



Bone shadow segmentation from ultrasound data for orthopedic surgery using GAN

Ahmed Z. Alsinan¹ · Vishal M. Patel² · Ilker Hacihaliloglu^{3,4}

Received: 19 November 2019 / Accepted: 20 June 2020 / Published online: 11 July 2020
© CARS 2020

Abstract

Purpose Real-time, two (2D) and three-dimensional (3D) ultrasound (US) has been investigated as a potential alternative to fluoroscopy imaging in various surgical and non-surgical orthopedic procedures. However, low signal to noise ratio, imaging artifacts and bone surfaces appearing several millimeters (mm) in thickness have hindered the wide spread adaptation of this safe imaging modality. Limited field of view and manual data collection cause additional problems during US-based orthopedic procedures. In order to overcome these limitations various bone segmentation and registration methods have been developed. Acoustic bone shadow is an important image artifact used to identify the presence of bone boundaries in the collected US data. Information about bone shadow region can be used (1) to guide the orthopedic surgeon or clinician to a standardized diagnostic viewing plane with minimal artifacts, (2) as a prior feature to improve bone segmentation and registration.

Method In this work, we propose a computational method, based on a novel generative adversarial network (GAN) architecture, to segment bone shadow images from in vivo US scans in real-time. We also show how these segmented shadow images can be incorporated, as a proxy, to a multi-feature guided convolutional neural network (CNN) architecture for real-time and accurate bone surface segmentation. Quantitative and qualitative evaluation studies are performed on 1235 scans collected from 27 subjects using two different US machines. Finally, we provide qualitative and quantitative comparison results against state-of-the-art GANs.

Results We have obtained mean dice coefficient (\pm standard deviation) of 93% (± 0.02) for bone shadow segmentation, showing that the method is in close range with manual expert annotation. Statistical significant improvements against state-of-the-art GAN methods (paired t -test $p < 0.05$) is also obtained. Using the segmented bone shadow features average bone localization accuracy of 0.11 mm (± 0.16) was achieved.

Conclusions Reported accurate and robust results make the proposed method promising for various orthopedic procedures. Although we did not investigate in this work, the segmented bone shadow images could also be used as an additional feature to improve accuracy of US-based registration methods. Further extensive validations are required in order to fully understand the clinical utility of the proposed method.

Keywords Orthopedic surgery · Segmentation · Ultrasound · Bone · Generative adversarial network · Deep learning

✉ Ahmed Z. Alsinan
ahmed.alsinan@rutgers.edu
Vishal M. Patel
vpatel36@jhu.edu
Ilker Hacihaliloglu
ilker.hac@soe.rutgers.edu

¹ Department of Electrical and Computer Engineering, Rutgers University, Piscataway, NJ, USA

² Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, USA

Introduction

Imaging has been an integral component of various surgical and non-surgical orthopedic procedures such as total knee replacement (TKR), intramedullary nail locking for femoral shaft fractures, pedicle screw insertion for spinal fusion

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

⁴ Department of Radiology, Rutgers University Robert Wood Johnson Medical School, New Brunswick, NJ, USA

surgery, lumbar neuraxial anesthesia, and epidural analgesia [25]. Current practice during these procedures relies on intra-procedure 2D fluoroscopy as the main imaging modality for localization and visualization of bones, fractures, implants, and surgical tool positions. However, with such projection imaging, surgeons and clinicians typically face considerable difficulties in accurately localizing bone fragments in 3D space and assessing the adequacy and accuracy of the procedure. This problem has been overcome with 3D fluoroscopy units, however, they are twice as expensive and not widely available as standard 2D units. Finally, fluoroscopy involves significant radiation exposure [25]. The limits to exposure to ionizing radiation should be kept at minimal in order to avoid potential long-term complications. In order to overcome some of these limitations and provide a safe alternative, 2D/3D US has emerged as a safe alternative while remaining relatively cheap and widely available [8]. US image data, however, is typically characterized by high levels of speckle noise, reverberation, anisotropy and signal dropout which introduce significant difficulties during interpretation of captured data. Limited field-of-view and being a user dependent imaging modality causes additional difficulties during data collection since a single-degree deviation angle by the operator can reduce the signal strength by 50% [8]. In order to overcome these difficulties automatic bone segmentation [8] and registration [21] methods have been developed. Most recently, methods based on deep learning have achieved successful results for segmenting bone surfaces [1,2,22,23]. However, these methods require large amounts of training data and accuracy decreases if the quality of the testing data is low or if testing data comes from a different vendor machine. In the context of bone imaging using US, high quality data represents high intensity bone surface followed by a low intensity region referred to as shadow region. Difficulties in acquiring high quality US images is an ongoing limitation of current US guided orthopedic procedures.

Acoustic shadows occur at the interfaces where there is a high impedance difference such as air-tissue, tissue-bone, and tissue-lesion. Bone shadow information can aid in the interpretation of the collected data and has been incorporated as an additional feature to improve the segmentation of bone surfaces from US data [3,5,8,23]. Real-time feedback of bone shadow information can also be used to guide the clinician to a standardized diagnostic viewing plane with minimal artifacts. Finally, shadow information can also be used as an additional feature for registering CT, MRI or statistical shape models (SSM) to US data [21]. However, poor transducer contact or wrong orientation of the transducer with respect to the imaged anatomy can lead to poor shadow appearance and resulting in misinterpretation of anatomy and failure of the computational method using the shadow feature

(Fig. 1). Therefore, the enhancement of shadow regions has been investigated and practical solutions have been offered.

Several groups have proposed computational methods to improve the appearance of shadow regions from US data. Karamalis et al. [14] have proposed a random-walk geometric technique, based on image intensity, that models the propagation path of an US signal along the scanline. The generated images were termed confidence map (CM) images. Shadow regions were extracted from the CM images by intensity thresholding. This approach was later extended for processing radio-frequency US data [15]. In [10], shadow images of the brain were extracted by entropy analysis along the scanline. Pixels with low entropy would be selected to form the shadow image [10]. The method was later incorporated into a spinous process segmentation framework [3]. In [11], statistics of B-mode and radio frequency (RF) US data were investigated and used for shadow detection. Mean dice similarity coefficient (DSC) of 0.90 and 0.87 were obtained for the RF and B-mode algorithms. Processing time was not reported. Although promising results in these earlier works were achieved, intensity-based approaches are not robust to typical imaging artifacts and affected by intensity variations. Changing the US machine acquisition settings, sub-optimal orientation of the transducer concerning the imaged anatomy, imaging complex shape anatomy (such as spine), or scanning patients with different body mass index results in the collection of low quality US data (Fig. 1) and decrease the success of intensity-based approaches. RF-based shadow detection overcomes some of the difficulties of intensity approaches, however, they require special hardware, or software, to access RF signal domain which is not available in most clinical US machines. In order to provide an intensity invariant alternative, methods based on local phase image information have been proposed for the enhancement of bone shadow region [8]. The method proposed in [8] uses local phase image features as an input to a L1 norm-based contextual regularization method which emphasizes uncertainty in the shadow regions. Quantitative analysis, performed on a manually selected region of interest (ROI) achieved a mean DSC of 0.88. The mean computation time was 9.3 seconds making the method not suitable for real-time applications. In [17] a weakly supervised method for acoustic confidence estimation for shadow regions from fetal US data was proposed. In particular, a shadow-seg module to extract generalized shadow features for a large range of shadow types in fetal US images under limited weak manual annotations was presented. Both a classification and a segmentation networks with attention layer mechanism were used. The reported average DSC, Recall, and Precision were 0.71, 0.72, 0.73 respectively.

In this paper, we propose a conditional GAN(cGAN)-based method for accurate real-time segmentation of bone shadow regions from in vivo US scans. Our specific con-

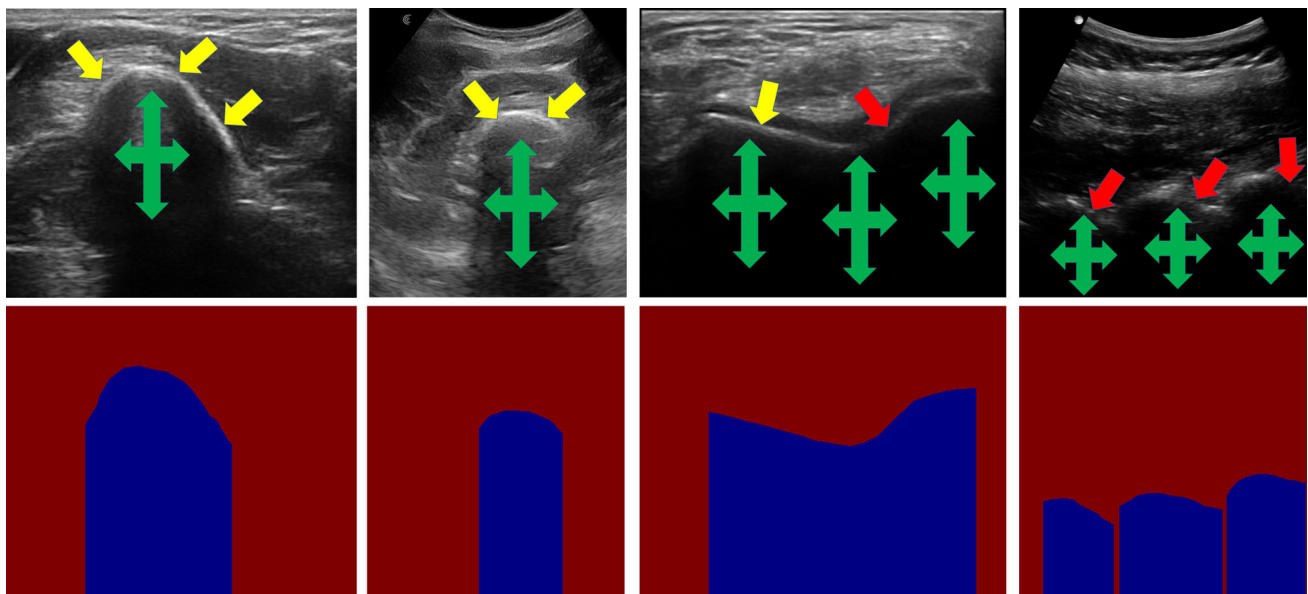


Fig. 1 Top row: From left to right in vivo B-mode US image of distal radius, femur, knee, and spine, respectively. Yellow arrows point to high intensity bone features. Red arrows point to the problematic low intensity bone features due to misalignment of the transducer or complex shape of the anatomy. Green arrow quads show the shadow region.

Bottom row: Manually segmented gold standard shadow images corresponding to B-mode data shown in the top row. In all the images blue color coded region is the shadow region and red color coded region is the soft tissue interface

tributions include: (1) A novel GAN architecture designed to perform accurate, robust and real-time segmentation of bone shadow images from in vivo US data. (2) We show how the segmented bone shadow regions can be used as an additional proxy to improve bone surface segmentation results of a multi-feature guided (CNN) architecture [1]. The significance of using shadow features-based segmentation is that they can be generated in real time as opposed to local phase image-based methods [1] which takes around one second. (3) We evaluate the proposed method on extensive in vivo data obtained from 27 volunteers using two different US imaging systems. We provide quantitative evaluation results against state-of-the-art GAN architectures.

Methods

Data acquisition

Upon obtaining the approval of the institutional review board (IRB), two imaging devices were used to collect data from 27 healthy subjects. Depth settings and image resolutions varied between 3–8 cm, and 0.12–0.19 mm, respectively:

1. Sonix-Touch US machine (Analogic Corporation, Peabody, MA, USA) with a 2D C5-2/60 curvilinear probe and L14-5 linear probe. Using this device we have collected 1000 scans from 23 subjects.

2. Clarius C3 hand-held wireless ultrasound probe (Clarius Mobile Health Corporation, BC, Canada). Using this device we have collected 235 scans from 4 subjects.

All the collected scans were scaled to a standardized size of 256×256 . The bone surfaces were manually segmented by an expert ultrasonographer. Gold standard bone shadow images were constructed automatically by investigating the intensity values from the manually segmented bone surfaces in the scanline direction. Region below the manually segmented bone surface is identified as shadow region. In total, we had 1235 B-mode US images categorized into four groups of bone structures: radius, femur, spine and tibia. We have performed fivefold cross validation on the Sonix-Touch data. No scans from one patient appeared in more than onefold. Training of the network architectures was performed using the Sonix-Touch data only. All the data, 235 scans in total, obtained from the Clarius C3 probe were used as test data.

Network architecture

Our architecture is based on the common GAN layout consisting of two co-existing neural networks; a generator G that attempts to generate synthetic samples and a discriminator D that tries to discriminate between generated synthetic samples and real ones [6]. In this work, we adopt the conditional aspect presented by [12] with our generator G and discriminator D both incorporating additional information into

account. Our proposed cGAN-based bone shadow segmentation and bone surface segmentation network architecture is shown in Fig. 2. The training of our proposed cGAN architecture follows the typical optimization problem [12] such that the discriminator D is trying to maximize and the generator G is trying to minimize the following objective \mathcal{G} :

$$\begin{aligned} \mathcal{G} = \arg \min_G \max_D \mathbb{E}_{\text{BM,GS}} [\log D(\text{BM}, \text{GS})] \\ + \mathbb{E}_{\text{BM},z} [\log (1 - D(\text{BM}, G(\text{BM}, z)))] \\ + \lambda \mathbb{E}_{\text{BM,GS},z} [\|\text{GS} - G(\text{BM}, z)\|_1] \end{aligned}$$

in which GS represents gold standard shadow images, z represents Gaussian noise in initial training but was applied as dropout on some layers in the convolution blocks, BS represents the segmented bone shadow image, GS represents the gold standard bone shadow image, and BM represents the B-mode US image. Different from a traditional GAN architecture, our actual generator G model was conditioned on the in vivo B-mode US image, BM, and is additionally tasked to generate BS images that are as close as possible to the GS images with the introduction of the L1-distance term as shown in the equation above. Our generator architecture is based on the common contractive-expansive design where the encoder maps the input image into a low-dimensional latent space, and the decoder maps the latent representation into the original space. It is trained to generate bone shadow, BS, images. However, unlike [12] where the generator was based on [20], we employ a different structure for the generator. Similar to [1], the input is processed through convolutional blocks, with each block consisting of several convolutional layers (Fig. 3). We incorporated skip connection and projection blocks similar to [16]. Our skip connection blocks, denoted as S , consist of a 1×1 convolution, a 3×3 convolution, and another 1×1 convolution with each convolution operation followed by batch normalization and leaky rectified linear unit (Leaky ReLU) activation. This process reduces and restores channel dimensions. In our design Leaky ReLU was used in the encoder and decoder. In [24], it has been shown that the Leaky ReLU achieves lower training and test errors compared to ReLU. Furthermore, Leaky ReLU attempts to overcome the 'dying ReLU (vanishing gradient)' problem by maintaining a small slope in the negative portion while training the piecewise constant gradient, making the network converge faster during training. This informed our choice of Leaky ReLU. A concatenated input and the aforementioned convolutions produce the output. As for our projection blocks, denoted as P , we add a 1×1 convolution to the projected input, and the rest is similar to the skip connection blocks. In the decoder, we replace all convolution operations by transposed-convolutions. We also use a stride of 2 transposed convolutions to upsample the feature maps. Therefore, these skip and projection

blocks with transposed-convolutions are denoted as S' and P' , respectively. Additionally, one difference in the decoder is that the batch normalization is followed by a dropout layer with a dropout rate of 50%. The architecture of the generator can be summarized as:

- **Encoder:** $S_{32} S_{32} P_{32} - S_{64} S_{64} P_{64} - S_{128} S_{128} P_{128} - S_{256} S_{256} P_{256} - S_{512} S_{512} P_{512}$
- **Decoder:** $S'_{512} S'_{512} P'_{512} - S'_{256} S'_{256} P'_{256} - S'_{128} S'_{128} P'_{128} - S'_{64} S'_{64} P'_{64} - S'_{32} S'_{32} P'_{32}$

In our discriminator model a two-input $N \times N$ PatchGAN-like discriminator [12] was used to essentially classify $N \times N$ patches of the input image as real or synthetic. Like the aforementioned generator, our discriminator architecture consists of five convolutional blocks, where a final convolution is applied to the last layer to map the 1-dimensional output before applying a Sigmoid function. Each batch normalization was followed by 0.2-slope Leaky ReLU. An Adam solver with a 0.0002 learning rate was used and the structure of the discriminator can be expressed as follows:

- **Discriminator:** $S_{32} S_{32} P_{32} - S_{64} S_{64} P_{64} - S_{128} S_{128} P_{128} - S_{256} S_{256} P_{256} - S_{512} S_{512} P_{512}$

While our proposed cGAN architecture was used to segment bone shadow regions BS, our bone surface segmentation network with its dual input proposed in [1] was used in our model to localize bone structures. The B-mode US image BM, and the segmented bone shadow image BS were used as input to our multi-feature CNN architecture. Feature maps extracted from both images are fused in a fusion layer at early (pixel level), mid (feature level) and late (classifier level) stages. Concatenation fusion was used as the fusion operation [1], which does not define any correspondence as it stacks feature maps at the same spatial locations across the feature channels. The multi-feature CNN architecture was trained separately from our proposed cGAN architecture using cross-entropy loss.

Quantitative evaluation

Bone shadow segmentation: The performance of our proposed design was compared against state-of-the-art GAN networks proposed in [12] and [18]. The depths of the networks were increased to a scale close to our proposed design. The bone shadow regions were also segmented, from the test data, using the local phase image-based bone shadow enhancement method proposed in [7]. In order to show the effectiveness of discriminator network we have obtained bone shadow segmentation results by only training our proposed generator network. Finally, to show the improvements achieved using a cGAN architecture over a traditional CNN

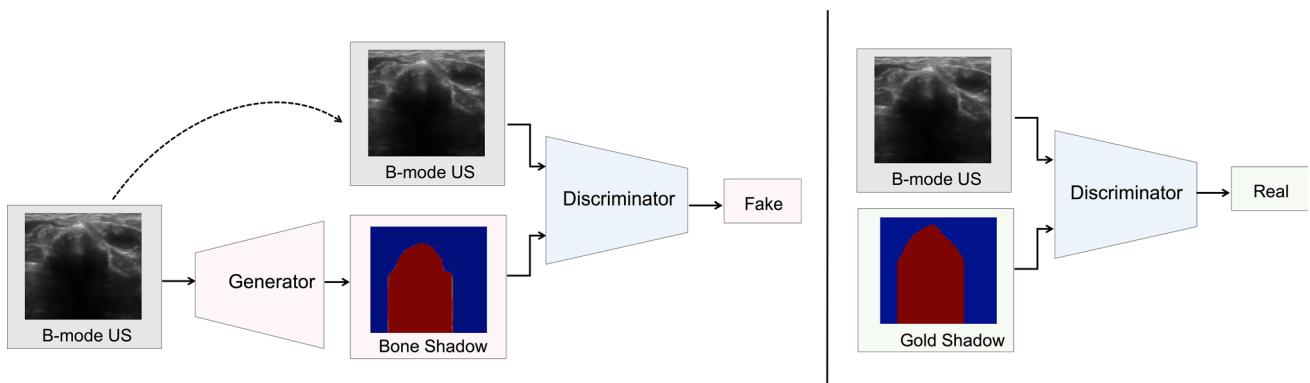


Fig. 2 Our proposed conditional GAN in which the discriminator learns to classify between real (gold standard shadow images, GS) and fake (generated bone shadow images, BS)

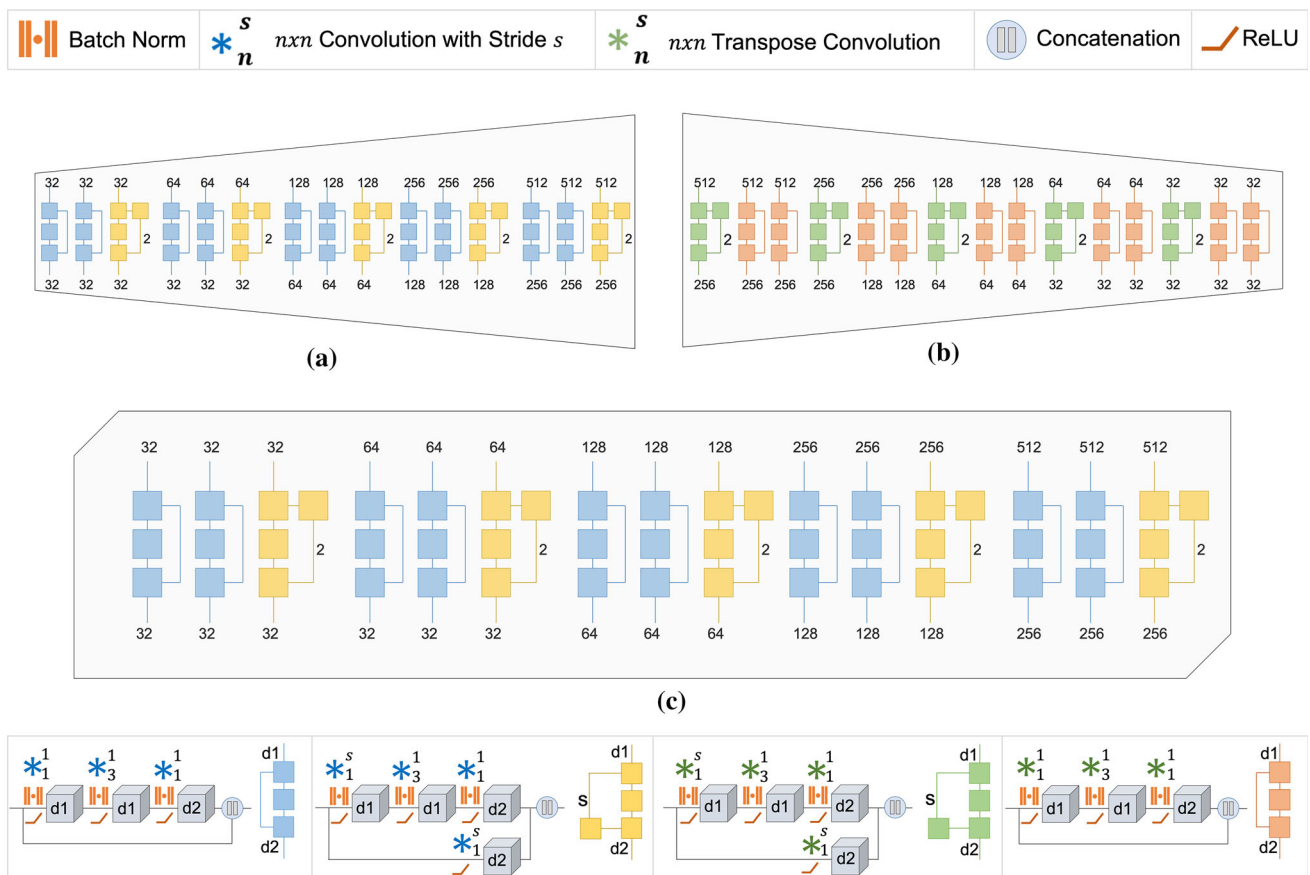


Fig. 3 An overview of our proposed cGAN architecture with its **a** generator’s encoder consisting of ten skip connection blocks (blue), in addition to five projection blocks (yellow) and **b** generator’s decoder consisting of ten transposed skip connection blocks (orange), in addition

to five transposed projection blocks (green). Depths of each convolutional layer are indicated in each block by d_1 and d_2 . **c** Our proposed patchGAN-like discriminator

architecture we trained the U-net network, proposed in [20], using B-mode US image features and gold standard bone shadow images. Based on [4,7,13,19], four error metrics were calculated in our testing set: Dice, Rand error, Hamming Loss, as well as the intersection over union (IoU). The

evaluation metrics are computed on the estimated probability maps, with grayscale color maps, and compared to the gold standard bone shadow images.

Bone surface segmentation: Bone shadow images segmented using our proposed design, [12] and [18], were used

as an additional feature to our multi-feature CNN architecture [1] for bone surface segmentation. Our method utilizes fusion of feature maps obtained from B-mode US data and bone shadow images. During the evaluation studies we investigate different fusion architectures: early, mid and late fusion [1]. We also investigate bone surface segmentation results if gold standard bone shadow images are used as an additional feature. The bone segmentation networks were trained to minimize the cross-entropy loss. We have used Adam Optimizer with batch size of 8 and a learning rate of 0.0002 for 36,000 iterations. In addition to the previously error metrics explained in this section, we also evaluate the average Euclidean distance (AED) error for the task of bone segmentation. AED was calculated between the automatically segmented bone surfaces and the manual expert segmentation [1].

Results

Our experiments were conducted using the Keras framework and Tensorflow as backend with an Intel Xeon CPU at 3.00GHz and an Nvidia Titan-X GPU with 8GB of memory. Our network converged in about 8 hours during the training process. Testing on average took 54 milliseconds in total for bone shadow and bone surface segmentation.

Quantitative results

Bone shadow segmentation: Table 1 shows the performance difference of bone shadow segmentation methods investigated. Overall our method outperforms previous state-of-the-art GAN architectures and the local phase-based bone shadow enhancement method [7]. The local phase image-based method proposed in [7] achieved the lowest DSC value (0.28). However, we would like to mention that in the original work of [7] a ROI, covering a bone interface spanning the full width of the image, was selected during quantitative evaluation. During our analysis we did not select a ROI and rather used the full B-mode US image. Our generator network, without the discriminator, achieved average dice value of 0.67. While adding the discriminator resulted in 39% improvement in Dice value. Our proposed cGAN architecture achieves 8% and 3% improvement, in DSC value, over the state-of-the-art GAN architectures proposed in [12,18] respectively. A paired *t*-test, for IoU, DSC and AED results at a %5 significance level, between our proposed network and the networks in [12,18] achieved *p* values less than 0.05 indicating that the improvements of our method are statistically significant. The improvement over the U-net architecture [20] was 46% for DSC value. We have also observed that our generator network, without discriminator, outperforms U-net [20] by 6% in Dice value.

Bone surface segmentation: Quantitative results for bone surface segmentation are presented in Table 2. The average numerical error calculations show that the late-fusion design had the lowest errors, and the highest average IoU and Dice (Table 2) (Note: In all tables, the results of the method that outperformed other methods were indicated in bold). A paired *t*-test, for IoU, DSC and AED results at a %5 significance level, between our proposed network and the networks in [12,18] achieved *p* values less than 0.05 indicating that the improvements of our method are statistically significant. There was no statistical significance when using gold standard bone shadow images and the bone shadow images generated using the proposed design for late fusion design. When using local phase image features as an additional feature for our multi-feature CNN architecture [1] the AED error was 0.30 mm compared to 0.11 mm when using the generated bone shadow images.

Qualitative results

Qualitative results of our proposed model are shown in Fig. 4. We demonstrate five examples of in vivo US B-mode images bone types, namely: femur, tibia, radius, knee, and spine, where red pixels indicate high prediction scores while blue pixels indicate low prediction scores for the prediction. Gold standard bone shadow images obtained by an expert are displayed followed by generated bone shadow results obtained using the convolutional network presented by Ronneberger et al. [20] and generative networks in [12,18] and our proposed model, as shown in Fig. 4d through g. In Fig. 4c, we demonstrate shadow results obtained using local phase-based ultrasound transmission maps method presented in [7]. Investigating the qualitative results we can conclude that our proposed method segments bone shadow images with minimal artifacts.

Discussion and conclusions

A method, based on a novel GAN, for real-time and accurate segmentation of bone shadow regions from in vivo US scans was proposed. Our model has two main networks: (1) a cGAN to generate bone shadow images and (2) a segmentation network that will take the generated bone shadow data in conjunction with B-mode US data for localization of bone surfaces. Our integral component of building the generator and discriminator was the skip and projection blocks. To the best of our knowledge, this was not previously investigated in the community. We also would like to mention that this is the first work proposing a novel cGAN architecture for the task of bone shadow segmentation. The projection blocks allow semantic information to be more efficiently passed forward in the network while progressively increasing fea-

Table 1 Bone shadow segmentation error metrics

Method	IoU%	Dice	Rand	Hamming
Dataset I - Sonix-Touch US machine				
LP-based transmission maps [7]	0.2350	0.2186	0.9993	0.8722
Ronneberger et al. [20]	0.5242	0.6504	0.9897	0.5649
Generator network only	0.5927	0.6839	0.9867	0.4019
Radford et al. [18]	0.8015	0.8972	0.6371	0.1951
Isola et al. [12]	0.8628	0.9404	0.6364	0.1501
Ours	0.9277	0.9603	0.4841	0.0873
Dataset II - Clarius C3 US probe				
LP-based transmission maps [7]	0.2670	0.2802	0.9983	0.8722
Ronneberger et al. [20]	0.4726	0.6374	0.9857	0.5272
Generator network only	0.5157	0.6755	0.9831	0.4848
Radford et al. [18]	0.7965	0.8620	0.6685	0.2034
Isola et al. [12]	0.8424	0.9015	0.6730	0.1575
Ours	0.9023	0.9354	0.5990	0.0976

Table 2 Bone segmentation error metrics

Method	IoU%	Dice	Rand	Hamming	AED (mm)
Dataset I - Sonix-Touch US machine					
Radford et al. [18] BM & BS early fusion	0.8511	0.9368	0.7507	0.1003	0.5863
Radford et al. [18] BM & BS mid fusion	0.8735	0.9166	0.6201	0.1297	0.3972
Radford et al. [18] BM & BS late fusion	0.8834	0.9223	0.5961	0.1165	0.3258
Isola et al. [12] BM & BS early fusion	0.8877	0.9248	0.5847	0.1122	0.3105
Isola et al. [12] BM & BS mid fusion	0.8958	0.9294	0.5646	0.1041	0.2976
Isola et al. [12] BM & BS late fusion	0.9076	0.9368	0.5430	0.0923	0.1776
Ours BM & LP late fusion	0.8892	0.9354	0.8290	0.1107	0.3059
Ours BM & GS late fusion	0.9779	0.9826	0.3062	0.0220	0.1059
Ours BM & BS early fusion	0.9670	0.9746	0.5775	0.0329	0.1164
Ours BM & BS mid fusion	0.9694	0.9770	0.4668	0.03054	0.1089
Ours BM & BS late fusion	0.9803	0.9833	0.3644	0.0196	0.1032
Dataset II - Clarius C3 US probe					
Radford et al. [18] BM & BS early fusion	0.8388	0.8995	0.6782	0.1611	0.8542
Radford et al. [18] BM & BS mid fusion	0.8513	0.9369	0.7512	0.1001	0.6753
Radford et al. [18] BM & BS late fusion	0.8731	0.9356	0.7474	0.1029	0.4227
Isola et al. [12] BM & BS early fusion	0.8655	0.9361	0.7496	0.1019	0.3814
Isola et al. [12] BM & BS mid fusion	0.8986	0.9409	0.8105	0.1013	0.2669
Isola et al. [12] BM & BS late fusion	0.9146	0.9442	0.5543	0.0853	0.1973
Ours BM & LP late fusion	0.8730	0.9250	0.8390	0.1269	0.4215
Ours BM & GS late fusion	0.9695	0.9781	0.4353	0.0304	0.1106
Ours BM & BS early fusion	0.9315	0.9555	0.5347	0.06843	0.1655
Ours BM & BS mid fusion	0.9526	0.9692	0.5081	0.0473	0.1306
Ours BM & BS late fusion	0.9625	0.9752	0.4908	0.0374	0.1129

BM B-mode US image, *BS* bone shadow image, *LP* local phase image, *GS* gold standard image

ture map sizes, compared to simple convolutions which is used in many designs including in [20]. By implementing these projection blocks, we allow to have more comprehen-

sive feature maps that improve the bone shadow generation. We have also extended the depth of the discriminator used in the state-of-the-art [12]. This is one of the reasons why

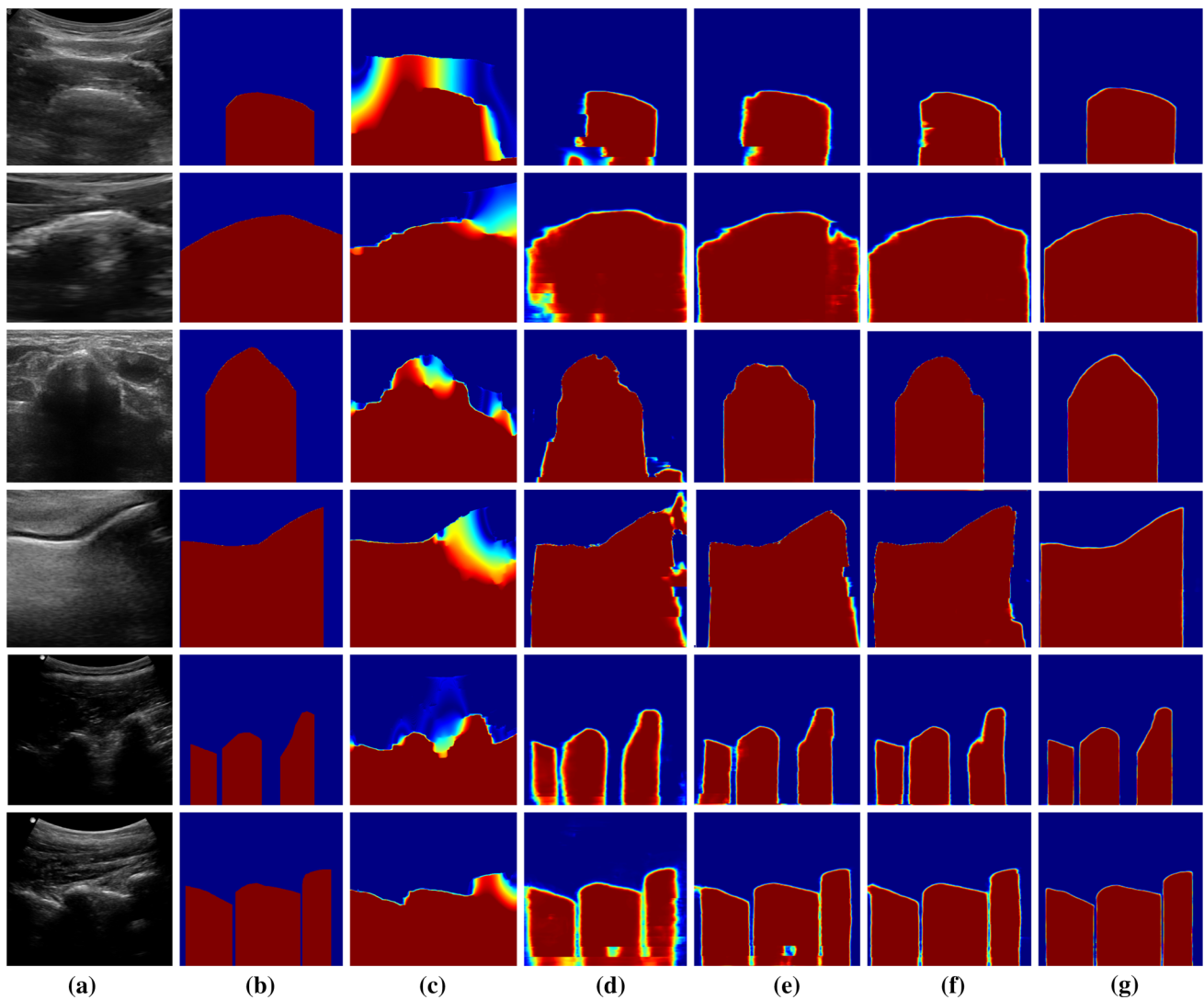


Fig. 4 Qualitative results for bone shadow segmentation. **a** In vivo B-mode US images of femur, tibia, radius, knee, and spine. **b** Gold standard bone shadow images. **c** Bone shadow results obtained using local phase-based ultrasound transmission maps method presented in

[7]. **d** Bone shadow results obtained using Ronneberger et al. [20] **e** Bone shadow results obtained using Radford et al. [18] **f** Bone shadow results obtained using Isola et al. [12]. **g** Bone shadow results obtained using our proposed cGAN

our cGAN outperformed other state-of-the-art networks on this testing data set. Based on these results, we can conclude that having a cGAN with prior information can significantly improve the results for the task at hand. In this study we have also shown the importance of adversarial training. The success of well trained CNN architectures is effected if the architecture is deployed on test data coming from different centers, vendors, or changing acquisition parameters. For US data, even when the machine is from the same vendor the image acquisition settings can be adjusted from one scanning procedure to the next. BMI of the patient, orientation of the transducer with respect to the imaged anatomy will also change the appearance of the collected data drastically. We have shown that GAN are more robust to these conditions. We

have also investigated how to combine information from bone shadow and B-mode US data by analyzing different fusion strategies. Our results demonstrate that for the task of bone segmentation fusing B-mode US and bone shadow features at a later stage outperforms early and mid fusion, specifically for the dataset obtained from Clarius C3 US probe. One of the advantages of the proposed work is that bone shadow features are obtained instantaneously making the computational time required suitable for real-time applications. Our future work will involve (1) extensive clinical validation of the proposed GAN-based method on data obtained from subjects who have differing pathology in their bone such as fracture or bone deformity such as scoliosis. We will also extend our network architecture to process volumetric US data [9].

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

- Alsina 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
- Baka N, Leenstra S, van Walsum T (2017) Ultrasound aided vertebral level localization for lumbar surgery. *IEEE Trans Med Imaging* 36(10):2138–2147
- Berton F, Cheriet F, Miron MC, Laporte C (2016) Segmentation of the spinous process and its acoustic shadow in vertebral ultrasound images. *Comput Biol Med* 72:201–211
- Cernazanu-Glavan C, Holban S (2013) Segmentation of bone structure in X-ray images using convolutional neural network. *Adv Electr Comput Eng* 13(1):87–94
- El-Hariri H, Mulpuri K, Hodgson A, Garbi R (2019) Comparative evaluation of hand-engineered and deep-learned features for neonatal hip bone segmentation in ultrasound. In: International conference on medical image computing and computer-assisted intervention. Springer, Berlin, pp. 12–20
- Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. In: Ghahramani Z, Welling M, Cortes C, Lawrence ND, Weinberger KQ (eds) *Advances in neural information processing systems* 27. Curran Associates, Inc., pp 2672–2680. <http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf>
- Hacihaliloglu I (2017) Enhancement of bone shadow region using local phase-based ultrasound transmission maps. *Int J Comput Assist Radiol Surg* 12(6):951–960
- Hacihaliloglu I (2017) Ultrasound imaging and segmentation of bone surfaces: a review. *Technology* 05(02):74–80
- Hacihaliloglu I, Guy P, Hodgson AJ, Abugharbieh R (2014) Volume-specific parameter optimization of 3d local phase features for improved extraction of bone surfaces in ultrasound. *Int J Med Robot Comput Assist Surg* 10(4):461–473
- Hellier P, Coupé P, Morandi X, Collins DL (2010) An automatic geometrical and statistical method to detect acoustic shadows in intraoperative ultrasound brain images. *Med Image Anal* 14(2):195–204
- Hu R, Singla R, Deeba F, Rohling RN (2019) Acoustic shadow detection: study and statistics of b-mode and radiofrequency data. *Ultrasound Med Biol* 45(8):2248–2257
- Isola P, Zhu JY, Zhou T, Efros AA (2017) Image-to-image translation with conditional adversarial networks. In: 2017 IEEE Conference on computer vision and pattern recognition (CVPR), pp. 5967–5976. IEEE
- Jain V, Bollmann B, Richardson M, Berger DR, Helmstaedter MN, Briggman KL, Denk W, Bowden JB, Mendenhall JM, Abraham WC, Harris KM, Kasthuri N, Hayworth KJ, Schalek R, Tapia JC, Lichtman JW, Seung SH (2010) Boundary learning by optimization with topological constraints. In: 2010 IEEE Conference on computer vision and pattern recognition (CVPR), pp. 2488–2495. IEEE
- Karamalis A, Wein W, Klein T, Navab N (2012) Ultrasound confidence maps using random walks. *Med Image Anal* 16(6):1101–1112
- Klein T, Wells WM (2015) Rf ultrasound distribution-based confidence maps. In: International conference on medical image computing and computer-assisted intervention, vol 9350. Springer, Berlin, pp. 595–602
- Laina I, Rupprecht C, Belagiannis V, Tombari F, Navab N (2016) Deeper depth prediction with fully convolutional residual networks. In: 2016 4th international conference on 3D vision (3DV), pp. 239–248. IEEE
- Meng Q, Housden J, Matthew J, Rueckert D, Schnabel JA, Kainz B, Sinclair M, Zimmer V, Hou B, Rajchl M, Toussaint N, Oktay O, Schlemper J, Gomez A (2019) Weakly supervised estimation of shadow confidence maps in fetal ultrasound imaging. *IEEE Trans Med Imaging* 38(12):2755–2767
- Radford A, Metz L, Chintala S (2016) Unsupervised representation learning with deep convolutional generative adversarial networks. In: 4th International conference on learning representations, ICLR 2016, San Juan, Puerto Rico, May 2–4, 2016, conference track proceedings. [arXiv:1511.06434](https://arxiv.org/abs/1511.06434)
- Rand WM (1971) Objective criteria for the evaluation of clustering methods. *J Am Stat Assoc* 66(336):846–850
- 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, Berlin, pp 234–241
- 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, Berlin, pp 271–297
- 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
- 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, vol 11073. Springer, Berlin, pp 134–142
- Xu B, Wang N, Chen T, Li M (2015) Empirical evaluation of rectified activations in convolutional network. [arXiv preprint arXiv:1505.00853](https://arxiv.org/abs/1505.00853)
- Zheng G, Nolte LP (2015) Computer-assisted orthopedic surgery: current state and future perspective. *Front Surg* 2:66

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.