Variational Intensity Cross Channel Encoder for Unsupervised Vessel Segmentation on OCT Angiography

Yihao Liu[a], Lianrui Zuo[a], Aaron Carass[a,b], Yufan He[a], Angeliki Filippatou[c], Sharon D. Solomon[d], Shiv Saidha[c], Peter A. Calabresi[c], and Jerry L. Prince[a,b]

[a]Department of Electrical and Computer Engineering, The Johns Hopkins University
[b]Department of Computer Science, The Johns Hopkins University
[c]Department of Neurology, The Johns Hopkins University School of Medicine
[d]Wilmer Eye Institute, The Johns Hopkins University School of Medicine

ABSTRACT

Deep learning approaches have been used extensively for medical image segmentation tasks. Training deep networks for segmentation, however, typically requires manually delineated examples which provide a ground truth for optimization of the network. In this work, we present a neural network architecture that segments vascular structures in retinal OCTA images without the need of direct supervision. Instead, we propose a variational intensity cross channel encoder that finds vessel masks by exploiting the common underlying structure shared by two OCTA images of the the same region but acquired on different devices. Experimental results demonstrate significant improvement over three existing methods that are commonly used.

Keywords: OCTA, Vessel segmentation, Variational autoencoder, Unsupervised segmentation, Image synthesis

1. INTRODUCTION

Optical coherence tomography (OCT) based angiography (OCTA) is a recently developed, non-invasive imaging modality that provides insight into retinal vascular morphology. Current clinical use of OCTA relies on measurements like the mean vessel density computed on binarized *en-face* angiography of the superfical or deep vascular plexus.[1,2] The binarization, however, depends greatly on the quality of the scans, the binarization method that is applied, and the algorithm used to generate the OCTA.[3,4] Although there have been studies on artifacts and their removal in OCTA,[1,5–7] it is currently difficult to remove all image artifacts that are present in OCTA. In a typical workflow, poor quality images are excluded by human reviewer, in order to get more consistent measurements.[2,8] Additionally, to segment vessels from OCTA manually is extremely difficult because of the widespread presence of artifacts, poor image contrast, and the abundance of fine capillary details that must be traced.

Deep learning has been used extensively in medical image segmentation tasks.[9–11] However, investigation of applying deep learning to segmenting vessels in retinal OCTA images is limited in the literature,[12,13] because of the above stated challenges. In addition, a trained network model is dedicated to the scanner where the training data is acquired. Usually, it cannot generalize to images from other scanners, due to the contrast and noise level differences as well as the presence of various artifacts. In this paper, we propose the variational intensity cross channel encoder (VICCE) which produces artifact robust and consistent binarization across different devices. Our proposed architecture is similar to a traditional variational autoencoder[14] but with cross channel "reconstruction" of the target image. The latent space produced by the encoder has the same spatial dimension as the input, effective encoding the intensity information. VICCE is trained without any manual delineated data and is capable of producing state-of-the-art vessel segmentations in the presence of various artifacts.

Send correspondence to Yihao Liu (yliu236@jhu.edu)

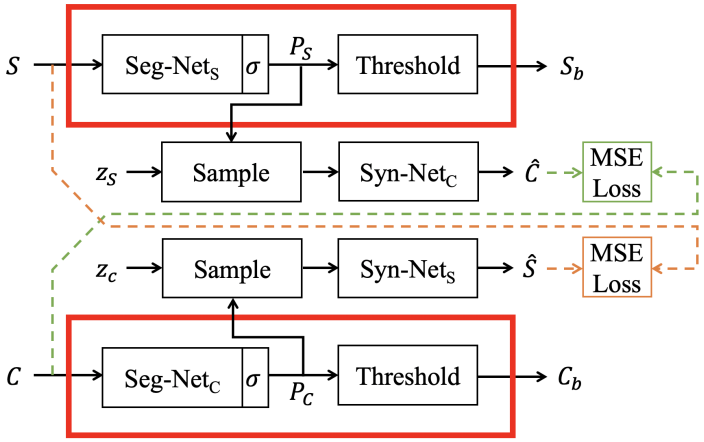


Figure 1. Diagram of the overall network architecture. S and C stand for Spectralis OCTA and Cirrus OCTA, and $\hat{S}$ and $\hat{C}$ are the synthetic results. The binarization of S and C, denoted as S_b and C_b, are obtained by thresholding the Bernoulli probability P_S and P_c at each pixel location. The symbol σ represents a sigmoid function attached right after each `Seg-Net`. The sample block represent the "reparameterization trick" for Bernoulli random variable, z_S and z_C are fields of uniform random variables

2. DATASET

The data used in this work were obtained from a Spectralis OCTA (Heidelberg Engineering, Heidelberg, Germany) scanner and a Cirrus OCTA (Carl Zeiss Meditec Inc, Dublin, CA) scanner. Thirty three subjects with multiple sclerosis (MS) and two healthy controls (HCs) are included. For every subject, each eye was scanned by both devices and the superficial vascular plexus (SVP) angiograms were automatically generated and exported from the devices. Each scan covers a 3mm $\times$ 3mm area centered at the fovea.

3. METHOD

Our overall network architecture is shown in Fig. 1, two final segmentation networks (`Seg-Net`$_S$ and `Seg-Net`$_C$) are used to produce the desired segmentation map after training (red boxes): in the Spectralis channel, `Seg-Net`$_S$ takes a Spectralis angiogram (S) as input and outputs a probability map P_S that indicates the presence of vascular structure in S; similarly for the Cirrus channel, `Seg-Net`$_C$ takes a Cirrus angiogram (C) and outputs probability map P_C. Both P_S and P_C have the same spatial size as their input, and they are thresholded at intensity 0.5 to give the final binary segmentation S_b and C_b.

In cases where direct supervision is available, training the segmentation networks is straight forward: first compute a loss between the `Seg-Net` output and the given ground truth, then use back propagation to optimize the network parameters. In this work, in absence of any ground truth, we introduce two synthesis networks: `Syn-Net`$_C$ for the Spectralis channel, and `Syn-Net`$_S$ for the Cirrus channel that take the output of the two segmentation networks. Through the synthesis networks, P_S is used to synthesize the Cirrus image $\hat{C}$ and P_C is used to synthesize the Spectralis image $\hat{S}$. Therefore, training VICCE requires pairs of OCTA superficial vascular plexus angiograms exported from two different devices—Spectralis and Cirrus. One serves as the input to the `Seg-Net` and the other as the synthesis ground truth, for each channel. To provide a pixel correspondence between the input (S and C) and the synthesis target ($\hat{C}$ and $\hat{S}$), we pre-process those paired angiograms by registering the Cirrus to the Spectralis using mutual information as a cost metric under an affine transformation. The registered Cirrus and Spectralis are then cropped to have matching fields of view.

This cross channel synthesis design allows the synthesis loss, computed as the mean squared error between the synthetic angiogram and the real angiogram, to be back-propagated to `Seg-Net`$_S$ and `Seg-Net`$_C$, forcing them to extract useful information from the input to synthesize across modalities. Such information is, in fact, the underlying vascular structure shared by S and C. Compared with a regular autoencoder, VICCE replaces self reconstruction with cross channel synthesis. This provides additional benefits—artifact and noise suppression—because the input images cannot have the same artifacts and noise pattern as the target, capturing those elements

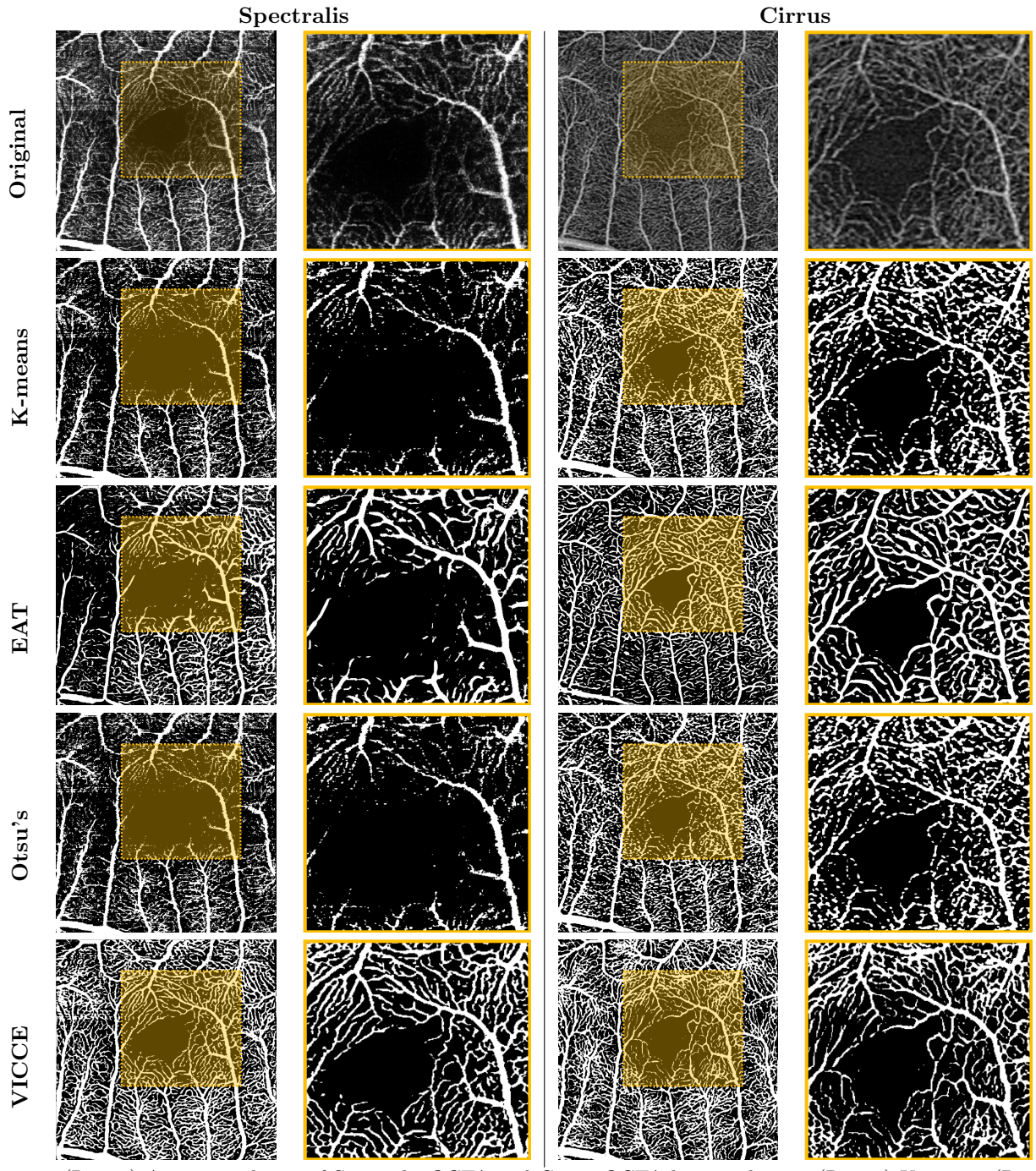
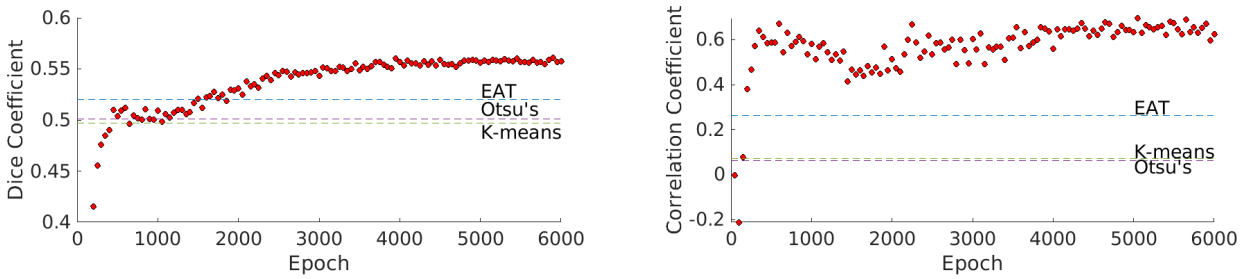


Figure 2. (Row 1) A registered pair of Spectralis OCTA and Cirrus OCTA binarized using (Row 2) K-means, (Row 3) Enhancement by Frangi followed by local adaptive thresholding, (Row 4) Global Otsu's method, and (Row 5) the proposed method. The shaded yellow patch in the first and third column are magnified and shown in the second and fourth column for better visualization.

will increase the synthesis loss, and thus the networks reduce their presence. Because the latent space (P_S and P_C) has the same spatial dimension as the input, and the grayscale level of the latent space is now encoded as Bernoulli random variables, we interpret `Seg-Nets` as intensity encoders that encode the grayscale level of the input image.



(a) **(b)**

Figure 3. The performance of VICCE for the validation data over the training epochs. **(a)** Dice coefficient of S_b and C_b, and **(b)** the correlation coefficient of the computed vessel density.

Because $\texttt{Seg-Net}_\texttt{S}$ and $\texttt{Seg-Net}_\texttt{C}$ are not supervised, the probability maps are not intrinsically driven to be near 0 for background and near 1 for vessels. Without such a constraint, it is difficult to find appropriate thresholds to yield the desired binary segmentations. VICCE solves this problem by introducing a variational sampling stage prior to the synthesis networks. Sampling is accomplished by a "reparameterization trick" for Bernoulli random variables.[15,16] At location (x, y) in the Spectralis channel the output of sampling is given as

$$\sigma\left\{\ln(z_S(x,y)) - \ln(1 - z_S(x,y)) + \ln(P_S(x,y)) - \ln(1 - P_S(x,y))\right\}, \tag{1}$$

where z_S is a field of uniform random variables, and σ is the sigmoid function. This output is interpreted as a random sample of the Bernoulli distribution with probability $P_S(x, y)$. During training, z_S and z_C are resampled for each training pair. The effect of this random sampling is to create "harder" probability maps P_S and P_C, which in turn makes it possible to choose 0.5 as the threshold for producing the binarized images S_b and C_b. This is because a "harder" P_S and P_C is more likely to produce predictable samples, and thus better synthetic images.

4. RESULTS

4.1 Network and training details

After pre-processing, poorly registered pairs were manually identified and excluded. The resultant 65 pairs of registered Spectralis-Cirrus angiograms were divided into 43 pairs for training, 15 pairs for validation, and 7 pairs for testing. Images were resized to 512×512 pixels. We augmented our data with random horizontal and vertical flips. Random transpose is not used because motion artifacts are always present as horizontal lines. Our four networks use a five-level U-net architecture.[17] The downsample operations were achieved by 2×2 maxpooling. There were 8 channels of feature maps at the finest resolution. After downsample, the number of channels was doubled. Skip connections were use at each of the top four levels. All convolutional layers had a receptive field of 3×3 with zero padding followed by instance normalization. For upsampling, the transpose convolution used in the traditional U-net was replaced by bilinear interpolation and a 3×3 convolutional layer. Before each downsample and upsample operation, a dropout layer with dropout rate 0.5 is applied. We find the dropout layer essential for producing segmentations with connected capillary networks, as shown in Fig. 5.

All four networks were trained for $6,000$ epochs using Adam optimizer. The learning rate for the two segmentation networks was set to 0.001 and for the two synthesis networks was 0.00005. This encourages denoising and artifact removal to be done by the segmentation networks, which in turn encourages artifact and noise suppressed binarization. A sigmoid layer was attached after $\texttt{Seg-Net}$s to ensure P_S and P_C were between $[0, 1]$. To prevent the networks from entering the other local minima, where background is bright and vessels are dark, we added an MSE loss between the probability map (P_S and P_C) and the input (S and C). This loss decayed and had almost no effect after 200 epochs. We assumed no prior probability of the presence of vessel, thus there was no KL-divergence loss term for the Bernoulli probability map.

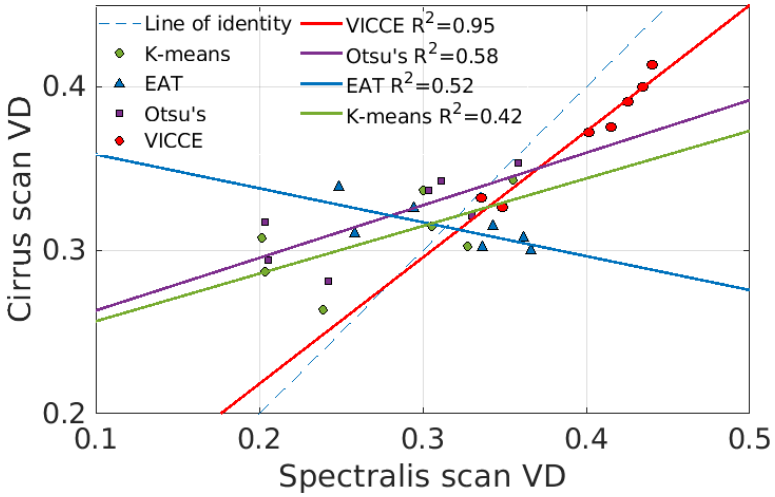


Figure 4. The vessel density for seven pairs of test images by VICCE and three other methods. The least square fitted lines are shown in the same color as the data points.

4.2 Validation and testing

We compare VICCE with three other commonly used methods: Global K-means binarization, which is also the default ImageJ binarization method; Enhancement by Frangi filter followed by local adaptive threshold (EAT); and Global Otsu's method. Methods similar to EAT have been previously reported.[18,19] From the 15 pairs of validation data, the Dice coefficient[20] of the segmented vessels and the correlation coefficient of the vessel density (VD) are computed between Spectralis OCTA and Cirrus OCTA. Because the same underlying anatomy structure are imaged, we expect those metrics to be the same regardless of the scanner type. We find that even without enforcing the similarity between two binary maps, both metrics are higher with VICCE after $2,000$ training epochs (Fig. 3(a,b)). We used the model at $5,000$ epochs and tested on 7 pairs of Spectralis and Cirrus angiograms. An example of binarization produced by VICCE and the other methods is shown in Fig. 2. A major banding artifact is present in the Spectralis OCTA image, which result in the Spectralis binarized image being unuseable for K-means, EAT, and Otsu's methods. Whereas the VICCE result successfully recovers the capillaries even within the region of the artifacts. The vessel densities of all 7 pairs of test images are shown in Fig. 4. For each method, the least square fitted lines are shown in the same color as the data points. It is evident that VICCE line (red) is closer to the line of identity. It proves that VICCE produces more consistent vessel density measures across the two scanners. The coefficient of determination between VD computed from the Cirrus versus Spectralis images are much higher ($R^2 = 0.95$) than the three competing methods. Which suggests the potential of combining vessel density measures from scans that are acquired using different scanners.

In addition to the binarized images, we also produce the synthetic images ($\hat{C}$ and $\hat{S}$) during testing time using the `Syn-Nets` (Fig. 5). Notice that the synthetic images are almost noise and artifact free, with detailed capillary networks. This further demonstrates the advantage of cross channel synthesis, and also the possibility of using VICCE to produce harmonized OCTA images, similar to those used in magnetic resonance imaging.[21]

5. CONCLUSION

We proposed a novel network architecture for unsupervised segmentation. It is trained using registered paired data but each channel can be used individually after training. In the application of vessel segmentation in OCTA, it is evident that the binarized images from different modality using VICCE are similar and do not suffer as much from artifacts as do the other three comparing methods. The results indicate much greater consistency between the computed measurements derived from the Cirrus and Spectralis OCTA images—the two devices are essentially equivalent when using the VICCE binarization.

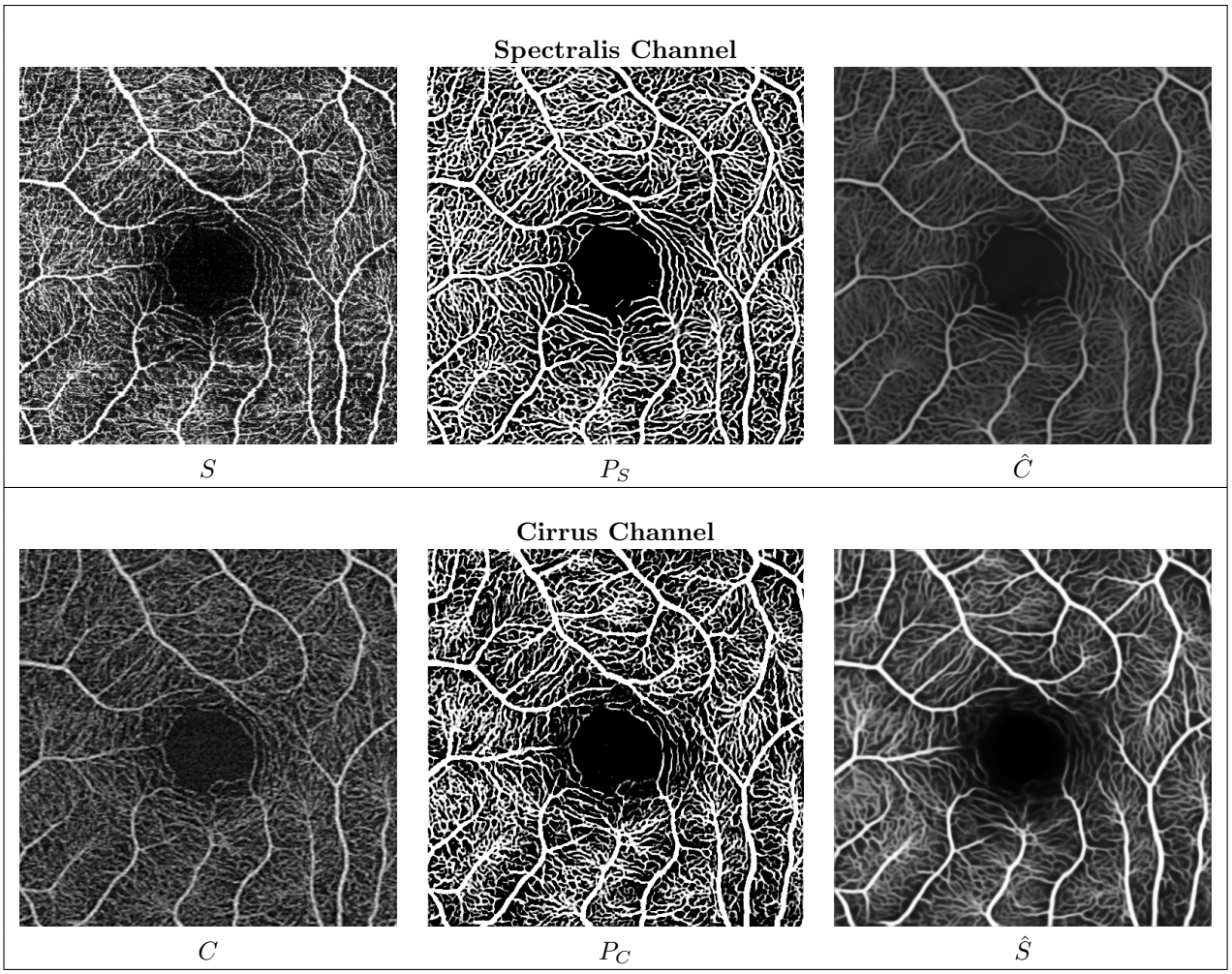


Figure 5. The result of synthesis during test time. In Spectralis channel, the Spectralis image (S) passes through $\texttt{Seg-Net}_\texttt{S}$ to produce P_S, which is used to synthesize its corresponding Cirrus image ($\hat{C}$) through $\texttt{Syn-Net}_\texttt{C}$. Similarly in the Cirrus channel, a Cirrus image (C) is used as input to VICCE and produce P_C and synthesize its Spectralis counterpart ($\hat{S}$).

VICCE architecture can be easily extended to more than two modalities. Either by adding more channels or by using a universal encoder to separate vessel segmentation and scanner dependent information, and a universal decoder. Future work will focus on adding more constraints to the `Syn-Net`s for a more interpretable latent space, currently it is achieved by using a small learning rate on the `Syn-Net`s.

6. ACKNOWLEDGMENT

This work was supported by the NIH/NEI grant R01-EY024655 and NIH/NINDS grant R01-NS082347.

REFERENCES

[1] Fenner, B. J., Tan, G. S., Tan, A. C., Yeo, I. Y., Wong, T. Y., and Cheung, G. C., "Identification of imaging features that determine quality and repeatability of retinal capillary plexus density measurements in oct angiography," *British Journal of Ophthalmology* **102**(4), 509–514 (2018).

[2] Murphy, O. C., Kwakyi, O., Iftikhar, M., Zafar, S., Lambe, J., Pellegrini, N., Sotirchos, E. S., Gonzalez-Caldito, N., Ogbuokiri, E., Filippatou, A., et al., "Alterations in the retinal vasculature occur in multiple sclerosis and exhibit novel correlations with disability and visual function measures," *Multiple Sclerosis Journal* , 1352458519845116 (2019).

[3] Jia, Y., Tan, O., Tokayer, J., Potsaid, B., Wang, Y., Liu, J. J., Kraus, M. F., Subhash, H., Fujimoto, J. G., Hornegger, J., et al., "Split-spectrum amplitude-decorrelation angiography with optical coherence tomography," *Optics express* **20**(4), 4710–4725 (2012).

[4] An, L. and Wang, R. K., "In vivo volumetric imaging of vascular perfusion within human retina and choroids with optical micro-angiography," *Optics Express* **16**(15), 11438–11452 (2008).

[5] Falavarjani, K. G., Al-Sheikh, M., Akil, H., and Sadda, S. R., "Image artefacts in swept-source optical coherence tomography angiography," *British Journal of Ophthalmology* **101**(5), 564–568 (2017).

[6] Spaide, R. F., Fujimoto, J. G., and Waheed, N. K., "Image artifacts in optical coherence angiography," *Retina (Philadelphia, Pa.)* **35**(11), 2163 (2015).

[7] Liu, Y., Carass, A., Filippatou, A., He, Y., Solomon, S. D., Saidha, S., Calabresi, P. A., and Prince, J. L., "Projection artifact suppression for inner retina in oct angiography," in [*2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*], 592–596, IEEE (2019).

[8] Yarmohammadi, A., Zangwill, L. M., Diniz-Filho, A., Suh, M. H., Manalastas, P. I., Fatehee, N., Yousefi, S., Belghith, A., Saunders, L. J., Medeiros, F. A., et al., "Optical coherence tomography angiography vessel density in healthy, glaucoma suspect, and glaucoma eyes," *Investigative ophthalmology & visual science* **57**(9), OCT451–OCT459 (2016).

[9] He, Y., Carass, A., Liu, Y., Jedynak, B. M., Solomon, S. D., Saidha, S., Calabresi, P. A., and Prince, J. L., "Deep learning based topology guaranteed surface and mme segmentation of multiple sclerosis subjects from retinal oct," *Biomedical optics express* **10**(10), 5042–5058 (2019).

[10] He, Y., Carass, A., Liu, Y., Jedynak, B. M., Solomon, S. D., Saidha, S., Calabresi, P. A., and Prince, J. L., "Fully convolutional boundary regression for retina oct segmentation," in [*International Conference on Medical Image Computing and Computer-Assisted Intervention*], 120–128, Springer (2019).

[11] Han, S., He, Y., Carass, A., Ying, S. H., and Prince, J. L., "Cerebellum parcellation with convolutional neural networks," in [*Medical Imaging 2019: Image Processing*], **10949**, 109490K, International Society for Optics and Photonics (2019).

[12] Prentašić, P., Heisler, M., Mammo, Z., Lee, S., Merkur, A., Navajas, E., Beg, M. F., Šarunic, M., and Lončarić, S., "Segmentation of the foveal microvasculature using deep learning networks," *Journal of biomedical optics* **21**(7), 075008 (2016).

[13] Heisler, M., Chan, F., Mammo, Z., Balaratnasingam, C., Prentasic, P., Docherty, G., Ju, M., Rajapakse, S., Lee, S., Merkur, A., et al., "Deep learning vessel segmentation and quantification of the foveal avascular zone using commercial and prototype oct-a platforms," *arXiv preprint arXiv:1909.11289* (2019).

[14] Kingma, D. P. and Welling, M., "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114* (2013).

[15] Jang, E., Gu, S., and Poole, B., "Categorical reparameterization with gumbel-softmax," *arXiv preprint arXiv:1611.01144* (2016).

[16] Maddison, C. J., Mnih, A., and Teh, Y. W., "The concrete distribution: A continuous relaxation of discrete random variables," *arXiv preprint arXiv:1611.00712* (2016).

[17] Ronneberger, O., Fischer, P., and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," in [*International Conference on Medical image computing and computer-assisted intervention*], 234–241, Springer (2015).

[18] Kim, A. Y., Chu, Z., Shahidzadeh, A., Wang, R. K., Puliafito, C. A., and Kashani, A. H., "Quantifying microvascular density and morphology in diabetic retinopathy using spectral-domain optical coherence tomography angiography," *Investigative ophthalmology & visual science* **57**(9), OCT362–OCT370 (2016).

[19] Zhang, M., Hwang, T. S., Dongye, C., Wilson, D. J., Huang, D., and Jia, Y., "Automated quantification of nonperfusion in three retinal plexuses using projection-resolved optical coherence tomography angiography in diabetic retinopathy," *Investigative ophthalmology & visual science* **57**(13), 5101–5106 (2016).

[20] Dice, L. R., "Measures of the amount of ecologic association between species," *Ecology* **26**(3), 297–302 (1945).

[21] Dewey, B. E., Zhao, C., Reinhold, J. C., Carass, A., Fitzgerald, K. C., Sotirchos, E. S., Saidha, S., Oh, J., Pham, D. L., Calabresi, P. A., et al., "Deepharmony: A deep learning approach to contrast harmonization across scanner changes," *Magnetic resonance imaging* **64**, 160–170 (2019).