International Journal of Computer Assisted Radiology and Surgery (2020) 15:1127–1135
https://doi.org/10.1007/s11548-020-02184-1

ORIGINAL ARTICLE

Robust real-time bone surfaces segmentation from ultrasound using a local phase tensor-guided CNN

Puyang Wang[1] · Michael Vives[2] · Vishal M. Patel[1] · Ilker Hacihaliloglu[3]

Received: 18 November 2019 / Accepted: 23 April 2020 / Published online: 19 May 2020

Abstract

Purpose Automatic bone surfaces segmentation is one of the fundamental tasks of ultrasound (US)-guided computer-assisted orthopedic surgery procedures. However, due to various US imaging artifacts, manual operation of the transducer during acquisition, and different machine settings, many existing methods cannot deal with the large variations of the bone surface responses, in the collected data, without manual parameter selection. Even for fully automatic methods, such as deep learning-based methods, the problem of dataset bias causes networks to perform poorly on the US data that are different from the training set.

Methods In this work, an intensity-invariant convolutional neural network (CNN) architecture is proposed for robust segmentation of bone surfaces from US data obtained from two different US machines with varying acquisition settings. The proposed CNN takes US image as input and simultaneously generates two intermediate output images, denoted as local phase tensor (LPT) and global context tensor (GCT), from two branches which are invariant to intensity variations. LPT and GCT are fused to generate the final segmentation map. In the training process, the LPT network branch is supervised by precalculated ground truth without manual annotation.

Results The proposed method is evaluated on 1227 in vivo US scans collected using two US machines, including a portable handheld ultrasound scanner, by scanning various bone surfaces from 28 volunteers. Validation of proposed method on both US machines not only shows statistically significant improvements in cross-machine segmentation of bone surfaces compared to state-of-the-art methods but also achieves a computation time of 30 milliseconds per image, 98.5% improvement over state-of-the-art.

Conclusion The encouraging results obtained in this initial study suggest that the proposed method is promising enough for further evaluation. Future work will include extensive validation of the method on new US data collected from various machines using different acquisition settings. We will also evaluate the potential of using the segmented bone surfaces as an input to a point set-based registration method.

Keywords Ultrasound · Orthopedics · Surgery · Bone · Segmentation · Deep learning

Introduction

In order to provide a radiation-free, real-time, cost-effective imaging alternative for intra-operative fluoroscopy ultrasound (US) has been incorporated into various computer-assisted orthopedic surgery (CAOS) procedures such as percutaneous scaphoid fixation and pelvic ring facture surgery [1]. US-based guidance systems for non-surgical procedures such as epidural anesthesia and/or spinal blocks have also been developed [2,3].

Nonetheless, problems such as high levels of noise, imaging artifacts, limited field of view, and bone boundaries appearing several millimeters (mm) in thickness have hin-

✉ Ilker Hacihaliloglu
ilker.hac@soe.rutgers.edu

Puyang Wang
pwang47@jhu.edu

Michael Vives
vivesmj@njms.rutgers.edu

Vishal M. Patel
vpatel36@jhu.edu

1 Johns Hopkins University, Baltimore, USA

2 Rutgers New Jersey Medical School, Newark, USA

3 Department of Biomedical Engineering, Rutgers University, Piscataway, NJ 08854, USA

dered the wide spread adaptability of US-guided CAOS systems. These difficulties prohibited the use of US as a stand-alone intra-operative imaging modality, and focus was given on developing automated bone segmentation, enhancement [4–7] and intra-operative US-based image registration [8] methods. Our groups main focus is the development of an US-based CAOS system where automatically extracted bone surfaces are used for continuous real-time guidance. Therefore, complete, accurate, and robust segmentation of bone surfaces is of paramount importance.

Early work for segmenting bone surfaces from US utilized image intensity and gradient information [1]. However, these methods are not robust for processing low contrast bone surfaces and are affected by acquisition settings, image artifacts, and body mass index (BMI) of the patient. To address this challenge, local phase-based bone surface enhancement methods have been proposed [1]. Local phase information is extracted by filtering the B-mode US data in frequency domain using bandpass quadrature filters. Most common filters include Log-Gabor filter, monogenic filter, and local phase tensor filter [1]. The enhanced bone surfaces were localized using post-processing methods such as dynamic programming [1] or simple bottom up ray-casting. Although phase-based approaches are more robust to image artifacts and low-contrast bone surfaces, successful segmentation depends on the robustness of the post-processing method used. Furthermore, local phase-based methods require the optimization of band-pass quadrature filter parameters which requires large processing time making the methods not suitable for real-time processing [4,9].

Due to some recent advances in deep learning, deep learning-based methods have shown to provide much better bone segmentation given enough training data. In [10], a modified version of U-net was used for localizing vertebra bone surfaces. However, low-quality bone surfaces were excluded from the validation and testing procedure. Recently, various filtered feature-guided methods [5–7] were proposed. These methods propose to incorporate filtered features, such as local phase tensor image or enhanced bone shadow image, into the convolutional neural network (CNN) by either using early feature fusion or late fusion operations. Particularly, the multi-feature-guided CNN in [5] takes US image, local phase tensor image, bone shadow enhanced image, and local phase image as concatenated input. And it demonstrated the state-of-the-art performance when testing on different US machines. However, it was shown that average computational time for additional input local phase and shadow enhancement was 2 seconds, making real-time application impossible. In summary, despite the fact that methods based on deep learning produce robust and accurate results, the success rate is dependent on either: (1) consistent and high-quality US scans used for training and testing [10] or (2) additional computation time required for image filtering [5].

High-quality US data refer to US images where bone surfaces appear sharp with high intensity followed by intensity dropout representing the bone shadow interface. During the data collection, quality of the US images and machine plays an important role in acquiring high-quality US data. Most of the clinically available US machines are equipped with high-quality transducers. However, this is not valid with the point-of-care portable low-cost transducers. Furthermore, manual operation of the transducer introduces additional difficulties during data collection since a single-degree deviation angle by the operator can reduce the signal strength by 50% [1]. Most of previously proposed deep learning methods are trained on high-quality US data from single US machine. If the acquisition involves low bone surface contrast and image quality, the ability of complete and accurate segmentation always decreases dramatically.

In this paper, in order to address the problem of segmenting bone data more robustly, we propose a novel local phase tensor [11]-guided CNN architecture for bone surface segmentation from US data of various qualities. In order to improve the computation time of multi-feature-guided CNN [5], our proposed framework accommodates a local phase tensor (LPT) network that is trained to capture contrast and noise-invariant local phase information. To further improve the robustness and suppress non-bone responses of LPT network, a global context tensor (GCT) network that focuses on learning global context is proposed. The two subnetworks share a common encoder, and their outputs are fused to generate the final segmentation map. To take the full advantages of both LPT and GCT, we propose and evaluate three different fusion methods including addition, multiplication, and concatenation. The fundamental difference of leveraging LPT image between multi-feature-guided CNN [5] and proposed method is that our proposed method sees LPT image as a supervision signal during training instead of a input feature to the network. This allows the further optimization of LPT image toward better bone segmentation. Thus, the proposed method can provide robust, accurate, and real-time segmentation for bone US data.

Method

Terminology: In the following sections, we use LPT in italic font to represent the ground truth local phase tensor image, $\hat{LPT}$ for approximated LPT and LPT in non-italic font for general concept including both LPT and .

In our proposed robust real-time bone segmentation network architecture (Fig. 1), we first construct an encoder with a series of CNN blocks using convolution (Conv), batch normalization (BN) [12], and rectified linear unit (ReLU) [13] to capture mainly low-level features from the input US scan. Then, the network splits into two branches to generate the

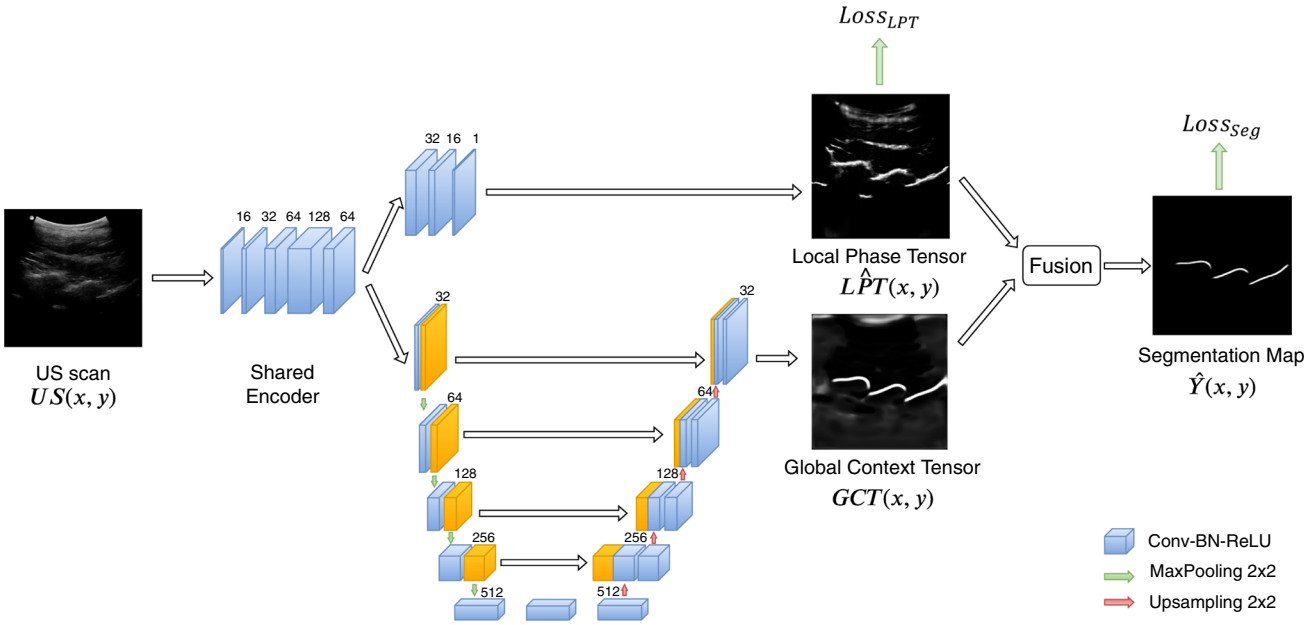

Fig. 1 Overview of the proposed method. 3×3 zero-padded convolutions are used for all convolution layer. The network, including LPT and GCT, is trained jointly in an end-to-end fashion with Loss_{LPT} and Loss_{Seg}

estimated local phase tensor images, denoted as $\hat{\text{LPT}}(x, y)$, and global context tensor images, denoted as $\text{GCT}(x, y)$, separately. Different from the original LPT in [11,14], $\hat{\text{LPT}}(x, y)$ is the approximation of LPT using CNN. The LPT network output provides a local phase image response of contrast information that is independent of not only the US transducer but even the quality of the scans. Hence, the LPT network can be seen as a general boundary indicator. On the other hand, GCT network provides bone-related global context information. Considering the expected feature level of each branch, the LPT and GCT subnetworks feature different network architectures to extract low- and mid-level features, respectively. Finally, the two outputs are combined together using different fusion methods (addition, multiplication, and concatenation) to generate the final segmentation map. To ensure the capability of LPT subnetwork outputs intensity-invariant local phase features, we employ Loss_{LPT} that takes precalculated local phase tensor images, denoted as $\text{LPT}(x, y)$, using its definition as the ground truth which we explain in the next section.

Local phase tensor

In the proposed work, local phase information is obtained using a gradient energy tensor filter. This information is used to construct local phase tensor image, denoted as $\text{LPT}(x, y)$, which highlights the contrast change and weak edges including bone surfaces in US scans. Because LPT is derived in a non-data-driven way, the calculation of $\text{LPT}(x, y)$ is independent of imaging devices which provides a perfect feature for cross-machine bone segmentation. In the proposed framework, the LPT network is designed to learn a mapping function between the input image and its $\text{LPT}(x, y)$ image calculated by three Conv–BN–ReLU blocks (Fig. 1). This enables weakly-supervised training of the LPT subnetwork without manual annotations.

Given a B-mode US image, denoted as $\text{US}(x, y)$, $\text{LPT}(x, y)$ is obtained from odd (T_{odd}) and even (T_{even}) filter responses using [11]:

$$\begin{aligned} T_{\text{even}} &= [\boldsymbol{H}(\text{US}_{\text{DB}}(x, y))]\,[\boldsymbol{H}(US_{DB}(x, y))]^T, \\ T_{\text{odd}} &= -0.5 \times ([\nabla \text{US}_{\text{DB}}(x, y)]\left[\nabla\nabla^2 \text{US}_{\text{DB}}(x, y)\right]^T \\ &\quad + \left[\nabla\nabla^2 \text{US}_{\text{DB}}(x, y)\right][\nabla \text{US}_{\text{DB}}(x, y)]^T), \end{aligned} \tag{1}$$

where T_{even} and T_{odd} represent the symmetric and asymmetric features of $\text{US}(x, y)$, respectively. $\mathbf{H}$, ∇, and ∇^2 represent the Hessian, Gradient, and Laplacian operations, respectively. In order to improve the enhancement of bone surfaces located deeper in the image and mask out soft tissue interfaces close to the transducer, $\text{US}(x, y)$ image is masked with a distance map and band-pass filtered image using the Log-Gabor filter [11,14]. The resulting image, from this operation, is represented as $\text{US}_{\text{DB}}(x, y)$. The final $\text{LPT}(x, y)$ image is obtained using the instantaneous phase ϕ:

$$\phi = \text{ang}\left(s_{\text{even}}\sqrt{\text{Trace}(T_{\text{even}})} + i \cdot s_{\text{odd}}\sqrt{\text{Trace}(T_{\text{odd}})}\right) \tag{2}$$

$$\text{LPT}(x, y) = \sqrt{T_{\text{even}}^2 + T_{\text{odd}}^2} \times \cos(\phi), \tag{3}$$

where $s_{\text{even}} = -\text{sign}(\mathbf{o}^T[\mathbf{H}\text{US}_{\text{DB}}(x, y)]\mathbf{o})$, $s_{\text{even}} = -\text{sign}(\mathbf{o}^T[\nabla\text{US}_{\text{DB}}(x, y)])$ and $\mathbf{o}$ is the orientation vector obtained from gradient energy tensor (GET) filter [11,14].

In order to regularize the network to approximate Eq. 3, we employ the following loss function between the estimated $\hat{\text{LPT}}(x, y)$ and the ground truth $\text{LPT}(x, y)$:

$$\text{Loss}_{\text{LPT}} = \frac{1}{WH}\sum_{x=1}^{W}\sum_{y=1}^{H}|\text{LPT}(x, y) - \hat{\text{LPT}}(x, y)|, \tag{4}$$

where it is assumed that LPT and $\hat{\text{LPT}}$ are of size $W \times H$.

Global context tensor

As mentioned above, the LPT network can be seen as a general boundary indicator. Although it can effectively enhance the bone surfaces in the US image, non-bone boundaries such as soft tissues will also be highlighted (Fig. 3). To overcome the drawbacks of the LPT network, we propose a global context tensor (GCT) subnetwork for extracting the missing bone-related global context information in LPT. Unlike locally computed LPT, the GCT network requires a larger receptive field to extract high-level features. Thus, we use the widely used contractive-expansive design with skip connections similar to U-net [15].

Given the output features from the shared encoder, it is fed through a 4-stage maxpooling and upsampling U-net with half the feature maps compared to the original U-net. The output of GCT network is a 2D image, denoted as $\text{GCT}(x, y)$, of same size as $\hat{\text{LPT}}(x, y)$.

Although there is no direct supervision on GCT (no ground truth), GCT is indirectly supervised by the final segmentation loss which is discussed in following section and the goal is to refine the coarse segmentation of $\hat{\text{LPT}}(x, y)$ through a fusion layer.

LPT and GCT fusion

Despite that the shared encoder and LPT network are guided by back propagating Loss_{LPT}, GCT must be properly regularized by incorporating it into the end-to-end bone segmentation framework. Therefore, some kind of fusion for LPT and GCT should be applied to generate the final predicted bone segmentation map $\hat{Y}$. The following three fusion methods (addition, multiplication, and concatenation) to generate $\hat{Y}$ are proposed for our framework:

$$\hat{Y} = \text{Sigmoid}(\hat{\text{LPT}} + \text{GCT}) \tag{5}$$

$$\hat{Y} = \text{Sigmoid}(\hat{\text{LPT}} \cdot \text{GCT}) \tag{6}$$

$$\hat{Y} = \text{Sigmoid}(w_1 \cdot \hat{\text{LPT}} + w_2 \cdot \text{GCT} + b). \tag{7}$$

Note that for concatenation in Eq. 7, it is implemented by concatenation along channel dimension followed by a linear layer with learnable weight (w_1, w_2) and bias b that are optimized during the training process as part of the network. For all fusion methods, we define the segmentation loss using the binary cross-entropy loss:

$$\text{Loss}_{\text{Seg}} = -\frac{1}{WH}\sum_{x=1}^{W}\sum_{y=1}^{H} Y(x, y) \cdot \log \hat{Y}(x, y) + (1 - Y(x, y)) \cdot (1 - \log \hat{Y}(x, y)), \tag{8}$$

where Y is ground truth segmentation mask.

Dataset and experiments

With the institutional review board (IRB) approval, 25 healthy volunteers were included in the study. We have collected a total of 1042 different US images using SonixTouch US machine (Analogic Corporation, Peabody, MA, USA) using 2D C5-2/60 curvilinear and L14-5 linear transducer. In order to collect new test data not used for training, we have recruited 3 new subjects and collected a total of 185 US scans using a handheld wireless US system (Clarius C3, Clarius Mobile Health Corporation, BC, Canada). Image resolution varied between 0.1mm and 0.15mm depending on the depth setting. Because of differences in transducer design and images reconstruction pipeline, the US scans from Clarius C3 have lower image quality in terms of bone imaging. The following bones were scanned: knee, femur, radius, and spine. Bone surfaces from the collected data were manually segmented by an expert ultrasonographer in order to generate the gold standard surfaces.

A random split of US images based on subjects from SonixTouch in training (80%) and testing (20%) sets was applied. The training set consists of a total of 834 images obtained from SonixTouch only. The remaining 208 images from SonixTouch and all 185 images from Clarius C3 were used for testing. During the random split of the SonixTouch dataset, the training and testing data did not include the same patient scans. All images including ground truth LPT are normalized to $[-1, 1]$ before feeding to the networks. For training, the overall loss is defined as:

$$\text{Loss}_{\text{Seg}} = \text{Loss}_{\text{Seg}} + \lambda \cdot \text{Loss}_{\text{LPT}}, \tag{9}$$

where λ is balancing weight of two losses. We searched for the optimal λ by varying it from 0 to 1 using 10% of training set as validation set. We observed that the segmentation performance will be severely impacted when λ is either less than 0.01 or larger than 0.25. But when λ is inside that range, the result stays relatively stable. Therefore, we empirically set it to 0.1 for all experiments. ADAM stochastic optimization

[16] with batch size of 16 and a learning rate of 0.001 is used for learning the weights.

For validation and comparison, two reference methods were selected: original U-net [15] and state-of-the-art for bone segmentation [5] (MFGCNN). For the proposed method, we included four configurations with three different fusion method and one ablation study in which Loss_{LPT} is not added. All these methods were implemented and evaluated by segmenting collected data. By thresholding the estimated segmentation map, we used the center pixels along each scanline as a single bone surface. The quality of the localization was evaluated by computing average Euclidean distance (AED) between the two surfaces along each scanline. We also evaluated the bone segmentation methods in terms of recall, precision, and their harmonic mean, the F-score. True positive is considered with 1-mm tolerance. Bone surface point outside 1-mm tolerance is excluded from AED error.

Results

Quantitative results

The AED results in Table 1 show that all variations of the proposed method achieve comparable bone surface localization performance against the stat-of-the-art method, MFGCNN for both datasets. GCT-only represents the network without LPT branch and has a sigmoid activation layer added to the end. Note that training set only contains images from US machine (SonixTouch), while testing is performed on both including the low-quality images from handheld wireless US scanner (Clarius C3). It is worth mentioning that although AED results do not show significant difference across different methods, 0.3-mm (2 pixels) error for segmented bone surfaces can be very well accepted for US-based segmentation.

The average recall and precision rates as well as F-scores in Table 1 clearly demonstrate the robustness of our proposed method for both datasets. By adding Loss_{LPT}, the F-score on Clarius C3 is boosted from 0.817 to 0.882 compared to 0.841 for MFGCNN (paired t test $p < 0.05$). Similar significant improvements are also observed for precision and recall rates (paired t test $p < 0.05$). Although both MFGCNN and the proposed method utilize LPT features, MFGCNN does not have direct supervision on leveraging/extracting information of LPT features. LPT features are used as supervision signals to generate potentially better intermediate features for bone segmentation. This is done by optimizing the estimated $\hat{\text{LPT}}$ over final segmentation loss Loss_{Seg}.

It is also shown that fusion by addition outperforms multiplication and concatenation in terms of recall, precision, and F score. The possible reason for concatenation fusion method is that it can easily let the estimated GCT dominate the final segmentation without leveraging the information in LPT. For example, in concatenation fusion equation Eq. 7, the network can solely depend on GCT by simply putting w_1 to near zero without increasing the training loss. This conjecture is supported by the observation that optimized $w_1 = 0.0031$ and $w_2 = 7.5725$ after training. As for multiplication fusion, it requires GCT subnetwork to precisely localized the false predicted pixels in $\hat{\text{LPT}}$ in order to correct it by flipping the sign since $\hat{\text{LPT}}$ is in the range of $[-1, 1]$, while for addition fusion, false predictions in $\hat{\text{LPT}}$ can be possibly corrected by uncertain prediction from GCT.

The box plot for recall rates of U-net, MFGCNN, and the proposed method with addition as fusion in Fig. 2 further demonstrates the superior cross-machine bone surface segmentation ability of providing more complete result. Note that from the AED and recall rate results on SonixTouch and Clarious C3 data in Table 1 and Fig. 2, one can observe that all methods suffer performance drops when tested on Clarious C3 data. This is the result of low bone surface contrast of Clarious C3.

Qualitative results

Qualitative results in Fig. 3 show that our method achieves improved and complete segmentation results, while U-net and MFGCNN suffer from missing bone segments which is crucial for US-based intra-operative guidance using features extracted from US data. Average computational time for our method is 30 ms compared to 2 seconds for the complete MFGCNN framework. This is an improvement of 98.5% over MFGCNN [5]. The comparison in terms of image quality between two US machines can be made by comparing first and last rows in Fig. 3 that both are femur bone US scans. The US image (last row) from Clarius C3 shows lower bone surface contrast compared to SonixTouch. However, both estimated $\hat{\text{LPT}}$ images can enhance bone surfaces very well despite of intensity/gradient difference in two US images.

Although LPT is used as supervision signal to regulate the estimated $\hat{\text{LPT}}$, we observed that the estimated LPT in our proposed network can produce a better coarse bone segmentation inside bone regions. We demonstrated this in Fig. 4 by showing both LPT and ground truth bone segmentation. It is clear to see that estimated $\hat{\text{LPT}}$ has more clear and complete bone enhanced signal around bone surfaces.

Conclusion

In order to make US an essential imaging modality in orthopedics clinically acceptable accuracy and robustness of guidance system need to be ensured. Therefore, complete, accurate, and robust bone segmentation is of paramount

Table 1 AED (mm), standard deviation of AED, recall, precision, and F-scores for the proposed and state-of-the-art methods

	Method	U-net [15]	MFGCNN [5]	GCT-only	Proposed					
	CAT				✓			✓		
	ADD					✓				✓
	MUL						✓		✓	
	LPT							✓	✓	✓
SonixTouch	AED	0.371	0.332	0.358	0.340	0.334	0.359	0.332	0.347	**0.325**
	Std	0.172	0.164	0.169	0.158	0.152	0.173	0.148	0.167	0.158
	Recall	0.878	0.939	0.917	0.934	0.935	0.932	0.941	0.940	**0.948**
	Precison	0.809	0.809	0.808	0.81	0.805	0.812	0.806	0.811	**0.815**
	F-score	0.842	0.869	0.859	0.869	0.865	0.868	0.868	0.870	**0.877**
Clarious C3	AED	0.500	**0.401**	0.453	0.416	0.429	0.438	0.422	0.427	0.410
	Std	0.351	0.201	0.310	0.217	0.228	0.243	0.237	0.235	0.215
	Recall	0.698	0.787	0.712	0.750	0.765	0.727	0.817	0.803	**0.847**
	Precison	0.880	0.902	0.894	0.897	0.882	0.906	0.898	0.911	**0.920**
	F-score	0.779	0.841	0.793	0.817	0.819	0.807	0.856	0.854	**0.882**

All methods are trained only on SonixTouch data. CAT, ADD, and MUL denote three fusion methods: concatenation, addition, and multiplication. LPT is the option of adding Loss_{LPT}. Best number across all methods is in bold font

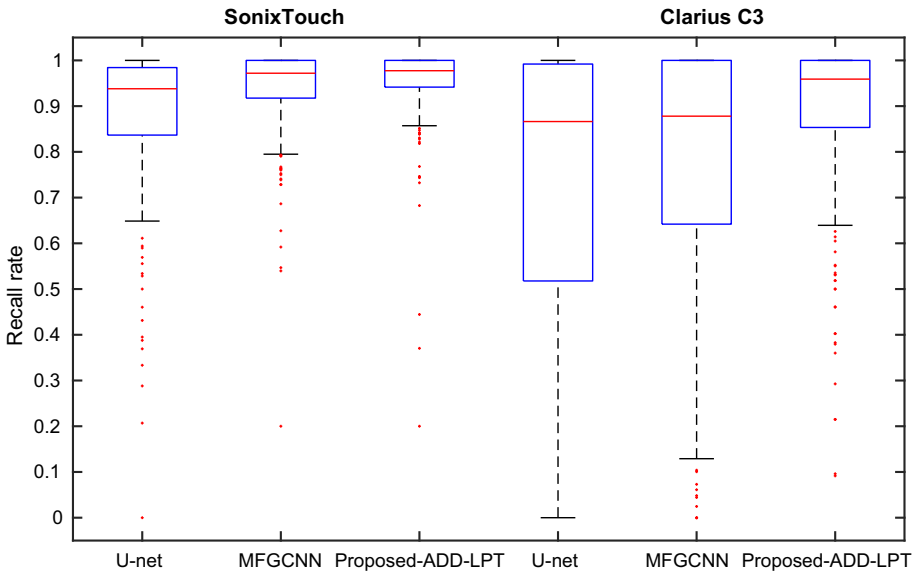

Fig. 2 Recall rates reflect the completeness of segmentation results which is the main drawback of existing bone segmentation methods when testing on data from different US machine

importance for US-based orthopedic surgical and non-surgical procedures where automatically extracted bone surfaces are used for continuous real-time guidance.

We proposed an end-to-end local phase-guided framework that enables robust and accurate bone surfaces segmentation for US-based computer-assisted orthopedic procedures. The main novelty of our work lies in (1) the integration of learning local phase tensor (LPT) and global context tensor (GCT) into a single network, (2) the design of fusion method of LPT and GCT to improve cross-machine segmentation performance of various bone imaging quality, and (3) the first systematic design of a fully automatic real-time framework for robust multi-machine LPT-guided bone surface segmentation from US images. It is critical for an automatic US segmentation algorithm to maintain robust performance on various US machines without any modification. Furthermore, this is the first study proposing a CNN-based local phase image generation network which we believe is an important contribution in the field of US-based orthopedic procedures.

Through validation, we demonstrate the state-of-the-art sensor adaption capability of our proposed method by separating training and testing data into two US machines. our proposed method achieves an AED of 0.32mm and 0.41mm, respectively, with a processing time of 30ms (98.5% improvement over state-of-the-art in processing speed [5]).

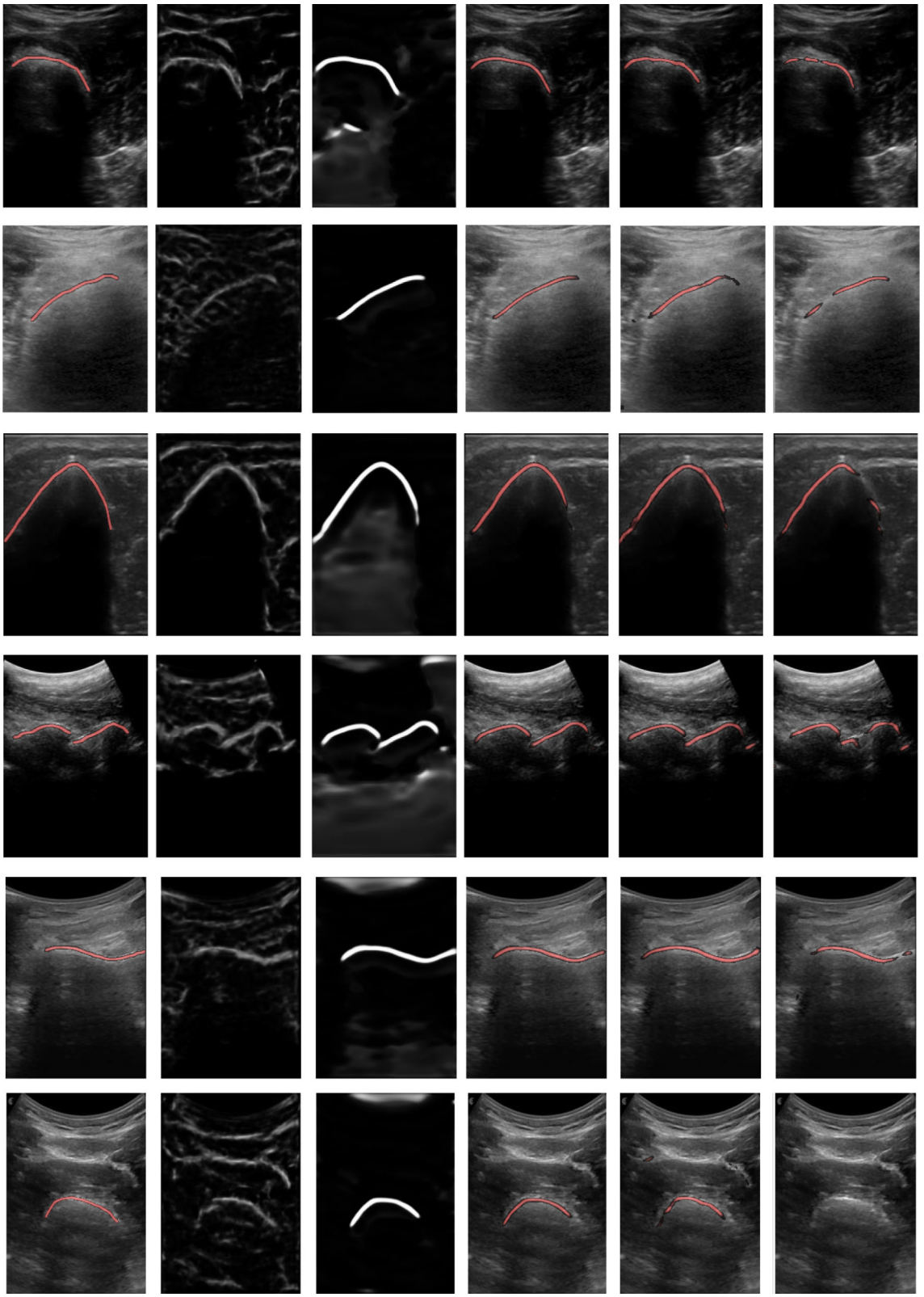

Fig. 3 Bone segmentation results. Top four and bottom two rows show in vivo B-mode US scans from SonixTouch and Clarius C3 US machines, respectively. Manual expert augmentations are shown in the first column. Note that U-net completely fails to segment bone surface in bottom sample imaged using Clarius C3

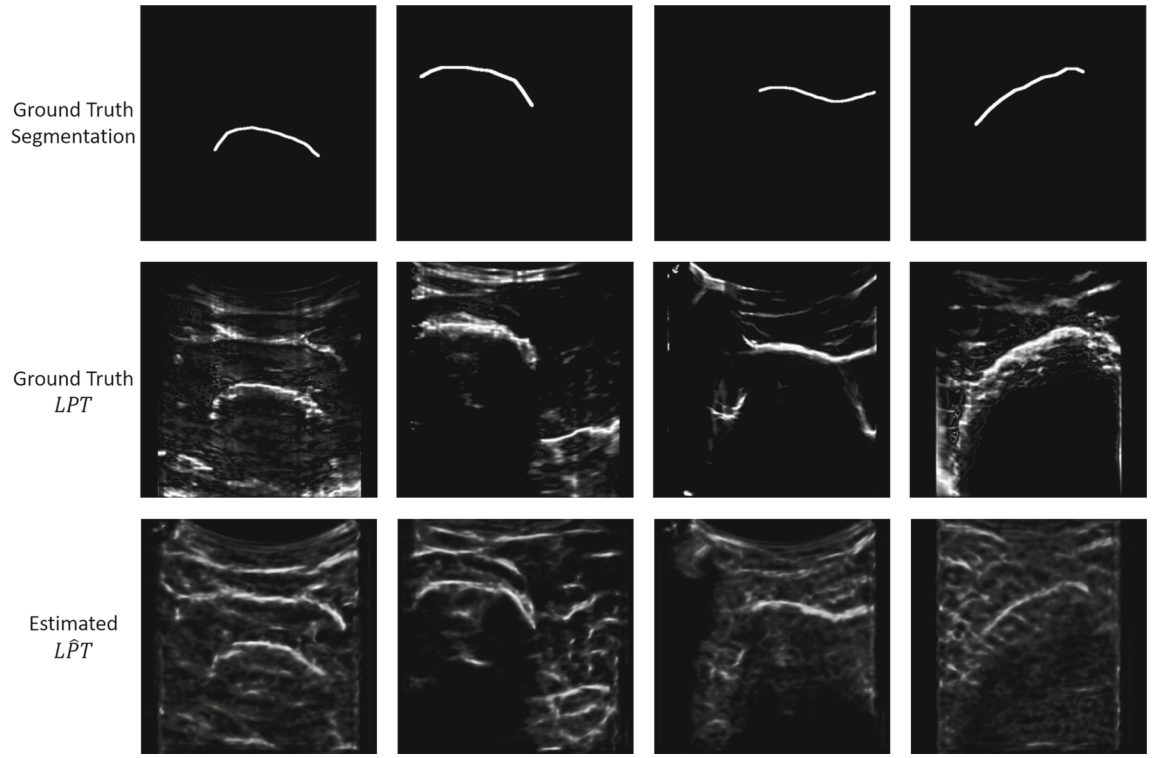

Fig. 4 Comparison of ground truth LPT and estimated LP̂T. Ground truth bone segmentation is shown at the top row for reference

One limitation of our study is that the uncertainty present in the gold standard labels (manual expert segmentation), resulting from intra- and inter-user variability, was not investigated. These errors can have direct impact on the accuracy of the developed deep neural networks. Another limitation is that only two US machines were used to collect the data. In order to fully investigate the true generalization of our method, more US machines should be used for data collection. We also would like to mention that one of the ongoing limitations in US bone segmentation research is that there are still no publicly available large data sets on which different algorithms could be trained and evaluated on. One solution to this could be joint effort in order to construct a publicly available database. Such an effort was most recently proposed in [17]. Our future work will include the extensive evaluation of our proposed method on this publicly available dataset. We will also investigate the incorporation of the extracted bone surfaces into a registration method.

Funding This work was supported in part by 2017 North American Spine Society Young Investigator Award.

Compliance with ethical standards

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

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

References

1. Hacihaliloglu I (2017) Ultrasound imaging and segmentation of bone surfaces: a review. Technology 5(02):74–80
2. Yamauchi M, Kawaguchi R, Sugino S, Yamakage M, Honma E, Namiki A (2009) Ultrasound-aided unilateral epidural block for single lower-extremity pain. J Anesth 23(4):605–608
3. Seitel A, Sojoudi S, Osborn J, Rasoulian A, Nouranian S, Lessoway VA, Rohling RN, Abolmaesumi P (2016) Ultrasound-guided spine anesthesia: feasibility study of a guidance system. Ultrasound Med Biol 42(12):3043–3049
4. Anas EMA, Seitel A, Rasoulian A, John PS, Pichora D, Darras K, Wilson D, Lessoway VA, Hacihaliloglu I, Mousavi P, Rohling R, Abolmaesumi P (2015) Bone enhancement in ultrasound using local spectrum variations for guiding percutaneous scaphoid fracture fixation procedures. Int J Comput Assist Radiol Surg 10(6):959–969
5. Wang P, Patel VM, Hacihaliloglu I (2018) Simultaneous segmentation and classification of bone surfaces from ultrasound using a multi-feature guided CNN. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 134–142
6. Alsinan AZ, Patel VM, Hacihaliloglu I (2019) Automatic segmentation of bone surfaces from ultrasound using a filter-layer-guided CNN. Int J Comput Assist Radiol Surg 14(5):775–783

7. 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. Int J Comput Assist Radiol Surg 13(11):1707–1716
8. Schumann S (2016) State of the art of ultrasound-based registration in computer assisted orthopedic interventions. In: Zheng G, Li S (eds) Computational radiology for orthopaedic interventions. Springer, Cham, pp 271–297
9. Hacihaliloglu I, Guy P, Hodgson AJ, Abugharbieh R (2015) Automatic extraction of bone surfaces from 3D ultrasound images in orthopaedic trauma cases. Int J Comput Assist Radiol Surg 10(8):1279–1287
10. Baka N, Leenstra S, van Walsum T (2017) Ultrasound aided vertebral level localization for lumbar surgery. IEEE Trans Med Imaging 36(10):2138–2147
11. Hacihaliloglu I, Rasoulian A, Rohling RN, Abolmaesumi P (2014) Local phase tensor features for 3-D ultrasound to statistical shape+ pose spine model registration. IEEE Trans Med Imaging 33(11):2167–2179
12. Ioffe S, Szegedy C (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. In: Proceedings of the 32nd international conference on machine learning (ICML-15), pp 448–456
13. Nair V, Hinton GE (2010) Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th international conference on machine learning (ICML-10), pp 807–814
14. Hacihaliloglu I (2017) Localization of bone surfaces from ultrasound data using local phase information and signal transmission maps. In: International workshop and challenge on computational methods and clinical applications in musculoskeletal imaging. Springer, pp 1–11
15. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 234–241
16. Kingma D, Ba J (2015) Adam: a method for stochastic optimization. In: Proceedings of the international conference on learning representations (ICLR)
17. Pandey P, Patel H, Guy P, Hacihaliloglu I, Hodgson AJ (2019) Preliminary planning for a multi-institutional database for ultrasound bone segmentation. EPiC Ser Health Sci 3:297–300