Automatic segmentation of spinal ultrasound landmarks with U-net using multiple consecutive images for input

Victoria Wu[1], Tamas Ungi[1], Kyle Sunderland[1], Grace Pigeau[1], Abigael Schonewille[1], Gabor Fichtinger[1, 2]

[1]Laboratory for Percutaneous Surgery, School of Computing, Queen's University, Kingston, Canada
[2]Department of Surgery, School of Medicine, Queen's University, Kingston, Canada

ABSTRACT

PURPOSE: Scoliosis screening is currently only implemented in a few countries due to the lack of a safe and accurate measurement method. Spinal ultrasound is a viable alternative to X-ray, but manual annotation of images is difficult and time consuming. We propose using deep learning through a U-net neural network that takes consecutive images per individual input, as an enhancement over using single images as input for the neural network.

METHODS: Ultrasound data was collected from nine healthy volunteers. Images were manually segmented. To accommodate for consecutive input images, the ultrasound images were exported along with previous images stacked to serve as input for a modified U-net. Resulting output segmentations were evaluated based on the percentage of true negative and true positive pixel predictions.

RESULTS: After comparing the single to five-image input arrays, the three-image input had the best performance in terms of true positive value. The single input and three-input images were then further tested. The single image input neural network had a true negative rate of 99.79%, and a true positive rate of 63.56%. The three-image input neural network had a true negative rate of 99.75%, and a true positive rate of 66.64%.

CONCLUSION: The three-image input network outperformed the single input network in terms of the true positive rate by 3.08%. These findings suggest that using two additional input images consecutively preceding the original image assist the neural network in making more accurate predictions.

1. PURPOSE

Scoliosis, an excessive curvature of the spine, is a disease that affects approximately 2-3% of the population. It is most commonly found in children and adolescents, where curvature progresses until the spine is fully developed. Severe curves often require surgical fusion of the vertebrae, a long surgical process that can risk further complications and future surgery. One common method to reduce the need for surgery is bracing of the spine to prevent the curve from worsening. However, effective bracing requires early referral, which is often prevented due to the lack of a safe and affordable instrument for spinal curvature measurement. Current methodologies most commonly involve X-rays of the spine, exposing children and adolescents to ionizing radiation, and limiting the accessibility of scoliosis measurement.

One proposed alternative to X-rays for spinal curvature measurement is ultrasound. Ultrasound is a safe and affordable measurement modality, with increasing accessibility to recent release of pocket-sized ultrasound machines. Spinal curvatures can be accurately computed from ultrasounds, by measuring the vertebral angle from a reference line[1]. However, ultrasound images are not as clear as x-ray images; it is often difficult and time consuming to distinguish bone from the rest of the ultrasound image. In order to be able to effectively use the ultrasound images, they need to be manually contoured by a physician, radiologist, or trained expert. This process is often slow and tedious. A viable tool to aid with the segmentation of bone in ultrasound images is convolutional neural networks. Specifically, the U-net is a convolutional neural network architecture that contains a contracting path to capture feature contexts, then an expansive path for more precise localization. The U-net is especially effective for the segmentation of biomedical images as it can generate accurate segmentation based on fewer training images than networks typically used for classification or localization[2]. U-nets have successfully segmented spinal ultrasound images for vertebral level localization[3].

When a person is performing manual segmentation of spinal ultrasounds, we observed that the user performing the segmentation often looks back on previous ultrasound images to acquire general information about the anatomy, and to assess consistency between consecutive images. Recurrent neural networks are a class of neural networks that are suitable for temporal and sequential data; however, they are not as effective as convolutional neural networks for processing and segmenting spatial data. Current convolutional neural network architectures do not take advantage of previous image frames, as these networks typically process each image individually. We propose a network architecture and method of curating the data that allows the U-net convolutional neural network to examine multiple consecutive images as a single input.

2. METHODS

2.1 Experimental setup

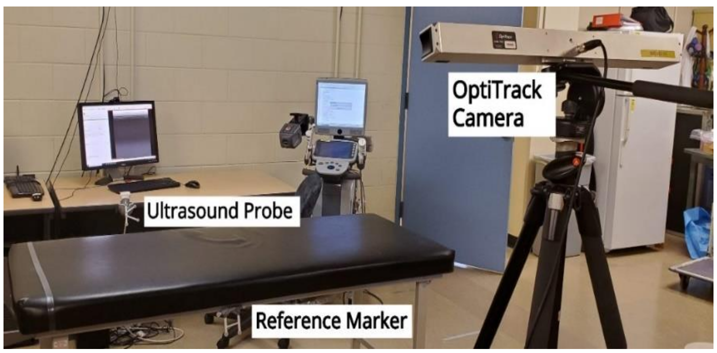


Figure 1: *Left.* The experimental setup of the optically tracked ultrasound system.

The sonographer collected ultrasound data from volunteer subjects, lying in the prone position. Scans were obtained using the curvilinear US probe (Telemed MicrUs EXT-1H, Vilnius, Lithuania). The operator scanned the ultrasound probe down the spine in the sagittal plane. Nine healthy volunteer subjects between the ages of 16 and 30 were scanned. Additionally, females are ten times more likely to suffer from scoliosis than males. This ratio was reflected in the volunteers scanned; in total, 8 females and 1 male were scanned. All image data was acquired using an optically tracked ultrasound. The ultrasound was tracked with an OptiTrack Trio, with tracking markers fixed to the ultrasound probe and also to the bench on which the subjects lay, to act as a reference marker (Figure 1). The OptiTrack Trio and ultrasound machine were connected to a computer using PLUS toolkit [4]. Images were then streamed from PLUS into the open source software 3D Slicer[i], using the SlicerIGT extension [5]. In total, seven of the scans were used for training data, totaling 3,856 images. The additional two scans were used for testing data, giving 717 images. Consequently, the two scans used for testing data had never previously been seen by the network, ensuring a non-bias result.

Figure 2: Scanning of volunteer subject

2.2 Data curation

Image data from all 9 participants were manually segmented using 3D Slicer. We created an extension called *Single Slice Segmentation* for 3D Slicer[ii], which allowed users to contour individual images from ultrasound scans with greater ease. The scans were exported as a series of png images (Figure 3). In addition, the four images preceding each segmented image were exported as well. All the images were converted into numpy arrays for network input. Five separate series of numpy arrays were created based on the exported images. The first series contained the segmentations and the single

[i] https://www.slicer.org/
[ii] https://github.com/SlicerIGT/aigt

corresponding images in one channel. The subsequent arrays contained the segmentations and the corresponding image, along with a number of images preceding it. Arrays were created containing one to four preceding images, giving arrays that contained two to five channels in total. All image processing and data curation was done in Python, using the OpenCV Library[iii].

.

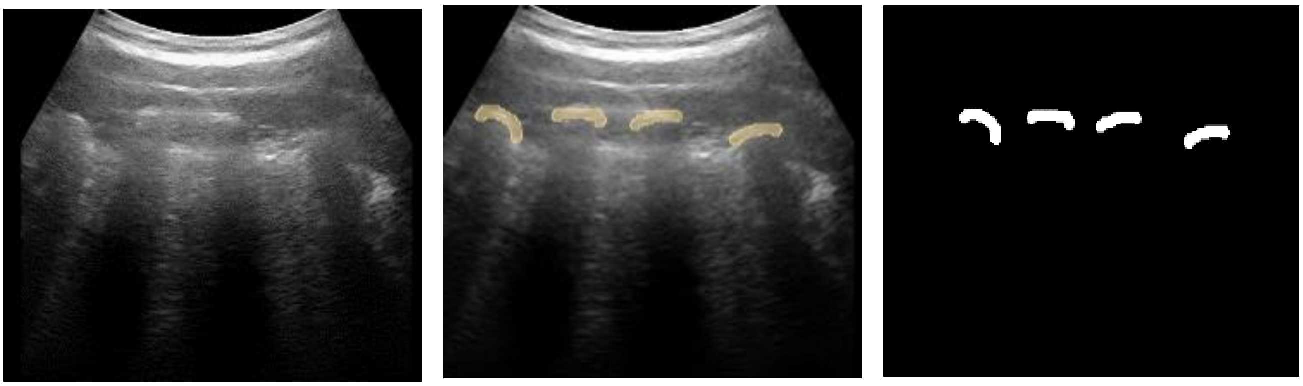


Figure 3: *Left.* Ultrasound image. *Middle.* Segmented ultrasound image. *Right.* Input ground truth segmentation

2.3 Convolutional neural network architecture

The convolutional neural networks used were based on the U-net architecture [2]. The U-net was trained and tested with the single input images, then altered to be able to take in inputs with two to five channels of previous images. A batch generator was used in each neural network to generate additional input images for network based on slight variations and rotations on existing input images. In addition, all input images were shuffled to further regularize training. The neural networks were then trained for 100 epochs with 20% of the training images used for validation of each epoch. The network was then tested with the testing data. This training and testing process were run for the single image input, followed by each of the two to five-image inputs.

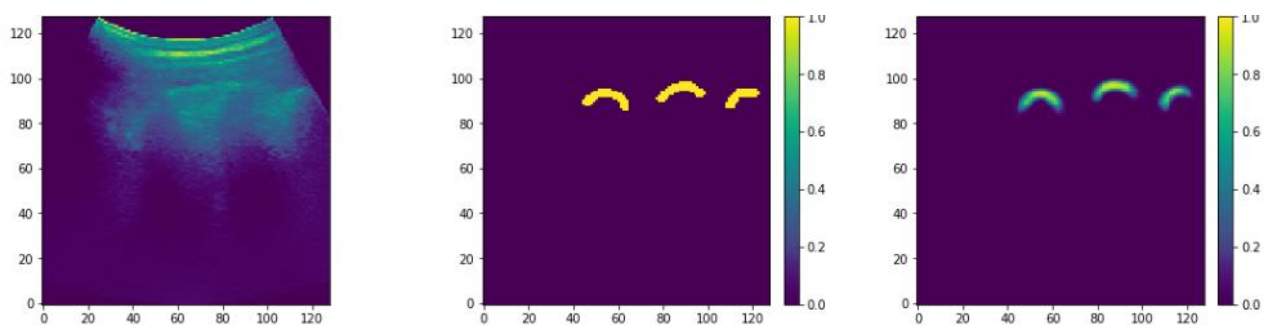


Figure 4: *Left.* Ultrasound image. *Middle.* Input segmentation. *Right.* Generated prediction

---

[iii] https://opencv.org/

2.4 Segmentation evaluation metrics

The segmentations generated by the convolutional neural network were thresholded at a level of 0.25, creating a clear cut segmentation with no uncertainty around the borders. This threshold value was chosen after testing a series of values, as it provided the best results in terms of both accuracy metrics, and visual appearance. An acceptable segmentation region was created by adding 2mm around the borders of the ground truth segmentation; this accounts for the varying brush sizes that a user may use when segmenting. The true positive rate of the segmentation was determined by comparing the predicted segmentation with the ground truth segmentation. False positives were any segmentation predicted that was not in the acceptable ground truth segmentation. The true negative rate was calculated as difference between the ground truth segmentation and the false positives predicted (Figure 5).

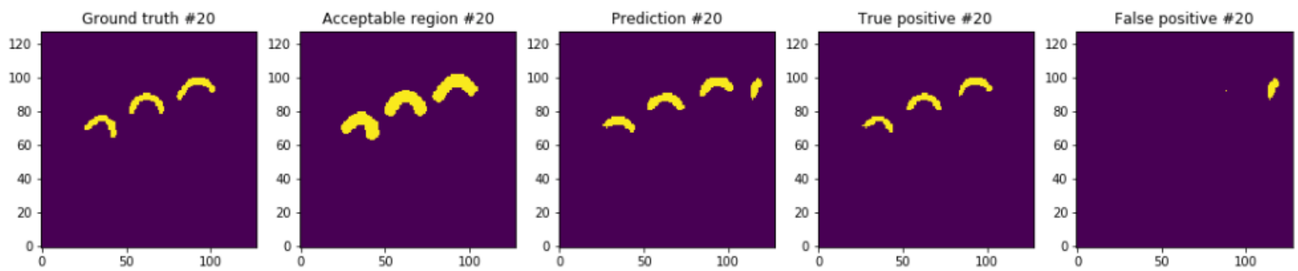


Figure 5: Comparison between the ground truth and predicted segmentation

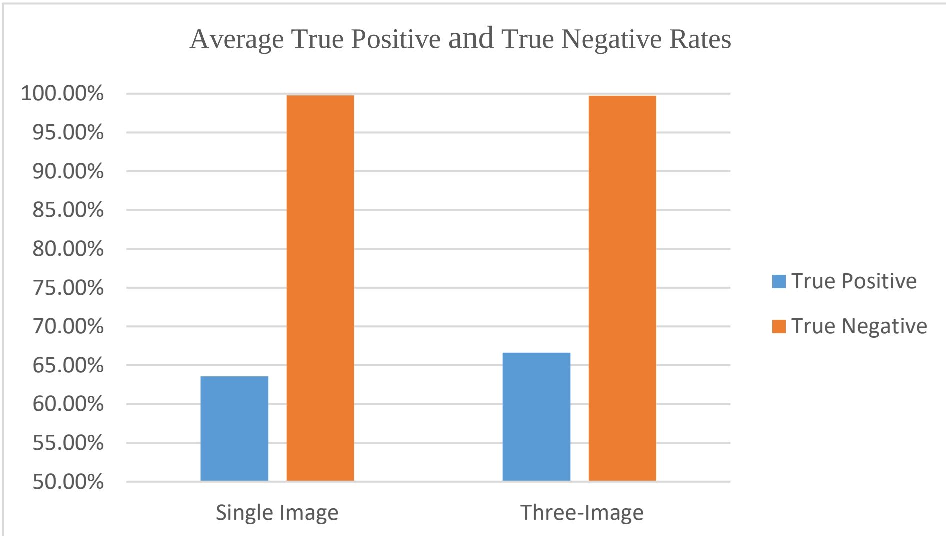
3. RESULTS

After testing all five input arrays, the three-image input had the greatest true positive accuracy (Figure 6). Based on this information, further tests were conducted comparing single image input and three-image input. The training and testing process were run an additional 4 times for the single and three-image input, totaling 5 tests each. The average true positive and true negative rates from these 5 experiments was then computed.

	Single Image	Two-Image	Three-Image	Four-Image	Five-Image
True Positive	62.89%	61.94%	68.19%	64.42%	60.42%
True Negative	99.78%	99.83%	99.75%	99.86%	99.79%

Figure 6: True positive and true negative rates of single to five-image inputs

When the experiment was run five times using the single image segmentation, the average true positive rate was 63.56% and the average true negative rate was 99.79%. When the experiment was run five times using the three-image input, the average true positive rate was 66.64% and the average true negative rate was 99.75% (Figure 7). Overall, the three-image input network outperformed the single image input network on average by 3.08%.



Experiment Number	**Single Image**		**Three-Image**	
	True Positive	**True Negative**	**True Positive**	**True Negative**
1	63.28%	99.82%	71.27%	99.69%
2	64.04%	99.80%	65.07%	99.77%
3	61.11%	99.82%	64.64%	99.79%
4	66.17%	99.73%	66.84%	99.75%
5	63.20%	99.80%	65.36%	99.77%
Average	63.56%	99.79%	66.64%	99.75%

Figure 7: Comparison of average true positive (blue) and true negative (orange) rates between single image (left) and three-image input (right)

4. DISCUSSION

Spinal ultrasound images are mainly background noise, with the spine only being a small part of each image. As a result, even though some segmentations may be inaccurate, the true negative value will always be high and over 99%; an ultrasound image with no predicted segmentation would still receive a high true negative value. Therefore, quality of the generated segmentations is better seen through the true positive evaluation metric. Furthermore, although all scans were performed on healthy volunteers, this work can be extrapolated to scoliosis patients as well, as the actual difference in the ultrasound scans is not very distinct.

The visual appearance of the resulting segmentations is very similar to that of the ground truth manual segmentation. While the true positive metric was only around 65%, the network was still able to accurately segment transverse processes. This suggests that the missing true positive pixels come from inaccuracies in the shapes of the segmented processes, rather than spinal processes that had been entirely missing from the generated segmentation.

Additionally, the adjustment from the U-net to a multi-image U-net is straightforward. Only the number of input channels into the network is altered; the rest of the neural network is exactly the same. This change is robust and easy to implement. As a result, the multi-image U-net is viable and easy to apply to other domains where U-net is being used for segmentation.

5. NEW OR BREAKTHROUGH WORK TO BE PRESENTED

We presented a new approach to segmentation of spinal ultrasound images. By providing one to four previous images as additional channels in the input array, we achieved better results. This reflects the human ability to scroll through ultrasound scans and use information about the general anatomy of the ultrasound from previous images for segmentation.

6. CONCLUSIONS

This experiment presents the workflow and convolution neural network architecture for the segmentation of spinal ultrasound images. Based on the study, using three consecutive images per input into a U-net architecture convolutional neural network enhances segmentation results by 3.08% when compared to using single image inputs. This is a marginal improvement, suggesting that the capabilities of this methodology of examining consecutive images is limited. However, despite the modest improvement, segmentations generated from the single input and three-input networks are both quite accurate, suggesting that U-net is a viable architecture for spine ultrasound segmentation.

ACKNOWLEDGEMENTS

Victoria Wu was supported by CIHR Undergraduate Summer Studentship Award.Gabor Fichtinger is supported as a Canada Research Chair in Computer-Integrated Surgery. This work was funded, in part, by NIH/NIBIB and NIH/NIGMS (via grant 1R01EB021396-01A1 - Slicer+PLUS: Point-of-Care Ultrasound) and by CANARIE's Research Software Program.

7. REFERENCES

[1] Ungi T, King F, Kempston M, Keri Z, Lasso A, Mousavi P, Rudan J, Borschneck DP, Fichtinger G. "Spinal curvature measurement by tracked ultrasound snapshots," Ultrasound Med Biol 40(2), 447-54 (2014).

[2] Ronneberger O, Fischer P, Brox T. "U-net: Convolutional networks for biomedical image segmentation," Med Image Comput Comput Assist Interv, 234-241 (2015).

[3] Baka N, Leenstra S, van Walsum T. "Ultrasound aided vertebral level localization for lumbar surgery," IEEE T Med Imaging 36(10), 2138-47 (2017).

[4] A. Lasso, T. Heffter, A. Rankin, C. Pinter, T. Ungi and G. Fichtinger, "PLUS: Open-Source Toolkit for Ultrasound-Guided Intervention Systems," IEEE T Bio-Med Eng 61(10), 2527-2537 (2014).

[5] Ungi T, Lasso A, Fichtinger G. "Open-source platforms for navigated image-guided interventions," Med Image Anal. 33, 181-6 (2016).