Research Article | Vol. 10, No. 7 | 1 Jul 2019 | BIOMEDICAL OPTICS EXPRESS 3484
Biomedical Optics EXPRESS

Segmentation of mouse skin layers in optical coherence tomography image data using deep convolutional neural networks

TIMO KEPP,[1,2,*] CHRISTINE DROIGK,[3] MALTE CASPER,[4,5] MICHAEL EVERS,[4,5] GEREON HÜTTMANN,[4] NUNCIADA SALMA,[5] DIETER MANSTEIN,[5] MATTIAS P. HEINRICH,[1] AND HEINZ HANDELS[1]

[1] *Institute of Medical Informatics, University of Lübeck, Lübeck, Germany*
[2] *Graduate School for Computing in Medicine and Life Sciences, University of Lübeck, Lübeck, Germany*
[3] *Institute for Signal Processing, University of Lübeck, Lübeck, Germany*
[4] *Institute of Biomedical Optics, University of Lübeck, Lübeck, Germany*
[5] *Cutaneous Biology Research Center, Massachusetts General Hospital, Boston, MA, USA*
[*] *kepp@imi.uni-luebeck.de*

Abstract: Optical coherence tomography (OCT) enables the non-invasive acquisition of high-resolution three-dimensional cross-sectional images at a micrometer scale and is mainly used in the field of ophthalmology for diagnosis as well as monitoring of eye diseases. Also in other areas, such as dermatology, OCT is already well established. Due to its non-invasive nature, OCT is also employed for research studies involving animal models. Manual evaluation of OCT images of animal models is a challenging task due to the lack of imaging standards and the varying anatomy among models. In this paper, we present a deep-learning algorithm for the automatic segmentation of several layers of mouse skin in OCT image data using a deep convolutional neural network (CNN). The architecture of our CNN is based on the U-net and is modified by densely connected convolutions. We compared our adapted CNN with our previous algorithm, a combination of a random forest classification and a graph-based refinement, and a baseline U-net. The results showed that, on average, our proposed CNN outperformed our previous algorithm and the baseline U-net. In addition, a reduction of outliers could be observed through the use of densely connected convolutions.

1. Introduction

Optical Coherence Tomography (OCT) is a non-invasive imaging modality that enables the acquisition of high-resolution 3D volume images of biological tissue [1]. The principle of OCT is similar to that of ultrasound imaging, except that light is used instead of sound waves. The image acquisition is based on light with short coherence length, which is utilized to measure the distance of partially reflecting structures in biological tissues using an interferometer. The resulting 1D signal is called A-scan. Multiple laterally combined A-scans build a 2D cross-sectional image called B-scan. A 3D OCT volume then is composed of a series of adjacent B-scans. Due to its penetration depth of a few millimeters at a resolution at micrometer scale, OCT allows a highly detailed representation of smallest structures. OCT is a rapidly growing imaging technique that is used in various biological and medical disciplines. Particularly in ophthalmology, OCT has been established as an imaging modality for the diagnosis of retinal diseases since the early 1990s [2]. Later, OCT was also introduced in dermatology where it represents a non-invasive alternative to biopsy and histology [3, 4]. Besides its application in clinical routine, OCT is also used in research studies involving animals as *in vivo* models instead of humans [5–7].

Especially the evaluation of OCT images of animal models is a challenging task. Missing standardized image acquisition procedures, as well as varying anatomies of the animal models,

#363011 https://doi.org/10.1364/BOE.10.003484
Journal © 2019 Received 22 Mar 2019; revised 24 May 2019; accepted 24 May 2019; published 21 Jun 2019

Biomedical Optics EXPRESS

complicate the evaluation. Therefore, quantitative analysis of OCT images, which often includes manual segmentation, is very time-consuming and impractical. This motivated the development of semi- as well as fully automated segmentation methods, which have been described in preliminary works. Due to the widespread use of OCT in ophthalmology, research has been focused on layer segmentation of the human retina [8–16]. Specifically, graph-based models segmented retinal layers with high accuracy. Garvin et al. integrated boundary conditions into the graph-theoretical model in order to obtain smooth surfaces as well as to allow predefined distances between the individual layers [9]. However, segmentation errors occur if larger pathologies are present. To address this problem, a combination of machine learning and mathematical modeling was investigated. In general, the machine learning algorithm's prediction was used as initialization, e.g. random forests (RF), to support the graph-based segmentation [12, 17]. Due to their widespread popularity in computer vision, research interest in deep learning methods, especially convolutional neural networks (CNNs), has grown rapidly in medical image analysis. In contrast to classical machine learning approaches, deep learning methods learn hierarchical features directly from training data. The resulting classifiers represent complex functions that can even capture highly varying training data. Fang et al. [18] used a CNN in combination with graph search to segment retinal layers. Graph search was initialized with boundary classifications of the patch-based CNN. In [19] Ronneberger et al. presented the U-net for the segmentation of biomedical image data. The architecture of this fully convolutional network (FCN) consists of a multiscale encoder and decoder part linked via skip connections, which efficiently capture large context and enable accurate dense predictions, avoiding further steps such as graph search [18]. Ronneberger et al. were able to show that they can efficiently train the U-net with only a very small number of training data using advanced augmentation techniques. Due to these properties, the U-net represents the architectural basis in numerous subsequent works [13–16].

Furthermore, multiple papers proposed algorithms for the automated layer segmentation of different tissue types, such as retina [20, 21], skin [22–26] or even cartilage [27], in OCT using animal models. In their work [24], Sheet et al. segmented OCT images of mouse skin and compared the results with corresponding histologies. They employed a transfer learning approach to integrate statistical physical models into an RF classifier. In a subsequent study [25], Sheet et al. used denoising auto-encoders to learn tissue-specific speckle presentations for classification.

In our previous work [26], we proposed an algorithm for segmenting the subcutaneous fat layer in OCT image data of the mouse skin. An RF classifier was used to segment single B-scans. Similar to [12, 17], we integrated RF predictions into a graph-based approach. The algorithm can be summarized in four steps as follows: After several preprocessing steps (e.g. denoising, intensity homogenization and flattening) normalized 2D OCT B-scans were segmented using

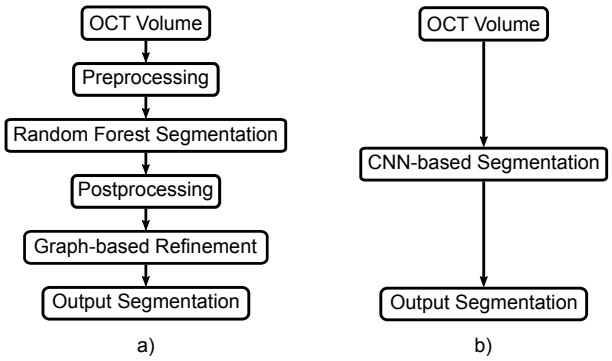

Fig. 1. Flowcharts of our previous approach [26] a), where we used a random forest classifier in combination with graph-based refinement, and our new CNN-based method b).

Biomedical Optics EXPRESS

an RF classifier. Subsequently, segmentation errors of the RF were corrected by heuristic postprocessing steps. Finally, an additional graph-based refinement step took neighborhood information into account and extended the approach to the three-dimensional space which allows the segmentation of a whole OCT volume at once. A flowchart of this approach is shown in Fig. 1. Further details of the algorithm can be found in [26]. The algorithm was used in the work of Salma et al. [7], which investigated the thickness change of the subcutaneous fat layer after localized cold exposure (selective cryolipolysis) [28] using a mouse model. Despite the fact that our algorithm already provides accurate and robust segmentation results for the subcutaneous fat layer, it is difficult to adapt it to new research issues, such as the segmentation of further structures or the customization for a different animal model. In this case, it would be necessary to modify the graph-based model as well as the RF classification with respect to the selected handcrafted features. Moreover, the preprocessing of the image data would have to be modified, too.

In this paper, we extend our work [26] and present a deep learning algorithm for segmenting multiple layers as well as tattoo areas in OCT image data using a CNN. Unlike in [26], the CNN takes the raw B-scans of the OCT volume as input and outputs a pixelwise prediction for each class label. This significantly reduces the number of processing steps (see Fig. 1). Furthermore, the CNN-based approaches can be more easily adapted to new segmentation problems compared to other works [9, 12, 17, 26], since e.g. no adjustments to handcrafted features or modifications of graph models have to be made. Due to the already mentioned accurate and robust segmentation performance of the U-net, we use this architecture as a basis for our algorithm. Our CNN-based segmentation approach significantly outperforms our previous approach [26] in terms of segmentation accuracy. By using densely-connected convolutions [29] it can be shown that the accuracy of segmentation accuracy increases compared to baseline U-net and becomes more robust against outliers.

2. Materials and methods

2.1. Image dataset

The same OCT image data set as in [26] was used for this work. It consisted of 20 OCT volumes of the inguinal region of four different C57BL/6 mice. OCT images were acquired before, immediately after as well as 30 days after cosmetic treatment. All volumes were acquired using a Telesto II OCT system (Thorlabs, Inc.) at a central wavelength of 1310 nm, an axial resolution of 3.4 μm and a transversal resolution of 6.5 μm. With an image volume size of 1024×1307×1307 voxel, the total field of view is 3.5 × 8.5 × 8.5 mm. A sample spacer was used during acquisition, which was placed directly on the mouse skin to keep the tissue in place and to reduce breathing motion. In addition, water was applied between the sample spacer and the mouse skin, which reduces the scattering of the incident light beam. Fig. 2 shows an example OCT B-scan with the associated FOV image and histology of the corresponding tissue structures. To ensure that the same area can be imaged over a longer period, the mice were tattooed before the first examination to guarantee relocalization. Due to strong motion and shadow artifacts, two volumes had to be excluded, resulting in a total number of 18 OCT volumes, which means that an additional volume was used as in [26]. Training, as well as evaluation of supervised machine learning algorithms, require ground truth data. Due to its high effort, the manual segmentation of entire OCT volumes was not feasible in this work. Therefore, four representative B-scans per volume were selected and manually segmented by two human experts. For this work, the software tool ITK-Snap [30] was used. All pixels of each B-scan were segmented into the following five classes: (epi-)dermis layer (DL) ●, subcutaneous fat layer (SFL) ●, fascia and muscle layer (FML) ●, tattoos ● and background. Since the epidermis layer is difficult to detect at contact points of the sample spacer (see Fig. 2), epidermis and dermis layer were treated as one class. Regions above or below the target structures were considered as background. In total, the image data set consisted of 72 B-scans, each with two individual expert annotations.

Biomedical Optics EXPRESS

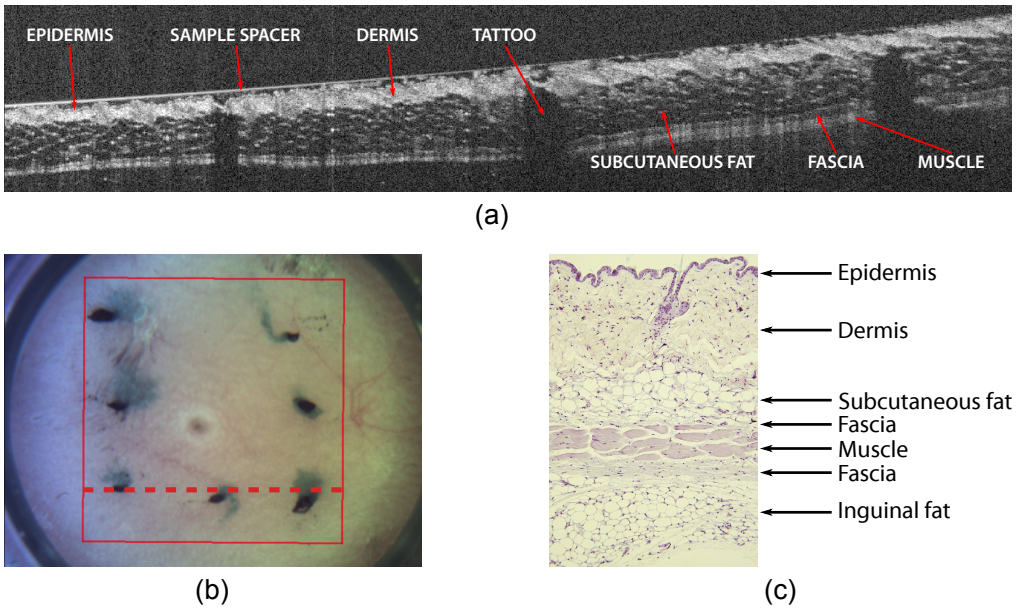

Fig. 2. OCT scan of mouse skin of the inguinal area. (a) B-scan of inguinal mouse skin. Note that the sample spacer is visible as a bright horizontal line in the upper part of the image, whereby it is difficult to distinguish between epidermis and dermis at contact points. The red rectangle in the overview image (b) visualizes the field of view and the dashed line indicates the scan position. A histological section of mouse skin is given in (c).

2.2. Deep learning algorithm

In this paper, we present a deep learning algorithm that automatically segments OCT volumes of the mouse skin. Our algorithm consists of a CNN-based on the architecture of the U-net. The CNN segments an OCT volume by processing each B-scan independently. After being processed, the predicted B-scan segmentations are merged into a single output volume. The architecture of our CNN is described in 2.2.1 and the training in 2.2.2.

2.2.1. Network architecture

The proposed CNN architecture is inspired by the U-net [19], which consists of a contracting encoder and an expanding decoder branch (see Fig. 3). On one hand, the encoder receives an image as input and sequentially generates feature maps on multiple scales and abstraction levels, resulting in a multi-level, multi-resolution feature representation. On the other hand, the decoder receives the feature representations of the encoder and gradually generates an output prediction of every image pixel in the original resolution. In the following, we describe the baseline U-net, which shows only minor changes compared to [19].

The layers of the encoder consist of two 3×3 convolutions, each followed by a rectified linear unit (ReLU) as activation function. A 2×2 average pooling with a stride of two halves the image resolution at the end of each layer. Furthermore, after each pooling operation, the number of feature channels is doubled. Starting with 32 feature channels in the first encoder layer, this results in a number of 512 feature channels in the lowest encoder layer. In the decoder branch, features traverse every double convolution block, halving the number of feature channels per scale level. Furthermore, the feature map resolution is gradually increased using bilinear upsampling [31] in order to match the feature map resolution of the encoder's respective scale level. Subsequently, the upscaled features are concatenated with the corresponding encoder features of the same scale

Biomedical Optics EXPRESS

level via skip connections before they pass the next double convolution block. Skip connections maintain detailed contextual information and enhance gradient flow during backpropagation. The classification layer is realized by a 1×1 convolution where the number of output features equals the number of classes. Finally, a softmax activation function transforms the features of the output layer into probabilities that indicate to which class a pixel belongs to. Batch normalization is applied after each 3×3 convolution for faster convergence during training [32]. In total, encoder and decoder consist of five and four layers respectively.

For this work, an additional skip connection is added around each double convolution block of the introduced baseline U-net (see Fig. 3). Instead of an identity mapping using a residual connection [33], we concatenate the input channels of each individual double convolution block with its output. The number of output channels of the second 3×3 convolution and an additional 1×1 convolution in the skip connection are each set to f/2 where f is the number of output feature channels of each convolution block (see Fig 3). This ensures that f does not change due to the feature concatenation. The use of such densely-connected (DC) blocks has several advantages. They reduce the problem of the disappearing gradient, enhance feature propagation and encourage feature reuse [29].

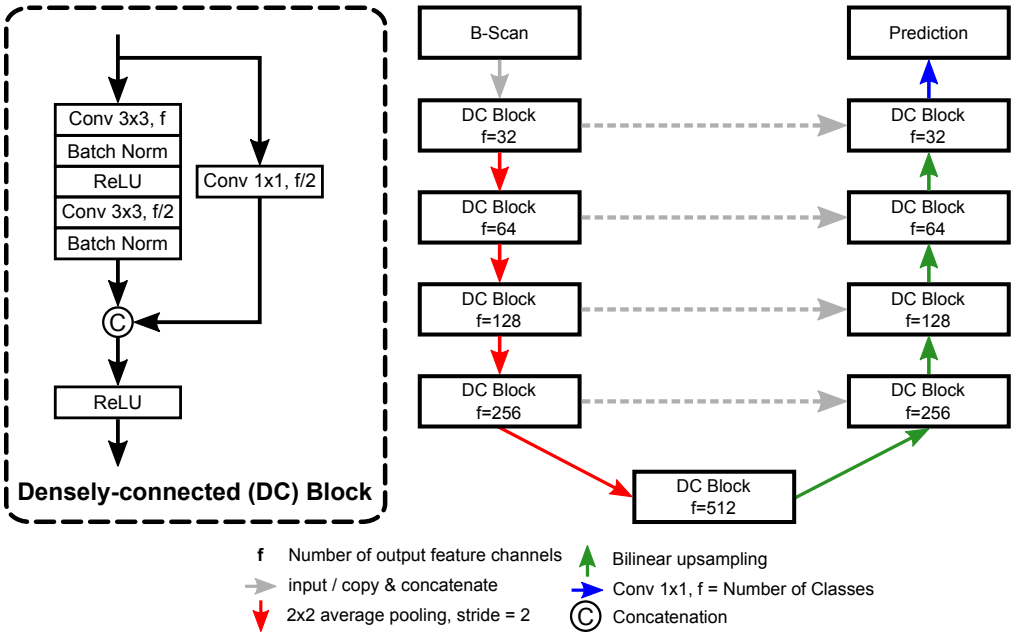

Fig. 3. Architecture of our proposed DCU-net.

2.2.2. Training

We train our network in an end-to-end manner using stochastic gradient descent with Adam optimization [34]. The exponential decay rates β_1 and β_2 are set to 0.9 and 0.999, respectively. The training starts with an initial learning rate of 10^{-4} which is gradually adjusted by an exponential scheduler with decay rate $\gamma = 0.99$.

The Dice metric is utilized as loss function for optimizing the CNN during training [35]. It quantifies the overlap between prediction and ground truth segmentation. More specifically, the

Biomedical Optics EXPRESS

generalized version [36] of the Dice loss is used:

$$\mathcal{J}_{\mathrm{DSC}} = \frac{2\sum_l \alpha_l \sum_{\boldsymbol{x}\in\Omega} p_l(\boldsymbol{x}) g_l(\boldsymbol{x}) + \varepsilon}{\sum_l \alpha_l \sum_{\boldsymbol{x}\in\Omega} p_l^2(\boldsymbol{x}) + \sum_{\boldsymbol{x}\in\Omega} g_l^2(\boldsymbol{x}) + \varepsilon} \tag{1}$$

where $p_l(\boldsymbol{x})$ and $g_l(\boldsymbol{x})$ represent values of the pixel location $\boldsymbol{x}$ of the image domain Ω of the predicted and the ground truth segmentation, respectively. Furthermore, the generalized Dice loss uses an additional weighting factor α_l that provides invariance to each individual class to effectively counteract the class imbalance. Inverse class frequency weighting is used for α_l. The parameter ε ensures numerical stability and is set to 10^{-5}.

We trained each fold for a total of 300 epochs. Furthermore, weight initialization of the CNN is realized utilizing the approach of Glorot et al. [37]. To maximize the GPU usage, we set the batch size to five during training. Furthermore, online data augmentation is performed to deal with the limited amount of training data and to prevent overfitting. We therefore randomly apply horizontal flips and rotations ($\pm 7^\circ$) to each input B-scan image. In order to increase computational efficiency, each B-scan is first cropped to remove unnecessary background and then resized. In addition, zero-padding is used, which results in a final resolution of 256×656 pixels for each input B-scan. Network architecture, as well as training and test procedures, are implemented in Python 3.6 using the PyTorch v0.41 framework [38]. Training is performed on a NVidia GeForce 1080 TI using CUDA v8.0 with cuDNN v5.1.

3. Experiments

3.1. Comparative methods

In order to demonstrate the performance of our proposed CNN-based segmentation approach using DC blocks (DCU-net), we evaluated it against our previously developed algorithm, where an RF classier was used in combination with a graph-based refinement [26]. This approach is referred to as RF+GC in the following. Compared to [26], we extended the graph-based refinement of the RF+GC algorithm for further investigations in the meantime, allowing us now to segment both the DL and tattoo regions. However, modeling the FML turned out to be difficult, therefore it is not computed by the RF+GC algorithm in this work. Tattoo areas were not the focus of a particular research question, but we would like to use them in future works as landmarks for the comparison of different time points. Furthermore, the advantage of DC blocks was evaluated. Therefore, DCU-net was compared with the baseline U-net architecture (U-net) as described in 2.2.1.

In order to fully utilize the data set, we ran an 18-fold cross validation (18FCV). For this purpose, the OCT volumes were divided into non-overlapping training, validation and test data sets for each run in a ratio of 13-4-1. Note that we used the four selected B-scans of each OCT volume (see Sec. 2.1) for the 18FCV. The 18FCV was carried out individually for each expert's ground truth. This means that 18 different CNNs or RFs were trained per expert dataset. No validation set was used for the 18FCV of the RF+GC algorithm.

3.2. Evaluation metrics

Three different metrics are employed for a quantitative assessment of segmentation accuracy. Dice similarity coefficient (DSC) is utilized to quantify the overlap between the predicted and the ground truth labels. The calculation of the DSC takes into account only the intersection of two sets of points, and not the distance to outer points. Also the size of the segmented areas has an effect on the DSC, since misclassifications have a stronger impact on smaller areas than on larger ones. Therefore we additionally use the average symmetric surface distance (ASSD) in this work. Let $S_P = \{p_0, \ldots, p_{n_1}\}$ and $S_G = \{q_0, \ldots, q_{n_2}\}$ be subsets of a predicted segmentation P and a

Biomedical Optics EXPRESS

ground truth G with $S_P \subseteq P$ and $S_G \subseteq G$ containing surface points. The surface distance SD between S_P and S_G is then defined as:

$$SD(S_P, S_G) = \sum_{i=0}^{n_2} \min_{0 \leq j < n_1} \|p_j - q_i\|_2. \tag{2}$$

The surface distance can then be used to determine the ASSD:

$$ASSD = \frac{SD(S_P, S_G)}{2n_2} + \frac{SD(S_G, S_P)}{2n_1}. \tag{3}$$

In addition, the Hausdorff distance (HD) is also reported for outlier detection:

$$HD = \max[SD(S_P, S_G), SD(S_G, S_P)]. \tag{4}$$

4. Results

4.1. Qualitative analysis of segmentation performance

The segmentation results of the algorithms were visually examined and compared with the corresponding ground truth for qualitative analysis. For illustration two B-scans were selected (Figs. 4 and 5).

Overall, there was a high visual agreement between the segmentation results of the algorithms and the ground truth segmentations of the experts, especially for the DL. More noticeable deviations from ground truth could be observed for the SFL and FML since they are more difficult to segment than the DL. One of the reasons for this is the higher variability of SFL and FML and the difficulty to distinguish boundaries. Especially the differentiation between SFL and FML is very diffuse. The reason for this is that SFL and FML are much lower in contrast to DL and more affected by signal extinction as a result of their lower spatial position. The tattoo areas were segmented by all methods with varying performances. In comparison to the CNN-based algorithms, RF+GC segmented tattoo regions with low accuracy. Both U-net and our DCU-net segmented tattoos showed high agreement to the ground truth. Furthermore, the U-net segmented the tattoo regions with a slightly higher visual agreement to ground truth than our proposed DCU-net. Despite the better segmentation performance of the CNN-based algorithms, they showed more outliers which are more strongly represented in segmentations of the U-net (see Figs. 4(g) and 5(i)) than in those the DCU-net (see Figs. 4(j) and 5(k)). The vast majority of outliers are located in areas of deeper tissue layers, where inguinal fat and fascia tissue also occur and may be classified by CNNs as SFL or FML. These errors were compensated by the graph-based refinement of the RF+GC algorithm.

4.2. Quantitative analysis of segmentation performance

The quantitative analysis of the segmentation performance of the algorithms was based on the metrics described in Sec. 3.2, which were calculated using the manual expert segmentations. Means and standard deviations for each metric are shown in Table 1. In addition, box plots visualize the result distributions for all classes except for tattoos in Fig. 6. In addition, the inter-rater reliability (IRR) between the manual annotations of both experts was calculated. The Wilcoxon signed-rank test with a significance level of 5% was employed to identify significant differences.

The DCU-net showed a significant better segmentation performance for DSC and ASSD compared to the RF+GC algorithm for all classes. In addition to minor improvements for the DL and SFL classes, an increase in DSC of 110% for the tattoo class compared to the RF+GC algorithm was observed. The ASSDs were reduced by 8% for the DL and by 11% for the SFL. The U-net also showed good results for DSC and ASSD. For the tattoo class, the U-net

Biomedical Optics EXPRESS

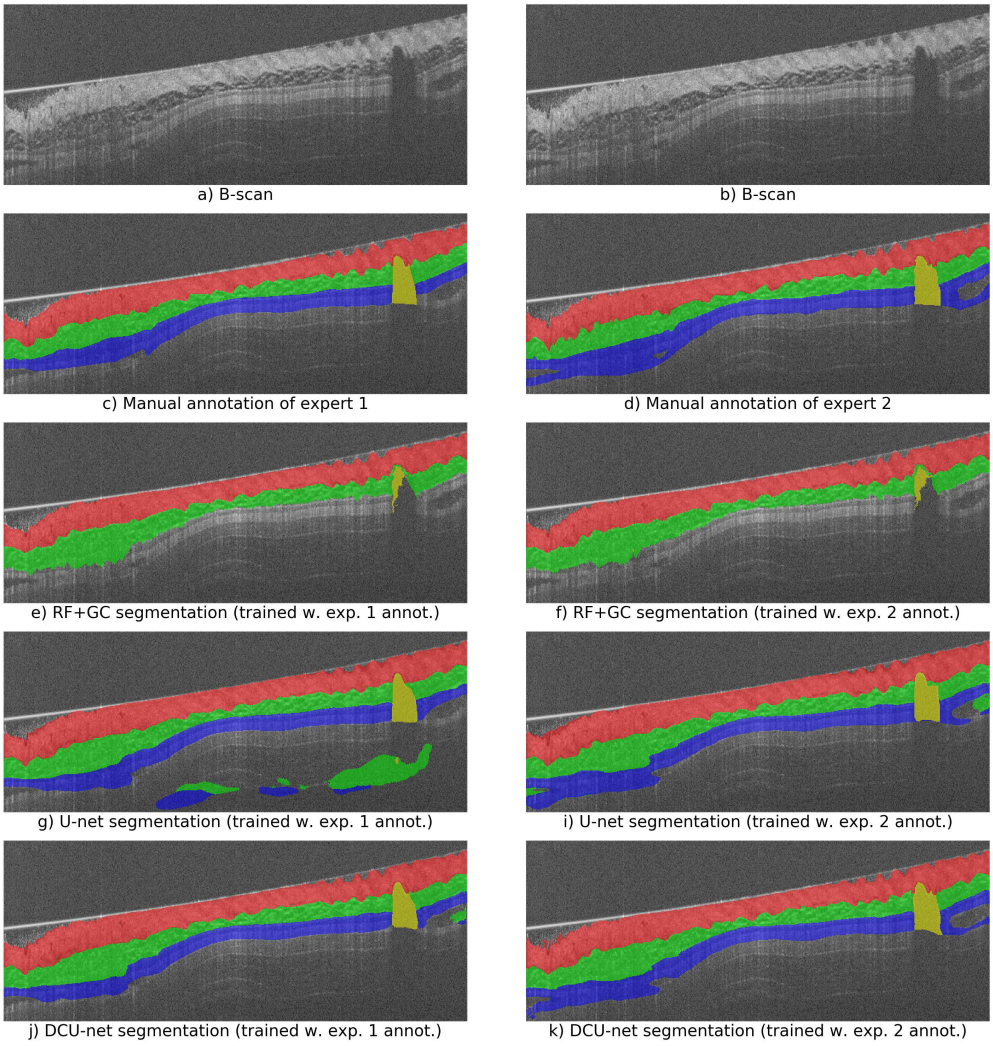

Fig. 4. Segmentation results of a single OCT B-scan. The B-scan without segmentation overlay is shown in a)/b), with expert annotations in c)/d) and with segmentations by RF+GC, U-net and DCU-net in e)/f), g)/h) and i)/j), respectively. e), g), j) or f), i), k) show segmentation results of the 18FCV which was performed with annotations of expert 1/2. Color mapping is explained as follows: DL ●, SFL ●, FML ● and tattoos ●. Note that the FML was not segmented by the RF+GC algorithm.

achieved the best result with a DSC of 0.718, which is an improvement of 12% compared to the DCU-net result and an improvement of 135% compared to the RF+GC algorithm. However, no significant improvements in the remaining DSC and ASSD results of the U-net compared to the RF+GC algorithm could be observed. Furthermore, the DCU-net showed a significantly better segmentation performance for all metrics except the DSC for the FML and tattoo class compared to the U-net.

In contrast to the CNNs, the RF+GC algorithm achieved the lowest HDs, which differ significantly from those of the DCU-net and the U-net. The increased robustness of the algorithm could already be seen in the qualitative evaluation (see Sec. 3.1) and is explained by the

Biomedical Optics EXPRESS

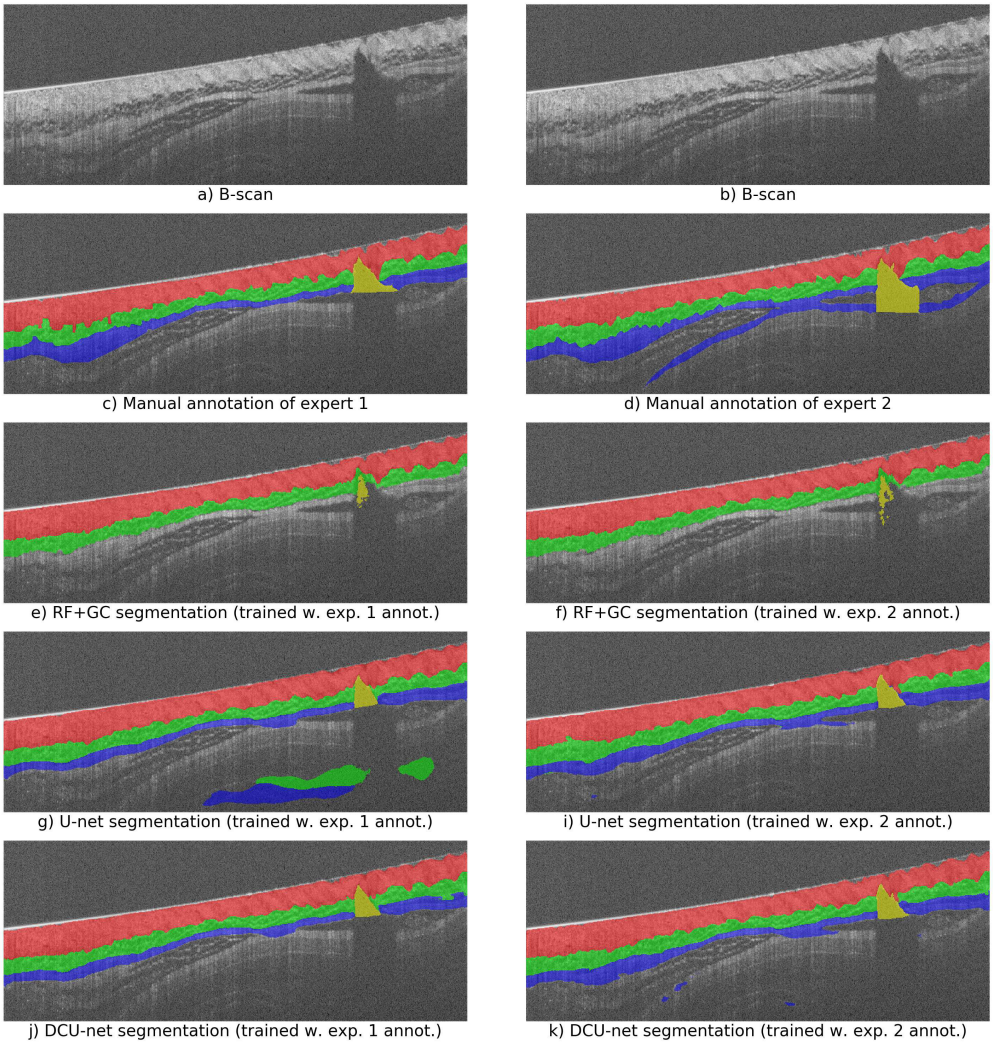

Fig. 5. Segmentation results of a single OCT B-scan. The B-scan without segmentation overlay is shown in a)/b), with expert annotations in c)/d) and with segmentations by RF+GC, U-net and DCU-net in e)/f), g)/h) and i)/j), respectively. e), g), j) or f), i), k) show segmentation results of the 18FCV which was performed with annotations of expert 1/2. Color mapping is explained as follows: DL ●, SFL ●, FML ● and tattoos ●. Note that the FML was not segmented by the RF+GC algorithm.

graph-based refinement. Both DCU-net and U-net produced more outliers (see Figs. 4 and 5), which is confirmed by the significantly higher HDs. Also for the HDs, the DCU-net provided significantly better results than the U-net.

With a few exceptions, no significant difference between DSC and ASSDs in relation to the IRR could be found. Only for the SFL predictions the DCU-net showed significantly higher ASSDs to ground truth than the IRR. For the remaining classes there are no significant differences in DSC and ASSD between the RF+GC and the U-net. Due to the presence of outliers, the HDs of all algorithms were significantly larger compared to the IRR. The U-net only achieved the same segmentation performance as the experts for the DSC. In contrast, the U-net segmented

Biomedical Optics EXPRESS

Table 1. Average results of the two 18FCVs are presented, conducted with annotations of the respective expert. The best performances are printed in bold. The distances for ASSD and HD are given in μm (pixel height and width correspond to 6.8 μm and 13 μm, respectively). Note that the FML class was not computed by the RF+GC algorithm. In addition, the inter-rater reliability (IRR) between both experts was determined.

		DL 🔴	SFL 🟢	FML 🔵	Tattoo 🟡
DSC	RF+GC	0.945 ± 0.015	0.910 ± 0.061	–	0.305 ± 0.248
	U-net	0.946 ± 0.013	0.920 ± 0.054	0.873 ± 0.067	**0.718 ± 0.270**
	DCU-net	**0.950 ± 0.013**	**0.925 ± 0.058**	**0.874 ± 0.071**	0.639 ± 0.264
	IRR	0.949 ± 0.014	0.937 ± 0.028	0.881 ± 0.052	0.583 ± 0.380
ASSD	RF+GC	9.98 ± 4.10	14.60 ± 9.78	–	–
	U-net	10.20 ± 3.17	18.53 ± 17.61	20.48 ± 22.80	–
	DCU-net	**9.16 ± 3.23**	**13.00 ± 9.37**	**17.12 ± 21.01**	–
	IRR	9.05 ± 3.13	9.96 ± 4.13	11.97 ± 7.17	–
HD	RF+GC	**86.67 ± 44.46**	**140.84 ± 89.69**	–	–
	U-net	152.14 ± 101.76	261.72 ± 212.63	265.01 ± 207.40	–
	DCU-net	113.61 ± 84.99	155.48 ± 123.54	**222.92 ± 181.59**	–
	IRR	87.82 ± 46.74	95.10 ± 48.85	128.67 ± 87.90	–

tattoo areas significantly better than the experts. The RF+GC algorithm only reaches the expert level for the DL class' DSC scores.

5. Discussion

In this work, we presented a deep learning algorithm for the segmentation of different tissues and tattooed areas of mouse skin in OCT image data. A CNN was used based on the architecture of the U-net. In addition to minor architectural changes, such as average pooling or bilinear upsampling, we extended the standard convolution blocks to densely-connected blocks (see Fig. 3). We evaluated our approach against a previously developed algorithm (RF+GC). In addition, we tested our approach against a baseline U-net to demonstrate the advantages of densely-connected feature extraction. DSC, ASSD and HD were used as metrics for the evaluation. We were able to show that the segmentation performance of our new CNN-based DCU-net method is very accurate (cf. Tab. 1). DCU-net showed both higher DSC scores and lower ASSDs compared to RF+GC. The DCU-net achieved on average better results than the baseline U-net except for the tattoo class. Furthermore, it could be shown that the RF+GC algorithm makes less precise predictions in areas with high variability, such as tattoo areas or the SFL. In particular, modeling tattoo areas or the SFL is a challenging task and difficult to describe using handcrafted features (RF) and structure assumptions (GC). In contrast, the graph model of the RF+GC algorithm offers robust segmentation, which can be seen in the comparatively small HDs. Compared to our previous method, which involves four steps (see Fig. 1), (DC)U-net is able to jointly learn both expressive local features and an accurate global classifier which robustly segments target structures without pre- or postprocessing within a single model. Furthermore, the DC blocks not only reduce the number of parameters to be learned, but also improve the segmentation accuracy in the area of object contours and increase the robustness against outliers. This becomes obvious when ASSDs and HDs are compared between the standard U-net and the DCU-net.

Biomedical Optics EXPRESS

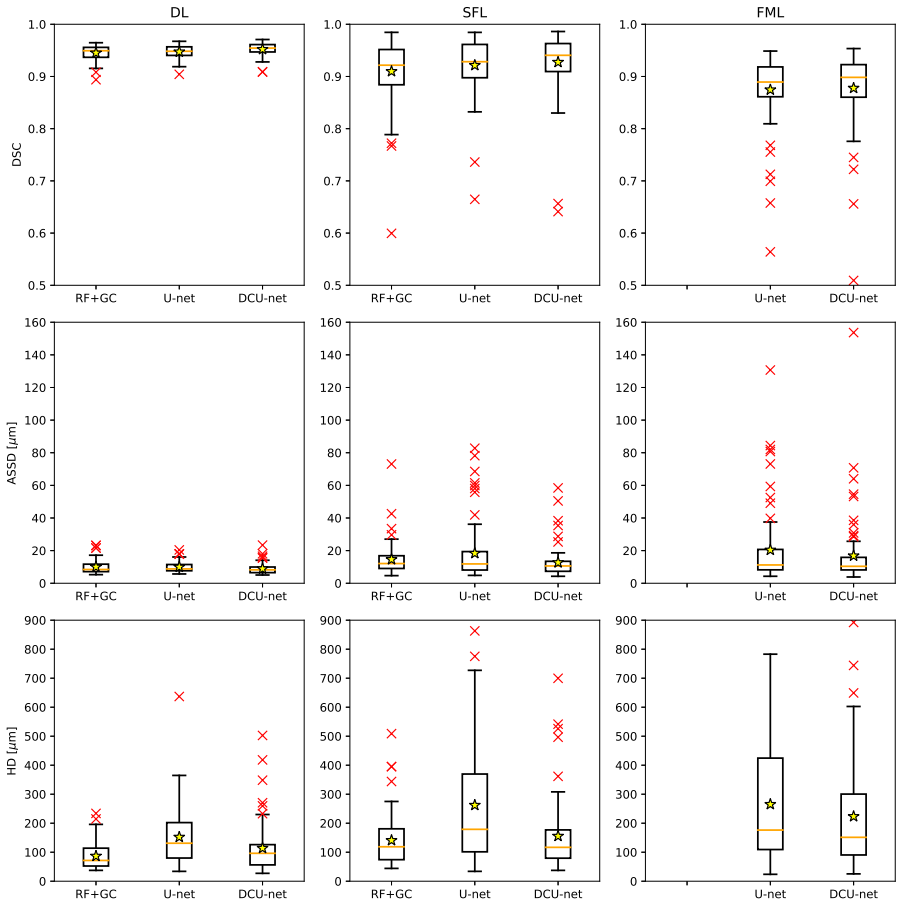

Fig. 6. Quantitative analysis of the comparative algorithms. Box plots represent the result distributions of the averaged metrics for DL, SFL and FML from both 18FCVs. Note that the FML was not calculated by the RF+GC algorithm.

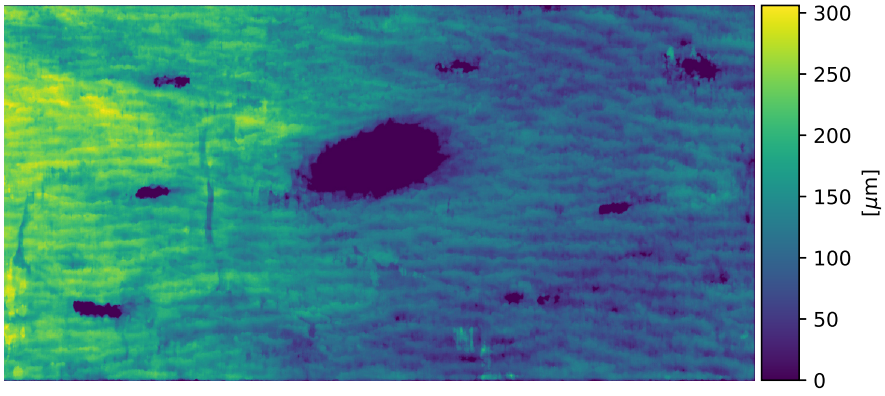

Fig. 7. Thickness map visualization of the SFL. Each single B-scan of the OCT volume was segmented by the DCU-net. Remaining outliers were eliminated using minor postprocessing steps.

Biomedical Optics EXPRESS

The qualitative evaluation showed that the predictions of the CNN-based algorithms had both smaller (DCU-net) and larger (U-net) topology errors in contrast to those of the RF+GC algorithm. To address this problem, we plan to integrate topology-preserving conditions directly into the loss function in future works. In addition, we plan to extend our approach to three-dimensional space so that the DCU-net, similar to the RF+GC algorithm, receives further neighborhood information as input.

Due to its high segmentation accuracy, our algorithm is able to provide the basis for quantitative measurements, e.g. the thickness measurement of individual skin layers (see Fig. 7), which are important for studies such as in [7]. Furthermore, the DCU-net can be more easily adapted to new research questions or animal models in comparison to the RF+GC approach, which simplifies the practical use.

Funding

This work was partly supported by the Graduate School for Computing in Medicine and Life Sciences funded by Germany's Excellence Initiative [DFG GSC 235/2].

Disclosures

The authors declare that there are no conflicts of interest related to this article.

References

1. D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito *et al.*, "Optical coherence tomography," Science **254**, 1178–1181 (1991).
2. M. R. Hee, J. A. Izatt, E. A. Swanson, D. Huang, J. S. Schuman, C. P. Lin, C. A. Puliafito, and J. G. Fujimoto, "Optical coherence tomography of the human retina," Arch. Ophthalmol. **113**, 325–332 (1995).
3. J. Welzel, "Optical coherence tomography in dermatology: a review," Ski. Res. Technol. Rev. article **7**, 1–9 (2001).
4. T. Gambichler, G. Moussa, M. Sand, D. Sand, P. Altmeyer, and K. Hoffmann, "Applications of optical coherence tomography in dermatology," J. Dermatol. Sci. **40**, 85–94 (2005).
5. A. J. Singer, Z. Wang, S. A. McClain, and Y. Pan, "Optical coherence tomography: a noninvasive method to assess wound reepithelialization," Acad. Emerg. Medicine **14**, 387–391 (2007).
6. S. M. Srinivas, J. F. de Boer, B. H. Park, K. Keikhanzadeh, H.-E. L. Huang, J. Zhang, W. G. Jung, Z. Chen, and J. S. Nelson, "Determination of burn depth by polarization-sensitive optical coherence tomography," J. Biomed. Opt. **9**, 207–213 (2004).
7. N. Salma, M. Evers, M. J. Casper, C. Droigk, T. Kepp, H. Handels, and D. Manstein, "Mouse model of cold-induced localized fat loss (selective cryolipolysis)," in *Lasers in Surgery and Medicine,* vol. 50 (Wiley, 2018), pp. S19–S20.
8. K. Li, X. Wu, D. Z. Chen, and M. Sonka, "Optimal surface segmentation in volumetric images-a graph-theoretic approach," IEEE Trans. on Pattern Anal. Mach. Intell. **28**, 119–134 (2006).
9. M. K. Garvin, M. D. Abràmoff, R. Kardon, S. R. Russell, X. Wu, and M. Sonka, "Intraretinal layer segmentation of macular optical coherence tomography images using optimal 3-d graph search," IEEE Trans. Med. Imaging **27**, 1495–1505 (2008).
10. S. J. Chiu, X. T. Li, P. Nicholas, C. A. Toth, J. A. Izatt, and S. Farsiu, "Automatic segmentation of seven retinal layers in sdoct images congruent with expert manual segmentation," Opt. Express **18**, 19413–19428 (2010).
11. R. Kafieh, H. Rabbani, M. D. Abramoff, and M. Sonka, "Intra-retinal layer segmentation of 3d optical coherence tomography using coarse grained diffusion map," Med. Image Anal. **17**, 907–928 (2013).
12. S. J. Chiu, M. J. Allingham, P. S. Mettu, S. W. Cousins, J. A. Izatt, and S. Farsiu, "Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema," Biomed. Opt. Express **6**, 1172–1194 (2015).
13. A. G. Roy, S. Conjeti, S. P. K. Karri, D. Sheet, A. Katouzian, C. Wachinger, and N. Navab, "ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks," Biomed. Opt. Express **8**, 3627–3642 (2017).
14. F. G. Venhuizen, B. van Ginneken, B. Liefers, M. J. van Grinsven, S. Fauser, C. Hoyng, T. Theelen, and C. I. Sánchez, "Robust total retina thickness segmentation in optical coherence tomography images using convolutional neural networks," Biomed. Opt. Express **8**, 3292–3316 (2017).
15. S. Apostolopoulos, S. De Zanet, C. Ciller, S. Wolf, and R. Sznitman, "Pathological oct retinal layer segmentation using branch residual u-shape networks," in International Conference on Medical Image Computing and Computer-Assisted Intervention, (Springer, 2017), pp. 294–301.

Biomedical Optics EXPRESS

16. S. K. Devalla, P. K. Renukanand, B. K. Sreedhar, G. Subramanian, L. Zhang, S. Perera, J.-M. Mari, K. S. Chin, T. A. Tun, N. G. Strouthidis *et al.*, "DRUNET: a dilated-residual U-Net deep learning network to segment optic nerve head tissues in optical coherence tomography images," Biomed. Opt. Express **9**, 3244–3265 (2018).
17. B. J. Antony, A. Lang, E. K. Swingle, O. Al-Louzi, A. Carass, S. Solomon, P. A. Calabresi, S. Saidha, and J. L. Prince, "Simultaneous segmentation of retinal surfaces and microcystic macular edema in sdoct volumes," in *Medical Imaging 2016: Image Processing,* vol. 9784 (International Society for Optics and Photonics, 2016), p. 97841C.
18. L. Fang, D. Cunefare, C. Wang, R. H. Guymer, S. Li, and S. Farsiu, "Automatic segmentation of nine retinal layer boundaries in oct images of non-exudative amd patients using deep learning and graph search," Biomed. Opt. Express **8**, 2732–2744 (2017).
19. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical Image Computing and Computer-assisted Intervention (Springer, 2015), pp. 234–241.
20. A. Yazdanpanah, G. Hamarneh, B. R. Smith, and M. V. Sarunic, "Segmentation of intra-retinal layers from optical coherence tomography images using an active contour approach," IEEE Trans. Med. Imaging **30**, 484–496 (2011).
21. A. Mishra, A. Wong, K. Bizheva, and D. A. Clausi, "Intra-retinal layer segmentation in optical coherence tomography images," Opt. Express **17**, 23719–23728 (2009).
22. K. W. Gossage, T. S. Tkaczyk, J. J. Rodriguez, and J. K. Barton, "Texture analysis of optical coherence tomography images: feasibility for tissue classification," J. Biomed. Opt. **8**, 570–576 (2003).
23. P. Pande, S. Shrestha, J. Park, M. J. Serafino, I. Gimenez-Conti, J. L. Brandon, Y.-S. Cheng, B. E. Applegate, and J. A. Jo, "Automated classification of optical coherence tomography images for the diagnosis of oral malignancy in the hamster cheek pouch," J. Biomed. Opt. **19**, 086022 (2014).
24. D. Sheet, A. Chaudhary, S. P. K. Karri, D. Das, A. Katouzian, P. Banerjee, N. Navab, J. Chatterjee, and A. K. Ray, "In situ histology of mice skin through transfer learning of tissue energy interaction in optical coherence tomography," J. Biomed. Opt. **18**, 090503 (2013).
25. D. Sheet, S. P. K. Karri, A. Katouzian, N. Navab, A. K. Ray, and J. Chatterjee, "Deep learning of tissue specific speckle representations in optical coherence tomography and deeper exploration for in situ histology," in 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI) (IEEE, 2015), pp. 777–780.
26. T. Kepp, C. Droigk, M. Casper, M. Evers, N. Salma, D. Manstein, and H. Handels, "Segmentation of subcutaneous fat within mouse skin in 3d oct image data using random forests," in Medical Imaging 2018: Image Processing, vol. 10574 (SPIE, 2018), pp. 1057426–1–1057426–8.
27. J. Rogowska and M. E. Brezinski, "Image processing techniques for noise removal, enhancement and segmentation of cartilage oct images," Phys. Med. Biol. **47**, 641 (2002).
28. D. Manstein, H. Laubach, K. Watanabe, W. Farinelli, D. Zurakowski, and R. R. Anderson, "Selective cryolysis: A novel method of non-invasive fat removal," Lasers Surg. Med. **40**, 595–604 (2008).
29. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (IEEE, 2017), pp. 4700–4708.
30. P. A. Yushkevich, J. Piven, H. Cody Hazlett, R. Gimpel Smith, S. Ho, J. C. Gee, and G. Gerig, "User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability," Neuroimage **31**, 1116–1128 (2006).
31. A. Odena, V. Dumoulin, and C. Olah, "Deconvolution and checkerboard artifacts," Distill **1**, e3 (2016).
32. S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," https://arxiv.org/abs/1502.03167.
33. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (IEEE, 2016), pp. 770–778.
34. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," https://arxiv.org/abs/1412.6980.
35. F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in 2016 Fourth International Conference on 3D Vision (3DV) (IEEE, 2016), pp. 565–571.
36. C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support,* (Springer, 2017), pp. 240–248.
37. X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS, 2010), pp. 249–256.
38. A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," in 31st Conference on Neural Information Processing Systems (NIPS, 2017), pp. 1–4.