

Force Sensing Guided Artery-Vein Segmentation via Sequential Ultrasound Images

Yimeng Geng^{1,2}, Gaofeng Meng^{1,2,3}, Mingcong Chen^{3,4}, Guanglin Cao^{1,2,3},
Mingyang Zhao³, Jianbo Zhao^{1,2}, and Hongbin Liu^{1,3,5}(✉)

¹ State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China

² School of Artificial Intelligence, University of Chinese Academy of Sciences

³ Centre for Artificial Intelligence and Robotics, HK Institute of Science & Innovation, Chinese Academy of Sciences, Hong Kong SAR

⁴ City University of Hong Kong, Hong Kong SAR

⁵ School of Biomedical Engineering and Imaging Sciences, King's College London, London, UK

liuhongbin@ia.ac.cn

Abstract. Accurate identification of arteries and veins in ultrasound images is crucial for vascular examinations and interventions in robotics-assisted surgeries. However, current methods for ultrasound vessel segmentation face challenges in distinguishing between arteries and veins due to their morphological similarities. To address this challenge, this study introduces a novel force sensing guided segmentation approach to enhance artery-vein segmentation accuracy by leveraging their distinct deformability. Our proposed method utilizes force magnitude to identify key frames with the most significant vascular deformation in a sequence of ultrasound images. These key frames are then integrated with the current frame through attention mechanisms, with weights assigned in accordance with force magnitude. Our proposed force sensing guided framework can be seamlessly integrated into various segmentation networks and achieves significant performance improvements in multiple U-shaped networks such as U-Net, Swin-unet and Transunet. Furthermore, we contribute the first multimodal ultrasound artery-vein segmentation dataset, Mus-V, which encompasses both force and image data simultaneously. The dataset comprises 3114 ultrasound images of carotid and femoral vessels extracted from 105 videos, with corresponding force data recorded by the force sensor mounted on the US probe. The code and dataset can be available at <https://www.kaggle.com/datasets/among22/multimodal-ultrasound-vascular-segmentation>.

Keywords: Force fusion · Sequential ultrasound images · Artery-Vein segmentation.

1 Introduction

Vascular intervention surgery (VIS) is a minimally invasive surgical technique involving catheter or device insertion into blood vessels [6]. This real-time and

low-risk surgery plays a crucial role in treating various diseases such as coronary artery diseases [22] and peripheral vascular diseases [2]. Accurate segmentation of arteries and veins is vital in VIS, enabling clinicians to precisely identify target vessels for intervention and minimizing the risk of inadvertent injury to adjacent structures [19]. As a non-invasive, radiation-free and real-time modality, ultrasound imaging is widely used in VIS [7]. Therefore, artery-vein segmentation in ultrasound images is crucial for enhancing the safety and efficacy of VIS.

Ultrasound vessel segmentation, a subset of medical image segmentation, is developing rapidly, primarily because of the rapid evolution of deep learning techniques [24]. In particular, a series of U-shaped networks have demonstrated outstanding performance [21,12,3,4,5] and been widely applied in ultrasound vessel segmentation. Xie et al. [25] directly utilized UNet [21] to segment carotid artery lumen. Blanco et al. [1] presented a modified UNet to exploit adjacency information in spatially neighboring ultrasound images for intravascular ultrasound images segmentation. Groves et al. [8] employed both Mask-RCNN [10] and UNet to segment carotid artery (CA) and internal jugular vein (IJV).

However, current ultrasound vessel segmentation networks either solely focus on morphological features within a single image [8], making it challenging to differentiate arteries and veins that are morphologically similar, or only segment vessels without distinguishing between arteries and veins [18,25,1]. In clinical practice, doctors typically differentiate arteries from veins by palpating the patient’s skin and observing the degree of vascular deformation in ultrasound images [14]. During this process, arteries exhibit minimal deformation under compression, while veins undergo significant deformation and can even be completely occluded [14]. Inspired by this method, we propose a novel force sensing guided approach for ultrasound artery-vein segmentation. This method aims to precisely segment arteries and veins using tactile force data, especially in scenarios involving robotic ultrasound systems.

Robotic Ultrasound Systems (RUS) can improve the accuracy and safety of VIS through accurate image guidance and robotic arm operation, and provide better operating experience for doctors [9,16,17,15]. We set up a simple RUS in our experiments as shown in Fig. 1(a). The sensor mounted on the US probe can collect real-time force data, which is employed as a prior to identify key frames that represent maximum deformation. These key frames effectively capture and store vascular deformation information, aiding in distinguishing the distinct deformability of arteries and veins as shown in Fig. 1(b). Then we can retrieve the stored information by computing the correlation between key frames and the current frame to get enhanced features. To validate the effectiveness of the proposed method, we also contribute a benchmark dataset containing both ultrasound images and force data called Mus-V.

The main contributions are: (1) We present a novel force sensing guided segmentation approach that utilizes force data to assist in artery-vein segmentation, leading to enhanced segmentation accuracy. (2) Our method is versatile and demonstrates substantial performance improvement across multiple segmentation networks. (3) We contribute the first publicly available ultrasound

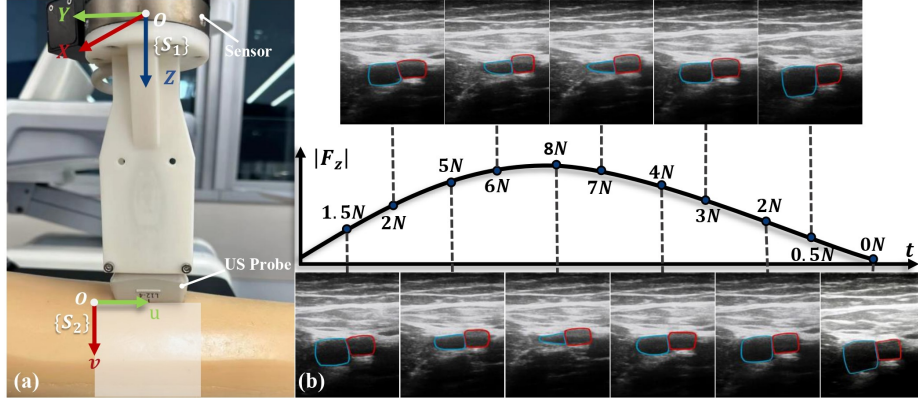


Fig. 1. (a) Hardware setup: a robotic arm, a force/torque sensor and an ultrasound probe. $\{S_1\}$ and $\{S_2\}$ respectively stand for coordinate systems of the sensor and the US image (b) An illustration of data acquisition. The force along the z-axis, $|F_z|$, initially increases and then decreases over time t . Eleven points are selected along the curve, with corresponding force magnitudes and ultrasound scans. Veins are labelled in blue, while arteries are labelled in red. Notably, veins exhibit higher deformability compared to arteries.

artery-vein segmentation dataset with force data, providing a baseline for future researches.

2 Ultrasound Vascular Dataset with Force data

We introduce Mus-V: the first ultrasound artery-vein segmentation dataset containing both force and image data. The dataset collects ultrasound videos of carotid and femoral blood vessels, with the arteries and veins labelled separately for vascular analysis and identification. In addition, the dataset also records the force data collected by the force sensor mounted on the ultrasound probe during the collection of ultrasound videos. The structure of the robotic ultrasound system consists of an ultrasound (US) probe, a robotic arm carrying the US probe and a force/torque sensor mounted on the US probe (see Fig. 1(a)). In each data collection session, an ultrasound video is recorded in conjunction with corresponding force data. The contact force between the US probe and the skin is systematically adjusted to vary gradually. Meanwhile, the deformation of blood vessels is also strengthened or reduced as Fig. 1(b) shows. Subsequent to acquisition, the ultrasound videos and force data undergo aligning, cleaning, downsampling, and annotation.

A total of 3114 ultrasound images were sampled from 105 ultrasound videos collected from 11 healthy volunteers. The training and validation datasets comprise 2203 and 911 images, respectively. The force data is organized in a matrix format, with each row corresponds to an ultrasound image. Each column, from

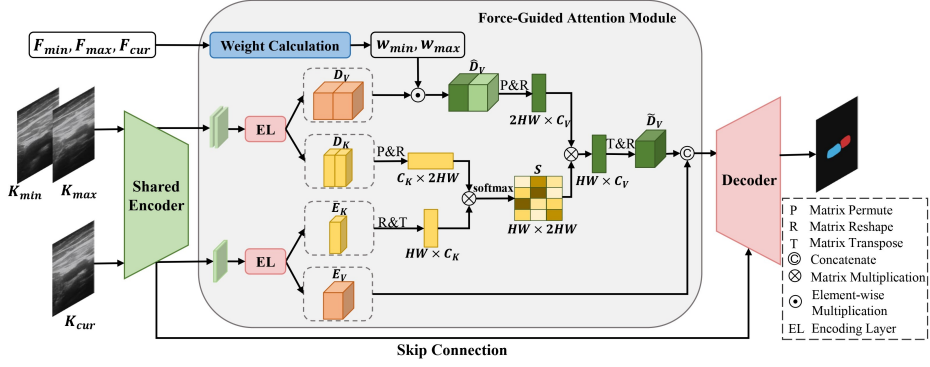


Fig. 2. Illustration of our proposed Force Sensing Guided Segmentation Framework. Selected key frames and the current frame are fed into a shared encoder to extract features. The force-guided attention module captures relation between embeddings of key frames and current frames, generating an enhanced embedding $\tilde{D}_V$. The value embedding of current frame E_V is concatenated with $\tilde{D}_V$ to get final segmentation result through the decoder. Skip connection is applied between the encoder and the decoder.

the first to the last, represents the force components in the X, Y, and Z directions, as well as the moment components in the X, Y, and Z directions, respectively.

3 Method

3.1 Overview

The most obvious difference between arteries and veins in US images lies in the extent of deformation when contact force varies. Therefore, capturing the deformation of vessels over time is crucial for accurate segmentation, which requires to compare different frames within the same video. Deformation magnitude correlates directly with force magnitude, thereby we utilize contact force data to identify key frames for reference. The selected key frames exhibit the most substantial degree of vascular deformation. Specific method for selecting key frames is described in Sec. 3.2. These key frames aid in differentiation between arteries during the subsequent segmentation process.

Image segmentation networks based on the UNet architecture typically consist of an encoder that downsamples features and a decoder that upsamples features. Skip connections are utilized between different stages of the encoder and the decoder to fully leverage information at various resolutions. In the proposed fusion approach, key frames and the current segmentation frame are firstly fed into a shared encoder for feature extraction, with the final stage of the encoder outputting features with low resolution and high channel numbers. Subsequently, these features are processed by our proposed force-guided attention module. The

fused features then enter the decoder to restore the original resolution and obtain the mask. This design allows for the encoder and decoder to be replaced with any U-shaped network, thereby enhancing the versatility of the method.

3.2 Force-Driven Keyframe Selection

In the Robotic Ultrasound System, magnitude of contact force varies continuously, and changes between two consecutive frames may not be particularly evident. The magnitude of deformation is directly positively correlated with magnitude of force along the z-axis. Therefore, we utilize force along the z-axis to select key frames that contain the most significant deformation information. The target ultrasound image for segmentation is denoted as K_{cur} . K_{cur} is one frame from an ultrasound video with a duration of T and $\mathcal{K} = \{K_0, \dots, K_T\}$ indicates the set of all the frames in the same video. Correspondingly, $\mathcal{F} = \{F_0, \dots, F_T\}$ represents the absolute value of force along the z-axis collected with $\mathcal{K}$.

As illustrated in Eq. 1, we select the image frame corresponding to the maximum force, denoted as $K_{\max}$, and the frame corresponding to the minimum compression force, denoted as $K_{\min}$. These frames, $K_{\min}$ and $K_{\max}$ are referred to as key frames.

$$\begin{cases} K_{\min} = K_{\arg \min_t F_t} \\ K_{\max} = K_{\arg \max_t F_t} \end{cases} \quad (1)$$

Therefore, $K_{\min}$ corresponds to the most dilated form of the blood vessel, while $K_{\max}$ corresponds to the most compressed form.

3.3 Force-Guided Attention Module

The output of the encoder has high-dimensional channels which are computationally expensive. Inspired by [23], we firstly apply an encoding layer to the output for channel reduction. As illustrated in Fig. 2, the encoding layer including a 1x1 convolution and a 3x3 convolution will generate a key feature and a value feature. We respectively stack the key features and value features related to key frames $K_{\min}$ and $K_{\max}$, and obtain a 4-dimension key feature $D_K \in R^{N \times C_K \times H \times W}$ and a 4-dimension value feature $D_V \in R^{N \times C_V \times H \times W}$, where $N = 2$. C_K and C_V represent the number of channels for key and value features respectively. E_K and E_V are the corresponding key and value feature related to K_{cur} .

For frames in proximity to a key frame, the reference value of the key frame is negligible. For example, if $K_{cur} = K_{\max}$, then $K_{\max}$ will fail to provide any meaningful information. Therefore, we propose a force-based dynamic weighting strategy, assigning a higher weight to the key frame with lower similarity to the current frame. Detailed calculation of weights is illustrated in Sec. 3.4.

After obtaining force-based weights $w_{\min}$ and $w_{\max}$, we perform element-wise multiplication between the weights and D_V to get weighted values $\hat{D}_V$. Then, E_K is reshaped and transposed to $E_K \in R^{C_K \times M}$, $M = H \times W$. Meanwhile, D_K and $\hat{D}_V$ are permuted and reshaped to $\hat{D}_K \in R^{C_K \times NM}$ and $\hat{D}_V \in R^{C_V \times NM}$.

Subsequently, we conduct matrix multiplication between D_K and E_K , and utilize a softmax layer to compute the attention map $S \in R^{N \times M}$,

$$S_{ij} = \frac{\exp(D_K^i \cdot E_K^j)}{\sum_{j=1}^E \exp(D_K^i \cdot E_K^j)} \quad (2)$$

A larger value of S_{ij} indicates that the i^{th} position in D_K has a greater influence on the j^{th} position in E_K Memory. After obtaining the attention map S , we multiply S and $\hat{D}_V$ to get the force-guided enhanced feature $\tilde{D}_V$. Finally, we combine $\tilde{D}_V$ with the value feature of the current frame. For simplicity, we utilize feature concatenation to get the aggregated multimodal feature f as follows:

$$f = \text{concat}(\tilde{D}_V, E_V) \quad (3)$$

3.4 Force-Based Dynamic Weights

Adhering the principle that a higher weight is assigned to the key frame with lower similarity to the current frame, we design a weight calculation formula based on force magnitudes. The dynamic weights are directly proportional to the difference in force magnitudes and lies within the range of $[0, 1]$,

$$w_{\min} = \frac{F_{cur} - F_{\min}}{F_{\max} - F_{\min}} \quad (4)$$

$$w_{\max} = \frac{F_{\max} - F_{cur}}{F_{\max} - F_{\min}} \quad (5)$$

where $w_{\min}$ and $w_{\max}$ are force-based dynamic weights for $K_{\min}$ and $K_{\max}$. For each input, a new pair of force-based weights is computed and applied. F_{cur} , $F_{\min}$, $F_{\max}$ are force respectively corresponding to K_{cur} , $K_{\min}$ and $K_{\max}$.

4 Experiments

4.1 Implementation and Setup

Our experiments are implemented based on PyTorch on a RTX 4090 GPU. We employ the ReduceLROnPlateau [13] policy to automatically adjust the learning rate and utilize RMSprop [11] as the optimizer. Momentum, weight decay and batch size are set to 0.9, 1e-8 and 4 for all experiments, respectively. We set the maximum number of epochs as 120 for experiments. For data augmentation, we apply random horizontal and vertical flip, random intensity transformation, random translation, random rotation for input images and key frames for all experiments. All images are resized to 256x256. We also utilize Torchio [20] to add random bias fields and random noise to input images and key frames. To address the issue of class imbalance, we employed weighted cross-entropy loss with weights of 1, 30.7, and 23.1 for background, arteries and veins, respectively.

Table 1. Comparison results of different methods on the Mus-V dataset.

Method	mIoU(%) $\uparrow$	Dice Coefficient $\uparrow$	FLOPs(M) $\downarrow$
Swin-unet[4]	68.36	0.7968	0.774×10^4
UNet[21]	69.23	0.8140	4.019×10^4
Transunet[5]	74.38	0.8499	3.225×10^4
FG-Swin-unet	69.56	0.8033	1.855×10^4
FG-UNet	72.02	0.8280	8.016×10^4
FG-Transunet	75.63	0.8547	8.867×10^4

Table 2. Ablation study results. The w/o KFS represents that force-driven key frame selection is removed and two frames temporally before the current frame are selected as key frames. The w/o FBW means that force-based dynamic weights are removed.

Method	mIoU(%) $\uparrow$	Dice Coefficient $\uparrow$
UNet[21]	69.23	0.8140
FG-UNet w/o KFS&FBW	70.22	0.8205
FG-UNet w/o FBW	71.36	0.8244
FG-UNet	72.02	0.8280

4.2 Comparison with State-of-the-Arts

We conduct comprehensive experiments by applying force sensing guidance to three different U-shaped networks, namely UNet, Swin-unet and Transunet. The force-guided attention module is inserted as a “neck” between the encoder and the decoder of three basic networks. For Transunet, we choose to use the R50-Vit-B16 version. Table 1 displays the performance metrics and FLOPs of all the methods. The network applied with force sensing guidance is named with the prefix “FG” followed by the network name. In conclusion, Force Sensing Guided TranUnet (FG-Transunet) achieves the highest performance, attaining 75.63% mIoU and 0.8547 dice coefficient. Additionally, the application of force sensing guidance results in performance enhancements of 1.2%, 2.79% and 1.25% mIoU in Swin-UNet, UNet and Transunet, respectively. Qualitative results are presented in Fig. 3

4.3 Ablation Studies

We conduct several ablation experiments to validate the effectiveness of all the components of the proposed approach, utilizing UNet [21] as our baseline. To save computational resources and training time, we also exploit Force Sensing Guided UNet (FG-UNet) as the base model for all ablation experiments.

To validate the effectiveness of the force-based dynamic weights, we conducted experiments by removing the weights multiplication. The experimental

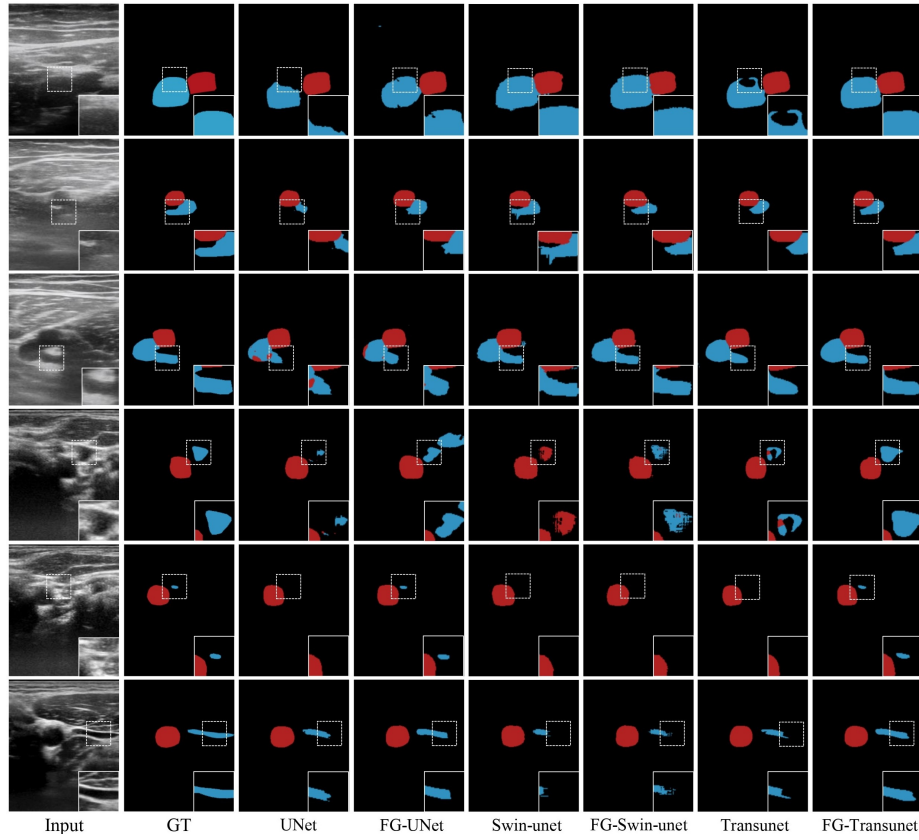


Fig. 3. Qualitative results. The results demonstrate that integrating force sensing guidance effectively enhances the performance of three baseline networks. Specifically, our method improves the model’s ability to differentiate between arteries (red) and veins (blue), and ability to segment small veins.

results are shown in Table 2. After removing the force-based weights, the mIoU decreases by 0.66%.

To investigate the effect of the proposed force-driven key frame selection strategy, we conduct an experiment that simply selecting two frames preceding the current frame in time as key frames. The results are displayed in Table 2. The performance has decreased significantly in the aspect of both mIoU and dice coefficient after removing the key frame selection. The mIoU decreases by 1.14% and the dice coefficient decreases by 0.0039.

5 Conclusion

In this study, we introduced a pioneering force sensing guided approach for ultrasound artery-vein segmentation. Our method includes innovative compo-

nents such as force-driven key frames selection, force-guided attention module and force-based dynamic weights. Extensive experiments demonstrated that our proposed approach led to notable performance enhancements across multiple U-shaped networks. Furthermore, we presented Mus-V, the first publicly available ultrasound artery-vein segmentation dataset including images and force. Future directions include further optimizing the performance by exploring additional methodologies, as well as expanding the capacity of Mus-V.

Acknowledgments. This work was supported partially by the National Natural Science Foundation of China (Grant No.62376267) and the InnoHK programme.

Disclosure of Interests. The authors have no competing interests.

References

1. Blanco, P.J., Ziemer, P.G., Bulant, C.A., Ueki, Y., Bass, R., Räber, L., Lemos, P.A., García-García, H.M.: Fully automated lumen and vessel contour segmentation in intravascular ultrasound datasets. *Medical image analysis* **75**, 102262 (2022) [2](#)
2. Bradbury, A.W., Ruckley, C., Fowkes, F., Forbes, J., Gillespie, I., Adam, D., et al.: Bypass versus angioplasty in severe ischaemia of the leg (basil): multicentre, randomised controlled trial. *Lancet* **366**(9501), 1925–1934 (2005) [2](#)
3. Cai, S., Tian, Y., Lui, H., Zeng, H., Wu, Y., Chen, G.: Dense-unet: a novel multi-photon in vivo cellular image segmentation model based on a convolutional neural network. *Quantitative imaging in medicine and surgery* **10**(6), 1275 (2020) [2](#)
4. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: *European conference on computer vision*. pp. 205–218. Springer [2](#), [7](#)
5. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.J.a.p.a.: Transunet: Transformers make strong encoders for medical image segmentation (2021) [2](#), [7](#)
6. Da, L., Zhang, D., Wang, T.: Overview of the vascular interventional robot. *The International Journal of Medical Robotics and Computer Assisted Surgery* **4**(4), 289–294 (2008) [1](#)
7. Dalili, D., Pracon, G.: Role of ultrasound in diagnostic and interventional musculoskeletal imaging. *Journal of Ultrasonography* **21**(85), 84–85 (2021) [2](#)
8. Groves, L.A., VanBerlo, B., Veinberg, N., Alboog, A., Peters, T.M., Chen, E.C.: Automatic segmentation of the carotid artery and internal jugular vein from 2d ultrasound images for 3d vascular reconstruction. *International journal of computer assisted radiology and surgery* **15**, 1835–1846 (2020) [2](#)
9. von Haxthausen, F., Böttger, S., Wulff, D., Hagenah, J., García-Vázquez, V., Ipsen, S.: Medical robotics for ultrasound imaging: Current systems and future trends. *Current Robotics Reports* **2**(1), 55–71 [2](#)
10. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: *Proceedings of the IEEE international conference on computer vision*. pp. 2961–2969 (2017) [2](#)
11. Hinton, G., et al.: Neural networks for machine learning. Coursera Lecture Slides (2012), <http://www.cs.toronto.edu/~hinton/nn2012/nn2012.html> [6](#)
12. Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., Han, X., Chen, Y.W., Wu, J.: Unet 3+: A full-scale connected unet for medical image segmentation. In: *ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. pp. 1055–1059. IEEE (2020) [2](#)

13. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International conference on machine learning. pp. 448–456. pmlr **6**
14. Komasa, N., Mihara, R., Hattori, K., Minami, T.J.B.B.J.o.A.: Evaluation of artery and vein differentiation methods using ultrasound imaging among medical students **117**(6), 832–833 (2016) **2**
15. Langsch, F., Virga, S., Esteban, J., Göbl, R., Navab, N.: Robotic ultrasound for catheter navigation in endovascular procedures. In: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 5404–5410 (2019). <https://doi.org/10.1109/IROS40897.2019.8967652> **2**
16. Li, K., Xu, Y., Meng, M.Q.H.: An overview of systems and techniques for autonomous robotic ultrasound acquisitions. IEEE Transactions on Medical Robotics and Bionics **3**(2), 510–524 (2021) **2**
17. Ma, G., Oca, S.R., Zhu, Y., Codd, P.J., Buckland, D.M.: A novel robotic system for ultrasound-guided peripheral vascular localization. In: 2021 IEEE International Conference on Robotics and Automation (ICRA). pp. 12321–12327 (2021). <https://doi.org/10.1109/ICRA48506.2021.9561924> **2**
18. Mishra, D., Chaudhury, S., Sarkar, M., Soin, A.S.: Ultrasound image segmentation: a deeply supervised network with attention to boundaries. IEEE Transactions on Biomedical Engineering **66**(6), 1637–1648 (2018) **2**
19. Müller-Wille, R., Wildgruber, M., Sadick, M., Wohlgemuth, W.A.: Vascular anomalies (part ii): interventional therapy of peripheral vascular malformations. In: RÖFo-Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren. vol. 190, pp. 927–937. © Georg Thieme Verlag KG (2018) **2**
20. Pérez-García, F., Sparks, R., Ourselin, S.J.C.M., Biomedicine, P.i.: Torchio: a python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning **208**, 106236 (2021) **6**
21. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18. pp. 234–241. Springer **2**, **7**
22. Serruys, P.W., De Jaegere, P., Kiemeneij, F., Macaya, C., Rutsch, W., Heyndrickx, G., Emanuelsson, H., Marco, J., Legrand, V., Materne, P., et al.: A comparison of balloon-expandable-stent implantation with balloon angioplasty in patients with coronary artery disease. New England Journal of Medicine **331**(8), 489–495 (1994) **2**
23. Wang, H., Wang, W., Liu, J.: Temporal memory attention for video semantic segmentation. In: 2021 IEEE International Conference on Image Processing (ICIP). pp. 2254–2258. IEEE **5**
24. Wang, Y., Ge, X., Ma, H., Qi, S., Zhang, G., Yao, Y.: Deep learning in medical ultrasound image analysis: a review. IEEE Access **9**, 54310–54324 (2021) **2**
25. Xie, M., Li, Y., Xue, Y., Huntress, L., Beckerman, W., Rahimi, S., Ady, J., Roshan, U.: Vessel lumen segmentation in carotid artery ultrasounds with the u-net convolutional neural network. In: 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 2680–2684. IEEE (2020) **2**