EPiC Series in Health Sciences

Volume 3, 2019, Pages 6–10

CAOS 2019. The 19th Annual Meeting of the International Society for Computer Assisted Orthopaedic Surgery

Spine Surface Segmentation from Ultrasound Using Multi-feature Guided CNN

Ahmed Z. Alsinan[1], Michael Vives[3], Vishal Patel[5], Ilker Hacihaliloglu [2,4]

[1] Department of Electrical and Computer Engineering, Rutgers University, USA
[2] Department of Biomedical Engineering, Rutgers University, USA
[3] Department of Orthopedics, Rutgers New Jersey Medical School, USA
[4] Department of Radiology, Rutgers Robert Wood Johnson Medical School, USA
[5]Department of Electrical and Computer Engineering, Johns Hopkins University, USA
ahmed.alsinan@rutgers.edu ilker.hac@soe.rutgers.edu

Abstract

Accurate, robust, and real-time segmentation of bone surfaces is an essential objective for ultrasound (US) guided computer assisted orthopedic surgery (CAOS) procedures. In this work, we present a convolutional neural network (CNN)-based technique for segmenting spine surfaces from in vivo US scans. Proposed design utilizes fusion of feature maps extracted from multimodal images to abate sensitivity to variations caused by imaging artifacts and low intensity bone boundaries. In particular, our multimodal inputs consist of B-mode US images and their corresponding local phase filtered counterparts. Validation studies performed on 261 in vivo US scans obtained from 10 subjects achieved a mean localization accuracy of 0.1 mm with an F-score of 97%. Comparison against state-of-the-art CNN networks show an improvement of 89% in bone surface localization accuracy.

1 Introduction

Real-time, 2D/3D ultrasound (US) provides a safe and cost-effective alternative to fluoroscopy for intra-operative navigation during percutaneous pedicle screw insertion (PPSI) in spinal fusion surgery. Nonetheless, low signal-to-noise ratio (SNR), blurred and thick bone surface appearance, and imaging artifacts, present in the collected US scans, have hindered the design of an US-based PPSI system.

In order to provide a solution to these difficulties, focus has been given to develop automated US bone segmentation and enhancement methods that are robust and computationally inexpensive for US guided CAOS procedures. Most-recently, various groups have investigated methods based on deep learning. In (Baka, Leenstra, & van Walsum, 2017), a network architecture based on U-net of (Ronneberger , Fischer , & Brox , 2015), was investigated for segmenting vertebra bone surfaces. Reported precision, recall, and F-score rates were 0.88, 0.94, and 0.90 respectively. Also based on (Ronneberger , Fischer , & Brox , 2015), a deep learning network architecture was developed by (Salehi,

Prevost, Moctezuma, & Navab, 2017) for segmentation of bone surfaces from US data. Although localization accuracy was not reported, the recall and precision rates for the proposed method were 0.87. In (Villa , et al., 2018), an algorithm based on fully convolutional networks (FCN), where B-mode US and local phase image features were used, was proposed. The reported recall, precision, and F-score values were 62%, 64%, 57%.

In this work, we evaluate the performance of our newly proposed CNN architecture for segmentation of spine bone surfaces. Our design utilizes fusion of feature maps and employs multi-modal images to abate sensitivity to variations caused by imaging artifacts and low intensity bone boundaries (Alsinan, Patel, & Hacihaliloglu, 2019).We perform quantitative and qualitative validation on in vivo spine US data collected from 10 subjects.

2 Methods

2.1 Data Acquisition

After obtaining institutional ethics board approval, a total of 261 B- mode US images, from 10 subjects, were collected. Data augmentation (by means of image rotation) was performed on this dataset to obtain 1,044 B-mode US images in total. All bone surfaces were manually segmented by an expert ultrasound technician.

2.2 Multi-feature guided CNN Architecture

We developed our proposed CNN architecture based on the common contractive-expansive design. First, we resize the input B-mode US image $US(x, y)$ and its complementary local phase filtered image $LP(x, y)$ based on (Hacihaliloglu, Enhancement of bone shadow region using local phase-based ultrasound transmission maps, 2017). In our proposed design, each input image would connect to an independent primary network, and a secondary network (Alsinan, Patel, & Hacihaliloglu, 2019). In each network, the input image is convolved in the encoder by convolutional layers with 3×3 filters (same padding convolutions) each followed by a rectified linear unit (ReLU) and a 2×2 maxpooling. Whereas in the decoder path, transposed-convolutions of same kernel size and paddings are applied and upsampled. The encoder maps the input image into a low-dimension latent space, and the decoder maps the latent representation into the original space. The proposed network layers specifications are depicted in Figure 1. In the primary network, the input image is a B-mode US image $US(x, y)$, while in the secondary network, the input is a local phase filtered image $LP(x, y)$ that proceeds through the aforementioned convolutional, and max pooling layers. Feature maps extracted from both networks are fused in a late fusion stage. This classifier level model was implemented in which high-level features from each network are concatenated. A 3×3 convolution with sigmoid activation is performed on the output of the fused layer to generate the final segmented probability distribution. (Hazirbas, Ma, Domokos, & Cremers, 2016)

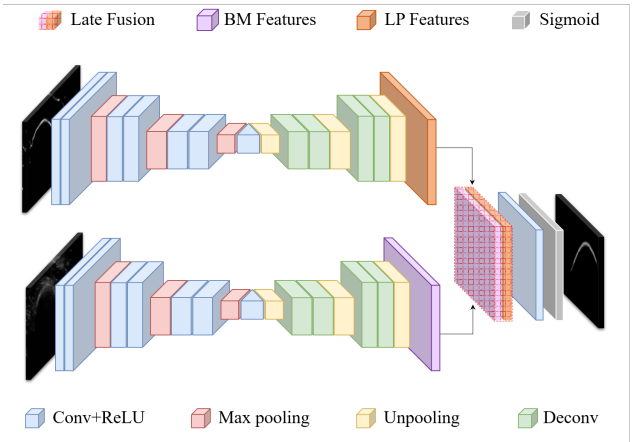


Figure 1. Proposed multi-modal fusion architecture.

2.3 Training and Testing

The performance of our proposed design was compared against the U-net network proposed in (Ronneberger , Fischer , & Brox , 2015) with its depth increased to a scale close to our proposed design. We have also trained the U-net network by using same input image features, B-mode US and local phase images. The networks were trained using a training set of 912 B-Mode US images and their corresponding local phase filtered images. The remaining 132 B-mode US images were reserved for testing the performance of the networks. During the random split of the dataset, the training and testing data did not include the same patient scans. This process was repeated five times, with each training and testing data randomized from our dataset. All the networks were trained to minimize the cross-entropy loss. Based on (Rand, 1971), and (Cernazanu-Glavan & Holban, 2013), five error metrics were calculated in our testing set; namely, F-score, Rand error, Hamming Loss, as well as the IoU and average Euclidean distance (AED) error as the bone localization error. Bone localization was achieved by thresholding the estimated probability segmentation map and using the center pixels along each US scanline as a single bone surface. AED error was calculated between the automatically segmented bone surfaces and the manual expert segmentation.

3 Results

3.1 Bone Segmentation Quantitative Results

Investigating Figure 2-(a) we can observe that the proposed CNN architecture outperforms the state-of-the-art in all error metrics investigated. A paired t-test, at a %5 significance level, between our designed network and the two U-net designs investigated achieved p-values less than 0.05 indicating that the improvements of our method are statistically significant.

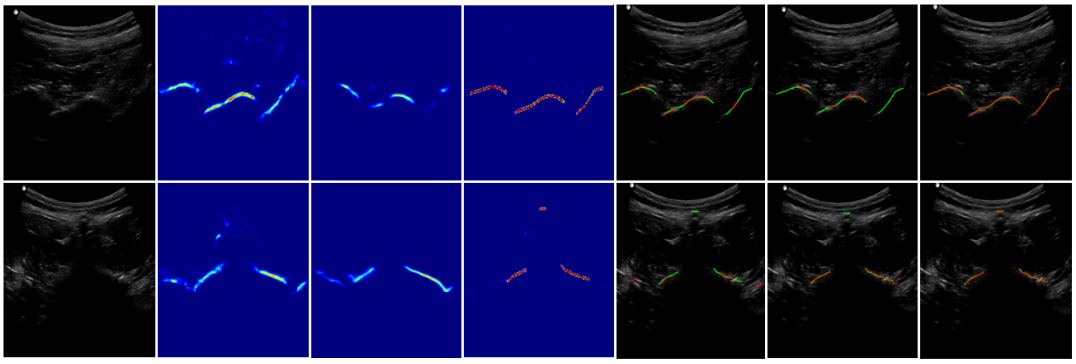
3.2 Bone Segmentation Qualitative Results

Qualitative results of our method show high prediction scores (red pixels) for the segmented bone surfaces while the investigated U-net designs have low prediction scores (light blue pixels) (Fig.2 (b-2)-(b-4). In addition, bone localization results against expert manual localization, are presented in

Figure 2 (b-5) to (b-7). Compared to U-net, our proposed design achieved significantly improved alignment with the expert bone localization.

Methods	IoU%	F-score	Rand	Hamming	AED (mm)
Ronneberger et al. - US B-mode only	0.803022	0.866023	0.662321	0.196977	2.9653
Ronneberger et al. - US B-mode & LP	0.851270	0.936892	0.750253	0.062402	0.9372
Ours - US B-mode & LP	**0.969489**	**0.977450**	**0.448840**	**0.030511**	**0.1097**

(a)



(b-1) (b-2) (b-3) (b-4) (b-5) (b-6) (b-7)

Figure 2: (a) Table showing quantitative results. (b-1) Two US B-mode images of in vivo spine bones. (b-2) Network segmentation output obtained from U-net trained with B-mode US images only. (b-3) Network segmentation output obtained from U-net trained with B-mode US images and Local-phase filtered images. (b-4) Network segmentation output obtained from proposed design. (b-5), (b-6), (b-7) Bone localization in (red) overlaid with manual expert localization (green): (b-5) U-net trained with B-mode US image, (b-6) U-net trained with B-mode US image and Local-phase image, (b-7) our proposed localization.

4 Discussion and Future Work

In this study, a multimodal CNN architecture was proposed for B-mode US bone segmentation. Our network incorporated local phase images in conjunction with B-mode US data. Quantitative and qualitative validation were performed against state-of-the-art U-net (Ronneberger , Fischer , & Brox , 2015). It was demonstrated that incorporating local phase bone image features improves the performance of the segmentation task. Particularly, it was observed that the late fusion of spatial-phase features resulted in higher bone segmentation probability outcomes. Our future work will involve further validations prior to utilizing the proposed method clinically. In addition, improving the computational cost of local phase feature extraction would be essential.

Acknowledgement

This work was supported by the North American Spine Society 2017 Young Investigator Basic Research Grant.

References

Alsinan, A., Patel, V., & Hacihaliloglu, I. (2019). Automatic segmentation of bone surfaces from ultrasound using a filter-layer-guided CNN. *International journal of computer assisted radiology and surgery*, 1-9.

Baka, N., Leenstra, S., & van Walsum, T. (2017). Ultrasound aided vertebral level localization for lumbar surgery. . *IEEE transactions on medical imaging 36(10)*, 2138–2147 .

Cernazanu-Glavan, C., & Holban, S. (2013). Segmentation of bone structure in x-ray images using convolutional neural network. *Adv. Electr. Comput. Eng*, 87-94.

Hacihaliloglu, I. (2017, June). Enhancement of bone shadow region using local phase-based ultrasound transmission maps. *International journal of computer assisted radiology and surgery, 12*((6)), pp. 951-60.

Hacihaliloglu, I. (2018). Localization of Bone Surfaces from Ultrasound Data Using Local Phase Information and Signal Transmission Maps. *Computational Methods and Clinical Applications in Musculoskeletal Imaging*, pp 1-11.

Hazirbas, C., Ma, L., Domokos, C., & Cremers, D. (2016). Fusenet: Incorporating depth into semantic segmentation via fusion-based cnn architecture. *Asian Conference on Computer Vision* (pp. 213-228). Springer.

Jain, V., Bollmann, B., Richardson, M., Berger, D. R., Helmstaedter, M. N., Briggman, K. L., . . . Abraham, W. C. (2010). Boundry Learning by optimization with topological constraints. *Computer Vision and Pattern Recognition (CVPR)* (pp. 2488-2495). IEEE.

Organization, W. H. (2003). *The burden of musculoskeletal conditions at the start of the new millenium: report of a who scientific group.* WHO Technical Report Series 919.

Ozdemir, F., Ozkan, E., & Goksel, O. (2016). Graphical modeling of ultrasound propagation in tissue for automatic bone segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 256–264.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, 846-850.

Ronneberger , O., Fischer , P., & Brox , T. (2015). U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, 234–241.

Salehi, M., Prevost, R., Moctezuma, J. L., & Navab, N. (2017). Precise ultrasound bone registration with learning-based segmentation and speed of sound calibration. *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 682-690). Springer.

The Burden of Musculoskeletal Diseases in the United States (BMUS). (2014). *United States Bone and Joint Initiative*, Third. Retrieved 8 25, 2017, from United States Bone and Joint Initiative: The Burden of Musculoskeletal Diseases in the United States (BMUS): http://www.boneandjointburden.org/

Valada, A., Vertens, J., Dhall, A., & Burgard, W. (2017). Adapnet: Adaptive semantic segmentation in adverse environmental conditions. *IEEE International Conference on Robotics and Automation (ICRA)* (pp. 4644-4651). IEEE.

Villa , M., Dardenne, G., Nasan, M., Letissier, H., Hamitouche, C., & Stindel , E. (2018). FCN-based approach for the automatic segmentation of bone surfaces in ultrasound images. *International journal of computer assisted radiology and surgery*, 13(11):1707-16.