

Ana Estrada Lugo*, Niclas Bockelmann and Felix von Haxthausen

Sequential U-Net Architecture for Automatic Femoral Artery Segmentation in Ultrasound Images

Abstract: This work compares three different approaches to automatically segment the femoral artery from 2D ultrasound images. Two of the architectures follow a sequential structure, where each ultrasound image is considered a slice of the whole vessel volume, and its previous segmentation result will be part of the input, thus leading to a spatial prior. The Dice score on test data show a better performance on the baseline U-Net (0.819) compared to the sequential U-Net approaches (0.633, 0.725) for the femoral artery segmentation. This could be attributed to the misalignment of the slices being used in those networks. A possible improvement could be assumed in the implementation of a spatially calibrated and tracked ultrasound probe. Overall, these results indicate promising approaches for an automatic segmentation of the femoral artery using 2D ultrasound data.

Keywords: Automatic Segmentation, Ultrasound, Femoral Artery, Spatial Prior.

<https://doi.org/10.1515/cdbme-2021-1034>

1 Introduction

Vascular diseases such as the peripheral arterial disease (PAD) are one of the most frequent causes of death in western nations [1]. In terms of diagnosis and treatment of PAD, these often require the administration of contrast agents and continuous X-Ray radiation during the procedures. This leads to potential kidney damage for the patient caused by the nephrotoxic contrast agents and carcinogenic radiation exposure for patients and physicians. For this reason, ultrasound (US) is an

excellent option to guide endovascular procedures, due to its low cost, real-time capabilities and its low risk associated with the absence of radiation. To further improve outcomes during these procedures, automated analysis of US images is desired. A precise automatic segmentation of the femoral artery within the US images is a first step towards this direction. In recent years, the gold standard in research for medical image segmentation relies on deep learning based approaches, especially the U-Net architecture [2]. With 2D US images of the femoral artery acquired consecutively over time using a handheld US probe, a stacked but not aligned volume can be created since no spatial tracking of the probe is conducted. In this scheme, a native 3D deep learning approach could be impeded. However, the position of regions of interest for segmentation should not change vastly from slice to slice. In this paper, we try to leverage this sequential fashion as prior spatial information by using the segmentation results of the previous slice as additional input to different variations of a U-Net while iterating through the stacked volume.

2 Materials and methods

2.1 Data acquisition

Data acquisition was performed by an experienced sonographer with a 2D linear US probe (L12-3, Philips Healthcare) along with a US station (EPIQ7, Philips Healthcare) and images were transferred to computer using a network protocol provided by the manufacturer. The complete dataset consisted of US images along the femoral artery from five healthy volunteers between ages of 27 and 31. The scan for each one of the subjects started below the inguinal ligament following a superior-to-inferior movement and ending at the adductor hiatus. This protocol was repeated three times for each proband, resulting in a total of 10,557 2D US images (Table 1) with a size of 512 x 323 pixels and an intensity range of (0,255). Binary ground truth masks consisting of two

*Corresponding author: Ana Estrada Lugo: University of Lübeck, Lübeck, Germany, E-mail: ana.estradalugo@student.uni-luebeck.de

Niclas Bockelmann and Felix von Haxthausen: Institute for Robotics and Cognitive Systems, University of Lübeck, Lübeck, Germany.

classes, the background and the femoral artery (0,1), were manually segmented with the guidance of an experienced sonographer. Once every image had its corresponding mask, training, validation and testing datasets were defined, with an approximate split of 60 %, 20 % and 20 %, respectively.

Table 1: Summary of images per proband.

	Run 1	Run 2	Run 3	Total
Proband 1	733	767	669	2,169
Proband 2	1098	938	917	2,953
Proband 3	522	530	548	1,600
Proband 4	704	756	683	2,143
Proband 5	566	551	575	1,692
<i>Total</i>				10,557

2.2 Segmentation Networks

2.2.1 U-Net Baseline

To allow a comparison of the approaches used, the well-known U-Net architecture [2] is used as a baseline and referred to by *U-Net Baseline*. Conformed by five levels, each block in the encoder path has a two 3x3 convolution, each followed by an activation function; in this case, a rectified linear unit (ReLU) and a 2x2 max pooling with a downsampling stride of 2. In the decoder path, each block consists of a 2x2 upconvolution, a concatenation with the output block having the same shape in the encoder path and a double 3x3 convolution, each followed by a ReLU.

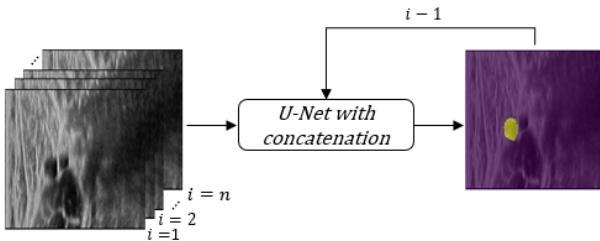


Figure 1: U-Net sequential architecture proposals iterate through the volume slice by slice, using the segmentation of the previous slice with the current US image to form an input to the network.

2.2.2 Sequential Architectures

The following approaches are considered sequential U-Net architectures. In these cases, the 2D US images were stacked to form a continuous representation of the femoral artery. The sequential U-Nets iterate through the volume slice by slice, each time using the segmentation of the previous slice with the current US image to form an input to the network, as shown in Figure 1. This way, prior spatial information is incorporated into the network to potentially improve segmentation results.

Since commonly sequential U-Net approaches work on native 3D data and are more complex [3], two simplified methods that could reflect the image acquisition were investigated: *U-Net Concatenated Input* and *U-Net Concatenated Bottleneck*. The former concatenates the previous slice segmentation and the current US image forming a two-channel input to the U-Net. In the latter, the segmentation and US image are processed by their own encoder and the concatenation occurs at the bottleneck of the U-Net (Figure 2). This aims to independently extract the spatial information from the previous segmentation without compromising the extraction of the most characteristic features of the US image.

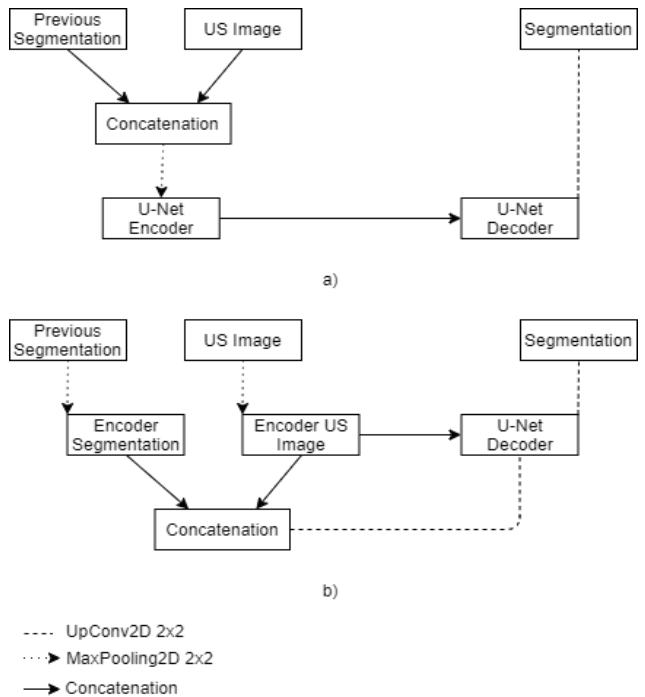


Figure 2: Schematics of the proposed sequential approaches a) *U-Net Concatenated Input*, b) *U-Net Concatenated Bottleneck*.

2.3 Training

The networks were coded in Python with PyTorch as the preference deep learning framework and trained on NVIDIA DGX A100. Images from our datasets were down-sampled from 512x323 to 128x128, in order to decrease memory and computational requirements. The networks were trained using a Binary cross entropy loss with a starting learning rate (LR) of 0.0001. If our validation loss did not decrease after four epochs, the LR was multiplied by a factor of 0.1. Each network training consisted of 30 epochs with a batch size of 25. Probands 1, 3 and 4 sets were used to train the network. Validation and test sets were defined by probands 2 and 5, respectively and were solely used for evaluating the performance of the networks.

3 Results and discussion

The segmentations of our different approaches were assessed and compared to our manual segmented ground truth by calculating the Dice score. The resulting Dice scores are 0.633, 0.725 and 0.819 for the *U-Net Concatenated Input*, *U-Net Concatenated Bottleneck* and *U-Net Baseline*, respectively. Comparing the results between our approaches quantitatively (Table 2) as well as qualitatively (Figure 3) the *U-Net Baseline* achieved the highest performance with a Dice score of 0.819. Surprisingly, the Dice score for the sequential architectures was not as high as expected.

Table 2: Comparison of Dice scores between the implemented approaches.

Approach	Dice Score
<i>U-Net Baseline</i>	0.819
<i>U-Net Concatenated Input</i>	0.633
<i>U-Net Concatenated Bottleneck</i>	0.725

The reason might be that the stacked images that represent the 3D volume, were not spatially aligned, causing some notable discrepancies on the location of the vessel from one slice to another, making it harder to our network to learn the vessel features. Since we consider the stacked slices as a volume, it is expected that one image is the continuation of the other. In the *U-Net Concatenated Input*, this misalignment is considered as part of the input and by consequence carried on. In the *U-Net Concatenated Bottleneck*, this misalignment error does not impact the results as much due to the separate feature

extraction. The vessel segmentation is not compromised because the US image has a greater impact in the prediction.

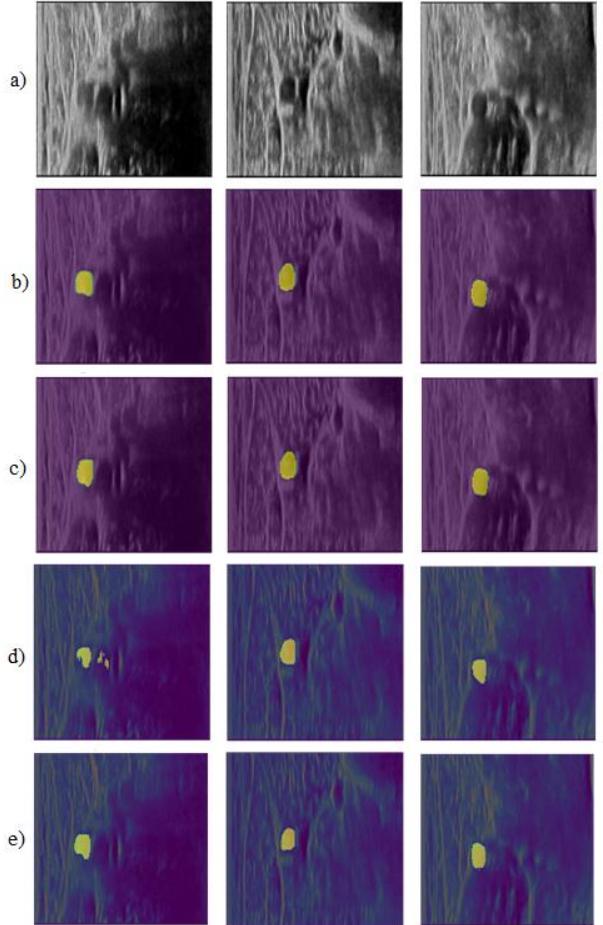


Figure 3: Comparison between a) original US Image b) ground truth, c) *U-Net Baseline*, d) *U-Net Concatenated Input*, e) *U-Net Concatenated Bottleneck* for three different sample images.

It is also important to remark, that due to the surrounding tissue and its anatomy, the femoral artery varies significantly from person to person. Considering this, our current number of probands for training and validation could be increased, along with a greater age range between the volunteers. The latter improvement would let us assess the network accuracy taking in consideration noteworthy vascular conditions associated with age, such as vascular calcification.

4 Conclusion

For the used dataset, the *U-Net Baseline* was the most accurate approach to segment the femoral artery due to its highest Dice score. Although our ideas of the *U-Net Concatenated Input*

and *U-Net Concatenated Bottleneck* are promising, the evaluation results can be improved. In order to do this, a spatially calibrated and tracked ultrasound probe must be implemented for the US scan. Another idea to improve the misalignment between slices would be the integration of a spatial transformer that allows us to align an image prior the segmentation. Overall, an algorithm for segmentation of the femoral artery in 2D ultrasound images was implemented that might support physicians during vascular diagnostics and interventions.

Author Statement

Research funding: Parts of this work were supported by the German Federal Ministry of Education and Research (BMBF, project KI-LAB-Lübeck, grant number: 01IS19069 and project Nav EVAR, grant number: 13GW0228).

Conflict of interest: Author declare that there were no conflicts of interest.

Informed consent: Informed consent has been obtained from all individuals included in the presented study.

Ethical approval: The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration, and has been approved by the authors' institutional review board or equivalent committee.

References

- [1] S. Mendis, P. Puska, and B. Norrving. Global atlas on cardiovascular disease prevention and control. WHO, 2011. ISBN : 978 92 4 156437 3.
- [2] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention (MICCAI), vol. 9351 of LNCS, pages 234–241. Springer, 2015.
- [3] A. Novikov, et al. "Deep sequential segmentation of organs in volumetric medical scans." IEEE transactions on medical imaging 38.5 (2018): 1207-1215
- [4] S. Liu, Y. Wang, X. Yang, B. Lei, L. Liu, S. Xiang Li, D. Ni and T. Wang Deep learning in medical ultrasound analysis: A review. Eng J 2019;5(2):261–275.
- [5] L.J. Brattain, B.A. Telfer, M. Dhyani, J.R. Grajo, A.E. Samir Machine learning for medical ultrasound: status, methods, and future opportunities. Abdom Radiol 2018;43:786–799. Available: <https://doi.org/10.1007/s00261-018-1517-0>
- [6] L.A. Groves, B. VanBerlo, N. Veinberg , A. Alboog, T.M. Peters, and E.C.S. Chen, Automatic segmentation of the carotid artery and internal jugular vein from 2D ultrasound images for 3D vascular reconstruction. Int J CARS 2020;15(11):1835–1846.. Available: <https://doi.org/10.1007/s11548-020-02248-2>
- [7] B.R. Thomson, J. Nijkamp, O. Ivashchenko, F. van der Heijden, P. Fischer, J.N. Smit, N.F.M. Kok, K.F.D. Kuhlmann, T.J.M. Ruers and M. Fusaglia, Hepatic vessel segmentation using a reduced filter 3D U-Net in ultrasound imaging. In: International Conference on Medical Imaging with Deep Learning 2019, arXiv:1907.12109v1, 2019.
- [8] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.C. Fillion-Robin J, S. Pujol, et. al 3D Slicer as an Image Computing Platform for the Quantitative Imaging Network. Magn Reson Imaging 2012;30: 1323-1341.