Retinal Blood Vessel Segmentation Using a Fully Convolutional Network – Transfer Learning from Patch- to Image-Level

Taibou Birgui-Sekou, Moncef Hidane, Julien Olivier, Hubert Cardot

▶ **To cite this version:**

Taibou Birgui-Sekou, Moncef Hidane, Julien Olivier, Hubert Cardot. Retinal Blood Vessel Segmentation Using a Fully Convolutional Network – Transfer Learning from Patch- to Image-Level. International Workshop on Machine Learning in Medical Imaging, Sep 2018, Granada, Spain. pp.170-178, 10.1007/978-3-030-00919-9_20 . hal-03405807

HAL Id: hal-03405807
https://hal.science/hal-03405807

Submitted on 4 Mar 2022

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

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

Retinal Blood Vessel Segmentation using a Fully Convolutional Network – Transfer learning from patch- to image-level

Taibou Birgui Sekou[1,3], Moncef Hidane[1,3], Julien Olivier[1,3], and Hubert Cardot[2,3]

[1] Institut National des Sciences Appliquées Centre Val de Loire, Blois France,
[2] Université de Tours, Tours, France
[3] LIFAT EA 6300, Tours, France

Abstract. Fully convolutional networks (FCNs) are well known to provide state-of-the-art results in various medical image segmentation tasks. However, these models usually need a tremendous number of training samples to achieve good performances. Unfortunately, this requirement is often difficult to satisfy in the medical imaging field, due to the scarcity of labeled images. As a consequence, the common tricks for FCNs' training go from data augmentation and transfer learning to patch-based segmentation. In the latter, the segmentation of an image involves patch extraction, patch segmentation, then patch aggregation.
This paper presents a framework that takes advantage of all these tricks by starting with a patch-level segmentation which is then extended to the image level by transfer learning. The proposed framework follows two main steps. Given a image database $\mathcal{D}$, a first network $\mathcal{N}_P$ is designed and trained using patches extracted from $\mathcal{D}$. Then, $\mathcal{N}_P$ is used to pre-train a FCN $\mathcal{N}_I$ to be trained on the full sized images of $\mathcal{D}$. Experimental results are presented on the task of retinal blood vessel segmentation using the well known publicly available DRIVE database.

Keywords: Retinal blood vessel segmentation · fully convolutional neural networks · transfer learning

1 INTRODUCTION

The human vascular system is an important risk biomarker in a large number of diseases. In particular, retinal blood vessels serve as a cue to diagnose diabetic retinopathy, age-related macular degeneration and glaucoma. As the eye shares neural and vascular similarities with the brain, its vascularization also offers a direct window to cerebral pathology.

Manual delineation of blood vessels from images by ophthalmologists is a tedious task. It is also subject to inter- and intra-operator variability. To alleviate this difficulty, an intensive body of work has concentrated on developing automatic retinal blood vessel segmentation (RBVS) techniques. Fraz *et al.* [1]

presented a review and a taxonomy of the proposed methods in the field, up to 2012. A recent reviews is presented in [2].

State-of-the-art methods for RBVS are supervised and mainly based on deep learning. They associate a label to each pixel in the image, indicating whether it belongs to a vessel or to the background.

In [3], each pixel is labeled using a (preprocessed) surrounding $m \times m$ patch (a small neighborhood centered around it). The classification is performed using a freely designed deep convolutional neural network. The authors also presented an interesting variation of their method where they framed the segmentation task as a structured inference problem. This leads to a deep neural network that predicts the class assignments for all pixels in a small window, of size $s \times s$ (with $s < m$), located inside the input patch. The same idea of structured prediction is employed in [4] to train a FCN using patches extracted from the training images. In [6, 7], patch-based methods are proposed using discriminative dictionary learning techniques for RBVS.

In general, working at the patch level eases the possibility of learning arbitrarily designed deep networks since from few images one can extract millions of (overlapping) patches. However, a patch aggregation step, which may be time consuming, is needed to obtain the segmentation of an entire image. Moreover, a patch based segmentation labels a pixel using a restrained view and does not take advantage of the extra information located in other parts of the image. Instead, it is also possible to work at the image level, that is learning on complete images and segmenting each test image in one forward pass across the network.

On the task of RBVS, Mo et al. [8] proposed a VGG-like[12] network and included an additional contextual information in the network by aggregating the segmentation of different layers. The proposed model working at the image level, they performed a data augmentation technique to boost the database size. Still, the size of their training set was very small compared to the number of parameters in their network. Thus, the authors initialize their network using pre-trained weights from ImageNet.

Transfer learning from VGG, or other architectures, might lead to good performance but diminishes the degree for freedom when designing new networks since the transfered parts of the network must remain unchanged. This reduces the field of exploration and research.

This paper introduces a transfer learning based framework to train arbitrarily designed FCNs even on relatively small sized databases. The framework is tested on the task of RBVS using an example of freely designed FCN. The proposed framework consists of two steps. Given a database $\mathcal{D}$. In the first step, a fully convolutional network $\mathcal{N}_P$ is designed and trained using patches extracted from the training images of $\mathcal{D}$. The second step consists in re-using the weights of $\mathcal{N}_P$ to pre-train a FCN $\mathcal{N}_I$ which takes as inputs the full sized images of $\mathcal{D}$. Full size image segmentation means that an image is segmented in one forward pass, which is more practical in the medical field than aggregating extracted patches. In this manner, one can train various network architectures first at the patch level then transfer the weights to segment images in one forward pass. We

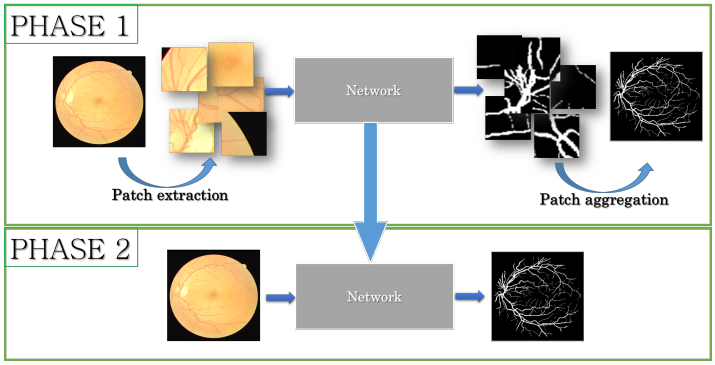

Fig. 1: The proposed framework. Phase 1: The network is fed with patches extracted from the original images. Phase 2: The network is fine-tuned with the full sized images.

experimentally show that our network outputs outstanding results on the well known publicly available DRIVE dataset.

The overview of the paper is as follows. Section 2 presents our framework and the proposed network in terms of its architecture. In Section 3, we present all the experimental aspects of the work including the dataset, the training set generation and the results along with a discussion. Section 4 sums up the paper and introduces possible future work.

2 PROPOSED MODEL

This section presents the core components of our proposition: the different steps of the framework and the network.

2.1 Fully convolutional networks and transfer learning

Deep neural networks generally consist of multiple layers of different type and purpose. The convolution or the fully connected layers aim at learning efficient patterns in the training set. The patterns are either discriminative (e.g. classification tasks) or generative (e.g. generative models). Layers that perform pooling, non-linearity or batch- normalization inject more robustness in the model and provide a way for the network i) to be more invariant to some change in the input, and ii) to have better generalization power. A detailed presentation of common layers of a deep network is presented in [11].

The proposed framework is mainly based on fully convolutional networks [13] which are particular deep neural networks that do not include fully connected layers. As a consequence, the spatial and structural information of the input can be preserved throughout the network. A FCN, by construction, only imposes its inputs to share the same number of dimensions and channels. For example, a FCN trained on an input of shape $10 \times 10 \times 3$ to output a shape 5×5 can,

theoretically, be applied on an input of shape $20 \times 20 \times 3$ to output 10×10. On the task of medical image segmentation, to train FCNs, one usually makes use of a transfer learning technique. Transfer learning consists in reusing some weights of a network $\mathcal{N}_s$ trained on a source database $\mathcal{D}_s$ as a starting point for a network $\mathcal{N}_t$ to be trained on a target database $\mathcal{D}_t$ [14].

In the following, the proposed framework is introduced. It makes use of *i*) the FCNs shape preservation between the input and output, and *ii*) the transfer learning from patch- to image-level.

2.2 The proposed framework

As aforementioned, RBVS is a major phase in the diagnostic system. It needs to be as fast and accurate as possible. And, this applies to all medical image segmentation tasks. To guarantee a fast segmentation time, we need to segment each image at once instead of working at the patch level.

On the other hand, the number of images makes the training phase highly challenging, thus the need for a good starting point. On RBVS tasks, Mo et al. [8] initialized their network with pre-trained weights learned on a large-scale natural image dataset (ImageNet).

In this paper, we introduce another pre-training possibility. We propose to initialize the network with weights learned from the patches of the same dataset. The proposed framework consists of two phases as depicted on Fig. 1.

Phase 1 (patch level) In this phase, we first create a patch training set $\mathcal{D}_P$ by extracting a large number of patches from the training images. Then, a fully convolutional network $\mathcal{N}_P$ is freely designed and fed with samples of $\mathcal{D}_P$. Given a patch of shape $m \times n$ the network $\mathcal{N}_P$ will output its segmentation map of the same shape. Hence, at this level, one need to aggregate patches of different location to segment an entire image (*see* First row of Fig. 1).

Phase 2 (image level) At this stage, we already have a network that outputs good results at the patch level. The goal is to have similar performance but at the image level. To do so, we first design a network $\mathcal{N}_I$ that takes as an input an entire image and produces its overall segmentation mask. Suppose, the images are of shape $h \times w$, then the only difference between a patch and an image is the number of pixel, their number of dimension and channel being the same. As aforementioned, FCNs can process inputs of similar shapes regardless of their number of pixels. Our transfer learning consists in reusing all the weights of $\mathcal{N}_P$ in $\mathcal{N}_I$. In other words, the only difference between $\mathcal{N}_P$ and $\mathcal{N}_I$ is the size of their inputs. The network $\mathcal{N}_I$ can the be trained using the full sized images of $\mathcal{D}$. This training process can be seen as a fine-tuning of $\mathcal{N}_P$ at the image level.

In the following, the patch level network refers to $\mathcal{N}_P$ while the one fine-tuned at the image level refers to $\mathcal{N}_I$.

2.3 Neural Network Architecture

In this work, we present a fully convolutional auto-encoder-like network depicted on Fig. 2. The encoder phase (first eleven layers) extracts high-level abstract

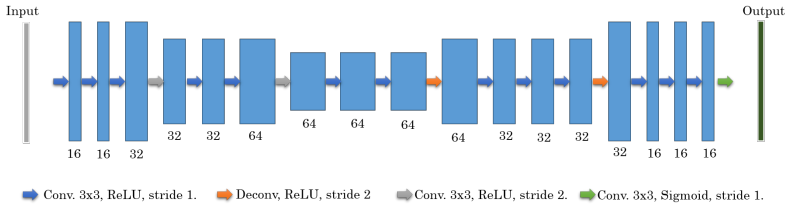

Fig. 2: Network architecture.

patterns. From these underlying structures, the decoder phase tries to recover the segmentation of an input.

The network consists of convolution and deconvolution layers. The down-samplings are performed with convolution of stride 2. And, to recover the initial input size, the deconvolution layers which apply transposed convolutions with a certain stride, are used. All convolutions are followed by a ReLU activation except the last one which uses a Sigmoid to output a class probability for each pixel of the input. The network may have the same look as other well know architectures but has less weights and less down-sampling levels. This architecture has been experimentally selected and performs better than very deep and complex propositions of the literature on the task of retinal blood vessel segmentation.

A classical cross-entropy, which is based on the distance between probability distribution, is used as loss function.

3 EXPERIMENTS

We applied the proposed method on the task of RBVS. The evaluation is performed on the publicly available DRIVE dataset. This section presents practical details about the training including the normalizations and the patches extraction procedure. Afterwards, our numerical and some qualitative results are exposed and discussed.

The DRIVE[4] (Digital Retinal Images for Vessel Extraction) [15] contains RGB fundus image of size 585×564×3. The mask image delineating the field of view (FoV) of each image is also provided. The DRIVE dataset is divided into two sets of 20 images: the training and testing sets.

3.1 Data Preparation and Network Training

We applied three operations on each image before any procedure: 1) gray-scale conversion, 2) gamma correction (with gamma set to 1.7), and 3) Contrast Limited Adaptive Histogram Equalization (CLAHE). The database is boosted by adding for each image its vertically and horizontally flipped version.

[4] http://www.isi.uu.nl/Research/Databases/DRIVE/

Table 1: Results on the DRIVE database.

Methods	AUC	Spec	Sens	Acc
Proposed-patches	**98.01**	**98.37**	76.65	**95.58**
Proposed-images	97.87	97.82	79.90	95.52
Images-nopretrain	97.01	98.27	72.91	95.02
Birgui S. et al. [7] (JDCL)**	-	96.32	**80.60**	94.93
Dasgupta et al. [4]*	97.44	98.01	76.91	95.33
Javidi et al. [6]**	-	97.02	72.01	94.50
Liskowski et al. [3]*	97.90	98.07	78.11	95.35
Mo et al. [8]*	97.82	97.80	77.79	95.21
Vega et al.[5]*	-	96.00	74.44	94.12
Zhang et al. [9]	96.36	97.25	77.43	94.76

*deep learning — **dictionary learning

At the patch level, to avoid storing the entire patch dataset, their are extracted on the fly. That is, at each epoch and for each image, we extract randomly 3200 patches of shape 32×32, where half of them are centered on a vessel pixel and the other half are centered on a background pixel.

At the image level, if needed, we concatenate a zero matrix to an image to ensure an output with the same size. In other words, if the input image is of size (584×565) we zero-pad the second dimension to obtain (584×568) so that the down-sampling by 4 in the network will be straightforward.

The network is implemented using the Keras library. The training is carried out on a GPU Nvidia GeForce GTX 1080 Ti, with 64 batch size when using patches and 1 at the image level. The *adadelta* [16] learning algorithm is adopted. We performed 15 epochs at the patch level and 300 at the image level.

3.2 Results

The proposed model is compared to the most recent and state-of-the-art methods using the Area Under the ROC Curve (**AUC**) metric. The latter is a commonly used metric for RBVS. The AUC score is an important metric in the sense that it aggregates metrics of various threshold. Let TP, TN, FP, and FN respectively denote the number of true positive, true negative, false positive, and false negative. We computed with a 0.5 threshold the sensitivity $\mathbf{Sens} = \frac{TP}{TP+FN}$, the specificity $\mathbf{Spec} = \frac{TN}{TN+FP}$ and the accuracy $\mathbf{Acc} = \frac{TP+TN}{TP+TN+FP+FN}$.

Our numerical results are presented on Table 1 and are discussed in the next section. On Table 1, *Proposed-patches* are the results obtained with the network trained only on patches and *Proposed-images* are the results achieved when the network is fine-tuned at the image level. The line *Images-nopretrain* is added to present the results at the image level without pretraining from the patches, obtained after convergence on a validation set (300 epochs).

Figure 3 illustrates some qualitative results of the proposed method.

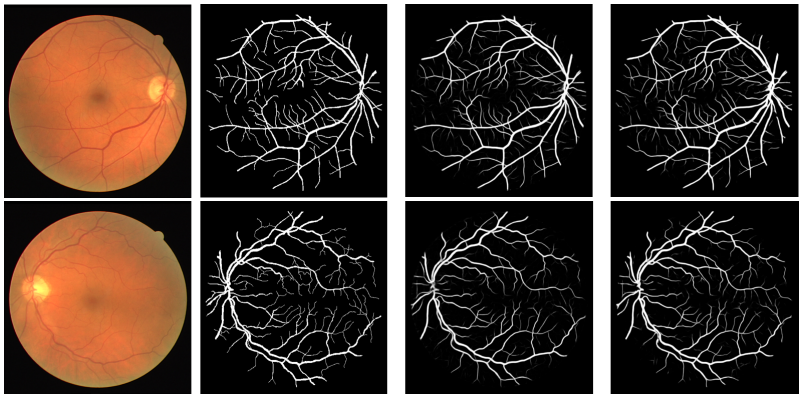

Fig. 3: Qualitative results.From left to right: RGB original image, ground-truth, segmentation at patch level, segmentation after fine-tuning at image level.

3.3 Discussion

Firstly, note that, the comparison of RBVS methods is rather biased in the sense that the field of view usually differ from one method to another. For example, the results in [3] are computed on an eroded version of the field of view. Thus, we focused on reaching competitive results and proposing an interesting training procedure that can be generalized to any medical image segmentation task.

The numerical results show that our metrics are state-of-the-art on the DRIVE dataset. On the one hand, we reach the results in [3] when working at the patch level. That is, the network is efficiently trained at the patch level in the Phase 1 of the proposed framework. On the other hand, at the image level, our transfer learning from the patch level outperforms the one from the VGG proposed in [8]. Moreover, we notice that, at least with this network, at the image level it is better to work with transfer learning than without (*see* Proposed-images and Images- nopretrain on the Table 1).

Using the framework, One can also see that the results obtained at the image level are rather close to the ones from the patch level. When trained at the patch level, the network is constrained to be precise in a small window (*i.e.* the patch). While at the image level, the constraint window becomes much larger and the network may miss some fine details. A loss function that consider the output's size or the classes' balance may improve the metrics at the image level.

4 SUMMARY & PERSPECTIVES

We proposed a fully convolutional network training framework and applied it on the task of retinal blood vessel segmentation. First, The framework is employed to train a freely designed FCN using patches extracted from the training images. Then, to meet the real-time necessity of the medical field, we fine-tuned the network using the full size images.

The training at the patch level being the first step, future work may include more ways to improve the latter such as data augmentation and preprocessing. Furthermore, we plan to examined the results on various networks such as residual networks and on other medical image modalities. Moreover, detailed studies of the patch and image levels and their correlation are left for future work.

References

1. M.M. Fraz et al., Blood vessel segmentation methodologies in retinal images A survey, Computer Methods and Programs in Biomedicine, vol. 108, no. 1, pp. 407–433, 2012.
2. Chetan L. Srinidhi, P. Aparna, and Jeny Rajan, Re- cent Advancements in Retinal Vessel Segmentation., J. Medical Systems, vol. 41, no. 4, pp. 70:1–70:22, 2017.
3. Pawel Liskowski et al., Segmenting Retinal Blood Vessels With Deep Neural Networks., IEEE Trans. Med. Imaging, vol. 35, no. 11, pp. 2369– 2380, 2016.
4. Avijit Dasgupta and Sonam Singh, A fully convolu- tional neural network based structured prediction ap- proach towards the retinal vessel segmentation., in ISBI. 2017, pp. 248-251, IEEE.
5. Roberto Vega et al., Retinal vessel extraction using Lattice Neural Net- works with dendritic processing., Comp. in Bio. and Med., vol. 58, pp. 2030, 2015.
6. Malihe Javidi et al., Vessel segmentation and microaneurysm detection us- ing discriminative dictionary learning and sparse rep- resentation., Computer Methods and Programs in Biomedicine, vol. 139, pp. 93108, 2017.
7. Taibou Birgui Sekou, Moncef Hidane, Julien Olivier, and Hubert Cardot, Segmentation of Retinal Blood Vessels Using Dictionary Learning Techniques., in FIFI/OMIA@MICCAI, M. Jorge Cardoso et al., Eds. 2017, vol. 10554 of Lecture Notes in Computer Science, pp. 8391, Springer.
8. Juan Mo and Lei Zhang, Multi-level deep supervised networks for retinal vessel segmentation., Int. J. Com- puter Assisted Radiology and Surgery, vol. 12, no. 12, pp. 21812193, 2017.
9. Jiong Zhang et al., Robust Retinal Vessel Segmentation via Locally Adaptive Derivative Frames in Orientation Scores., IEEE Trans. Med. Imaging, vol. 35, no. 12, pp. 26312644, 2016.
10. Olaf Ronneberger, Philipp Fischer, and Thomas Brox, U-Net: Convolutional Networks for Biomedical Im- age Segmentation., in MICCAI (3), Nassir Navab et al., Eds. 2015, vol. 9351 of Lecture Notes in Computer Science, pp. 234241, Springer.
11. Ian Goodfellow, Yoshua Bengio, and Aaron C. Courville, Deep Learning., Adaptive computation and machine learning. MIT Press, 2016.
12. Karen Simonyan and Andrew Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recog- nition, CoRR, vol. abs/1409.1556, 2014.
13. Shelhamer, E., Long, J., Darrell, T. ,"Fully Convolutional Networks for Semantic Segmentation". In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 4, pp. 640–651 (2017).
14. Yosinski J, Clune J, Bengio Y, and Lipson H. How transferable are features in deep neural networks? In Advances in Neural Information Processing Systems 27 (2014).
15. Joes Staal et al., Ridge- based vessel segmentation in color images of the retina., IEEE Trans. Med. Imaging, vol. 23, no. 4, pp. 501509, 2004.
16. Matthew D. Zeiler, ADADELTA: An Adaptive Learn- ing Rate Method, CoRR, vol. abs/1212.5701, 2012.