ORIGINAL ARTICLE

Automatic segmentation of the carotid artery and internal jugular vein from 2D ultrasound images for 3D vascular reconstruction

Leah A. Groves[1] · Blake VanBerlo[2] · Natan Veinberg[2] · Abdulrahman Alboog[3] · Terry M. Peters[4] · Elvis C. S. Chen[4]

Abstract

Purpose In the context of analyzing neck vascular morphology, this work formulates and compares Mask R-CNN and U-Net-based algorithms to automatically segment the carotid artery (CA) and internal jugular vein (IJV) from transverse neck ultrasound (US).

Methods US scans of the neck vasculature were collected to produce a dataset of 2439 images and their respective manual segmentations. Fourfold cross-validation was employed to train and evaluate Mask RCNN and U-Net models. The U-Net algorithm includes a post-processing step that selects the largest connected segmentation for each class. A Mask R-CNN-based vascular reconstruction pipeline was validated by performing a surface-to-surface distance comparison between US and CT reconstructions from the same patient.

Results The average CA and IJV Dice scores produced by the Mask R-CNN across the evaluation data from all four sets were 0.90 ± 0.08 and 0.88 ± 0.14. The average Dice scores produced by the post-processed U-Net were 0.81 ± 0.21 and 0.71 ± 0.23, for the CA and IJV, respectively. The reconstruction algorithm utilizing the Mask R-CNN was capable of producing accurate 3D reconstructions with majority of US reconstruction surface points being within 2 mm of the CT equivalent.

Conclusions On average, the Mask R-CNN produced more accurate vascular segmentations compared to U-Net. The Mask R-CNN models were used to produce 3D reconstructed vasculature with a similar accuracy to that of a manually segmented CT scan. This implementation of the Mask R-CNN network enables automatic analysis of the neck vasculature and facilitates 3D vascular reconstruction.

Keywords Deep learning · US · Surface reconstruction · Surgical guidance · Vasculature · Automatic segmentation

✉ Leah A. Groves
  lgroves6@uwo.ca

  Terry M. Peters
  tpeters@robarts.ca

  Elvis C. S. Chen
  chene@robarts.ca

1 School of Biomedical Engineering, Western University, London, Canada

2 Schulich School of Medicine and Dentistry, Western University, London, Canada

3 Department of Anesthesia and Perioperative Medicine, Western University, London, Canada

4 Department of Medical Biophysics, School of Biomedical Engineering, Robarts Research, London, Canada

Introduction

Percutaneous internal jugular vein (IJV) needle insertions are used to access the central venous system [4]. Carotid artery (CA) punctures are one of the most common and severe complications that occur during IJV cannulation [4]. Ultrasound-(US)-guided needle insertions have the potential to reduce complications by providing clinicians with a real-time cross-sectional view of the neck anatomy to visualize the relationship between the IJV and CA in 2D [9,21]. The fact that neck vasculature is extremely variable across the patient population [9,23] has motivated research efforts in the development of advanced US-based surgical navigation systems [2,10], along with the characterization of neck vasculature morphology to further assist and improve central venous cannulation (CVC) [9,23].

Specifically, US imaging has been used to analyze the effect of the anatomical relationship between the IJV and CA on CVC [9,23], and the relationship between head rotation and diameter of the vessels [16,25]. Since US produces real-time images and does not carry the risks associated with ionizing radiation, obtaining these images carries minimal risk for the patient. For these applications, anatomical structures must be segmented from the US images. While the gold standard for segmentation is often established by manual segmentation, a process that is labor-intensive and sensitive to human error [17], patient data derived from 2D US alone have limitations, as a single cross-sectional slice cannot adequately represent the entire structure. One example of a measurement that requires 3D information is the assessment of the variability of the location of the CA bifurcation, which to date has been performed using excised vessels from cadavers [15,26]. Vascular dissection is a time-consuming process that sacrifices the structural integrity and normal physiological properties found *in vivo*. Automatic segmentation of the vessels from US in 3D that reflects the patient positioning at the time of an intervention would therefore be ideal.

The degree of manual analysis required to quantify trends in vascular anatomy has prompted work such as automatic segmentation of the media-adventitia and lumen-intima boundaries of the CA from 3D US images [28], the inner lumen of the CA in a longitudinal orientation [27], and CA plaques [24]. As far as we are aware, there is no method in current literature to simultaneously and automatically delineate both the IJV and CA within a 2D transverse US image. Such a procedure would allow for the automatic analysis of the morphology and anatomical relationships of these vessels and to enable accurate reconstruction of 3D volumes of the neck vasculature without exposing the patient to radiation, removing barriers for further research on the morphology of neck vasculature. Other applications of these vascular reconstructions include, but are not limited to, real-time intra-operative use or during preoperative planning to augment guidance for CVC. Therefore, the secondary motivation of this work is the development of 3D models of the vasculature, which could be used to develop a more clinically relevant navigation system, while maintaining 3D information.

A U-Net convolutional neural network (CNN) architecture was applied developed to automatically segment regions of interest associated with the CA [24,28]. U-Net is a semantic segmentation architecture, trained to provide pixel-wise label maps [20]. Each pixel is classified as either the background or one of the foreground classes that were provided during training [20]. For certain U-Net applications, false segmentations occur due to the inability of the network to differentiate between regions that contain pixels of a specific class and regions that contain pixels with similar features to the class of interest. Two methods to compensate for this issue of false segmentation include: i) post-processing steps to retain the largest segmentation [27], or ii) cropping the input to a region of interest (ROI) that contains only the anatomy of interest [24]. The Mask R-CNN architecture provides an alternative method to segment the CA and IJV with the potential of returning fewer or no false segmentations [11]. Mask R-CNN was inspired by Faster R-CNN [11] for object detection [19] and consists of two stages. In the first, a region proposal network (RPN) determines possible bounding boxes that may contain objects of interest. In the second, two components execute in parallel, receiving the region proposals from the RPN as input. The first component, inspired by Faster R-CNN, predicts object class and bounding box localization, while the second predicts pixel segmentation for each ROI [11]. Therefore, segmentations of the IJV and CA can be automatically predicted without the requirement of pre- or post-processing the data. Mask R-CNN has recently been applied to medical image processing tasks including the detection and segmentation of meniscus tears [6] and segmentation of the prostate gland and prostatic lesions in MRI images [7]. Other applications include a modified Mask R-CNN for breast tumor detection and segmentation in US images [14]. These successes have motivated the investigation of a Mask R-CNN deep learning solution to automatically segment the CA and IJV from tracked 2D US images and reconstruct the 3D vessels' surfaces for guiding intra-operative interventions.

The objectives of this research are twofold. First, we aim to develop an automatic segmentation framework capable of delineating both the CA and IJV from transverse US images, with an accuracy comparable to that obtained by manual segmentation. We then aim to formulate a vessel reconstruction pipeline to utilize these automatic vascular segmentation and spatial tracking to reconstruct the 3D geometries of the CA and IJV, with a accuracy comparable to that provided by reconstructions from CT angiography. These capabilities have the potential to automate vascular measurements in 2D and 3D and to improve US-guided needle interventions.

Materials and methods

Data collection

All images were collected using the Ultrasonix US scanner (SonixTouch, BK Medical, USA) with the L-14-5 Linear US transducer. As vascular structures can be as deep as 5.5 cm [9], an imaging depth of 6 cm was used to acquire neck vascular US as it should include all human vascular configurations. This US probe was spatially calibrated [5] and tracked using a magnetic tracker (Aurora Tabletop, NDI, Canada). The US calibration provides the spatial pose of the US image with respect to the magnetic tracker's coordinate system, scaled to the true size of the US field of view. The scanning proto-

col was defined as follows. The scan started between the two heads of the sternocleidomastoid muscle just above the clavicle, ending at the mandibular border, and proceeding in an inferior-to-superior direction. The images from these scans were recorded using the PLUS Server [13]. Nine (9) normal control US scans of healthy volunteers were performed by a medical student specifically trained in this procedure, with each subject being imaged in two positions employed in clinical practice: supine on a horizontal table, and head lowered $-15°$ below horizontal. A third-year anesthesia resident performed an additional 6 scans on patients in a local hospital, with patients laying horizontally in a standard hospital bed. The CA and IJV were manually segmented from these US images by a medical student with experience in US neck imaging using 3D Slicer, such that each image had a corresponding mask for both the CA and IJV.[1]

The complete dataset comprises 2439 US images from 15 subjects containing cross-sectional views of the neck vascular anatomy. The US images are stored as 8-bit bitmaps, having pixel intensities in the range of [0, 255]. All images were thresholded, with all grey levels less than 25 being mapped to 0 and all those above 75 being mapped to 75. To perform fourfold cross-validation, this dataset was partitioned into 4 unique training, test and validation sets. Each training set comprised a unique combination of scans and their masks from 11 subjects (70–78% of the dataset). Each test and validation set consisted of unique combinations of images from both a normal control and a patient, as well as their respective labels. The test and validation sets comprise 15–23% and 5–7% of the dataset, respectively. The number of images included in each dataset is summarized in Table 1. No images included in neither the test and nor validation sets were used to train the network, as they were employed solely for evaluation. The number of images in each set varies as the number of images with clear vascular representations differs for each subject, and some have both left and right scans. Each of these training sets were augmented by randomly scaling by a factor in the range 0.8 to 1.2 and rotating by an angle in the range of $-15°$ to $15°$, to produce images that represent possible variation that may occur during scanning. These transformations were automatically performed during training. During this process, the test and validation sets were used to evaluate the Dice score of the trained model to form a baseline accuracy across normal and patient data. However, for analysis, the images within the training and test sets for each fold were reorganized based on whether they had been derived from a normal control or patient subject. The images that comprise these control and patient datasets were not used to train the fold that they would be evaluating. These control and patient images will be analyzed using the Dice score, recall, and precision. This control patient split was selected

to provide a more in-depth analysis on the applications of these networks on control and patient data independently, as well as on the overall accuracy across a mixed cohort.

Deep learning segmentation

Computational hardware used for training the networks included an Intel® Xeon® E5-2683 v4 CPU at 2.1 GHz and 2 NVIDIA® Tesla® P100 GPUs with 12 GB of memory each. All code was written in Python and executed on SHARCNET (Compute Canada's High Performance Computing Network). We trained two neural network models: one with the Mask R-CNN architecture and the other with U-Net CNN for automatic vessel segmentation [11]. Both networks were trained using identical datasets. Memory and computational requirements during training and inference were decreased by resampling the images from 589×374 to 256×256 pixels with bilinear interpolation.

The implemented U-Net architecture was motivated by the standard U-Net encoder–decoder architecture [20]. The encoder consisted of 3 blocks of 2 convolutions with a kernel size (k) of 3, followed by a max pooling layer with $k = 2$. The bottleneck consisted of 2 consecutive convolutions with $k = 3$, while the decoder consisted of 3 blocks of up-convolutions and 2 subsequent convolutions with $k = 3$. The decoder's blocks also received residual connections from the output of blocks in the encoder of the same shape. ReLU was used as the activation function for all intermediate layers. The output layer was a single convolution with $k = 1$ that employed the softmax activation function over the background and classes, producing an output with the same shape as the input image. The network was trained to minimize the categorical cross-entropy loss function. The learning rate (α) was set to 0.0001 at the start of training. During training, if the validation loss did not decrease after the most recent 3 epochs, α was multiplied by 0.5. To encourage regularization, early stopping was applied to halt training when the validation loss did not decrease over the 10 most recent epochs [18]. Each fold was trained over the following number of epochs: set A ran for 33 epochs, set B ran for 19 epochs, set C ran for 27 epochs, and set D ran for 19 epochs. As U-Net is susceptible to false segmentations, a connected-component post-processing algorithm was applied to keep the largest connected segmentation for both the IJV and CA and remove all other segmentations, as done in work by Xie. et al. [27].

A Mask R-CNN model requires ground truth segmentation masks and bounding boxes for training. The bounding boxes were generated automatically by calculating the smallest rectangle that would enclose an individual vessel segmentation, defined by a 4-tuple consisting of two (x, y) coordinate pairs. The input to the Mask R-CNN model was the resized raw US image. The output of the model was a series of 256×256 masks, bounding boxes, and classes for

---

[1] https://github.com/VASST/AIVascularSegmentation.

Table 1 Summary of number of images allocated to each dataset used to train and evaluate the networks

Dataset	Fold A	Fold B	Fold C	Fold D
Training images	1717 (81%)	1799 (81%)	1902 (87%)	1819 (81%)
Test images	552 (95%)	475 (88%)	352 (80%)	481 (83%)
Validation images	170 (53%)	165 (47%)	185 (42%)	139 (53%)
Control images	623	498	360	474
Patient images	99	142	177	146

The percentage of normal control images in the training, test, and validation sets is denoted in brackets. The number of normal and patient images for each fold is the images that will be used to evaluate the networks

each predicted vessel instance. In the rare case that there were more than two object masks predicted by the network, we considered only the two that the network predicted with the highest confidence. The code to define and train the neural network model was adapted from Matterport's Mask R-CNN implementation, which was built using the Keras library with the TensorFlow backend [1]. No changes were made to the core Mask R-CNN architecture. Our model segments objects of two classes: CA and IJV. Although the image background may be considered as a third class, no background segmentation masks are actually predicted by the network. Matterport's implementation [1] offered the choice between ResNet-50 and ResNet-101 as the backbone of the network. ResNet-50 was chosen here because it contains significantly fewer parameters, lending itself to faster training and prediction time [12]. Multiple hyperparameters were tuned by performing several training experiments and adjusting the value of one while keeping others constant. The square anchor boxes used in the RPN had side lengths of 8, 16, 32, 64, and 128 pixels. Sixty-four regions of interest (ROIs) were fed to mask and classifier heads of the network for each image. The RPN non-max suppression threshold was set to 0.7. The learning rate (α) was set to 0.001 at the start of training. During training, if the validation loss did not decrease after the most recent 15 epochs, α was multiplied by 0.75. The batch size was 16 and was spread equally across 2 GPUs during training. The model was trained for 100 epochs to minimize the Mask R-CNN loss function, defined as: $L = L_{cls} + L_{box} + L_{mask}$, where L_{cls} and L_{box} are defined as they were for Fast R-CNN [11], L_{cls} is the categorical cross-entropy loss for object classification, and L_{box} is the smooth L1 loss for bounding box localization [8]. Localization is defined as a 4-tuple consisting of an (x, y) coordinate, width, and height. L_{mask} is the mean per-pixel binary cross-entropy loss across segmentation masks for both classes [11]. The neural network was trained to minimize the loss function. The object segmentation with the highest probability is selected for each class (Fig. 1).

Vessel reconstruction

The automatically segmented label masks and tracking information were used to reconstruct the vessels in 3D. The calibrated spatial tracking data provide the pose of each image in 3D such that the automatic segmentations extracted from each image can be positioned with respect to the field of view of the image where it was captured. Three-dimensional binary morphological hole filling, with an annulus shaped kernel of size [30, 30, 30], was used to fill the gaps between the slices [22]. A 3D Gaussian blur filter with an α of 0.5 was applied to smooth the vessels, as visually depicted in Fig. 2. The four trained Mask R-CNN algorithms were used to obtain surface reconstructions on a patient left-side scan which was not used to train or evaluate any of the folds. The reconstruction algorithm was evaluated through surface-to-surface distance comparisons between the US and CT reconstructed vessels after rigid surface-based registration. The US scanning protocol consistently collected scans beginning just superior to the clavicle. The CT scan segmentations started just superior to the clavicle and ended at approximately the same location as the most superior US image. The point data from these volumes were used to perform an iterative closest point registration [3] that solves for the smallest root mean-squared error between the CT and US volumes, such that they are in a common coordinate system for comparison. The volume and surface area (SA) of the reconstructed vessels from US and CT were calculated. These values were expressed as a ratio of the metric extracted from US to the metric extracted from CT. The smaller value was used as the numerator as to not bias the average.

Results

Fourfold cross-validation was performed, whereby all 2439 collected and segmented images were allocated into training, test, and validation sets in four unique combinations. During training the test and validation sets, each comprised one patient and one normal control scan. The images that were excluded from training were reorganized into patient and control datasets for evaluation. The manual and automatic
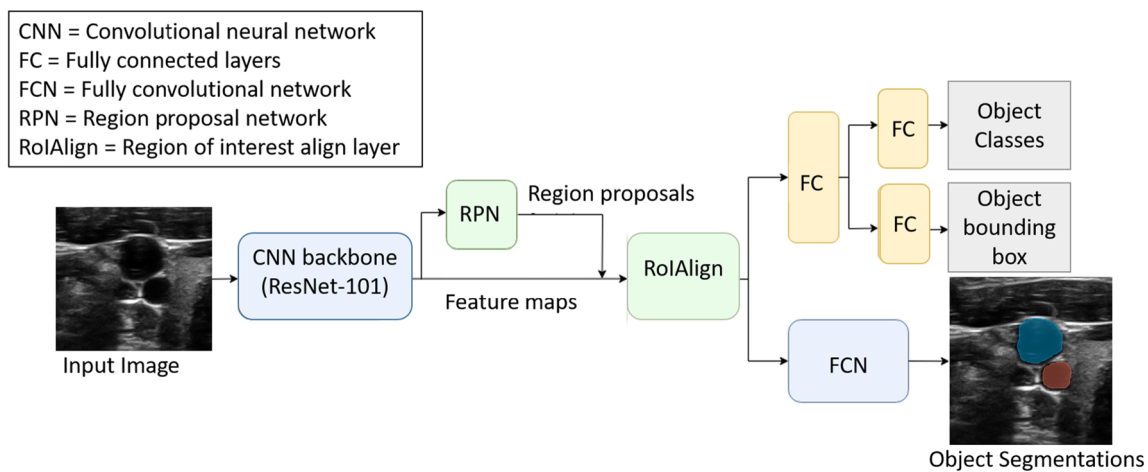
Fig. 1 The Mask R-CNN architecture depicting the CNN backbone, the region proposal network, and the RoIAlign layer. The "box head" is a series of fully connected layers that outputs the predicted class and bounding box for each object. The "mask head" outputs binary segmentation masks for each object instance
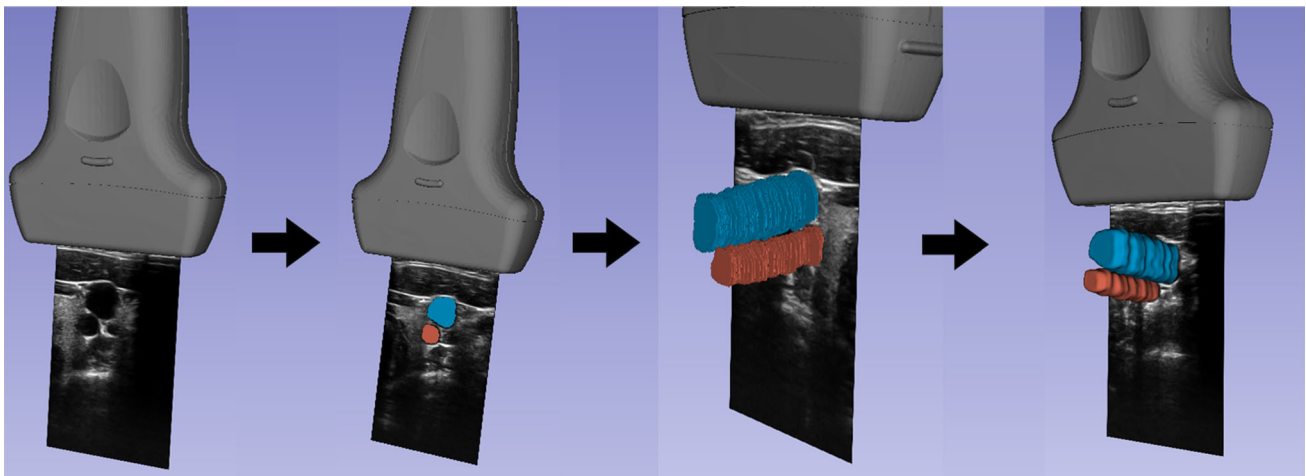

Fig. 2 Visual depiction of the reconstruction process. The first image showing a calibrated US image positioned and scaled to capture the field-of-view of the US image when captured. The second image is a segmented and calibrated US image, where the CA and IJV have been delineated. The third image depicts a vascular skeleton where each image in the tracked scan has been segmented and these segmentations are spatially calibrated to form a skeleton. The final image depicts the closed surface reconstructed vessels after the application of binary morphological hole filling and Gaussian blur smoothing

segmentations produced by the Mask R-CNN and U-Net algorithms were compared by calculating the Dice score, recall, and precision across each class. These results along with the average across all folds and all evaluation images are summarized in Figs. 3, 4, 5. Four sample images were selected to show the potential issues that occur with the U-Net segmentation, and post-processing is depicted in Fig. 6.

The surface-to-surface distance between the registered vessel models from all fourfold is depicted in Figs. 7 and 8, where colors progress from blue (cool) to red (hot) as the distance increases. The SA and volume ratios of the values extracted from the four Mask R-CNN reconstructions and the CT vessels and the average are summarized in Table 2.

The four representative vasculature reconstructions are visualized with respect to the calibrated US image for reference in Fig. 9. These subjects did not have associated neck CT scans, and therefore, a more comprehensive analysis could not be performed.
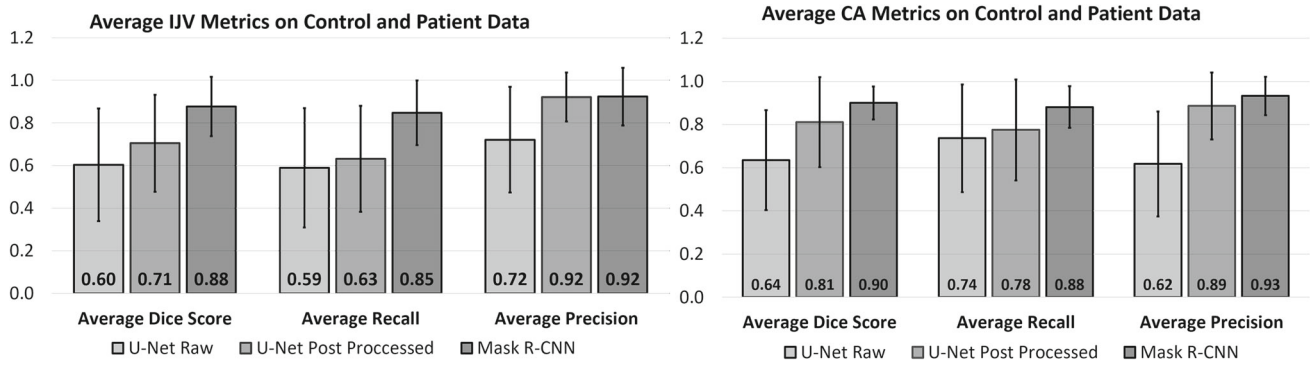
Fig. 3 Summary of the Dice, recall, and precision averaged across the patient and control data from all fourfold for the raw U-Net, post-processed U-Net, and Mask R-CNN. These data are presented for the IJV and CA separately
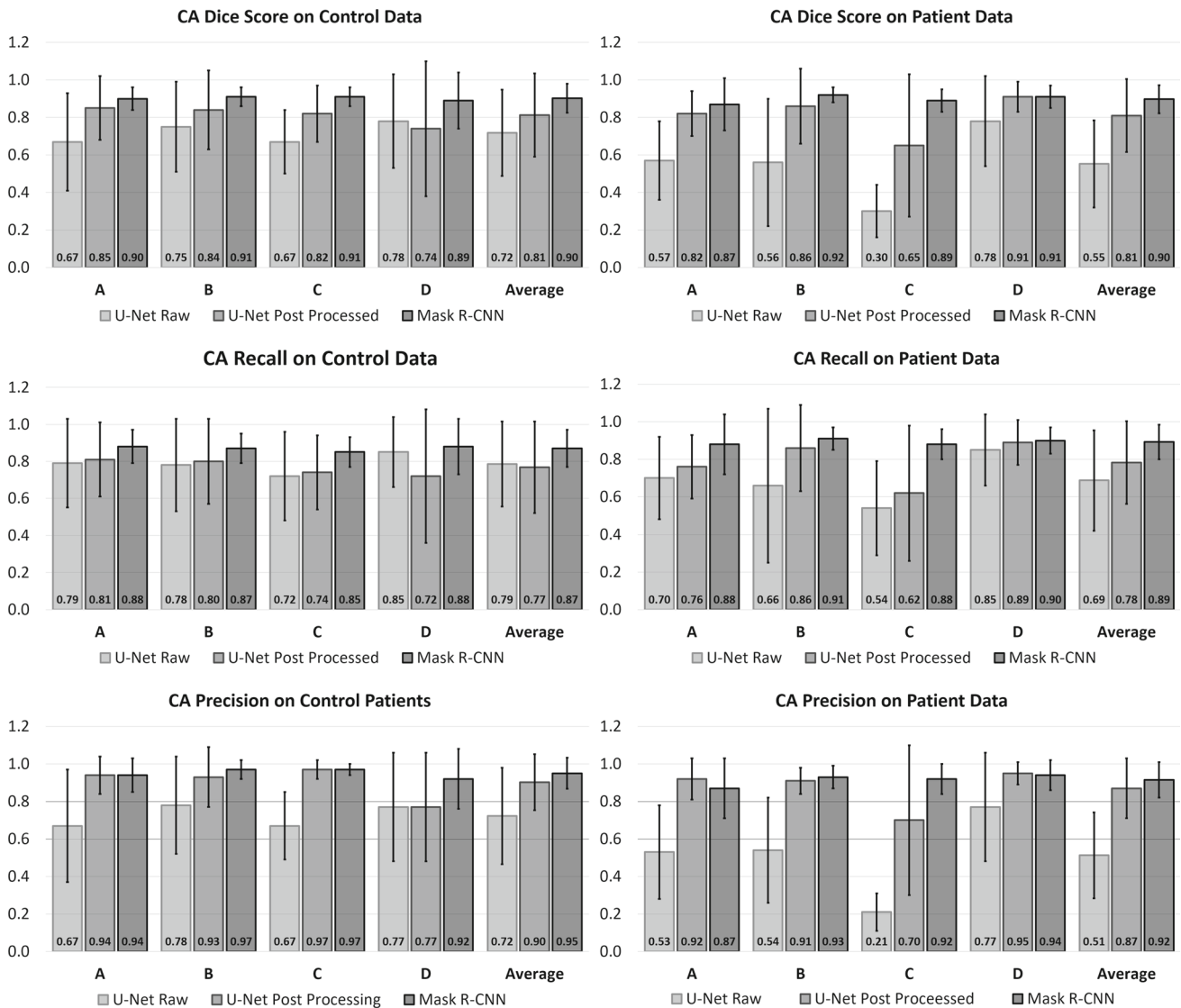

Fig. 4 Average Dice, recall, and precision for the CA from each of the fourfold. These results are reported separately for normal and patient data
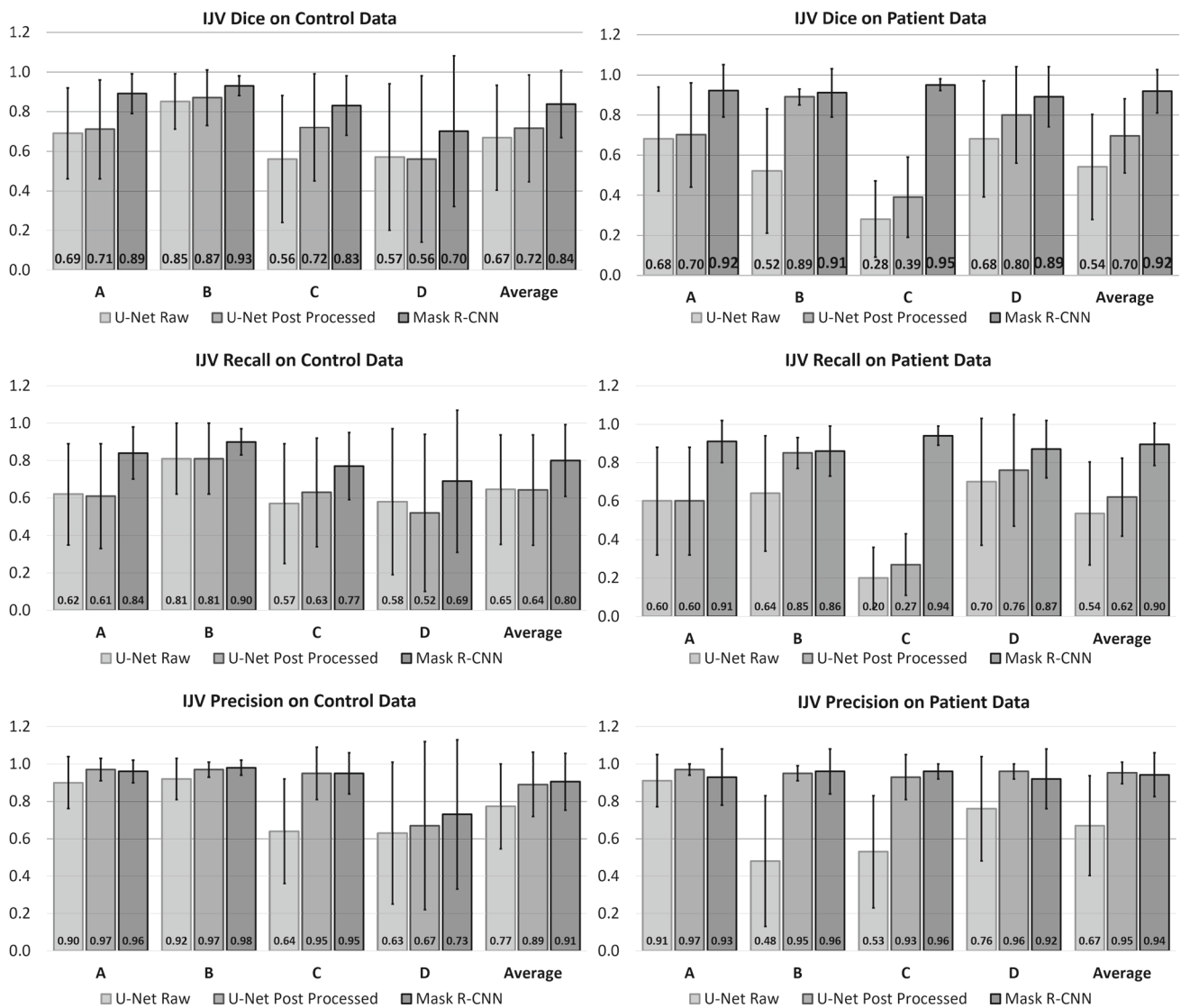
Fig. 5 Average Dice, recall, and precision for the IJV from each of the fourfold. These results are reported separately for normal and patient data

Discussion

In this work, we compare U-Net and Mask R-CNN algorithms both capable of automatically segmenting the CA and IJV from transverse US images. These segmentations can be used to obtain automatic vascular measurements or perform vascular surface reconstruction used for vascular morphology analysis or surgical navigation.

U-Net is a semantic segmentation algorithm where each pixel is assigned to a class. Our implementation produces a label map where each pixel has been assigned to one of three classes: background, CA, or IJV. The raw output of the U-Net may produce multiple clusters of pixels labeled as either the CA or the IJV, with some pixels being misclassified as seen in Fig. 6. These erroneous segmentations motivated using a post-processing step to identify one segmentation for each

of the CA and IJV classes. A major factor that contributes to the high number of false segmentations is the non-unique appearance of the neck vascular structures under US. The CA and IJV are vascular trunks with several branching vessels that have similar features under US. The CA and IJV are the major vascular structures in the neck and should be the largest vascular structures in the US images acquired. For this reason, similar to the work of Xie. et al. [27], we applied a post-processing step that identifies the largest connected-component for each of the CA and IJV classes. The average Dice score for the CA and IJV for the post-processed U-Net is 0.71 ± 0.23 and 0.81 ± 0.21, respectively. Applying this post-processing step improved the Dice score by 0.11 and 0.17, compared to the raw output, for the IJV and CA, respectively. This post-processing algorithm fails in cases where an erroneous segmentation has the largest number of
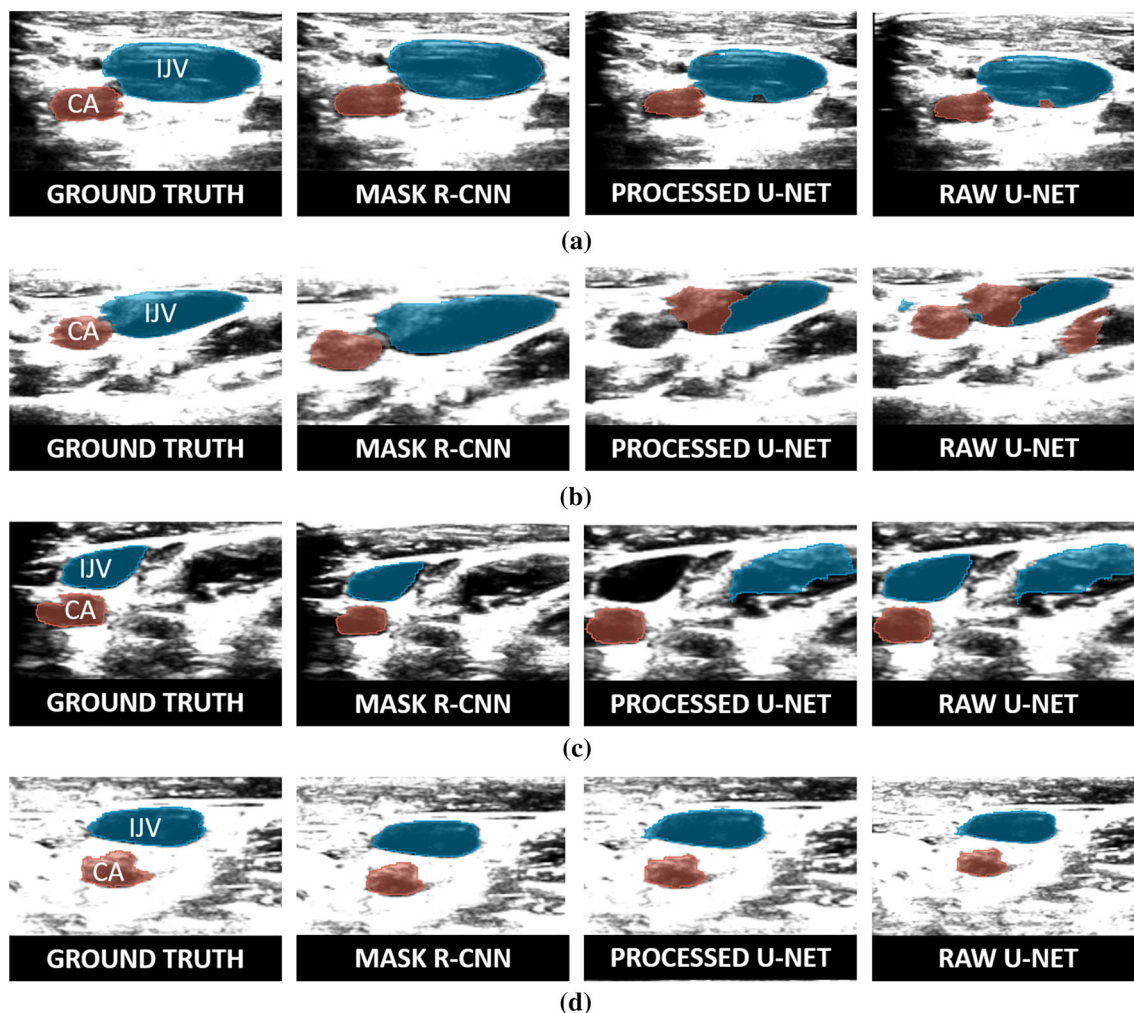
(a)

(b)

(c)

(d)

Fig. 6 Four sample images with their respective outputs from the Mask R-CNN and U-Net with and without processing. Row **a** shows an example of a small cluster of misclassified pixels from the U-Net. Row **b** provides an example of a large group of pixels that have been misclassified as CA when they should be IJV in the U-Net output. Row **c** depicts a vessel-like structure that has been misclassified as the IJV and the post-processing algorithm selecting this false segment as the IJV. Row **d** depicts an image that has accurate outputs across all 3 algorithms. Despite the U-Net output from images a-c containing erroneous segmentations the Mask R-CNN produced accurate segmentations across all sample images

connected components, and thus, the post-processing selects the wrong cluster of pixels (Fig. 6c). Moreover, the U-Net output commonly misclassifies pixels between the CA and IJV (Fig. 6b), an issue that would persist regardless of the post-processing algorithm applied. Both of these issues contribute to the small change in Dice scores. As the accuracy of the post-processed U-Net was still lower than desired, for this application we investigated the use of Mask R-CNN.

The Mask R-CNN contains a regional proposal subnetwork that identifies bounding boxes within the image where segmentations are most likely to occur. The algorithm then segments these structures within the bounding box and returns a probability that they belong to the class to which they have been assigned by the label. The out-

put of our Mask R-CNN algorithm selects the segmentation with the highest probability of belonging to the CA and IJV class. Thus, our algorithm returns a single fully connected segmentation for the CA and IJV based on a trained statistical probability with reduced number of misclassified pixels. The average Dice score for the IJV and CA for the Mask R-CNN is 0.88 ± 0.14 and 0.90 ± 0.08, respectively. The Mask R-CNN improved the Dice score by 0.17 and 0.09 compared to the post-processed U-Net, for the IJV and CA, respectively. For US imaging segmentation problems where features in the image are not unique, or where many structures similar to the structure of interest reside will likely experience similar issues with the U-Net approach. The Mask R-CNN thus serves as a good alternative to U-Net in these

Fig. 7 IJV surface-to-surface distances between the reconstructed US and the ground truth CT for all fourfold. The color progresses to warm colors as distances increase
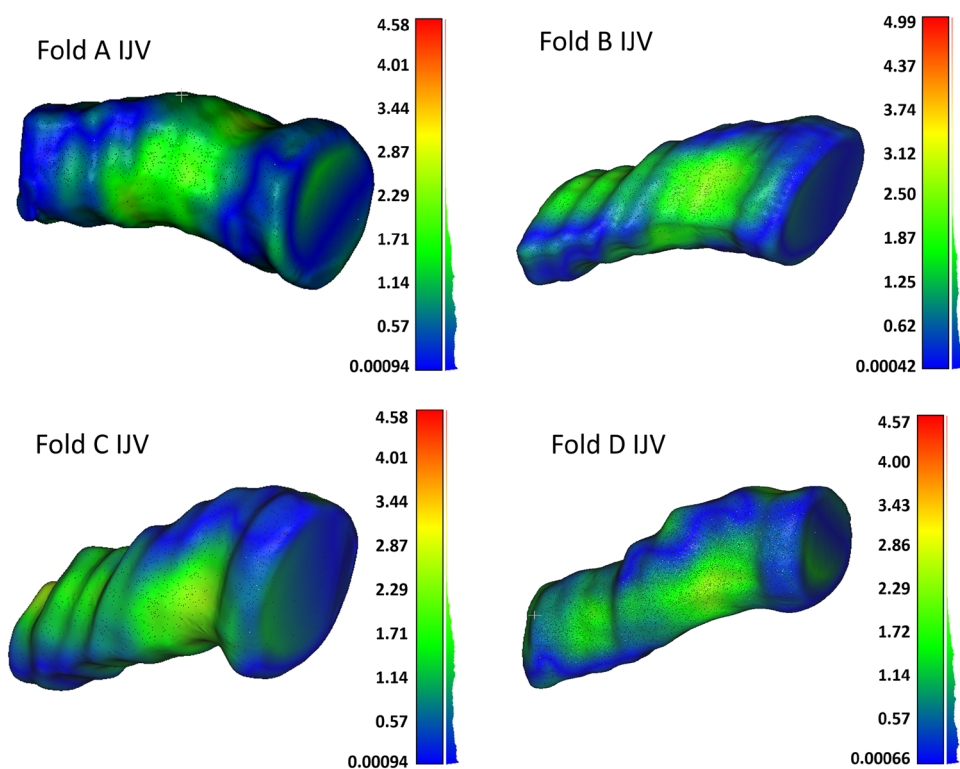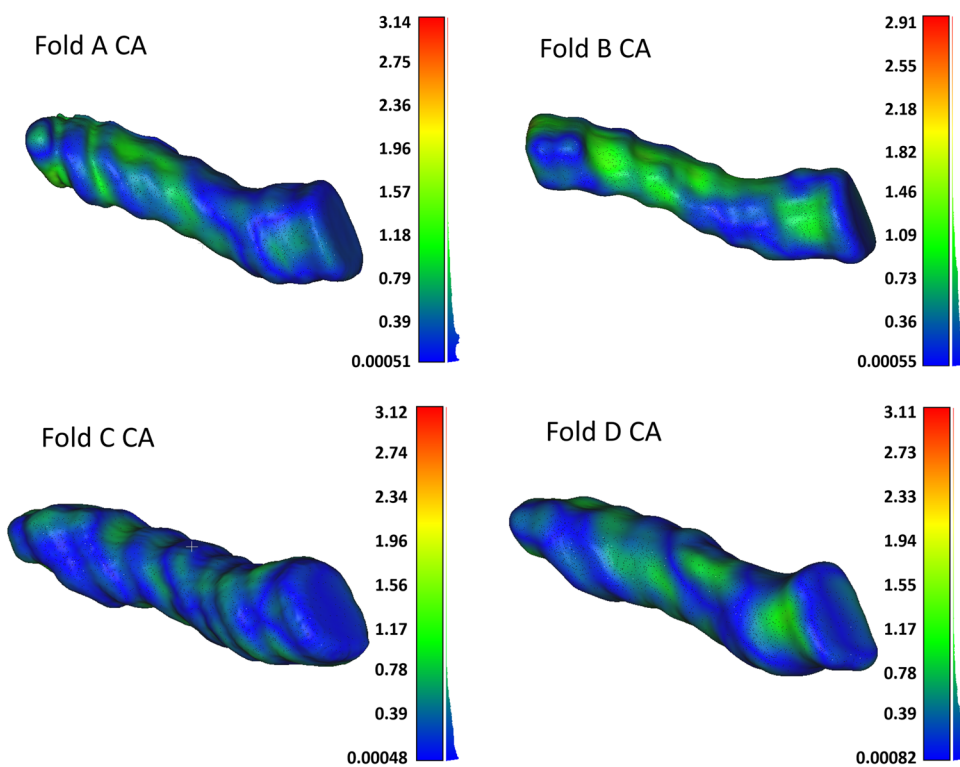

Fig. 8 CA surface-to-surface distances between the reconstructed US and the ground truth CT. The color progresses to warm colors as distances increase
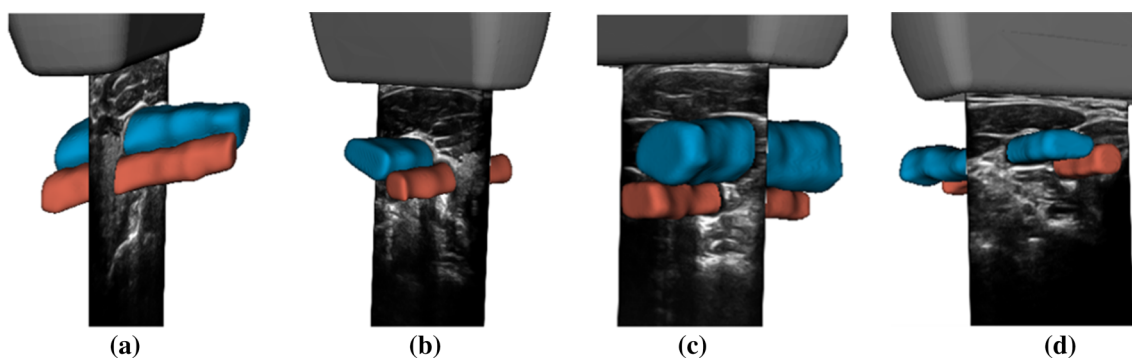
Fig. 9 Each letter (**a**–**d**) represents a unique human subject who was not used to train the algorithm used to produce the segmentations

cases as the architecture is similar to U-Net but does not require post-processing and allows for these segmentations to be selected based on statistical probability. However, the Mask R-CNN model has higher computational requirements when compared to U-Net. The use of a high-end GPU would likely allow the vascular reconstructions to be obtained in near real time. Overall, the Mask R-CNN achieved average Dice scores, recall, and precision of values above 0.85, which are sufficiently accurate to be used for vascular reconstruction and measurements pertaining to the relationship between vessels.

We used all four trained Mask R-CNN networks to obtain vascular US surface reconstructions of the CA and IJV on a patient scan that was not part of the training or evaluation datasets. Each reconstruction was compared to a manually segmented CT scan of the same patient's vasculature, using a surface-to-surface distance analysis (Figs. 7, 8). The CA is slightly more accurate than the IJV, as the IJV is susceptible to deformation under the pressure of the US probe during scanning, and thus is more representative of the true accuracy of the reconstruction. We calculated the ratio of the SA and volume values extracted from the US to the values from the CT reconstructed vessels, as summarized in Table 2. On average, the SA ratio was 0.94 and 0.88, for the CA and IJV, respectively. The average volume ratio was 0.86 for both the CA and IJV. The errors present in the Mask R-CNN results are typically in the form of a loss of detail at the border of the vessel lumen. These small details have minor effects on the ability to use these reconstructions for surgical navigation or vascular measurements. With the majority of points being within 2 mm of the CT reconstructed vessels and with sub-millimeter difference in metrics produced, this algorithm is capable of producing accurate vascular reconstructions.

In future work, we intend to perform a comprehensive accuracy analysis of our reconstructed vasculature through comparing to a larger cohort of patient CT scans. We also aim to apply this vascular reconstruction pipeline to guide central line insertions. Additionally, the multi-class segmentation using Mask R-CNN can trivially be extended

Table 2 Summary of the SA and volume ratio between the metrics produced from the US reconstructions from the four trained networks to the metrics extracted from the CT segmented vasculature

Fold	Fold A	Fold B	Fold C	Fold D	Average
CA SA ratio	0.97	0.84	0.96	0.98	0.94
IJV SA ratio	0.94	0.85	0.81	0.93	0.88
CA volume ratio	0.87	0.70	0.94	0.93	0.86
IJV volume ratio	0.95	0.82	0.75	0.93	0.86

to include additional pathologies and anatomical structures. One possible extension for future work is segmentation of calcified plaques. Plaques have a non-unique appearance in US images. Relying on a network such as U-Net or algorithms based on feature detection would likely result in many incorrect segmentations of plaques. As the size of plaques can vary drastically, a more rigorous post-processing selection algorithm is required. The Mask R-CNN is more suitable than U-Net for this type of application, as it provides a statistical method for selecting the appropriate segmentation, which is important for multi-class segmentation problems where features are not inherently unique to the structure of interest. Furthermore, segmentation of plaques could be framed as an instance segmentation problem, in which the Mask-RCNN was designed to accomplish. The ability to automatically segment pathologies and visualize them with respect to the CA and IJV US reconstructions would provide improved surgical guidance with no harm to patients. As a result, measurements related to pathologies such as total plaque volume or common locations of plaque within the CA may be determined. We also aim to validate the usefulness of 3D reconstructed models for surgical navigation or planning.

Conclusions

In this work, we compared Mask R-CNN and U-Net algorithms developed to automatically segment the CA and IJV

from transverse US images. The Mask R-CNN algorithm was more accurate than the U-Net alternative and achieved average Dice scores of 0.88 ± 0.14 and 0.90 ± 0.08, for the IJV and CA, respectively. The Mask R-CNN-based vascular reconstruction pipeline was accurate compared to the CT equivalent with majority of distances between the surfaces being less than 2 mm. These reconstructions were able to produce accurate metrics with the average ratio of the volume produced by the US to the volume produced by the CT being 0.86 for both the CA and IJV. This work can be used to analyze neck vasculature morphology in both 2D and 3D. Furthermore, the 3D models can be used for surgical planning or surgical navigation. Overall, we have developed and evaluated a highly accurate Mask R-CNN algorithm for instance segmentation of the CA and IJV in transverse US images.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Abdulla W (2017) Mask R-CNN for object detection and instance segmentation on Keras and Tensorflow
2. Ameri G, Baxter JSH, Bainbridge D, Peters TM, Chen ECS (2018) Mixed reality ultrasound guidance system: a case study in system development and a cautionary tale. Int J Comput Assist Radiol Surg 13(4):495–505
3. Besl PJ, McKay ND (1992) Method for registration of 3-d shapes. In: Sensor fusion IV: control paradigms and data structures, vol 1611. International Society for Optics and Photonics, pp 586–606
4. Chao A, Lai CH, Chan KC, Yeh CC, Yeh HM, Fan SZ, Sun WZ (2014) Performance of central venous catheterization by medical students: a retrospective study of students' logbooks. BMC Med Educ 14(1):168
5. Chen ECS, Peters TM, Ma B (2016) Guided ultrasound calibration: where, how, and how many calibration fiducials. Int J Comput Assist Radiol Surg 11(6):889–898
6. Couteaux V, Si-Mohamed S, Nempont O, Lefevre T, Popoff A, Pizaine G, Villain N, Bloch I, Cotten A, Boussel L (2019) Automatic knee meniscus tear detection and orientation classification with Mask-RCNN. Diagn Interv Imaging 100(4):235–242
7. Dai Z, Carver E, Liu C, Lee J, Feldman A, Zong W, Pantelic M, Elshaikh M, Wen N (2020) Segmentation of the prostatic gland and the intraprostatic lesions on multiparametic MRI using Mask R-CNN. Adv Radiat Oncol 5:473–481
8. Girshick R (2015) Fast R-CNN. In: Proceedings of the IEEE international conference on computer vision 2015 (ICCV 2015), pp 1440–1448
9. Gordon AC, Saliken John C, Johns D, Owen Richardand Gray RR (1998) US-guided puncture of the internal jugular vein: complications and anatomic considerations. J Vasc Interv Radiol 9(2):333–338
10. Groves L, Li N, Peters TM, Chen ECS (2019) Towards a mixed-reality first person point of view needle navigation system. In: Essert C, Zhou S, Yap PT, Khan A, Shen D, Liu T, Peters TM, LH Staib (eds) Medical image computing and computer assisted intervention (MICCAI 2019). Springer, Berlin, pp 245–253
11. He K, Gkioxari G, Dollár P, Girshick R (2017) Mask R-CNN. In: 2017 IEEE international conference on computer vision (ICCV), pp 2980–2988
12. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR), pp 770–778
13. Lasso A, Heffter T, Rankin A, Pinter C, Ungi T, Fichtinger G (2014) PLUS: open-source toolkit for ultrasound-guided intervention systems. IEEE Trans Biomed Eng 61(10):2527–2537
14. Liu J, Li P (2018) A Mask R-CNN model with improved region proposal network for medical ultrasound image. In: Huang DS, Jo KH, Zhang XL (eds) Intelligent computing theories and application. Springer, Berlin, pp 26–33
15. Lo A, Oehley M, Bartlett A, Adams D, Blyth P, Al-Ali S (2006) Anatomical variations of the common carotid artery bifurcation. ANZ J Surg 76(11):970–972
16. Merritt RL, Hachadorian ME, Michaels K, Zevallos E, Mhayamaguru KM, Closser Z, Derr C (2018) The effect of head rotation on the relative vascular anatomy of the neck: implications for central venous access. J Emerg Trauma Shock 11(3):193–196
17. Niessen WJ, Bouma CJ, Vincken KL, Viergever MA (2000) Error metrics for quantitative evaluation of medical image segmentation. In: Klette R, Stiehl HS, Viergever MA, Vincken KL (eds) Performance characterization in computer vision. Springer, Berlin, pp 275–284
18. Prechelt L (2012) Early stopping—but when? In: Neural networks: tricks of the trade, 2nd ed. Springer, Berlin, pp 53–67
19. Ren S, He K, Girshick R, Sun J (2015) Faster R-CNN: towards real-time object detection with region proposal networks. In: Cortes C, Lawrence ND, Lee DD, Sugiyama M, Garnett R (eds) Advances in neural information processing systems, Curran Associates, Inc., pp 91–99
20. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics), vol 9351. Springer, Berlin, pp 234–241
21. Saugel B, Scheeren TWL, Teboul JL (2017) Ultrasound-guided central venous catheter placement: a structured review and recommendations for clinical practice. Crit Care 21(1):225
22. Soille P (2004) Morphological image analysis. Springer, Berlin
23. Turba UC, Uflacker R, Hannegan C, Selby JB (2005) Anatomic relationship of the internaljugular vein and the common carotid artery applied to percutaneous transjugular procedures. CardioVasc Interv Radiol 28(3):303–306
24. Ukwatta E, Awad J, Buchanan D, Parraga G, Fenster A (2012) Three-dimensional semi-automated segmentation of carotid atherosclerosis from three-dimensional ultrasound images. In:

Medical imaging 2012: computer-aided diagnosis, vol 8315, p 83150O. International Society for Optics and Photonics

25. Wang W, Liao X, Chen ECS, Moore J, Baxter JSH, Peters Terry M, Bainbridge D (2019) The effects of positioning on the volume/location of the internal jugular vein using 2-dimensional tracked ultrasound. J Cardiothor Vasc Anesth 34:920–925

26. Woldeyes DH (2014) Anatomical variations of the common carotid artery bifurcations in relation to the cervical vertebrae in Ethiopia. Anat Physiol Curr Res 4(3). https://doi.org/10.4172/2161-0940.1000143

27. Xie M, Li Y, Xue Y, Shafritz R, Rahimi SA, Ady JW, Roshan UW (2019) Vessel lumen segmentation in internal carotid artery ultrasounds with deep convolutional neural networks. In: 2019 IEEE international conference on bioinformatics and biomedicine (BIBM). IEEE, pp 2393–2398

28. Zhou R, Fenster A, Xia Y, Spence JD, Ding M (2019) Deep learning-based carotid media-adventitia and lumen-intima boundary segmentation from three-dimensional ultrasound images. Med Phys 46(7):mp.13581

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.