Segmentation of Vessels in Ultra High Frequency Ultrasound Sequences Using Contextual Memory

Tejas Sudharshan Mathai[1(✉)], Vijay Gorantla[2], and John Galeotti[1]

[1] The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA
tmathai@andrew.cmu.edu

[2] Department of Surgery, Wake Forest Institute for Regenerative Medicine, Winston-Salem, NC 27101, USA

Abstract. High resolution images provided by Ultra High Frequency Ultrasound (UHFUS) scanners permit the vessel-based measurement of the Intimal-Media Thickness (IMT) in small vessels, such as those in the hand. However, it is challenging to precisely determine vessels in UHFUS sequences due to severe speckle noise obfuscating their boundaries. Current level set-based approaches are unable to identify poorly delineated boundaries and are not robust against varying speckle noise. While recent neural network-based methods, including recurrent neural networks, have shown promise at segmenting vessel contours, they are application specific and do not generalize to datasets acquired from different scanners, such as a traditional High Frequency Ultrasound (HFUS) machine, with different scan settings. Our goal for a segmentation approach was the accurate localization of vessel contours, and generalization to new data within and across biomedical imaging modalities. In this paper, we propose a novel ultrasound vessel segmentation network (USVS-Net) architecture that assimilates features extracted at different scales using Convolutional Long Short Term Memory (ConvLSTM) and segments vessel boundaries accurately. We show the results of our approach on UHFUS and HFUS sequences. To show broader applicability beyond US, we also trained and tested our approach on a Chest X-Ray dataset. To the best of our knowledge, this is the first learning-based approach to segment deforming vessel contours in both UHFUS and HFUS sequences.

Keywords: Ultrasound · Vasculature · Segmentation · Deep learning

1 Introduction

Intima-Media Thickness (IMT) is a parameter that quantifies risk in clinical applications, such as atherosclerotic plaque buildup [1]. In particular however,

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-32245-8_20) contains supplementary material, which is available to authorized users.

D. Shen et al. (Eds.): MICCAI 2019, LNCS 11765, pp. 173–181, 2019.
https://doi.org/10.1007/978-3-030-32245-8_20

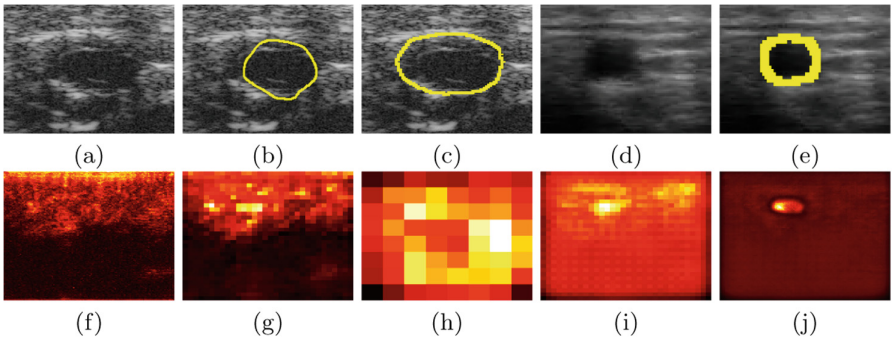

Fig. 1. (a) Still frame capturing a pulsating vessel acquired using UHFUS; (b) Segmentation (yellow contour) from a level set method [6] bleeds into the tissue region due to poor boundary contrast; (c) Final segmentation from the proposed USVS-Net; (d) Frame acquired using HFUS (zoomed), and (e) its associated final vessel segmentation; Activations of the network for the vessel imaged in (a) at different network depths: (f) downsampling level 1; (g) downsampling level 3; (h) downsampling level 5; (i) upsampling level 3; (j) upsampling level 1.

it can be used to track the functional progress of hand transplant recipients, where the gold standard for monitoring changes is currently histopathology [2]. Recently, Ultra-High Frequency Ultrasound (UHFUS) has been shown to quantitatively measure IMT through the resolution of vessel structures at 0.03 mm within a shallow tissue depth of ∼1 cm [1]. However, this improved resolution is traded-off with an increase in speckle noise corrupting the vessel boundaries, which is in contrast to traditional ultrasound and high frequency ultrasound (HFUS) machines [1]. Furthermore, vessels at shallow depths contort themselves significantly (due to transducer pressure and motion) as opposed to vessels deeper in the body, such as the carotid artery [3,4]. Therefore, the key motivation of this work is the sub-mm localization of rapidly moving and pulsating vessel contours in UHFUS and HFUS sequences to compare changes in IMT over time.

Prior vessel-based segmentation approaches for ultrasound sequences fall into two categories: traditional and learning-based methods. Traditional approaches, such as state-of-the-art level set methods for HFUS and UHFUS [5,6], are quick to execute, but lack robustness needed in clinical use due to the fine-tuning of parameters. In contrast, learning-based approaches are resilient to changes in scan settings and variations in image quality. In particular, Convolutional Neural Networks (CNNs) [3,4] have made great strides in integrating features extracted at multiple scales through feature forwarding [7], residual learning [8], dilated convolutions [9] etc. However, these methods segment longitudinal vessels in ultrasound videos, and are task specific without adequately harnessing inter-frame vessel dynamics. Long Short Term Memory (LSTM) networks [10–15] intelligently combine multi-scale features to retain relevant features over video time steps, and only update the features when required. Some of these

approaches have shown applicability to microscopy [10], X-Ray [11] etc. We tested the performance of these Convolutional LSTM (ConvLSTM) methods on the challenging task of segmenting highly deformable vessel contours in UHFUS and HFUS sequences. However, they did not accurately segment vessel cross-sections (see supplementary material). An ideal approach would accurately segment boundaries, while generalizing to data within and across biomedical imaging modalities.

In this paper, we propose a novel ConvLSTM-based ultrasound vessel segmentation network called USVS-Net to segment transverse vessel cross-sections in UHFUS and HFUS sequences. This network was influenced by methods designed for different anatomies (retina [16], cornea [17], microscopy [10], X-Ray [11]). Validation of our method was conducted on 38 UHFUS and 6 HFUS sequences respectively.

Contribution. (1) We propose a novel USVS-Net architecture that outperforms current ConvLSTM networks on vessel segmentation tasks. (2) To gauge the potential broader applicability of our work to other biomedical domains, we trained and tested our method on the Montgomery County Chest X-Ray dataset [18] with comparable results to the state-of-the-art [11].

2 Methods

As seen in Fig. 2, the proposed USVS-Net design is comprised of two sections: a downsampling encoder and a ConvLSTM-based decoder. Our network design is different from the traditional U-Net [7] based segmentation models, which treat each frame in a sequence independently. ConvLSTM-based models implement a memory mechanism [10–15] that considers the inter-relation between video frames to retain vessel appearance over multiple scales for dense pixel-wise predictions. By combining the ConvLSTM in the decoder with the spatial context gathered in the encoder, spatio-temporal vessel-related features are estimated for improved segmentation.

Encoder. The encoder structure is inspired by the approaches in [16,17], which have shown applicability to retina and cornea tissue interface segmentation. The blocks in the encoder pull out meaningful representations of the vessel appearance over multiple scales using dilated convolutions [9] and residual connections [8]. As shown in Fig. 1, the feature maps characterized at the first few layers of the encoder depict finely defined properties (edges, corners etc.), which are low-level attributes, and are limited due to their smaller receptive field. At the deeper layers of the network, coarse, but complex attributes are seen with poorly defined contours. At this level, more of the image is seen on a global scale due to the larger receptive field. Residual connections and dilated convolutions gather more spatial information, especially relating to faintly discernible boundaries [17], and inculcate this information from one block to the next to prevent holes in the final segmentation. Yet, this hierarchical representation is not enough on its own to model the dynamics of vessel movement in a video sequence. By forwarding the feature maps extracted at different scales to the ConvLSTM cells,

which can retain relevant features of interest in memory, they can be integrated to produce segmentations of better quality and precision [10,11].

Decoder. Every encoder block forwards its output feature maps to a ConvLSTM unit in the decoder section. In this work, we incorporate the structured LSTM proposed in [12]. These LSTM cells consider the output of each encoder block as a single time step, and implement a memory mechanism wherein the features extracted at multiple scales are integrated in a coarse-to-fine manner. This is done by gating structures that carefully regulate the removal or addition of new information to the cell state. In this manner, global contextual information from the deepest encoder layer is observed by the LSTM unit first, and as the receptive fields are reduced, finer details about the vessel contour are added.

From Fig. 2, each LSTM unit uses three feature sets (input, hidden, and cell), and outputs information using three gates: forget, input, and output. The forget gate removes information from the cell state. The input gate determines the new information that will be incorporated into the cell state. Finally, the output of the LSTM unit is regulated by the output gate. Contrary to [12],

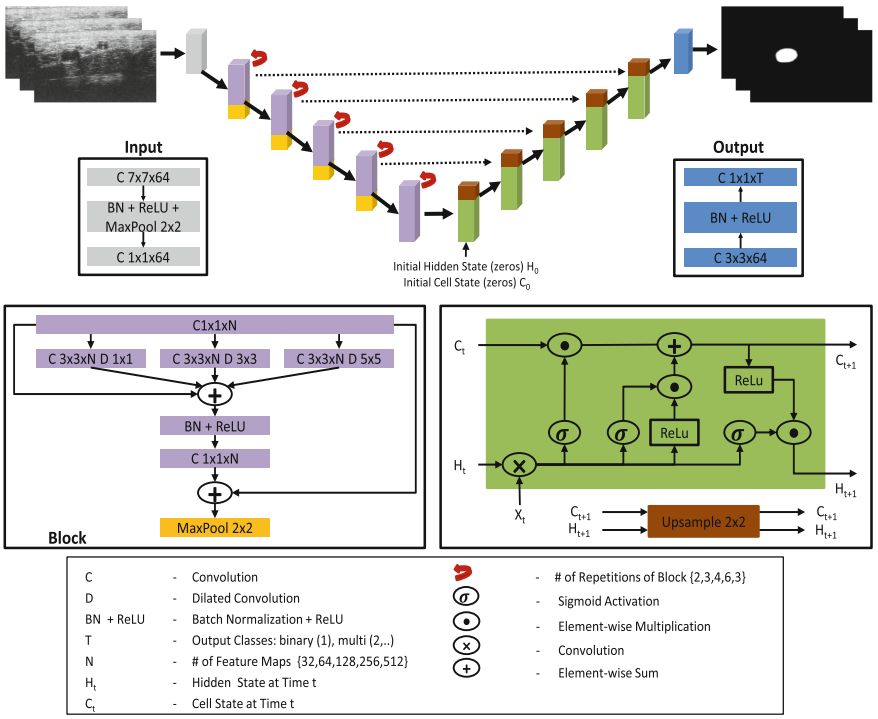

Fig. 2. The USVS-Net architecture contains encoding (purple) and decoding (green) sections. The encoder uses of residual connections and dilated convolutions to extract features, while the decoder uses structured ConvLSTM blocks to retain vessel shape attributes and segment the vessel. (Color figure online)

bi-directional LSTMs were not used in this work as our video sequences can be of arbitrary length with non-smooth vessel motion between consecutive frames, c.f. Fig. 3, making their implementation impractical. We employed convolution in the structured LSTM unit, and replaced the tanh operation with a ReLU as we empirically observed an improved segmentation accuracy. Similar to [11], the initial hidden and cell states were set to zero, and the hidden and cell states of the other LSTM units were upsampled from the LSTM unit below (see Fig. 2).

3 Experiments and Results

Data. Previously acquired (free-hand) deidentified video sequences from an existing research database [2,6] were used in this work, and they came from two scanners: a Visualsonics Vevo 2100 UHFUS machine (Fujifilm, Canada), and a Diasus HFUS scanner (Dynamic Imaging, UK). The UHFUS scanner provided a 50 MHz transducer with physical resolution of 30 µm and a pixel spacing of 11.6 µm. 58 UHFUS sequences were used, each containing 100 2D B-scans with dimensions of 832×512 pixels. The HFUS scanner had a 10–22 MHz transducer with a pixel spacing of 92.5 µm. 26 HFUS sequences were used, each containing a variable number of 2D B-scans (50–250) with dimensions of 280×534 pixels. All the sequences contained arteries of the hand (eg. superficial palmar arch) with a wide range of adjustable gain settings (40–70 dB). Extensive probe motions were also acquired, such as longitudinal scanning, beating vessels, out-of-plane vessel deformation etc. An expert grader annotated all the 84 UHFUS and HFUS sequences. To show general applicability, we also retrained and tested our architecture on the Montgomery County Chest X-Ray dataset [18], which contained 138 annotated images with 58 abnormal and 80 normal cases.

Setup. Of the 58 UHFUS sequences, 20 were chosen for training and the remaining 38 were used for testing. Similarly, from the 26 HFUS sequences, 20 were chosen for training and the remaining 6 were used for testing. We ran a 3-fold cross-validation for the vessel segmentation task. To simulate a clinical application, an ensemble of the two best models with the lowest validation loss (from a single fold) were used for testing. Similar to [11], we also ran a 3-fold cross validation for the lung segmentation task in the CXR dataset.

Baseline Comparisons. For the vessel segmentation task, we compared our errors against those from a level set-based method [6], and two LSTM-based segmentation approaches: DecLSTM [10] and CFCM34 [11]. For the lung segmentation task, we compared against the state-of-the-art CFCM34 model [11].

Training. Our sequences contained variable image sizes and training a ConvLSTM with full-sized images is limited by GPU RAM. We trained our USVS-Net by scaling each B-scan to 256×256 pixels. Data augmentation (elastic deformation, blurring etc.) was done to increase the training set to ∼120,000 images. To compare against [11], we used the generalized dice coefficient [11] loss with the ADAM optimizer [19], and set the batch size to 16 with a learning rate of 0.00001 for 30 epochs. The final pixel level probabilities were classified using the

Table 1. Segmentation error comparison for the UHFUS (top) and HFUS (bottom) sequences. (* 33/38 sequences successful)

Method	DSC	HD (mm)	MAD (mm)	DFPD	DFND	Prec	Rec
Traditional* [6]	81.13 ± 3.72	0.21 ± 0.05	0.06 ± 0.02	$\mathbf{3.08 \pm 1.68}$	8.71 ± 0.55	$\mathbf{96.44 \pm 2.56}$	72.03 ± 4.9
DecLSTM [10]	88.83 ± 3.74	0.15 ± 0.06	0.04 ± 0.03	6.76 ± 1.05	$\mathbf{5.35 \pm 1.4}$	87.54 ± 4.45	92.46 ± 3.93
CFCM34 [11]	88.45 ± 3.97	0.15 ± 0.07	0.04 ± 0.04	$\mathbf{6.41 \pm 1.21}$	5.51 ± 1.39	88.07 ± 4.83	91.31 ± 3.87
USVS-Net	$\mathbf{92.15 \pm 2.29}$	$\mathbf{0.11 \pm 0.03}$	$\mathbf{0.03 \pm 0.01}$	6.83 ± 1.13	6.33 ± 1.36	$\mathbf{91.76 \pm 3.78}$	$\mathbf{93.2 \pm 3.34}$
Traditional [6]	83.6 ± 5.47	0.47 ± 0.13	0.08 ± 0.04	$\mathbf{2.08 \pm 2.01}$	6.02 ± 0.51	$\mathbf{95.13 \pm 4.8}$	75.42 ± 7.49
DecLSTM [10]	88.34 ± 5.21	0.39 ± 0.1	0.05 ± 0.3	4.23 ± 0.97	5.61 ± 0.78	87.21 ± 3.15	83.94 ± 7.61
CFCM34 [11]	89.44 ± 3.34	0.36 ± 0.09	0.05 ± 0.02	3.74 ± 1.04	5.23 ± 0.62	94.21 ± 3.48	85.74 ± 5.51
USVS-Net	$\mathbf{89.74 \pm 3.05}$	$\mathbf{0.36 \pm 0.08}$	$\mathbf{0.04 \pm 0.02}$	4.98 ± 0.86	$\mathbf{4.53 \pm 1.03}$	88.63 ± 0.05	$\mathbf{91.52 \pm 0.05}$

Table 2. Segmentation error comparison (pixels) for the Montgomery County Chest X-Ray dataset.

Method	DSC	HD	MAD	DFPD	DFND	Prec	Rec
CFCM34 [11]	$\mathbf{97.01 \pm 1.82}$	11.05 ± 10.78	0.13 ± 0.31	6.67 ± 0.97	$\mathbf{6.39 \pm 0.98}$	96.93 ± 2.42	$\mathbf{97.25 \pm 2.67}$
USVS-Net	96.89 ± 1.80	$\mathbf{10.29 \pm 8.26}$	$\mathbf{0.10 \pm 0.19}$	$\mathbf{6.64 \pm 0.89}$	6.73 ± 1.04	$\mathbf{97.15 \pm 1.65}$	96.57 ± 2.97

softmax function, and the connected component in the foreground class was considered the segmentation. For the DecLSTM [10], we used RMSProp optimizer [10], weighted cross-entropy loss [7], and a learning rate of 0.0001 for 30 epochs.

Metrics. We compared each baseline's results against the expert annotation. The following metrics were calculated to quantify errors: (1) Dice Similarity Coefficient (DSC) [6], (2) Hausdorff Distance (HD) in millimeters [6], (3) Mean Absolute Deviation (MAD) in millimeters [11], (4) Definite False Positive and Negative Distances (DFPD, DFND) [6], (5) Precision (Prec.) and (6) Recall (Rec.) [11]. Videos and additional visualizations are provided in the supplementary material that detail the results of vessel segmentation in UHFUS and HFUS sequences.

4 Discussion

UHFUS Results. Our primary assumption was that a vessel would be present in every frame of the video sequence. From Table 1 (top), the traditional level set approach only succeeded in segmenting vessels in 33 of 38 sequences, while the LSTM-based methods successfully segmented vessels in all sequences. The proposed USVS-Net matched the expert annotations with the highest DSC, and lowest HD and MAD errors among all baselines. We estimated the statistical significance of our results using paired t-tests for every baseline, and determined that our results were statistically significant ($p < 0.05$) for all metrics except DFPD. Our largest HD error of 0.14 mm was $\sim$15$\times$ lower than the largest observed vessel diameter of 2.17 mm. Similarly, the average HD error was $\sim$10$\times$ lower than the smallest observed vessel diameter of 1.1 mm. Although our method slightly over-segmented the boundaries (outer adventitia) as evidenced

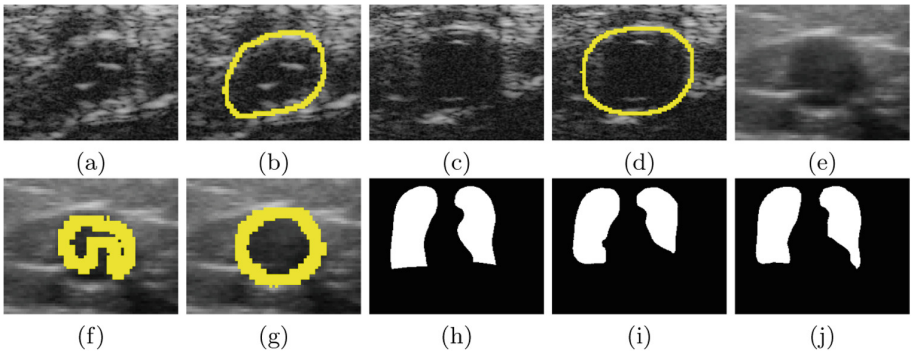

Fig. 3. (a) Frame 152 in a UHFUS sequence showing a completely contracted vessel, and (b) its associated segmentation; (c) Next frame 153 in the same sequence showing a patent vessel, and (d) its segmentation; (e) Zoomed view of a HFUS B-scan (gain set to maximum); (f) Segmentation by the CFCM34 [11] and (g) our segmentation result; (h) Ground truth lung segmentation from the CXR dataset; (i) Result from CFCM34, and (j) our result (note the improved segmentation due to better contextual information).

by the highest DFPD score, the low clinically relevant measures of HD and MAD were acceptable. Our primary intention for the USVS-Net was to segment vessels in UHFUS sequences, and through our results, we satisfactorily hit our target of sub-mm vessel localization in UHFUS sequences presenting with increased speckle, and large vessel motion.

HFUS Results. As seen in Table 1 (bottom), the performance of the CFCM34 and the USVS-Net is comparable. The USVS-Net edges out the CFCM34 with a higher DSC score, along with lower HD, MAD, and DFND errors, and a higher recall rate. We postulate that this is due to lower speckle and clearer contrast along the vessel boundaries. Again, we conducted paired t-tests to assess the statistical significance of our results, and report that the results were statistically significant ($p < 0.05$) for all metrics except DFPD and Precision. The largest HD error of 0.45 mm was $\sim$9.5$\times$ smaller than the largest observed vessel diameter of 4.35 mm, while the average HD error was $\sim$8$\times$ lower in contrast to the smallest vessel diameter of 2.9 mm. We note that the CFCM34 can be a useful alternative for clinical use in HFUS images, for which CFCM34 and USVS-Net could both be run (and results compared) for improved segmentation.

Chest X-Ray Results. To show the broader applicability of our approach to other biomedical imaging modalities, we took our network designed for UHFUS vessel segmentation, retrained, and validated it on CXR images. As seen in Table 2, the errors between the CFCM34 and the USVS-Net are comparable. CFCM34 has a higher DSC, lower DFND, and a higher recall rate, while we achieve slightly lower HD and MAD errors, lower DFPD and higher precision. As seen in Fig. 3, the dilated convolutions in the USVS-Net provide the utility of incorporating regions excluded by the CFCM34 in the final segmentation. The

increased contextual information available at the deepest layers of the network allowed it to segment the lung regions better.

Performance. The network training and testing was performed using Tensorflow on a desktop using a 3.5 GHz Intel i7 processor, 16 GB DDR3 RAM, and a NVIDIA Titan Xp GPU. The DecLSTM had 32.65 million parameters and a runtime of 5.83 s (58.3 ms per B-scan in 100 B-scan sequence). The CFCM34 had 49.16 million parameters and a runtime of 8.45 s (84.5 ms per B-scan in 100 B-scan sequence). The USVS-Net had 64.34 million parameters and a runtime of 9.95 s (99.5 ms per B-scan in 100 B-scan sequence). The level set method had a runtime of 2.03 s (20.31 ms per B-scan in 100 B-scan sequence), but yielded less accurate segmentations in contrast to the deep learning approaches. Testing was done with an ensemble of two models with the lowest validation loss.

5 Conclusion and Future Work

In this paper, we proposed a novel architecture called the USVS-Net that segmented transverse vessel cross-sections in challenging UHFUS and HFUS sequences. The performance of the USVS-Net surpasses current state-of-the-art ConvLSTM-based architectures on UHFUS sequences presenting with highly deformable vessels, but the CFCM34 is also a viable clinical alternative to USVS-Net for HFUS video segmentation. We have also shown broader applicability of our approach to a Chest X-Ray dataset. To the best of our knowledge, this is the first work targeting the segmentation of rapidly deforming vessels in UHFUS and HFUS sequences. In the future, we plan to embed a level set-based framework in a ConvLSTM-based architecture.

Acknowledgements. These awards helped us in gathering data and designing initial algorithms: NIH 1R01EY021641, DOD awards W81XWH-14-1-0371 and W81XWH-14-1-0370, NVIDIA Corporation GPU donations, Carnegie Mellon Center for Machine Learning in Health (CMLH). Patent pending, US 62/860,392.

References

1. Mohler III, E.R., et al.: High frequency ultrasound for evaluation of intimal thickness. J. Am. Soc. Echocardiogr. **22**(10), 1129–1133 (2009)
2. Gorantla, V., et al.: Acute and chronic rejection in upper extremity transplantation: what have we learned? Hand Clin. **27**(4), 481–493 (2011)
3. Menchon-Lara, R.M., et al.: Fully automatic segmentation of ultrasound common carotid artery images based on machine learning. Neurocomputing **151**(1), 161–167 (2015)
4. Shin, J.Y., et al.: Automating carotid intima-media thickness video interpretation with convolutional neural networks. In: CVPR, pp. 2526–2535 (2016)
5. Chaniot, J., et al.: Vessel segmentation in high-frequency 2D/3D ultrasound images. In: IEEE International Ultrasonics Symposium, pp. 1–4 (2016)

6. Mathai, T.S., Jin, L., Gorantla, V., Galeotti, J.: Fast vessel segmentation and tracking in ultra high-frequency ultrasound images. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11073, pp. 746–754. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00937-3_85
7. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
8. He, K., et al.: Deep residual learning for image recognition. In: IEEE CVPR, pp. 770–778 (2016)
9. Koltun, V., et al.: Multi-scale context aggregation by dilated convolutions. In: ICLR (2016)
10. Arbelle, S., et al.: Microscopy cell segmentation via convolutional LSTM networks. In: IEEE ISBI, pp. 1008–1012 (2019)
11. Milletari, F., Rieke, N., Baust, M., Esposito, M., Navab, N.: CFCM: segmentation via coarse to fine context memory. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11073, pp. 667–674. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00937-3_76
12. Gao, Y., et al.: Fully convolutional structured LSTM networks for joint 4D medical image segmentation. In: IEEE ISBI, pp. 1104–1108 (2018)
13. Zhang, D., et al.: A multi-level convolutional LSTM model for the segmentation of left ventricle myocardium in infarcted porcine cine MR images. In: IEEE ISBI, pp. 470–473 (2018)
14. Zhao, C., et al.: Predicting tongue motion in unlabeled ultrasound videos using convolutional LSTM neural network. In: IEEE ICASSP, pp. 5926–5930 (2019)
15. Basty, N., Grau, V.: Super resolution of cardiac cine MRI sequences using deep learning. In: Stoyanov, D., et al. (eds.) RAMBO/BIA/TIA -2018. LNCS, vol. 11040, pp. 23–31. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00946-5_3
16. Apostolopoulos, S., De Zanet, S., Ciller, C., Wolf, S., Sznitman, R.: Pathological OCT retinal layer segmentation using branch residual U-shape networks. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10435, pp. 294–301. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66179-7_34
17. Mathai, T.S., et al.: Learning to segment corneal tissue interfaces in OCT images. In: IEEE ISBI, pp. 1432–1436 (2019)
18. Jaeger, S., et al.: Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. Quant. Imaging Med Surg. **4**(6), 475–477 (2014)
19. Kingma, D., et al.: Adam: a method for stochastic optimization. In: ICLR (2015)