

# **HHS Public Access**

# Author manuscript

Med Image Comput Comput Assist Interv. Author manuscript; available in PMC 2021 December 22.

Published in final edited form as:

Med Image Comput Comput Assist Interv. 2021 ; 12901: 514–524. doi:10.1007/978-3-030-87193-2\_49.

# **LIFE: A Generalizable Autodidactic Pipeline for 3D OCT-A Vessel Segmentation**

**Dewei Hu**1, **Can Cui**1, **Hao Li**1, **Kathleen E. Larson**2, **Yuankai K. Tao**2, **Ipek Oguz**<sup>1</sup>

<sup>1</sup>Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, USA

<sup>2</sup>Department of Biomedical Engineering, Vanderbilt University, Nashville, TN, USA

# **Abstract**

Optical coherence tomography (OCT) is a non-invasive imaging technique widely used for ophthalmology. It can be extended to OCT angiography (OCT-A), which reveals the retinal vasculature with improved contrast. Recent deep learning algorithms produced promising vascular segmentation results; however, 3D retinal vessel segmentation remains difficult due to the lack of manually annotated training data. We propose a learning-based method that is only supervised by a self-synthesized modality named local intensity fusion (LIF). LIF is a capillary-enhanced volume computed directly from the input OCT-A. We then construct the local intensity fusion encoder (LIFE) to map a given OCT-A volume and its LIF counterpart to a shared latent space. The latent space of LIFE has the same dimensions as the input data and it contains features common to both modalities. By binarizing this latent space, we obtain a volumetric vessel segmentation. Our method is evaluated in a human fovea OCT-A and three zebrafish OCT-A volumes with manual labels. It yields a Dice score of 0.7736 on human data and  $0.8594 \pm 0.0275$ on zebrafish data, a dramatic improvement over existing unsupervised algorithms.

# **Keywords**

OCT angiography; Self-supervised; Vessel segmentation

# **1 Introduction**

Optical coherence tomography (OCT) is a non-invasive imaging technique that provides high-resolution volumetric visualization of the retina [19]. However, it offers poor contrast between vessels and nerve tissue layers [9]. This can be overcome by decoupling the dynamic blood flow within vessels from stationary nerve tissue by decorrelating multiple cross-sectional images (B-scans) taken at the same spatial location. By computing the variance of these repeated B-scans, we obtain an OCT angiography (OCT-A) volume that has better visualization of retinal vasculature than traditional OCT [15]. In contrast to other techniques such as fluorescein angiography (FA), OCT-A is advantageous because it both provides depth-resolved information in 3D and is free of risks related to dye leakage or

potential allergic reaction [9]. OCT-A is popular for studying various retinal pathologies [4,14]. Recent usage of the vascular plexus density as a disease severity indicator [12] highlights the need for vessel segmentation in OCT-A.

Unlike magnetic resonance angiography (MRA) and computed tomography angiography (CTA), OCT-A suffers from severe speckle noise, which induces poor contrast and vessel discontinuity. Consequently, unsupervised vessel segmentation approaches [2,3,22,28,33] developed for other modalities do not translate well to OCT-A. Denoising OCT/OCT-A images has thus been an active topic of research [5,13,24]. The noise is compounded in OCT-A due to the unpredictable patterns of blood flow as well as artifacts caused by residual registration errors, which lead to insufficient suppression of stationary tissue. This severe noise level, coupled with the intricate detail of the retinal capillaries, leads to a fundamental roadblock to 3D segmentation of the retinal blood vessels: the task is too challenging for unsupervised methods, and yet, obtaining manual segmentations to train supervised models is prohibitively expensive. For instance, a single patch capturing only about 5% of the whole fovea (Fig. 4f) took approximately 30 h to manually segment. The large inter-subject variability and the vast inter-rater variability which is inevitable in such a detailed task make the creation of a suitably large manual training dataset intractable.

As a workaround, retinal vessel segmentation attempts have been largely limited to 2D images with better SNR, such as the depth-projection of the OCT-A [10]. This only produces a single 2D segmentation out of a whole 3D volume, evidently sacrificing the 3D depth information. Similar approaches to segment inherently 2D data such as fundus images have also been reported [17]. Recently, Liu et al. [20] proposed an unsupervised 2D vessel segmentation method using two registered modalities from two different imaging devices. Unfortunately, multiple scans of a single subject are not typically available in practice. Further, the extension to 3D can be problematic due to inaccurate volumetric registration between modalities. Zhang et al. proposed the optimal oriented flux [18] (OOF) for 3D OCT-A segmentation [32], but, as their focus is shape analysis, neither a detailed discussion nor any numerical evaluation on segmentation are provided.

We propose the local intensity fusion encoder (LIFE), a self-supervised method to segment 3D retinal vasculature from OCT-A. LIFE requires neither manual delineation nor multiple acquisition devices. To our best knowledge, it is the first label-free learning method with quantitative validation of 3D OCT-A vessel segmentation. Fig. 1 summarizes the pipeline. Our **novel contributions** are:

- **•** An en-face denoising method for OCT-A images via local intensity fusion (LIF) as a new modality (Sect. 2.1)
- **•** A variational auto-encoder network, the local intensity fusion encoder (LIFE), that considers the original OCT-A images and the LIF modality to estimate the latent space which contains the retinal vasculature (Sect. 2.2)
- **•** Quantitative and qualitative evaluation on human and zebrafish data (Sect. 3)

### **2 Methods**

#### **2.1 Local Intensity Fusion: LIF**

Small capillaries have low intensity in OCT-A since they have slower blood flow, and are therefore hard to distinguish from the ubiquitous speckle noise. We exploit the similarity of vasculature between consecutive en-face OCT-A slices to improve the image quality. This local intensity fusion (LIF) technique derives from the Joint Label Fusion [29] and related synthesis methods [7,24,27].

Joint label fusion (JLF) [29] is a well-known multi-atlas label fusion method for segmentation. In JLF, a library of K atlases with known segmentations  $(X_k, S_k)$  is deformably registered to the target image *Y* to obtain  $(\hat{X}_k, \hat{S}_k)$ . Locally varying weight maps are computed for each atlas based on the local residual registration error between *Y*  and  $\hat{X}_k$ . The weighted sum of the  $\hat{S}_k$  provides the consensus segmentation *S* on the target image.

JLF has been extended to joint intensity fusion (JIF), an image synthesis method that does not require atlas segmentations. JIF has been used for lesion in-painting [7] and crossmodality synthesis [27]. Here, we propose a JIF variant, LIF, performing fusion between the 2D en-face slices of a 3D OCT-A volume.

Instead of an external group of atlases, for each 2D en-face slice *X* of a 3D OCT-A volume, adjacent slices within an R-neighborhood  $\{X_{-R}, ..., X_{+R}\}$  are regarded as our group of 'atlases' for *X*. Note that the atlas  $X_0$  is the target X itself, represented as the image with a red rim in Fig. 1. We perform registration using the greedy software [31]. While closely related, we note that the self-fusion method reported in [24] for tissue layer enhancement is not suitable for vessel enhancement as it tends to substantially blur and distort blood vessels [13].

Similar to a 1-D Gaussian filter along the depth axis, LIF has a blurring effect that improves the homogeneity of vessels without dilating their thickness in the *en-face* image. Further, it can also smooth the speckle noise in the background while raising the overall intensity level, as shown in Fig. 2a/2b. In order to make vessels stand out better, we introduce the contrast enhanced local intensity fusion  $(CE-LIF)^1$  in Fig. 2c. However, intensity fusion of en-face images sacrifices the accuracy of vessel diameter in the depth direction. Specifically, some vessels existing exclusively in neighboring images are inadvertently projected on the target slice. For example, the small red box in Fig. 2 highlights a phantom vessel caused by incorrect fusion. As a result, LIF and CE-LIF are not appropriate for direct use in application, in spite of the desirable improvement they offer in visibility of capillaries (e.g., large red box). In the following section, we propose a novel method that allows us to leverage LIF as an auxiliary modality for feature extraction during which these excessive projections will be filtered out.

<sup>1</sup><https://pillow.readthedocs.io/en/stable/reference/ImageEnhance.html>.

Med Image Comput Comput Assist Interv. Author manuscript; available in PMC 2021 December 22.

#### **2.2 Cross-Modality Feature Extraction: LIFE**

Liu et al. [20] introduced an important concept for unsupervised feature extraction. Two depth-projected 2D OCT-A images,  $M_1$  and  $M_2$ , are acquired using different devices on the same retina. If they are well aligned, then aside from noise and difference in style, the majority of the anatomical structure would be the same. A variational autoencoder (VAE) is set as a pix2pix translator from  $M_1$  to  $M_2$  in which the latent space  $L_{12}$  keeps full resolution.

$$
L_{12} = f_e(M_1)
$$
 and  $M'_2 = f_d(L_{12})$  (1)

If  $M_2$  is well reconstructed ( $M'_2 \approx M_2$ ), then the latent feature map  $L_{12}$  can be regarded as the common features between  $M_1$  and  $M_2$ , namely, vasculature. The encoder  $f_e$  is considered a segmentation network (Seg-Net) and the decoder  $f_d$  a synthesis network (Syn-Net).

Unfortunately, this method has several drawbacks in practice. Imaging the same retina with different devices is rarely possible even in research settings and unrealistic in clinical practice. Furthermore, the 3D extension does not appear straightforward due to the differences in image spacing between OCT devices and the difficulty of volumetric registration in these very noisy images. In contrast, we propose to use a single OCT-A volume X and its LIF,  $X_{LIF}$  as the two modalities. This removes the need for multiple devices or registration, and allows us to produce a 3D segmentation by operating on individual en-face OCT-A slices rather than a single depth-projection image. We call the new translator network local intensity fusion encoder (LIFE).

Figure 3 shows the network architecture. To reduce the influence of speckle noise, we train a residual U-Net as a denoising network (Dn-Net), supervised by LIF. For the encoder, we implement a more complex model (R2U-Net) [1] than Liu et al. [20], supervised by CE-LIF. As the decoder we use a shallow, residual U-Net to balance computational power and segmentation performance. The reparameterization trick enables gradient back propagation when sampling is involved in a deep network [16]. This sampling is achieved by  $S = \mu + \sigma$  $\cdot$  **e**, where  $\epsilon \in \mathcal{N}(0, 1)$ , and  $\mu$  and  $\sigma$  are mean and standard deviation of the latent space. The intensity ranges of all images are normalized to [0, 255]. To introduce some blurring effect, both  $L_1$  and  $L_2$  norm are added to the VAE loss function:

$$
Loss = a \sum_{i,j} |Y(i,j) - Y'(i,j)| + \frac{b}{N} \sum_{i,j} (Y(i,j) - Y'(i,j))^2
$$
 (2)

where  $(i, j)$  are pixel coordinates, N is the number of pixels, Y is CE-LIF and Y' is the output of Syn-Net.  $a = 1$  and  $b = 0.05$  are hyperparameters. Equation 2 is also used as the loss for the Dn-Net, with the LIF image as **Y**,  $a = 1$  and  $b = 0.01$ .

As discussed above, LIF enhances the appearance of blood vessels but also introduces phantom vessels because of fusion. The set of input vessel features  $\mathcal V$  in X will thus be a subset of  $\mathcal{V}_{LIF}$  in  $X_{LIF}$ . Because LIFE works to extract  $\mathcal{V} \cap \mathcal{V}_{LIF}$ , the phantom features that exist only in  $\mathcal{V}_{LIF}$  will be cancelled out as long as the model is properly trained without suffering from overfitting.

#### **2.3 Experimental Details**

**Preprocessing for Motion Artifact Removal.—**Decorrelation allows OCT-A to emphasize vessels while other tissue types get suppressed (Fig. 4a). However, this requires the repeated OCT B-scans to be precisely aligned. Any registration errors cause motion artifacts, such that stationary tissue is not properly suppressed (Fig. 4b). These appear as horizontal artifacts in *en-face* images (Fig. 4c). We remove these artifacts by matching the histogram of the artifact B-scan to its closest well-decorrelated neighbor (Fig. 4d).

**Binarization.—**To binarize the latent space  $L_{12}$  estimated by LIFE, we apply the  $2^{nd}$ Perona-Malik diffusion equation [26] followed by the global Otsu threshold [25]. Any islands smaller than 30 voxels are removed.

**Dataset.—**The OCT volumes were acquired with  $2560 \times 500 \times 400 \times 4$  pix. (spectral  $\times$  lines  $\times$  frames  $\times$  repeated frames) [6,23]. OCT-A is performed on motion- corrected [11] OCT volumes using singular value decomposition. We manually crop the volume to only retain the depth slices that contain most of the vessels near the fovea, between the ganglion cell layer (GCL) and inner plexiform layer (IPL). Three fovea volumes are used for training and one for testing. As the number of slices between GCL and IPL is limited, we aggressively augment the dataset by randomly cropping and flipping 10 windows of size [320, 320] for each en-face image. To evaluate on vessels differing in size, we labeled 3 interacting plexus near the fovea, displayed in Fig. 4e. A smaller ROI ( $120 \times 120 \times 17$ ) cropped in the center (Fig. 4f) is used for numerical evaluation.

To further evaluate the method, we train and test our model on OCT-A of zebrafish eyes, which have a simple vessel structure ideal for easy manual labeling. This also allows us to test the generalizability of our method to images from different species. Furthermore, the fish dataset contains stronger speckle noise than the human data, which allows us to test the robustness of the method to high noise. 3 volumes ( $480 \times 480 \times 25 \times 5$  each) are labeled for testing and 5 volumes are used for training. All manual labelling is done on ITKSnap [30].

**Baseline Methods.—**Due to the lack of labeled data, no supervised learning method is applicable. Similar to our approach that follows the enhance + binarize pattern, we apply Frangi's multi-scale vesselness filter [8] and optimally oriented flux (OOF) [18,32] respectively to enhance the artifact-removed original image, then use the same binarization steps described above. We also present results using Otsu thresholding and k-means clustering.

**Implementation details.—**All networks are trained on an NVIDIA RTX 2080TI 11GB GPU for 50 epochs with batch size set to 2. For the first 3 epochs, the entire network uses the same Adam optimizer with learning rate of 0.001. After that, LIFE and decoder are separately optimized with starting learning rates of 0.002 and 0.0001 respectively in order to distribute more workload on the LIFE. Both networks decay every 3 epochs with at a rate of 0.5.

# **3 Results**

Figure 5 displays examples of extracted latent images. It is visually evident that LIFE successfully highlights the vasculature. Compared with the raw input, even delicate capillaries show improved homogeneity and separability from the background. Figure 6 illustrates 2D segmentation results within the manually segmented ROI, where LIFE can be seen to have better sensitivity and connectivity than the baseline methods. Figure 6 also shows a 3D rendering (via marching cubes [21]) of each method. In the middle row, we filtered out the false positives (FP) to highlight the false negatives (FN). These omitted FP areas are highlighted in yellow in the bottom row. It is easy to see that these FPs are often distributed along horizontal lines, caused by unresolved motion artifacts. Hessian-based methods appear especially sensitive to motion artifacts and noise; hence Frangi's method and OOF introduce excessive FP. Clearly, LIFE achieves the best preservation in small capillaries, such as the areas highlighted in white boxes, without introducing too many FPs.

Figure 7 shows that LIFE has superior performance on the zebrafish data. The white boxes highlight that only LIFE can capture smaller branches.

Figure 8 shows quantitative evaluation across B-scans, and Table 1 across the whole volume. Consistent with our qualitative assessments, LIFE significantly ( $p \ll 0.05$ ) and dramatically (over 0.20 Dice gain) outperforms the baseline methods on both human and fish data.

Finally, we directly binarize LIF and CE-LIF as additional baselines. The Dice scores on the human data are 0.5293 and 0.4892, well below LIFE (0.7736).

# **4 Discussion and Conclusion**

We proposed a method for 3D segmentation of fovea vessels and capillaries from OCT-A volumes that requires neither manual annotation nor multiple image acquisitions to train. The introduction of the LIF modality brings many benefits for the method. Since LIF is directly computed from the input data, no inter-volume registration is needed between the two modalities input to LIFE. Further, rather than purely depending on image intensity, LIF exploits local structural information to enhance small features like capillaries. Still, there are some disadvantages to overcome in future research. For instance, LIFE cannot directly provide a binarized output and hence the crude thresholding method used for binarization influences the segmentation performance.

### **Acknowledgements.**

This work is supported by NIH R01EY031769, NIH R01EY030490 and Vanderbilt University Discovery Grant Program.

# **References**

- 1. Alom MZ, Hasan M, Yakopcic C, Taha TM, Asari VK: Recurrent residual CNN based on u-net (r2u-net) for medical image segmentation. arXiv preprint arXiv:1802.06955 (2018)
- 2. Aylward SR, Bullitt E: Initialization, noise, singularities, and scale in height ridge traversal for tubular object centerline extraction. IEEE Trans. Med. Imaging 21(2), 61–75 (2002). 10.1109/42.993126 [PubMed: 11929106]

- 3. Bozkurt F, Köse C, Sarı A: A texture-based 3d region growing approach for segmentation of ica through the skull base in cta. Multimedia Tools Appl. 79(43), 33253–33278 (2020)
- 4. Burke TR, et al. : Application of oct-angiography to characterise the evolution of chorioretinal lesions in acute posterior multifocal placoid pigment epitheliopathy. Eye 31(10), 1399–1408 (2017) [PubMed: 28983094]
- 5. Devalla SK, et al. : A deep learning approach to denoise OCT images of the optic nerve head. Sci. Rep 9(1), 1–13 (2019) [PubMed: 30626917]
- 6. El-Haddad MT, Bozic I, Tao YK: Spectrally encoded coherence tomography and reflectometry: Simultaneous en face and cross-sectional imaging at 2 gigapixels per second. J. Biophotonics 11(4), e201700268 (2018) [PubMed: 29149542]
- 7. Fleishman GM, et al. : Joint intensity fusion image synthesis applied to MS lesion segmentation. In: MICCAI BrainLes Workshop, pp. 43–54 (2017)
- 8. Frangi AF, Niessen WJ, Vincken KL, Viergever MA: Multiscale vessel enhancement filtering. In: Wells WM, Colchester A, Delp S (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 130–137. Springer, Heidelberg (1998). 10.1007/BFb0056195
- 9. Gao S, et al. : Optical coherence tomography angiography. IOVS 57(9), OCT27–OCT36 (2016)
- 10. Giarratano Y, et al. : Automated and network structure preserving segmentation of optical coherence tomography angiograms. arXiv preprint arXiv:1912.09978 (2019)
- 11. Guizar-Sicairos M, Thurman ST, Fienup JR: Efficient subpixel image registration algorithms. Opt. Lett 33(2), 156–158 (2008) [PubMed: 18197224]
- 12. Holló G: Comparison of peripapillary oct angiography vessel density and retinal nerve fiber layer thickness measurements for their ability to detect progression in glaucoma. J. glaucoma 27(3), 302–305 (2018) [PubMed: 29303879]
- 13. Hu D, Malone J, Atay Y, Tao Y, Oguz I: Retinal OCT denoising with pseudo-multimodal fusion network. In: MICCAI OMIA, pp. 125–135 (2020)
- 14. Ishibazawa A, et al. : OCT angiography in diabetic retinopathy: a prospective pilot study. Am. J. Ophthalmol 160(1), 35–44 (2015) [PubMed: 25896459]
- 15. Jia Y, et al. : Split-spectrum amplitude-decorrelation angiography with optical coherence tomography. Opt. Express 20(4), 4710–4725 (2012) [PubMed: 22418228]
- 16. Kingma DP, Welling M: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)
- 17. Lahiri A, Roy AG, Sheet D, Biswas PK: Deep neural ensemble for retinal vessel segmentation in fundus images towards achieving label-free angiography. In: IEEE EMBC, pp. 1340–1343. IEEE (2016)
- 18. Law MWK, Chung ACS: 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). 10.1007/978-3-540-88693-8\_27
- 19. Li M, Idoughi R, Choudhury B, Heidrich W: Statistical model for oct image denoising. Biomed. Opt. Express 8(9), 3903–3917 (2017) [PubMed: 29026678]
- 20. Liu Y, et al. : Variational intensity cross channel encoder for unsupervised vessel segmentation on oct angiography. In: SPIE Medical Imaging 2020: Image Processing, vol. 11313, p. 113130Y (2020)
- 21. Lorensen WE, Cline HE: Marching cubes: a high resolution 3d surface construction algorithm. SIGGRAPH Comput. Graph 21(4), 163–169 (1987)
- 22. Lorigo LM, et al. : CURVES: curve evolution for vessel segmentation. Med. Image Anal 5(3), 195–206 (2001) [PubMed: 11524226]
- 23. Malone JD, El-Haddad MT, Yerramreddy SS, Oguz I, Tao YK: Handheld spectrally encoded coherence tomography and reflectometry for motion-corrected ophthalmic OCT and OCT-A. Neurophotonics 6(4), 041102 (2019) [PubMed: 32042852]
- 24. Oguz I, Malone JD, Atay Y, Tao YK: Self-fusion for OCT noise reduction. In: SPIE Medical Imaging 2020: Image Processing, vol. 11313, p. 113130C (2020)
- 25. Otsu N: A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern 9(1), 62–66 (1979)

- 26. Perona P, Malik J: Scale-space and edge detection using anisotropic diffusion. IEEE Trans. Pattern Anal. Mach.Intell 12(7), 629–639 (1990)
- 27. Ufford K, Vandekar S, Oguz I: Joint intensity fusion with normalized cross-correlation metric for cross-modality MRI synthesis. In: SPIE Medical Imaging 2020: Image Processing, vol. 11313 (2020)
- 28. Vasilevskiy A, Siddiqi K: Flux maximizing geometric flows. IEEE Trans. Pattern Anal. Mach. Intell 24, 1565–1578 (2001)
- 29. Wang H, Suh JW, Das SR, Pluta JB, Craige C, Yushkevich PA: Multiatlas segmentation with joint label fusion. IEEE PAMI 35(3), 611–623 (2012)
- 30. Yushkevich PA, et al. : User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 31(3), 1116–1128 (2006) [PubMed: 16545965]
- 31. Yushkevich PA, Pluta J, Wang H, Wisse LE, Das S, Wolk D: Fast automatic segmentation of hippocampal subfields and medial temporal lobe subregions in 3T and 7T T2-weighted MRI. Alzheimer's Dement. 7(12), P126–P127 (2016)
- 32. Zhang J, et al. : 3d shape modeling and analysis of retinal microvasculature in oct-angiography images. IEEE TMI 39(5), 1335–1346 (2020)
- 33. Zhao S, Tian Y, Wang X, Xu P, Deng Q, Zhou M: Vascular extraction using mra statistics and gradient information. Mathematical Problems in Engineering 2018 (2018)





Overall pipeline. \* indicates LIF and CE-LIF provide supervision for Dn-Net and LIFE in training process respectively.



**Fig. 2.** 

Modalities of en-face OCT-A. **Large red box** highlights improvement in capillary visibility. **Small red box** points out a phantom vessel.







### **Fig. 4.**

**(a-d)** Horizontal motion artifacts and their removal. **(e,f)** Three manually labelled vessel plexuses. Only the branches contained within the ROI were fully segmented; any branches outside the ROI and all other trees were omitted for brevity.



# **Fig. 5.**

The latent image  $L_{12}$  from LIFE considerably improves vessel appearance.



#### **Fig. 6.**

2D slice and 3D marching cubes rendering of segmentation results on human retina, with Gaussian smoothing,  $\sigma = 0.70$ . **Red, green, blue** show three different branches; **yellow** highlights false positives. LIFE is the only method that can recover the 3D structure and connectivity of the capillaries outside the largest vessels without causing excessive FP (yellow). **White boxes** highlight LIFE's improved sensitivity.



# **Fig. 7.**

Segmentation result of zebrafish retina with the same rendering setting in Fig. 6.



#### **Fig. 8.**

Quantitative result evaluation for **(left)** human and **(right)** zebrafish data. TPR: true positive rate, FPR: false positive rate, Acc: accuracy.

# **Table 1.**

Quantitative evaluation of human and zebrafish segmentation. TPR: true positive rate, FPR: false positive rate. Bold indicates the best score per column.

