Road Detection from Remote Sensing Imagery

Pantelis Kaniouras
student #4944089
P.Kaniouras@student.tudelft.nl

1st supervisor: Liangliang Nan — liangliang.nan@tudelft.nl
2nd supervisor: Roderik Lindenbergh — r.c.lindenbergh@tudelft.nl
external supervisor: Frido Kuijper — frido.kuijper@tno.nl

January 10, 2020

1 Introduction

The road network is the core and essential mode of transportation, providing a variety of services to the human civilization. Possession of detailed and reliable digital road network datasets can support a wide number of applications such as vehicle navigation, urban planning, updating of geographic information systems, traffic management, road monitoring, crisis response, disaster management and many more. Up until recently, digital road network datasets were created in a semi-automatic way followed by meticulous manual extraction [23]. It is a time consuming, expensive and labor-intensive procedure. As a consequence, it is unfeasible to satisfy the needs of modern times with manual work, given the popularity of location-based services and applications. In order for those activities to maintain their continuous functionality, an automated method is needed to acquire and update digital datasets of the dynamic road network structure.

Technological advances in the field of remote sensing offer the opportunity for spatial information extraction from an abundance of high-resolution image data that faithfully depict the earth's surface [32]. Although scientists have been trying to solve the problem of road detection for more than 30 years now [1], there hasn't been a flawless technique that can generalize the desired output under all different situations that occur in the built environment. The reason is that road networks are intricate structures. Their observation in remote sensing imagery can be either locally blocked by a variety of objects, like trees, buildings, cars and other street furniture or confused by a neighbouring location that has similar texture. Furthermore, their intensity pixel values can vary due to the difference of atmospheric conditions, seasonality of data acquisition and shadows of objects. A plethora of factors and their interrelation has to be taken into consideration to create a robust road extraction algorithm.

Recently, as a state-of-the-art machine learning technique, deep learning [16] made a major breakthrough in conventional computer vision tasks such as image classification, object detection, semantic segmentation and instance segmentation [20, 30, 13, 31, 11, 12, 38, 45, 7, 8, 6, 14, 33, 35, 29]. As a result, a large number of researchers started using deep learning techniques to solve remote sensing problems [21], as well as the road detection task [41]. Because of its supremacy of modeling complex nonlinear relationships between variables, deep learning surpassed conventional road detection algorithms [40]. However, the entire automation of the road extraction procedure is still not feasible. Extracted road networks using deep learning techniques frequently contain noise, artefacts, isolated road segments or miss information, making them inadequate for real-world applications, requiring significant amount of manual labor to correct the errors [3].

Road extraction with deep learning using remote sensing data is usually posed as either a semantic segmentation task, where each pixel of a remote sensing image is classified according to the class that it belongs -road or non-road-, or as a road center-line vector graph extraction, both followed by post-processing steps to refine their results. The main advantage of semantic segmentation approaches is the preservation of geometric properties of the road network, since every classified road image pixel corresponds to a specific spatial extent, while the main advantage of the vector graph extraction approaches is the preservation of topological properties, owing to their design to preserve road connectivity. Apart from the aforementioned difficulties and properties of the road network structure that complicate the procedure of road detection, inaccuracies using deep learning techniques appear for two more, but very important reasons. First, because topology is generally ignored during the pixel-wise semantic segmentation task, meaning that every pixel is handled individually and second, because

ground truth or reference data used to train the deep learning models contain inaccuracies (label scarcity , omission, noise) [36, 26], which unfortunately mislead neural networks into making incorrect estimations.

Road networks are structures for which prior knowledge exists. For example, it is known that roads have consistent width, are almost always made of concrete and are continuous (i.e. any location of the network should be able to be reached by any other location of the network). Existing approaches haven't utilized all the properties and known information of the road network to enhance their training phase. The focus of this research is to utilize this knowledge and retain both topological and geometrical properties of the road network, as much as possible, meliorating current road detection algorithms. The collection of each property is equally important and essential for all applications. When combined, they can provide a structured and detailed representation of the road network, useful for safe localization and motion planning. In order to achieve that, I intend to combine the two different deep learning techniques used for road detection (semantic segmentation and vector road-center-line graph extraction) into one, unified and novel model, using the Multi-task learning approach [19]. Enriching the amount of context that each task receives for its training phase should improve the performance of each task, and thus, improve the performance of road detection.

1.1 Scientific relevance

The topic of road detection using deep learning techniques is highly correlated with the scientific field of the Geomatics for the Built Environment study program. A robust and working algorithm will allow the acquisition of essential geographical knowledge about the built environment by employing an innovative and advanced technique.

2 Related work

As pointed out earlier, many algorithms have been developed to extract road networks from remote sensing images. To confine the range of the literature review, only pioneering approaches that explore deep learning techniques were examined.

2.1 Pioneer Road Detection approaches

Mnih and Hinton [27] used restricted Boltzmann machines (RBMs) to detect road areas from high resolution aerial images. To improve their results, they applied both a pre-processing and a post-processing step. During the pre-processing step, they reduced the dimensionality of the input data, while during the post-processing, they removed disconnected blotches and filled in the gaps or holes in the roads. Zhang et al. [43] created a Deep Residual U-Net model, by extending the U-Net architecture [33], adding short-cut connections between the CNN layers, producing a semantic segmentation output depicting the road network. Cheng et al. [9] created a cascaded end-to-end CNN that both detects and extracts road center-lines. Their method is divided into two networks, one dealing with the road detection and one with the center-line extraction exploiting information collected during the inferring process of the first network. In the final step, a thinning algorithm is applied to refine the extracted center-line network. Buslaev et al. [4] proposed a fully convolutional neural network consisting of a ResNet-34 pre-trained on ImageNet encoder and a decoder similar to that from a vanilla U-Net model. Additionally, they designed a loss function that simultaneously considers binary cross entropy loss and intersection over union (IoU) loss to improve their predictions. Another study that utlized U-Net is the work of Sun et al [37], who created a model using stacked

U-Nets with multiple output. They also incorporated a hybrid loss function to confront the issue of unbalanced classes of training data. To improve performance, they also implemented post-processing steps, like shorterst path search with hierarchical thresholds. Zhou et al. [44] designed a semantic segmentation model, called D-LinkNet, based on an encoder-decoder structure, dilated convolution to enlarge the receptive field of the feature points maintaining resolution and pretrained encoder especially for the road detection task. The D-LinkNet was the winner of the DeepGlobe 2018 road challenge [10]. Mattyus et al. [22] created a method that improves road connectivity with post-processing steps after the prediction stage, which is based on a CNN segmentation, applying heuristics to missing connections or isolated road segments. Although it is a method that produces excellent results when the prediction output is accurate, it doesn't perform well when it has many errors, which is the usual case due to occlusion, shadows and reasons explained above. Mosinska et al. [28] proposed the topology-aware loss function, which is a new loss function for delineation of curvilinear structures. Replacing the regular pixel-wise loss function, e.g. cross entropy, with their own in a U-Net model, topology structure information is taken into account and better road extraction performance is achieved. Ventura et al. [39] proposed a method, built on top of a semantic segmentation classification, that iteratively connects road segments inferred in neighbouring image patches, thus resulting in a fully connected network. Li et al. [17] proposed PolyMapper, an algorithm for direct extraction of topological maps as a collection of building footprints and roads, using raw aerial imagery. They combined the principle of a maze solving algorithm, commonly known as the left-hand or right-hand rule, with a CNN-RNN deep learning model that first detects the points of interest, like intersection points or seed points, and then sequentially connects them according to their conditional probability distribution to belong to the road network.

2.2 Baseline methods

During the process of the literature review, two studies were distinguished by their performance and ability to maintain specific properties of the road network, those of topology and geometry. These two studies will be used as inspiration for the creation of my own road detection algorithm as well as baselines for its evaluation. The following sections give a detailed description of their functionality as the sub-tasks of my own deep learning model will have the same logic.

2.2.1 RoadTracer: Automatic Extraction of Road Networks from Aerial Images

RoadTracer [2] is an algorithm that directly constructs a graph representation of the road network using an iterative search procedure controlled by a CNN-based decision function. The construction of the road network graph starts from a seed location known to be on the road network, and sequent points are added according to the search procedure. The decision function is invoked at each step of the search to figure out the next best action to take: either add and walk towards a new node to the road network, or return to the previous node and continue from there.

More specifically, the search algorithm starts with an input of (v_0, B), where v_0 is the seed location node, and B is a bounding box that defines the area of interest. The search algorithm also keeps a graph G and a stack of nodes S, that both initially contain only v_0. The current location of the search algorithm is represented by S_{top}, which is the top node of the stack S. On every step, the decision function uses graph G, S_{top} and an image centered at S_{top}'s location to determine if it should walk a fixed distance D (authors of the paper used $D = 12$) forward

from S_{top}, keeping a certain direction, or go back to node S_{top-1}. The direction is also selected by the decision function from a set of known a angles distributed in $[0, 2\pi)$. The new location is added as a node onto S, and as a node along with an edge on the graph G. If the decision function chooses to stop, S_{top} is popped from S. When S is emptied, the road network graph G is complete.

The CNN decision function is evidently the most crucial part of the RoadTracer method, since all the aforementioned actions originate from it. It is implemented with a CNN, which takes the centered RGB image at S_{top}'s location as input, concatenated with one additional channel. This fourth channel is the graph G, as it is constructed until the current step. G is rendered by drawing anti-aliased lines along the edges of the graph that are presented withing the centered image extend. The output has two components. The first one is the final action to be taken, either walking or stopping, and the second one is the angle that describes the direction of the walking. The decision function output holds the only post-processing step that is required for this method, which is the selection of the threshold value that will distinguish the outputs of either walking or stopping, which has to be selected manually.

To train the CNN-based decision function, training examples were dynamically generated by running the search algorithm as the decision function during training. Given an input region (v_0, B), training starts by initializing an instance of the search algorithm G, S. On every individual training step, like during inference, the CNN has to decide on an action based on the constructed output, and update G and S based on that decision. The ground truth of that action is determined according to how an "oracle" decision function that makes optimal decisions using the ground truth graph $G*$ would react. The CNN corrects itself and trains to learn that action.

The algorithm contains some small improvement steps such as a simple merging step of earlier explored paths and guidance when walking outside of the bounding box area. The main limitations of this approach are its inability to keep geometrical properties of the road, like shape and width, since it directly constructs a vector graph representation of the road network, the creation of abundant vertices and edges, increasing the complexity factor of the graph, its tendency to fail at intersections, at roads with high curvature and finally, with greater length. This happens because the road graph construction is based on the sequential decisions made by the CNN decision function. Once the CNN makes a mistake, the algorithm will consequently produce incorrect road segments or won't succeed to recognize existing road segments.

2.2.2 Improved Road Connectivity by Joint Learning of Orientation and Segmentation

Batra et al. [3] developed a two stage road detection method to enhance the connectivity of the extracted road network. During the first stage, a joint learning module by stacking multi-branch encoder-decoder structure is implemented, aiming to allow the flow of information between two related tasks, those of per-pixel road segmentation and road orientation. During the second stage, a connectivity refinement model is applied, to connect small gaps and remove false positive occurrences.

The additional task of orientation learning introduces a connectivity constraint in the encoded representation, due to the fact that accurate road orientation predictions foster connected, and not isolated, road segments. Joint learning of related tasks (i.e. road segmentation and road orientation) results to more generalizable features [5, 15]. In order to train the model to learn

how to identify the road orientations, orientation ground-truths were created from the reference road vector data. To keep it simple and consistent, driving directions were ignored and road orientation was always calculated as a vector starting from left to right and from top to bottom as a unit vector in sequential pixels.

Although orientation supervision increases the connectivity in the predicted road network, intricate situations such as bridges, parking lots and highway cross-roads cause inaccurate or fail to produce orientation predictions. This was the motivation to create the connectivity refinement step, which considers missing or fake road segments as corrupted ground-truth masks, that will be restored by the pre-trained refinement network. In pre-training phase, an input remote sensing image, a corrupted ground truth mask and the previous prediction image are concatenated and feeded as input to the refinement model. Ultimately, the neural network is trained to encode available context, filling missing road segments. Later on, the pre-trained model is further fine-tuned by replacing the manually corrupted ground truth masks with the segmentation outputs.

The stacked multi-branch module has three parts. The first part is the shared encoder, the second one is the repetitive fusion with multi-branch and the third one is the prediction branches of each task. This module simultaneously learns a robust common representation in the shared encoder, creates predictions for segmentation and orientation of roads and permits the information flow from one task to the other, enhancing road connectivity. The capability of producing intermediate outputs at different scales, allows to use multi-scale loss function to guide the model training.

3 Research questions

3.1 Objectives

The main research question for this study is:

> *To which extent is it possible to utilize topological and geometrical constraints to improve road detection using deep learning techniques?*

The goal of this research is to examine how topological and geometrical prior knowledge can assist the remote sensing task of road detection using deep learning techniques. In order to do that, a new deep learning model will be designed, able to exchange and leverage information between different sub-tasks, thereby benefiting from the capabilities of each sub-task. The following sub-questions will be relevant:

- *Can geometric prior knowledge improve road detection?*
- *Can topological prior knowledge improve road detection?*
- *How to incorporate prior knowledge or geometric constraints into a deep learning model?*
- *How to combine two different deep learning models with different architecture and aim into one, unified model?*
- *What are the limitations of a model that combines two different models into one, unified model?*

3.2 Scope of research

This thesis will focus on the development of a new deep learning based approach incorporating prior knowledge and geometric constraints for road detection, which is expected to outperform state-of-the-art methods (i.e., the two baseline approaches). Furthermore, this method will be limited for evaluation only on the selected datasets described in 6.2. If there is enough time, the extension of the proposed model's capabilities will be explored.

4 Methodology

In this study, two models that were described in section 2.2 will be used as inspiration for the creation of a new and novel model to solve the task of road detection. Since each model has different architecture and design, it is not possible to create a new model using one as a skeleton and exploit their strongest points, aiming to surpass their performance. In order to combine their abilities into one, unified approach, a new model inspired by the baseline methods and Multi-task learning will be used.

4.1 Overview of Multi-task learning in Deep Neural Networks

Multi-Task learning, as the name betrays, aims to solve multiple tasks simultaneously, by taking advantage of the relationships between different tasks. The main principle is, that if there are n tasks with an acceptable degree of relation between them (not all tasks need to be related to each other), Multi-Task Learning (or MTL) will improve the performance of a model -i.e. of each task-, by utilizing the knowledge obtained from every participating task [42].

MTL is beneficial because a more generalized representation of the features or attributes of the input data can be learnt by a multi-task learning model, than by a normal or single-tasked deep learning model. Normal deep learning models aim to optimize their objective function by fine-tuning hyperparameters till the performance of a single task cannot be further increased. In a multi-task learning model hyperparameters are fine-tuned according to an objective function that considers the loss functions of auxiliary tasks, aiming to increase their accuracy rates concurrently.

Apart from its generalization property, MTL leads to better results also due to the regularization ability expressed by inducing bias into the models. When trying to optimize multiple tasks, hypotheses that favor all the n tasks are preferred and thus, the risk of overfitting is significantly reduced, while the model becomes more capable on handling random noise during training [5].

The two most commonly used techniques to apply MTL in Deep Neural Networks are the Hard Parameter Sharing and the Soft Parameter Sharing (of hidden layers) techniques. In Hard Parameter Sharing, a common hidden layer(s) (e.g. a common encoder) is used for all the auxiliary tasks. Towards the end of the model, several layers are dedicated to exclusive tasks. In Soft Parameter Sharing layers are not shared. Each participating model has its own sets of weights and biases. These parameters are regularized in order to become more similar and able to represent each task individually. Intermediate extracted features or end results can be shared between different tasks to enhance the performance of the multi-task model [34].

4.2 Multi-task learning applied to Road Detection

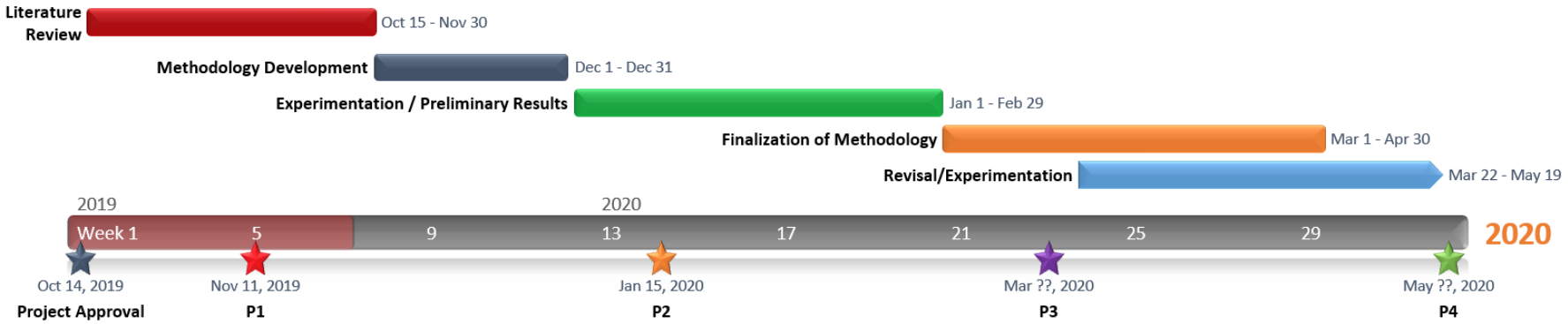
To solve the problem of road detection, MTL will be used to combine a model aiming to solve the task of road centerline vector graph extraction with a model aiming to solve the task of semantic segmentation. The first task represents the preservation of topology, and the second the preservation of geometry or road surface. The architecture and aim of each model is different and thus, Hard Parameter Sharing is not recommended as an initial option.

In this study, the first technique that will be used is the Soft Parameter Sharing. In the early stages of the new model development, the initial models that will keep their architecture and will be assisted either by the extracted intermediate or end results of the other model. To mitigate common drawbacks of Multi-task Learning and especially, those of Soft Parameter Sharing, counter-measures will be taken. Specifically, in order to let the model determine in what way the task-specific networks leverage the knowledge of the other task, the work of Misra et al [24] could be used, who answered this problem by creating the "cross-stitch units" that learn a linear combination of the output of the previous layer. Another very good idea was that of Kendal et al. [15], where they noticed that the performance of multi-task learning systems is highly dependent on the relative weighting between each task's objective function or loss function. Therefore, they proposed a method that weighs multiple loss or objective functions by taking into consideration the homoscedastic uncertainty of each task, overcoming the need of manual setting.

However, since deep learning models are problem-based, success is not guaranteed. Depending on the model's performance, different solutions might be implemented, such as the creation of a brand new neural network using reasoning and algorithms of the two baseline neural networks, or the extension of the multi-task learning model by adding another few layers after the output layers of each tasks, aiming to combine their outputs.

5 Time planning

The following Gantt chart illustrates the thesis's schedule, mainly focused on the algorithm's implementation. Documentation will coincide with the development of the algorithm, to avoid loss of information and record every step in detail. The exact dates of P3 and P4 will be determined in the future.



6 Tools and datasets used

6.1 Programming language

The programming language used for the implementation of the proposed model will be the Python Programming Language. It has a concise, simple and readable syntax, is platform independent, supported by a wide community and offers a wide variety of reliable libraries and frameworks, dedicated to AI-based projects. It is the best option to increase productivity and efficiency during the implementation of a prototype model.

6.2 Data

Three challenging datasets that have been widely used by researchers trying to solve the problem of road network detection will be used. Their selection accounts for the fact that they were created by well-trained engineers, making them a reliable choice, and offer the chance to compare results of the proposed methodology with results provided by other scientists.

6.2.1 DeepGlobe

The DeepGlobe road dataset was created for the DeepGlobe Road Extraction Challenge [10]. The dataset contains 6226 RGB satellite images, collected by DigitalGlobe's satellite from three different areas: Thailand, Indonesia, and India. 4696 images are destined for training purposes and 1530 for validation. Their format is GeoTiff, with size 1024x1024 pixels and ground resolution of 50cm/pixel. Each satellite image is paired with a mask image for road labels. The mask is a gray-scale image, with white standing for road pixel, and black standing for background. Mask images are not flawless, because manual annotation of images is rather costly, especially in urban regions. The creators of the datasets mention that small roads within farmlands were intentionally not annotated.
In order to use the dataset for the experiments, the mask images were binarized with a threshold of 128, because their values weren't pure 0 and 255. Furthermore, image data augmentation was applied on-the-fly, to artificially enrich the dataset with multiple images.

Figure 1: Example image from the DeepGlobe Dataset

6.2.2 SpaceNet

The SpaceNet road dataset was created for the SpaceNet Challenge: Road Extraction and Routing [18]. The dataset contains 2567 satellite images, collected by DigitalGlobe's satellite from four different cities: Paris, Las Vegas, Shanghai, and Khartoum. Their format is GeoTiff (16-bit), with size 1300x1300 pixels and ground resolution of 30cm/pixel. Their corresponding road network ground truth data is provided in the form of vector data (line-strings), representing the center line of roads. Furthermore, the road labels correspond to different road types (Motorway, Primary, Secondary, Tertiary, Residential, Unclassified, Cart Tracks) from the four cities, having diverse road widths and visual appearance. Each image may have multiple line-strings and each line- string consists of pixel coordinates X Y depicting road center line points in the 2D image plane, assuming top-left corner as the origin.
In order to use the dataset for the experiments, the dataset was divided into 2000 images for training and 567 for testing, after converting all images from 16-bit to 8-bit format. The division of the dataset was done in a way that each city would equally contribute to train and test the algorithm (80% - 20% respectively). Furthermore, image data augmentation was applied on-the-fly, to artificially enrich the dataset with multiple images.

Figure 2: Example image from the SpaceNet Dataset

6.2.3 Massachusetts Roads Dataset

The Massachusetts Roads Dataset was built by Mnih et al. [25]. It was the first publicly available dataset for convolutional neural network training addressing the road detection problem. It consists of 1171 aerial images of the state of Massachusetts, from which 14 were randomly selected for validation and 49 for testing by the creator. Their format is GeoTiff, with size 1500x1500 and ground resolution of 1m/pixel. It covers an area of approximately 2600 square kilometers in total, and introduces a wide variety of urban, suburban, and rural regions. The mask images for road labels were generated by rasterizing road center lines obtained from the OpenStreetMap project. According to the creator, a thickness of 7 pixels and no smoothing was used.
In order to use the dataset for the experiments, image data augmentation was applied on-the-fly, to artificially enrich the dataset with multiple images and make them usable for the network.

Figure 3: Example image from the Massachusetts Roads Dataset

References

[1] Ruzena Bajcsy and Mohamad Tavakoli. "Computer Recognition of Roads from Satellite Pictures". In: *IEEE Transactions on Systems, Man, and Cybernetics* 6 (1976), pp. 623–637.

[2] Favyen Bastani et al. "Unthule: An Incremental Graph Construction Process for Robust Road Map Extraction from Aerial Images". In: *CoRR* abs/1802.03680 (2018). arXiv: 1802.03680. URL: http://arxiv.org/abs/1802.03680.

[3] Anil Batra et al. "Improved Road Connectivity by Joint Learning of Orientation and Segmentation". In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2019.

[4] Alexander Buslaev et al. "Fully Convolutional Network for Automatic Road Extraction from Satellite Imagery". In: June 2018, pp. 197–1973. DOI: 10.1109/CVPRW.2018.00035.

[5] Rich Caruana. "Multitask Learning: A Knowledge-Based Source of Inductive Bias". In: *ICML*. 1993.

[6] Liang-Chieh Chen et al. "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs". In: *CoRR* abs/1606.00915 (2016). arXiv: 1606.00915. URL: http://arxiv.org/abs/1606.00915.

[7] Liang-Chieh Chen et al. "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation". In: *CoRR* abs/1802.02611 (2018). arXiv: 1802.02611. URL: http://arxiv.org/abs/1802.02611.

[8] Liang-Chieh Chen et al. "Rethinking Atrous Convolution for Semantic Image Segmentation". In: *CoRR* abs/1706.05587 (2017). arXiv: 1706.05587. URL: http://arxiv.org/abs/1706.05587.

[9] Guangliang Cheng et al. "Automatic Road Detection and Centerline Extraction via Cascaded End-to-End Convolutional Neural Network". In: *IEEE Transactions on Geoscience and Remote Sensing* PP (Mar. 2017), pp. 1–16. DOI: 10.1109/TGRS.2017.2669341.

[10] Ilke Demir et al. "DeepGlobe 2018: A challenge to parse the earth through satellite images". In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. Vol. 2018-June. May 2018, pp. 172–181. ISBN: 9781538661000. DOI: 10.1109/CVPRW.2018.00031. arXiv: 1805.06561. URL: http://arxiv.org/abs/1805.06561%20http://dx.doi.org/10.1109/CVPRW.2018.00031.

[11] Ross B. Girshick. "Fast R-CNN". In: *CoRR* abs/1504.08083 (2015). arXiv: 1504.08083. URL: http://arxiv.org/abs/1504.08083.

[12] Ross B. Girshick et al. "Rich feature hierarchies for accurate object detection and semantic segmentation". In: *CoRR* abs/1311.2524 (2013). arXiv: 1311.2524. URL: http://arxiv.org/abs/1311.2524.

[13] Kaiming He et al. "Mask R-CNN". In: *CoRR* abs/1703.06870 (2017). arXiv: 1703.06870. URL: http://arxiv.org/abs/1703.06870.

[14] Simon Jégou et al. "The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation". In: *CoRR* abs/1611.09326 (2016). arXiv: 1611.09326. URL: http://arxiv.org/abs/1611.09326.

[15] Alex Kendall, Yarin Gal, and Roberto Cipolla. "Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics". In: *CoRR* abs/1705.07115 (2017). arXiv: 1705.07115. URL: http://arxiv.org/abs/1705.07115.

[16] Yann LeCun, Y. Bengio, and Geoffrey Hinton. "Deep Learning". In: *Nature* 521 (May 2015), pp. 436–44. DOI: 10.1038/nature14539.

[17] Zuoyue Li, Jan Dirk Wegner, and Aurélien Lucchi. "Topological Map Extraction from Overhead Images". In: (2018). arXiv: 1812.01497. URL: http://arxiv.org/abs/1812.01497.

[18] Dave Lindenbaum and Todd Bacastow. "SpaceNet : A Remote Sensing Dataset and Challenge Series". In: (). arXiv: arXiv:1807.01232v3.

[19] Shikun Liu, Edward Johns, and Andrew J Davison. "End-to-End Multi-task Learning with Attention". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019, pp. 1871–1880.

[20] Wei Liu et al. "SSD: Single Shot MultiBox Detector". In: To appear. 2016. URL: http://arxiv.org/abs/1512.02325.

[21] Lei Ma et al. "Deep learning in remote sensing applications: A meta-analysis and review". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 152 (2019), pp. 166–177. ISSN: 0924-2716. DOI: https://doi.org/10.1016/j.isprsjprs.2019.04.015. URL: http://www.sciencedirect.com/science/article/pii/S0924271619301108.

[22] Gellert Mattyus, Wenjie Luo, and Raquel Urtasun. "DeepRoadMapper: Extracting Road Topology from Aerial Images". In: *Proceedings of the IEEE International Conference on Computer Vision* 2017-Octob (2017), pp. 3458–3466. ISSN: 15505499. DOI: 10.1109/ICCV.2017.372.

[23] Greg Miller. *The Huge, Unseen Operation Behind the Accuracy of Google Maps*. June 2017. URL: https://www.wired.com/2014/12/google-maps-ground-truth/.

[24] Ishan Misra et al. “Cross-stitch Networks for Multi-task Learning”. In: *CoRR* abs/1604.03539 (2016). arXiv: 1604.03539. URL: http://arxiv.org/abs/1604.03539.

[25] Volodymyr Mnih. “Machine Learning for Aerial Image Labeling”. PhD thesis. University of Toronto, 2013.

[26] Volodymyr Mnih and Geoffrey Hinton. “Learning to Label Aerial Images from Noisy Data”. In: *Proceedings of the 29th International Coference on International Conference on Machine Learning*. ICML’12. Edinburgh, Scotland: Omnipress, 2012, pp. 203–210. ISBN: 978-1-4503-1285-1. URL: http://dl.acm.org/citation.cfm?id=3042573.3042603.

[27] Volodymyr Mnih and Geoffrey E Hinton. “Learning to Detect Roads in High-Resolution Aerial Images”. In: (2009), pp. 1–14.

[28] Agata Mosinska et al. “Beyond the Pixel-Wise Loss for Topology-Aware Delineation”. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 1 (2018), pp. 3136–3145. ISSN: 10636919. DOI: 10.1109/CVPR.2018.00331. arXiv: 1712.02190.

[29] George Papandreou et al. “Weakly- and Semi-Supervised Learning of a DCNN for Semantic Image Segmentation”. In: *CoRR* abs/1502.02734 (2015). arXiv: 1502.02734. URL: http://arxiv.org/abs/1502.02734.

[30] Joseph Redmon et al. “You Only Look Once: Unified, Real-Time Object Detection”. In: *CoRR* abs/1506.02640 (2015). arXiv: 1506.02640. URL: http://arxiv.org/abs/1506.02640.

[31] Shaoqing Ren et al. “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”. In: *CoRR* abs/1506.01497 (2015). arXiv: 1506.01497. URL: http://arxiv.org/abs/1506.01497.

[32] John A. Richards. In: *Remote Sensing Digital Image Analysis*. 2013. ISBN: 978-3-642-30061-5. DOI: 10.1007/978-3-642-30062-2. URL: https://www.springer.com/gp/book/9783642300615#aboutBook.

[33] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *CoRR* abs/1505.04597 (2015). arXiv: 1505.04597. URL: http://arxiv.org/abs/1505.04597.

[34] Sebastian Ruder. “An Overview of Multi-Task Learning in Deep Neural Networks”. In: *CoRR* abs/1706.05098 (2017). arXiv: 1706.05098. URL: http://arxiv.org/abs/1706.05098.

[35] Evan Shelhamer, Jonathon Long, and Trevor Darrell. “Fully Convolutional Networks for Semantic Segmentation”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (May 2016), pp. 1–1. DOI: 10.1109/TPAMI.2016.2572683.

[36] Suriya Singh et al. “Self-Supervised Feature Learning for Semantic Segmentation of Overhead Imagery”. In: *BMVC*. 2018.

[37] Tao Sun et al. “Stacked U-nets with multi-output for road extraction”. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. Vol. 2018-June. 2018. ISBN: 9781538661000. DOI: 10.1109/CVPRW.2018.00033.

[38] Towaki Takikawa et al. “Gated-SCNN: Gated Shape CNNs for Semantic Segmentation”. In: *CoRR* abs/1907.05740 (2019). arXiv: 1907.05740. URL: http://arxiv.org/abs/1907.05740.

[39] Carles Ventura et al. “Iterative Deep Learning for Road Topology Extraction”. In: (2018). arXiv: 1808.09814. URL: http://arxiv.org/abs/1808.09814.

[40] Weixing Wang et al. *A review of road extraction from remote sensing images*. June 2016. DOI: 10.1016/j.jtte.2016.05.005.

[41] Yongyang Xu et al. "Road extraction from high-resolution remote sensing imagery using deep learning". In: *Remote Sensing* 10.9 (Sept. 2018). ISSN: 20724292. DOI: 10.3390/rs10091461.

[42] Yu Zhang and Qiang Yang. "An overview of multi-task learning". In: *National Science Review* 5.1 (Sept. 2017), pp. 30–43. ISSN: 2095-5138. DOI: 10.1093/nsr/nwx105. eprint: http://oup.prod.sis.lan/nsr/article-pdf/5/1/30/24164435/nwx105.pdf. URL: https://doi.org/10.1093/nsr/nwx105.

[43] Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. "Road Extraction by Deep Residual U-Net". In: *CoRR* abs/1711.10684 (2017). arXiv: 1711.10684. URL: http://arxiv.org/abs/1711.10684.

[44] Lichen Zhou, Chuang Zhang, and Ming Wu. "D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction". In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. Vol. 2018-June. 2018, pp. 192–196. ISBN: 9781538661000. DOI: 10.1109/CVPRW.2018.00034.

[45] Yi Zhu et al. "Improving Semantic Segmentation via Video Propagation and Label Relaxation". In: *CoRR* abs/1812.01593 (2018). arXiv: 1812.01593. URL: http://arxiv.org/abs/1812.01593.