ANNEX 17: DEEP SEMANTIC SEGMENTATION FOR BUILT-UP AREA EXTRACTION AND MAPPING FROM SATELLITE IMAGERY

Athuraliya, C. D.;Ramasinghe, Sameera;Lokanathan, Sriganesh;

;

IDRC Grant/ Subvention du CRDI: 108008-001-Leveraging Mobile Network Big Data for Developmental Policy

Annex 17: Deep Semantic Segmentation for Built-up Area Extraction and Mapping from Satellite Imagery

C.D. Athuraliya, Sameera Ramasinghe, Sriganesh Lokanathan

INTRODUCTION

Urban population and urban areas continue to grow. Today, 54% of the world's population lives in urban areas and it is projected to be 66% by 2050. Governments and practitioners are in a continuous struggle to capture this growth for better urban planning, management and various other purposes. Built-up areas play a vital role in any urban setting. They serve varying purposes from human shelters to entertainment. Currently traditional methods prevail in the exercise of understanding the distribution and the growth of built-up areas. These estimates are also used as proxy measures for other urban parameters such as income levels. Nevertheless the process of mapping built-up areas is still predominantly conducted using traditional methods (such as surveys and census) that are infrequent and costly. Multiple efforts have been made to complement or replace traditional methods with novel techniques, however there is a significant usability gap between novel methods and traditional methods in real world use cases.

With the recent advancements and successes in applying deep learning (a sub-field of machine learning) for image analysis, many novel techniques have been proposed to use deep learning for remote sensing data analysis. Yet the extensive training data requirement for deep learning had not been addressed until recent years in remote sensing data domain. Datasets such as SAT-4 and SAT-6 and SpaceNet have reasonably addressed this issue and enabled efficiently applying novel deep learning and machine learning techniques on remote sensing data.

In this work we propose a deep semantic segmentation based method for built-up area extraction and mapping from satellite imagery. Furthermore we propose post-processing techniques from traditional image processing to generate more usable land cover/building maps for practitioners by exploiting certain properties of built-up areas in satellite imagery. We see this as an open challenge due to that fact that even the best performing segmentation results are not visually appealing or engaging due to coarse results. Our method attempts to push raw segmentation results closer to the maps used in real world use cases where domain experts need to understand built-up areas in a given geographical area. We acknowledge the fact that there is an accuracy trade off between raw segmented result and usable map output. But compared to the current low- resolution, costly maps used by authorities, especially in developing countries, the proposed solution offers a better solution.

The dataset we used for this work was SpaceNet which is a publicly available free satellite imagery dataset with pixel level annotations. The dataset includes satellite imagery from five areas of interest (AOI), Rio De Janeiro, Paris, Las Vegas, Shanghai and Khartoum covering 5555 km^2 in the original dataset. This dataset is provided through a collaboration between DigitalGlobe, CosmiQ Works, and NVIDIA. We applied multiple preprocessing steps on both raw satellite imagery as well as on annotations before feeding them into neural network architecture.

METHODOLOGY

Fully Convolutional Networks

Literature suggests that semantic segmentation is a more apt approach compared to bounding box object detection or single image recognition. Intuitively, we needed to understand exactly where our regions of interest (ROI) are by analysing remote sensing data. This is more close to the boundary estimation problem than to bounding box estimation. In our work we utilized a variant of Fully Convolutional Network (FCN) neural network architecture which has been extensively used for semantic segmentation problems. FCN is a re-architectured and fine-tuned classification architecture for direct, dense prediction for semantic segmentation. FCN can operate on an input image of any size and it can produce an output of corresponding dimensions. The architecture we used in this work is based on a classification net which is converted into a fully convolutional net that produces coarse output maps.

Morphological Operations

In image processing, morphological processing implies applying a set of non-linear operators on an input image. This involves applying a structuring element (also known as kernel) on the input image and producing an output image of the same size. Input image and the kernel is typically combined by a set operator such as intersection, union, inclusion, complement. Most common morphological operations are erosion, dilation, opening and closing. Each of these operations generates an output image with desired feature enhancements.

Different variants of morphological operators can be used depending on the problem we are trying to solve. As it was mentioned earlier, the object class we are trying to identify in satellite imagery, built-up areas, has a unique property to it. As opposed to typical object classes such as persons, indoor objects, vehicles where semantic segmentation has been largely applied, built-up areas in satellite imagery typically forms proper geometric shapes: polygons and circles. Out of these two geometric shapes, rectangular shape is the most common in built-up area ground truth and this also agrees with our general understanding. This property lead us to leverage morphological operators to improve segmentation results given our focus on generating more usable built-up area maps. Additionally, morphological operations also act as a noise removal technique to remove small scale false positives as well as to reduce coarseness and holes in true positives.

Our initial attempt was to apply morphological operators as a post-processing step. This is the more common and intuitive method of applying morphological operators in traditional image processing. We applied erosion and dilation on the semantically segmented binary image and visually inspected the change in the final result. The kernel size was the only parameter we adjusted in this case. It significantly depends on segmented image size and the desired granularity of built-up area segmentations. Small kernel sizes lead to minimum or no improvements in results and less noise removal, large kernel sizes caused detail loss in final results. One possible extension to this work would be implementing a quantitative method to determine the optimal kernel size automatically by taking other parameters into account.

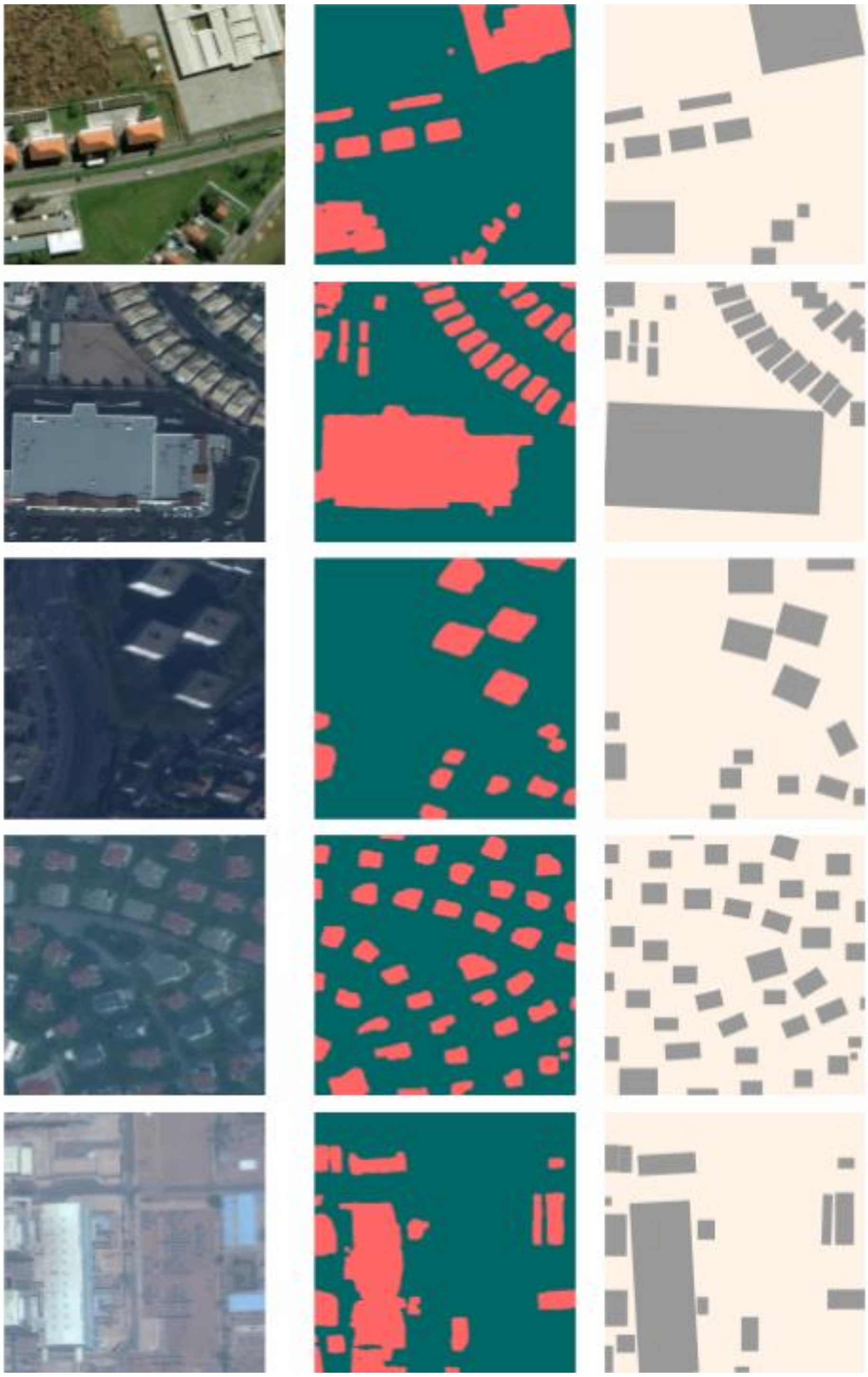
Applying morphological operations as an independent post-processing step yielded better results. Our next attempt was to integrate morphological operators into the FCN architecture itself and observe any improvements in results by altering network parameters in the training process. the main intuition behind this approach is refine the learning by calculating loss after applying morphological operations which will eventually lead the network to train with morphological operations. Results of these experiments demonstrated that training without morphological operations yields better accuracies which led us to continue with the post-processing approach.

Improved Boundary Estimation Using Contours

As discussed earlier, we applied multiple post-processing techniques from image processing to improve the raw segmentation results. A contour is a closed curve that joins all continuous points of the same intensity along a boundary, hence contours are typically obtained from edges. Contours are mainly used to analyse the shape of an object. Rotated bounding rectangles were estimated for each building polygon using contour estimations. These bounding rectangles are final built-up area estimations for each satellite imagery tile.

RESULTS AND DISCUSSION

Quantitatively we obtained a mean Intersection over Union (IoU) of 81.4% by applying training FCN without morphological operations on the full dataset. IoU is the most widely used accuracy measure to evaluate semantic segmentation results. With morphological layers the mean IoU was decreased by approximately 10%. We have also given a comparison of reference images from the dataset for visual inspection which shows the improvement in final result when training was done with morphological layer.

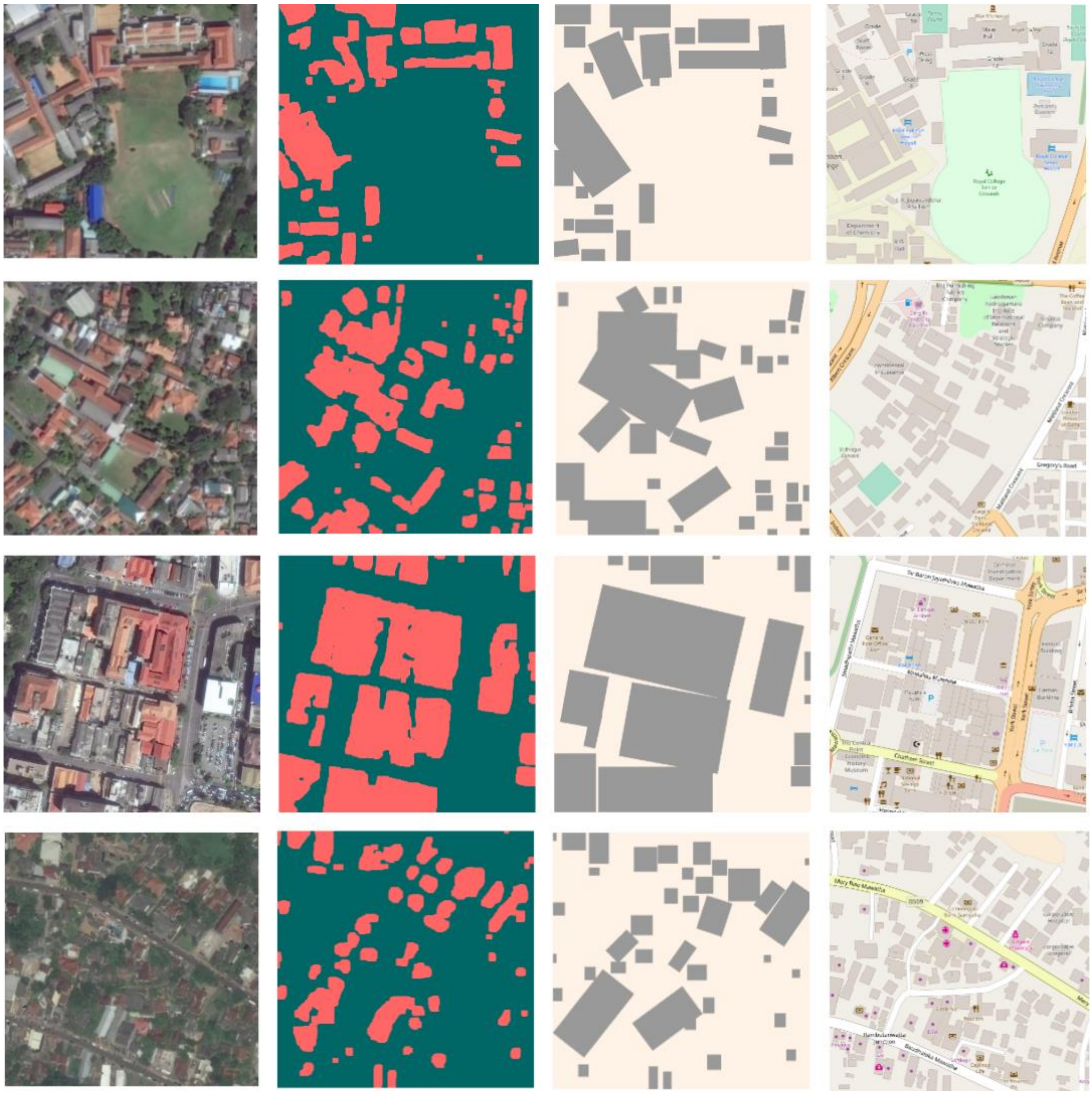


Final result for reference image tiles from 5 AOI row-wise. Left column is image tiles, middle column is built-up area segmentations with binary morphology, right column is final result after rotated bounding rectangles

In this work we proposed a modified FCN architecture to improve semantic segmentation operation on satellite imagery for built-up area extraction and mapping. Leveraging a unique property of built-up areas in satellite imagery, we proposed a post-processing pipeline with morphological operations and other image processing layers to improve the segmentation result generated by FCN architecture. We also present the quantitative and qualitative results whilst acknowledging the accuracy trade off between raw segmented result and improved result for better usability. As was mentioned earlier, our focus was to not only improve the segmentation of built-up areas, but to also bridge the gap between existing extraction techniques and actual land cover/built-up area maps used by practitioners.

This work can be further extended in many directions. One major limitation we noticed in the post-processing results is in handling concave boundaries. This was mainly due to the contour box approach. Our experiments with concave hull for boundary estimation did not yield satisfactory results due to the limitations in handling different concave polygons. One possible extension would be to implement an improved concave hull algorithm for boundary estimation of built-up areas to estimate more precise building footprints. In the literature we can see more recent semantic segmentation architectures applied to relatively small datasets. It would be interesting to see how those results hold on a large dataset with varying features. Another possible extension would be a novel CNN architecture inspired by region proposal networks such as R-CNN to directly predict building boundaries. A quantitative method to automatically select kernel size is also can be explored in future work.

As a further step, we applied the same pipeline for multiple satellite imagery tiles obtained from different areas of Sri Lanka: Colombo and Gampaha. We retrieved high resolution satellite imagery for these areas from DigitalGlobe Maps API web interface. Resolution was comparable but lower than SpaceNet imagery. Respective building footprints, obtained from OpenStreetMap were used to validate the results given below qualitatively by visual inspection. We see several potential use cases of this work in socio economic classification and urban planning such as building density as a proxy measure for socio economic level and building distribution for urban area estimates respectively. Same pipeline can be further extended with deep learning and computer vision techniques to estimate other indicators, e.g. roads and rooftop material.



Results for different areas of Sri Lanka: Colombo and Gampaha