Sensing Urban Space: from Street View Recognition, Event Inference to Understand Urban Behavior

Fan Zhang¹, Hui Lin²*

¹The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong
²The Chinese University of Hong Kong, Shenzhen Research Institute, Shenzhen, Guangdong Province, China

*Corresponding author, e-mail: huilin@cuhk.edu.hk

Keywords: Street Image Recognition; Event Inference; Deep Learning; Urban Design

Objective: The proposed study aims to build a theoretical and technical framework, which mainly include street view image feature extraction and scene recognition, dynamic event inference from time-series street image, in order to provide insight into understanding the physical and social structure of inherently complex and dynamic urban space from the perspective of urban spatial-temporal heterogeneity, further to facilitate spatial knowledge discovery in different scales of urban space. Specific objectives of the research are as below:

1) Convolutional Neural Networks (CNNs) based street view image training and feature extraction for scenes semantic classification and recognition: This study introduces a workflow to realize street view image segmentation, feature extraction and scenes semantic classification by combining technologies in Computer Vision and Deep Learning network. A series of well-trained CNNs will be used to discover, segment and annotate static targets such as roads, buildings, trees as well as dynamic features such as pedestrians, vehicles, time of day, weather.

2) A model to describe the heterogeneity of urban physical space: This study provides a quantitative method to build a dimensional space, in order to measure differences between diverse street scenes and landscapes with different urban area. All the features like road grades, buildings styles, green areas will be taken into account as dimensions to represent urban similarity or peculiarity.

3) A feasible method for context-awareness and event inference based on video cameras: The methodology will serve for the quantitative assessment of the dynamic urban behavior.

Background: Urban space are of great heterogeneous and often unequal in regard to the environmental quality of the neighborhoods, the beauty of their architecture, and the liveliness of their streets, among other evaluative dimensions. However, our ability to understand the effect of a city’s built environment has been limited by the lack of quantitative data on urban perception. Until recently, as the development of deep learning network and computer vision, which provide a powerful tool to simulate and mining inherent structure and knowledge inside the massive dataset, it is possible to deeply understand the dynamic change of urban physical structure and social behavior, by training CNNs with huge number of street view images. It gives urban designers the chance to seek quantitative connection between spatial variables and its inner latent feathers.

Methodology: The methodology of this research comprises the following three steps: (1) By leveraging SegNet, which is a deep convolutional encoder-decoder architecture for semantic pixel labelling model, to realize street view image segmentation, feature extraction and scenes semantic classification: (2) through web-cameras as pervasive sensing tool to support an semantic and computational process to the visual quality assessment in urban spaces through quantitative spatial variables; (3) incorporating urban dimensional space model and time-series street view images to detect and infer the events; (4) identifying the quantitative interdependency between the spatial variables of the urban
environment and the urban spatial-temporal heterogeneity to support a close-to-life urban design process.

The methodology is applied in and evaluated by two case studies oriented to the typical urban spaces, one residential area and another open-square space. The results of this project lead to a better understanding of urban spatial cognition and urban spatial-temporal heterogeneity, and additionally improve the current urban design procedures.

ACKNOWLEDGMENT

This research has been partially supported by the Direct Grant for research of CUHK (4052048), and the Social Science Collaborative Research Fund of CUHK (6903777).

REFERENCES