Roadmap Generation using a Multi-Stage Ensemble of Deep Neural Networks with Smoothing-Based Optimization

CVPR 2018 SALT LAKE CITY • JUNE 18-22

Dragos Costea[1,2,*] Alina Marcu[2,3,*] Emil Slușanschi[1] Marius Leordeanu[1,2,3]

[1] University Politehnica of Bucharest [2] Autonomous Systems [3] *Simion Stoilow* Institute of Mathematics of the Romanian Academy

Introduction

- Road detection and vectorization is essential for automatic map generation. High resolution, daily RGB satellite imagery is available. Manual annotation is expensive and time consuming.
- Accurate maps enable localization and navigation for unmanned ground and aerial vehicles (UGVs and UAVs). Shorter routes result in faster and more economical transportation.
- We propose an automatic map generation method that yields good results in terms of pixelwise accuracy, while also producing a vectorized map.

Our Approach

We use a multi-stage ensemble of U-Net-like networks with multiple dilation rates. As road junctions are important for overall graph structure, we also train an intersection detector.

Steps:

- Extract road segmentation maps from trained U-Net-like networks.
- Detect intersections.
- Concatenate our networks results (for both roads and intersections).
- Train another network on top of previous predictions.
- Create road graph from the final predictions.
- Add missing links inferred from the graph.

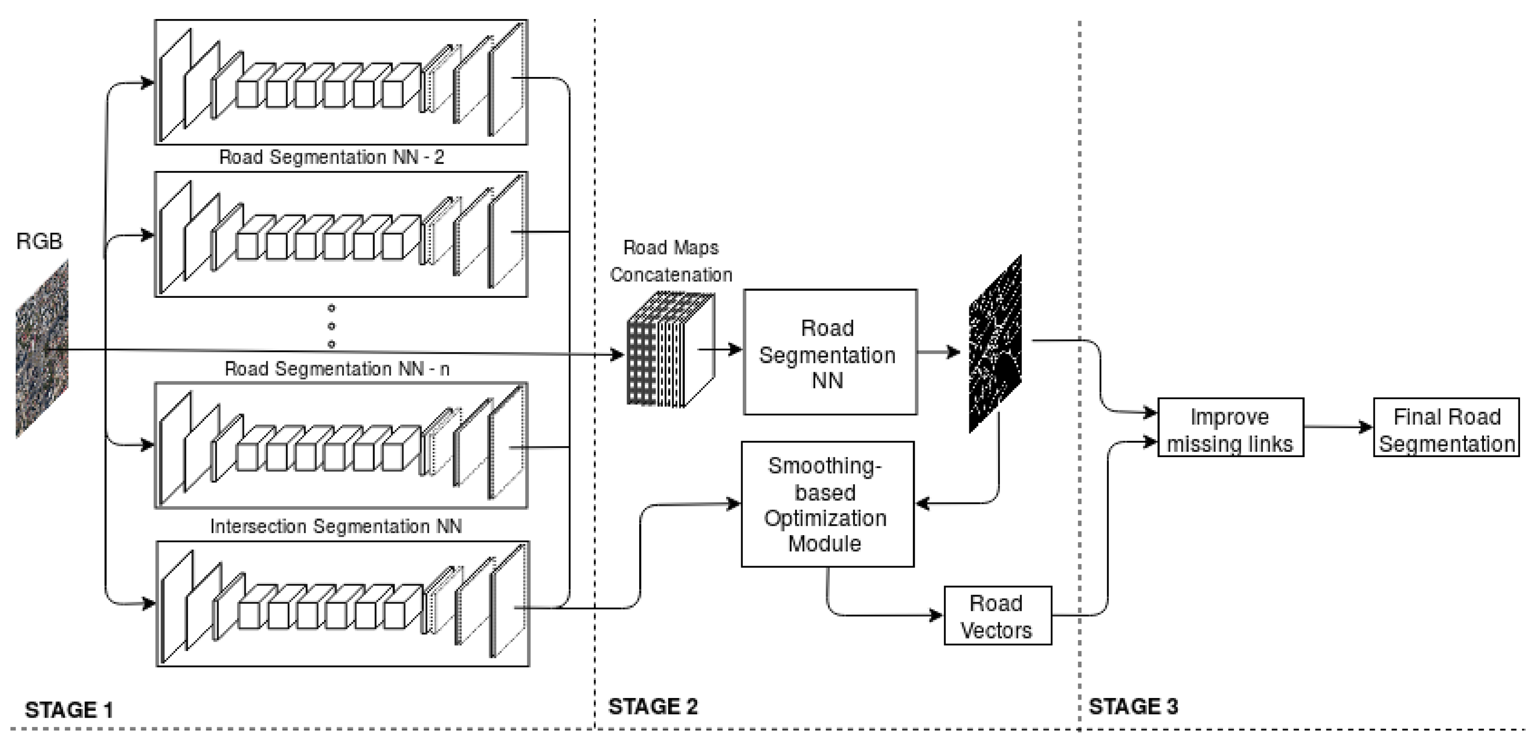
Multi-Stage Road Detection

Stage 1: We detect roads using an ensemble of U-nets with variable dilation rates for the bottleneck:

- Max dilation 32 (1, 2, 4, 8, 16, 32)
- Max dilation 48 (1, 2, 4, 8, 16, 32, 48)
- Max dilation 64 (1, 2, 4, 8, 16, 32, 48, 64)

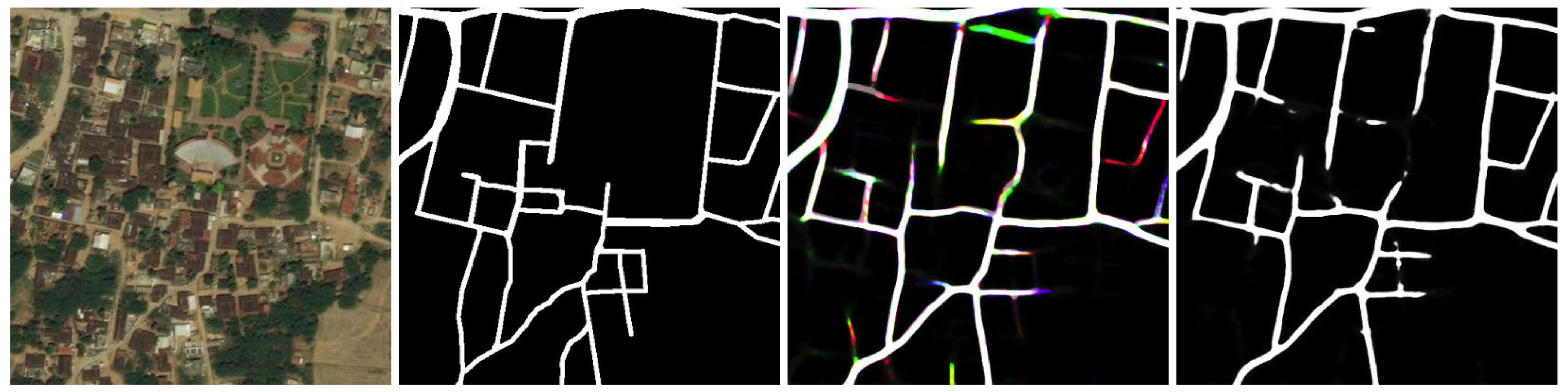
Stage 2: The generated road maps, along with the RGB and detected intersections are concatenated and fed to the next stage U-net-like architecture trained for road segmentation.

Stage 3: We generate road vectors from the improved segmentation map and further use these vectors to add missing links.



*Equal contribution

Qualitative Results



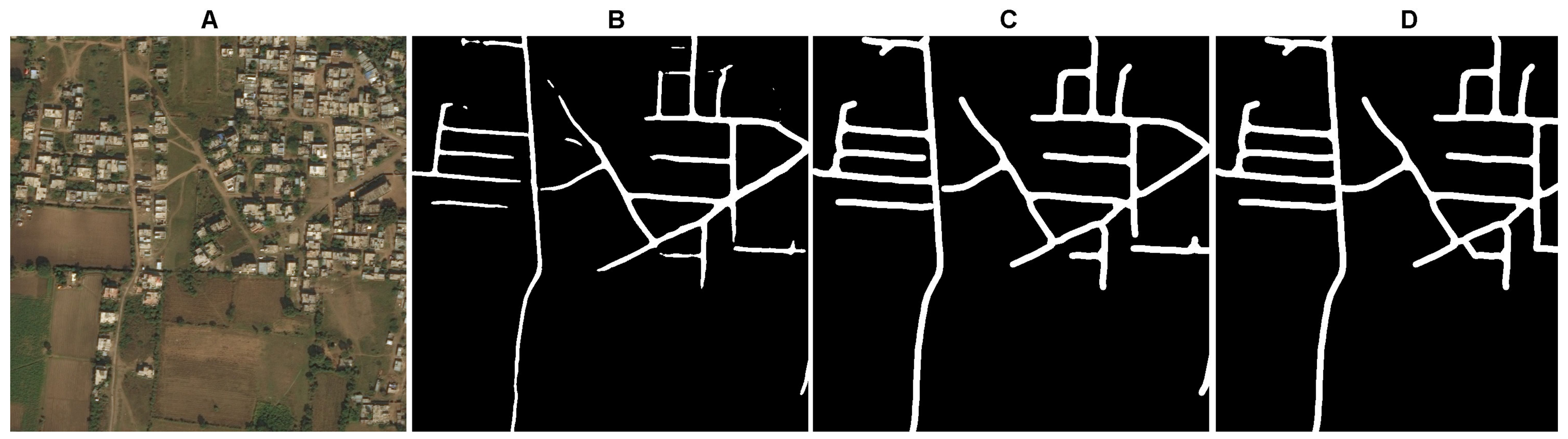
(A) RGB input, (B) Ground truth, (C) Ensemble 1 (Stage 1), shown as sum of 4 CNN outputs: Max dilation 32 - blue, Max dilation 48 - green, Max dilation 64 - red, Max dilation 32, same width thin - grey. (D) Ensemble 2, the output of Stage 2.

Problems: wrong or missing label, prolonged roadside occlusion, ample road width variations. It also highlights the gains of multiple dilation rates and fixed width training.

Adding Missing Links

We vectorize the final binary image from the NN using SBO [1] and infer missing links based on the graph structure.

This would be difficult to achieve by means of traditional techniques (such as distance transform) - see roundabouts, junctions.



Qualitative results of stage 3 of our method (adding missing links). (A) RGB input, (B) binary mask from CNN, (C) plotted road vectors from SBO, (D) final segmentation with missing links added to the road vectors.

Road Thickness Study

- Thin roads and large pavement areas generally yield poor detection performance.
- We investigate the impact of label thickness on detection and trained additional networks with thicker roads.
- Roads as thick connections are better than having a variable or thin road that misses out road segments.

Width type	Thickness	IoU Our Training	IoU Our Validation
Same width	Thin (≈ 4m)	0.6282	0.6123
	Thick (2x thin)	0.6918	0.6751
Variable	Thin (original)	0.6432	0.6483
	Thick (2x thin)	**0.7254**	**0.6889**

More details about our work: https://sites.google.com/site/aerialimageunderstanding/

Experiments

Table 1. Roads segmentation results on the official validation set, 1243 images (**Round 1**).

Model	Iteration	IoU Validation
Max dilation 32	1	0.5924
	2	0.5975
Max dilation 48	1	0.6039
	2	**0.6058**

Table 2. Roads segmentation results. Results reported on 5603 training images and 623 validation images, randomly selected from the original training set, for which ground truth was provided.

Model	Our Training		Our Validation	
	IoU	F1	IoU	F1
Max dilation 32	0.6432	0.7824	0.6483	0.7883
Max dilation 48	0.6577	0.7913	0.6601	0.7957
Max dilation 64	**0.6591**	**0.7919**	**0.6640**	**0.7966**

Table 3. Intersection segmentation results (report IoU scores).

Input	Our Training	Our Validation
RGB only	0.2840	0.1724
Roads only	0.5627	0.5517
RGB + Roads	**0.7112**	**0.6492**

Table 4. Roads segmentation results using our ensembles.

Model	Our Training		Our Validation	
	IoU	F1	IoU	F1
Ensemble 1	0.6356	0.7749	0.6345	0.7769
Ensemble 2	**0.7287**	**0.8506**	**0.6920**	**0.8239**
Ensemble 1+2	0.6514	0.7882	0.6412	0.7829

Table 5. Roads segmentation results on the official testing set, 1101 images. The results were provided by the submission site (**Round 2**). Results reported after adding the missing links.

Model	IoU Testing
Baseline [2]	0.545
Ensemble 1	0.5788
Ensemble 2	**0.5862**
Ensemble 1+2	0.5785

References

[1] D. Costea, A. Marcu, E.-I. Slusanschi, and M. Leordeanu. Creating roadmaps in aerial images with generative adversarial networks and smoothing-based optimization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2100–2109, 2017.

[2] I. Demir, K. Koperski, D. Lindenbaum, G. Pang, J. Huang, S. Basu, F. Hughes, D. Tuia, and R. Raskar. Deepglobe 2018: A challenge to parse the earth through satellite images. *arXiv preprint arXiv:1805.06561*, 2018.