**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
Main Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2021

Deep Learning for Automatic Segmentation of Hybrid Optoacoustic Ultrasound (OPUS) Images

Lafci, Berkan ; Mercep, Elena ; Morscher, Stefan ; Dean-Ben, Xose Luis ; Razansky, Daniel

Abstract: The highly complementary information provided by multispectral optoacoustics and pulse-echo ultrasound have recently prompted development of hybrid imaging instruments bringing together the unique contrast advantages of both modalities. In the hybrid optoacoustic ultrasound (OPUS) combination, images retrieved by one modality may further be used to improve the reconstruction accuracy of the other. In this regard, image segmentation plays a major role as it can aid improving the image quality and quantification abilities by facilitating modeling of light and sound propagation through the imaged tissues and surrounding coupling medium. Here, we propose an automated approach for surface segmentation in whole-body mouse OPUS imaging using a deep convolutional neural network (CNN). The method has shown robust performance, attaining accurate segmentation of the animal boundary in both optoacoustic and pulse-echo ultrasound images, as evinced by quantitative performance evaluation using Dice coefficient metrics.

DOI: <https://doi.org/10.1109/tuffc.2020.3022324>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-211673>

Journal Article

Accepted Version

Originally published at:

Lafci, Berkan; Mercep, Elena; Morscher, Stefan; Dean-Ben, Xose Luis; Razansky, Daniel (2021). Deep Learning for Automatic Segmentation of Hybrid Optoacoustic Ultrasound (OPUS) Images. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 68(3):688-696.

DOI: <https://doi.org/10.1109/tuffc.2020.3022324>

Deep Learning for Automatic Segmentation of Hybrid Optoacoustic Ultrasound (OPUS) Images

Berkan Lafci¹, Elena Merčep², Stefan Morscher², Xosé Luís Deán-Ben¹, and Daniel

Razansky^{1*}

¹ Institute for Biomedical Engineering and Institute of Pharmacology and Toxicology,
University of Zurich and ETH Zurich, Switzerland

² iThera Medical GmbH, Munich, Germany

* E-Mail: daniel.razansky@uzh.ch

ABSTRACT

The highly complementary information provided by multispectral optoacoustics and pulse-echo ultrasound have recently prompted development of hybrid imaging instruments bringing together the unique contrast advantages of both modalities. In the hybrid optoacoustic ultrasound (OPUS) combination, images retrieved by one modality may further be used to improve the reconstruction accuracy of the other. In this regard, image segmentation plays a major role as it can aid improving the image quality and quantification abilities by facilitating modeling of light and sound propagation through the imaged tissues and surrounding coupling medium. Here we propose an automated approach for surface segmentation in whole-body mouse OPUS imaging using a deep convolutional neural network (CNN). The method has shown robust performance, attaining accurate segmentation of the animal boundary in both optoacoustic and pulse-echo ultrasound images, as evinced by quantitative performance evaluation using Dice coefficient metrics.

I. INTRODUCTION

Optoacoustics (OA) is gaining maturity as a high performance biomedical imaging technique. Both clinical and small animal multispectral optoacoustic tomography (MSOT) scanners are being employed by a steadily growing number of groups in various fields including cancer research [1], [2], [3] cardio-vascular and inflammatory diseases [4], [5], neuroscience [6], [7] and drug development [8], [9]. The unique advantages of MSOT for biological investigations stem from its inherent capacity for spectroscopic differentiation of light-absorbing substances (chromophores) with high resolution deep in scattering mammalian tissues [10], [11]. The functional and molecular imaging capabilities of MSOT can greatly benefit from hybridization with pulse-echo ultrasound (US), which can provide additional information on tissue anatomy and blood flow [12], [13]. Efficient hybrid OPUS imaging was recently implemented with concave arrays of cylindrically-focused transducers, which allowed mitigating limited-view and side-lobe effects in both MSOT and pulse-echo US reconstructions [14]. By implementing a full-ring array tomographic imaging geometry, it was further possible to render speed of sound and attenuation maps from transmission US measurements [15].

Image formation in both US and MSOT is commonly based on coherent summation of the detected acoustic waveforms. Sequential transmission from each transducer element and subsequent image formation based on the synthetic transmit aperture (STA) method has been shown suitable for pulse-echo US beamforming with the concave array systems used in MSOT [14]. Speed of sound variations in the imaged object may lead to misplacement of the compounded images and subsequent loss of spatial resolution. Due to its shared acoustic propagation path, MSOT reconstructions are challenged by a similar problem since assumption of uniform speed of sound in the imaged tissue and the surrounding medium has been shown to result in significant distortions, blurring and loss of contrast in the images [16]. Thereby, proper segmentation of the tissue surface and assignment of correct acoustic properties are essential for rendering high quality MSOT reconstructions [17], [18]. Knowledge of the object's boundaries can also aid proper modelling of the excitation light distribution [19], [20], thus further contribute to easier interpretation and quantification of the reconstructed images.

Image segmentation is often challenged by low contrast, noise, and other experimental factors. In pulse-echo US images, common artifacts are related to attenuation, speckle noise or shadowing, which may result in missing boundaries [21]. Efficient segmentation of MSOT images is similarly hampered by the relatively low intrinsic contrast of large anatomical

structures and tissue boundaries [22]. One potential solution consists in the application of active contour models for the segmentation of MSOT images, which was further used for enhancing imaging performance by accounting for differences in speed of sound and light fluence attenuation [20]. Methods based on deep learning (DL) have been increasingly utilized for medical image processing over the last two decades owing to their good performance in classification and function approximation, particularly, in the case of pre-processing (e.g. construction and restoration), segmentation, registration and recognition [23]. The advent of DL methods consisting of more hidden layers and providing higher level of abstraction has further enhanced the performance of convolutional neural networks (CNNs) for medical image analysis.

Herein, we propose a CNN based on the U-Net architecture [24] for segmenting images acquired with a hybrid optoacoustic ultrasound (OPUS) imaging scanner. A large number of manually-segmented OA images was employed for efficient training of the network, whose performance was then evaluated on images rendered with both modalities by calculating Dice coefficient metrics [25] of the difference between the automatically segmented and ground truth (manually segmented) images. The network performance was further compared with those of active contour segmentation and of a CNN trained with both OA and US images.

II. MATERIALS AND METHODS

The key system components and the general workflow are schematically depicted in Fig. 1. In short, pulse-echo US and OA images were acquired from living mice non-invasively with a small-animal OPUS scanner. The outer boundaries were manually segmented (labelled) from the reconstructed images and served as ground truth for training a CNN based on the U-Net architecture. The performance of the CNN was subsequently compared with that of the active contour segmentation method by quantifying differences in the segmentation results with the ground truth approach using Dice coefficient metric. A more detailed description of the methods is provided below.

A. Convolutional Neural Network

CNNs use image kernels to extract image features and learn the optimal kernels by updating them during training for a specific task. Kernel combinations using different sizes and interconnections lay the foundation of the CNN architectures [26]. The CNN architecture used

in this work was first proposed in [24] and referred to as the U-Net architecture. U-Net architecture features a combination of encoders and decoders, where every encoder layer contains convolutional kernels followed by steps of batch normalization and rectified linear unit (ReLU) [27]. In our case, the network used 256x256 pixel image with one channel as input, increased the channel size of the image and downsampled the image in the encoder part. In the decoder part, the network up-sampled and concatenated the image while the channel size was reduced. The network was further adapted to the input image sizes and layers using the PyTorch framework, as shown in Fig. 2. After the application of two convolutional layers having 3x3 kernels with stride 1, a max pooling layer having 2x2 kernels with stride 2 was applied. The decoder layers first applied transposed convolution with kernel size 3x3, stride 1 and up-sampled the input with a factor of 2 using bilinear interpolation. Then, the same sized encoder layer was concatenated with the decoder layer. After concatenation, the convolutional layer with 3x3 kernels on inputs that have the stride value of 1 was applied followed by the batch normalization and ReLU.

Augmentation methods were further applied in order to increase the amount of training data and ensure variability in the dataset. Specifically, this was done by random scaling, rotation and shifting of the images. The network used a loss function combined in different ratios (0.25-0.75, 0.5-0.5, 0.75-0.25) of binary cross entropy and soft Dice loss. The loss function was optimized by stochastic gradient descent (SGD) with learning rate 0.01 and momentum 0.99. The training was done with batch size 5 over 100 epochs. The initialization of the weights was done by the Xavier uniform initializer [28].

B. Ground Truth Manual Segmentation

Manual segmentation still remains a common practice in image analysis applications to obtain ground truth data for evaluating the performance of image segmentation approaches [29], [30] or providing data for CNN training [31]. In this work, manual segmentation of all OA and US images was done under supervision of an experienced biologist well trained in the imaging field. It was performed by drawing polygons around the structures followed by spline interpolation of the polygon for smooth contour appearance. The final ground truth data contained binary 300x300 pixels images, where “1” and “0” corresponds to pixels located inside and outside the mouse, respectively.

C. Active Contour by Edge Detection

The CNN-based segmentation performance was evaluated and compared with the active contour by edge detection (ACED) method, an automatic segmentation approach widely employed in medical image analysis [32], [33], [34]. Specifically, we used the implementation of active contour model introduced by Li et al. [35] combined with an automatic initialization method by edge detection and circle fitting algorithm. Automatic initialization facilitated the application of active contour segmentation as an automatic method not affected by user-dependent inputs. This ensured a fair comparison with the CNN, which is also an automatic segmentation method. The algorithm steps as well as intermediate outputs are shown in Fig. 3.

The ACED algorithm received as input OA or US images as a square array of 256x256 pixels (Fig. 3a). The images were first downsampled to 150x150 pixels to reduce the computation time whereas the pixel intensities were converted to 8-bit range between 0 and 255. Edge detection was implemented to overcome any dependency of the initial guess upon the user. For this, the Canny edge detector [36] was applied after smoothing the image using Gaussian filter with kernel size 3 and sigma 0.5. A binary image was then created with the pixels on the edges labelled as “1” and the background labelled as “0” (Fig. 3b). The outliers and the non-connected components in the pixels erroneously detected as edges were removed by applying morphological operations of dilation and erosion with a disc-shaped structuring element of 3-pixel size (Fig. 3c). A circle that has the minimum radius of all possible circles enclosing all the edge pixels was then selected (Fig. 3d). This circle was subsequently fed into the active contour function along with the number of iterations (20) and step size (1 pixel). The optimal number of iterations and step size that maximize Dice coefficient for the entire dataset was selected heuristically. The deformable spline contour evolves based on the minimization of an energy function with a distance regularization term and an external energy that drives the motion of the contour toward desired locations (Fig. 3e) [35].

D. Experimental Data Sets

All images were acquired with the commercial small animal imaging systems inVision 256-TF and inVision 512-echo (iThera Medical GmbH, Munich, Germany). All procedures involving animal care and experimentation were conducted in full compliance with the animal handling guidelines and with approval from the institutional review board. The detected OA signals for 800 nm excitation wavelength were band-pass filtered with 50 kHz and 6.5 MHz cut-off frequencies and deconvolved with the electrical impulse response of the transducer to account for its

limited bandwidth and phase distortions [10]. A two-dimensional filtered back-projection algorithm [16] was subsequently used to reconstruct cross-sectional images from the OA pressure signals with a pixel size and field of view (FOV) set to 100 μm and 30x30 mm^2 , respectively.

The US signals were acquired using STA by repeating transmission events for every single element of the ring-shaped detection array operating at 5 MHz central frequency. Signals from every transmission event were then beamformed using delay-and-sum method to generate low-resolution images. The final high-resolution US images were then rendered by compounding low-resolution images that were generated by the individual transmission events. Overall, STA with sub-aperture size of 128 elements in combination with spatial compounding were applied, as described elsewhere [37]. The final US images were generated in a 30x30 mm^2 FOV with a pixel size of 100 μm . Details on the training and test datasets for OA and US images are provided below.

1. *Training Datasets:* The dataset was carefully selected by assigning all acquired images to either training or test sets based on the mouse ID. This represents a standard procedure that avoids accidental use of images from the same mouse for both training and testing, thus prevents network from overfitting the data. Training was first performed using exclusively OA images. Subsequently, a combination of OA and US images was used for training. Note that the number of available US images was lower. In total, the OA training data contained 174 images (12 mice) for the brain region, 97 images (13 mice) for the kidney region and 108 images (14 mice) for the liver region. The US training data contained 98 images (13 mice) and 78 images (13 mice) from liver and kidney regions, respectively. The brain, kidney and liver data sets were subsequently combined to generate training dataset, i.e., training was eventually performed using 379 OA and 176 US images in total for brain, kidney and liver regions. A fragment of the training data was further separated for validation purposes during the training process. The ratio between training and validation sets was chosen as 0.80.

2. *Test Datasets:* The datasets used for evaluation of segmentation methods included both OA and US images, as summarized in Table 1.

Table 1. Summary of the test data sets.

Anatomical region	Number of test images	
	OA mode	US mode
Brain	28 (4 mice)	9 (2 mice)
Kidney	38 (5 mice)	16 (5 mice)

Liver	33 (5 mice)	30 (5 mice)
Total	99	55

Test datasets further included 5 brain, 4 kidney and 4 liver images with artifacts and 19 liver images acquired with mouse positioned off-center in the imaging chamber that are included in Table 1. Those data sets were included in order to test the performance of different segmentation algorithms under sub-optimal data acquisition conditions. Presence of prominent reflection artifacts in the OA images of the brain region can be ascribed to air-filled cavities or bones [38], while the off-center positioning of the animal in the imaging chamber may occur due to operation by inexperienced users.

E. Evaluation Criteria

Final evaluation was done using the Dice coefficient, also called the overlap index [25]. The Dice coefficient can be calculated by pixel-wise comparison based on the true positive (TP), false positive (FP) and false negative (FN) ratios as follows

$$DICE = \frac{2TP}{2TP + FP + FN}.$$

III. RESULTS

The segmentation performance of the CNN trained exclusively with OA data and the ACED algorithm for exemplary test datasets of the brain (Fig. 4a-c), liver (Fig. 4d-f), and kidney (Fig. 4g-i) cross-sections is shown. It can readily be observed that the CNN method outperformed the ACED approach in all representative anatomical regions. It was further possible to accurately segment the mouse surface in images containing indistinct boundaries (Fig. 4c, i) or non-circular shapes (Fig. 4f). This observation is quantitatively confirmed by the higher Dice coefficients obtained using CNN. The poorer performance of the ACED algorithm may be attributed to the step of circle initialization, when high-intensity regions may erroneously be classified as edges (Fig. 4b), or to the step of the active contour evolution, subject to over-fitting caused by low-contrast regions in the image (Figs. 4e,h). Manual initialization and parameter optimization on a per-image basis may yet recover some of the ACED performance to the detriment of the loss of an automated approach. A robust analysis was ensured by empirically optimizing the parameters to provide an optimal performance for the entire dataset rather than for individual images.

The superior performance of the CNN trained exclusively with OA data over ACED was corroborated with sub-sets of images whose quality was compromised by sub-optimal data acquisition conditions, which resulted in image artifacts and/or diminished resolution with the animal positioned off-center in the imaging chamber (Fig. 5). Specifically, images of the brain and liver regions (first and second row in Fig. 5) with arc-shaped artifacts as well as an image of the kidney region (third row of Fig. 5) with sub-optimal positioning of animal (off-center) are shown. The performance of the CNN was barely affected in those cases (Figs. 5c, f, and i), whereas the performance of the ACED algorithm drastically deteriorated. For two out of the three analyzed anatomical regions, circle initialization failed as the “arc”-shape reflection artefacts at the periphery of the FOV were classified as edges. This has led to selecting a larger initial circle and erroneous inclusion of pixels lying outside the mouse body into the segmented region (Fig. 5b, e). The CNN-based algorithm performed equally well also when the mouse was positioned off-center (Fig. 5i), with yielded Dice score of 0.98 versus 0.81 for the ACED method (Fig. 5h).

The applicability of the CNN trained exclusively on OA images for US image segmentation is exemplified in Fig. 6. Good segmentation performance was observed in the liver and kidney cross-sections (Fig. 6f, i), attaining Dice coefficients of 0.95 and 0.95, respectively. However, the CNN was less efficient when segmenting brain cross-sections (Fig. 6c). This may be attributed to the presence of high-contrast features from the acoustically reflecting but optically transparent protective waterproof membrane (Fig. 6a), not present in the OA training images. In all cases, the ACED algorithm performed poorly owing to the failed initialization step in the presence of reflections outside the mouse body, erroneously classified as edges.

The overall performance of both algorithms for all the test datasets is summarized in Fig. 7. The CNN algorithm that was trained with a loss function ratio with 0.5-0.5 of binary cross entropy and soft Dice loss outperformed ACED for segmentation of OA images with a corresponding Dice score of 0.95 ± 0.04 and a very narrow range of outliers, versus 0.78 ± 0.11 for the ACED method. Both methods showed inferior performance for segmentation of US images, in particularly with respect to the number of outliers, whereas CNN still yielded better results than ACED with Dice score of 0.80 ± 0.08 versus 0.55 ± 0.10 .

Different types of noise (Gaussian, salt & pepper and speckle) was further added to OA images to demonstrate robustness of the proposed algorithm. The network achieved Dice scores of 0.94 ± 0.03 in the presence of Gaussian noise (0 mean, 0.1 variance) for OA images and 0.78 ± 0.10 for US images. The salt and pepper noise (0.004 noise density) decreased the

mean Dice scores to 0.94 ± 0.02 for OA images and 0.76 ± 0.09 for US images compared to noise free images. The network showed an inferior performance for OA images in the presence of speckle noise, i.e. when randomly generated numbers between 0 and 255 were added. Specifically, it achieved Dice scores of 0.93 ± 0.03 for OA images and 0.81 ± 0.06 for US images.

We further analyzed the effects of loss functions on training. For this, the network was trained with different combinations of binary cross entropy and soft Dice loss. It achieved Dice score of 0.96 ± 0.02 for OA images and 0.83 ± 0.08 for US images for a 0.25-0.75 ratio of binary cross entropy and soft Dice loss ratio. On the other hand, it achieved Dice score of 0.95 ± 0.02 for OA images and 0.80 ± 0.11 for US images when considering a ratio 0.75-0.25 of binary cross entropy and soft Dice loss ratio. As expected, it appears that a higher weight in the soft Dice loss results in an increased Dice score on the test images.

The CNN was further trained by additionally including the available 176 US images from the kidney and liver cross sections. In this case, no improvement in the overall performance for OA image segmentation was observed. In fact, the measured Dice coefficient was 0.95 ± 0.4 , approximately the same as previously measured when training exclusively with the OA images (Fig. 7). In addition, the number of outliers increased in OA test images. On the other hand, the performance of US image segmentation degraded when including US images into training. The network trained with both OA and US images reached a Dice coefficient of 0.76 ± 0.09 , i.e. lower than the network trained only with OA images. We ascribe the inferior performance of the CNN trained with a combination of OA and US data to the relatively low number of available US images. We believe that the fact that the network trained with OA data can additionally accurately segment US images represents an important advantage of the suggested processing approach for hybrid OPUS imaging.

IV. DISCUSSION AND CONCLUSIONS

This work reports on a new methodology for automatic segmentation of the dual-modality images acquired with hybrid optoacoustic ultrasound (OPUS) imaging systems by means of the U-Net CNN architecture. It has shown robust performance in accurate segmentation of mouse boundary in both OA and pulse-echo US images, as evinced by quantitative evaluation of the CNN performance using Dice coefficient metrics based on the difference between the automatically segmented and ground truth images. In contrast, performance of the active contour based segmentation approach is severely compromised by

image artifacts, contrast variations as well as off-center positioning of the mouse. However, the ultimate goal of medical image segmentation is elimination of user dependency without sacrificing segmentation accuracy. In this respect, the proposed CNN approach has shown great promise in accurate segmentation in cross-sections corresponding to the mouse brain, liver and kidney regions. Better performance of the network may potentially be achieved by separating brain, kidney and liver datasets for every training process [39]. However, the data from three different anatomical regions was mixed to demonstrate the robustness of the network for different contrast and shapes. As accurate labeling of large amount of data is generally a time-consuming process, in this work we concentrated on showcasing performance in three selected anatomical regions. However, the CNN approach could be seamlessly applied to whole-body mouse segmentation as well as segmentation of distinct organs.

We further demonstrated applicability of the CNN trained on OA images for segmentation of pulse-echo US images, although this approach may have limitations due to different contrast exhibited by the two modalities. However, our results indicate that additional training with the US data did not improve the segmentation accuracy, although larger amount of datasets may need to be taken into account to draw unequivocal conclusions. In future studies, we will explore the possibility of feeding matching pairs of OA and US in the same network to increase the training performance. For this, a hybrid imaging system enabling simultaneous image acquisition in both modalities is required. The suggested approach can potentially be applied to hybrid images acquired by other types of OPUS imaging systems. For instance, multi-segment arrays consisting of concave and linear parts have been recently shown to provide an outstanding performance for multi-modal imaging [12], [40]. This particular embodiment employs different parts of the array for OA and pulse-echo US imaging, which may imply separate CNN training. The OPUS modality has also been extended to mapping the speed of sound and acoustic attenuation in tissues by using a transmission US approach [15], which may equally be used to facilitate more efficient segmentation. Three-dimensional systems providing both OA and US images have also been suggested for both small animal tomography and hand-held scanning in humans [41], [42]. Much like in the two-dimensional case, accurate segmentation of the tissue boundary can be used to improve spatial resolution and facilitate image co-registration.

The advent of DL methods represents a major shift from traditional machine learning approaches [43]. In addition to showing unprecedented performance in computer vision tasks, DL networks have been applied for processing of images from different modalities in the field of medical imaging, including CT [44], MRI [45], PET [46] and US [47], [48], [49], [50]. The

proposed CNN approach for automatic OPUS image segmentation can be used to more accurately compensate for image artifacts originating from light attenuation in deep tissues as well as speed of sound heterogeneities. Apart from image segmentation, DL methods have been shown to significantly enhance other signal and image processing steps, including tomographic reconstructions from sparse OA data [51], [52], [53], image artifact removal [54], [55] and quantitative analysis of OA images [56]. Considering the generally high complexity of the OA image reconstruction pipeline, in particular for multispectral methods like MSOT, the DL approaches are expected to play a major role in facilitating accurate image reconstruction and analysis.

ACKNOWLEDGEMENT

The authors acknowledge funding support from the Swiss Data Science Center (project C19-04) and the European Research Council under Consolidator Grant ERC-2015-CoG-682379.

REFERENCES

- [1] A. Taruttis, G. M. van Dam, and V. Ntziachristos. Mesoscopic and macroscopic optoacoustic imaging of cancer. *Cancer Research*, 75(8):1548–1559, 2015.
- [2] K. S. Valluru and J. K. Willmann. Clinical photoacoustic imaging of cancer. *Ultrasonography (Seoul, Korea)*, 35(4):267–280, 10 2016.
- [3] M. Mehrmohammadi, S. J. Yoon, D. Yeager, and S. Y. Emelianov. Photoacoustic imaging for cancer detection and staging. *Current molecular imaging*, 2(1):89–105, 03 2013.
- [4] X. L. Dean-Ben, S. J. Ford, and D. Razansky. High-frame rate four dimensional optoacoustic tomography enables visualization of cardiovascular dynamics and mouse heart perfusion. *Scientific Reports*, 5:10133 EP –, 07 2015.
- [5] S. Iskander-Rizk, A. F. W. van der Steen, and G. van Soest. Photoacoustic imaging for guidance of interventions in cardiovascular medicine. *Physics in Medicine & Biology*, 64(16):16TR01, Aug 2019.
- [6] X. L. Dean-Ben, G. Sela, A. Lauri, M. Kneipp, V. Ntziachristos, G. G. Westmeyer, S. Shoham, and D. Razansky. Functional optoacoustic neuro-tomography for scalable whole-brain monitoring of calcium indicators. *Light: Science & Applications*, 5:e16201 EP –, 12 2016.
- [7] J. Yao and L. V. Wang. Photoacoustic brain imaging: from microscopic to macroscopic scales. *Neurophotonics*, 1(1):011003, 05 2014.
- [8] D. Razansky, N. Deliolanis, C. Vinegoni, and V. Ntziachristos. Deep tissue optical and optoacoustic molecular imaging technologies for preclinical research and drug discovery. *Current pharmaceutical biotechnology*, 13:504–22, 01 2012.
- [9] Y. Zhang, J. Yu, A. R. Kahkoska, and Z. Gu. Photoacoustic drug delivery. *Sensors (Basel, Switzerland)*, 17(6):1400, 06 2017.
- [10] R. Ma, A. Taruttis, V. Ntziachristos, and D. Razansky. Multispectral optoacoustic tomography (msot) scanner for whole-body small animal imaging. *Opt. Express*, 17(24):21414–21426, Nov 2009.
- [11] M. P. Mienkina, C. S. Friedrich, N. C. Gerhardt, M. F. Beckmann, M. F. Schiffner, M. R. Hofmann, and G. Schmitz. Multispectral photoacoustic coded excitation imaging using unipolar orthogonal golay codes. *Opt. Express*, 18(9):9076–9087, Apr 2010.
- [12] X. L. Dean-Ben, E. Merčep, and D. Razansky. Hybrid-array-based optoacoustic and ultrasound (opus) imaging of biological tissues. *Applied Physics Letters*, 110(20):203703, 2017.
- [13] R. Nuster, N. Schmitner, G. Wurzing, S. Gratt, W. Salvenmoser, D. Meyer, and G. Paltauf. Hybrid photoacoustic and ultrasound section imaging with optical ultrasound detection. *Journal of Biophotonics*, 6(67):549–559, 2013.

- [14] E. Merčep, G. Jeng, S. Morscher, P. C. Li, and D. Razansky. Hybrid optoacoustic tomography and pulse-echo ultrasonography using concave arrays. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 62(9):1651–1661, Sept 2015.
- [15] E. Merčep, J. L. Herraiz, X. L. Dean-Ben, and D. Razansky. Transmission–reflection optoacoustic ultrasound (tropus) computed tomography of small animals. *Light: Science & Applications*, 8(1):18, 2019.
- [16] C. Lutzweiler and D. Razansky. Optoacoustic imaging and tomography: Reconstruction approaches and outstanding challenges in image performance and quantification. *Sensors*, 13(6):7345–7384, 2013.
- [17] J. Jose, R. G. H. Willemink, W. Steenbergen, C. H. Slump, T. G. van Leeuwen, and S. Manohar. Speed-of-sound compensated photoacoustic tomography for accurate imaging. *Medical Physics*, 39(12):7262–7271, 2012.
- [18] X. L. Dean-Ben, V. Ntziachristos, and D. Razansky. Effects of small variations of speed of sound in optoacoustic tomographic imaging. *Medical Physics*, 41(7):073301, 2014.
- [19] F. M. Brochu, J. Brunner, J. Joseph, M. R. Tomaszewski, S. Morscher, and S. E. Bohndiek. Towards quantitative evaluation of tissue absorption coefficients using light fluence correction in optoacoustic tomography. *IEEE Transactions on Medical Imaging*, 36(1):322–331, Jan 2017.
- [20] S. Mandal, X. L. Dean-Ben, and D. Razansky. Visual quality enhancement in optoacoustic tomography using active contour segmentation priors. *IEEE Transactions on Medical Imaging*, 35(10):2209–2217, Oct 2016.
- [21] J. A. Noble and D. Boukerroui. Ultrasound image segmentation: a survey. *IEEE Transactions on Medical Imaging*, 25(8):987–1010, Aug 2006.
- [22] C. Lutzweiler, R. Meier, and D. Razansky. Optoacoustic image segmentation based on signal domain analysis. *Photoacoustics*, 3(4):151 – 158, 2015.
- [23] Z. Shi, L. He, K. Suzuki, T. Nakamura, and H. Itoh. Survey on neural networks used for medical image processing. *International journal of computational science*, 3 1:86–100, 2009.
- [24] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of LNCS, pages 234–241. Springer, 2015.
- [25] L. R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.
- [26] Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio. Object recognition with gradient-based learning. In *Shape, Contour and Grouping in Computer Vision*, pages 319–, London, UK, UK, 1999. Springer-Verlag.
- [27] V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In *ICML*, 2010.

- [28] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. *Journal of Machine Learning Research - Proceedings Track*, 9:249–256, 01 2010.
- [29] K. Tingelhoff, K. W. G. Eichhorn, I. Wagner, M. E. Kunkel, A. I. Moral, M. E. Rilk, F. M. Wahl, and F. Bootz. Analysis of manual segmentation in paranasal ct images. *European Archives of Oto-Rhino-Laryngology*, 265(9):1061–1070, Sep 2008.
- [30] X. Jiang, C. Marti, C. Irniger, and H. Bunke. Distance measures for image segmentation evaluation. *EURASIP Journal on Advances in Signal Processing*, 2006(1):035909, Mar 2006.
- [31] F. C. Monteiro and A. C. Campilho. Performance evaluation of image segmentation. In *Image Analysis and Recognition*, pages 248–259. Springer Berlin Heidelberg, 2006.
- [32] F. Derraz, M. Beladgham, and M. Khelif. Application of active contour models in medical image segmentation. In *International Conference on Information Technology: Coding and Computing*, 2004. Proceedings. ITCC 2004., pages 675–681 Vol.2, April 2004.
- [33] R. Boscolo, M. S. Brown, and M. F. McNitt-Gray. Medical image segmentation with knowledge-guided robust active contours. *Radio- Graphics*, 22(2):437–448, 2002. PMID: 11896232.
- [34] X. Qian, J. Wang, S. Guo, and Q. Li. An active contour model for medical image segmentation with application to brain ct image. *Medical physics*, 40(2):021911–021911, 02 2013.
- [35] C. Li, C. Xu, C. Gui, and M. D. Fox. Distance regularized level set evolution and its application to image segmentation. *IEEE Transactions on Image Processing*, 19(12):3243–3254, Dec 2010.
- [36] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, Nov 1986.
- [37] E. Merčep, N. C. Burton, J. Claussen, and D. Razansky. Whole-body live mouse imaging by hybrid reflection-mode ultrasound and optoacoustic tomography. *Opt. Lett.*, 40(20):4643–4646, Oct 2015.
- [38] X. L. Dean-Ben, R. Ma, D. Razansky, and V. Ntziachristos. Statistical approach for optoacoustic image reconstruction in the presence of strong acoustic heterogeneities. *IEEE Transactions on Medical Imaging*, 30(2):401–408, Feb 2011.
- [39] K. Jnawali, B. Chinni, V. Dogra, and N. Rao. Automatic cancer tissue detection using multispectral photoacoustic imaging. *International Journal of Computer Assisted Radiology and Surgery*, 15(2):309–320, 2020.
- [40] E. Merčep, X. L. Dean-Ben, and D. Razansky. Combined pulse-echo ultrasound and multispectral optoacoustic tomography with a multi-segment detector array. *IEEE Transactions on Medical Imaging*, 36(10):2129–2137, Oct 2017.
- [41] V. Neuschmelting, N. C. Burton, H. Lockau, A. Urich, S. Harmsen, V. Ntziachristos, and M. F. Kircher. Performance of a multispectral optoacoustic tomography (msot) system

equipped with 2d vs. 3d handheld probes for potential clinical translation. *Photoacoustics*, 4(1):1 – 10, 2016.

[42] T. F. Fehm, X.L. Dean-Ben, and D. Razansky. Four dimensional hybrid ultrasound and optoacoustic imaging via passive element optical excitation in a hand-held probe. *Applied Physics Letters*, 105(17):173505, 2014.

[43] D. C. Cireşan, U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber. Flexible, high performance convolutional neural networks for image classification. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Two, IJCAI'11*, pages 1237–1242. AAAI Press, 2011.

[44] P. F. Christ, F. Ettlinger, F. Grün, M. E. A. Elshaer, J. Lipkov'a, S. Schlecht, F. Ahmaddy, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, F. Hofmann, M. D'Anastasi, S. A. Ahmadi, G. Kaissis, J. W. Holch, W. H. Sommer, R. Braren, V. Heinemann, and B. H. Menze. Automatic liver and tumor segmentation of ct and mri volumes using cascaded fully convolutional neural networks. *CoRR*, abs/1702.05970, 2017.

[45] Z. Akkus, A. Galimzianova, A. Hoogi, D. L. Rubin, and B. J. Erickson. Deep learning for brain mri segmentation: State of the art and future directions. *Journal of Digital Imaging*, 30(4):449–459, Aug 2017.

[46] P. Blanc-Durand, A. Van Der Gucht, N. Schaefer, E. Itti, and J. O. Prior. Automatic lesion detection and segmentation of 18f-fet pet in gliomas: A full 3d u-net convolutional neural network study. *PLOS ONE*, 13(4):1–11, 04 2018.

[47] M. H. Yap, G. Pons, J. Martí, S. Ganau, M. Sentís, R. Zwigelaar, A. K. Davison, and R. Martí. Automated breast ultrasound lesions detection using convolutional neural networks. *IEEE Journal of Biomedical and Health Informatics*, 22(4):1218–1226, July 2018.

[48] P. Looney, G. N. Stevenson, K. H. Nicolaides, W. Plasencia, M. Molloholli, S. Natsis, and S. L. Collins. Fully automated, real-time 3d ultrasound segmentation to estimate first trimester placental volume using deep learning. *JCI insight*, 3(11):e120178, 06 2018.

[49] M. Amiri, R. Brooks, and H. Rivaz. Fine tuning u-net for ultrasound image segmentation: Which layers? In *Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data*, pages 235–242, Cham, 2019. Springer International Publishing.

[50] M. Philip, A. Sowmya, H. Avnet, A. Ferreira, G. Stevenson, and A. Welsh. Convolutional neural networks for automated fetal cardiac assessment using 4d b-mode ultrasound. pages 824–828, 04 2019.

[51] N. Davoudi, X.L. Dean-Ben, and D. Razansky. Deep learning optoacoustic tomography with sparse data. *Nature Machine Intelligence*, 2019.

[52] S. Antholzer, M. Haltmeier, and J. Schwab. Deep learning for photoacoustic tomography from sparse data. *Inverse Problems in Science and Engineering*, 04 2017.

- [53] A. Hauptmann, F. Lucka, M. Betcke, N. Huynh, J. Adler, B. Cox, P. Beard, S. Ourselin, and S. Arridge. Model-based learning for accelerated, limited-view 3-d photoacoustic tomography. *IEEE Transactions on Medical Imaging*, 37(6):1382–1393, June 2018.
- [54] D. Allman, A. Reiter, and M. A. L. Bell. A machine learning method to identify and remove reflection artifacts in photoacoustic channel data. In *2017 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4, Sep. 2017.
- [55] S. Guan, A. A. Khan, S. Sikdar, and P. V. Chitnis. Fully dense unet for 2d sparse photoacoustic tomography artifact removal. *IEEE Journal of Biomedical and Health informatics*, 2018.
- [56] C. Yang, H. Lan, H. Zhong, and F. Gao. Quantitative photoacoustic blood oxygenation imaging using deep residual and recurrent neural network. *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 741–744, 2019.

FIGURE CAPTIONS

Figure 1. Key components and workflow of the conducted study. Images were acquired by multi-spectral optoacoustic tomography small animal scanners (Models inVision 256-TF and inVision 512-echo, iThera Medical GmbH, Munich, Germany) and manually segmented to create ground truth data. Deep learning (DL)-based segmentation method or active contour method are applied on the acquired images. The results are evaluated by comparing the output of the segmentation algorithms with manually segmented images using Dice coefficient.

Figure 2. The U-Net Convolutional Neural Network Architecture used for image segmentation. The network uses 256x256 pixel image with one channel as an input, increases the channel size of the image and downscales the image in the encoder part (left). In the decoder part, the network up-samples and concatenates the image while the channel size is reduced (right).

Figure 3. Algorithmic steps of the active contour by edge detection (ACED) segmentation approach. (a) The originally reconstructed optoacoustic image. (b) Output of Canny edge detector. (c) Image after morphological operations of dilation/erosion. (d) Fitted circle on pixels detected in the previous step. (e) The final result of the active contour segmentation.

Figure 4. Segmentation results for OA images obtained with manual segmentation (left), ACED (middle) and CNN (right) for cross-sections from the brain (a-c), liver (d-f), and kidney (g-i) regions. The corresponding Dice coefficients are also shown.

Figure 5. Segmentation results for OA images obtained with manual segmentation, ACED and CNN for mouse cross-sections through brain (a-c), liver (d-f), and kidney (g-i) regions that showcase sub-optimal (non-centered) animal positioning (kidney) and artifacts (liver, brain). The corresponding Dice coefficients are also shown.

Figure 6. Segmentation results for US images obtained with manual segmentation, ACED and CNN (trained on OA images) for mouse cross-sections through brain (a-c), liver (d-f), and kidney (g-i) regions with the corresponding Dice coefficients.

Figure 7. Boxplots illustrating the Dice coefficient variability for the segmented OA and US images obtained with ACED, CNN trained with only OA images and CNN trained with combination of OA and US images.

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6

Figure 7