Detecting Roads from Satellite Imagery in the Developing World

Yoni Nachmany
Radiant Earth Foundation
Washington, D.C., USA
yoni.nachmany@radiant.earth

Hamed Alemohammad
Radiant Earth Foundation
Washington, D.C., USA
hamed@radiant.earth

Abstract

Advances in computer vision are improving the ability to accurately extract structured information from frequent and high-resolution satellite imagery, shedding light on global challenges and furthering Sustainable Development Goals. While these advances, along with increased availability of high capacity computational resources, result in improved models, lack of diverse training data significantly limits applications of these models to certain geographical regions. We review state-of-the-art models for road detection using satellite imagery, and compare predictions of two models (one trained in Las Vegas, USA and another in Khartoum, Sudan) in Khartoum. This comparison shows the need for regionally trained models using local training data. Finally, we outline a roadmap to use transfer learning and regional models in cities that do not have human verified labels.

1. Introduction

Accurate and up-to-date road maps are essential for several impactful applications globally. In particular, the lack of good maps in parts of the developing world results in inefficient resource allocation for urban planning and disaster response operations, as well as public health efforts such as vaccination campaigns.

High-resolution satellite imagery is a valuable source of data for detecting and mapping features such as roads and building footprints. Moreover, the frequent revisits of satellites help capture and record sudden changes that occur after major events or longer developments that occur over time. However, foundational mapping of these features is still a largely manual process of labeling and validating. Open crowdsourcing platforms such as OpenStreetMap (OSM) regularly compile labels from users globally, and organizations like the Humanitarian OSM Team mobilize volunteers for urgent situations, such as after natural disasters, to map these features using pre- and post-event satellite imagery. Such efforts are effective: since 2014, over 41,000 contributors have labeled 13 million kilometers of roads [15], and the OSM community started updating roads within 48 hours of the 2010 Haiti earthquake, ultimately producing the default map for other organizations [10]. However, crowdsourcing alone is not scalable: millions of kilometers of roads remain unmapped [15], and even with 5,300 mappers tracing 30,000 kilometers of roads in Puerto Rico after Hurricane Maria, it took a month to deliver the map of roads and buildings [25].

Computer vision and deep learning techniques can help provide faster updates over larger areas than solely crowdsourced or ground-based methods, especially in dynamic situations. Development Seed's team produced a machine learning model that accelerated professional human mapping of the electricity grid in Pakistan by a factor of around 15 to 20 [26]. However, the lack of geographically diverse training data is a barrier to applying these techniques in many developing countries, and two of the most popular public image datasets for pretraining models "exhibit an observable amerocentric and eurocentric representation bias," with better predictions for images in the West [20].

Spatial patterns of cities and towns, and consequently roads, as well as road types and pavements, are diverse globally. This results in a low prediction accuracy for models trained in one region and used in another region. Even within a country, differences exist, and Development Seed's model trained in one part of Pakistan made more errors on terrain types underrepresented in the training data [5].

In this study, we present how a model trained on labels from Las Vegas, USA fails to predict roads in Khartoum, Sudan. We then use human verified labels in Khartoum from the SpaceNet dataset [18], and build a road detection algorithm for Khartoum with hyperparameter tuning. We analyze the results of both models in Khartoum and discuss potential factors influencing performance. Lastly, we present the next steps required to fine-tune this model for road prediction in another city in Africa using transfer learning. The ultimate goal of this work is to accurately predict road segments in regions that lack human verified labels.

2. Previous Work

Automatically detecting roads in aerial imagery is challenging due to tree occlusions, building shadows, and varied atmospheric and ground conditions that can cause discontinuities in predictions, limiting their applicability for routing problems. The use of convolutional neural networks trained on large datasets to capture more spatial context of road networks dates back to Mnih and Hinton in 2010 [17], and in the last several years, better models [27] and datasets [24] have been introduced for road detection, as well as novel metrics and approaches that optimize for connectivity.

The DeepGlobe 2018 Satellite Image Understanding Challenge, which had a workshop co-located with CVPR 2018 and aimed to bridge computer vision research and remote sensing analysis, included a track on road extraction, which was formulated as a binary classification of pixels (i.e. semantic segmentation). The high-resolution data, captured at a ground sampling distance of 50 cm/pixel, covers a diverse sample of climates and topologies in urban and rural areas of Thailand, Indonesia, and India. For a different track on land cover classification, the DeepGlobe challenge featured the first public dataset of sub-meter imagery primarily from rural areas. The last track, on building detection, used data from the SpaceNet corpus, which hosts large swaths of labeled satellite imagery from Rio de Janeiro, Las Vegas, Paris, Shanghai, and Khartoum on Amazon Web Services (AWS) for free, and was the first challenge to include significant amounts of data from Asia and Africa [9].

The organizers of DeepGlobe, consisting of researchers from organizations like Facebook, DigitalGlobe, and the MIT Media Lab, ran baselines like those used in [10]. Those deep architectures included variants of VGG [21], U-Net [19], ResNet [13], and SegNet [3] (see Figure 1). In the process of tackling the problem that 75% of the world's roads lack adequate street addressing systems by creating a generative algorithm, Facebook researchers first created binary predictions of roads from three channel satellite images. A version of DeepLab [7] had higher precision and recall than SegNet, but showed signs of overfitting. The DeepGlobe dataset is larger and more diverse, and the organizers achieved the best baseline with a modified version of DeepLab with a ResNet18 backbone and Focal Loss [14]. The DeepGlobe road extraction challenge used a pixel-based Intersection over Union (IOU) score between the predicted and ground truth binary masks as their metric, and their baseline model achieved an IoU score of 0.545. The winner of the track reached a score of 0.6577 using a LinkNet with a pretrained encoder and dilated convolution [28].

The SpaceNet challenge on road extraction featured the same imagery as its building challenge: Level 2A atmospherically-corrected data from the DigitalGlobe WorldView-3 satellite in a continuous image strip, collected at 30 cm resolution pan-sharpened, in 3-band and 8-band GeoTiffs. Multispectral remote sensing data contributes to the richness of earth imagery, but AI models trained on everyday images do not handle the non-visible spectrum, though work on moving beyond RGB is ongoing [1].

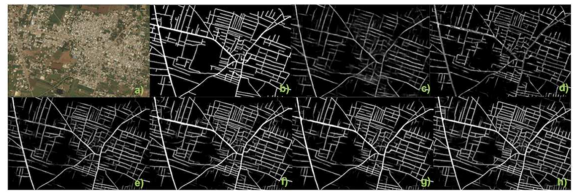

Figure 1. Comparison of NN Models. An example (a) satellite image and (b) ground truth; and road predictions using (c) VGG; (d) U-Net; (e) ResNet50; (f) ResNet101; (g) SegNet; and (h) DeepLab (source: [10]).

Unlike the DeepGlobe challenge, SpaceNet introduced a novel metric called Average Path Length Similarity (APLS) [22], which captures the difference between the predicted and ground truth graphs and penalizes discontinuity severely. The ground truth road centerlines were fully labeled and validated by experts, and annotated with road type, surface type, bridge, and lane number. The importance of label quality for connectivity was demonstrated by an experiment that showed that a model trained on SpaceNet achieved an APLS of 0.71 on Las Vegas, higher than the APLS of 0.59 from the model trained on OpenStreetMap[1].

SpaceNet's baseline algorithm [23] trained a U-Net to predict binary masks that were converted to graphs, which were post-processed for better connectivity, yielding an APLS of 0.49 [11]. The top 5 competitors followed a similar process [23], with the winner training an ensemble of segmentation models for a global road model that achieved an APLS of 0.6663 [11]. The highest performing city was Las Vegas, which received an APLS of 0.7977 for the winning model, and the lowest performing city was Khartoum, with an APLS of 0.6093[2], as dirt roads are challenging to detect and often confused with dirt paths [23] (see Figure 2).

A noteworthy submission came from Bastani, who co-authored RoadTracer and attempted to adapt it to the SpaceNet challenge, but later switched to an approach similar to DeepRoadMapper, albeit without its post-processing[3]. RoadTracer directly extracts road networks from aerial imagery, avoiding the noise introduced by the intermediary segmentation, using an iterative CNN-based search starting from a single known road location [4]. At a 5% error

[1] https://simplecore.intel.com/nervana/wp-content/uploads/sites/53/2018/05/IntelAIDC2018_Adam_Van_Etten_METROPOLIS_5_24.pdf

[2] https://github.com/SpaceNetChallenge/RoadDetector

[3] https://github.com/SpaceNetChallenge/RoadDetector/tree/master/fbastani-solution

Figure 2. Results from the winning implementation for the roads challenge. Top Left: A simple road network in a 400m 400m chip from the test set in Las Vegas; the blue line is the ground truth, the yellow line the proposal network, and the APLS score is 0.99. Top Right: A complex road network in Las Vegas; in the center of the graph network there is a disconnect where the divider is located. Bottom Left: A complex network in Shanghai; there are several missed streets in the center of the graph. Bottom Right: A low scoring road network in Khartoum; the proposal network misses several dirt roads, but performs well on the more established paved road network (source: [11]).

rate, RoadTracer captures 45% more junctions across 15 cities, mostly from North America and Europe, but including Tokyo. However, for SpaceNet data, the model struggled where roads were close together, such as parking lots. Bastani proceeded to implement DeepRoadMapper, out of the Uber Advanced Technologies Group. Sensors mounted on top of cars produce high definition but costly maps of cities at a small scale, whereas regular satellite imagery provides a more cost-effective, timely, and high coverage source of data. Rather than using complex heuristics to post-process segmentation results for connectivity, DeepRoadMapper frames the addition of missing connections as a shortest path problem with an efficient solution, using graph-theoretic notions similar to APLS. DeepRoadMapper proved effective on the TorontoCity dataset, but both the dataset and implementation are not publicly available [16].

Our focus lies in using the best open data and code to train state-of-the-art models that can be easily implemented and compared across standard metrics, in order to improve mapping on difficult but important areas in the developing world. Efforts have been made to predict road infrastructure quality in Africa [6], including work to classify the quality of roads in chips of high-resolution imagery from Kenya [6]. The generation of road maps, especially in rural areas, was identified as a critical application of machine learning by East African organizations, according to in-depth interviews with experts [8]. The challenges for road detection include rapid change due to construction and weather, the difficulty of identifying unpaved roads, and the lack of regional labeled data. The authors therefore note that "rapid frequent, satellite-imagery-based map making is thus a high value use of computer vision in East Africa."

3. Model and Training

DeepLab is a state-of-the-art deep learning model for semantic image segmentation, with the best results on the PASCAL VOC 2012 and Cityscapes datasets[4], as well as in a review of deep learning techniques applied to semantic segmentation [12]. DeepLab uses atrous convolution, also known as dilated convolution or convolution with upsampled filters, which enlarges the field of view of filters, integrating more context without increasing parameters or computation[5]. DeepLab also uses a Fully Connected Conditional Random Field (CRF), which allows for combining information about pixel class scores and the interaction between pixels, capturing long-range dependencies [12]. DeepLab-v3 introduced an encoder-decoder structure with the ability to control the resolution of encoder features.

Raster Vision, an open source Python framework for deep learning on satellite imagery, has support for TensorFlow's implementation of DeepLab built in, as well as configurations to support two network backbones: the faster MobileNet-v2 and the more powerful Xception-65[6]. Xception has relatively fewer parameters than VGG and Inception for faster training and inference[7]. The Xception-65 backbone is pretrained on ImageNet, with some modifications[8]. The MobileNet-v2 backbone is pretrained on the MS-COCO VOC 2012 augmented training set, with batch normalization already fine-tuned and no atrous spatial pyramid pooling or decoder modules for fast computation.

Raster Vision allows engineers to easily configure workflows for training models on large satellite images with a single file, outputting a bundled file for easy inference. The pipeline includes calculating dataset statistics, creating training chips, training, predicting, evaluating, and bundling. In addition, the Raster Vision examples code in-

[4] https://paperswithcode.com/task/semantic-segmentation

[5] http://liangchiehchen.com/projects/DeepLab.html

[6] https://docs.rastervision.io/en/0.8/misc.html#tensorflow-deeplab

[7] http://devseed.com/ml-grid-docs/methodology/2-machine-learning/

[8] https://github.com/tensorflow/models/blob/master/research/deeplab/g3doc/model_zoo.md

Base Learning Rate	Road F1 Score	Overall F1 Score
0.0001	0.276	0.274
0.001	0.567	0.690
0.002	0.580	0.693

Table 1. MobileNet-v2 hyperparameter tuning in Khartoum

Class	Precision	Recall	F1
Road	0.792	0.670	0.719
Background	0.883	0.761	0.778
Average	0.870	0.749	0.770

Table 2. Performance metrics for MobileNet-v2 in Khartoum

cludes a starting point for training SpaceNet models and tuning hyperparameters, as well as a bundled model file for Las Vegas (hereafter referred to as the Las Vegas model). The Las Vegas model was trained with the default hyperparameters of the MobileNet-v2 backbone and a learning rate of 0.001. In addition, other hyperparameters include a batch size of 8, a total of 100,000 steps, a chip size of 300, and 9 training chips per scene of 1300x1300 pixels[9].

We investigated the effect of network backbone for training a model using SpaceNet data in Khartoum, and for the selected network backbone, the learning rate was tuned, on a reduced number of steps (10,000 vs. 100,000) and scenes (128 vs. 283). At the default learning rate of 0.001, Xception-65 had a slightly higher F1 score for the road class than MobileNet-v2 (0.584 vs. 0.566) and an even smaller advantage for the overall F1 score. However, Xception-65 had a much larger total loss than MobileNet-v2 (0.5881 vs. 0.2891). Future work will explore the costs of bringing down the loss of Xception-65 to the level of MobileNet-v2, if attainable.

Given the MobileNet-v2 backbone, learning rates were tuned, and increasing the learning rate from 0.001 to 0.002 showed a slight improvement on the smaller experiment, though the full training yielded the best metrics with a learning rate of 0.001 for MobileNet-v2 (see Tables 1 and 2). We use the MobileNet-v2 with a learning rate of 0.001 as the base model in Khartoum (hereafter referred to as the Khartoum model) in the next section.

4. Discussion

The Las Vegas model is able to perform well on a variety of example scenes from Las Vegas, with diverse road patterns (Figure 3). Though a long building shadow in the top right image of Figure 3 prevents the successful prediction of a road, the model accurately predicts straight and curved roads between rows of homes, as well as roads around larger and smaller buildings. However, Khartoum data shows fundamentally different patterns, and the direct application of the Las Vegas model on Khartoum fails dramatically (Figure 4).

[9]https://github.com/azavea/raster-vision-examples

Figure 3. Four sample predictions in Las Vegas from the model trained in Las Vegas. Background shows the input imagery overlayed with the shaded road predictions from the model and red lines show the labels.

The Khartoum scene in Figure 4 on the left has one main asphalt road, whereas the other scenes include grids, lines, or more informal assortments of unpaved roads. In addition, the lighting, color, and arrangement of the Khartoum scenes is structurally different than those of Las Vegas, resulting from geographic, historical, and cultural factors. Even the regular grid in the third column cannot be detected by the Las Vegas model.

However, the Khartoum model almost perfectly learns the grid and still performs well on the less structured roads in the second column. Whereas the Las Vegas model could not reasonably predict roads in Khartoum, the Khartoum model produced a much more reasonable starting point for Khartoum and potentially other cities in Africa.

5. Conclusion

In this study we highlight state-of-the-art road detection models using satellite imagery, and investigate their performance in regions with limited human verified labels. In particular, we assess performance of a model in Sudan, which is trained using data in the United States. Diversity of urban texture results in very low prediction performance for such a model. However, using a set of labels from the SpaceNet dataset in Sudan, we train a model that is able to capture the spatial pattern of roads and successfully predict on new images.

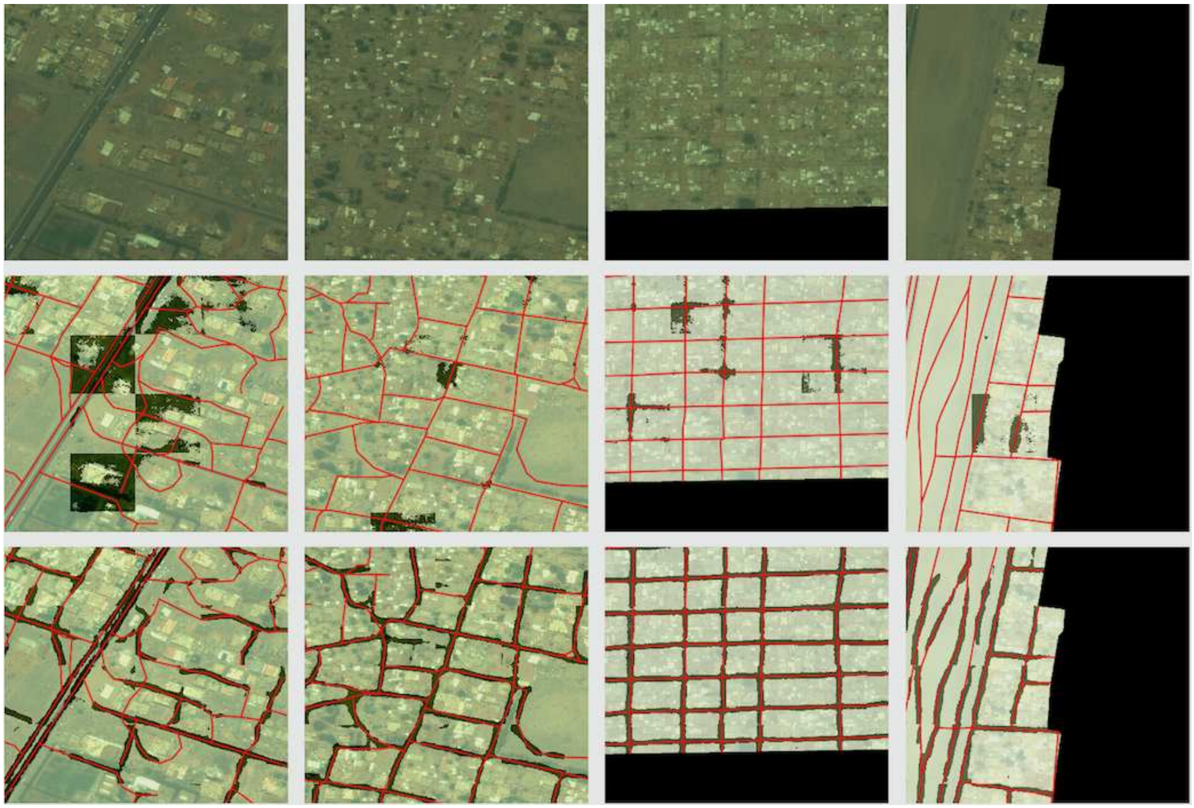

Figure 4. Prediction results in Khartoum for four different scenes. Top: input imagery, middle: prediction results (shaded) from the model trained in Las Vegas overlayed with labels (red lines) on top of the input imagery, bottom: prediction results (shaded) from the model trained in Khartoum overlayed with labels (red lines) on top of the input imagery.

In this study, we use the Raster Vision package, an open source deep learning library in Python that is designed for large scale training and predictions jobs using aerial imagery. Raster Vision has the experimental capacity to generate labels from vector tiles created from OSM data. OSM is a global dataset of crowdsourced labels that has been used in contests like ISPRS [2] and SpaceNet Buildings[10]. Future work will involve training and fine-tuning models with OSM data in other areas of interest; particularly, using the Khartoum model as a base and retraining with OSM data.

[10] `https://i.ho.lc/winning-solution-for-the-spacenet-challenge-joint-learning-with-openstreetmap.html`

References

[1] Nicolas Audebert, Bertrand Le Saux, and Sébastien Lefèvre. Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 140:20–32, jun 2018.

[2] Nicolas Audebert, Bertrand Le Saux, and Sébastien Lefèvre. Joint learning from earth observation and openstreetmap data to get faster better semantic maps. *CoRR*, abs/1705.06057, 2017.

[3] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12):2481–2495, dec 2017.

[4] Favyen Bastani, Songtao He, Sofiane Abbar, Mohammad Alizadeh, Hari Balakrishnan, Sanjay Chawla, David J. DeWitt, and Sam Madden. Unthule: An incremental graph construction process for robust road map extraction from aerial images. *CoRR*, abs/1802.03680, 2018.

[5] Drew Bollinger. Geo-diversity for better, fairer machine learning. `https://developmentseed.org/blog/2018/03/19/geo-diversity/{%}0A`, 2018.

[6] Gabriel Cadamuro, Aggrey Muhebwa, and Jay Taneja. Assigning a grade: Accurate measurement of road quality using satellite imagery. *CoRR*, abs/1812.01699, 2018.

[7] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):834–848, apr 2018.

[8] Milan Cvitkovic. Some requests for machine learning research from the east african tech scene. *CoRR*, abs/1810.11383, 2018.

[9] Ilke Demir, Krzysztof Koperski, David Lindenbaum, Guan Pang, Jing Huang, Saikat Basu, Forest Hughes, Devis Tuia, and Ramesh Raskar. Deepglobe 2018: A challenge to parse the earth through satellite images. *CoRR*, abs/1805.06561, 2018.

[10] lke Demir, Forest Hughes, Aman Raj, Kaunil Dhruv, Suryanarayana Murthy Muddala, Sanyam Garg, Barrett Doo, and Ramesh Raskar. Generative street addresses from satellite imagery. *ISPRS International Journal of Geo-Information*, 7(3), 2018.

[11] Adam Van Etten, Dave Lindenbaum, and Todd M. Bacastow. Spacenet: A remote sensing dataset and challenge series. *CoRR*, abs/1807.01232, 2018.

[12] Alberto Garcia-Garcia, Sergio Orts-Escolano, Sergiu Oprea, Victor Villena-Martinez, and José García Rodríguez. A review on deep learning techniques applied to semantic segmentation. *CoRR*, abs/1704.06857, 2017.

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[14] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollar. Focal loss for dense object detection. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[15] David Lindenbaum. Introducing the SpaceNet Road Detection and Routing Challenge and Dataset. `https://medium.com/the-downlinq/introducing-the-spacenet-road-detection-and-routing-challenge-and-dataset-7604de39b779`, 2017.

[16] Gellert Mattyus, Wenjie Luo, and Raquel Urtasun. Deeproadmapper: Extracting road topology from aerial images. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[17] Volodymyr Mnih and Geoffrey E. Hinton. Learning to detect roads in high-resolution aerial images. In Kostas Daniilidis, Petros Maragos, and Nikos Paragios, editors, *Computer Vision – ECCV 2010*, pages 210–223, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

[18] SpaceNet on Amazon Web Services (AWS). Datasets. the spacenet catalog. Last modified april 30, 2018. `https://spacenetchallenge.github.io/datasets/datasetHomePage.html`, 2018. [Accessed on Mar 1st, 2019].

[19] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Nassir Navab, Joachim Hornegger, William M Wells, and Alejandro F Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[20] Shreya Shankar, Yoni Halpern, Eric Breck, James Atwood, Jimbo Wilson, and D. Sculley. No classification without representation: Assessing geodiversity issues in open data sets for the developing world. In *NIPS 2017 workshop: Machine Learning for the Developing World*, 2017.

[21] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

[22] Adam Van Etten. SpaceNet Road Detection and Routing Challenge Part I. `https://medium.com/the-downlinq/spacenet-road-detection-and-routing-challenge-part-i-d4f59d55bfce`, 2017.

[23] Adam Van Etten. Broad Area Satellite Imagery Semantic Segmentation (BASISS). `https://medium.com/the-downlinq/broad-area-satellite-imagery-semantic-segmentation-basiss-4a7ea2c8466f`, 2018.

[24] Shenlong Wang, Min Bai, Gellért Máttyus, Hang Chu, Wenjie Luo, Bin Yang, Justin Liang, Joel Cheverie, Sanja Fidler, and Raquel Urtasun. Torontocity: Seeing the world with a million eyes. *CoRR*, abs/1612.00423, 2016.

[25] Nick Weir. Towards accelerating disaster response with automated analysis of overhead imagery. `https://towardsdatascience.com/towards-accelerating-disaster-response-with-automated-analysis-of-overhead-imagery-18c65731eaaf`, 2019.

[26] Mark Wronkiewicz. Mapping the electric grid. `https://medium.com/devseed/mapping-the-electric-grid-fe29f041d54e`, 2018.

[27] Z. Zhang, Q. Liu, and Y. Wang. Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):749–753, May 2018.

[28] Lichen Zhou, Chuang Zhang, and Ming Wu. D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.

A. Open Source Code for Transfer Learning

The inspiration for an open source exploration of deep transfer learning in the developing world arose from good initial results of a road detection model trained on the SpaceNet dataset and fine-tuned in Dar es Salaam with DigitalGlobe imagery and OpenStreetMap labels, for a FOSS4G 2018 demo from Development Seed[11]. Training data was prepared with Label Maker, which downloads and tiles OpenStreetMap information and satellite imagery[12]. The Pixel Decoder library trained a ResNet50 U-Net for four checkpoints with combinations of dice and log loss[13]. We also investigated the fast.ai library, to create a similar ResNet50 U-Net architecture for semantic segmentation. Raster Vision provided a clearer pipeline for pre-training on SpaceNet Khartoum and fine-tuning on Ghana OSM data.

[11] `https://github.com/Geoyi/FOSS_4g_Pixel_decoder`
[12] `https://github.com/developmentseed/label-maker`
[13] `https://github.com/developmentseed/label-maker`

Raster Vision facilitates training a model on geospatial data with the TensorFlow DeepLab backend using AWS Batch[14]. Another library, RoboSat, works with Slippy Map tiles to convert between geo-referenced data and fixed-size images. RoboSat can turn segmentation probabilities and masks into clean and simple geometries, even de-duping with OSM[15]. Raster Vision's segmentation can now generate polygons[16]. APLS can be used with vectors, but we focus on precision, recall, and F1 of rasters, comparing RoboSat's ResNet50 U-Net with Lovasz loss to Raster Vision's DeepLab model. We hand-labeled part of a test scene in Ghana for validation.

[14] https://docs.rastervision.io/en/0.9/
[15] https://github.com/mapbox/robosat
[16] https://www.azavea.com/blog/2019/04/15/raster-vision-0-9-release-candidate/