# GAN-based Realistic Bone Ultrasound Image and Label Synthesis for Improved Segmentation*

Ahmed Z. Alsinan[1], Charles Rule[2], Michael Vives[3], Vishal M. Patel[4], and Ilker Hacihaliloglu[5,6]

[1] Department of Electrical and Computer Engineering, Rutgers University, Piscataway, NJ
[2] Department of Computer Science, Rutgers University, Piscataway, NJ
[3] Department of Orthopedics, Rutgers New Jersey Medical School, Newark, NJ
[4] Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD
[5] Department of Biomedical Engineering, Rutgers University, Piscataway, NJ
[6] Department of Radiology, Rutgers University Robert Wood Johnson Medical School, New Brunswick, NJ

**Abstract.** To provide a safe alternative, for intra-operative fluoroscopy, ultrasound (US) has been investigated as an alternative safe imaging modality for various computer assisted orthopedic surgery (CAOS) procedures. However, low signal to noise ratio, imaging artifacts and bone surfaces appearing several millimeters (mm) in thickness have hindered the wide spread application of US in CAOS. In order to provide a solution for these problems, research has focused on the development of accurate, robust and real-time bone segmentation methods. Most recently methods based on deep learning have shown very promising results. However, scarcity of bone US data introduces significant challenges when training deep learning models. In this work, we propose a computational method, based on a novel generative adversarial network (GAN) architecture, to (1) produce synthetic B-mode US images and (2) their corresponding segmented bone surface masks in real-time. We show how a duality concept can be implemented for such tasks. Armed by two convolutional blocks, referred to as self-projection and self-attention blocks, our proposed GAN model synthesizes realistic B-mode bone US image and segmented bone masks. Quantitative and qualitative evaluation studies are performed on 1235 scans collected from 27 subjects using two different US machines to show comparison results of our model against state-of-the-art GANs for the task of bone surface segmentation using U-net.

**Keywords:** orthopedic surgery · segmentation · ultrasound · bone · generative adversarial network · deep learning

* This work was supported in part by 2017 North American Spine Society Young Investigator grant.

## 1 Introduction

Segmentation of bone surfaces from intra-operative US data is an important step for US-guided CAOS procedures. Due to the success of deep learning methods in medical image analysis, recent research has focused on the use of convolutional neural networks (CNNs) for accurate, robust, and real-time segmentation of bone surfaces [1, 14]. However, scarcity of data size, due to a lack of standardized data and patient privacy concerns, is a major challenge in applying deep learning methods in the medical imaging field. This is specifically a challenge due to the fact that US is not a standard imaging modality in CAOS and US-guided CAOS procedures are not common. Another limiting factor is the manual data collection procedure: sub-optimal orientation of the US transducer with respect to the imaged bone anatomy will result in the acquisition of low quality bone scans [4].

Increasing the size of existing datasets through data augmentation in order to improve models' performance is extensively investigated by various researchers [9]. Earlier work has focused on the introduction of hand crafted image transformations such as random rotations, translations, nonlinear deformations. However, such augmentation methods are limited in their ability to mimic real variations and are highly sensitive to the parameter choice [17]. While transfer learning methods [11], that first train on large datasets then fine-tune on smaller datasets achieve state-of-the-art results on natural image datasets, these methods often do not suit medical image data and offer relatively little benefit to performance [11]. This is especially very problematic for bone US data since its very limited compared to larger medical data such as chest X-ray images. This gap in performance is due to the difference between medical images' features and natural images' features. Furthermore, medical images are often 3D, and there is no streamlined way to transfer 2D feature knowledge into 3D feature knowledge. One approach to overcome this problem is by using unsupervised feature extractors that have only been trained on medical images, however, this requires the target network architecture to be similar to the feature extractors' source architecture, which is uncommon. Image generation methods have recently become a popular solution for the challenge of creating large amounts of training data for deep learning [13]. Generative Adversarial Networks (GANs) have been used in diverse contexts such as unsupervised representation learning [10], image-to-image translation [5] and unsupervised domain adaptation of multi-modal medical imaging data [6]. This groundwork of successful research demonstrates GANs' potential for augmenting small datasets of medical images.

In this work, we propose a computational method, based on a GAN architecture specifically designed to (1) produce synthetic B-mode bone US images and (2) generate their corresponding segmented bone surfaces which can be used as labels. Based on [8] and [16], we show that a duality concept can be adopted for such tasks when implemented by two convolutional blocks, referred to as self-projection and self-attention blocks. We have conducted quantitative and qualitative evaluation studies on 1235 scans collected from 27 subjects using two different US machines. Furthermore, we show comparison results of our model

against state-of-the-art GANs presented in [10] and [5] for the task of generating B-mode bone US images. We also evaluate bone surface segmentation accuracy using synthesized B-mode bone US images generated by the networks investigated when tested on Ronneberger's et. al. [12] U-net architecture. Our work is the first report for generating simultaneous B-mode bone US data and corresponding segmentation labels which we believe to be a novel contribution in the field of US-guided CAOS.

## 2 Proposed Method

### 2.1 Network Architecture

Our architecture is based on the common GAN layout utilizing two co-existing neural networks; a generator $G$ that generate synthetic samples and a discriminator $D$ which attempts to discriminate between these generated synthetic samples and real ones [3]. The generator network transforms some pure random noise vectors $z$ (typically a Gaussian) sampled from a prior distribution $p_z(z)$ into new samples such that $\mathbf{x} = G(\mathbf{z})$. The generated image $x_g$ is expected to resemble the real images $x_r$. On the other hand, the discriminator $D$ has both: (1) real samples with distribution $p_r(x)$ as well as (2) generated samples with distribution $p_g(x)$ and its output $y_s = D(\mathbf{x})$. The gradient information is back-propagated from the discriminator to the generator and hence, the generator optimizes its parameters to generate better images. Gradient-based methods have been proposed to train such a GAN as saddle point optimization problem. However, an imbalance between the training of the generator and the discriminator might occur if the Jensen–Shannon (JS) divergence was used [15] and the discriminator will more likely be too strong, which makes the generator weakly-trained. Moreover, the problem of mode collapse would arise when the distribution $p_g(x)$ learned by the generator was based on limited modes of the real samples distribution $p_r(x)$. This results in weak and limited generations of images. The training of our proposed GAN follows the typical optimization problem such that the discriminator $D$ is trying to maximize and the generator $G$ is trying to minimize the following objective function $\mathcal{L}(D, G)$:

$$\min_G \max_D \mathcal{L}(D, G) = \underset{x_r \sim p_{r(x)}}{E} [\log D(x, y)] + \underset{z \sim p_{z(z)}}{E} [\log(1 - D(x, z))] ;$$

In our generator architecture design 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 both US images and their corresponding segmentation images. We adopt the duality concept presented by [8] with our generator $G$ and discriminator $D$ both incorporating dual information into account. Therefore, our proposed GAN architecture generates segmentation masks/label in addition to the synthesized B-mode US images. This is achieved by modifying the GAN architecture to use two-channel images. In vivo real B-mode US data was assigned to the first channel and expert bone segmentation

was assigned to the second channel. Based on [7] and [16], we also employ a self-projection and self-attention blocks into the GAN model as shown in Figure 1. Our input is processed through convolutional blocks, with each block consisting of several convolutional layers. Our projection blocks, denoted as $P$, we add a $1\times1$ convolution to the projected input that is fed-forward through a $1\times1$ convolution, a $3\times3$ convolution, and another $1\times1$ convolution with each convolution operation followed by batch normalization and rectified linear unit (ReLU) activation. We also use a stride of 2 convolutions to upsample the feature maps. On the other hand, our self-attention block, denoted as $A$, consists of a $1\times1$ convolution (followed by by batch normalization and Leaky ReLU activation) that is (1) multiplied by a transposed $1\times1$ convoluted replica resulting in an attention map and (2) multiplied by the attention map to generate self-attention feature maps. The self-attention approach helps modeling wider range image regions. With self-attention features, the generator can associate fine details at every location and associate them with similar portions of the image. In addition, the discriminator can now enforce complicated geometric constraints relative to the overall image [16]. The architecture of the generator can be summarized as:

- **encoder:** $A32$ $P32$ - $A64$ $P64$ - $A128$ $P128$ - $A256$ $P256$ - $A512$ $P512$
- **decoder:** $A512$ $P512$ - $A256$ $P256$ - $A128$ $P128$ - $A64$ $P64$ - $A32$ $P32$

In our discriminator model a two-input $N\times N$ PatchGAN-like discriminator [5] was used to classify $N\times N$ patches of the input image as real or synthetic. Our discriminator architecture consists of five convolutional blocks, with a final convolution is applied to the last layer to map the 1-dimensional output before applying a Sigmoid function. Batch normalization operations were 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:** $A32$ $P32$ - $A64$ $P64$ - $A128$ $P128$ - $A256$ $P256$ - $A512$ $P512$

## 3 Experimental Results

### 3.1 Data Acquisition

To conduct our experiments that particularly target the problem of data limitation in the US-guided CAOS field, we have collected 1235 in vivo B-mode US images categorized into four groups of bone structures: radius, femur, spine and tibia. Data were collected upon obtaining the approval of the institutional review board (IRB). Depth settings and image resolutions varied between 3-8 cm, and 0.12-0.19 mm, respectively. All the collected scans were scaled to a standardized size of $256\times256$ and manually segmented by an expert ultrasonographer. Two imaging devices were used to collect data:

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

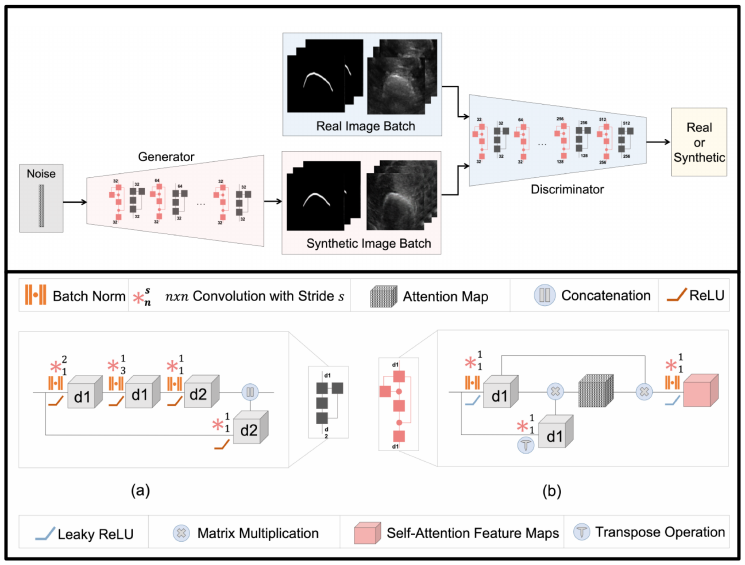


**Fig. 1.** Top: An overview of our proposed GAN architecture with its self-projection and attention blocks based generator and patchGAN-like discriminator. Bottom: Our proposed (a) self-projection block and (b) self-attention block.

have collected 1000 scans from 23 subjects. 400 scans from the Sonix touch, using random split, were used for training the GANS, 300 scans were used for training the U-net, and 300 scans were used for testing. We repeated this process 3 times and during random split same patient data was not included in the training and testing data.

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 Clarius data was used for testing.

We conducted our experiments 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 GAN converged in about 2 hours during the training process. Testing on average took 35 milliseconds. For our experiments, the proposed network and those presented in [10] and [5] were implemented as per the recommendations by their respective authors. For consistency, we used an Adam solver with learning rate of 0.0002, an exponential decay rate for the first and second moment estimates of $\beta_1 = 0.5$ and $\beta_2 = 0.999$, with a mini-batch SGD for all models considered.

### 3.2 Quantitative Results

Quantitative evaluation of our proposed GAN architecture was performed against three methods [10, 5, 2]. In order to show that the synthesized US images are useful for improving the performance of a supervised segmentation network we use the well known U-net architecture described in [12]. We would like to mention that the U-net architecture used in this work is not the main contribution of this work, since the synthesized images can be used in conjunction with other CNN-based network architectures [1, 14]. If a GAN architecture captures the target distribution correctly it should generate a new set of training images (synthesized images) that should be indistinguishable from the in vivo real B-mode US data. Therefore, a U-net trained on either of these datasets, assuming they have the same size, should produce similar results. To evaluate this we have performed the following studies: (1) train U-net using limited in vivo real B-mode US data and test using in vivo real B-mode US data, (2) train U-net using limited in vivo real B-mode US data together with synthesized B-mode US data and test using in vivo real B-mode US data, (3) train U-net using synthesized B-mode US data and test U-net using in vivo real B-mode US data, (4) train U-net using real in vivo B-mode US data and test U-net using synthesized B-mode US data. Bone segmentation results are evaluated by calculating Dice, Rand error (Rand), the structural similarity index (SSIM), Hamming Loss, intersection over union (IoU) and average Euclidean distance (AED) [1].

Table 1 shows the performance of bone surface segmentation when U-net [12] is trained on various combinations of in vivo real B-mode US data and synthesized B-mode US data. We observe that adding synthesized images to the real in vivo B-mode US images improves the accuracy over the corresponding real-only counterpart. Overall our method outperforms previous state-of-the-art GAN architectures. In particular, it achieves 7%/7% and 5%/4% improvement for both data sets (Sonix/Clarius), in IoU value, over the GAN architectures proposed in [10], [5] respectively. A paired t-test, for IoU, Dice and AED results at a %5 significance level, between our proposed network and the networks in [10], [5] achieved p-values less than 0.05 indicating that the improvements of our method are statistically significant. Quantitative results presented in Tables 2-3 show that our proposed GAN architecture captures the target distribution better compared to the methods in [10, 5] achieving improved results for IoU, Dice, Rand and AED evaluation metrics. Results in Table 2 were obtained when U-net [12] was trained using 600 synthetic B-mode US data generated using the proposed and two other architectures [10, 5]. Testing was performed using 300 in vivo real B-mode US data obtained from Sonix Touch and 235 in vivo real B-mode US data obtained from Clarius probe. In Table 3 results were obtained when U-net [12] was trained using 535 in vivo real B-mode US data obtained from SonixTouch and Clarius probe. Testing was performed using 600 synthetic B-mode US data generated using the proposed method and two other GAN architectures [10, 5].

**Table 1.** Quantitative results for bone surface segmentation using U-net [12]. Testing was done using 300 in vivo real B-mode US data obtained from Sonix Touch for Dataset I. For Dataset II testing was performed using all the 235 scans collected from Clarius C3 US probe. Notation note: number of in vivo real B-mode US images/number of synthetic B-mode US images used for training- GAN method used.

| Method | IoU% | Dice | Rand | SSIM | Hamming | AED |
|---|---|---|---|---|---|---|
| Dataset I - Sonix-Touch US | | | | | | |
| 300/0 - N/A | 0.7703 | 0.8642 | 0.9264 | 0.1106 | 0.2280 | 0.9386 |
| 300/300 - Radford et. al. [10] | 0.8391 | 0.9036 | 0.8522 | 0.3588 | 0.1608 | 0.8146 |
| 300/300 - Isola et. al. [5] | 0.8516 | 0.9117 | 0.8477 | 0.5401 | 0.1483 | 0.5687 |
| 300/300 - Ours | **0.8977** | **0.9400** | **0.7899** | **0.7038** | **0.1022** | **0.2985** |
| 300/600 - Radford et. al. [10] | 0.8621 | 0.9183 | 0.7084 | 0.5540 | 0.3826 | 0.1378 |
| 300/600 - Arjovsky et. al. [2] | 0.8827 | 0.9255 | 0.6876 | 0.6038 | 0.1244 | 0.3220 |
| 300/600 - Isola et. al. [5] | 0.8943 | 0.9395 | 0.6657 | 0.7021 | 0.1035 | 0.2896 |
| 300/600 - Ours | **0.9309** | **0.9580** | **0.6125** | **0.7586** | **0.0690** | **0.1596** |
| Dataset II - Clarius C3 US | | | | | | |
| 300/0 - N/A | 0.7594 | 0.8564 | 0.9350 | 0.1086 | 0.2405 | 0.7821 |
| 300/300 - Radford et. al. [10] | 0.8128 | 0.8869 | 0.8678 | 0.2750 | 0.1871 | 0.7536 |
| 300/300 - Isola et. al. [5] | 0.8322 | 0.9126 | 0.8463 | 0.3483 | 0.1593 | 0.8211 |
| 300/300 - Ours | **0.8753** | **0.9193** | **0.8381** | **0.5861** | **0.1278** | **0.1970** |
| 300/600 - Radford et. al. [10] | 0.8458 | 0.9128 | 0.8483 | 0.4822 | 0.1486 | 0.6217 |
| 300/600 - Arjovsky et. al. [2] | 0.8531 | 0.9196 | 0.8104 | 0.5480 | 0.1311 | 0.4853 |
| 300/600 - Isola et. al. [5] | 0.8646 | 0.9214 | 0.7903 | 0.5728 | 0.1275 | 0.3482 |
| 300/600 - Ours | **0.9225** | **0.9536** | **0.7636** | **0.7408** | **0.0774** | **0.1583** |

### 3.3 Qualitative Results

Qualitative results of our proposed GAN model are shown in Figure 2. In each row of Figure 2, we demonstrate one example of in vivo real B-mode US image (four examples in total). Columns are labeled alphabetically where we show in (a)-right: real in vivo B-mode US images and in (a)-left: their corresponding bone surface segmentations obtained by an expert. Figure 2 columns (b) through (d) demonstrate synthetic B-mode US images (right) and their corresponding synthetic bone surface segmentations as generated by [10], [5] and our proposed model, respectively. Investigating the results we can infer that our proposed method results in fewer artifacts compared to the state-of-the-art [10, 5].

## 4 Discussion and Conclusion

In this paper, a novel GAN model for real-time and accurate B-mode bone US image generation is proposed. Our model has been implemented using two main components: (1) a generator that produces synthesized B-mode US as well as bone surface images and (2) a PatchGAN-like discriminator [5] that was used

**Table 2.** Quantitative results for bone surface segmentation. Results were obtained when U-net [12] was trained using 600 synthetic B-mode US data generated using the proposed method and [10, 5]. Testing was performed using 300 in vivo real B-mode US data (Sonix Touch) and 235 in vivo real B-mode US data (Clarius probe). Notation note: method used-blocks type.

| Method | IoU% | Dice | Rand | Hamming | AED |
|---|---|---|---|---|---|
| Radford et. al. [10] | 0.8471 | 0.9158 | 0.8483 | 0.1783 | 0.7133 |
| Isola et. al. [5] | 0.8625 | 0.9115 | 0.8284 | 0.1183 | 0.4540 |
| Ours-none | 0.6952 | 0.8068 | 0.9845 | 0.1967 | 0.9347 |
| Ours-self-projection only | 0.8356 | 0.9023 | 0.8615 | 0.1883 | 0.8053 |
| Ours-self-attention only | 0.8502 | 0.9104 | 0.8816 | 0.1668 | 0.5063 |
| Ours-self-projection & self-attention | **0.9054** | **0.9766** | **0.8169** | **0.1208** | **0.1852** |

**Table 3.** Quantitative results for bone surface segmentation. Results were obtained when U-net [12] was trained using 535 in vivo real B-mode US data obtained from Sonix Touch and Clarius probe. Testing was performed using 600 synthetic B-mode US data generated using the proposed method and two other GAN architectures [10, 5]. Notation note: number of synthetic B-mode images used for testing - method used.

| Method | IoU% | Dice | Rand | Hamming | AED |
|---|---|---|---|---|---|
| 600-B-mode-Radford et. al. [10] | 0.8726 | 0.9158 | 0.8464 | 0.1405 | 0.4610 |
| 600-B-mode-Isola et. al. [5] | 0.8933 | 0.9304 | 0.7629 | 0.1108 | 0.2814 |
| 600-B-mode-Ours | **0.9357** | **0.9640** | **0.7195** | **0.0496** | **0.1952** |

to classify $N \times N$ patches of the input images as real or synthetic. We have employed two integral components of building the generator and discriminator: a self-projection and self-attention blocks. With self-attention features the generator can associate fine details at every location and associate them with similar portions of the image. The main benefit of our self-attention blocks is that they leverage complementary features in distant portions of the image rather than local regions of fixed shape especially for images with complex structural patterns, e. g. US B-mode images. The relationship between near and far pixels is learned, which allows the model to focus on separated structurally relevant features. Since the task is to replicate the relationship between the US B-mode and segmentation images, our model's ability to span a larger region in the image to create features gives it an advantage over the classic GAN model, which is limited by its filter size. In a classic GAN model, the relationship between the segment and US features is likely to be diluted across local features, while in a self-attention model the relationship is preserved by these larger feature regions. Additionally, the self-attention discriminator used checks for consistency in features in distant areas, which enforces accurate reproduction of geometric patterns in the B-mode US images and leads to higher-quality augmented data. On the other hand, self-projection blocks allow semantic information to be more

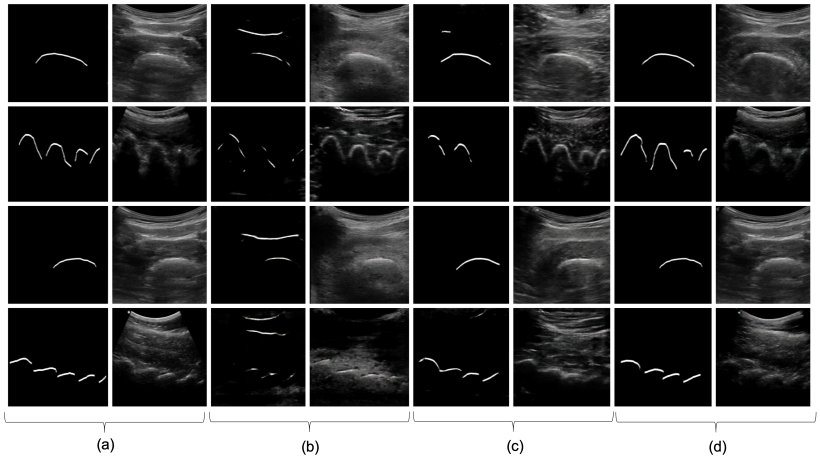


**Fig. 2.** Four examples of B-mode US images and their corresponding bone segmentation mask images are displayed in four rows. Columns are labeled alphabetically where we show in (a)-right: real in vivo B-mode US images and in (a)-left: their corresponding bone surface segmentations mask as obtained by an expert. Columns (b) through (d) demonstrate synthetic B-mode US images (right) and their corresponding synthetic bone surface segmentations as generated by [10], [5] and our proposed model.

efficiently passed forward in the network while progressively increasing feature map sizes, compared to simple convolutions. They allow us to have more comprehensive feature maps. Furthermore, self-projection blocks are also convolutional blocks, and therefore are computationally less expensive to train and infer on. To the best of our knowledge, this was not previously investigated for generating B-mode bone US images. Based on the quantitative results presented, we can conclude that having a self-attention mechanism can significantly improve the results for the image synthesis task at hand. Our future work will involve more extensive clinical validation of the proposed GAN model.

## References

1. Alsinan, A.Z., Patel, V.M., Hacihaliloglu, I.: Automatic segmentation of bone surfaces from ultrasound using a filter layer guided cnn. International Journal of Computer Assisted Radiology and Surgery **14**(5), 775–783 (2019)
2. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. In: Precup, D., Teh, Y.W. (eds.) Proceedings of the 34th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 70, pp. 214–223. PMLR, International Convention Centre, Sydney, Australia (06–11 Aug 2017)
3. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. Advances in neural information processing systems pp. 2672–2680 (2014)

4. Hacihaliloglu, I., Guy, P., Hodgson, A.J., Abugharbieh, R.: Volume-specific parameter optimization of 3d local phase features for improved extraction of bone surfaces in ultrasound. The International Journal of Medical Robotics and Computer Assisted Surgery **10**(4), 461–473 (2014)
5. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5967–5976. IEEE (2017)
6. Kamnitsas, K., Baumgartner, C., Ledig, C., Newcombe, V.F.J., Simpson, J.P., Kane, A.D., Menon, D.K., Nori, A., Criminisi, A., Rueckert, D., Glocker, B.: Unsupervised domain adaptation in brain lesion segmentation with adversarial networks (2016)
7. Laina, I., Rupprecht, C., Belagiannis, V., Tombari, F., Navab, N.: Deeper depth prediction with fully convolutional residual networks. In: 2016 Fourth International Conference on 3D Vision (3DV). pp. 239–248. IEEE (2016)
8. Neff, T., Payer, C., Štern, D., Urschler, M.: Generative adversarial networks to synthetically augment data for deep learning based image segmentation (05 2018). https://doi.org/10.3217/978-3-85125-603-1-07
9. Payer, C., Stern, D., Bischof, H., Urschler, M.: Regressing Heatmaps for Multiple Landmark Localization Using CNNs, Lecture Notes in Computer Science, vol. 9901, pp. 230–238. Springer International Publishing AG, Switzerland (10 2016)
10. Radford, A., Metz, L., Chintala, S.: 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 (2016), http://arxiv.org/abs/1511.06434
11. Raghu, M., Zhang, C., Kleinberg, J., Bengio, S.: Transfusion: Understanding transfer learning for medical imaging. In: Advances in Neural Information Processing Systems. pp. 3342–3352 (2019)
12. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
13. Shin, H.C., Tenenholtz, N.A., Rogers, J.K., Schwarz, C.G., Senjem, M.L., Gunter, J.L., Andriole, K.P., Michalski, M.: Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In: International workshop on simulation and synthesis in medical imaging. pp. 1–11. Springer (2018)
14. Villa, M., Dardenne, G., Nasan, M., Letissier, H., Hamitouche, C., Stindel, E.: Fcn-based approach for the automatic segmentation of bone surfaces in ultrasound images. International journal of computer assisted radiology and surgery **13**(11), 1707–1716 (2018)
15. Yadav, A.K., Shah, S., Xu, Z., Jacobs, D.W., Goldstein, T.: Stabilizing adversarial nets with prediction methods. In: 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings. OpenReview.net (2018), https://openreview.net/forum?id=Skj8Kag0Z
16. Zhang, H., Goodfellow, I., Metaxas, D., Odena, A.: Self-attention generative adversarial networks (2018)
17. Zhao, A., Balakrishnan, G., Durand, F., Guttag, J.V., Dalca, A.V.: Data augmentation using learned transformations for one-shot medical image segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 8543–8553 (2019)