2023 IEEE International Conference on Robotics and Automation (ICRA 2023)
May 29 - June 2, 2023. London, UK

Reslicing Ultrasound Images for Data Augmentation and Vessel Reconstruction

Cecilia G. Morales, Jason Yao, Tejas Rane, Robert Edman, Howie Choset, Artur Dubrawski

Abstract— **Robot-guided vascular access has the potential to deliver urgent medical care in situations where medical personnel are unavailable. However, this technique requires accurate and reliable segmentation of anatomical landmarks in the body. For the ultrasound imaging modality, obtaining large amounts of training data for a segmentation model is time-consuming and expensive. This paper introduces RESUS (RESlicing of UltraSound Images), a weak supervision data augmentation technique for ultrasound images based on slicing reconstructed 3D volumes from tracked 2D images. This technique allows us to generate views which cannot be easily obtained in vivo due to physical constraints of ultrasound imaging, and use these augmented ultrasound images to train a semantic segmentation model. We demonstrate that RESUS achieves statistically significant improvement over training with non-augmented images and highlight qualitative improvements through vessel reconstruction.**

I. INTRODUCTION

Prompt vascular access is crucial for trauma patients to receive anesthesia, medications, fluids, and diagnostic procedures [1], [2]. In severe cases of injury, hemorrhaging is a leading cause of death within the first hour [1]. Quick treatment is vital to save the life of a patient in critical condition, but access to treatment resources can be limited in remote or mass casualty situations. Robot-guided vascular access could enable necessary medical care to the patient within a short period of time where medical personnel are not yet readily available. Identifying femoral blood vessels is crucial for resuscitation, and a suitable medical imaging modality is necessary for autonomous needle insertion.

Medical imaging involves non-invasive procedures aimed to visualize anatomical features within the body. This work utilizes ultrasound as it allows real-time imaging, easy portability, no need for intravenous contrast agents, no ionizing radiation, and low-costs. By attaching an ultrasound transducer to the robot's end-effector and utilizing the method described in [3], we can scan the femoral region to identify the exact location of the vessels. Using a robotic ultrasound system ensures accuracy, stability, maneuverability, and tracking the location of the transducer during ultrasound image acquisition.

Nonetheless, automatically segmenting anatomical landmarks in the body such as arteries or veins is a challenging task. One must consider the noise and artifacts in images, inter-patient and machine variability, as well as natural and injury-inflicted anatomical differences. Acquiring comprehensive sets of training data for all these scenarios would be time-consuming, difficult and expensive, since only professionally trained personnel are able to adjudicate them.

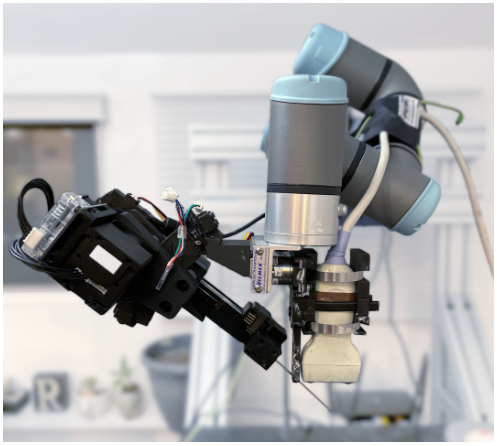

Fig. 1. Our Mechanical Ultrasound Scanning System, which consists of a linear array ultrasound transducer attached to the end-effector of a 6-DoF Universal Robots UR3e serial manipulator.

In this work, we introduce RESUS (**RES**licing of **U**ltra**S**ound Images), a weak supervision data augmentation technique for ultrasound images. Specifically, the method is designed to reconstruct 3D volumes made up of tracked 2D ultrasound images. The volume can then be resliced to reveal 2D images in planes that would not be accessible due to the physical restrictions of the scanning process, i.e., a plane parallel to the surface of the skin. These new reconstructed ultrasound images are then used to train a semantic segmentation U-Net network [4] such that it is able to generalize to different anatomical variations and slight altercations, artifacts, and varying image resolutions. A flow diagram of the method can be found in (Fig. 2).

We collect ultrasound images with a robotic arm through a series of experiments on a medical imaging phantom as well as real-world data. We compare our method with the predictions of a U-Net trained on non-augmented images, a U-Net pretrained on ImageNet, and a U-Net trained on random augmentation images. RESUS achieves statistically significant, respectively 8%, 5%, and 23% increases in Intersection over Union (IoU) score, also known as Jaccard Index, in segmenting images.

The main contributions of our work are: 1) Developing a novel method that augments ultrasound images to train on

*This work was supported by U.S. Department of Defense contracts W81XWH-19-C0083, W81XWH-19-C0101, and W81XWH-19-C-0020

Authors are with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA {cgmorale, jlyao, tejasr, redman, choset, awd}@andrew.cmu.edu

new imaging perspectives which are otherwise inaccessible due to the physical constraints of the system; 2) Creating a novel volume reconstruction technique to adapt to the data collected; 3) Evaluating the effectiveness of our method both qualitatively and quantitatively in increasing the IoU score; 4) Achieving better vessel reconstructions affordably using weak supervision on real-world data; 5) Increasing the availability of data and easing the labelling task.

II. BACKGROUND AND RELATED WORK

A. Volume Reconstruction Strategies

Volume reconstruction is the process of creating a 3D model from 2D ultrasound images and their corresponding positions. The two main reconstruction methods are pixel-based and voxel-based. The pixel nearest neighbor (PNN) method is the most common pixel-based method [5]. Each pixel in the 2D images is directly mapped to the nearest voxel in the 3D volume. It is simple and computationally less expensive but it results in the loss of important details and produce a blocky appearance due to the low resolution of the 2D images [6]. Sugano et al. employed the pixel method to reconstruct blood vessels from ultrasound images by positional calibration of the ultrasound probe [7]. Since their method did not include error detections and was only verified on a simply shaped artificial vessel, they extended their work to obtained a volume data of blood vessel by scanning the conventional 2D probe and a state-of-art 3D probe to extract bifurcation points by structural analysis [8].

Voxel-based methods use 2D ultrasound frames to reconstruct a detailed 3D volume by interpolating or averaging pixel values [5]. This method considers the spatial relationship between voxels and pixels, resulting in accurate images. However, it requires more computation and may be sensitive to the quality of 2D images. Voxel-nearest neighbor (VNN) and distance-weighted (DW) are common methods of computing the voxel values. VNN preserves original texture patterns but generates large artifacts and speckle noise. DW suppresses speckle noise but smooths out the volume, losing information. We use the squared-distance-weighted (SDW) reconstruction algorithm proposed by Huang et al. [9], an extension of DW, to reduce the smoothing effect and preserve details in the reconstructed volume.

B. Semantic Segmentation using Deep Neural Networks

Manual segmentation of images is a time-consuming and labor-intensive task that can introduce bias since it involves subjective opinions. To address these issues in medical image analysis, deep learning techniques have been developed to accurately and efficiently segment organs and lesions from medical images.

One of the most popular deep learning techniques used for medical image segmentation is the U-Net [10]. The U-Net architecture consists of two paths, an encoder and a decoder, which provide classification and localization information, respectively. This structure meets the requirements of medical image segmentation with its fast training speeds and ability to yield reliable models using small amounts of training data. Various strategies have been applied to modify the U-Net structure to address different segmentation problems, resulting in improvements in feature enhancement, training speed optimization, training accuracy, feature fusion, small sample training sets, and generalization improvement [11], [10].

Antecedently, researchers have made modifications and developed new methods to improve vessel segmentation using U-Nets, such as image orientation and probe pose classification. One successful strategy by Jiang et al. was the use of a CNN composed of three U-Nets with an additional averaging segmentation network from US images in different orientations. This method was for the purpose of segmenting the carotid vessel wall using 3D carotid ultrasound images [12]. Another method was implemented by Barr et al., who created a study to evaluate deep learning methods for vessel segmentation and ultrasound pose classification to assess if electromagnetic tracking-based methods could be replaced [13]. They used a U-Net for vessel segmentation and a shallow CNN architecture to classify the pose of the ultrasound probe. A second classifier architecture was also tested that used the U-Net output as the CNN input. However, one of the major limitations of the study was that all experiments were done on a single venous access phantom. Lastly, when the use of Doppler[1] is available, combining it with the input image, to enter the U-Net can help with identifying the artery versus vein [14].

Another deep learning approach adopted by [15] used the You Only Look Once (YOLO) v3-Tiny, a classification network to detect and compute bounding boxes on arteries, veins, and bifurcations. The network was first trained on ImageNet, then, adapted to the ultrasound images via transfer learning. Furthermore, affine transformations and random intensities were used as data augmentation techniques to improve the generalization performance. Nonetheless this method lacked the major segmentation features of vessels.

C. Augmentations

When working with deep neural networks for semantic segmentation, data augmentation is a relatively simple and commonly used method for generalizing across unseen domains and to account for invariances. Many data augmentation methods for medical images involve simple image transformations, such as flips, skips, skews and blurs. Smistad et al. highlighted the use of different augmentations in the context of ultrasound images, and found that rotation, shadow, and elastic deformation gave the best results [16].

A novel data augmentation technique for ultrasound images was developed by [17], where a Bayesian temporal-based segmentation network was used to output both the segmentation and epistemic uncertainty maps. The epistemic uncertainty maps, are used to obtain knowledge of the model's spatial weaknesses, and are then passed into an augmentation module which is responsible for generating the

[1] A noninvasive test that can be used to estimate the blood flow through your blood vessels by bouncing high-frequency ultrasound off circulating red blood cells.

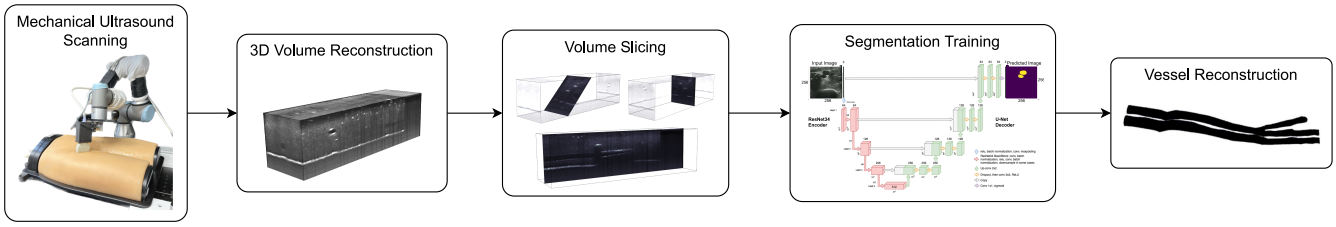

Fig. 2. Flow diagram of the proposed method. We collect 2D ultrasound images from a linear array ultrasound transducer attached to the end-effector of the 6-DoF Universal Robots UR3e serial manipulator. A 3D volume is reconstructed from these images, and then resliced to generate frames, some of which would have been inaccessible during scanning. These frames are then added to the training data to obtain the segmentation model, which identifies the vessels. The segmented vessels inform vessel reconstruction needed to find the optimal needle insertion location.

synthetic images for further training. This method was able to enhance the generalizability of the 3D U-Net, but at the cost of high memory usage and long computation time, therefore not being feasible for real-time field applications, or large amounts of data.

III. METHOD

In our approach, a robot with an ultrasound probe in its end-effector (Fig.1) scans the femoral region of the leg and reconstructs a volume that is then sliced to reveal 2D images that would have not been accessible for the ultrasound due to the physical restrictions of the scanning process. This produces a diverse set of views of the vessel that help generalize to different patients. Those images are used to train a vessel segmentation model. Once the vessels are segmented, they can be reconstructed and used to find appropriate needle insertion sites.

A. Dataset

Our method was first tested on a medical imaging phantom, CAE Blue Phantom anthropomorphic gel model. Images were scanned from both the right side and left side of the phantom. Once the pipeline was ready, we transitioned to experiments in a surgical environment at the University of Pittsburgh Medical Center (UPMC) with live pigs under anesthesia. The experiments involving live animals have been conducted in accordance with the Institutional Animal Care and Use Committee (IACUC) protocol approved by the cognizant authority. We collected ultrasound images of the femoral arteries and veins from eight different pigs, 500 to 2070 images each, using the 6-DoF Universal Robot UR3e serial manipulator to move the probe used for scanning. In-line volume reconstructions were achieved with five of the aforementioned pigs. The probe used was the Fukuda Denshi portable point-of-care scanner (POCUS) with 5MHz linear transducer with a maximum depth of 5cm. A total of 14,207 images were collected and labelled by expert clinicians using the Computer Vision Annotating Tool (CVAT) [18]. Robot Operating System (ROS) [19] was used to capture the robot poses along with ultrasound images that were time-synchronized concurrently.

B. Volume Definition

To reconstruct a 3D volume, we first select a voxel volume and choose a subset of image frames that exclude irrelevant data such as stationary images and extraneous

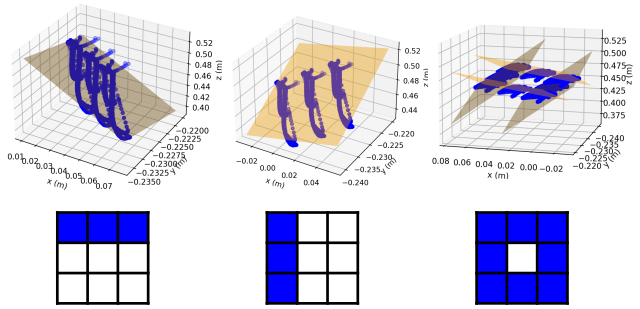

Fig. 3. Point clouds created from transformations of the top (left), left (center), and all sides (right) of the 3×3 image. The planes are fitted with RANSAC on each side separately.

movement. We treat all pixels in the selected frames as points in a point cloud and transform them to their positions in space. Assuming the resulting point cloud is shaped like a parallelepiped, we choose one of its corners as the new origin and use its adjacent vertices as the new basis vectors. We then perform a change of coordinates on all the points in the point cloud to the new volume coordinates and scale each dimension up by the desired number of voxels. This assumption works well for scanning linear vessels in the leg, but for zigzagging patterns, the data must be separated based on the direction of motion. This separation ensures that the resulting point cloud is still shaped like a parallelepiped and allows for the identification of the relevant vertices.

Finding the relevant vertices of the parallelepiped is the main challenge of this method. To address this, we fit planes to each side of the parallelepiped using a modified RANSAC [20] algorithm. This algorithm repeatedly samples 3 points randomly from the point cloud and defines a plane through them. It then computes the distance of that plane to all the points in the point cloud and counts the number of inlier points within a certain distance threshold. The plane with the greatest number of inliers after a set number of iterations is chosen as the best-fit plane.

To reduce the number of points in the point cloud, we first scale each image to have dimensions of 3×3 pixels, with the corner pixels of the smaller image ending up in the same coordinates as the corners of the original image. We then use three pixels from each image corresponding to the four sides of the image to create point clouds for the four side faces of the volume (see Figure 3).

For the two base faces, we first find the normal vector by running RANSAC on the point cloud created by only per-

forming the rotation transform on each image and omitting the translation. This plane represents the overall orientation of all the image frames, which is then translated to the coordinates of the start and end frames to close out the volume.

We calculate the relevant vertices by finding the intersections between these six planes. The origin is chosen to be the intersection point closest to the top-left corner of the start frame. The vertex along the vector representing the new z axis is the point closest to the top-left corner of the end frame, and the x and y vectors are chosen according to the right-hand rule.

C. Volume Reconstruction

Once the volume has been defined, we can fill in the values. Our reconstruction strategy uses the SDW algorithm from Huang et al. [9], which assigns the intensity for each voxel in the volume by considering all the image pixels that fall into the spherical region within radius r of the voxel. A voxel at coordinate $\vec{V}_C$ has intensity

$$I(\vec{V}_C) = \frac{\sum_{k=0}^{n} W_k I(\vec{V}_P^k)}{\sum_{k=0}^{n} W_k} \tag{1}$$

where $I(\vec{V}_P^k)$ is the intensity of the kth pixel out of n falling within the spherical region at coordinate $\vec{V}_P^k$. W_k is the weight of the kth pixel, defined by

$$W_k = \frac{1}{(d_k + \alpha)^2} \tag{2}$$

where d_k is the distance between $\vec{V}_C$ and $\vec{V}_P^k$ and α is a positive parameter to adjust the reconstruction.

After the initial reconstruction of the 3D volume from the image pixel data, the volume may contain holes that need to be filled. To address this issue, we employ the same algorithm used in the initial reconstruction, but with nearby voxels instead of image pixels. The hole-filling process involves multiple passes, with the radius of influence, denoted by r, being increased after each pass up to a certain limit.

Our implementation keeps track of the numerators and denominators for each voxel separately. We loop over each image frame and add every pixel's contribution to nearby voxels. For each frame, we transform the pixel vectors in the image to their corresponding coordinates in the volume. This is done by first applying the corresponding ROS transform to obtain coordinates in terms of the robot base and then applying a second transform from the robot base coordinates to the volume coordinates, as defined in the previous section. Specifically, for each 2D pixel vector $\vec{P}$, we have:

$$\vec{V}_P = \begin{bmatrix} n_x & 0 & 0 \\ 0 & n_y & 0 \\ 0 & 0 & n_z \end{bmatrix} X^{-1} \begin{bmatrix} R & \vec{b} - \vec{u} \end{bmatrix} \begin{bmatrix} \vec{P} \\ 0 \\ 1 \end{bmatrix} \tag{3}$$

where R is the 3×3 rotation matrix and $\vec{b}$ is the 3×1 translation vector from the ultrasound image frame to the robot base frame, $\vec{u}$ is the new origin, X is the 3×3 matrix with columns consisting of the new basis vectors,

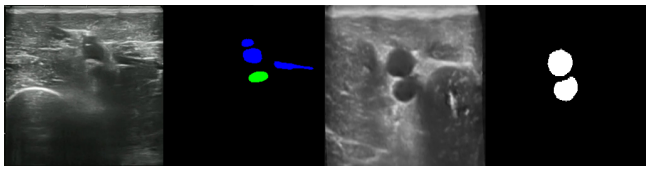

Fig. 4. Raw ultrasound image and label (left) and sliced augmented image and label (right)

and n_x, n_y, n_z are the number of voxels for each dimension of the volume. We augment the initial pixel vector to be able to perform matrix multiplication and vector addition with a single multiplication. Then, each pixel's distance and intensity to all voxels within a radius r is calculated, and its contribution to each voxel's overall summation is added.

After processing all the images, the algorithm identifies hole locations as the voxels with a denominator of zero. These locations are temporarily set to 1, and the intensity of each voxel is calculated by division. During the hole-filling stage, the algorithm is modified to exclude contributions from voxels that are also holes.

To reconstruct the ultrasound images and label classes, each class (e.g., vein, artery, and background) was processed separately, with each pixel assigned a value of 1 if it belonged to the corresponding class and 0 otherwise. The final label volume was obtained by assigning the label with the highest value to each voxel.

D. Volume Slicing

From the reconstructed volumes, we now extract the augmented ultrasound images and labels. To achieve this, the volumes are rotated around the x and y axes and fit within a new output volume using spline interpolation. The output volume is then expanded to include all the data from the original volume, with missing data filled with zeros.

Then, vertical slices are taken along the z axis of the rotated output volume to generate the augmented images. These vertical slices of the rotated volume correspond to oblique slices of the original volume. Each augmented image slice is first checked to ensure that the percentage of pixels corresponding to the original volume (i.e., nonzero pixels) meets a certain threshold to ensure the augmented image contains enough useful information. If threshold is met, the maximum amount of zero pixels are cropped from each side of the image, and then the image is resized to 256×256 pixel grid. For our results, the volumes were rotated $\pm 30^\circ$ around both axes with steps of 10°. Vertical slices were taken once every 5 voxels, and our threshold was 40% for the proportion of nonzero pixels needed in the augmented image.

E. Training

Our goal is to develop semantic segmentation of the femoral artery and vein in ultrasound images for the purpose of 3D reconstruction of the vessels for the robot to find an adequate needle insertion point. To that end, the deep learning network architecture that we use is a U-Net with an image classification model, ResNet34 [21], backbone in PyTorch [22] as the encoder using the Segmentation

Models library [23]. ResNet34 was chosen due to its relatively high training speed and low memory demands [24]. ResNet34 consists of 34 layers with (3x3) convolutional filters using same padding, max-pooling layers and fully-connected layers, ending with a sigmoid activation function used to ascertain the output class. We train our network until the validation Dice loss converges with batch size 8 on lower resolution (256×256) images. We use the Adam optimizer[25] with a learning rate of 0.0001. The network was trained on NVIDIA RTX A6000 GPUs.

The model is evaluated using leave-one-subject-out cross-validation protocol, in which each pig's data is used once as a test set while the remaining data forms the training set. Augmented images from RESUS are only used for training. To avoid over-fitting, images resliced from the test pig are removed from the training set.

F. Vessel Reconstruction

Using the method described in III-C, we can input the predictions to reconstruct the vessels at the exact location they were scanned, thus creating an internal map of the vessels. To differentiate between different types of vessels (artery vs. vein), we use the knowledge of body's anatomy.

IV. EXPERIMENTS

In this study, we conducted two sets of experiments to evaluate the performance of our proposed RESUS augmentation method for vessel segmentation in ultrasound images.

In the first experiment, we tested our method using a medical phantom by creating a 3D volume and reslicing it. We trained two different models: one on unaugmented ultrasound images in the transverse view and the other on the unaugmented images along with the resliced images from the reconstructed volumes. We evaluated the models using two different test sets: the first consisted of unaugmented images only, while the second included both unaugmented and resliced images representing various variations in real-world data, such as longitudinal vessels. We reconstructed the vessels using the model predictions and compared them to the ground truth both qualitatively and quantitatively using the IoU segmentation score.

In the second experiment, we transitioned to real-world data by training and testing on ultrasound images collected from live pigs. We tested our RESUS augmentation method against models trained on images collected from different subjects to assess the inter-subject generalization capabilities of the models. Additionally, we tested models pretrained on ImageNet and those using common augmentation techniques for medical images, including horizontal and vertical flips, Gaussian noise addition, sharpening of the images, random brightness, and random contrast. We reconstructed the vessels using the model predictions and compared them visually to the ground truth. We also used the IoU score metric to evaluate the accuracy of segmentation.

It is worth noting that we were unable to directly compare our RESUS method with the Bayesian temporal-based segmentation network proposed by [17]. This is due to the fact that their method was designed to handle a relatively small batch size and was limited to training on 60 images at a time, whereas RESUS utilizes a larger batch size with the sliced images.

V. RESULTS AND ANALYSIS

To evaluate the efficacy of our method we use qualitative and quantitative approaches. We observe that our method consistently surpasses the rest of the models at vessel reconstruction and removal of artifacts. As it can be observed in Figure 5C, RESUS reconstruction is visually closest to the ground truth. The U-Net prediction has artifacts that could be confused by a robot for a vessel, which could be life threatening in a real-life scenario. Quantitatively, the proposed method yields about 20% improvement in IoU score for the phantom vessels.

The IoU score of 0.186 obtained when training on ultrasound and testing on both US and RESUS images can be attributed to the limited generalization ability of the model when it encounters new data that deviates from the training set. However, incorporating resliced images into the training set can significantly improve the IoU score to 0.779 by allowing the model to learn and adapt to the diverse appearances of the data.

Evaluated on in-vivo animal data, the IoU score is on average about 8% higher than for the method using only unaugmented images and higher than attainable with other medical image segmentation techniques such as transfer learning with ImageNet (about 5%) and random augmentations (about 23%). Paired t-test performed to assess significance of the observed improvements, yields p-values of 0.034, 0.049 and 0.00046, respectively. Thus, we can confidently conclude that the proposed approach prevalently improves accuracy of image segmentation. Table II summarizes quantitative results.

Our method improves the aforementioned approaches by reslicing a 3D reconstructed volume to create augmented data which is more intuitive and interpretable than performing random transformations for data augmentation. This allows us to re-generate unseen views in the form of raw images, such as a longitudinal slice where all the raw images were transverse images, and create slight variations between vessels that could be seen in a new patient. By generating a variety of reconstructed views we increase the volume and comprehensiveness of training data for the segmentation network, as the single reconstructed volume can generate many different altercations of the vessels and resulting images. Transfer learning, on the other hand, may not be effective when there is a significant difference between the domain images, such as in the case of medical ultrasound images and images from ImageNet. Moreover, it is likely that the RESUS transformations were more realistic and closely matched the characteristics of the data compared to random medical transformations that may not accurately reflect the features present in this particular dataset.

There are several limitations to our work that may have contributed to the seemingly low IoU scores. Firstly, the

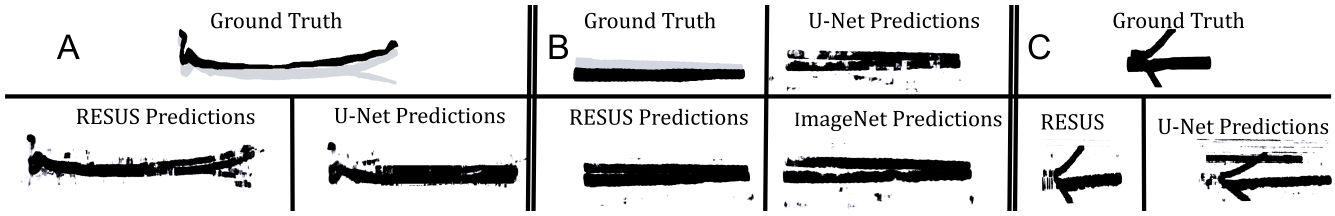

Fig. 5. A) Pig 1 Vessel Reconstruction; B) Pig 4 Vessel Reconstruction; C) Phantom Vessel Reconstruction

labels had many flaws, which affected the success rates of the models. Additionally, the ultrasound data for some pigs was stationary while in others it was zigzagging, making it challenging to create a volume. As a result, we only had five volumes of small sections of the data to train on for the eight pigs. Lastly, the experiments schedule was sparse to allow for technical developments of the prime purpose of the animal study, some of which mildly affected the collected ultrasound data. Even with the aforementioned obstacles, by comparing all qualitative results as shown in Figure 5(A& B) with the different methods, we can observe that the RESUS segmentation is again the closest to the ground truth. By augmenting the training set with data obtained from our method, our vessel reconstruction results better represent the shape and bifurcations of the vessels, and the resulting segmentations are more accurate than alternative approaches under consideration. These improvements can increase the likelihood of successful needle insertions for various kinds of treatments in trauma care in the field.

VI. CONCLUSION

This paper presents RESUS, a novel augmentation method for improving image segmentation specifically in modalities with multiple 2D slices forming a 3D volume, such as ultrasound. Our experiments with in-vivo animal data demonstrate that RESUS outperforms models trained on unaugmented data, models trained on ImageNet, and models trained using commonly used medical imaging augmentation techniques.

The main goal of RESUS is to enable intelligent use of ultrasound imaging as a sensor for robotic control, which is currently limited by the need for large volumes of training data and the high cost of acquiring expert labels. RESUS addresses these challenges by providing affordable and effective image augmentations specifically designed for ultrasound images. This enables high-quality ultrasound image segmentation, which in turn can automate important tasks in critical and trauma care.

Currently, we rely on anatomical knowledge to differentiate between the femoral artery and vein, but in the future, we plan to leverage models of deformation for automatic vein-artery differentiation and integrate our tool with the needle insertion controller. While the IoU score may not be the best evaluation metric for our image segmentation, we consistently achieve the best vessel reconstruction qualitatively. We will explore alternative evaluation metrics in the future.

Finally, the current robotic setup is not yet suitable for deployment in field emergency care due to various practical considerations such as form factor, weight, power needs, sterility, safety, and FDA certification, all of which must be addressed for real-world deployment.

TABLE I

IOU SCORE OF SEGMENTATIONS TESTED ON THE PHANTOM

	Train US	Train US and RESUS
Test US	0.536	**0.731**
Test US and RESUS	0.186	**0.779**

TABLE II

IOU SCORES OF SEGMENTATIONS TESTED ON THE PIGS. COMPARISON OF PERFORMANCES AMONG FOUR MODELS: (1) TRAINED ON JUST US, (2) PRETRAINED ON IMAGENET AND TRAINED ON US, (3) TRAINED ON US AND RANDOM MEDICAL AUGMENTATIONS, AND (4) MODEL TRAINED ON US PLUS THE IMAGES CREATED BY RESUS.

	U-Net	ImageNet	Rnd Aug	RESUS
Pig 1 ($n = 850$)	0.3685	0.591	0.383	**0.610**
Pig 2 ($n = 951$)	0.587	**0.607**	0.537	0.603
Pig 3 ($n = 1307$)	0.631	0.691	0.324	**0.693**
Pig 4 ($n = 900$)	0.513	0.590	0.350	**0.665**
Pig 5 ($n = 900$)	**0.665**	0.408	0.319	0.610
Pig 6 ($n = 500$)	0.7483	**0.785**	0.536	0.7529
Pig 7 ($n = 2070$)	0.422	0.509	0.262	**0.571**
Pig 8 ($n = 1200$)	0.574	0.531	0.555	**0.601**
AVERAGE	0.557	0.592	0.408	**0.638**

ACKNOWLEDGMENTS

We would like to thank Nico Zevallos, Dr. Michael R. Pinsky, and Dr. Hernando Gomez for gathering the data for our experiments, and Mononito Goswami for thoughtful suggestions on this manuscript.

REFERENCES

[1] K. Verhoeff, R. Saybel, P. Mathura, B. Tsang, V. Fawcett, and S. Widder, "Ensuring adequate vascular access in patients with major trauma: a quality improvement initiative," *BMJ Open Quality*, vol. 7, no. 1, p. e000090, Jan 2018. [Online]. Available: https://qir.bmj.com/lookup/doi/10.1136/bmjoq-2017-000090

[2] C. Buckenmaier and P. F. Mahoney, *Vascular Access and Infusion Devices for Combat Anesthesia*. Office of the Surgeon General, United States Army, 2015, p. 63–73.

[3] R. Goel, F. Abhimanyu, K. Patel, J. Galeotti, and H. Choset, "Autonomous ultrasound scanning using bayesian optimization and hybrid force control," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE Press, 2022, p. 8396–8402. [Online]. Available: https://doi.org/10.1109/ICRA46639.2022.9812410

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.

[5] F. Mohamed and C. V. Siang, "A survey on 3d ultrasound reconstruction techniques," in *Artificial Intelligence*, M. A. Aceves-Fernandez, Ed. Rijeka: IntechOpen, 2019, ch. 4. [Online]. Available: https://doi.org/10.5772/intechopen.81628

[6] X. Chen, T. Wen, X. Li, W. Qin, D. Lan, W. Pan, and J. Gu, "Reconstruction of freehand 3D ultrasound based on kernel regression," *Biomed Eng Online*, vol. 13, p. 124, Aug 2014.

[7] Y. Sugano, S. Onogi, A. Bossard, T. Mochizuki, and K. Masuda, "Development of a 3d reconstruction of blood vessel by positional calibration of ultrasound probe," in *The 5th 2012 Biomedical Engineering International Conference*, 2012, pp. 1–4.

[8] K. Masuda, A. Bossard, Y. Sugano, T. Kato, and S. Onogi, "Reconstruction and error detection of blood vessel network from ultrasound volume data," in *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, 2013, pp. 497–501.

[9] Q. Huang, Y. Zheng, M. Lu, and Z. Chi, "Development of a portable 3d ultrasound imaging system for musculoskeletal tissues," *Ultrasonics*, vol. 43, no. 3, pp. 153–163, 2005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0041624X04002185

[10] X. X. Yin, L. Sun, Y. Fu, R. Lu, and Y. Zhang, "U-Net-Based Medical Image Segmentation," *J Healthc Eng*, vol. 2022, p. 4189781, 2022.

[11] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, "U-net and its variants for medical image segmentation: A review of theory and applications," *IEEE Access*, vol. 9, p. 82031–82057, 2021. [Online]. Available: http://dx.doi.org/10.1109/ACCESS.2021.3086020

[12] M. Jiang, J. D. Spence, and B. Chiu, "Segmentation of 3d ultrasound carotid vessel wall using u-net and segmentation average network," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, 2020, pp. 2043–2046.

[13] C. Barr, R. Hisey, T. Ungi, and G. Fichtinger, "Ultrasound probe pose classification for task recognition in central venous catheterization," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, 2021, pp. 5023–5026.

[14] A. I. Chen, M. L. Balter, T. J. Maguire, and M. L. Yarmush, "Deep learning robotic guidance for autonomous vascular access," *Nature Machine Intelligence*, vol. 2, pp. 104–115, 2020.

[15] L. J. Brattain, T. T. Pierce, L. A. Gjesteby, M. R. Johnson, N. D. DeLosa, J. S. Werblin, J. F. Gupta, A. Ozturk, X. Wang, Q. Li, B. A. Telfer, and A. E. Samir, "AI-Enabled, Ultrasound-Guided Handheld Robotic Device for Femoral Vascular Access," *Biosensors (Basel)*, vol. 11, no. 12, Dec 2021.

[16] E. Smistad, K. F. Johansen, D. H. Iversen, and I. Reinertsen, "Highlighting nerves and blood vessels for ultrasound-guided axillary nerve block procedures using neural networks," *J. Med. Imaging (Bellingham)*, vol. 5, no. 4, p. 044004, Oct. 2018.

[17] E. Chen, H. Choset, and J. Galeotti, "Uncertainty-based adaptive data augmentation for ultrasound imaging anatomical variations," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, 2021, pp. 438–442.

[18] B. Sekachev, N. Manovich, M. Zhiltsov, A. Zhavoronkov, D. Kalinin, B. Hoff, TOsmanov, D. Kruchinin, A. Zankevich, DmitriySidnev, M. Markelov, Johannes222, M. Chenuet, a andre, telenachos, A. Melnikov, J. Kim, L. Ilouz, N. Glazov, Priya4607, R. Tehrani, S. Jeong, V. Skubriev, S. Yonekura, vugia truong, zliang7, lizhming, and T. Truong, "opencv/cvat: v1.1.0," Aug. 2020. [Online]. Available: https://doi.org/10.5281/zenodo.4009388

[19] Stanford Artificial Intelligence Laboratory et al., "Robotic operating system." [Online]. Available: https://www.ros.org

[20] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, p. 381–395, jun 1981. [Online]. Available: https://doi.org/10.1145/358669.358692

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[22] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019, pp. 8024–8035. [Online]. Available: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

[23] P. Iakubovskii, "Segmentation models pytorch," 2019.

[24] J. Howard *et al.*, "fastai," https://github.com/fastai/fastai, 2018.

[25] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2015. [Online]. Available: http://arxiv.org/abs/1412.6980