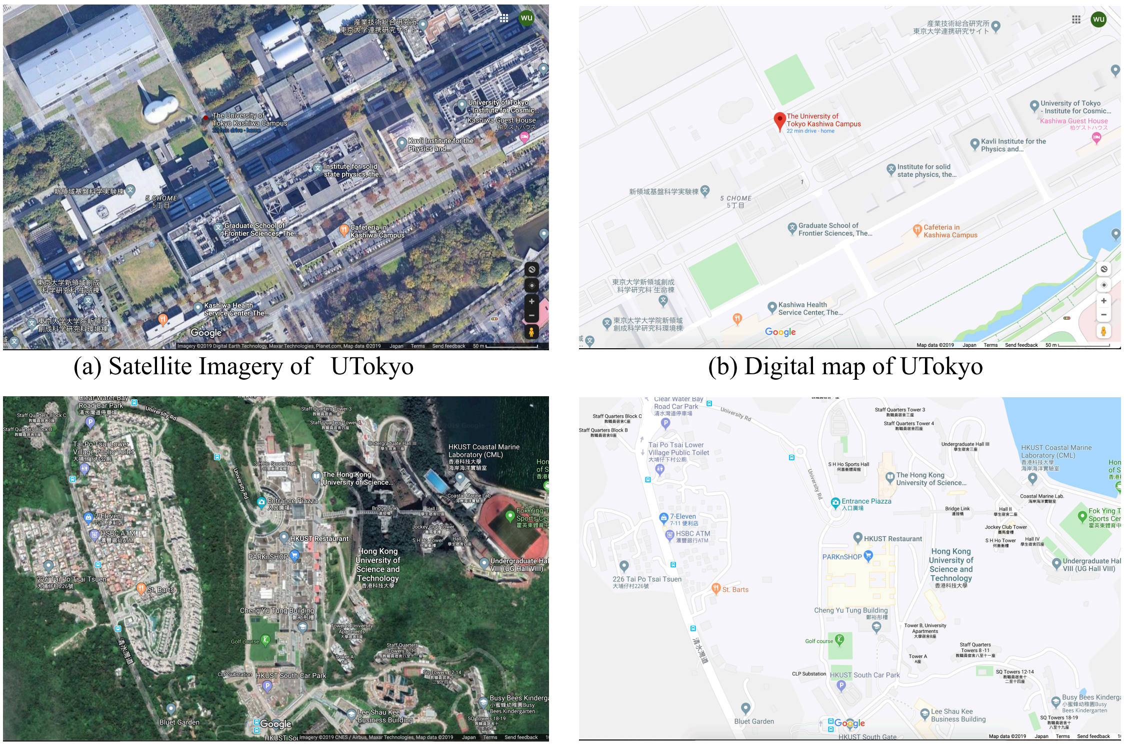
論 文 の 内 容 の 要 旨

論文題目　Automatic Mapping Through Deep Convolutional Neural Networks（ディープニューラルネットワーク（CNN）による自動マッピング）

氏　　名　伍広明（WU GUANGMING）

The distributions and changes of natural and artificial surfaces, such as grasslands, forests, buildings and roads, are essential for many applications such as urban planning, navigation, land-used management, and forest monitoring. Traditionally, this information was obtained by labor-intensive and time-consuming field surveys. The ability to achieve precise and cost-efficient updating of land cover is a long-existing demand for remote sensing. Over the last few years, with the emerging of innovative technologies, the cost as well as difficulty of capturing very high resolution (VHR) aerial imagery has significantly declined. Thus, robust and precise methods for automatic generation of digital maps from captured aerial or satellite imagery are the core of the whole solution.

As shown in Figure 1, compared to satellite images, the digital maps simplified and summarized the RGB values satellite images into two main categories: polygon features (*e.g.*, buildings, green lands, and lakes) as well as line features (*e.g.*, roads and building outline). Besides, even with significantly different appearance of satellite images from different locations, the digital maps should maintain similar style. Thus, to achieve automatic mapping, there are mainly three tasks: (1) Polygon Feature Extraction, (2) Line Feature Extraction, and (3) Model Transfer. The (1) and (2) make sure that the methods should be able to extract both polygon and line features efficiently. With (3), method built for one location should be also work for another location or the same location with different appearance.



(a) Satellite Imagery of UTokyo
(b) Digital map of UTokyo
(c) Satellite Imagery of HKUST
(d) Digital map of HKUST

Figure 1. Example of Satellite images and corresponding digital images from Google Maps. (a) Satellite image of kashiwa campus, The University of Tokyo, (b) Digital map of kashiwa campus, The University of Tokyo, (c) Satellite image of The Hong Kong University of Science and Technology, and (d) Digital map of The Hong Kong University of Science and Technology.

To achieve polygon feature extraction, deep-learning methods, especially fully convolutional networks (FCNs), has become a popular option. Compared with traditional solutions, these approaches have shown promising generalization capabilities and precision levels in various datasets of different scales, resolutions, and imaging conditions. To achieve superior performance, a lot of research has focused on constructing more complex or deeper networks. However, using an ensemble of different fully convolutional models to achieve better generalization and to prevent overfitting has long been ignored. In this research, we design four stacked fully convolutional networks (SFCNs), and a feature alignment framework for multi-label land- cover segmentation. The proposed feature alignment framework introduces an alignment loss of features extracted from basic models to balance their similarity and variety. Experiments on a very high resolution (VHR) image dataset with six categories of land-covers indicates that the proposed SFCNs can gain better performance when compared to existing deep learning methods. In the 2nd variant of SFCN, the optimal feature alignment gains increment of 4.2% (0.772 vs 0.741), 6.8% (0.629 vs 0.589), and 5.5% (0.727 vs 0.689) for its f1-score, jaccard index, and kappa coefficient, respectively.

For efficient line feature extraction, many indirect or direct approaches have been proposed. Recent years, due to the rapid development of deep convolutional networks, line feature ex- traction can also treat as a special polygon feature extraction task that deals with extremely biased distribution of negative and positive pixels. The existing methods are mainly focus on the network design that misalignments and rotations presented in manually created annotations are long ignored. Due to the very limited positive samples, the misalignments and rotations significantly reduce the correctness of pixel-to-pixel loss that might leads to gradient explosion. To overcome this, we propose a nearest feature selector (NFS) to dynamically re-align the pre- diction and slightly misaligned annotations. The NFS can seamlessly integrate into existing loss functions and prevent misleading by errors or misalignment of annotations. Experiments on a largescale image dataset with centered buildings and corresponding building outlines indicates that the additional NFS brings higher performance when compared to existing naive loss functions. In classic L1 loss, the addition of NFS gains increments of 8.8% of f1-score, 8.9% of kappa coefficient, and 9.8% of jaccard index.

Generalization is an important criterion of the algorithm. A well-generalized model should work properly on various areas and data source. Due to the limited training data, current data- driven algorithms, including deep convolutional networks (DCNs), are very sensitive to training data that cannot be applied to new data directly. Differ to existing methods that are trying to improve model generation capability using limited data, we propose an integrated pipeline to generated testing data that shares similar characteristic of training data. Our model transfer pipeline is consisting of super-resolution and colorization model that can convert low-resolution panchromatic images from satellite into high-resolution color images. Experiments on an image dataset with satellite and corresponding aerial imagery shows that pre-trained model achieves significantly higher performance on images processed by integrated pipeline. Compared to the performance on original satellite images, even with slightly decline of precision (4.4%), the addition of SR and colorization leads to 63.6%, 29.6%, 76.6%, and 51.8% increments of recall, f1-score, jaccard index, and kappa coefficient, respectively.