論 文 の 内 容 の 要 旨

論文題目　Semantic Segmentation for Multi-Source Remote Sensing Imagery based on Convolutional Neural Networks

(Convolutional Neural Networksを用いた多様なリモートセンシング画像の領域分割に関する研究)

氏　名　郭　直霊

In this dissertation, we creatively investigated the feasibility of applying deep learning methods [1] in different semantic segmentation tasks [2] via multi-source remote sensing imagery. The comprehensive researches including village mapping, urban building extraction, super-resolution integrated method, change detection, slum mapping, map segmentation, etc., are conducted. The proposed methods mentioned above are developed by our open source computer vision package named as GeoVision, which contains subpackage GeoSeg and GeoSR, to facilitate the development of the deep learning based segmentation and super-resolution models, respectively. The framework is shown in Figure 1.

In village mapping, we present the Ensemble Convolutional Neural Network (ECNN), an elaborate CNN frame formulated based on ensembling state-of-the-art CNN models [3], to identify village buildings from open high-resolution remote sensing (HRRS) images. First, to optimize and mine the capability of CNN for village mapping and to ensure compatibility with our classification targets, a few state-of-the-art models were carefully optimized and enhanced based on a series of rigorous analyses and evaluations. Second, rather than directly implementing building identification by using these models, we exploited most of their advantages by ensembling their feature extractor parts into a stronger model called ECNN based on the multiscale feature learning method. Finally,

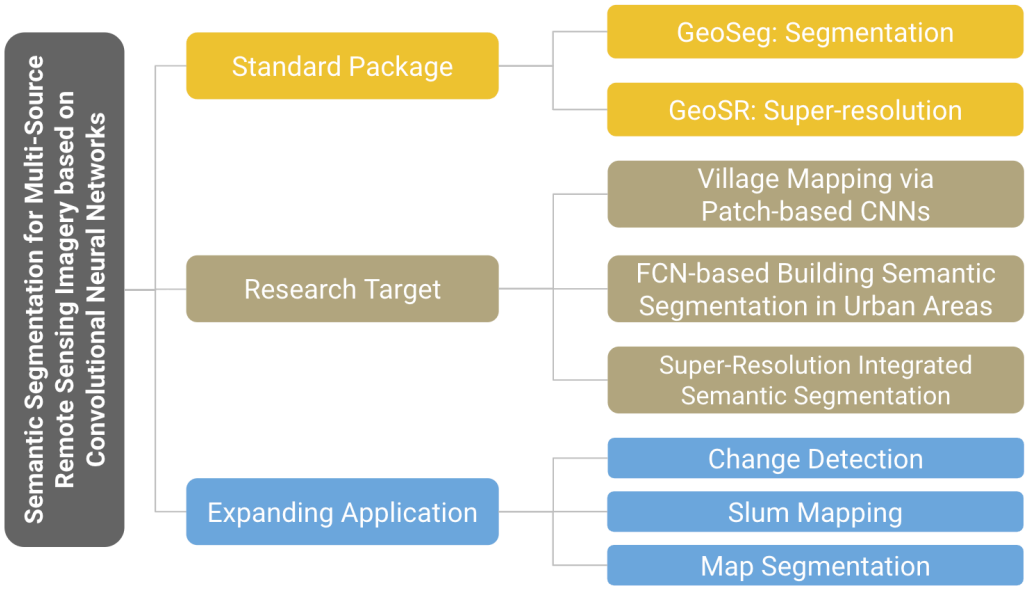

Figure 1. Framework of the dissertation

the generated ECNN was applied to a pixel-level classification frame to implement object identification. The experimental results obtained from the test area in Savannakhet province, Laos, prove that the proposed ECNN model significantly outperforms existing methods, improving overall accuracy from 96.64% to 99.26%, and kappa from 0.57 to 0.86.

As for urban building semantic segmentation part, we investigate the feasibility of applying FCN-based method [4] in conducting map semantic segmentation. Here, high resolution aerial imagery of Tokyo, which provide sufficient information about land features, as a representative sample to perform data source. To mitigate the impact of color difference on segmentation performance, the color transform methods are utilized in image preprocessing as well. In terms of DCNNs model, a specific deep learning architecture named concatenate feature pyramid networks (CFPN), which is a variant of FCN, is proposed based on feature concatenation and feature pyramid methods [5]. Given the variety of the buildings as well as the limited training dataset, CFPN model is deliberately designed in lightweight structure with relatively few parameters, which could be trained easily. Meanwhile, with the help of feature concatenation and feature pyramid, CFPN is capable of extracting adequate robust feature from complex texture to perform building segmentation with high accuracy, and outperform other baselines by 3.55% to 7.89%.

Since multi-source remote sensing imagery has become widely accessible owing to the development of data acquisition systems, we address the challenging task of the semantic segmentation of buildings via multi-source remote sensing imagery with different spatial resolutions. Unlike previous works that mainly focused on optimizing the segmentation model, which did not enable the severe problems caused by the unaligned resolution between the training and testing data to be fundamentally solved, we propose to integrate Super-resolution (SR) techniques [6] with the existing framework to enhance the segmentation performance. The feasibility of the proposed method was evaluated by utilizing representative multi-source study materials: high-resolution (HR) aerial and low-resolution (LR) panchromatic satellite imagery as the training and testing data, respectively. Instead of directly conducting building segmentation from the LR imagery by using the model trained using the HR imagery, the deep learning-based super-resolution (SR) model was first adopted to super-resolved LR imagery into SR space, which could mitigate the influence of the difference in resolution between the training and testing data. The experimental results obtained from the test area in Tokyo, Japan, demonstrate that the proposed SR-integrated method significantly outperforms that without SR, improving the Jaccard index and kappa by approximately 19.01% and 19.10%, respectively. The results confirmed that the proposed method is a viable tool for building semantic segmentation, especially when the resolution is unaligned.

After that, we expand the proposed methods mentioned above to more challenging applications including change detection [7], slum mapping [8], and map semantic segmentation [9]. As for change detection, color normalization, SR, and image registration methods are adopted to balance the training and testing datasets, after that, by adopting proposed CFPN model and image difference, the identification of land change can be achieved. In terms of slum mapping, here CFPN is adopted to perform multi-class semantic segmentation, the impact of resolution on slum segmentation is discussed as well. Furthermore, the important GIS-related task: map semantic segmentation, which aims at digitising historical maps is also applied by our deep learning model.

The experimental results reveal that our proposed method can serve as a viable tool for semantic segmentation tasks via multi-source remote sensing imagery with high accuracy and efficiency.

Reference:

[1] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." nature 521, no. 7553 (2015): 436-444.

[2] Bischke, Benjamin, Patrick Helber, Joachim Folz, Damian Borth, and Andreas Dengel. "Multi-task learning for segmentation of building footprints with deep neural networks." In 2019 IEEE International Conference on Image Processing (ICIP), pp. 1480-1484. IEEE, 2019.

[3] Sainath, Tara N., Abdel-rahman Mohamed, Brian Kingsbury, and Bhuvana Ramabhadran. "Deep convolutional neural networks for LVCSR." In 2013 IEEE international conference on acoustics, speech and signal processing, pp. 8614-8618. IEEE, 2013.

[4] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431-3440. 2015.

[5] Lin, Tsung-Yi, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. "Feature pyramid networks for object detection." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2117-2125. 2017.

[6] Ledig, Christian, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken et al. "Photo-realistic single image super-resolution using a generative adversarial network." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4681-4690. 2017.

[7] Zhu, Zhe. "Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications." ISPRS Journal of Photogrammetry and Remote Sensing 130 (2017): 370-384.

[8] Wurm, Michael, Hannes Taubenböck, Matthias Weigand, and Andreas Schmitt. "Slum mapping in polarimetric SAR data using spatial features." Remote sensing of environment 194 (2017): 190-204.

[9] Guo, Zhiling, Hiroaki Shengoku, Guangming Wu, Qi Chen, Wei Yuan, Xiaodan Shi, Xiaowei Shao, Yongwei Xu, and Ryosuke Shibasaki. "Semantic segmentation for urban planning maps based on U-Net." In IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 6187-6190. IEEE, 2018.