



HAL
open science

BRIGHTEARTH ROADS: TOWARDS FULLY AUTOMATIC ROAD NETWORK EXTRACTION FROM SATELLITE IMAGERY

Liuyun Duan, Willard Mapurisa, Maxime Leras, Leigh Lotter, Yuliya
Tarabalka

► **To cite this version:**

Liuyun Duan, Willard Mapurisa, Maxime Leras, Leigh Lotter, Yuliya Tarabalka. BRIGHTEARTH ROADS: TOWARDS FULLY AUTOMATIC ROAD NETWORK EXTRACTION FROM SATELLITE IMAGERY. IGARSS2024, Jul 2024, ATHENE, Greece. hal-04593369

HAL Id: hal-04593369

<https://inria.hal.science/hal-04593369>

Submitted on 20 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

BRIGHT EARTH ROADS: TOWARDS FULLY AUTOMATIC ROAD NETWORK EXTRACTION FROM SATELLITE IMAGERY

Liuyun Duan^a, Willard Mapurisa^b, Maxime Leras^a, Leigh Lotter^b, Yuliya Tarabalka^a

^a LuxCarta Technology, Mouans Sartoux, France ^b LuxCarta South Africa, Cape Town, South Africa

ABSTRACT

The modern road network topology comprises intricately designed structures that introduce complexity when automatically reconstructing road networks. While open resources like OpenStreetMap (OSM) offer road networks with well-defined topology, they may not always be up to date worldwide. In this paper, we propose a fully automated pipeline for extracting road networks from very-high-resolution (VHR) satellite imagery. Our approach directly generates road line-strings that are seamlessly connected and precisely positioned. The process involves three key modules: a CNN-based neural network for road segmentation, a graph optimization algorithm to convert road predictions into vector line-strings, and a machine learning model for classifying road materials. Compared to OSM data, our results demonstrate significant potential for providing the latest road layouts and precise positions of road segments.

Index Terms— Remote sensing, deep learning, semantic segmentation, road network, road materials.

1. INTRODUCTION

With the development of internet and electronic information technologies, the scale of the earth is no longer as immense as before in our imagination or the view of telescopes. We stay ambitious to duplicate the world into a virtual digital twin. Buildings and roads are two of the top interesting man-made object classes that are indispensable to extract. Applications such as smart cities, intelligent navigation, and simulation require large-scale high accuracy road networks with rich attributes to conduct realistic experiences. The precision of road axis positions, correctness of topology and geometry, and attributes such as number of lanes, width, materials become critically important for most applications. Thanks to the continuously intensive developments of remote sensing image processing [1, 2] and deep learning technologies, fully automatic road extraction becomes more and more promising [3, 4]. Recent road extraction works based on deep learning can be sorted into two major groups: pixel-based road segmentation with a post-processing module to convert the raster prediction to road axis in line-strings [5, 6, 7]; graph reconstruction directly from images by iterative prediction of next

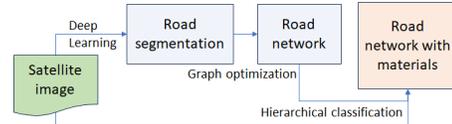


Fig. 1. The proposed road extraction pipeline from a single input satellite image.

connecting point or by detecting interesting points of road topology with/without a segmentation guidance [8, 9, 10, 11]. However, most of these existing methods often do not handle relatively simple scenarios with sparse intersections such as in residential areas. Incompleteness and broken connections are the most common problems when applying these methods to large-scale road network extraction. In this paper we propose an improved CNN-based road segmentation method and a graph optimization algorithm to generate road networks with mathematically explainable controls of the quality of the geometry and the topology. The precise road geometry and reliable topology also aid in the classification of road materials.

We summarize our main contributions as following:

- 1) Effective and efficient automatic road extraction pipeline that leverages both the advantages of high performance deep learning segmentation and robust and controllable remote sensing image processing technology.
- 2) New road segmentation architecture that improves the prediction quality by leveraging multiple learning tasks in one framework.
- 3) New road network reconstruction method based on graph optimization that generates OSM-like road structures with geometrical compactness and regularity.
- 4) High performance road material classification module by applying a light machine learning model with a small amount of training data, which is supported by the accurate road network vector layer generated by the previous two points.

We describe the technical details of the proposed pipeline in the following three sections: road segmentation, road network reconstruction and road material classification.

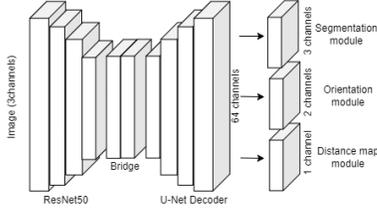


Fig. 2. The proposed new architecture for road segmentation from a single satellite image.

2. ROAD SEGMENTATION

The extraction of road semantics is subject to several challenges such as occlusion from buildings, trees and other roads, abrupt changes in color and materials along the same road and road-like human-made objects. To overcome these issues, the combination of a precise and worldwide dataset, a state of the art neural network design and a multi-class loss is needed.

2.1. Data Preparation

Our dataset comprises 130 km² areas of 30cm and 50cm satellite imagery, covering 108 zones worldwide. Each image was manually annotated with polygons of sub-pixel precision Ground Truth to match sidewalks and road limits. The road polygons were rasterized into a three-class mask that represents road interiors, road contours and others. Rasterizing these polygons without considering overlapping roads, as seen in interchanges, was insufficient. We added information to the polygons to determine the relative road levels locally.

2.2. Neural Network Design

The architecture of our road model is based on ResNet for the backbone and U-Net for the decoding part [5]. We have three modules responsible for the last deconvolutions: the segmentation module that outputs the three classes (road interior, road contour and others); the orientation module that outputs local orientations [12]; and a distance map module [13]. The orientation module, trained with a smoothing loss, helps to segment obstructed roads in a continuous way. The distance map module benefits both interior and contour segmentation. It is redundant with the contour, but it offers flexibility to define the best position of contours without too much penalty when missing the contour by one pixel. This is in contrast to IoU (Intersection over Union) or cross-entropy losses on the three classes segmentation. Figure 2 illustrates the idea of our neural network design.

2.3. Training

Our model is trained on the dataset with a training/validation ratio of 85/15. The segmentation module is trained using IoU

and cross-entropy losses with more weight on contours to balance classes. The orientation module is trained using a differentiable loss on predicted angles so that a predicted orientation of π is equivalent to a predicted orientation of 0. The distance map module is trained using L2 norm pixel-wise to penalize greater differences. Examples of our road segmentation are shown in Figure 4, second column.

3. ROAD NETWORK RECONSTRUCTION

Following the segmentation of the roads, the reconstruction of the road network is performed, where the results of the segmentation are vectorized. The goal of reconstruction of the road network is to reconstruct road vectors that accurately follow the center of the road. This process is composed of five processes; namely road skeletonization, linear road recovery, denoising, road smoothing, and road circle reconstruction.

3.1. Road Skeletonization

From the segmentation results, a road skeleton [14, 15] is created and vectorized as the first approximation of the road center. A graph is created to represent the vectorized skeleton. The vectorized skeleton is characterized by a stair stepping effect as a result of the raster to vector conversion.

3.2. Linear Roads Recovery

To remove the stair stepping effect, an incremental linear least squares line fitting is performed on the vector skeleton in between road junctions to create linear road vectors and remove stair stepping. A junction is any point where three or more roads intersect. Incremental least squares result in linear road vectors represented by the minimum number of vertices between junctions.

3.3. De-Noising

After recovering the initial linear road vectors, a cleaning operation is performed, termed denoising. The cleaning operation removes invalid roads by identifying closed road loops that cover small areas, since small areas do not constitute meaningful road features, with the exception of circles. By identifying loops with areas below a defined threshold in the graph of road vectors, noise and errors are removed. This is achieved by converting the loops into linear road segments. Since some of the loops detected are traffic circles, Hough circle detection [16] is performed on all detected loops. The raster sections from the segmentation results covered by the detected loops are used as input in verifying the presence of circles. Any loops that cover circles are recorded and used in the final step to reconstruct traffic circles, while all other loops are removed and replaced by linear road segments. This operation removes the noise from the road network. Additionally, dangling roads below a given threshold are filtered out.

3.4. Road Smoothing

One major drawback of raster skeletons is the inability to preserve T-junctions and correct road junction topology. Subsequently, the linear roads recovered in the second step have distorted junctions. After denoising, T-junctions are then reconstructed. This is achieved by imposing a smoothness constraint at the junctions. The constraint is that at intersections, at least two incident vertices to a junction and the junction vertex itself must be collinear. Enforcing this constraint on the vector road graph reconstructs the junctions of the road and results in smooth roads. The defined constraint ensures curves with smoothly changing curvature that is not disrupted by incorrect topology at junctions are reconstructed. Thus long roads with smooth curves are reconstructed. More complex road junctions are reconstructed incrementally.

3.5. Traffic Circle Reconstruction

Finally, circle fitting is performed on loops that cover circles that were detected in the second step 3.3. The original loops are replaced with the detected circles, thereby completing road reconstruction. Furthermore, the sections detected as double lanes are reconstructed by replacing the detected double lane center line with duplicate roads on either side of the double lane detected center line.

Examples of road networks generated by our method are shown in Figure 4, third column.

4. ROAD MATERIAL CLASSIFICATION

A hierarchical classification method is proposed to distinguish the material of generated road networks. Roads are classified initially into two main surface classes: processed and unprocessed. Processed roads refer to those made out of concrete, tar or other man-made, solid surfaces. Unprocessed roads include those made up of dirt, sand, gravel or other naturally occurring materials. To allow for better statistical representations to be extracted from the corresponding images, a small buffer of 2m is applied to the line-strings to create polygon segments to collect road material pixels from the input satellite images. We apply two iterations of the Support Vector Machine (SVM) classifier to train the processed and unprocessed road classification based on 148km² (RGB) and 483km² (RGB-Nir) archived data with hand crafted road material classes tagged in the Ground Truth road networks.

To further distinguish the subclasses: gravel and sand for unprocessed roads, we introduce our BrightEarth Land Use Land Cover (LULC) layer [17] (generated from the Sentinel-2 global mosaic) to determine the road materials in a semantic context. We search the land cover labels within the radius of 1 km² of each unprocessed road. If the surrounding is mostly barren land or water areas, then the probability of this road being sand is higher than gravel.



Fig. 3. Road materials classification examples. Green for processed roads and red for unprocessed roads (a), pink for gravel roads (b) and orange for sand roads (c).

The greatest advantage of the proposed SVM scheme is that the model can be trained with a small amount of data while providing a high generality and robustness to various application scenarios. Our experiments show an overall road precision of 89.99%, and recall of 84.77% for road materials classification. Figure 3 shows visual examples.

5. RESULTS AND DISCUSSION

We show the performance of our road extraction pipeline by testing on three different styles of cities: Timbuktu in Mali, Amman in Jordan and Aden in Yemen. All images are in 50 cm spatial resolution. The total testing area is 33.4 km². All the tests are conducted on a local PC with a CPU AMD K19, RAM 16G and GPU NVIDIA GeForce RTX 3060 with 4G memory. We evaluate the quality of our automatically extracted road networks both quantitatively and qualitatively.

5.1. Quantitative Statistics

Our Ground Truths are road polygons that depict along the road contours in the images. To evaluate the quality of our automatically reconstructed road networks, we manually created three sets of road networks as linear Ground Truths: Timbuktu in Mali, Amman in Jordan and Aden in Yemen. We buffer the generated road networks and the Ground Truths with a radius of 2 meters as a relaxation in all the comparisons, denoted as $\{r_{buff}\}$ and $\{gt_{buff}\}$. We demonstrate the geometrical and topological quality by analyzing the object-wise Precision, Recall, F1score and Hausdorff distance. The accuracy of the positions of our extracted roads is measured by calculating the average Hausdorff distance of the true positive roads and their matched references in the Ground Truth. Here are the definitions of our measure metrics:

true positive for each correctly generated road r_{buff} , at least 50% of its area is covered by $\{gt_{buff}\}$;

false positive for each redundantly generated road r_{buff} , less than 50% of its area is covered by $\{gt_{buff}\}$;

false negative for each missing Ground Truth road gt_{buff} , less than 50% of its area is covered by $\{r_{buff}\}$.

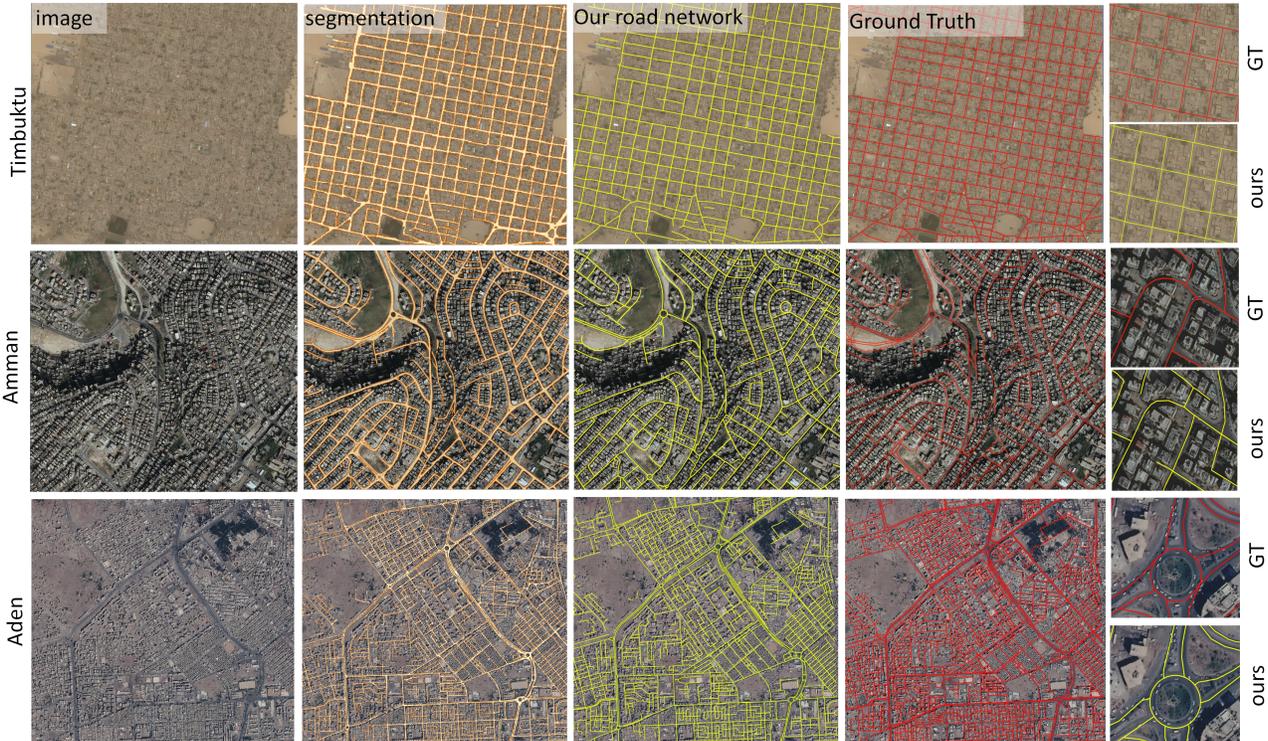


Fig. 4. Road networks generated by our pipeline over Timbuktu, Amman and Aden (top to bottom rows), with input satellite images, road segmentation, our extracted road networks, the Ground Truth, and crops (left to right columns). In the crops, Ground Truth are marked in red and our results in yellow.

city	GT Length	Precision	Recall	F1score	Hausdorff distance
Timbuktu	70.8	0.94	0.77	0.82	0.65m
Amman	351.5	0.86	0.77	0.81	0.58m
Aden	281.7	0.87	0.68	0.74	0.46m
Average		0.87	0.73	0.78	0.54m

Table 1. Quantitative evaluations on the three testing areas. *GT length* is the total length of the Ground Truth roads in kilometer. *Average* values in the last column are weighted averages that take into account of *GTLength*.

To further demonstrate the quality of our reconstructed road networks, we increase the buffer radius from 2m to 3m, relaxing the constraint of Hausdorff distance. This allows us to include more roads that are topologically matched with the Ground Truth, even if they are slightly shifted away from the centerline of roads. The average Precision, Recall, F1score are **0.93**, **0.79**, **0.85**, and the average Hausdorff distance is **0.86m**.

5.2. Qualitative Results

Figure 4 illustrates the visualization quality of our results, including input images, road segmentation, and extracted road networks. The crops demonstrate the regularity of the geometry and the potential of our method to preserve the correct topology.

6. CONCLUSIONS AND PERSPECTIVES

We propose a practical fully automatic pipeline for extracting road networks from VHR satellite images. Our pipeline provides geometrical regularity, topological correctness, and material classes. The neural network model we proposed labels road pixels with precise contours, enabling the proposed graph optimization method to generate clean road networks. Our method offers a practical and reliable solution for applications such as simulation that require strictly operational road networks. We aim to enhance the completeness of our road segmentation and continue exploring AI-aided solutions to tackle challenging scenarios, such as hyper-modern areas with elevated roads and bridges that are intertwined.

7. REFERENCES

- [1] Carsten Steger, Clemens Glock, Wolfgang Eckstein, Helmut Mayer, and Bernd Radig, "Model-based road extraction from images," in *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Armin Gruen, Olaf Kuebler, and Peggy Agouris, Eds., Basel, 1995, pp. 275–284, Birkhäuser Basel.
- [2] Christian Heipke, Helmut Mayer, C. Wiedemann, and Olivier Jamet, "Evaluation of automatic road extraction," *Inter. Arch. Photogramm. Remote Sens.*, vol. 32, 10 1997.
- [3] Ziyi Chen, Liai Deng, Yuhua Luo, Dilong Li, José Marcato Junior, Wesley Nunes Gonçalves, Abdul Awal Md Nurunnabi, Jonathan Li, Cheng Wang, and Deren Li, "Road extraction in remote sensing data: A survey," *International Journal of Applied Earth Observation and Geoinformation*, vol. 112, pp. 102833, 2022.
- [4] Renbao Lian, Weixing Wang, Nadir Mustafa, and Liqin Huang, "Road extraction methods in high-resolution remote sensing images: A comprehensive review," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 5489–5507, 2020.
- [5] Zhengxin Zhang, Qingjie Liu, and Yunhong Wang, "Road extraction by deep residual u-net," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 749–753, 2018.
- [6] Abolfazl Abdollahi, Biswajeet Pradhan, and Abdullah Alamri, "Vnet: An end-to-end fully convolutional neural network for road extraction from high-resolution remote sensing data," *IEEE Access*, vol. 8, pp. 179424–179436, 2020.
- [7] Lichen Zhou, Chuang Zhang, and Ming Wu, "D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.
- [8] Yang He, Ravi Garg, and Amber Roy Chowdhury, "Td-road: Top-down road network extraction with holistic graph construction," in *Computer Vision – ECCV 2022*, Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner, Eds., Cham, 2022, pp. 562–577, Springer Nature Switzerland.
- [9] Yong-Qiang Tan, Shang-Hua Gao, Xuan-Yi Li, Ming-Ming Cheng, and Bo Ren, "Vecroad: Point-based iterative graph exploration for road graphs extraction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
- [10] Gaetan Bahl, Mehdi Bahri, and Florent Lafarge, "Single-shot end-to-end road graph extraction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2022, pp. 1403–1412.
- [11] Gellert Mattyus, Wenjie Luo, and Raquel Urtasun, "Deeproadmapper: Extracting road topology from aerial images," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.
- [12] Nicolas Girard, Dmitriy Smirnov, Justin Solomon, and Yuliya Tarabalka, "Polygonal building segmentation by frame field learning," 2021.
- [13] Shusil Dangi, Cristian A. Linte, and Ziv Yaniv, "A distance map regularized cnn for cardiac cine mr image segmentation," *Medical Physics*, vol. 46, no. 12, pp. 5637–5651, Oct. 2019.
- [14] Ta-Chih Lee, Rangasami L Kashyap, and Chong-Nam Chu, "Building skeleton models via 3-d medial surface axis thinning algorithms," *CVGIP: Graphical Models and Image Processing*, vol. 56, no. 6, pp. 462–478, 1994.
- [15] Tongjie Y Zhang and Ching Y. Suen, "A fast parallel algorithm for thinning digital patterns," *Communications of the ACM*, vol. 27, no. 3, pp. 236–239, 1984.
- [16] Jaroslav Borovicka, "Circle detection using hough transforms documentation," *COMS30121-Image Processing and Computer Vision*, vol. 48, 2003.
- [17] S. Tripodi, N. Girard, G. Fonteix, L. Duan, W. Mapurisa, M. Leras, F. Trastour, Y. Tarabalka, and L. Laurore, "Brightearth: Pipeline for on-the-fly 3d reconstruction of urban and rural scenes from one satellite image," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. V-3-2022, pp. 263–270, 2022.