Check for updates

Received: 23 June 2022 | Revised: 2 August 2022 | Accepted: 7 September 2022

DOI: 10.1002/jbio.202200191

JOURNAL OF BIOPHOTONICS

LETTER

Improved segmentation of collagen second harmonic generation images with a deep learning convolutional neural network

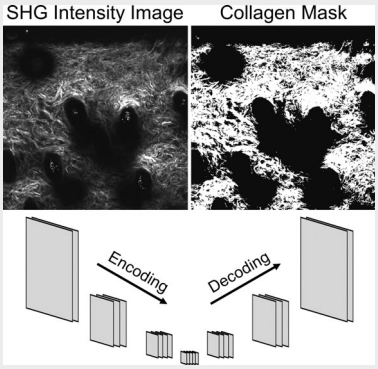
Alan E. Woessner | **Kyle P. Quinn**

Department of Biomedical Engineering, University of Arkansas, Fayetteville, Arkansas, USA

Correspondence
Kyle P. Quinn, Department of Biomedical Engineering, University of Arkansas, Fayetteville, AR, USA.
Email: kyle@quinnlab.org

Funding information
Arkansas Biosciences Institute; National Institute of Biomedical Imaging and Bioengineering, Grant/Award Numbers: R00EB017723, R01EB031032; National Institute of General Medical Sciences, Grant/Award Number: P20GM139768; National Institute on Aging, Grant/Award Number: R01AG056560; National Science Foundation, Grant/Award Number: 1846853

Abstract



Collagen fibers play an important role in both the structure and function of various tissues in the human body. Visualization and quantitative measurements of collagen fibers are possible through imaging modalities such as second harmonic generation (SHG), but accurate segmentation of collagen fibers is difficult for datasets involving variable imaging depths due to the effects of scattering and absorption. Therefore, an objective approach to segmentation is needed for datasets with images of variable SHG intensity. In this study, a U-Net convolutional neural network (CNN) was trained to accurately segment collagen-positive pixels throughout SHG *z*-stacks. CNN performance was benchmarked against other common thresholding techniques, and was found to outperform intensity-based segmentation algorithms within an independent dataset, particularly at deeper imaging depths. These results indicate that a trained CNN can accurately segment collagen-positive pixels within a wide range of imaging depths, which is useful for quantitative SHG imaging in thick tissues.

KEYWORDS
collagen, convolutional neural network, image segmentation, second harmonic generation

1 | INTRODUCTION

Second harmonic generation (SHG) microscopy is a nonlinear optical technique that has been used to analyze collagen organization for a broad range of biomedical research applications including skin biomechanics and aging [1, 2], ovarian cancer [3] and cardiovascular disease [4]. One key advantage of SHG microscopy is the ability to nondestructively quantify the three-dimensional (3D) organization of collagen fibers due to its intrinsic depth-sectioning capabilities. Many algorithms have previously been developed and used to quantify collagen fiber orientation or organization within 2D and 3D SHG image data, such as Fourier, Hough and curvelet transforms, as well as gradient-based techniques [5].

Abbreviations: CNN, convolutional neural network; GPU, graphics processing unit; ROC, receiver-operating characteristic; SHG, second harmonic generation; TNR, true negative rate; TPEF, two photon excited fluorescence; TPR, true positive rate.

 © 2022 Wiley-VCH GmbH.

Additionally, techniques like polarization-sensitive SHG and circular dichroism SHG, can probe other aspects of collagen organization at each pixel [6, 7]. Although SHG microscopy allows for 3D depth-resolved imaging of tissues and quantitative analysis of collagen structure at multiple scales, segmentation of collagen-positive regions within a single image or 3D *z*-stack is required for quantification of various metrics related to SHG microscopy such as average SHG intensity of fibers, collagen fiber orientation and fiber length [8, 9]. This is not a trivial process, because the collected signal becomes increasingly attenuated below the tissue surface due to photon scattering and absorption [10, 11]. Furthermore, the SHG images may also contain a signal from other tissue constituents in addition to collagen. These challenges make segmentation of collagen fibers difficult using SHG intensity-based thresholding [10].

Deep learning neural networks can provide a more accurate solution to automated collagen segmentation [12]. Neural networks are a type of artificial intelligence that can be trained to complete a certain task through an error optimization process. Convolutional neural networks (CNNs) are a type of artificial intelligence that has become increasingly popular for biomedical image analysis and segmentation [13]. CNNs work by recognizing object-related patterns within images, which are learned during network training to improve pattern recognition and segmentation accuracy. Recently, the presence of pixel-wise semantic segmentation CNN architectures, such as U-Net, has dramatically increased in the field of image analysis, allowing for more advanced 2D and 3D image segmentation [13, 14]. Additionally, open-source neural network-based toolboxes such as WekaSegmentor [15] and iLastik allow the ability to train neural networks on user-defined ground truth images, but may not be appropriate for densely-labeled training images, such as collagen-positive pixels within SHG image volumes [16]. The goal of this study was to train a U-Net CNN architecture to accurately classify collagen-positive pixels within an SHG image volume, and evaluate whether a trained CNN can accurately segment collagen fibers deep within tissue. This type of analysis provides an easily accessible and relatively fast process for automatically identifying collagen fibers within a highly scattering 3D tissue.

2 | MATERIALS AND METHODS

2.1 | Image dataset generation

Tissue samples consisting of excised ventral skin of young (n = 15; 4 months) and aged (n = 16; 23 months) C57BL/6J mice were prepared by carefully cleaning the epidermis of debris and resecting the hypodermis [2]. Multiphoton image volumes (512 × 512 pixels, 1.144 μm/pixel X-Y resolution; 13-bit intensity) containing SHG signal (855 nm excitation, ≤440 nm emission) were collected using a multiphoton microscope (Bruker; Middleton, Wisconsin) equipped with a Ti:Sapphire laser (Spectra-Physics; Mountain View, California) and a 20×, 1.0 NA water-dipping objective (Olympus; Tokyo, Japan). Image volumes were collected in either 1 μm or 2.5 μm *z*-steps and spanned total depths of 130 to 170 μm.

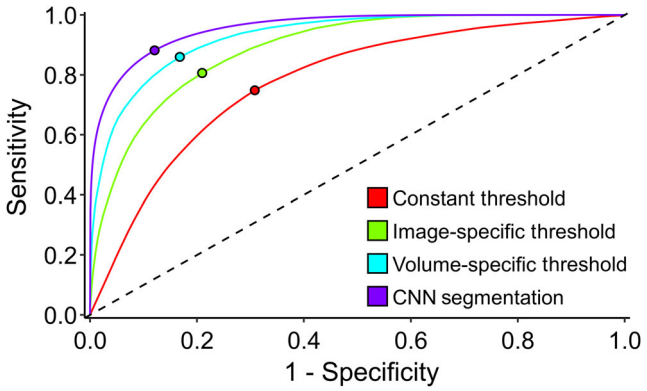
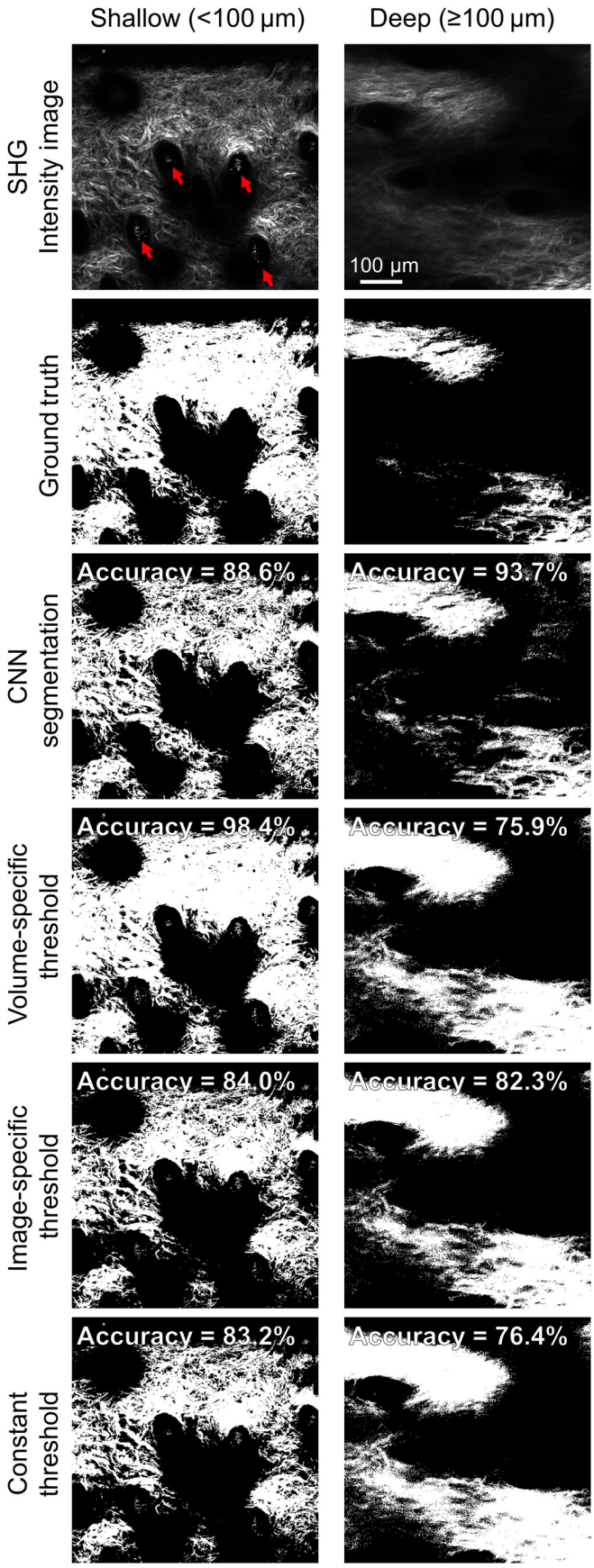


FIGURE 1 Optimal threshold values (colored circles) for each segmentation technique were computed from their corresponding ROC curves

Accurate ground truth collagen-positive masks were created by first manually adjusting intensity thresholds for each 2D image slice within a 3D *z*-stack until the thresholded collagen-positive mask accurately reflected the collagen-positive pixels within the intensity image. Emphasis was placed on highlighting any pixels containing collagen SHG signal rather than delineating the contours of individual fibers. Next, regions containing signal not associated with collagen fibers (eg, noise and signal from hair) were manually removed from the collagen-positive masks by comparing the mask to the original intensity image, resulting in the final ground truth images. To reduce memory usage but still accurately extract features relating to collagen fibers, small patches (64 × 64 pixels) of the SHG intensity images and corresponding ground truth masks were used in network training and data were sampled in increments of 5 μm within each image *z*-stack [14]. The entire dataset (~613 000 image patches) was found to have significantly more background pixels compared to collagen-positive pixels, and only patches that contained at least 10% collagen-positive pixels within the ground truth images were considered for training, validation and testing. The resulting image patch dataset (138 836 image patches) contained ~55% and ~45% collagen-positive pixels and background pixels, respectively. Images were randomly

18640648, 2022, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jbio.202200191 by University Of Arkansas Library, Wiley Online Library on [31/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



assigned to either a training (70%), validation (20%) or testing (10%) dataset, and all image patches within a single image z-stack were assigned to the same dataset.

2.2 | Network architecture and training

A traditional U-Net CNN architecture containing four encoding and decoding blocks was initialized using PyTorch [17]. Briefly, each encoding and decoding block consists of two 2D convolution layers followed by a rectified linear unit (ReLU) layer. For all convolution layers, each filter had a size of 3×3 pixels, and the number of filters for each sequential encoding block was doubled from 64 filters to 512 filters. A similar process was followed for each decoding block such that the final decoding block contained 64 filters. Between each encoding block, a 2×2 2D max pooling layer was used to downsample feature maps by half. Similarly, each decoding block was followed by an 2D up-sampling layer consisting of a interpolation layer with a scale factor of 2. For this network, the input images consisted of the raw 13-bit intensity images, which were normalized by (2^{13})-1 to ensure network input values between 0 and 1. Network training was performed on an RTX 2070 graphics processing unit (GPU), and an adaptive moment (ADAM) optimizer ($\eta = 0.001$, $\beta = [0.9, 0.999]$) was used to adjust weights and biases within the network [18]. After each epoch of training, the network accuracy was assessed with the validation set and a reduce-on-plateau scheduler reduced the learning rate at the end of every epoch based on the validation set accuracy to ensure overfitting did not occur. The network was trained for 10 epochs, which was based on when the network accuracy did not change between epochs. Additionally, for each epoch, all input images had a 50% chance of getting horizontally and/or vertically flipped. The output of the CNN is a pixel-wise collagen-positive probability map ranging between 0 and 1, and the loss was calculated using a pixel-wise binary cross-entropy loss algorithm [19].

FIGURE 2 Representative collagen-positive maps for shallow (<100 μm) and deep (>100 μm) z-depths (left and right columns, respectively). From top to bottom: SHG images, corresponding ground truth segmentation, collagen-positive CNN segmentation, segmentation with a volume-specific intensity threshold, image-specific intensity threshold and constant intensity threshold. Red arrows within representative shallow intensity image indicate signal from hair

18640648, 2022, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jbio.202200191 by University Of Arkansas Library, Wiley Online Library on [31/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.3 | Network performance

To quantify network performance, the network output for the testing dataset was benchmarked against three different automated thresholding techniques to generate collagen-positive masks: a "constant intensity" threshold, where all intensity values greater than a single intensity value were considered collagen-positive, as well as "image-specific" and "volume-specific" thresholds where intensities were considered collagen-positive if they were greater than the mean SHG intensity either a 2D image or 3D image volume multiplied by a scaling factor. To identify values that resulted in the best-case performance for each thresholding technique, a receiver operating characteristic (ROC) curve for the entire image dataset was produced by adjusting either the constant intensity value, the scaling factor for the mean SHG intensities, or the collagen-positive threshold value for the CNN probability map (Figure 1). Optimal values were determined by finding the largest Youden's index for each technique [20]. Using optimal threshold values for all techniques, the number of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) pixels within the dataset were counted, where correctly segmented collagen-positive pixels were considered true positive. Additionally, the true positive rate (TPR),

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{1}$$

true negative rate (TNR),

$$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}, \tag{2}$$

and accuracy,

$$\mathrm{accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{3}$$

were calculated for the testing dataset. To account for the slight imbalance between classes [14], a dice coefficient, or F1 score, was calculated as,

$$\mathrm{F1\ score} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{4}$$

To assess the accuracy of each technique with respect to imaging depth, accuracy was calculated for each 512 × 512 pixel intensity image (Figure 2), and mean accuracies were calculated at each separate imaging depth in 5 μm z-steps. The performance of the network was compared individually for all depths, as well as depths <100 μm and ≥100 μm. For depth-related comparisons, a two-way repeated measures ANOVA was performed with depth and segmentation technique as fixed effects and each tissue sample was treated as a random effect. Additionally, a similar statistical model with age and depth as fixed effects was used to determine the influence of age on segmentation. A post-hoc Dunnett test was used to make direct comparisons between the CNN accuracy and all segmentation techniques. Statistical analysis along with all data visualization was performed using R (R Core Team; Vienna, Austria).

3 | RESULTS AND DISCUSSION

Network training took 92 minutes to complete, and the trained network can segment a 512 × 512 pixel image in <1 second when utilizing a GPU. To benchmark network performance, optimal threshold values were first determined for each automated segmentation technique based on comparison to the ground truth segmentation. Benchmarking was performed using the testing dataset, and was quantified using TPR, TNR, accuracy and F1 score (Table 1). Overall, the constant intensity threshold technique was found to perform the worst for all metrics (63.1%, 38.1%, 50.9% and 0.568, respectively), which is potentially due to biological variability among samples (eg, sex and age) or day-to-day variations in laser power during image collection. These variations can be alleviated by using a volume-specific mean intensity to segment images which results in improved performance metrics (87.9%, 83.8%, 85.9% and 0.865, respectively). However, the trained CNN outperformed all thresholding techniques with respect to TNR (90.5%), accuracy (88.5%) and F1 score (0.885). For binary classification applications, the final probability map is segmented using some threshold value, typically 0.5, to determine the class of each pixel [17]. Of note, we found that the optimal value for determining collagen-positive pixels using the trained CNN was 0.56, indicating that there must be careful consideration in the threshold value used for final binary segmentation (Figure 1).

Collagen fiber SHG image intensities become attenuated as a function of imaging depth after ~100 μm, primarily due to photon scattering. To evaluate the performance of the trained CNN with respect to imaging depth from the tissue surface, a mean accuracy was calculated from all full-field images (512 × 512 pixels) corresponding to a specific depth (Figure 3). Spanning all imaging depths, the trained CNN achieved an average accuracy of 93.7 ± 5.87%, which is significantly better than the image-specific threshold technique (82.0 ± 10.6%; $P < .001$), and similar to the constant

18640648, 2022, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jbio.202200191 by University Of Arkansas Library, Wiley Online Library on [31/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 The trained CNN outperformed all other segmentation techniques in the testing dataset

Segmentation method	TPR	TNR	Accuracy	F1 score
Constant threshold	63.08%	38.08%	50.89%	0.5683
Image-specific threshold	84.06%	75.51%	79.89%	0.8107
Volume-specific threshold	87.92%	83.76%	85.89%	0.8646
CNN segmentation	86.54%	90.51%	88.48%	0.8850

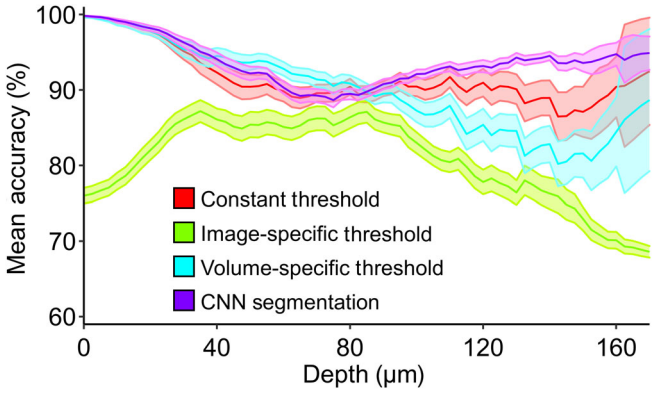


FIGURE 3 CNN segmentation algorithm performs better than typical thresholding techniques with increasing imaging depth due to image intensity attenuation from tissue scattering. For all depths, the trained CNN was found to have a mean accuracy of ≥90%. Corridors represent standard error

intensity (92.2 ± 9.61%) and volume-specific (91.4 ± 9.93%) techniques ($P = 1.000$). It should also be noted that the image-specific threshold performs substantially worse that other segmentation techniques particularly at depths less than 40 μm. This discrepancy is likely due to the use of an average intensity threshold on a per slice basis within an image volume, which lacks any context on if fibers should exist at a particular depth. At imaging depths <100 μm, the mean accuracy of the CNN (93.9 ± 6.35%) was significantly higher than the image-specific threshold accuracy (83.3 ± 10.7%; $P < .001$), and nearly identical to the constant intensity accuracy (93.1 ± 9.11%) and volume-specific threshold accuracy (93.9 ± 7.78%; $P = 1.000$). Interestingly, the accuracy of the CNN at depths ≥100 μm (93.1 ± 4.17%) was significantly higher than the image-specific (78.5 ± 9.44%; $P < .001$) and volume-specific technique (84.3 ± 11.7%; $P = .003$), and slightly improved compared to the constant intensity thresholding technique (89.6 ± 10.4%; $P = .484$).

There are well known changes in collagen fiber microstructure with increased age [21], but there were no significant differences in the accuracy of the trained CNN between young and aged skin ($P = .35$), indicating that the trained network does not demonstrate any age-related bias in detecting collagen-positive pixels. However, it is important to consider the resolution of the input images used to train the network. CNNs are trained to detect low-level and high-level features within the input images, in this case within a collagen fiber network. The fiber features learned by the network are likely specific to the magnification of the input images, so image resizing images of different magnifications may be necessary to ensure accurate results. Alternatively, transfer learning with the trained CNN can be employed to retrain the network for images at significantly different magnifications or to adapt this CNN for use in significantly different fiber networks (eg, elastin or Type II Collagen). Nonetheless, these results suggest that the trained CNN is capable of segmenting collagen-positive pixels at a wide range of imaging depths with improved accuracy and precision over typical segmentation techniques.

4 | CONCLUSION

Automated image segmentation and the use of artificial intelligence, particularly CNNs, are becoming increasingly popular tool in the field of biomedical image analysis. Manual segmentation or thresholding of biomedical images is time consuming and subjective, requiring days to weeks to accurately segment large image sets. In this study, we showed that a CNN can be trained with SHG images without the aid of additional information (eg, two photon excited fluorescence) to accurately distinguish collagen fibers from other image features (eg, hair) within the 3D image z-stacks (Figure 2). This method for segmenting collagen-positive pixels within an image volume can be useful for accurately quantifying and comparing collagen fiber organization. Although this CNN was only trained on images of mouse skin, the ability to segment collagen fibers with relatively high accuracy via raw intensity images from a single detection channel indicates that this trained CNN can be easily utilized for other areas of research that utilize SHG microscopy.

AUTHOR CONTRIBUTIONS

Alan E. Woessner and Kyle P. Quinn contributed to study design and conceptualization. Alan E. Woessner performed all data collection and analysis. Both authors were involved in manuscript revision and final approval.

18640648, 2022, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jbio.202200191 by University Of Arkansas Library, Wiley Online Library on [31/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ACKNOWLEDGMENTS

This research was funded by NIH grant numbers R00EB017723, R01AG056560, R01EB031032, the Arkansas Integrative Metabolic Research Center (P20GM139768), as well as NSF grant number 1846853 and the Arkansas Biosciences Institute.

CONFLICTS OF INTEREST

The authors declare no financial or commercial conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study, as well as the CNN trained in this study, are available from the corresponding author upon reasonable request.

ORCID

Alan E. Woessner https://orcid.org/0000-0001-8400-7813

Kyle P. Quinn https://orcid.org/0000-0002-6876-3608

REFERENCES

[1] S. Bancelin, B. Lynch, C. Bonod-Bidaud, G. Ducourthial, S. Psilodimitrakopoulos, P. Dokladal, J. M. Allain, M. C. Schanne-Klein, F. Ruggiero, *Sci. Rep.* **2015**, *5*, 17635.

[2] A. E. Woessner, J. D. Jones, N. J. Witt, E. A. Sander, K. P. Quinn, *Front. Bioeng. Biotechnol.* **2021**, *9*, 642866.

[3] S. Alkmin, R. Brodziski, H. Simon, D. Hinton, R. H. Goldsmith, M. Patankar, P. J. Campagnola, *Cancers* **2020**, *12*, 1390.

[4] I. Tandon, K. P. Quinn, K. Balachandran, *Front. Cardiovasc. Med.* **2021**, *8*, 688513.

[5] J. S. Bredfeldt, Y. Liu, C. A. Pehlke, M. W. Conklin, J. M. Szulczewski, D. R. Inman, P. J. Keely, R. D. Nowak, T. R. Mackie, K. W. Eliceiri, *J. Biomed. Opt.* **2014**, *19*, 016007.

[6] I. Gusachenko, V. Tran, Y. Goulam Houssen, J. M. Allain, M. C. Schanne-Klein, *Biophys. J.* **2012**, *102*, 2220.

[7] X. Chen, C. Raggio, P. J. Campagnola, *Opt. Lett.* **2012**, *37*, 3837.

[8] X. Chen, O. Nadiarynkh, S. Plotnikov, P. J. Campagnola, *Nat. Protoc.* **2012**, *7*, 654.

[9] A. M. Stein, D. A. Vader, L. M. Jawerth, D. A. Weitz, L. M. Sander, *J. Microsc.* **2008**, *232*, 463.

[10] P. J. Campagnola, C. Y. Dong, *Laser Photonics Rev.* **2011**, *5*, 13.

[11] T. Yasui, M. Yonetsu, R. Tanaka, Y. Tanaka, S. Fukushima, T. Yamashita, Y. Ogura, T. Hirao, H. Murota, T. Araki, *J. Biomed. Opt.* **2013**, *18*, 31108.

[12] K. de Haan, Y. Rivenson, Y. Wu, A. Ozcan, *Proc. IEEE* **2020**, *108*, 30.

[13] J. C. Gore, *Magn. Reson. Imaging* **2020**, *68*, A1.

[14] J. D. Jones, M. R. Rodriguez, K. P. Quinn, *Lasers Surg. Med.* **2021**, *53*, 1086.

[15] I. Arganda-Carreras, V. Kaynig, C. Rueden, K. W. Eliceiri, J. Schindelin, A. Cardona, H. Sebastian Seung, *Bioinformatics* **2017**, *33*, 2424.

[16] S. Berg, D. Kutra, T. Kroeger, C. N. Straehle, B. X. Kausler, C. Haubold, M. Schiegg, J. Ales, T. Beier, M. Rudy, K. Eren, J. I. Cervantes, B. Xu, F. Beuttenmueller, A. Wolny, C. Zhang, U. Koethe, F. A. Hamprecht, A. Kreshuk, *Nat. Methods* **2019**, *16*, 1226.

[17] O. Ronneberger, P. Fischer, T. Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*, **2015**. https://doi.org/10.48550/arXiv.1505.04597

[18] D. P. Kingma, J. Ba, *Adam: A Method for Stochastic Optimization*, **2014**, arXiv:1412.6980.

[19] Z. Zhang, M. R. Sabuncu, *Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels* **2018**, https://doi.org/10.48550/ARXIV.1805.07836.

[20] W. J. Youden, *Biometrics* **1950**, *6*, 172.

[21] M. J. Blair, J. D. Jones, A. E. Woessner, K. P. Quinn, *Adv. Wound Care (New Rochelle)* **2020**, *9*, 127.

How to cite this article: A. E. Woessner, K. P. Quinn, *J. Biophotonics* **2022**, *15*(12), e202200191. https://doi.org/10.1002/jbio.202200191

18640648, 2022, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jbio.202200191 by University Of Arkansas Library, Wiley Online Library on [31/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License