09.019.

Long, J., Shelhamer, E., and Darrell, T., 2015: Fully Convolutional Networks for Semantic Segmentation, *Ieee T Pattern Anal*, 39, 640-651, 10.1109/CVPR.2015.7298965.

Maggiori, E., Tarabalka, Y., Charpiat, G., and Alliez, P., Can Semantic Labeling Methods Generalize to Any City? The Inria Aerial Image Labeling Benchmark, *Igarss IEEE International Geoscience & Remote Sensing Symposium*, 10.1109/IGARSS.2017.8127684.

Mnih, V., 2013: Machine Learning for Aerial Image Labeling, University of Toronto (Canada).

Qian, S., Mengxi, L., Xiaoping, L., Penghua, L., Pengyuan, Z., Jinxing, Y., and Xia, L., 2019: Domain Adaption for Fine-Grained Urban Village Extraction From Satellite Images, *Ieee Geosci Remote S*, 17, 1-5, 10.1109/LGRS.2019.2947473.

Ronneberger, O., Fischer, P., and Brox, T., 2015: U-Net: Convolutional Networks for Biomedical Image Segmentation, *Springer, Cham*, 10.1007/978-3-662-54345-0_3.

Weiguo, X. and Mengjia, N., 2021: Quantitive Study on the Spatial Characteristics of the Settlements: Take 4 Villages in Xiahuayuan District, Zhangjiakou City as Examples, *Contemporary Architecture*, 28-31.

Xin, H., Liangpei, Z., and Tingting, Z., 2013: Building Change Detection From Multitemporal High-Resolution Remotely Sensed Images Based on a Morphological Building Index, *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*, 7, 105-115.

Xincheng, P., 2013: *A quantitative approach to the planar form of traditional rural settlements*, Southeast University Press

Xincheng, P., Yingjia, W., and Qian, H., 2020: Analysis of Quantitative Methods of Obtaining the Boundary Shape of Rural Settlements, *Architecture & Culture*, 189-191.

Ya-Juan, L. I., Hu, Y. U., Chen, T., Jing, H. U., and Cui, H. Y., 2016: Livelihood changes and evolution of upland ethnic communities driven by tourism: a case study in Guizhou Province, southwest China, *J Mt Sci-Engl*, 13, 1313-1332.

Yuan, Y., Chen, X., and Wang, J., 2020: *Object-Contextual Representations for Semantic Segmentation*, European Conference on Computer Vision, 10.1007/978-3-030-58539-6_11.

Yun, W., 2009: *The concept of space in the structure of traditional settlements*, China Architecture & Building Press

Zhe, L., Su, S., Chuanqi, Z., Xinxin, T., Yinxin, Z., and Yan, L., 2019: The Optimized Access to the Digital Museum of Traditional Chinese Villages Exploring and Quantifying the Wisdom of Traditional Villages by Three-Dimensional Computation, *Architectural Journal*, No.605, 80-86.