Augmenting Data When Training a CNN for Retinal Vessel Segmentation: How to Warp?

Américo Oliveira*, Sérgio Pereira*† and Carlos A. Silva*
*CMEMS-UMinho Research Unit, University of Minho, Guimarães, Portugal
Email: a68396@alunos.uminho.pt, csilva@dei.uminho.pt
†Centro Algoritmi, University of Minho, Braga, Portugal

***Abstract*—The retinal vascular condition is a trustworthy biomarker of several ophthalmologic and cardiovascular diseases, so automatic vessel segmentation is a crucial step to diagnose and monitor these problems. Deep Learning models have recently revolutionized the state-of-the-art in several fields, since they can learn features with multiple levels of abstraction from the data itself. However, these methods can easily fall into overfitting, since a huge number of parameters must be learned. Having bigger datasets may act as regularization and lead to better models. Yet, acquiring and manually annotating images, especially in the medical field, can be a long and costly procedure. Hence, when using regular datasets, people heavily need to apply artificial data augmentation. In this work, we use a fully convolutional neural network capable of reaching the state-of-the-art. Also, we investigate the benefits of augmenting data with new samples created by warping retinal fundus images with nonlinear transformations. Our results hint that may be possible to halve the amount of data, while maintaining the same performance.**

***Index Terms*—Data augmentation, Convolutional neural network, Retinal blood vessel segmentation**

I. Introduction

The eye is an organ remarkably sensitive to disorders in the human vascular system. Defects on retinal vasculature are often related with cardiovascular or ophthalmologic diseases, including arteriosclerosis, hypertension, age-related macular degeneration, diabetic retinopathy, and glaucoma [1]. Thus, retinal vessel segmentation is crucial for properly diagnose, treat, and monitor these problems. Even though manual segmentation remains indispensable, it is increasingly seen as an outdated task. Moreover, it is time-consuming and requires experienced specialists for reliability sake. Hence, automatic and accurate segmentation is becoming vital. This configures a complex task not only due to abrupt variations in the morphology and arrangement of vessels but also for abnormalities coming from diseases. As if that was not enough, retinal images commonly present poor quality [2].

At large, all previous approaches to this problem can be classified as either unsupervised or supervised learning.

On the one hand, unsupervised methods often rely on prior knowledge on the vascular structure and image intensity profiles. Among these strategies, we may include mathematical morphology [3], matched filtering techniques [4], and active contour based models [5]. One advantage of unsupervised approaches is that they waive the expensive manual annotations for training the models; in theory, this makes them well-suited for large unannotated datasets.

On the other hand, supervised methods require labeled data to learn; in retinal vessel segmentation, every pixel must be annotated as vessel or non-vessel. Thus, before these models can be applied to the test set, a classifier undergoes a learning process during a training stage. A large diversity of classifiers has been used in this field, including k-Nearest Neighbors [6], Gaussian Mixture Models [7], and Artificial Neural Networks [8]. Despite the disadvantage of requiring a training stage, these methods usually perform better than the previous ones. Still, their performance is often strongly related with the existence of effective features. If the amount of available data is reduced, extracting relevant information may even be unfeasible.

The previous supervised approaches all share one critical step: the feature engineering stage. Alternatively, Deep Learning-based approaches can learn features with multiple levels of abstraction from the data itself, bypassing this task [9]. Deep Learning-based methods have recently beaten the state-of-the-art in several fields. On the retinal vessel segmentation one, this kind of approach has also emerged. Melinscak et al. [10] addressed vessel segmentation by classifying the central pixel of an image patch using a 10-layer convolutional neural network (CNN). Contrasting, Li et al. [11] transformed the segmentation task into a cross-modality data transformation problem, where the vessel map is given by transforming the image according to a certain mapping function. The parameters of this function are learned by a 5-layer fully deep neural network. In [12], Liskowski et al. tested a comprehensive set of CNN architectures. Using a structured prediction scheme to highlight context information given by the neighborhood of each pixel, the authors reported the best results of a Deep Learning-based methodology applied to retinal vessel segmentation so far.

In fact, new CNN architectures are frequently emerging. This hints a paradigm change according to which people are now optimizing architectures, instead of designing handcrafted features which may be problem dependent, requiring expert knowledge and try-and-error approaches. One of these new CNN architectures is the so-called Fully Convolutional Network (FCN) [13], [14], that allows us to segment a full patch of pixels in one forward pass, while imposing segmentation regularization.

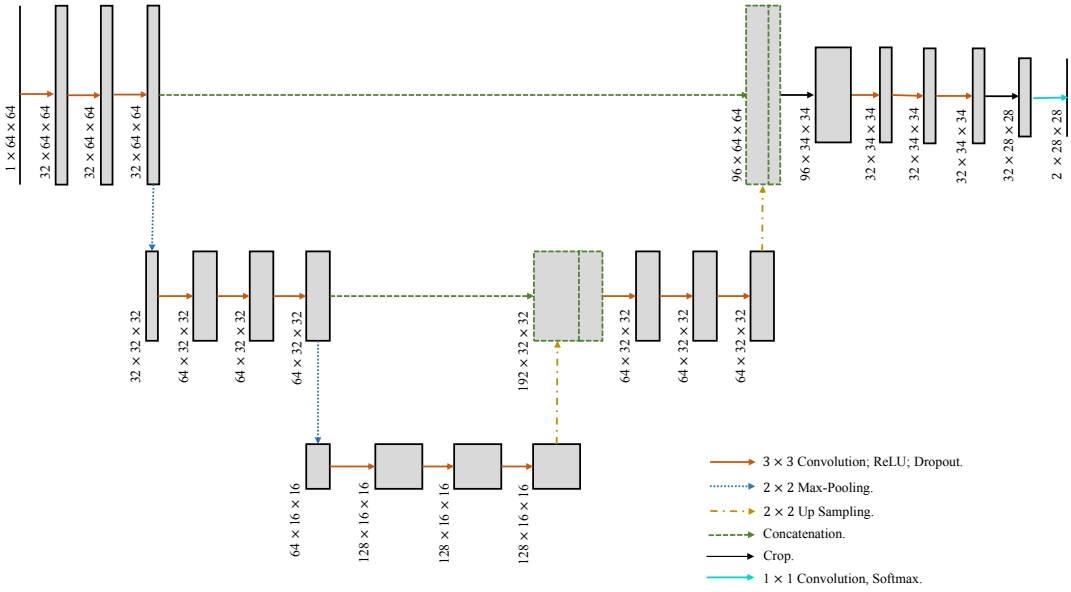

Fig. 1. Architecture of the implemented FCN.

Although it has unusual potential, Deep Learning continues to reveal some setbacks. One of the most significant is the huge amount of parameters that must be learned, which is prone to overfitting. Having a big amount of data may increase regularization, but acquiring and, especially, manually segmenting images may be a time-consuming task; hence, one solution may rely on synthesizing new data, taking as baseline already existing samples. Artificial data augmentation has already been useful to achieve better segmentation performances when using CNNs in other contexts [14], [15], [16]. The main contributions of this paper are the following. We implement a FCN capable of reaching the current state-of-the-art in retinal vessel segmentation; also, we study two nonlinear warping artificial data augmentation approaches, aiming to highlight the importance of artificial data augmentation in the context of medical image segmentation.

The remaining of this paper is organized as follows. In Section II we present the methodology. Results and discussion are shown in Section III. Finally, in Section IV, we outline the main conclusions.

II. Materials and Methods

A. Dataset

The presented approach was validated in the DRIVE database [6], consisting of 40 publicly available color images, 7 of which belonging to pathological individuals. The images were acquired using a Canon CR5 non-mydriatic 3-CCD camera with resolution of 768x584 pixels, and 8 bits per channel. The original set was divided into the training and test sets, each of them consisting of 20 images. Multiple experts manually segmented the dataset, but we use the ground truth identified as first manual for all images, as standard when using DRIVE.

B. Segmentation Method

We use a FCN and a blockwise segmentation methodology, where all the pixels of one patch are segmented at once. When applying data augmentation, each patch is replicated individually after being extracted. Next, we detail the key points of our implementation.

1) Pre-processing: We extract 2D patches from the green channel, and normalize them with zero mean and unit variance. No other pre-processing is applied.

2) Fully Convolutional Network and training: The implemented FCN is based on the U-net proposed by Ronneberger et al. [14]. Both a contracting and expanding pathways can, thus, be identified. In the contracting part, pooling is applied in order to summarize neighboring features and create higher level representations. In the expanding path, later, the feature maps are upsampled to the original resolution, enabling correspondence. When up-scaling deeper representations to the pixel space, we may lose the finer details, such as small objects, or sharper edges. Thus, lower level feature maps are then concatenated with the up-scaled ones. This is followed by convolutional layers that learn how to combine these data flows.

The implemented FCN architecture is presented in Fig. 1. As non-linear activation, we use Rectifier Linear Units (ReLU) [17], defined as $f(x) = \max(0, x)$, where x is the input. Additionally, after each convolutional layer with 3×3 kernels and ReLU activation, we use Dropout [18] with probability of $p = 0.3$. It works by removing random nodes with probability p in each training step. In this way, it acts as training regularization, preventing overfitting and node's co-adaptation. The Cross Entropy was defined as the loss function to be minimized in order to train the CNN. To that end, we used Stochastic Gradient Descent with learning rate of 0.01 as optimizer.

From the total 20 training subjects, we used 18 for training the FCN, and 2 as validation set. The FCN was implemented using Keras with Theano backend, and cuDNN 5.1.

C. Data Augmentation

In this study, artificial data augmentation consists in taking a real 2D patch and non-linearly transform it into a similar, but different one. We note that the same operation must be applied to both the image patch and the respective annotated patch to keep the correspondence with the ground truth.

Two non-linear warpings were implemented: Simard and Ronneberger transformations. The former is based on the transformation proposed by Simard et al. [19]. This deformation is created by uniformly generating a random displacement field $U(x, y) = rand(-1, +1)$ for each dimension of the image, where $rand(-1, +1)$ is a random number lying between -1 and +1. Afterwards, the fields (2 in the case of a 2D patch) are convolved with a Gaussian filter of standard deviation σ, and then multiplied by a scaling factor α that controls the intensity of the deformation. We used $\alpha = 16$, and $\sigma = 2.5$ (Fig. 2, center). From here, we will denote this data transformation as Trans1.

The second non-linear transformation is based on the work of Ronneberger et al. [14]. In this case, we begin by defining a set of N control points in a 2D grid. For those points, a deformation vector is sampled from a Gaussian distribution with standard deviation σ. Finally, the deformation vectors are interpolated to the remaining pixels. We set $N = 3$, so there are 3 control points in each dimension of the image (3×3 in total); σ was fixed on 15 (Fig. 2, right). This data transformation will be denoted as Trans2.

III. Results and Discussion

All the presented variants were evaluated in the test set of the DRIVE database. The obtained results are shown in Table I (with each implemented variant numbered from 1 to 4), where we compare them with other state-of-the-art performances, in terms of sensitivity (SEN), specificity (SPEC), area under the ROC curve (AUC), and accuracy (ACC). All metrics range from 0 to 1. Higher values mean better results.

As baseline, we started by training the network without data augmentation. In the proposed variants 1 and 2, we extracted 2750 and 5550 training patches from each image, respectively; thus, no artificial data is included in training. As we can observe, doubling the number of training samples (variant 2) improves results in almost all metrics, since more data has been shown to the network. Furthermore, to the best of our knowledge, variants 1 and 2 outperform the best reported to date results by a Deep Learning-based methodology in terms of sensitivity and accuracy, respectively.

Variants 3 and 4 result from artificially synthesizing new training samples using Trans1 and Trans2, respectively. Thereby, we picked variant 1 and duplicated its data (by non-linear warping with Trans1 and Trans2), in order to compare it with extracting the total number of training patches directly from the images (variant 2). From Table I, we can observe

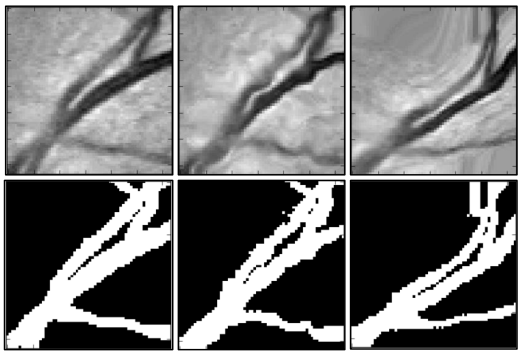

Fig. 2. Image transformations. On top, we show the image patches. On bottom, we reveal the corresponding ground-truths. From left to right: original patch, Simard transformation [19], and Ronneberger transformation [14].

that both artificial data augmentations lead to better metrics in terms of SPEC, ACC, and AUC, when compared with variant 1. We can argue that the network benefited from having more data, although SEN decreased. Hence, the artificial data introduced relevant information that was not initially available. This suggests that the implemented transformations can both be rewarding ways of augmenting data.

Particularly, we can observe that Trans1 (proposed by Simard et al. [19]) allowed the network to achieve better performances than Trans2 (proposed by Ronneberger et al. [14]) in all metrics, excepting SPEC (where the difference is probably negligible). In fact, augmenting data with artificially generated samples through Trans1 is even similar to do it with real samples (please compare variants 2 and 3). This suggests that similar results can be obtained using half of the data.

In Fig. 3, we show an example of segmentation for each of the already described variants.

All tests were conducted on a desktop equipped with a NVIDIA GeForce GTX 1080 GPU, an Intel Core i7-5930k 3.5 GHz (x12) processor, 32 GB of RAM, and running Linux Mint 18 OS. A full image segmentation takes approximately 25 seconds.

TABLE I
Results obtained in the DRIVE test set (best results are shown in bold)

	Method	SEN	SPEC	ACC	AUC
Unsupervised	Mendonça et al. [3]	0,7344	0,9764	0,9452	N.A
	Azzopardi et al. [4]	0,7655	0,9704	0,9442	0,9614
	Zhao et al. [5]	0,7420	0,9820	0,9540	0,8620
Supervised	Staal et al. [6]	N.A	N.A	0,9441	0,9520
	Soares et al. [7]	N.A	N.A	0,9466	0,9614
	Marin et al. [8]	0,7067	0,6944	0,9452	0,9588
	Melinscak et al. [10]*	0,7276	0,9785	0,9466	0,9749
	Li et al. [11]*	0,7569	0,9816	0,9527	0,9738
	Liskowski et al. [12]*	0,7811	0,9807	0,9535	**0,9790**
Proposed*	1) 2750 real	**0,8073**	0,9730	0,9517	0,9749
	2) 5500 real	0.7810	0.9800	**0.9543**	0.9768
	3) 2750 real + 2750 Trans1	0,7409	0,9853	0,9539	0,9755
	4) 2750 real + 2750 Trans2	0,7315	**0,9855**	0,9529	0,9750

*Deep Learning-based methods.

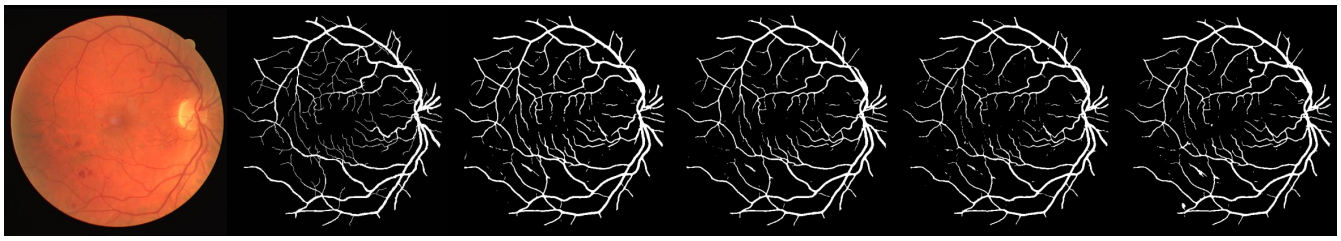

Fig. 3. Image segmentation examples. From left to right: original image, ground truth, and outputs of variants 1, 2, 3, and 4, respectively.

IV. Conclusions

Deep Learning-based methods are achieving promising results in several fields. Despite their potential, yet, they can still be impaired by the lack of available data for training.

In this work, we implemented a FCN that beats the state-of-the-art in terms of accuracy, sensitivity and specificity. Also, we studied the effect of applying elastic transformations for artificial data augmentation when segmenting retinal vessels.

When the amount of data was reduced by half, the global performance got worse. However, when the initial training set was restored by adding artificially generated samples, this behavior was reversed. In fact, we could state that increasing data with artificially generated samples, or with real ones, led to similar results, especially when using the elastic transformation proposed in [19]. This may allow us to reach effective results, while significantly reducing the amount of data typically required for a Deep Learning-model to be effective.

Acknowledgment

This work is supported by FCT with the reference project UID/EEA/04436/2013, by FEDER funds through the COMPETE 2020 Programa Operacional Competitividade e Internacionalização (POCI) with the reference project POCI-01-0145-FEDER-006941. Sérgio Pereira was supported by a scholarship from the Fundação para a Ciência e Tecnologia (FCT), Portugal (scholarship number PD/BD/105803/2014).

References

[1] M. D. Abràmoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE Reviews in Biomedical Engineering*, vol. 3, pp. 169–208, 2010.

[2] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "Blood vessel segmentation methodologies in retinal images–a survey," *Comput Meth Prog Bio*, vol. 108, no. 1, pp. 407–433, 2012.

[3] A. M. Mendonca and A. Campilho, "Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction," *IEEE T Med Imaging*, vol. 25, no. 9, pp. 1200–1213, 2006.

[4] G. Azzopardi, N. Strisciuglio, M. Vento, and N. Petkov, "Trainable cosfire filters for vessel delineation with application to retinal images," *Med Image Anal*, vol. 19, no. 1, pp. 46–57, 2015.

[5] Y. Zhao, L. Rada, K. Chen, S. P. Harding, and Y. Zheng, "Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retinal images," *IEEE T Med Imaging*, vol. 34, no. 9, pp. 1797–1807, 2015.

[6] J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. Van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE T Med Imaging*, vol. 23, no. 4, pp. 501–509, 2004.

[7] J. V. Soares, J. J. Leandro, R. M. Cesar, H. F. Jelinek, and M. J. Cree, "Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification," *IEEE T Med Imaging*, vol. 25, no. 9, pp. 1214–1222, 2006.

[8] D. Marín, A. Aquino, M. E. Gegúndez-Arias, and J. M. Bravo, "A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features," *IEEE T Med Imaging*, vol. 30, no. 1, pp. 146–158, 2011.

[9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[10] M. Melinščak, P. Prentašić, and S. Lončarić, "Retinal vessel segmentation using deep neural networks," in *VISAPP 2015 (10th International Conference on Computer Vision Theory and Applications)*, 2015.

[11] Q. Li, B. Feng, L. Xie, P. Liang, H. Zhang, and T. Wang, "A cross-modality learning approach for vessel segmentation in retinal images," *IEEE T Med Imaging*, vol. 35, no. 1, pp. 109–118, 2016.

[12] P. Liskowski and K. Krawiec, "Segmenting retinal blood vessels with deep neural networks," *IEEE T Med Imaging*, vol. 35, no. 11, pp. 2369–2380, 2016.

[13] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE T Pattern Anal*, 2016.

[14] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.

[15] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in mri images," *IEEE T Med Imaging*, vol. 35, no. 5, pp. 1240–1251, 2016.

[16] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Deep convolutional neural networks for the segmentation of gliomas in multi-sequence mri," in *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer, 2015, pp. 131–143.

[17] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807–814.

[18] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting." *J Mach Learn Res*, vol. 15, no. 1, pp. 1929–1958, 2014.

[19] P. Y. Simard, D. Steinkraus, J. C. Platt *et al.*, "Best practices for convolutional neural networks applied to visual document analysis." in *ICDAR*, vol. 3. Citeseer, 2003, pp. 958–962.