AUTOMATED NATIONAL URBAN MAP EXTRACTION

Hasan Nasrallah
National Center for
Remote Sensing - CNRS
Beirut, Lebanon

Abed Ellatif Samhat
Faculty of Engineering
Lebanese University
Hadath, Lebanon

Cristiano Nattero
WASDI
Luxembourg

Ali J. Ghandour[*]
National Center for
Remote Sensing - CNRS
Beirut, Lebanon

ABSTRACT

Developing countries usually lack the proper governance means to generate and regularly update a national rooftop map. Using traditional photogrammetry and surveying methods to produce a building map at the federal level is costly and time consuming. Using earth observation and deep learning methods, we can bridge this gap and propose an automated pipeline to fetch such national urban maps. This paper aims to exploit the power of fully convolutional neural networks for multi-class buildings' instance segmentation to leverage high object-wise accuracy results. Buildings' instance segmentation from sub-meter high-resolution satellite images can be achieved with relatively high pixel-wise metric scores. We detail all engineering steps to replicate this work and ensure highly accurate results in dense and slum areas witnessed in regions that lack proper urban planning in the Global South. We applied a case study of the proposed pipeline to Lebanon and successfully produced the first comprehensive national building footprint map with approximately 1 Million units with an 84% accuracy. The proposed architecture relies on advanced augmentation techniques to overcome dataset scarcity, which is often the case in developing countries.

Index Terms— Urban map, Rooftop Instance Segmentation, Slum Areas

1. INTRODUCTION

Buildings' footprints extraction from aerial imagery is an important step for many urban applications. Fully automated extraction and recognition of buildings footprint geometries can also be used for a wide range of scientific applications in various domains such as solar rooftop potential estimation, solid waste management, and air pollution modeling, among others.

Classifying pixels into semantic and instance objects in urban areas satellite images is currently undergoing important attention in the research community, in addition to development efforts in the industry. Remote sensing images are complex data, characterized by the form of heterogeneous regions with large intra-class variations and often lower inter-class variations [1]. Deep learning significantly reduces the time required to achieve such tasks because of its capabilities in automatically extracting meaningful and well-defined features and patterns present in large scenes.

Given the aforementioned time and cost advantage over other traditional methods like on-site measurements or pure image processing, we introduce a multi-class high-accuracy segmentation architecture for extracting rooftop geometries from satellite images using deep convolutional neural networks.

The contribution of this paper is three-folds: *(i)* present a fully automated pipeline to extract urban map at the national level, *(ii)* translate the problem in hand to a multi-class segmentation approach by adding two intermediate classes (border and spacing) and *(iii)* finally apply the proposed workflow to Lebanon as a demo.

2. RELATED WORK

Many methods have been proposed to solve the problem of building segmentation from aerial imagery. In [2], the authors use gated graph convolutional neural networks to produce a transcribed signed distance map (TSDM), which is then converted into a semantic segmentation mask of buildings. In [3], the authors propose two plug-and-play modules to generate spatial- and channel-augmented features for semantic segmentation from satellite images. In [4] the authors use augmentations like slicing, rescaling, and rotations and additional GIS data to improve building footprint extraction. In [5], the authors use siamese networks to segment and classify buildings present in pre- and post-disaster images. However, a more direct approach is presented in [6] where the authors use only a semantic segmentation network with an additional output mask that designates the spacing between very close buildings to separate building instances.

[*] Correspondence: aghandour@cnrs.edu.lb.

3. METHODOLOGY

3.1. Multi-class Segmentation

In order to achieve high national wide segmentation score, especially in dense and slum areas, we define the following three output classes:

Building Class Pixels belonging to the interiors of any building polygon.

Border Class Pixels belonging to the contours of the buildings. We create borders of width = 2 pixels.

Spacing Class Pixels belonging to the separation between very close buildings.

Our best model is trained for 100 epochs using Adam Optimizer [7] and the One-Cycle learning rate policy [8] using a Titan-Xp GPU. We used mixed precision training [9] and batch size = 16. During training, we apply random augmentations on images, including positional transformation, color, distortion, and noise transformations. Furthermore, to increase the robustness of the model, we use CutMix [10] data augmentation.

At inference time, each image undergoes 3 positional test time augmentations (horizontal-flip, vertical-flip and 180° rotation) to allow the receptive field of the model to analyze the same image in 4 different ways. Inference at a country level is computationally expensive, and hence several techniques were implemented to speed up this process.

3.2. Study Area

To create our dataset, we relied on a 35372x28874 GeoTIF image that covers the city of Tyre in the South of Lebanon. The labeling process took about 100 hours and the final dataset is made up of 1,352 images of 512x512 and 10,000 buildings. We split the data set into five folds.

We then automatically generate ground-truth border masks using Algorithm 1. To create the spacing mask between close buildings, we also followed several preprocessing techniques that will not be mentioned here for space limitation.

As for the test set, we selected 30 Areas of Interest (AoI) of 1024x1024 size from different cities in Lebanon including Beirut, Saida, Jounieh, Jbeil and Tripoli. This would help to assess the generalizability of our model over the whole Lebanese geographical area. These AoI's contained dense and slum urban regions (Beirut and Tripoli), areas with proper urban planning (Saida), and some suburban and rural areas (Jbeil).

4. EXPERIMENTAL RESULTS

In our experiments, Efficient-Net-B3 was found to perform better in terms of accuracy and variance compared to other members of the Efficient-Net family members and other encoders such as ResNet34, ResNeXt50, InceptionV4, InceptionResNetV2 and DPN92.

Algorithm 1 Border Mask Creation

```
Given a list of N Polygon Coordinates : $Polys = [P_1, P_2, ...P_N]$;
$H \leftarrow 1024$
$W \leftarrow 1024$
$AllBordersMask \leftarrow NumpyZerosArray((H, W))$ ▷ HxW array of zeros
for P in Polys do
    $TempMask \leftarrow NumpyZerosArray((H, W))$
    Draw & Fill Polygon $P$ in $Temp$ with ones
    $N \leftarrow 2$
    $ker \leftarrow Square(3)$
    $ErodedMask \leftarrow Erosion(TempMask, N, ker)$ ▷ Perfom Binary Erosion on the Temporary Mask with a 3x3 squared kernel for N=2 times
    $CurrentBorderMask \leftarrow Temp \bigoplus Eroded$ ▷ XOR operation get the border mask of Polygon P
    Extract Border pixels coordinates from $CurrentBorderMask$ & assign ones to these coordinates in $AllBordersMask$
end for
```

We also inspect the effect of mixed data augmentations such as Cutmix [10] and MixUp [11]. CutMix was introduced in [10] and was used for classification purposes. When using Cutmix, the IoU score increased from 73.2% to 75.1% and the distance between the Train and Validation scores decreases considerably. However, using Mixup does not improve model performance. Table 1 summarizes various combinations of hyperparameters experimented.

To further test the proposed pipeline, we sampled various tiles from different regions of Lebanon, including Beirut, Saida, Byblos and Tripoli, for which we ran the proposed pipeline and calculated the F-score results as shown in Table 2. In Figure 2, we qualitatively show the instance segmentation results for those different regions where each building is presented in a different color. The results in Table 2 show that the muti-class model approach provides up to 14% increase in F-score. This gain is achieved by the better ability of the multi-class model to separate very close buildings.

The Nadir angle for very dense areas with tall buildings, such as Beirut, is a major factor that affects the performance of the model, which explains the drop in F-score over this region. In future work, we plan to tackle this issue and provide practical workarounds. For regions with proper urban planning, such as Saida, our model provides an outstanding F-score of 81.5%.

Furthermore, to further investigate the performance per image for the single- vs multi-class methods, we draw the violin plots in Figure 1 to compare the probability density for

N	Backbone	AMP	Mixed Augs	Scheduler	Fscore
1	EffNet-B2	✗	✗	PolyLr	83
2	EffNet-B3	✓	✗	PolyLr	**84.3**
3	EffNet-B4	✓	✗	PolyLr	83.8
4	EffNet-B3	✓	MixUp	PolyLr	83.2
5	EffNet-B3	✓	CutMix	PolyLr	**85.6**
6	EffNet-B3	✓	CutMix	CycLR	85.2
7	EffNet-B3	✓	CutMix	CycLR WM	85.2
8	EffNet-B3	✓	CutMix	1Cycle	**87.6**

Table 1: Pixel Fscore results on our validation dataset showing a series of improvements done by searching for the best hyperparameters like backbone architecture, automatic mixed precision, mixed augmentation and learning rate schedulers.

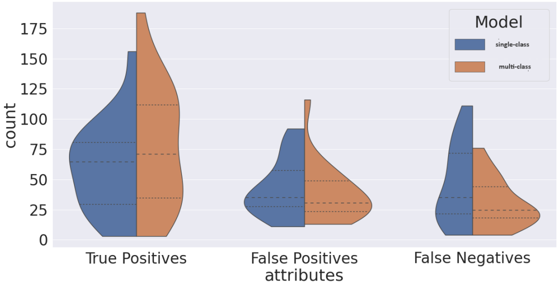


Fig. 1: Violin Plots showing the True Positive, False Positive and False Negative probability density distributions to compare the instance segmentation performance between the single and multi-class models.

the True Positive (TP), False Positive (FP), and False Negative (FN) predictions per image over the test dataset. The dashed line represents the median of each distribution and the dotted lines represent the Inter-Quartile Ranges (IQR). The IQR represents the 'Middle 50%' distribution of the data, so it provides a sense of the concentration of values. Figure 1 shows that the median and IQR boundaries of the True Positive are larger for the multi-class model. Moreover, the median and IQR boundaries of False Positive and False Negative plots are smaller for multi-class model. The data distribution shown in the violin plot emphasizes the gains of the proposed model at test time.

5. CONCLUSION

Region	single-class F-score(%)	multi-class F-score(%)
Saida	67.40	81.56
Tripoli	56.15	70.20
Byblos	56.68	67.03
Beirut	49.07	55.10

Table 2: A comparison between single-class and multi-class F-score results on images from different regions.

In this paper, we presented how we used a multi-class segmentation model to extract building footprints from satellite imagery at a national level with a demo for Lebanon. The proposed pipeline has impact and use cases over several technologies such as national solar potential map, utilities planning at local government level, urban agglomeration, air polution modeling, solid waste modeling, and many others.

6. REFERENCES

[1] Rasha Alshehhi, Prashanth Reddy Marpu, Wei Lee Woon, and Mauro Dalla Mura, "Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 130, pp. 139 – 149, 2017.

[2] Y. Shi, Q. Li, and X. X. Zhu, "Building extraction by gated graph convolutional neural network with deep structured feature embedding," in *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, 2020, pp. 3509–3512.

[3] Lichao Mou, Yuansheng Hua, and Xiao Xiang Zhu, "A relation-augmented fully convolutional network for semantic segmentation in aerial scenes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[4] Weijia Li, Conghui He, Jiarui Fang, Juepeng Zheng, Haohuan Fu, and Le Yu, "Semantic segmentation-based building footprint extraction using very high-resolution

satellite images and multi-source gis data," *Remote Sensing*, vol. 11, no. 4, 2019.

[5] Alexey Trekin, German Novikov, Georgy Potapov, Vladimir Ignatiev, and Evgeny Burnaev, "Satellite imagery analysis for operational damage assessment in emergency situations," *CoRR*, vol. abs/1803.00397, 2018.

[6] Vladimir Iglovikov, Selim Seferbekov, Alexander Buslaev, and Alexey Shvets, "Ternausnetv2: Fully convolutional network for instance segmentation," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.

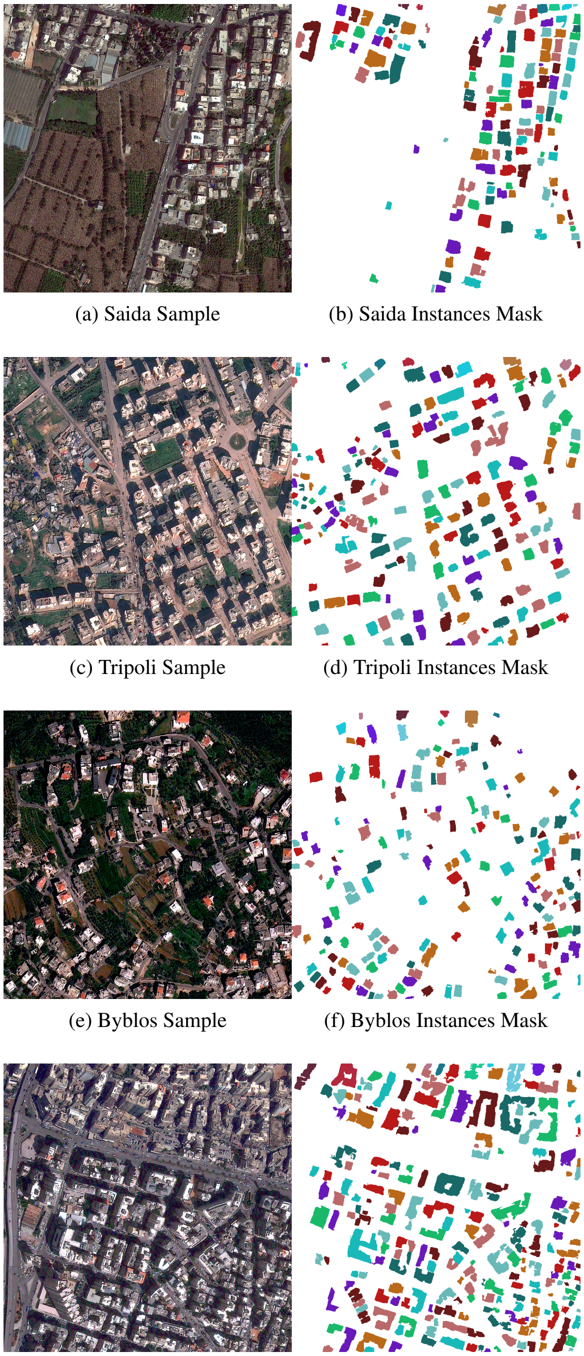
[7] Diederik P. Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2015.

[8] Leslie N. Smith and Nicholay Topin, "Super-Convergence: Very Fast Training of Neural Networks Using Large Learning Rates," *arXiv e-prints*, p. arXiv:1708.07120, Aug. 2017.

[9] Paulius Micikevicius, Sharan Narang, Jonah Alben, Gregory Diamos, Erich Elsen, David Garcia, Boris Ginsburg, Michael Houston, Oleksii Kuchaiev, Ganesh Venkatesh, and Hao Wu, "Mixed precision training," in *International Conference on Learning Representations*, 2018.

[10] S. Yun, D. Han, S. Chun, S. J. Oh, Y. Yoo, and J. Choe, "Cutmix: Regularization strategy to train strong classifiers with localizable features," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 6022–6031.

[11] Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, and David Lopez-Paz, "mixup: Beyond empirical risk minimization," in *International Conference on Learning Representations*, 2018.



(a) Saida Sample (b) Saida Instances Mask

(c) Tripoli Sample (d) Tripoli Instances Mask

(e) Byblos Sample (f) Byblos Instances Mask

(g) Beirut Sample (h) Beirut Instances Mask

Fig. 2: Sample images and their corresponding instance segmentation mask using multi-class inference. The buildings are represented in different colors to emphasize the fact that each instance has its own identifier.