Research Paper

Semantic Segmentation of Street View Map on the Cityscape of Heritage in the Historical and Cultural City

Peilun LI, School of Architecture, Harbin Institute of Technology; Key Laboratory of Cold Region Urban and Rural Human Settlement Environment Science and Technology, Ministry of Industry and Information Technology, China

Zhiqing ZHAO, School of Architecture, Harbin Institute of Technology; Key Laboratory of Cold Region Urban and Rural Human Settlement Environment Science and Technology, Ministry of Industry and Information Technology, China

Yuling CHEN, School of Architecture, Harbin Institute of Technology; Key Laboratory of Cold Region Urban and Rural Human Settlement Environment Science and Technology, Ministry of Industry and Information Technology, China

Abstract

The expression and perception of cityscape has significance influence on the preservation and regeneration of heritage. However, due to the urban sprawl in urbanizing, a large number of commercial exploitation have resulted in a series of obstacles on the expression and perception of heritage features. In order to study the importance of the expression and perception of the heritage features on the cityscape, this paper selected a typical historical and cultural city Harbin in China as the research object. In the current work, the large-scale investigation of cityscape was analyzed on the basis of street view map auditing from a human point of view. The relationship of cityscape between the features of heritage and cityscape were investigated using the semantic segmentation of machine vision, and Harbin was taken as a case study to examine whether variations exist between different areas. Results showed a significant coupling in the current city between cityscape and features of heritage. Therefore, according to the results of the relevant analysis, the paper confirmed the key indicators of the perception and expression of heritage, and advances some urban planning strategies to avoid the negative effects that urban sprawl brings to cityscape on heritage. The results of this work will provide a comprehensive understanding of the preservation and regeneration of the heritage.

Keywords

Semantic Segmentation, Street View Map, Cityscape of Heritage, Historical and Cultural City

1. Introduction

In China, rapid urbanization has brought about large-scale changes in the style of urban streets. This large-scale change has formed a diversified pattern of cityscape in China, and the cities rated as historical and cultural cities have become representatives of the diversified trend to a certain extent. As an important basis for election historic and cultural cities, streets are also an important spatial carrier of historical expression and cultural perception. Studying cityscape and spatial characteristics from streets, discovering influencing factors and dynamic mechanisms has been a commonly used approach. At the same time positive cityscape growth is related to a variety of heritage values, such as social, cultural and ecological values. Due to the reduction of cityscape conflicts, positive cityscape environment can also generate certain cultural space, as well as social and economic benefits from value recognition. However, the rapid expansion in the past has caused the phenomenon of cityscape disorder in historical and cultural cities, especially the conflict with the original cityscape of the heritage. The invasion of commercial culture, the increase of street capacity and the reproduction of a single model have hindered the cultural expression and perception of heritage.

It has been shown that the built environment of the streets is related to the expression and perception of the historic heritage landscape. A coordinated built environment can promote the transmission of information about historic heritage. In addition to basic transportation functions, streets are also important spaces for crowds to perceive information and generate regional images. Many street elements also contribute to the transmission of historical and cultural information of heritage, and promote the expression of information in the form of cityscape of historic heritage.

Most of the existing studies focus on the streets where heritage is concentrated, while ignoring the cityscape around the heritage and at the city scale. In addition, there are fewer studies on the differences in the cityscape of heritage core streets, heritage surrounding streets and regional streets in historical and cultural cities. In this paper, we take Harbin city, a historical and cultural city in China, as an example, and use semantic segmentation of deep learning to carry out large-scale streetscape element identification and statistics, explore the street style features and influencing factors of different types of historical and cultural districts, and combine ELO dichotomous evaluation to derive the influencing factors that need to be optimized urgently, and put forward relevant optimization suggestions.

2. Review

In early studies, researchers have recognized that street views are well suited for assessing the characteristics of the built environment (Kelly, Wilson, Baker, Miller, & Schootman, 2013) and have also been embraced by numerous fields. Over the years, SVI has been used for real estate valuation (Law, Page, & Russell, 2019), population studies (Gebru et al., 2017), data collection, pedestrian counts (Cheng, Yin,Wang & Shao, 2015), understanding crime (McKee et al., 2017),analyzing accessibility (Hara, Le, & Froehlich, 2013), GIS analytics (Biljecki & Ito, 2021) and mapping infrastructure deficiencies (Chang et al., 2017). In the study of historical and cultural heritage, it has been discussed in depth by scholars in the areas of building façade (Shalunts et al., 2011), heritage interest elements (Llamas et al. 2017), architectural style (Xu et al. 2014), and streetscape assessment and marking (Kang et al. 2018). In summary, the current assessment of heritage is mainly focused on the excavation of elements of heritage itself, and lacks large-scale measurement and perception of heritage and concentrated areas. The purpose of this paper is to conduct a large-scale assessment of heritage core areas using streetscape images, and to explore the strength of historical and cultural expressions of heritage and the elements of its involved landscape in the streetscape perspective based on the strength of subjective perception of heritage, so as to provide guidance for heritage optimization in historical and cultural cities and promote the cultural heritage and connotation of cities.

3. Material

3.1. Study area

The built environment of a street is inseparable from the city's history and culture (Wang, Zhao & Zhang, 2019). In China, the declared historical and cultural city should also have more than two historical and cultural streets within its protection area, and requires particularly richly preserved heritage, concentrated patches of historical buildings, and preserved traditional patterns and historical styles. Historic and cultural area are the starting point of a city's development, recording the spatial development of the city's veins and historical deposits, as well as an important part of the city's spatial structure and function, playing an important role in the growth, development and prosperity of the city. The built environment of the street is an important basic part of the historical and cultural district, as well as an important carrier of historical heritage, with important historical significance and research value.

In this study, three historic districts of Harbin, a famous historical and cultural city, are selected as research objects. Harbin was formed as an important engineering hub for the construction of the Middle East Railway by Russia in 1898. At the early stage of urban development, the spatial structure of the city was profoundly influenced by Russian urban planning ideas at that time (Zhang et al. 2018). The historical streets of Harbin are characterized by districts, which are mainly divided into three parts. The Nangang District, Central Avenue, and the Jingyu Street represent typical commercial-type districts, combined commercial and residential districts, and administrative mixed districts respectively in the process of historical development, and the development orientation of the three districts has shaped the development basis of the city's history and culture (Figure 1).

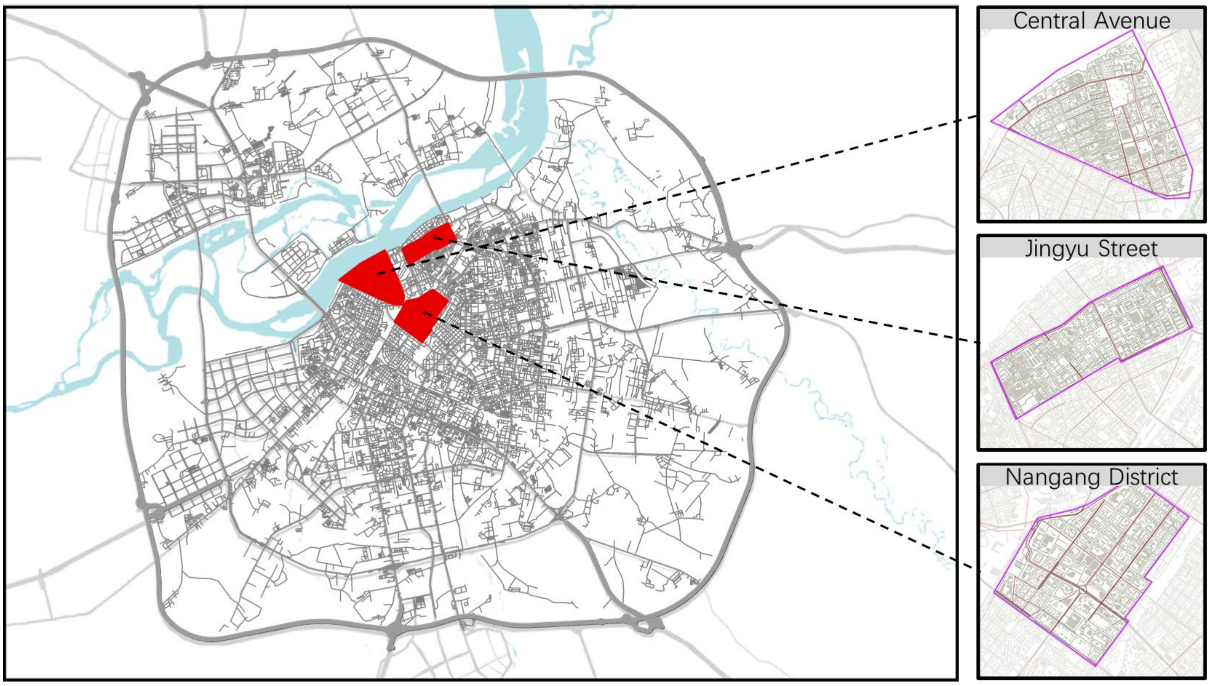


Figure 1. Harbin Street and Heritage Core Area. Source: Author.

3.2. Data Collection

The research data collected from Baidu map lbsyun platform, apply for developer status for the data call of street view image in Harbin, use the code with python 3.7+, set the parameters such as image size, latitude and longitude coordinates, send HTTP request to access Baidu map panoramic static map service, and get the retrieved data returned in jpg format (Table 1). The retrieved coordinate points are obtained on the basis of the road line data in Harbin, and the Harbin roads are densified at an interval of 30m, and the latitude and longitude of the fold point of the densified line segment is taken as the collection parameter of the street view image; the image size is set to 1024*512 as the parameter. Instead of using images for each basic direction, we queried the cubic panorama (including metadata), which provides a more comprehensive description of the cityscape (Helbich et al. 2021). Such parameter settings have the same camera equipment, the same hardware parameters and shooting angles to ensure the objectivity of the subsequent classification and detection of the architectural landscape content of the cityscape image. We collected 247,936 images and performed cleaning work on the image data in order to avoid the interference of winter snow, ice cover and tree trunks on the recognition of street view, and finally 216,831 images of street scenes were selected for the study. (Figure 2).

Table 1. Baidu Map Web Panorama Service API Detailed Parameters. Source: Baidu Map Web

Parameter	Profile
ak	User's access key. Only browser-side ak and Android/IOS SDK's ak are supported, server-side ak does not support sn verification method
mcode	Security code. If ak for Android/IOS SDK, this parameter is required.
width	image width, in the range [10,1024]
height	height of the image, in the range [10,512]
location	Coordinates of the panorama position point. Coordinate format: lng<longitude>, lat<latitude>
coordtype	Coordinate type of the panoramic location point, currently supports bd09ll (Baidu coordinates), wgs84ll (GPS coordinates) and gcj02 (google, gaode, soso coordinates)
poiid	POI's id, this attribute is usually obtained through place api interface, poiid sets the display scene of panorama together with panoid and location, the priority is: poiid>panoid>location. where the panoramic viewpoint obtained according to poiid is the best.
panoid	Panorama id, panoid sets the display scene of panorama together with poiid and location, the priority is: poiid>panoid>location.
heading	Horizontal view, range [0,360]
pitch	Vertical view angle, range [0,90]
fov	Horizontal range, range [10,360], fov=360 to display the whole panorama

Figure 2. Harbin SVI Collection Point. Source: Author.

3.3. Assessment of street view image

3.3.1 deep learning for image segmentation

Predictive statistics and visualization by visual deep learning models are used to determine the intensity magnitude of the historical information expressed under each street feature. In order to perform a comparative audit of streets through images, we use a semantic segmentation method of deep learning to segment the semantic attributes of images and count the proportion of pixels in the image for each semantic information. Six key street features, namely buildings, roads, sky, greenery, sidewalks, and cars, are extracted from the street image to obtain quantitative measures of each spatial element in the street view (Ye et al., 2019). Recent advances in deep learning have shown that fully convolutional networks (FCN) can predict the semantic attributes of each pixel in an image to produce natural target-level segmentation results. (Long et al., 2015; Badrinarayanan et al., 2017). FCN divides the street view image into multiple sub-scenes, each buildings, roads, trees, or other objects (Figure 3). Together with a Softmax layer-based (Shelhamer et al., 2014; Zheng et al., 2016) pixel-by-pixel loss calculation, the area ratio of each visual element in the image is generated by calculating the number of pixels in each segmentation mask (Yao et al., 2019).

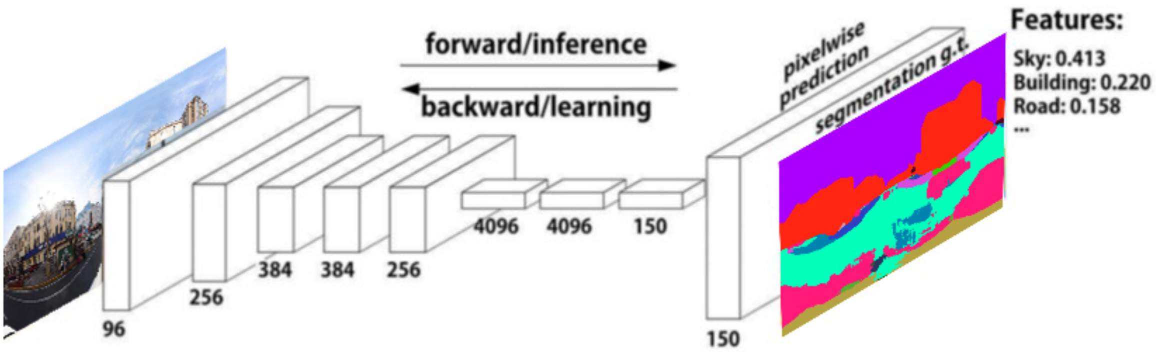


Figure 3. The Input and Output of the Fully Convolutional Network (FCN). Source: Semantic segmentation net by Yao et al., 2019.

3.3.2 ELO Dichotomous Sorting

The street images were screened for historical and cultural features by dichotomous sorting of ELO. Streetscape images in three historic areas of Harbin city were selected as the basis for the street view paradigm selection of the target cityscape. The scorer for two-by-two comparison was written by python 3.7, and the comparison results were quantified and integrated into a score ranking one by one using ELO rating algorithm (Figure 4). ELO's dichotomous sorting method has been extensively validated in terms of image and human perception, and is able to weaken the judgmental confusion in the judging process in terms of perceptions and intentions, iterating the results step by step in a two-choice manner (Zhang, Zhou, Liu. et al., 2018). Planning designers with five years of experience were selected to rank the most historic street in the reserve, and when the results were stabilized through several iterations, the final scores were obtained for all sample photos.

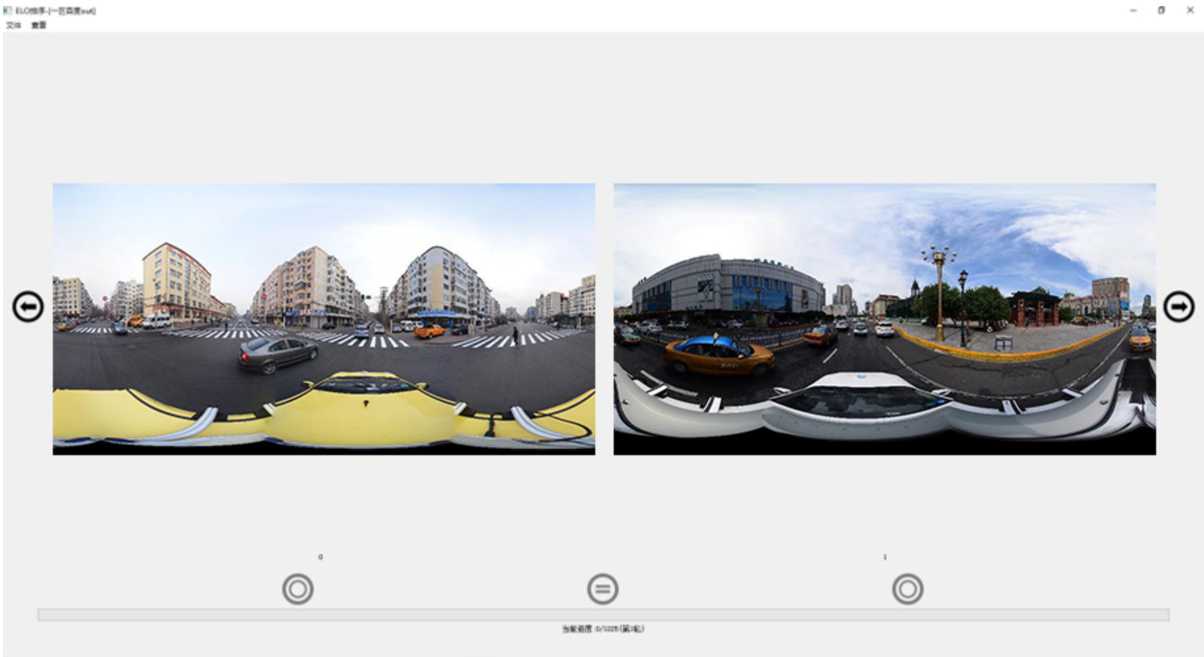


Figure 4. ELO Dichotomous Sorting. Source: Author.

3.3.3 Object detection and image gradient heat map

Computer vision is somewhat related to biological vision in that the feature areas in the images are the main basis for making judgments. Model prediction visualization is performed by pixel gradient heat to determine the intensity magnitude of heritage to express historical information under each street feature. To reduce the interference brought by street features and to keep the street features consistent for deep learning of target detection. Grad-CAM can be used to locate the areas associated with labels in the image, and to some extent can explain the CNN model, using pixel gradient heat to reflect the intensity of features in the pixel population during prediction. The images selected by ELO classification are trained with Pytorch for deep learning of target detection, and the labelImg is used for box marking of heritage in street view images. Based on more cutting-edge deep learning networks for image object detection, the comparison algorithms include Faster-Rcnn, RetinaNet, SSD, Yolov3-Spp and other methods. In order to obtain larger accuracy, Faster-Rcnn was used for training the detection network, and ResNeXt-101 was used as the backbone to determine the pre-training weights.

4. Results

4.1. SVI image segmentation and comparison

We collected and processed 216,831 images of street scenes within the Harbin Ring Road. The images were taken in the summer of 2020, with good vegetation and no interference from snow and ice. Semantic segmentation was used to distinguish the images into categories and to count the proportion of the object size, such as buildings, roads, sky, plants, sidewalks, and cars. The recognition results are summarized in the form of folded points on the statistical road line segments, and the mean and standard deviation of the road segments are counted separately, so as to judge the current situation of streets in Harbin.

In terms of the mean values of the elements, the old center of Harbin has a larger proportion of building and car elements in the cityscape compared with the surrounding new area, which may be related to the denser urban road network and traffic volume in the old area; the road and sky elements have a larger proportion in the streetscape of the new area, which is mainly developed and constructed with wider streets and building setbacks; the elements of greenery and sidewalks have a relatively balanced and relatively close to the vicinity of urban parks (Figure 5). In terms of the standard deviation of each element, the building and sky elements show a relatively discrete phenomenon and are mostly distributed in the old area, because in the process of renewal of the old area, high-rise buildings and partially widened roads cause a large change in the streetscape of the road section; the standard deviation of the cityscape of roads and cars shows a large dispersion, which is related to the density of the road network and the width of the roads in different construction periods; the standard deviation of plants and sidewalks is relatively stable, which is related to the density of the road network and the width of the roads in different construction periods; the standard deviation of plants and sidewalks is relatively The variance is relatively stable, which is related to the greening rate of road greening and site construction (Figure 6).

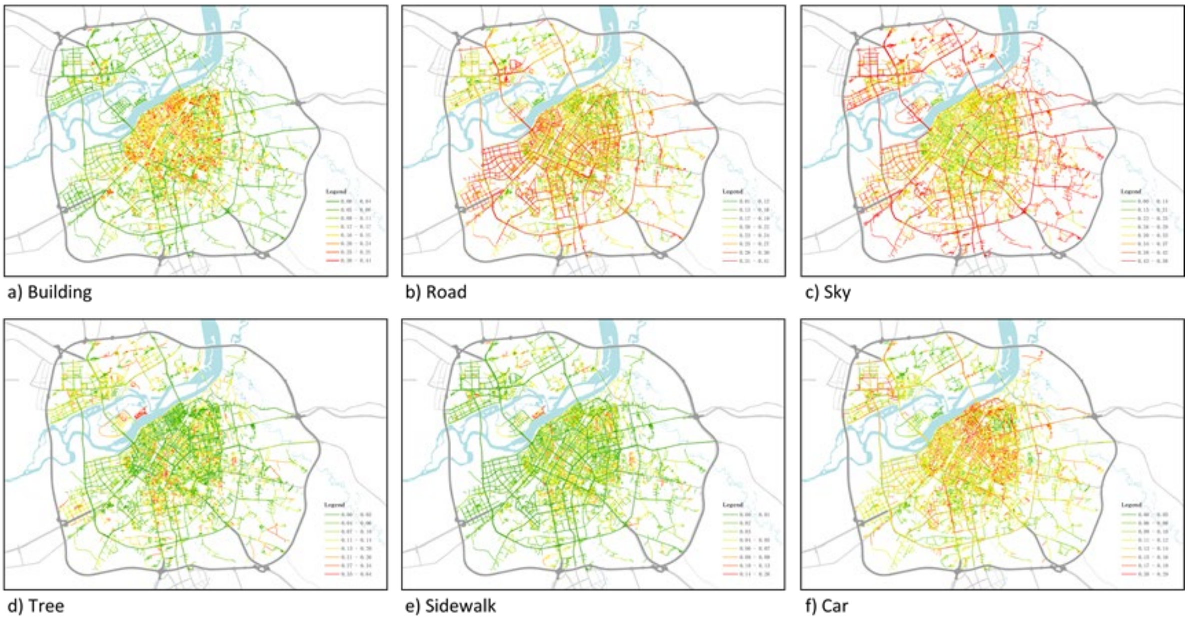


Figure 5. Average Value of Harbin Street Elements. Source: Author.

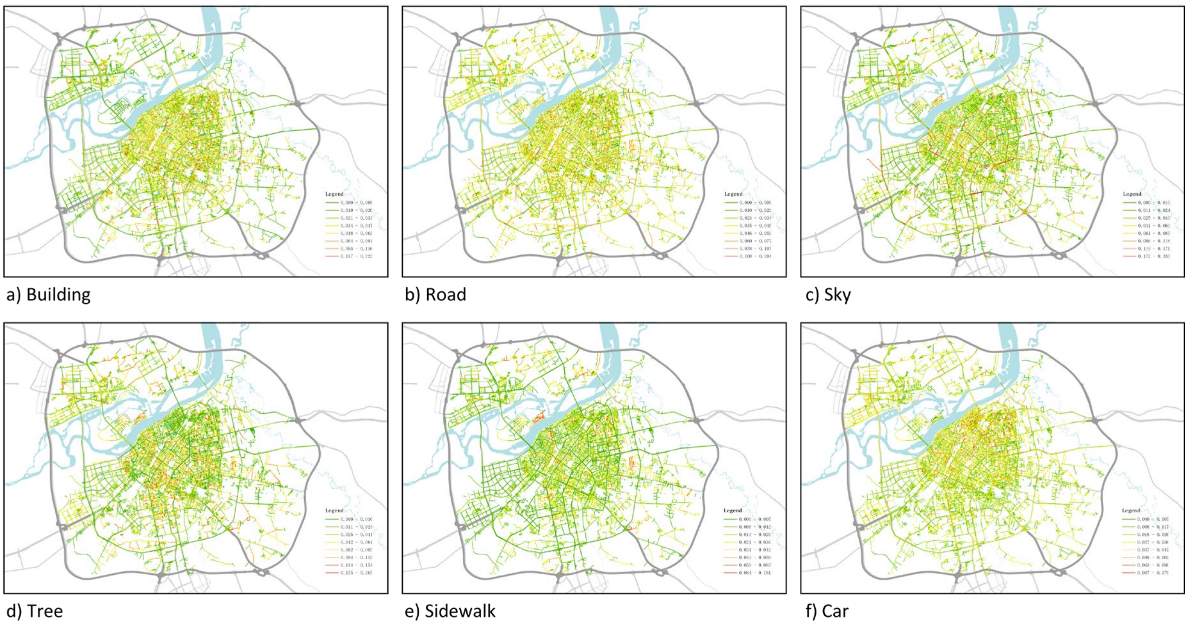
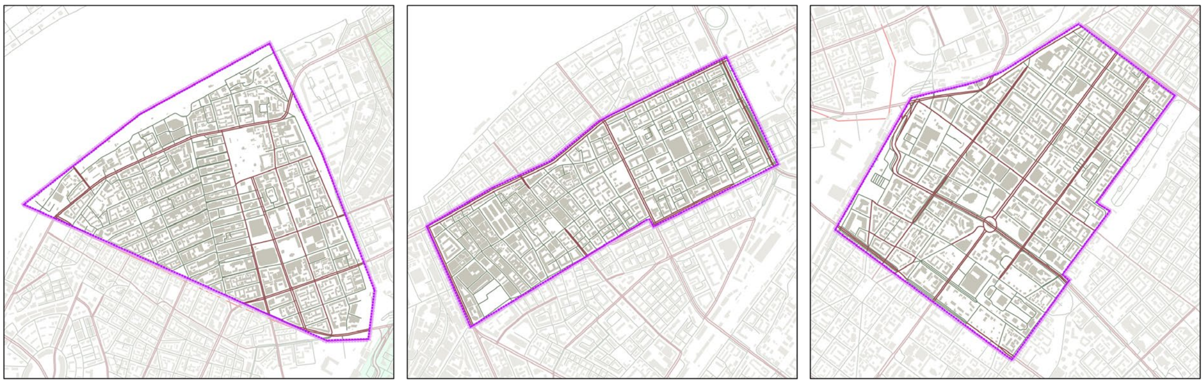


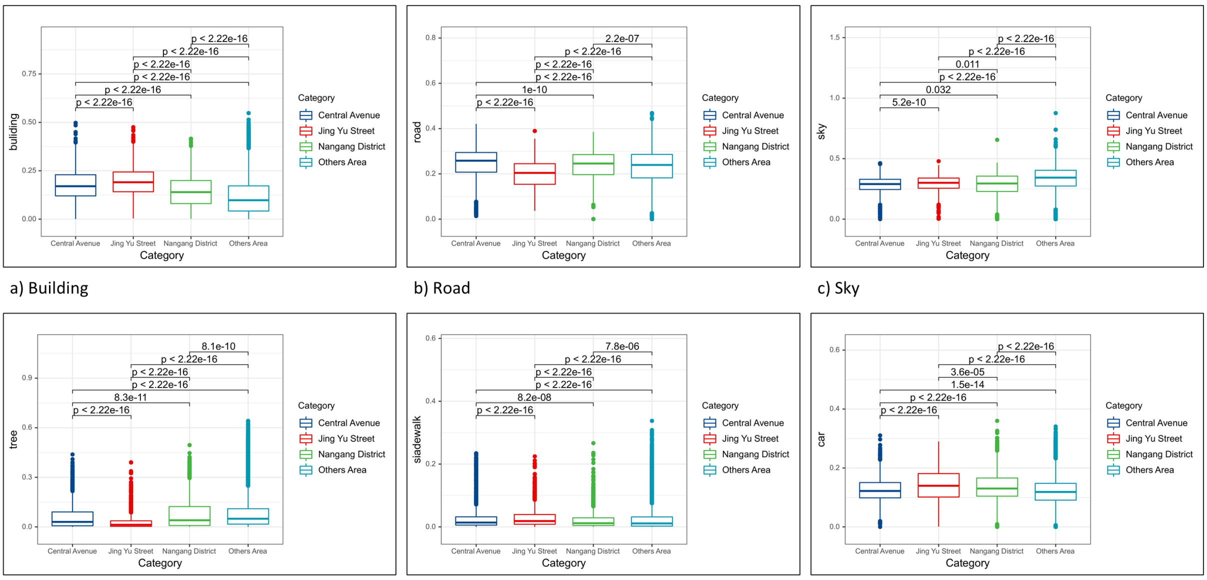
Figure 6. Standard Deviation of Harbin Street Elements. Source: Author.

The statistical analysis of the Heritage Core Preservation Area includes three parts of Nangang District, Central Avenue, and Jingyu Street, where the scale of buildings and streets still maintains the traditional pattern, although they have now undergone some functional replacement in response to new development needs (Figure 7). We compared the mean and standard deviation using the heritage core area with the other areas (Figure 8). In terms of the proportion of architectural elements in the cityscape, all three areas are higher than the rest of the city, which may be related to the higher street H/D in the heritage core area, but to a certain extent can better express the heritage information; in terms of road elements, the cityscape proportion of road elements in Central Avenue and Nangang District is more similar to that of the non-heritage area, but the proportion of road elements in Jingyu Street is relatively smaller; in terms of sky elements are lower in the core heritage area than in other areas, again related to the street height to width ratio, and in contrast to the building elements, the building spacing is a greater limiting factor; trees, sidewalks and cars all occupy a smaller proportion in them, and although there are small differences they are clearly insignificant.



a) Central Avenue b) Jingyu Street c) Nangang District

Figure 7. The Extent of the Heritage Core Area and the Street network. Source: Author.



a) Building b) Road c) Sky

d) Tree e) Sidewalk f) Car

Figure 8. Box-plot of the Elements within the Heritage Core Area. Source: Author.

4.2. ELO image sorting and object detection

Based on a audit work of the heritage core areas conducted by 30 planning designers with five years of experience, a ranking of streetscape images with a strong sense of history and culture in the core areas was obtained. 30 participants will compare two streetscape images from the three core areas and select the top 500 images as the training set for target detection using the ELO ranking. The participants use labelImg to label the areas that they think are strongly expressed according to their choice and use them as training labels. By aggregating the images and labels of 30 planning designers and integrating the labels of duplicate images and the groundtruth box, a total of 2317 streetscape images were obtained to represent the scenes of expression of the heritage core area (Figure 9). In the model training, 1622 images were randomly selected from the dataset for training, 695 images were selected for testing, and 30% of the images from the training sample were used as validation data to monitor the training status of the network. the training adopted a batch size of 8, trained with an initial learning rate of 0.01 and fixed 10 step size and All training, testing and validation were performed with PyTorch on 4 Nvidia RTX 3090 24GB GPUs. Due to the small image database, a migration learning approach was used to train the convolutional layer of the network, which achieved high accuracy in the first 50 epochs and showed continuous fluctuations in the next training (Figure 9).

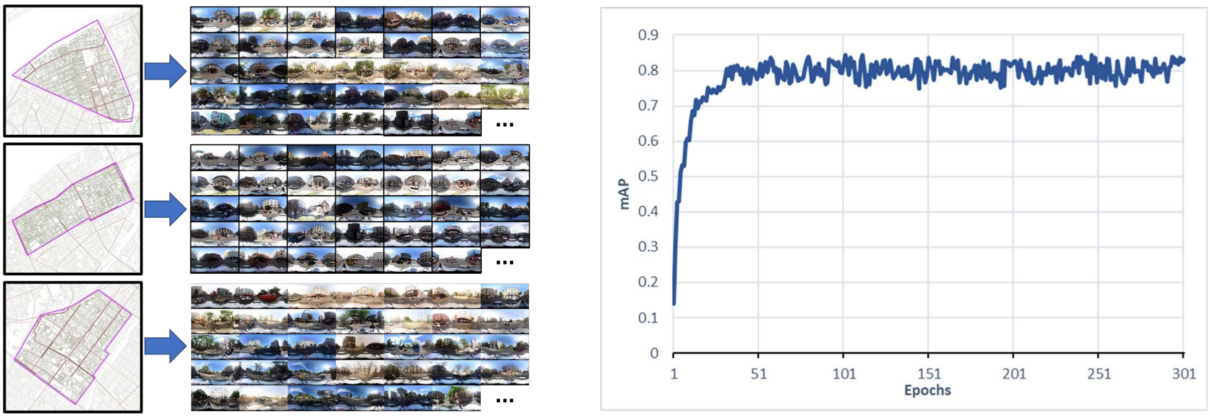


Figure 9. Example of Street View Selected by ELO (left) and Deep learning record for object detection(right). Source: Author.

4.3. Image gradient heat analysis

Building on the models of target detection, understanding the deep learning networks provides a visual interpretation that makes them easier to understand. Gradient-weighted Class Activation Mapping (Grad-CAM) generates a rough localization map through layers to highlight important regions in the image (Selvaraju et al., 2017; Chattopadhay et al., 2018). The Image gradient heat reveals that the areas of architectural heritage with a stronger sense of history and culture in the street are concentrated near the windows and the top of the building facade (Figure 10). These areas with large feature gradients cross at least one level of height of the street space, and therefore, these interfaces play an important role in the street, especially in the way heritage conveys historical and cultural information.

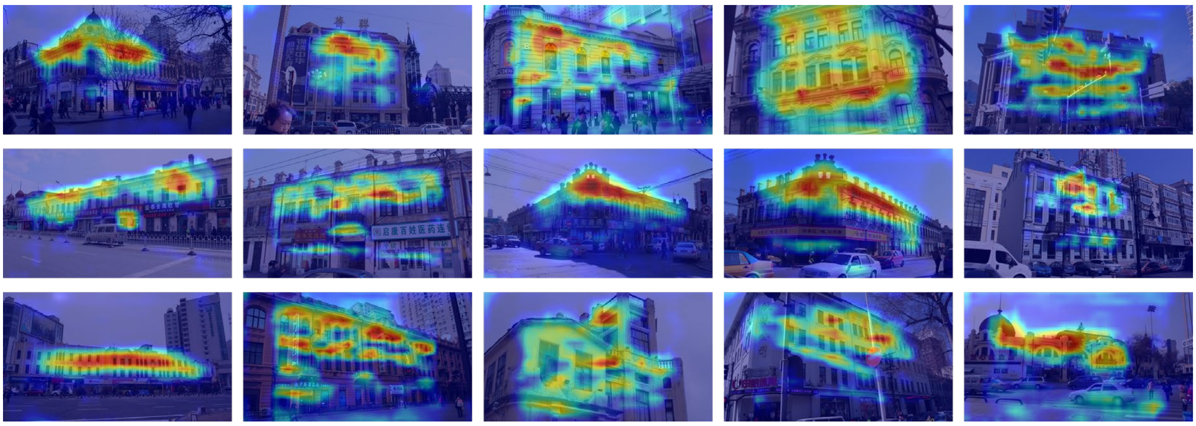


Figure 10. Example of Gradient-weighted Class Activation Mapping of images. Source: Author.

5. Discussion

5.1. Reduce facade shading

The occlusion of the building façade can cause a significant loss of heritage character, especially windows and higher parts of the façade. Through communication with the designers involved in the audit work, it was found that a large number of billboards in the streetscape interferes more strongly with the heritage appearance and influences their judgment. The heritage in use promotes spontaneous heritage maintenance. Considering the realistic function of the heritage, billboards can be installed on the first to second floors to ensure the promotion of the functions of shops along the street, but the premise is to ensure that it is in harmony with the heritage. However, above the third floor, it is recommended to remove the billboards to give a complete sight space to the characteristic elements of the heritage style to convey the historical and cultural information.

5.2. Upgrading of pedestrian walkways

The perception of cityscape from the human-scale perspective is the strongest. The street view in the study is mainly collected by professional research cars, whose paths are mainly concentrated on the motorway. Therefore, to a certain extent, it is difficult to reflect the observation state of the human-scale perspective. In the future, when wearable devices gradually become popular, the cityscape from a human perspective will gradually enhance the field. Promoting the optimization of sidewalks not only pulls in the distance between people and heritage, but also enhances the expression dimension of cityscape information, and strengthens the affinity of heritage rather than resembling artworks in museums. The optimization of sidewalks lies not only in the paving and street furniture, but also in the canyon space between the greenery and buildings along the street. This space can enhance the height of heritage in the human perspective, and moreover, it can generate a kind of looking-up perspective to enhance people's recognition of the valuable value of heritage.

5.3. Increase street greening

According to the analysis results, it is clear that most of the characteristic areas of the heritage landscape are concentrated in the higher parts of the façade. However, the spatial interface of the lower parts of the heritage landscape is relatively chaotic, which limits the communication of historical and cultural information. It is more environmentally friendly to be screened by vegetation. The green view of the street is important in many studies, in the fields of ecology, environment and health. More trees are rare and valuable for the streets and even the city, which represents that people, city and nature live in harmony, as does heritage.

6. Conclusion

We summarized the streetscape comparison between Harbin city and the heritage core area, and visualized the areas and features intensity of historical and cultural landscapes within the heritage core area in the streetscape images. The results show that the heritage core area differs significantly from the new urban area and is more influenced by factors such as street width-to-height ratio and street scale. Moreover, the specific expression of historic landscape information in the heritage core area relies primarily on areas with higher windows and facades. Although these findings are relevant to the field of architectural history and there are some deeper patterns that have not been verified, a large-scale assessment can observe the characteristics of cityscape information at the urban scale and can provide some specific grips for managers and designers.

Funding: National Natural Science Foundation of China—Study on the Characteristic Townscape Planning Paradigm within the Zone of Chinese Eastern Railway based on the Approaches of UAV Remote Sensing and AI Deep Learning (51878205).

7. References

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017) 'Segnet: A deep convolutional encoder-decoder architecture for image segmentation', *IEEE transactions on pattern analysis and machine intelligence*, 39(12), pp. 2481-2495.

Biljecki, F., & Ito, K. (2021) 'Street view imagery in urban analytics and GIS: A review', *Landscape and Urban Planning*, 215, p. 104217.

Chang, S., Wang, Z., Mao, D., Guan, K., Jia, M., & Chen, C. (2020) 'Mapping the essential urban land use in changchun by applying random forest and multi-source geospatial data', *Remote Sensing*, 12, p. 2488.

Chattopadhay, A., Sarkar, A., Howlader, P., & Balasubramanian, V. N. (2018) 'Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks', *In 2018 IEEE winter conference on applications of computer vision,* pp. 839-847.

Gebru, T., Krause, J., Wang, Y., Chen, D., Deng, J., Aiden, E. L., & Fei-Fei, L. (2017) 'Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States', *Proceedings of the National Academy of Sciences*, 114(50), pp. 13108-13113.

Hara, K., Le, V., & Froehlich, J. (2013) 'Combining crowdsourcing and google street view to identify street-level accessibility problems', *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 631–640.

Helbich, M., Poppe, R., Oberski, D., van Emmichoven, M. Z., & Schram, R. (2021) 'Can't see the wood for the trees? An assessment of street view-and satellite-derived greenness measures in relation to mental health', *Landscape and Urban Planning*, 214, p. 104181.

Kang, J., Körner, M., Wang, Y., Taubenböck, H., & Zhu, X. X. (2018) 'Building instance classification using street view images', *ISPRS journal of photogrammetry and remote sensing*, 145, pp. 44-59.

Kelly, C. M., Wilson, J. S., Baker, E. A., Miller, D. K., & Schootman, M. (2013) 'Using Google Street View to audit the built environment: inter-rater reliability results', *Annals of Behavioral Medicine*, 45(s1), pp. 108-112.

Law, S., Paige, B., & Russell, C. (2019) 'Take a look around: using street view and satellite images to estimate house prices', *ACM Transactions on Intelligent Systems and Technology*, 10(5), pp. 1-19.

Llamas, J., M Lerones, P., Medina, R., Zalama, E., & Gómez-García-Bermejo, J. (2017) 'Classification of architectural heritage images using deep learning techniques', *Applied Sciences*, 7(10), p. 992.

Long, J., Shelhamer, E., &Darrell, T., (2015) 'Fully convolutional networks for semantic segmentation', *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440.

McKee, P., Erickson, D. J., Toomey, T., Nelson, T., Less, E. L., Joshi, S., & Jones-Webb, R. (2017) 'The impact of single-container malt liquor sales restrictions on urban crime', *Journal of Urban Health*, 94, pp. 289–300.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017) 'Grad-cam: Visual explanations from deep networks via gradient-based localization', *In Proceedings of the IEEE international conference on computer vision,* pp. 618-626.

Shelhamer, E., Long, J., & Darrell, T. (2016) 'Fully Convolutional Networks for Semantic Segmentation', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4), pp. 640-651.

Wang, Q., Zhao, Z., &Zhang, B. (2019) 'Research on Capacity Control of Historic Urban Area in Harbin Based on Visual Landscape Analysis', Chinese Landscape Architecture, 35(02), pp. 59-63.

Xu, Z., Tao, D., Zhang, Y., Wu, J., & Tsoi, A. C. (2014) 'Architectural style classification using multinomial latent logistic regression', *In European Conference on Computer Vision*, pp. 600-615.

Yao, Y., Liang, Z., Yuan, Z., Liu, P., Bie, Y., Zhang, J., Wang, R., Wang J., & Guan, Q. (2019) 'A human-machine adversarial scoring framework for urban perception assessment using street-view images', *International Journal of Geographical Information Science*, 33(12), pp. 2363-2384.

Ye, Y., Zhang, Z., Zhang, X., & Zeng, W. (2019) 'Human-scale quality on streets: a large-scale and efficient analytical approach based on street view images and new urban analytical tools', *Urban Planning International*, 34(01), pp. 18-27.

Yin, L., Cheng, Q., Wang, Z., & Shao, Z. (2015) ''Big data' for pedestrian volume: Exploring the use of Google Street View images for pedestrian counts'. *Applied Geography*, 63, pp. 337–345.

Zhang, F., Zhou, B., Liu, L., Liu, Y., Fung, H. H., Lin, H., & Ratti, C. (2018) 'Measuring human perceptions of a large-scale urban region using machine learning', *Landscape and Urban Planning*, 180, pp. 148-160.

Zhang, B., Zhao, Z., Li, P., Wang, Q., & Zhang, X. (2018) 'Study on the Earlier Town's Planning under the Application of Spatial Syntax: Taking the Secondary Station-Located Towns as an Example', *Urban Development Studies*, 25(10), pp. 128-133.

Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., & Torr, P. H. (2015) 'Conditional random fields as recurrent neural networks', *In Proceedings of the IEEE international conference on computer vision,* pp. 1529-1537.