Iterative Deep Retinal Topology Extraction

Carles Ventura[1(✉)], Jordi Pont-Tuset[2], Sergi Caelles[2], Kevis-Kokitsi Maninis[2], and Luc Van Gool[2]

[1] Scene Understading and Artificial Intelligence Lab, Universitat Oberta de Catalunya, Barcelona, Spain
cventuraroy@uoc.edu

[2] Computer Vision Laboratory ETH Zürich, Zürich, Switzerland
{jponttuset,scaelles,kmaninis,vangool}@vision.ee.ethz.ch

Abstract. This paper tackles the task of estimating the topology of filamentary networks such as retinal vessels. 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 between 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 filamentary network, inspired by a human delineating a complex network with the tip of their finger. We perform a qualitative and quantitative evaluation on retinal veins and arteries topology extraction on DRIVE dataset, where we show superior performance to very strong baselines.

1 Introduction

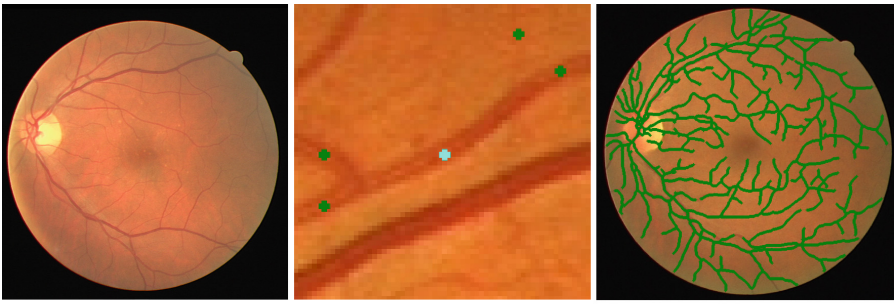
Deep learning has gone a long way since its jump to fame in the field of computer vision thanks to the outstanding results in the Imagenet [23] image classification competition back in 2012 [11]. We have witnessed the appearance of deeper [24] and deeper [10] architectures and the generalization to object detection [6,7,21]. 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. Many tasks have been tackled from this perspective since then: semantic instance segmentation [9,14], edge detection [29], medical image segmentation [15], etc.

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 [19], or the room layout [13] 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.

W. Bai et al. (Eds.): Patch-MI 2018, LNCS 11075, pp. 133–143, 2018.
https://doi.org/10.1007/978-3-030-00500-9_15

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



(a) (b) (c)

Fig. 1. Patch-based iterative approach for 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. (Color figure online)

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.

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 [28], or to detect vessel boundaries [1]. Model-based methods rely on strong assumptions about the geometric shapes of the filamentary structures [12,26]. Learning-based methods emerged for the task, using SVMs on line operators [22], fully-connected CRFs [20], gradient-boosting [2], classification trees [8], or nearest neighbours [25]. Closer to our approach, the most recent methods rely on Fully Convolutional Neural Networks (FCNs), to segment retinal blood vessels [5,15], or recover vascular boundaries [18]. 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.

More specifically, we train a CNN on small patches that localizes input and output points of the filaments within the patch (Fig. 1(b)). By iteratively

connecting these dots we obtain the global topology (graph) of the network (Fig. 1(c)). We tackle the extraction of the topology of retinal vessels (veins and arteries) from fundus images. We experiment on DRIVE dataset to show that our algorithm improves over some very strong baselines and provides accurate representations of the topology of vessels. To the best of our knowledge, we are the first to apply deep learning for tracing curvilinear structures. Code is available in https://github.com/carlesventura/iterative-deep-retinal.

2 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). The most relevant example of such approach is the VGG-based architecture used in DRIU [15] for vessel segmentation. 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, this method is blind to connectivity information among the points that lie in their predicted mask, since all points are assigned only a binary label.

Section 2.1 proposes a method that learns the connectivity of the elements at a local scale. Once the local model is learned, it is iteratively applied to the image, connecting previous predictions with next ones, and gradually extracting the topology of the network, as explained in Sect. 2.2. The evaluation metrics are presented in Sect. 2.3.

2.1 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. Current 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 in a patch are connected to a given input point.

More precisely, we take the architecture of stacked hourglass networks [19] 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 vessel). The output is a heatmap that predicts the probability of each location being connected to the central point of the patch.

Furthermore, the model is also trained to differentiate between the two types of vessel (arteries or veins), so the model is forced to learn not only the connectivity but also an artery-vein classification problem. To illustrate this idea, Fig. 2 shows some examples of connectivity for retinal images where we differentiate three types of models. Figures 2(a) and (b) compare two patches where all vessels that intersect the border patch have been marked (Fig. 2(a)), versus the ones that are connected to the vessel at the center of the patch (Fig. 2(b)). Figures 2(c) and (d) illustrate the difference between detecting the connectivity over any type of vessel (Fig. 2(c)), or forcing the connectivity to be over the same type of vessel (Fig. 2(d)).

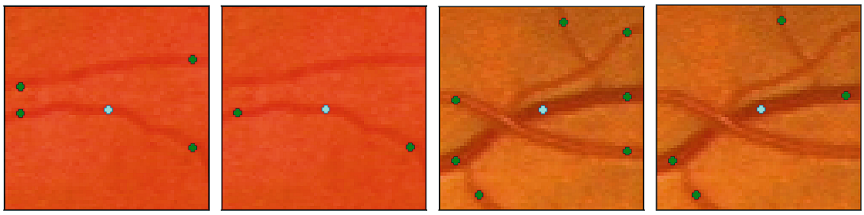


Fig. 2. Examples of training patches for connectivity. The green points represent the locations from the patch border connected with the vessel indicated by the blue point in the center. (Color figure online)

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. Note that the patch is local enough that a shortest path on the global model is reliable.

2.2 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. We start from the point with highest foreground probability, given by the global model. 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 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 [3] 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. 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.

Since in retinal images all vessels are connected through the optical disk, any vessel point from the image is reachable from any starting point used in the iterative approach. However, the algorithm has been generalized to tackle a problem with unconnected areas, e.g. a cropped retinal image where the entire retina is not visible and, therefore, there could be vessels 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.

2.3 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 [16] 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 [17] 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 (junctions refer to both crossovers and bifurcations). 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}. If the ratio between the length of p_{gt} and the length of p_{pred} is greater than 0.8 we consider that the ground truth path p_{gt} has been estimated without discontinuities. In Fig. 3, the two images on the left show examples where the ground truth segment have been estimated without discontinuities, whereas the two examples on the right are considered as not connected segments on the connectivity measure.

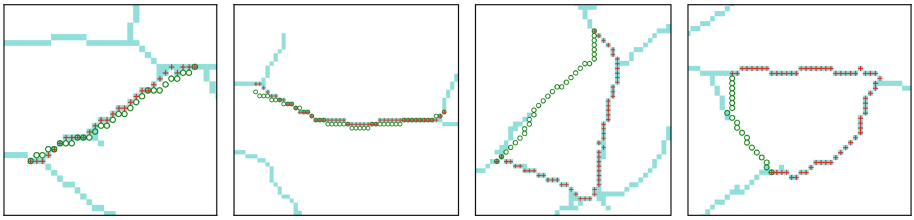


Fig. 3. Examples of good (on the left) and bad (on the right) connectivity. Green pixels represent ground truth connections, blue pixels represent predicted vessels with our iterative approach and red pixels represent the path found through predicted network. (Color figure online)

We propose to also have an F measure that combines precision P with connectivity C. The reason is that a high connectivity C value does not implies a good graph that defines the topology of the network. Whereas the connectivity measures the ratio of estimated segments without discontinuities, the precision measures how good the predicted locations along the segments are. 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.

3 Experiments

The experiments have been carried out on the DRIVE [27] dataset, which includes 40 eye fundus images and contains manual segmentation of the blood vessels by expert annotators. As a global model for segmentation, we use DRIU [15], which is the state of the art for retinal vessel segmentation.

Patch-level evaluation: To train the patch-level model for connectivity we randomly select 50 patches with size 64×64 pixels from each image of the training set, all of them centered on one of vertices of the graph annotations provided

by [4], which includes arteries and veins annotations. The ground-truth locations for the connectivity at the patch level are found by intersecting the vessels 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 a subset of the found locations, depending on the configuration:

- For the *non-connectivity* model, all the locations are considered (Fig. 2(a)).
- For the *connectivity* model, only the intersection points connected to the center along a path completely included in the patch are considered (Figs. 2(b) and (c)).
- For the *connectivity-av* model, only the intersection points connected to the center and belonging to the same type of vessel (artery or vein) as the vessel centered on the patch are considered (Fig. 2(d)).

The non-connectivity patch-level model reaches the best result (F = 82.1, P = 85.3, R = 79.1). The connectivity model, which has to tell apart those points connected with the patch center, achieves an only slightly worse performance (F = 80.4, P = 82.5, R = 78.4), despite the task being more complicated. The model that has also to distinguish between arteries and veins results in a more significant drop in the performance (F = 74.8, P = 75.9, R = 73.7), but it still keeps a very good result.

Figure 4 shows some visual results for the three type of configurations considered. In the first row, the model is able to differentiate the vessels connected to the patch center from those ones not connected (3rd and 4th column). In the second row, the model differentiates the vessels from the same type as the centered vessel (an artery) from those of different vessel type (see 4th and 5th column). The last row shows a failure case where the model correctly predicts the connectivity but it is not able to differentiate the arteries from the veins.

Iterative delineation: Once the patch-level model for connectivity has been trained, it is iteratively applied to extract the topology of the blood vessels networks from the eye fundus images. As a strong baseline we compare to extracting the morphological skeleton of detections binarized at different thresholds from the architecture proposed in DRIU [15], a VGG base network on which a set of specialized layers are trained to solve the retinal vessel segmentation task. 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 Sect. 2.2). Table 1 compares to DRIU for different thresholds: 224 (the optimal for vessel segmentation obtained in [15]), 200 (the optimal value for precision-recall boundary evaluation F^R) and 170 (the optimal value for precision-connectivity evaluation F^C). Our proposed iterative approach outperforms DRIU for connectivity in 6.6 points, which results on a improvement of 1.8 in the precision-connectivity evaluation measure F^C. Furthermore, both techniques are also compared with an upper bound and a lower bound: the former is the skeleton extracted from the ground truth vessel segmentation, and the latter results from evaluating the

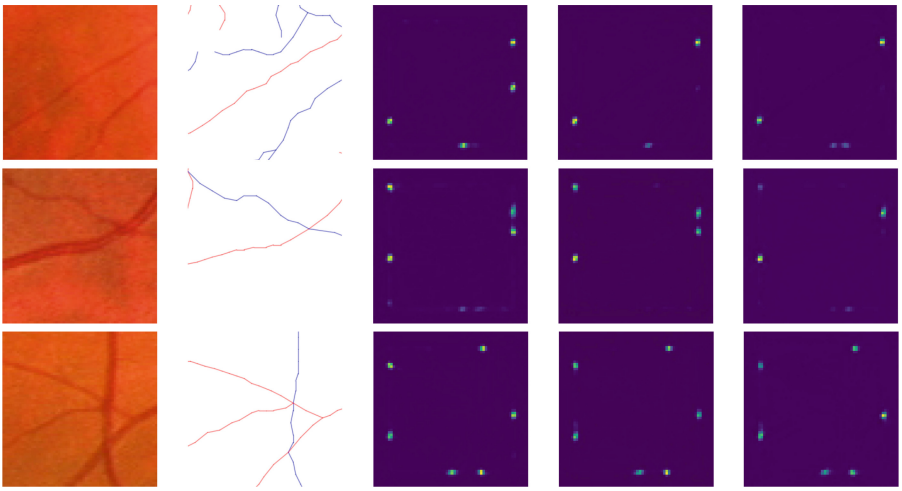


Fig. 4. Visual results of the patch-level models for eye fundus images. From left to right: eye fundus image, artery-vein annotation, output confidence for non-connectivity model, output confidence for connectivity model and output confidence for connectivity-av (artery-vein) model.

ground truth skeleton obtained from a different image. Our results are only 7.7 points below the upper bound in connectivity. The experiments have also been performed with other patch size values ($k = 32$ and $k = 128$) and the results do not change significantly, which shows the robustness of the patch size to the scale of the image. PSPNet [30], which is the state-of-the-art semantic segmentation method to date, has also been considered as a baseline. However, the results obtained by PSPNet in the DRIVE dataset are significanlty lower compared to DRIU [15] (see Table 1).

Figure 5 illustrates how the vessel network topology extraction evolves along the iterations of our proposed approach for one of the test images.

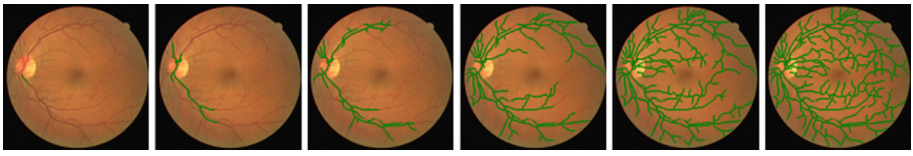


Fig. 5. Evolution of the vessel network in the iterative delineation.

Arteries and veins separation: For eye fundus images, we also pursue the objective of differentiating arteries and veins. The approach is similar to the iterative delineation proposed before, but now using the patch-level model for connectivity that also takes into account that the vessels connected have to be of the same type. We have referred before to this model as the *connectivity-av* model.

Table 1. Boundary Precision-Recall and Connectivity evaluation for vessels (left), arteries (top-right) and veins (bottom-right) in the DRIVE dataset

	P	R	C	F^R	F^C
DRIU-224 [15]	**97.3**	84.7	67.7	90.4	79.8
DRIU-200 [15]	93.8	90.6	74.0	**92.0**	82.7
DRIU-170 [15]	89.9	93.1	78.3	91.3	83.7
PSPNet [30]	92.8	69.9	49.7	79.5	64.7
Iterative (ours)	86.1	**94.1**	**84.9**	89.8	**85.5**
GT skel (upperbound)	95.6	99.3	92.6	97.4	94.1
Random (lowerbound)	44.2	45.9	21.8	44.9	29.2

	P	R	C	F^R	F^C
VGG-220	72.9	80.7	52.4	76.1	61.0
VGG-190	64.5	**88.2**	**65.4**	74.1	64.9
Iterative (ours)	**81.4**	75.3	63.0	**78.0**	**71.0**
VGG-230	70.8	79.1	42.2	74.2	52.9
VGG-180	57.4	**91.3**	**66.1**	70.2	61.5
Iterative (ours)	**72.0**	79.6	61.2	**75.4**	**66.2**

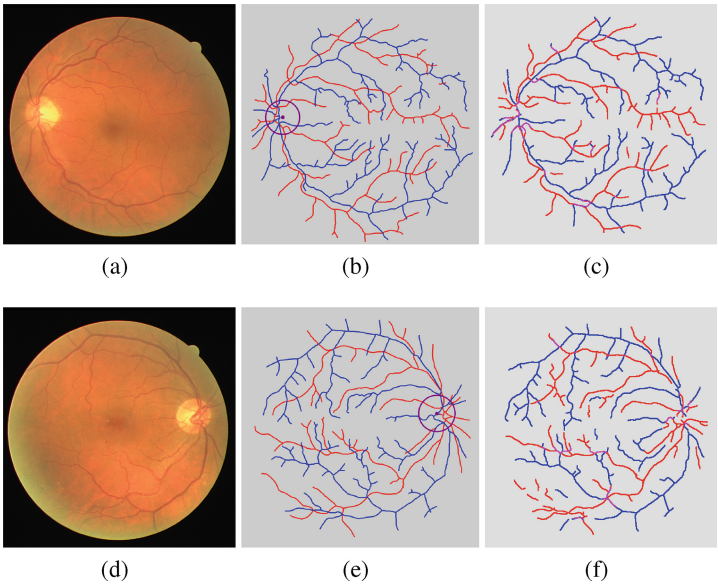


Fig. 6. Qualitative results on arteries and veins separation for two test images ((a) and (d)) comparing ground truth ((b) and (e)) with our method ((c) and (f)): veins in blue, arteries in red. (Color figure online)

As baseline, we have considered the same CNN architecture as in DRIU, i.e. a VGG-based architecture, but using the annotations for arteries and veins given by [4]. These annotations are only given at the graph level, so we build the ground-truth image by drawing one-pixel wide lines delineating the arteries and veins networks; which is different from the vessel segmentation pixel-accurate masks from DRIVE on which DRIU is usually trained. We train one global model for arteries and one for veins, and then we apply the delineation algorithm using the connectivity-av patch-level model. Table 1 shows the results obtained for arteries (top-right) and veins (bottom-right). In both cases, our iterative

approach reaches the best trade off between F^R and F^C. Figure 6 shows some qualitative results comparing the ground truth annotations with our method.

4 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 demonstrated the effectiveness of our technique on retinal vessels from fundus images. 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. Furthermore, we can also differentiate arteries and veins and extract their respective networks. A new F measure (F^C) that combines precision and connectivity has been proposed to evaluate the topology results. The experiments carried out on retinal images have obtained the best performance on F^C compared to strong baselines.

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.

References

1. Bankhead, P., Scholfield, C.N., McGeown, J.G., Curtis, T.M.: Fast retinal vessel detection and measurement using wavelets and edge location refinement. PloS one **7**, e32435 (2012)
2. Becker, C., Rigamonti, R., Lepetit, V., Fua, P.: Supervised feature learning for curvilinear structure segmentation. In: MICCAI (2013)
3. Dijkstra, E.: A note on two problems in connexion with graphs. Numerische Mathematik **1**, (1959)
4. Estrada, R., Allingham, M.J., Mettu, P.S., Cousins, S.W., Tomasi, C., Farsiu, S.: Retinal artery-vein classification via topology estimation. T-MI **34**, (2015)
5. Fu, H., Xu, Y., Lin, S., Kee Wong, D.W., Liu, J.: DeepVessel: retinal vessel segmentation via deep learning and conditional random field. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 132–139. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_16
6. Girshick, R.: Fast R-CNN. In: ICCV (2015)
7. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: CVPR (2014)
8. Gu, L., Cheng, L.: Learning to boost filamentary structure segmentation. In: ICCV (2015)
9. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. In: ICCV (2017)
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR (2016)
11. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: NIPS (2012)

12. Law, M.W.K., Chung, A.C.S.: Three dimensional curvilinear structure detection using optimally oriented flux. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008. LNCS, vol. 5305, pp. 368–382. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-88693-8_27
13. Lee, C.Y., Badrinarayanan, V., Malisiewicz, T., Rabinovich, A.: Roomnet: End-to-end room layout estimation. In: ICCV (2017)
14. Li, Y., Qi, H., Dai, J., Ji, X., Wei, Y.: Fully convolutional instance-aware semantic segmentation. In: CVPR (2017)
15. Maninis, K.-K., Pont-Tuset, J., Arbeláez, P., Van Gool, L.: Deep retinal image understanding. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 140–148. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_17
16. Martin, D.R., Fowlkes, C.C., Malik, J.: Learning to detect natural image boundaries using local brightness, color, and texture cues. In: TPAMI (2004)
17. Máttyus, G., Luo, W., Urtasun, R.: Deeproadmapper: Extracting road topology from aerial images. In: International Conference on Computer Vision (2017)
18. Merkow, J., Marsden, A., Kriegman, D., Tu, Z.: Dense volume-to-volume vascular boundary detection. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9902, pp. 371–379. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46726-9_43
19. Newell, A., Yang, K., Deng, J.: Stacked hourglass networks for human pose estimation. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9912, pp. 483–499. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46484-8_29
20. Orlando, J.I., Blaschko, M.: Learning fully-connected CRFs for blood vessel segmentation in retinal images. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014. LNCS, vol. 8673, pp. 634–641. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10404-1_79
21. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks. In: NIPS (2015)
22. Ricci, E., Perfetti, R.: Retinal blood vessel segmentation using line operators and support vector classification. T-MI **26**, (2007)
23. Russakovsky, O.: ImageNet large scale visual recognition challenge. IJCV **115**, 211–252 (2015)
24. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: ICLR (2015)
25. Sironi, A., Lepetit, V., Fua, P.: Projection onto the manifold of elongated structures for accurate extraction. In: ICCV (2015)
26. Soares, J.V., Leandro, J.J., Cesar Jr, R.M., Jelinek, H.F., Cree, M.J.: Retinal vessel segmentation using the 2-D gabor wavelet and supervised classification. T-MI **25**, (2006)
27. Staal, J., Abramoff, M., Niemeijer, M., Viergever, M., van Ginneken, B.: Ridge based vessel segmentation in color images of the retina. T-MI **23**, (2004)
28. Wang, Y., Narayanaswamy, A., Tsai, C.L., Roysam, B.: A broadly applicable 3-D neuron tracing method based on open-curve snake. Neuroinformatics **9**, 193–217 (2011)
29. Xie, S., Tu, Z.: Holistically-nested edge detection. IJCV **125**, (2017)
30. Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network. In: CVPR (2017)