Exploiting Deep Learning and Volunteered Geographic Information for Large-Scale Building Mapping

Jiangye Yuan

Building maps are critical geospatial data for various applications ranging from population estimation to disaster management. However, due to the high cost for large-scale mapping, such data are severely lacked in terms of quality, completeness, and sustainability, especially in the developing world. Satellite images provide a complete and cost-effective data source for creating geospatial data on a large scale. However, manually mapping buildings from satellite images is notoriously time- and effort- consuming. Although automatic methods have been studied for decades, at time of writing there is no existing commercial software that shows reliable performance on large geographic areas.

Given the difficulty of developing automated method, volunteered geographic information (VGI) emerges as an alternative solution. VGI relies on a large number of users to voluntarily contribute to manual mapping. As one of the most successful VGI examples, OpenStreetMap (OSM) is able to mobilize volunteers to quickly mapping objects from images, supporting various real-world applications. Despite its success, since map production is essentially based on manual work, the main option to improve mapping capabilities is to increase community engagement, which does not have an effective solution yet. In fact, the total number of mapped buildings in many countries (e.g., most Africa countries) remains minimal. More importantly, it is difficult to ensure map quality, which has been found to be significantly varying across different regions.

We introduce a new approach that utilizes deep neural networks and VGI data to reliably and efficiently extract buildings from satellite images. The approach has two novel components, described as follows.

1. Training deep neural networks requires large amounts of labeled data, which are expensive to collect. We utilize existing building footprints from VGI data to automatically generate labeled data. Since both maps and images are georeferenced, they can be converted to training samples where individual buildings are labeled. Images and maps are often not well aligned. To deal with this issue, we utilize a simple and efficient procedure that shifts maps to achieve maximum cross-correlation with images.

2. We design a special convolutional network that is well suited for the task. The network has a simple structure that integrates activation from multiple layers for pixel-wise prediction. The network takes input of arbitrary sizes, and processes images in an end-to-end manner. We propose to use the signed distance function to represent output, which provides two advantages over frequently used boundary maps and region maps. 1) Boundaries and regions are captured in a single representation and can be easily read out. 2) Training with this representation forces a network to learn more information about spatial layouts. The method has produced accurate results for country scale datasets.

In experiments, we map buildings for an area of 50,000 km^2 in Kano State in Nigeria. On OSM, there are about 80,000 buildings mapped in this area. Although the actual total building number is unavailable, the following comparison provides some quantitative insights into the scarcity of such data. Washington, D.C has twice as many as buildings on OSM, which is only one eighteenth as populated and one hundredth as large as this area. We use Worldview-2 satellite images with RGB bands and a spatial resolution of 0.5 meter resolution. The image set contains 30 satellite image strips of size 270,000 × 35,000 pixels. To compile training data, we collect OSM building layers of two cities, Kano, Nigeria and Yaounde, Cameroon, which have relatively complete building data. Kano is within the targeted region. OSM building layers are overlaid with corresponding images to yield a labeled dataset covering 51 km^2, which amount to 0.1% of the size of images be processed.

The limited quantity of labeled data poses a major challenge for training a network with good generalization abilities. To address this issue, we conduct two rounds of training, intervened with human feedback. We first train the network using only the OSM derived data. The trained model is tested on randomly selected areas. We then assign four image analysts to identify errors in the results and make corrections, resulting in new labeled images covering 10 km^2. The network is initialized with the previous model and retrained with the expanded training set. The overall training process is completed in 3 days. We use the trained network to process the entire image set, which takes 2 days. All computation is done using a single NVIDIA K80 GPU.

For quantitative evaluation, we select a random subset of results and compare against settlement maps, which are raster data with each 8 × 8 meter block labeled as human settlement or not. They are produced using a combination of a semi-automated tool and manual editing. The building extraction results achieve a precision rate of 75.7% and a recall rate of 75.5%. Results demonstrate that our approach combining deep learning techniques with VGI data provides a promising and highly scalable solution for mapping buildings in very large regions.