

arXiv:1808.09814v1 [cs.CV] 28 Aug 2018

# Iterative Deep Learning for Road Topology Extraction

Carles Ventura<sup>1</sup>

cventuraroy@uoc.edu

Jordi Pont-Tuset<sup>2</sup>

jponttuset@vision.ee.ethz.ch

Sergi Caelles<sup>2</sup>

scaelles@vision.ee.ethz.ch

Kevis-Kokitsi Maninis<sup>2</sup>

kmaninis@vision.ee.ethz.ch

Luc Van Gool<sup>2</sup>

vangool@vision.ee.ethz.ch

<sup>1</sup> Scene Understanding and Artificial

Intelligence Lab

Universitat Oberta de Catalunya

Barcelona, Spain

<sup>2</sup> Computer Vision Laboratory

ETH Zürich

Zürich, Switzerland

## Abstract

This paper tackles the task of estimating the topology of road networks from aerial images. Building on top of a global model that performs a dense semantical classification of the pixels of the image, we design a Convolutional Neural Network (CNN) that predicts the local connectivity among the central pixel of an input patch and its border points. By iterating this local connectivity we sweep the whole image and infer the global topology of the road network, inspired by a human delineating a complex network with the tip of their finger. We perform an extensive and comprehensive qualitative and quantitative evaluation on the road network estimation task, and show that our method also generalizes well when moving to networks of retinal vessels.

## 1 Introduction

Deep learning has gone a long way since its jump to fame in the field of computer vision thanks to the outstanding results in the Imagenet [25] image classification competition back in 2012 [10]. We have witnessed the appearance of deeper [26] and deeper [9] architectures and the generalization to object detection with the well-known trilogy of R-CNNs [9, 10, 23]. Convolutional Neural Networks (CNNs) have played a central role in this development.

A significant step forward was done with the introduction of CNNs for dense prediction, in which the output of the system was not a classification of an image or bounding box into certain categories, but each pixel would receive an output decision. The seminal fully convolutional network [12] was able to perform per-pixel semantic segmentation thanks to an architecture without fully connected layers (i.e. fully convolutional). Many tasks have been tackled from this perspective since then: semantic instance segmentation [11, 12], edge detection [25], medical image segmentation [16], etc.



Figure 1: Patch-based iterative approach for network topology extraction. *Left:* Current state at some point of the iterative approach where the patch-level model for connectivity is applied over the blue square. *Center:* Detections at the local patch for the points at the border (in red) connected to the central point (in blue). *Right:* Final result once the iterative approach ends.

Other tasks, however, have a richer output structure beyond a per-pixel classification, and a higher abstraction of the result is expected. Notable examples that have already been tackled by CNNs are the estimation of the human pose [71], or the room layout [63] from an image. The common denominator of these tasks is that one expects an abstracted model of the result rather than a set of pixel classifications.

This work falls into this category by bringing the power of CNNs to the estimation of the **topology of filamentary networks** such as road networks from aerial images. The structured output is of critical importance and priceless value in these applications: rather than knowing exactly which pixels in a aerial image are road or not, detecting whether two points are connected and how is arguably more informative.

If one thinks how humans would extract the topology of an entangled graph network from an image, it might quickly come to mind the image of them tracing the filaments with the finger and *sweeping* the connected paths continuously. Inspired by this, we propose an iterative deep learning approach that sequentially connects dots within the filaments until it *sweeps* all the visible network. Our approach naturally allows incorporating human corrections: one can simply restart the tracing from the corrected point.

More specifically, we train a CNN on small patches that localizes input and exit points of the filaments within the patch (see image in the middle from Figure 1). By iteratively connecting these dots we obtain the global topology (graph) of the network (see right image from Figure 1). We tackle the extraction of road networks from aerial photos. We experiment on the publicly available Massachusetts Roads dataset to show that our algorithm improves over some strong baselines and provides accurate representations of the network topology. Code is available in <https://github.com/carlesventura/iterative-deep-learning>.

## 2 Related Work

**Curvilinear Structure Segmentation and Tracing:** Tracing of curvilinear structures has been of broad interest in a range of applications, varying from blood vessel segmentation, roadmap segmentation, and reconstruction of human vasculature. Hessian-based methods rely on derivatives, to guide the development of a snake [53], or to detect vessel bound-

aries [1]. Model-based methods rely on strong assumptions about the geometric shapes of the filamentary structures [2, 4] or are based on directional morphological operators [3]. Learning-based methods emerged for the task, using support vector machines on line operators [5], fully-connected CRFs [6], gradient-boosting [7], classification trees [8], or nearest neighbours [9]. Closer to our approach, the most recent methods rely on Fully Convolutional Neural Networks (FCNs), to segment retinal blood vessels [9, 10], or recover vascular boundaries [11]. Different than all the aforementioned method that result in binary structure maps, our method employs deep learning to trace the entire structure of the curvilinear structures, recovering their entire connectivity map. Also related to our method, the authors of [3] trace blood vessels using directed graph theory. To the best of our knowledge, we are the first to apply deep learning for tracing curvilinear structures.

**Road Centerline Detection:** Centerline detection has also followed the trend of curvilinear structure segmentation, with early attempts on gradient-based methods getting outperformed when stronger machine learning techniques emerged [12, 13]. Sironi et al. [14] model the relationships between neighbouring patches to reach the decision for the centerlines. Most recent works employ deep learning techniques, and include results on the Torontocity dataset [15] (which has not been publicly released yet). [16] is the most recent work on extracting the road topology from aerial images, and proposes a post processing algorithm that reasons about missing connections in the extracted road topology from an initial segmentation. In contrast, in our paper we propose an approach that learns the connectivity of the roads at a local scale and is iteratively extended to the entire road network without relying on the results of an initial segmentation.

### 3 Our Approach

This section presents our approach, which combines a global scale for curvilinear structure segmentation and a local scale to estimate its connectivity. The current best approaches for curvilinear structure segmentation applies state-of-the-art deep learning techniques to obtain a segmentation map where each pixel is classified as belonging to the structure (foreground) or not (background). Despite their good performance in segmentation evaluation measures, one of the main drawbacks of these approaches is that they do not take any structure information into account. In particular, these methods are blind to connectivity information among the points that lie in their predicted mask, since all points are assigned only a binary label.

Section 3.1 proposes a method that learns the connectivity of the elements at a local scale. Given a patch of the image centered on a curvilinear structure, the model predicts the locations at the patch border connected with the centered structure. Figure 2-left shows some examples of how we formulate the local connectivity for aerial images: we learn to predict the points on the border of the patch (red dots) that are connected to the center pixel (blue dot). Once the local connectivity model is learned, it is iteratively applied to the image, connecting previous predictions with next ones, and gradually extracting the topology of the graph network, as explained in Section 3.2. We present our evaluation metrics in Section 3.3.

#### 3.1 Patch-level learning for connectivity

As introduced above, the goal is to train a model to estimate the local connectivity in patches. The concept of connectivity is not a property from single points but from pairs of pixels. Cur-



Figure 2: *Left*: Example of training patch for connectivity. The red points represent the locations from the patch border connected with the road indicated by the blue point in the center of the patch. *Right*: Shortest path on semantic segmentation to connect the locations detected at the patch border with the patch center.

rent architectures, however, are designed to estimate per-pixel properties rather than pairwise information. To solve this issue, the local network is designed to estimate which points from the border in a patch are connected to a given input point. Given a patch, therefore, we need to encode the position of the input and output points.

In the context of human pose estimation [21], for instance, points have been encoded as heatmaps with Gaussians centered on them. We follow the same approach and thus the output of our model is a per-pixel probability of being a connected point. Instead of encoding the input point by adding an extra input channel with a heatmap marking its position, we follow a simpler approach. We always place the input point at the center of the patch, thus avoiding the extra input channel and further simplifying the model. We see in the experiments that the model is indeed capable of learning that the central point is the input location we are interested in.

More precisely, we take the architecture of stacked hourglass networks [21] (also used for human pose estimation) to learn the patch-based model for connectivity. This architecture is based on a repeated bottom-up, top-down processing used in conjunction with intermediate supervision. Each bottom-up, top-down processing block is referred to as an hourglass module, which is related to fully convolutional networks that process spatial information at multiple scales but with a more symmetric distribution.

The network is trained using a set of  $k \times k$ -pixel patches from the training set with the pixel at the center of the patch belonging to the foreground, e.g. a pixel annotated as road for road segmentation. The output is a heatmap that predicts the probability of each location being connected to the central point of the patch.

We finally connect the border locations to the center locations by computing the shortest path through the semantic segmentation computed from the global model introduced before, as shown in Figure 2-right. Note that the patch is local enough that a shortest path on the global model is reliable.

## 3.2 Iterative delineation

Once the patch-level model for connectivity has been learned, the model is applied iteratively through the image in order to extract the topology of the network, as a human delineating an image with its fingers not to lose the track. We start from the point with highest foreground probability, given by the global model, as the starting point for the iterative sweeping approach. We then center a patch on this point and find the set of locations at the border of the patch that are connected to the center, with their respective confidence values, using the local patch model.

We discard the locations with a confidence value below a certain threshold and add the remaining ones to a bag of points to be explored  $B_E$ . For each predicted point, we store its location, its confidence value and its precedent predicted point (i.e. the point that was on the center of the patch when the point was predicted). The predicted point  $p$  from  $B_E$  with the highest confidence value is removed from  $B_E$  and inserted to a list of visited points  $B_V$ .

Then,  $p$  is connected to its precedent predicted point using the Dijkstra [1] algorithm over the segmentation probability map over the patch to find the minimum path between them.

We then iterate the process with a patch centered on  $p_c$  and the new predicted points over the confidence threshold are appended to  $B_E$  where they will *compete* against the previous points in  $B_E$  to be the next point to be explored. This process is iteratively applied until  $B_E$  is empty. Note that the list of visited points  $B_V$  is used to discard any point already explored and, therefore, to avoid revisiting the same points over and over again. In a patch centered on  $p_c$ , if a predicted point  $p_p$  belongs to a local neighbourhood of a point  $p_v \in B_V$  and  $p_v$  is the precedent point of  $p_c$ , then the predicted point  $p_p$  is discarded. Otherwise, if  $p_v$  is not the precedent point of  $p_c$  but  $p_p$  belongs to a local neighbourhood of  $p_v$ , then the predicted point  $p_p$  is considered to be connected with  $p_c$ , but  $p_p$  will not be considered for expansion.

The algorithm has been generalized to tackle a problem with unconnected areas, e.g. aerial road images may contain roads that are not connected between them and, therefore, there could be roads not reachable from a single starting point. To prevent that some part of the network topology may have not been extracted, we select a new starting point for a new exploration once the previous  $B_E$  is empty. We impose two constraints on the eligibility for a new starting point: (i) they have to be at a minimum distance of the areas already explored and (ii) their confidence value on the segmentation probability map has to be over a minimum confidence threshold. The iterative approach ends when there are no remaining points eligible for new starting points.

## 3.3 Topology evaluation

The output of our algorithm is a graph defining the topology of the input network, so we need metrics to evaluate their correctness. We propose two different measures for this: a *classical* precision-recall measure that evaluates which locations of the network are detected, and a metric to evaluate connectivity, by quantifying how many pairs of points are correctly or incorrectly connected.

To compute the classical precision-recall curve between two graphs, we build an image with a pixel-wide line sweeping all edges of the given graphs. We then apply the original precision-recall for boundaries [10] on these pair of images. Precision  $P$  refers to the ratio between the number of pixels correctly detected as boundary (true positives) and the number of pixels detected as boundary (true positives + false positives). Recall  $R$  refers to the ratio between the number of pixels correctly detected as boundary (true positives) and the number

of pixels annotated as boundary in the ground truth (true positive + false negative). We take the F measure between  $P$  and  $R$  as a trade-off metric.

The second measure is the connectivity  $C$ , inspired by the definition in [18] as the ratio of segments which were estimated without discontinuities. We define a segment in the graph as the curvilinear structure that connects two consecutive junctions in the ground-truth annotations, as well as connecting an endpoint and its closest connected junction. Two junctions are considered consecutive if there is no other junction within the line that connects them. Given the ground truth path between two consecutive junctions  $p_{gt}$ , the nearest point from the predicted network to each junction is retrieved. Then, the shortest path through the predicted network connecting the retrieved pair of points is computed, which is referred to as  $p_{pred}$ . The ratio between the length of  $p_{gt}$  and the length of  $p_{pred}$  is computed. If the ratio is greater than 0.8 we consider that the ground truth path  $p_{gt}$  has been estimated without discontinuities.

We propose to also have an F measure that combines precision  $P$  with connectivity  $C$ , the reason being that a high connectivity  $C$  value alone does not imply the correctness of the vessels, as it could be a result where all pixels are connected to everything. Adding a competing precision measure prevents this from happening.

For the rest of the paper,  $F^R$  stands for the F measure computed with recall and precision for boundaries values, whereas  $F^C$  stands for the F measure computed between connectivity and precision.

## 4 Experiments

The experiments for road topology extraction on aerial images have been carried out on the publicly available Massachusetts Roads Dataset [20], which includes 1108 images for training, 14 images for validation, and 49 images for testing. Each image is  $1500 \times 1500$  pixels in size, covering an area of 2.25 square kilometers.

**Patch-level evaluation:** To train the patch-level model for connectivity we randomly select 130 patches with size  $64 \times 64$  pixels from each image of the training set, all of them centered on any pixel annotated as road in the ground truth. The number of patches has been selected as a trade-off between having enough training data and diverse training data, i.e. avoiding very redundant patches. The ground-truth locations for the connectivity at the patch level are found by intersecting the skeletonized ground truth mask with a square of side  $s$  pixels (slightly smaller than the patch size) centered on the patch. The ground-truth output heatmap is then generated by adding some Gaussian peaks centered in the found locations. We obtain a precision value of 86.8%, a recall value of 82.2% and F=84.5% in a set of 700 patches randomly cropped from the validation images.

Figure 3 shows some results for the patch-level model for connectivity applied to patches. We can see some examples where the road connections are found despite the shadows of the trees or the similarity of the background with the road. Furthermore, it also learns not to detect roads that are visible on the image but they are not connected to the central road. Figure 4 illustrates other examples where the model fails with false or missing detections, e.g. a visible road not connected with the central road has been wrongly detected.

**Iterative delineation:** Once the patch-level model for connectivity has been trained, it is iteratively applied to extract the topology of the roads from the aerial images. As a baseline



Figure 3: Visual results for the patch-level model for connectivity applied to patches from the Massachusetts Roads dataset. The red crosses represent the detections.



Figure 4: Detection errors for the patch-level model for connectivity applied to patches from the Massachusetts Roads dataset. The red crosses represent the detections.

we compare to extracting the morphological skeleton of detections binarized at different thresholds from the architecture proposed in DRIU [16], a VGG base network on which a set of specialized layers are trained to solve a retinal vessel segmentation task. This architecture has been analogously trained for the road segmentation task presented in this paper. Our proposed iterative approach uses this VGG-based architecture as the global model to select the starting point and to connect the points detected by the patch-level model with the central point of the patch (see Section 3.2).

Table 1 shows the comparison between the global-based baseline for road segmentation based on a VGG architecture and our proposed iterative approach. The connectivity of the skeleton resulting from the VGG-based road segmentation is very low, with a maximum precision-connectivity value  $F^C = 49.3\%$  (a maximum precision-recall value  $F^R = 72.4\%$  is achieved with a different threshold). In contrast, our iterative delineation approach is able to reach a maximum precision-connectivity value  $F^C = 74.4\%$ , while also outperforming the classical precision-recall measure with  $F^R = 81.6\%$ .

Figure 5 illustrates how the road network topology delineation evolves along the iterations of our proposed approach for one of the test images. Figure 6 shows some qualitative results in comparison with the VGG-based road segmentation baseline and the ground-truth

	P	R	C	$F^R$	$F^C$
VGG-150	49.2	<b>94.7</b>	49.4	64.1	49.3
VGG-175	61.5	88.6	30.7	72.0	41.0
VGG-200	75.8	70.7	11.0	72.4	19.2
Iterative (ours)	<b>83.5</b>	80.8	<b>67.1</b>	<b>81.6</b>	<b>74.4</b>

Table 1: Boundary Precision-Recall and Connectivity evaluation in Massachusetts Roads dataset, where xxx in VGG-xxx refers to the threshold used on the output road segmentation from the VGG model before extracting the skeleton.



Figure 5: Evolution of the road network in the iterative delineation. The progress is displayed from left to right and from top to bottom.

annotations. Figure 7 shows some errors in the network topology delineation as some false detections on field (left image) and fluvial (central image) areas or missing detections on high density urban areas (right image).

Finally, some experiments have also been performed on vessel segmentation from retinal images to show that the proposed approach is also valid in other fields. Experiments performed on the DRIVE [30] dataset obtained similar precision-recall values in comparison with DRIU [16], which is the state-of-the-art, but significantly higher connectivity values (see Table 2). Figure 8 shows how the model has been trained and Figure 9 shows the evolution of the iterative approach on a retinal image.

	P	R	C	$F^R$	$F^C$
DRIU [16]	<b>97.3</b>	84.7	67.7	<b>90.4</b>	79.8
Iterative (ours)	86.1	<b>94.1</b>	<b>84.9</b>	89.8	<b>85.5</b>

Table 2: Quantitative results on DRIVE dataset for vessel segmentation.

## 5 Conclusions

In this paper we have presented an approach that iteratively applies a patch-based CNN model for connectivity to extract the topology of filamentary networks. We have demon-

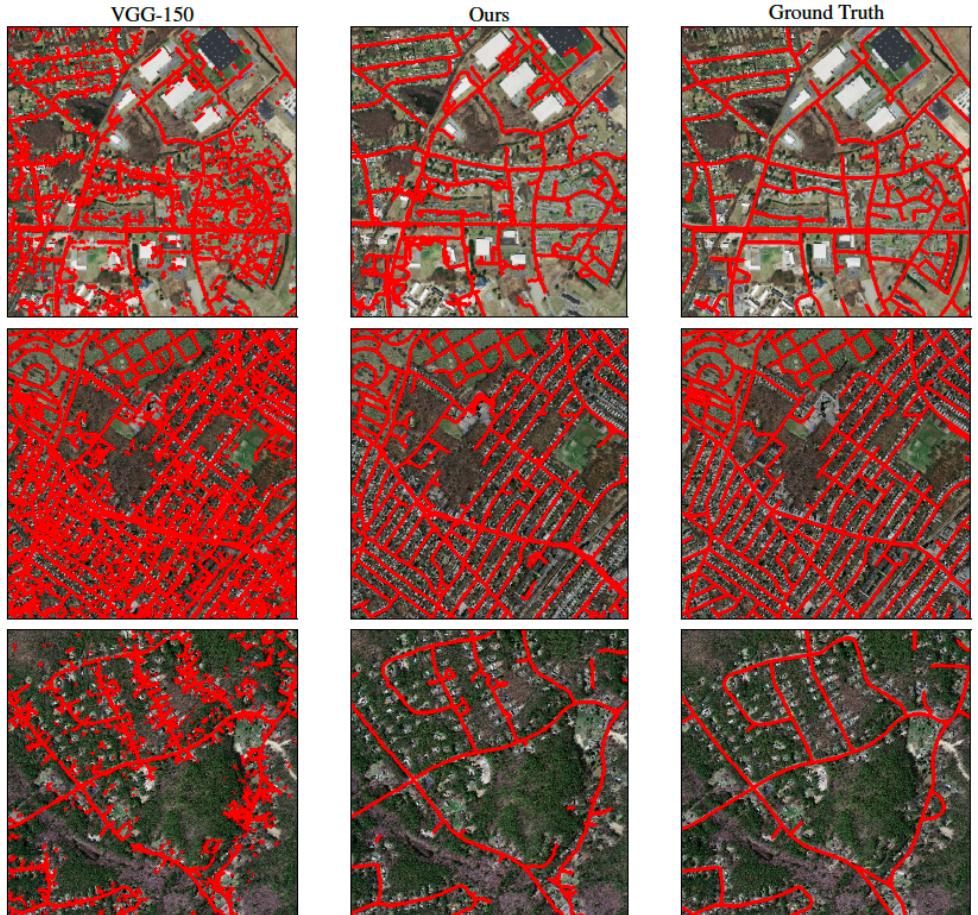


Figure 6: Results for road network topology extraction on the Massachusetts Roads dataset. From left to right: VGG-150, our iterative delineation approach and ground truth.

strated the effectiveness of our technique on road networks from aerial photos. The patch-based model is capable of learning that the central point is the input location and of finding the locations at the patch border connected to the center.

A new F measure ( $F^C$ ) that combines precision and connectivity has been proposed to evaluate the topology results. The experiments carried out on aerial images have obtained the best performance on  $F^C$  compared to strong baselines.

Finally, the proposed iterative patch-based model has also been validated on vessel network extraction from retinal images to show the robustness of the approach in another field.

## Acknowledgements

This research was supported by the Spanish Ministry of Economy and Competitiveness (TIN2015-66951-C2-2-R grant), by Swiss Commission for Technology and Innovation (CTI, Grant No. 19015.1 PFES-ES, NeGeVA) and by the Universitat Oberta de Catalunya.



Figure 7: False and missing detections for road network topology extraction on the Massachusetts Roads dataset.

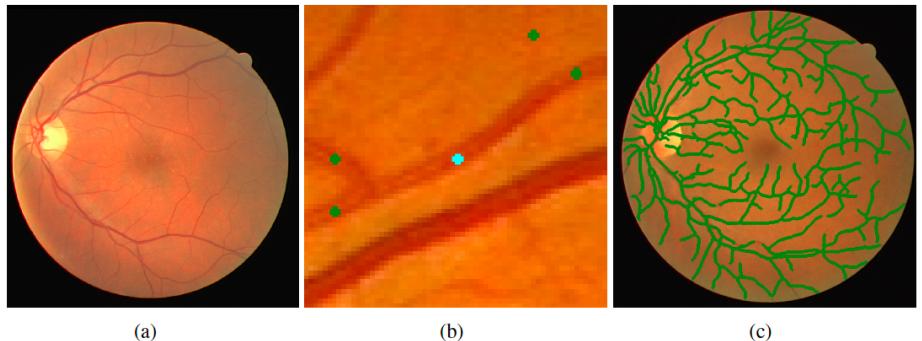


Figure 8: Patch-based iterative approach for vessel network topology extraction. (a): Input retinal image. (b): Detections at the local patch for the points at the border (in green) connected to the central point (in blue). (c): Final result once the iterative approach ends.



Figure 9: Evolution of the vessel network in the iterative delineation.

## References

- [1] Peter Bankhead, C Norman Scholfield, J Graham McGeown, and Tim M Curtis. Fast retinal vessel detection and measurement using wavelets and edge location refinement. *PloS one*, 2012.
- [2] Carlos Becker, Roberto Rigamonti, Vincent Lepetit, and Pascal Fua. Supervised feature learning for curvilinear structure segmentation. In *MICCAI*, 2013.
- [3] Li Cheng, Jaydeep De, Xiaowei Zhang, Feng Lin, and Huiqi Li. Tracing retinal blood vessels by matrix-forest theorem of directed graphs. In *MICCAI*, 2014.
- [4] E.W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1:269–271, 1959. URL <http://eudml.org/doc/131436>.
- [5] Huazhu Fu, Yanwu Xu, Stephen Lin, Damon Wing Kee Wong, and Jiang Liu. Deep-vessel: Retinal vessel segmentation via deep learning and conditional random field. In *MICCAI*, 2016.
- [6] Ross Girshick. Fast R-CNN. In *ICCV*, 2015.
- [7] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *CVPR*, 2014.
- [8] Lin Gu and Li Cheng. Learning to boost filamentary structure segmentation. In *ICCV*, 2015.
- [9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016.
- [10] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask R-CNN. In *ICCV*, 2017.
- [11] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, 2012.
- [12] Max WK Law and Albert CS Chung. Three dimensional curvilinear structure detection using optimally oriented flux. In *ECCV*, pages 368–382. Springer, 2008.
- [13] Chen-Yu Lee, Vijay Badrinarayanan, Tomasz Malisiewicz, and Andrew Rabinovich. Roomnet: End-to-end room layout estimation. In *ICCV*, 2017.
- [14] Yi Li, Haozhi Qi, Jifeng Dai, Xiangyang Ji, and Yichen Wei. Fully convolutional instance-aware semantic segmentation. In *CVPR*, 2017.
- [15] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *CVPR*, 2015.
- [16] Kevis-Kokitsi Maninis, Jordi Pont-Tuset, Pablo Arbeláez, and Luc Van Gool. Deep retinal image understanding. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 140–148. Springer, 2016.

- [17] David R Martin, Charless C Fowlkes, and Jitendra Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE transactions on pattern analysis and machine intelligence*, 26(5):530–549, 2004.
- [18] Gellért Mátyus, Wenjie Luo, and Raquel Urtasun. Deeproadmapper: Extracting road topology from aerial images. In *International Conference on Computer Vision*, 2017.
- [19] Jameson Merkow, Alison Marsden, David Kriegman, and Zhuowen Tu. Dense volume-to-volume vascular boundary detection. In *MICCAI*, 2016.
- [20] Volodymyr Mnih. *Machine Learning for Aerial Image Labeling*. PhD thesis, University of Toronto, 2013.
- [21] Alejandro Newell, Kaiyu Yang, and Jia Deng. Stacked hourglass networks for human pose estimation. In *ECCV*, 2016.
- [22] José Ignacio Orlando and Matthew Blaschko. Learning fully-connected crfs for blood vessel segmentation in retinal images. In *MICCAI*, 2014.
- [23] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *NIPS*, 2015.
- [24] Elisa Ricci and Renzo Perfetti. Retinal blood vessel segmentation using line operators and support vector classification. *IEEE Transactions on Medical Imaging*, 26(10):1357–1365, 2007.
- [25] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *IJCV*, 2015. doi: 10.1007/s11263-015-0816-y.
- [26] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *ICLR*, 2015.
- [27] Amos Sironi, Vincent Lepetit, and Pascal Fua. Multiscale centerline detection by learning a scale-space distance transform. In *CVPR*, 2014.
- [28] Amos Sironi, Vincent Lepetit, and Pascal Fua. Projection onto the manifold of elongated structures for accurate extraction. In *ICCV*, 2015.
- [29] João VB Soares, Jorge JG Leandro, Roberto M Cesar Jr, Herbert F Jelinek, and Michael J Cree. Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification. *IEEE Transactions on Medical Imaging*, 25(9):1214–1222, 2006.
- [30] J.J. Staal, M.D. Abramoff, M. Niemeijer, M.A. Viergever, and B. van Ginneken. Ridge based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging*, 23(4):501–509, 2004.
- [31] Silvia Valero, Jocelyn Chanussot, Jon Atli Benediktsson, Hugues Talbot, and Björn Waske. Advanced directional mathematical morphology for the detection of the road network in very high resolution remote sensing images. *Pattern Recognition Letters*, 31(10):1120–1127, 2010.

- [32] 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. In *International Conference on Computer Vision*, 2017.
- [33] Yu Wang, Arunachalam Narayanaswamy, Chia-Ling Tsai, and Badrinath Roysam. A broadly applicable 3-d neuron tracing method based on open-curve snake. *Neuroinformatics*, 9(2-3):193–217, 2011.
- [34] Jan D Wegner, Javier A Montoya-Zegarra, and Konrad Schindler. A higher-order crf model for road network extraction. In *CVPR*, 2013.
- [35] Saining Xie and Zhuowen Tu. Holistically-nested edge detection. *IJCV*, pages 1–16, 2017.