Automated Real Time Delineation of Supraclavicular Brachial Plexus in Neck Ultrasonography Videos: A Deep Learning Approach

ABSTRACT

Peripheral nerve blocks are crucial to treatment of post-surgical pain and are associated with reduction in perioperative opioid use and hospital stay. Accurate interpretation of sono-anatomy is critical for the success of ultrasound (US) guided peripheral nerve blocks and can be challenging to the new operators. This prospective study enrolled 227 subjects who were systematically scanned for supraclavicular and interscalene brachial plexus in various settings using three different US machines to create a dataset of 227 unique videos. In total, 41,000 video frames were annotated by experienced anaesthesiologists using partial automation with object tracking and active contour algorithms. Four baseline neural network models were trained on the dataset and their performance was evaluated for object detection and segmentation tasks. Generalizability of the best suited model was then tested on the datasets constructed from separate US scanners with and without fine-tuning. The results demonstrate that deep learning models can be leveraged for real time segmentation of supraclavicular brachial plexus in neck ultrasonography videos with high accuracy and reliability. Model was also tested for its ability to differentiate between supraclavicular and adjoining interscalene brachial plexus. The entire dataset has been released publicly for further study by the research community.

1.INTRODUCTION

Peripheral nerve blocks (PNB) are a critical component of multimodal approach for treating post-surgical pain. When compared to opioids alone, PNBs offer superior pain control and are associated with reduced perioperative opioid consumption, improved overall patient satisfaction and reduced hospital stay. [1,2] Over the past decade, use of ultrasound (US) has dramatically increased the scope of regional anesthesia. Ultrasound technology offers the ability to approach a target in real-time, enabling precise deposition of drug with increased success rates, reduced local anesthetic volume and fewer complications.[3] The success of US guided nerve blocks in turn, depends on accurate interpretation of sono-anatomy which can often be daunting for novice operators.[4] The classical approach of didactics, cadaveric demonstrations and 1:1 illustration are known to be cost-ineffective and time consuming. Simulation based techniques such as embedded electronic tutorials in US machines have consistently shown to be beneficial in enhancing sono-anatomical knowledge among trainees.[3,5,6] However, these techniques remain sporadic in practice with limitations in scope and can benefit from newer technologies.

Among the PNB targets to provide analgesia for upper limb surgeries, the two most commonly employed *bread and butter* techniques include brachial plexus blockade at interscalene and supraclavicular space.[7] These blocks target nerve roots and their divisions early on in their course, providing rapid and reliable anesthesia with minimal

needling and patient discomfort. The interscalene brachial plexus (ISC) is blocked at the level of nerve roots exiting from the cervical spinal cord and provides analgesia to shoulder region and upper arm whereas supraclavicular nerve block (SCBP) is performed inferiorly, lateral to the clavicular head of the sternocledomastoid muscle and provides analgesia to the upper arm, forearm and hand.[7,8] When performing these blocks under real-time US guidance, the clinician looks for distinct sono-anatomical patterns, such as, cluster of hypoechoic nodules or *bunch of grapes* next to subclavian artery for SCBP and a vertical chain of hypoechoic nodules or a *traffic light sign* sandwiched between two scaleni muscles for ISC. [9] However, the aforementioned patterns only serve as general guide and can vary considerably depending upon individual patient's anatomy and US probe orientation. Moreover, these two sites are situated very close to each other (within a few centimeters) in the neck and transition seamlessly into one another making it challenging to distinguish for novice operators.

In recent years, convolutional neural networks (CNN) utilizing large sized training datasets have shown remarkable success for object detection and segmentation in the field of medical imaging. [10,11,12] In regional anesthesia and musculoskeletal US, attempts have been made to identify individual nerves, muscles and blood vessels using CNNs.[13-18] However, most of these studies are limited by small sized training datasets that affects their accuracy and generalizability.

In 2016, a dataset containing 5,365 manually annotated US images of the ISC was made publicly available by *Kaggle* to train machine learning models for identification of the brachial plexus [13]. This dataset was used by Baby et al.[14] and Kakade et al. [15] to train and validate a U-Net model [19] with image preprocessing steps and achieved dice coefficient scores of 0.71 and 0.69 respectively. In 2021, Wu et al.[16] adopted Region-aware pyramid aggregation model with Adaptive pyramid fusion to achieve a dice coefficient of 0.74 with real-time performance. In 2022, Ding et al.[17] provided a separate dataset of 1,052 manually annotated US images of ISC with best model performance achieving dice coefficient of 0.51 and a runtime of 5.2 FPS.

Ultrasonography is generally considered a challenging modality for image segmentation mainly because 1.) It generates poor contrast images which are dependent on operator skills for image acquisition as well as interpretation, 2.) The speckle noise characteristic to US imaging yields a granular texture in the images which reduces the overall SNR [20]. 3.) In addition, for automated real-time image segmentation, the hazy boundaries and speckle noise further introduce additional time-complexity hurdles on image preprocessing. Image segmentation on the brachial plexus sites also demands minimizing false positives, since inadvertent needle injury to the high stakes surrounding structures such as lung or blood vessels can be catastrophic.

Supraclavicular brachial plexus blockade, even though closely anatomically related to ISC brachial plexus, provides anesthetic coverage to a distinct area and is devoid of

important side effects related to ISC block such as phrenic nerve paralysis and Horner syndrome.[7] However, despite its advantages and popularity among clinicians, SCBP segmentation has not been researched yet by the machine learning community. Furthermore, the two being closely related, there is a need to differentiate between these two adjacent structures for computer assisted diagnosis.

In the present study, we present an end-to-end DL approach for real time delineation of SCBP in neck US videos.This prospective study recruited 163 patients who were systematically scanned for SCBP using US. In addition, we also created a dataset of 33 ISC videos to evaluate the capability of the models to differentiate between SCBP and ISC. The resultant videos were then split into individual frames that were annotated by experienced anesthesiologists. A set of standard CNN models were trained as baseline models and their outputs were compared to the originally annotated images. In addition, two separate datasets of 15 videos each were constructed using different US scanners to test for generalizability of the model.

Our contributions can be summarized as follows:
1- We build a large dataset of brachial plexus ultrasound videos with segmentation annotations done by experienced anesthesiologists for all the frames in each video. To the best of our knowledge, this is also the largest ultrasound video dataset in terms of number of frames annotated.
2 - We built models capable of differentiating the two adjacent nerve blocks - SCBP and ISC in real-time.

3 - We demonstrate generalizability of our model on data acquired through different ultrasound machines by different operators.

2. METHODS

2.1 Dataset curation

This was a prospective study conducted jointly at the All India Institute of Medical Sciences, New Delhi, India and the International Institute of Information Technology, Hyderabad, India. This study follows the 'Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research'.[21] Institutional ethics committee approval was obtained prior to commencement of the study (IEC/18/1/2020,RP-02/2020). Patients of age 20-80 years were included in athis study. Exclusion criteria were patients known to have brachial plexus injury, brachial plexopathy, cervical trauma, limitations to neck movement or patient refusal. Patients were randomly approached in the preoperative holding area and were presented with written information regarding the study. After obtaining written informed consent, they were scanned with a linear high frequency probe of 12-15 MHz of US scanner eSaote MyLab One™(eSaote SPA, Genoa, Italy). Subjects were scanned on a 30-45 degree incline, with the head turned away from the side of interest with the overarching goal of focusing on the brachial plexus in the supraclavicular and interscalene space. US

probes were then tilted, rotated and slid around in the cranio-caudal and medial to

lateral directions to capture dynamic brachial plexus patterns in addition to adjoining

areas of the neck such as muscles, fascia and blood vessels.

Random low, medium and high gain as well as different depth settings were used to

increase the variation in the dataset. Table 1 summarizes the patient demographic

characteristics. All personal identifiable information was cropped. The resulting videos

were 542x562 pixel resolution.

Table 1. Data Set Patient Characteristics

Characteristic	Total	Supraclavicular	Interscalene
No. of Patients	196	157	33
No. of Videos*	196	157	33
Frames	34,926	29,129	4,622
Positive Data Frames	26,966	23,302	3,664
Gender			
Male	134 (68%)	107 (68%)	23(70%)
Female	62 (32%)	50 (32%)	10(30%)
Average Age	42 (13.6)	42 (13.5)	42 (13.9)
Average Height	170 (6.1)	171 (5.7)	169 (6.8)
Average BMI	25.3 (2.6)	25.4 (2.6)	25.2 (2.7)

Data is reported as n(%) or mean(±SD)
Positive data frames imply the presence of SCBP or ISC in a given frame.
6 videos contained only negative data and used only for training

Annotation

A custom video annotation tool was built using OpenCV software[22] for the purpose of

annotation, details of which have been uploaded on the GitHub website[23]. In a given

video, a random frame was annotated for SCBP by an expert anesthesiologist using

bounding boxes and was designated as seed frame. The bounding boxes were drawn to cover the approximate SCBP area and were not meant to annotate individual roots. In addition, boxes were drawn keeping in mind to stay clear of critical structures such as the subclavian artery and pleura in all frames. Subsequent frames were automatically annotated using the CSR-DCF and KCF object-tracking algorithms.[24,25] Each automated annotation was either approved or rejected by the annotator. If rejected, the bounding boxes were redrawn as new seeds for automated annotations. The individual boxes were then fused to create a single mask. The contour of this mask was then shrunk to better match the boundaries of the plexus using morphological geodesic active contour algorithm [26]. The parameters of the algorithm were chosen for each individual video separately and are available in the dataset. The final shrunken masks were labeled as "ground truth" masks (Fig. 1) which were used to train the neural network. The frames in which the primary annotator was uncertain, a second expert was called in to review the frames. If still uncertain of the presence or absence of SCBP or ISC, these frames were discarded to avoid noisy annotations.
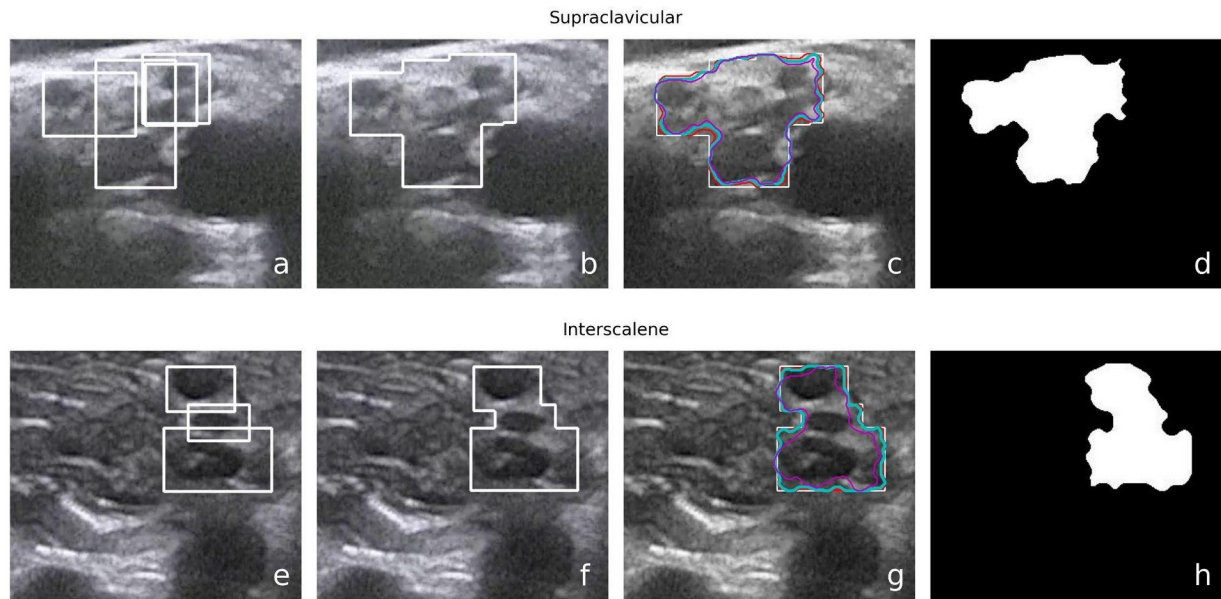
Fig 1. a,e) Hand drawn annotation bounding boxes to cover the plexus as a whole while avoiding critical structures such as the pleura and subclavian artery b,f) outer boundary of the union of boxes selected as the initial contour c,g) Various shrunken contours proposed by the morphological geodesic active contour algorithm using different parameters d,h) Best contour approximating the boundaries of the plexus chosen by experienced anaesthesiologists as the final ground truth

2.2 Data augmentation

Three data augmentations with random parameters were employed to enrich variations in the training data.

- Horizontal flip: Images were horizontally flipped with a 50% probability to mimic the left-right dexterity in SCBP sono-anatomical patterns.

- Rotation: Images were rotated by a random angle between -10 and 10 degrees to mimic the incline-angle variations of the US probe.

- Gamma intensity transformation [27]: The image intensities were transformed with random gamma values to mimic variations in US frequency gain. The original high gain images were transformed with a random gamma value between 1.5 and 2.0 to mimic medium and low gain images. The original low

gain images were transformed with a random gamma value between 0.5 and

0.75 to mimic high and medium gain images. The original medium gain images

were transformed with a random gamma value between 0.75 and 1.33 to mimic

high and low gain images.

2.3 Neural Network

From a wide range of available off-the-shelf segmentation models, we shortlisted two

models - DeepLabv3 [28] and UNet [19], and their enhancements - DeepLabv3+[29]

and UNet++ [30] based on established usage in medical image segmentation literature

[31, 32], real-time processing capability and low memory usage. All the four models use

50-layer residual-net as backbone[33]

One of the limitations while using traditional fully convolutional neural networks for

image segmentation tasks is that the spatial resolution is progressively downsampled

resulting in information loss and low resolution output masks with fuzzy object

boundaries. The DeepLabV3 model addresses this by using Atrous convolutions[34]

and Atrous Spatial Pyramid Pooling (ASPP)[35-37] which help in extracting information

with a wider receptive field and thus avoiding the loss of spatial resolution. The UNet

model mitigates the downsampled resolution problem through up-convolutions on the

expanding path and concatenations with features from the downsampling path, resulting

in a symmetrical U shaped network.

Model design and training was done in the Python programming language using the

PyTorch 1.8 deep learning framework[38]. The models were initialized with pretrained

Imagenet weights [39] and were trained using Adam optimizer[40] for 25-50 epochs with early stopping and batch size 16 on an Ubuntu 20.04 machine with an NVIDIA RTX 3090 GPU. Training took approximately 9 minutes per epoch and minimized pixel-level binary cross-entropy loss.

.

2.4 Internal validation:

For internal model validation, stratified 5-fold cross-validation was performed. The training:validation:test proportions were 61:19:20 for each fold. The stratified samples were chosen to keep roughly equal proportions for dexterity (left/right), frequency gains (high/medium/low), and gender (male/female) in each of the training,validation and test sets.

2.5 Generalizability

For testing generalizability of the model, two separate datasets of 15 videos each were constructed using different US machines. The first dataset was created using Sonosite M-Turbo™ (Fujifilm Sonosite Inc, Bothell, WA, USA) which is a cart based US scanner that produces videos of similar quality to the original machine in the main dataset. Six of the 15 videos were acquired with needle in-situ to test the performance of the model in the presence of needle. These 15 videos contained a total of 2,186 frames.

The second dataset was created using a relatively inexpensive portable handheld

machine Butterfly-IQ™ (Butterfly Network Inc., Burlington, MA, USA) which produces

videos of lower quality [41,42]. The 15 videos from the second machine contained a

total of 3,088 frames.

Models were evaluated on these datasets with and without fine-tuning.

2.6 Evaluation

The goal of this study is not pixel-accurate segmentation but rather to identify the

approximate location of the plexus in the US images. Therefore, the performance

evaluation was done using both object detection metrics and segmentation metrics.

For object detection metrics, the precision, recall and F-1 scores were calculated using

intersection over union (IoU) thresholds.[43] The IOU measures the area overlapping

between the prediction mask and the ground truth divided by the area of union between

them. The IOU is then compared to a given threshold (t) depending on which, the

predictions are classified as correct or incorrect. If the IoU ≥ t, then the prediction was

considered correct or true positive. If IOU < t, the detection was considered as incorrect

or false positive. Failure to detect ground truth in a given frame was considered false

negative.Smistad et. al used a 25% IOU threshold to delineate nerves and blood

vessels in the axillary space stating a goal of only highlighting the approximate location

of nerves and not pixel perfect nerve boundary delineation.[18] In computer vision literature, 50% IOU is used as a standard threshold for classifying the presence or absence of an object in an image. The IOU thresholds tested in the current study were set at 25% and 50%. The predictions too small to be regarded as SCBP or ISC were discarded. The threshold for this was set at 20% of the median area occupied by the ground truth in the entire dataset, i.e; 3240 pixels for median SCBP area and 914 pixels for ISC in resized 256x256 images.

The different spatial resolutions of different US machines were brought to common ground of width 256 pixels for comparison. The original dataset and Butterfly-IQ™ dataset were evaluated at 256x256 resolution and Sonosite M-Turbo™ dataset were evaluated at 256x192 resolution.

For segmentation metrics, dice coefficient was calculated in three different ways:

1. Dice coefficient for both plexuses in all frames of all videos: This measured the ability of the model to differentiate the two plexuses. When the model correctly identified the absence of nerve in any image i.e. in the true-negative case, 100% dice-score was assigned for that image. However, this inflated the dice coefficient.

2. Dice coefficient for SCBP was calculated using only SCBP videos and dice coefficient for ISC was calculated using only ISC videos. Here also, the frames with true-negative cases inflated the overall dice scores for the video.

3. Dice coefficients were calculated separately for SCBP and ISC using only the positive frames in SCBP and ISC videos respectively. The frames with true-negative cases were ignored and not taken for consideration.

For all the above metrics, the average metric of all considered videos is the final metric. The metric for individual video is the average metric of all frames in that video.

3. RESULTS

For a total of 163 videos, six did not contain SCBP and were used for training purposes only. In total, 29,129 image frames were annotated with 23,302 positive frames (SCBP present) and 5,827 negative frames (SCBP absent). Prediction runtime of the algorithm was within real-time constraints. The average run-rate was measured to be more than 30 fps using the NVIDIA RTX 3090 GPU.

Table 2, shows the average precision, recall and F score ± standard deviation at 25% and 50% IOU thresholds, and three dice coefficients as described in the Evaluation section for the four baseline models' performance on SCBP and ISC.

Table 2: Comparison of baseline models on Object Detection & Segmentation metrics

	DeepLabV3	DeepLabV3+	UNet	UNet++
Supraclavicular BP (SCBP)				
Object Detection at 25% IOU				
Precision	0.93 (0.17)	0.94 (0.15)	0.94 (0.16)	0.95 (0.16)
Recall	0.96 (0.07)	0.97 (0.13)	0.96 (0.07)	0.97 (0.08)
F-1 score	0.94 (0.16)	0.95 (0.13)	0.95 (0.15)	0.96 (0.15)
Object Detection at 50% IOU				
Precision	0.86 (0.22)	0.87 (0.23)	0.87 (0.21)	0.88 (0.21)
Recall	0.95 (0.11)	0.96 (0.06)	0.94 (0.09)	0.96 (0.10)
F-1 score	0.90 (0.20)	0.91 (0.20)	0.90 (0.19)	0.92 (0.19)
Dice Coefficient				
All videos	0.81 (0.12)	0.83 (0.12)	0.83 (0.11)	0.84 (0.11)
SCBP videos	0.78 (0.10)	0.80 (0.10)	0.79 (0.09)	0.81 (0.09)
SCBP positive frames	0.74 (0.10)	0.76 (0.10)	0.75 (0.09)	0.77 (0.09)
Interscalene BP (ISC)				
Object Detection at 25% IOU				
Precision	0.76 (0.44)	0.78 (0.32)	0.79 (0.33)	0.80 (0.36)
Recall	0.88 (0.16)	0.90 (0.22)	0.84 (0.37)	0.88 (0.26)
F-1 score	0.82 (0.44)	0.84 (0.31)	0.81 (0.34)	0.84 (0.36)
Object Detection at 50% IOU				
Precision	0.57 (0.36)	0.59 (0.31)	0.59 (0.35)	0.60 (0.36)
Recall	0.85 (0.28)	0.87 (0.23)	0.79 (0.36)	0.81 (0.34)
F-1 score	0.68 (0.36)	0.70 (0.32)	0.68 (0.37)	0.69 (0.36)
Dice Coefficient				
All videos	0.93 (0.15)	0.94 (0.16)	0.93 (0.15)	0.94 (0.15)
ISC videos	0.64 (0.19)	0.69 (0.19)	0.65 (0.19)	0.68 (0.19)
ISC positive frames	0.59 (0.20)	0.64 (0.19)	0.61 (0.20)	0.63 (0.19)

Data Reported as mean(±SD)
All models had ResNet-50 backbone

Figure 2 depicts the precision-recall curves at 25% and 50% IOU thresholds for UNet++
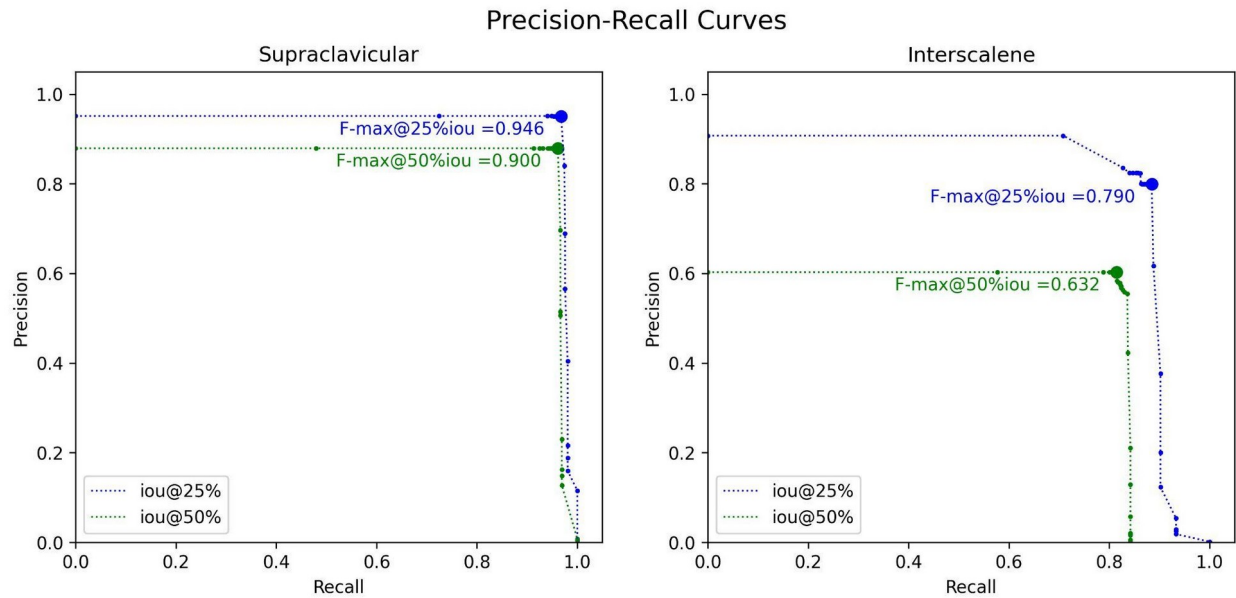
model on SCBP and ISC.

Figure 2. Precision-Recall Curves for IoU overlap criteria of 25% and 50%

Table 3 shows the performance of UNet++ model before and after fine-tuning on the

datasets created through Sonosite M-Turbo™ which is a similar quality machine and a

lower quality handheld machine i.e. Butterfly IQ™ scanner.

Table 3: UNet++ performance on different machines

	Sonosite M-Turbo™		Butterfly IQ™	
	Before FT	After FT	Before FT	After FT
Object Detection at 25% IOU				
Precision	0.99 (0.03)	0.99 (0.02)	0.77 (0.19)	0.96 (0.02)
Recall	0.96 (0.06)	0.99 (0.04)	0.76 (0.20)	0.98 (0.03)
F-1 score	0.97 (0.04)	0.99 (0.03)	0.76 (0.19)	0.97 (0.03)
Object Detection at 50% IOU				
Precision	0.80 (0.24)	0.95 (0.05)	0.23 (0.26)	0.60 (0.25)
Recall	0.90 (0.09)	0.98 (0.04)	0.42 (0.40)	0.96 (0.06)
F-1 score	0.85 (0.19)	0.96 (0.04)	0.30 (0.30)	0.74 (0.19)
Dice Coefficient				
All frames	0.69 (0.09)	0.78 (0.03)	0.55 (0.17)	0.67 (0.08)
Positive frames	0.69 (0.08)	0.77 (0.03)	0.45 (0.16)	0.66 (0.08)

FT: Fine-Tuning
Data is reported as n(%) or mean(±SD)

Supplementary video 1 shows model performance with comparison to ground truth.

Supplementary video 2 illustrates model predictions differentiating SCBP and ISC with

US probe movement along the neck.

Supplementary video 3 shows model interaction with needle insertion.

4. DISCUSSION

In the present work, we demonstrate a DL based approach that successfully delineates and tracks the SCBP across the given US videos of the neck with high accuracy and reproducibility. The prediction rate was greater than 30 fps, which is well above the real-time constraints. When applied to an unseen separate dataset constructed using a separate US scanner of similar quality, the model performance remained consistent. To the best of our knowledge, this is the first DL based approach trained on a large dataset that delineates SCBP in US videos with high precision and reliability. In addition, the model trained jointly on SCBP and ISC was able to successfully differentiate between SCBP and adjacent ISC as shown in Supplementary video 2.

Over the past two decades, ultrasonography has rapidly ascended to become indispensable in clinical practice due to its unique set of advantages, such as absence of ionizing radiation, portability and ability to produce real-time imaging which can be used for both diagnostics and interventions. With continuous advancement in computational power to give out faster results, the idea of leveraging DL to augment US experience is getting closer to realization, wherein algorithm segmentations can be superimposed upon US videos in real time that may assist clinicians in making decisions. However, for a robust training of DL algorithms and to achieve desired performance, a large training dataset is usually needed, that needs to be manually annotated by experts, often proving to be a laborious task. A common roundabout is

thus taken where human experts annotate one or two US images per clip to create a small to moderate sized ground truth dataset. This approach leaves DL models at a disadvantage because it does not incorporate the dynamic patterns of neural structures, as in our case SCBP and ISC, that continually change in shape and size with the US probe movement. Furthermore, *transfer learning*, a technique used to compensate for small amounts of training data is commonly applied wherein pre-trained models that identify real world natural image objects are further trained on small sized medical image datasets.[44, 45] Although successful to some extent, this leads to poor generalization to US images as the speckle noise seen in grayscale US images is not present in natural images. Also, the structures in US images tend to have fuzzy boundaries compared to well-defined boundaries seen in natural image objects. In the current study, we partially automated annotation using object tracking and active contour algorithms that helped anesthesiologists annotate a seemingly large image dataset in a comprehensive manner within a short period of time. Scanning protocols were prospectively laid out to capture variable ISC and SCBP patterns seen with different US probe orientations. Instead of using one or two images per video, sequential frames were annotated in different depth and gain settings for exhaustive training of our CNNs that we believe greatly adds to the accuracy and robustness of our approach.

Upon review of literature, various authors have attempted to delineate brachial plexus in ultrasound images at the interscalene and axillary spaces using machine learning. Utilizing the 5,365 training images of *Kaggle* dataset, Baby et al. [14] and Kakade et al. [15] implemented the UNet model to achieve dice coefficient of 0.71 and 0.69 respectively, but evaluation on the test dataset could not be done since the test dataset annotations were not released by Kaggle. Both of these approaches utilized time consuming image-processing steps that made the real-time inference infeasible. Wu et al improved upon this with Region-aware pyramid aggregation model and Adaptive pyramid fusion to achieve a dice coefficient of 0.74 at real-time performance [ref]. Ding et al. [17] created a new data set of 1,052 frames from 101 videos and publicly released its' training dataset of 955 frames from 91 videos containing annotations for ISC as well as surrounding anatomical structures i.e. muscles and blood vessels. They improved upon Mask-RCNN through a spatial local contrast feature extraction module and self-attention gate thereby utilizing the spatial information of surrounding anatomical structures and achieved a dice coefficient of 0.51 at 5.19 FPS. In comparison, a standard of-the-shelf UNet++ model trained on our 3,664 ISC frames from 33 videos achieved a dice coefficient of 0.68 at more than 30 FPS. However, the higher score in our study could be attributed to different image characteristics and bigger dataset in our study. Smistad et al.[18] used U-Net to delineate individual brachial plexus nerves in axillary space and reported an F-score ranging from 0.39 (radial nerve) to 0.73 (median

nerve) using 25% IoU overlap criterion with real-time inference capability. Our results for

SCBP showed a superior F-score of 0.95 and 0.90 for 25% and 50% IOU threshold

respectively.

It is noteworthy that none of the abovementioned studies evaluated their models for

generalizability by testing on a separate dataset curated from other US scanners. This

could be an important limitation since different US scanners tend to have different

image characteristics. The present study tested its model on two separate scanners - a

traditional cart based Sonosite M-Turbo™ with superior image quality (Fujifilm Sonosite

Inc, Bothell, WA, USA) and a hand held portable Butterfly IQ scanner that can be

plugged into smartphones and run with an app with a compromise of a lower image

quality. Our results showed a dice coefficient of 0.69 and 0.55 for delineating SCBP on

images generated from Sonosite and Butterfly US scanners respectively, without any

additional fine-tuning . The dice coefficient scores further improved to 0.78 and 0.67

respectively after fine-tuning.

The present study has its limitations. Our CNN was trained from a set of US videos

obtained by four anesthesiologists trained in regional anesthesia according to a well

defined protocol focusing only on two specific areas of the patient's neck, i.e; the

supraclavicular and interscalene spaces using one US scanner. Variations may come to

model performance in diverse clinical settings with different operators using different US

scanners as depicted in Table 3. That being said, performance on a new machine or patient population can be improved with fine-tuning on a small amount of annotated data from the concerned machine or population . Rapid and jerky probe movements may result in erroneous predictions that tend to improve when the probe movements are slowed and stabilized over one area. This is because of the motion artifacts that cause blurring of structures leading to erroneous interpretations. The demographic characteristics of our training set consisted mostly of average built subjects which may again lead to variable model performance when used on extremes of body habitus.

As an extension to our scanning protocol, a few videos were captured to assess model performance with needle in situ while performing nerve block as shown in supplementary video 3. The model performance remained consistent in the presence of the needle but dropped significantly when the needle entered the confines of the plexus. This may be explained as the entry of the needle changes the plexus anatomy that has not been incorporated in the training of models. This finding underscores the need for future studies and datasets that include videos of PNB being performed with needle in situ to train models that can further improve their performance in such conditions.

One of the major obstacles for DL based studies is acquiring a high fidelity dataset containing a well labeled ground truth that experts agree upon. As a resource for future

studies to train and validate their models, we are releasing our full dataset of 227 videos consisting of 41,000 annotated images of SCBP and ISC that can be accessed on the GitHub website.[23] To the best of our knowledge, this is one of the largest datasets of US images annotated by human experts for SCBP and ISC that could help future DL studies on ultrasonography. Our combined SCBP and ISC US video dataset further complements previous available datasets which have consisted only of isolated US video frames. We hope that further video based models that utilize temporal information can be developed by the research community using this dataset.

In conclusion, we provide a DL based approach for real time delineation of SCBP in US videos with high accuracy and reliability. This model may be incorporated in the US scanners that could potentially assist regional anesthesia trainees to administer US-guided SCBP nerve block. We also believe the contribution of an open labeled dataset will greatly enhance the development of new deep learning applications in regional anesthesiology and musculoskeletal ultrasonography.

Dataset and code availability:

The dataset and code used for annotation in this work can be freely downloaded from:

https://github.com/Regional-US/

Competing interests:

The author(s) declare no competing interests

References

[1] Joshi, G. Gandhi, K. Shah, N. Gadsden, J. & Corman, S.L. Peripheral nerve blocks in the management of postoperative pain: challenges and opportunities. *J Clin Anesth*. 2016; **35**:524-529 (2016).

[2] Chan, E.Y. Fransen, M. Parker. D.A. Assam, P.N. & Chua, N. Femoral nerve blocks for acute postoperative pain after knee replacement surgery. *Cochrane Database Syst Rev*. 2014 **(5)**:CD009941. (2014)

[3] Chen, X. X., et al. Ultrasound-Guided Regional Anesthesia Simulation Training: A Systematic Review. *Regional anesthesia and pain medicine*, *42*(6), 741–750 (2017).

[4] Bowness, J. El-Boghdadly, K. Burckett-St Laurent D. Artificial intelligence for image interpretation in ultrasound-guided regional anaesthesia. *Anaesthesia* **76**:602–7 (2021).

[5] Wegener, J.T. et al. Value of an electronic tutorial for image interpretation in ultrasound-guided regional anesthesia. *Reg Anesth Pain Med*. **38(1)**:44-49 (2013).

[6] Nix, C. M., et al. A scoping review of the evidence for teaching ultrasound-guided regional anesthesia. *Regional anesthesia and pain medicine*, *38***(6)**, 471–480. (2013).

[7]Neal, J.M. et al. Upper Extremity Regional Anesthesia. *Regional Anesthesia & Pain Medicine* **34**:134–70. (2009)

[8] Lanz, E., Theiss, D. & Jankovic, D. The Extent of Blockade Following Various Techniques of Brachial Plexus Block. In: Van Kleef, J.W., Burm, A.G.L., Spierdijk, J. (eds) Current Concepts in Regional Anaesthesia. Developments in Critical Care Medicine and Anaesthesiology, vol 7. Springer, Dordrecht.(1984)

[9] Yang, W.T. Chui, P.T. Metreweli, C. Anatomy of the normal brachial plexus revealed by sonography and the role of sonographic guidance in anesthesia of the brachial plexus. *American Journal of Roentgenology* **171**:1631–6 (1998)

[10] Litjens, G. et al. A survey on deep learning in medical image analysis. *Med. Image Anal.* **(42)** 60-88 (2017)

[11] Shen, D. Wu, G. Suk, H. Deep Learning in Medical Image Analysis. *Annu Rev Biomed Eng.* (**19**) 221–248 (2017)

[12] Ker, J. Wang, L. Rao, J. & Lim T. Deep Learning Applications in Medical Image Analysis. *IEEE Access* (**6**) 9375-9389 (2018)

[13] https://www.kaggle.com/c/ultrasound-nerve-segmentation/. Accessed on November 30, 2022

[14] Baby, M. & Jereesh, A.S. Automatic nerve segmentation of ultrasound images. In *2017 International Conference of Electronics, Communication and Aerospace Technology (ICECA)* **(1)** 107-112 (2017)

[15] Kakade, A. & Dumbali, J. Identification of nerve in ultrasound images using u-net architecture. In *2018 International Conference on Communication information and Computing Technology (ICCICT)*, 1–6 (2018)

[16] Wu, H., Liu, J., Wang, W., Wen, Z., & Qin, J. Region-aware Global Context Modeling for Automatic Nerve Segmentation from Ultrasound Images. *Proceedings of the AAAI Conference on Artificial Intelligence* **35(4)**, 2907-2915 (2021)

[17] Ding, Y. et al. MallesNet: A multi-object assistance based network for brachial plexus segmentation in ultrasound images. *Med. Image Anal.* **80** (2022).

[18] Smistad, E. Johansen, K.F. Iversen, D.H. & Reinertsen, I. Highlighting nerves and blood vessels for ultrasound-guided axillary nerve block procedures using neural networks. *Journal of Medical Imaging*. **5(4)**:044004 (2018)

[19] Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2015*, 234-241 (2015)

[20] Kapoor, A. & Singh, T. A brief review: Speckle reducing filtering for ultrasound images. *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)* 242-246 (2017)

[21] Luo, W., et al. Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View. *J Med Internet Res* **18(12)**:e323 (2016).

[22] https://opencv.org/ .Accessed on November 30, 2022

[23] https://github.com/Regional-US/ Accessed on November 30, 2022

[24] Lukezic, A. Vojir, T. Zajc, L.C. Matas, J. & Kristan, M. Discriminative Correlation Filter with Channel and Spatial Reliability. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (CVPR)* 6309-6318 (2017)

[25] Henriques, J.F. Caseiro, R. Martins, P. & Batista, J. High-speed tracking with kernelized correlation filters. In *IEEE Transactions on Pattern Analysis and Machine Intelligence* **(3)**:583-96 (2014).

[26] Márquez-Neila, P. Baumela, L. & Álvarez, L. A Morphological Approach to Curvature-based Evolution of Curves and Surfaces. In *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, (2014)

[27] Somasundaram, K. & Kalavathi, P. Medical Image Contrast Enhancement based on Gamma Correction. *International Journal of Knowledge Management and e-Learning*. **3** 15-18. (2011)

[28] Chen, L.C. Papandreou, G. Schroff, F. & Adam, H. Rethinking atrous convolution for semantic image segmentation. Preprint at https://arxiv.org/abs/1706.05587 (2017).

[29] Chen, L.C. Zhu, Y. Papandreou, G. Schroff F. & Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proc. Eur. Conf. Comput. Vis. (ECCV)* 801-818, (2018)

[30] Zhou, Z. Siddiquee, M. M. R. Tajbakhsh, N. & Liang, J. "UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Trans. Med. Imag.* (2019)

[31] Liu, X., Song, L., Liu, S., & Zhang, Y. A review of deep-learning-based medical image segmentation methods. *Sustainability* **13(3)**, 1224 (2021)

[32] Wang, R. et al. Medical image segmentation using deep learning: A survey. *IET Image Process*. **16**, 1243–1267 (2022)

[33] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 770-778).

[34] Chapter: Holschneider, M., Kronland-Martinet, R., Morlet, J., Tchamitchian, P.: A real-time algorithm for signal analysis with the help of the wavelet transform. In: Wavelets:Time-Frequency Methods and Phase Space. 289-297.(Springer, 1989)

[35] Grauman, K. & Darrell, T. The pyramid match kernel: Discriminative classification with sets of image features. In *International Conference on Computer Vision (ICCV)* (2005).

[36] Lazebnik, S. Schmid, C. & Ponce, J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2006)

[37] Chen, L.C. Papandreou, G. Kokkinos, I. Murphy, K. & Yuille, A. L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)* (2016).

[38] Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A. Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems. 2019;32.

[39] Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*.(CVPR) 248–55 (2009)

[40] Kingma, Diederik P. and Ba, Jimmy. Adam: A Method for Stochastic Optimization - In International Conference on Learning Representations (ICLR) 2015

[41] Le, M-P.T. et al. Comparison of four handheld point-of-care ultrasound devices by expert users. The Ultrasound Journal **14** (2022)

[42] Newhouse, S.M. Effing, T.W. Dougherty, B.D. D'Costa, J.A. & Rose, A.R. Is bigger really better? Comparison of ultraportable handheld ultrasound with standard point-of-care ultrasound for evaluating safe site identification and image quality prior to pleurocentesis. *Respiration*. **99(4)**:325–32 (2020)

[43] Padilla, R. Netto, S.L. da Silva, E.A.B. A Survey on Performance Metrics for Object-Detection Algorithms," In *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)* 237-242 (2020)

[44] Lu, J. et al. Transfer learning using computational intelligence: A survey. *Knowledge-based Systems 2015*;**80** 14–23 (2015)

[45] Malik, H. et al. A Comparison of Transfer Learning Performance Versus Health Experts in Disease Diagnosis From Medical Imaging. In *IEEE Access* **(8)** 139367-139386 (2020)

LEGEND

Table 1: Data Set Patient Characteristics

Table 2: Comparison of baseline models on Object Detection & Segmentation metrics

Table 3: UNet++ performance on different machines

Figure 1a and 1e: Hand drawn annotation bounding boxes to cover the plexus as a whole while avoiding critical structures such as the pleura and subclavian artery

Figure 1b and 1f: outer boundary of the union of boxes selected as the initial contour

Figure 1c and 1g: Various shrunken contours proposed by the morphological geodesic active contour algorithm using different parameters

Figure 1d and 1h: Best contour approximating the boundaries of the plexus chosen by experienced anaesthesiologists as the final ground truth

Figure 2: Precision-Recall Curves for IoU overlap criteria of 25% and 50%

Supplementary Video 1 : Algorithm segmentation (prediction mask) in blue compared to ground truth (red)

Supplementary Video 2 : Model predictions differentiating SCBP and ISC with US probe movement along the neck. Ground Truth (red), SCBP mask (blue) and ISC mask (yellow)

Supplementary Video 3 : Model interaction with needle insertion