2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

# MSF-GAN: Multi-Scale Fuzzy Generative Adversarial Network for Breast Ultrasound Image Segmentation

Kuan Huang[1], Yingtao Zhang[2], H. D. Cheng[1,⋆], Ping Xing[3]

*Abstract*— **Automatic breast ultrasound image (BUS) segmentation is still a challenging task due to poor image quality and inherent speckle noise. In this paper, we propose a novel multi-scale fuzzy generative adversarial network (MSF-GAN) for breast ultrasound image segmentation. The proposed MSF-GAN consists of two networks: a generative network to generate segmentation maps for input BUS images, and a discriminative network that employs a multi-scale fuzzy (MSF) entropy module for discrimination. The major contribution of this paper is applying fuzzy logic and fuzzy entropy in the discriminative network which can distinguish the uncertainty of segmentation maps and groundtruth maps and forces the generative network to achieve better segmentation performance. We evaluate the performance of MSF-GAN on three BUS datasets and compare it with six state-of-the-art deep neural network-based methods in terms of five metrics. MSF-GAN achieves the highest mean IoU of 78.75%, 73.30%, and 71.12% on three datasets, respectively.**

## I. Introduction

Breast cancer is the most common cancer (excluding skin cancers) and is the second leading cause of cancer death among US women. Automatic BUS image segmentation can provide radiologists a second opinion and help them make correct decisions and improve the diagnosis accuracy. It separates the tumor regions from the background automatically. However, BUS image segmentation is still a challenging task due to poor image quality and inherent speckle noise [1].

Many approaches have been proposed for BUS image segmentation. These approaches can be divided into five categories: thresholding algorithms, region-growing algorithms, watershed algorithms, graph-based algorithms, and deep neural network-based algorithms [2]. In [3], patch-based LeNet [4], U-Net [5], and FCN [6] perform well for BUS image segmentation on two BUS datasets. Shareef *et al*. [7] propose a small tumor-aware network to better segment breast tumors with different sizes by using kernels with three different sizes at each convolutional layer. Lei *et al*. [8] propose a boundary regularized convolutional encoder-decoder network for the segmentation of breast anatomical layers that is robust to speckle noise and posterior acoustic shadows. They further design a self-co-attention neural network that employs both spatial and channel attention modules to explore contextual relationships in BUS images and achieves better segmentation results [9]. In [10], a medical knowledge constrained deep learning + conditional random fields method is proposed for three-layer BUS image semantic segmentation. To further improve the performance of classic segmentation networks, researchers propose a Generative Adversarial Network (GAN) [11] which employs an adversarial network to guide the segmentation network to generate more accurate segmentation results. Xue *et al*. [12] further propose an adversarial network with multi-scale $L_1$ loss for image segmentation that can learn features in different scales and capture contextual relationships to boost the segmentation accuracy.

Despite the good performance of the above methods, they do not take the uncertainty in BUS images into account. In this study, we propose a novel multi-scale fuzzy generative adversarial network (MSF-GAN) for BUS image segmentation that uses uncertainty maps to train the discriminative network. Inspired by reference [12], the proposed MSF-GAN consists of a generative network (G-net) and a discriminative network (D-net) which respectively minimizes and maximizes the loss functions. The output of G-net is a segmentation map. The proposed MSF-GAN employs a fuzzy attentive feature generator and a multi-scale fuzzy entropy (MSF) module which can transform the segmentation maps and groundtruth maps into the fuzzy domain to measure uncertainty. The multi-scale fuzzy entropy (MSF) module can distinguish the difference in uncertainty maps from two inputs and help to train a better segmentation network. The major contributions of the proposed approach are: (1) Design a novel MSF-GAN for BUS image segmentation that outperforms six state-of-the-art deep neural network-based methods on three BUS datasets in terms of five metrics. (2) Design a fuzzy attentive feature generator to generate fuzzy attentive feature maps for the segmentation maps generated by G-net and groundtruth maps. (3) Design an MSF module to measure the uncertainty in segmentation maps and groundtruth maps and calculate a multi-scale $L_1$ loss on uncertainty maps to help to train the segmentation network.

## II. Method

The proposed MSF-GAN consists of a G-net for the generation of pixel-wise segmentation maps, a D-net for guiding G-net to generate more accurate segmentation maps, and a fuzzy attentive feature generator.

### A. Overview

The architecture of the proposed MSF-GAN is illustrated in Fig. 1. MSF-GAN employs a U-ResNet (a U-shape

⋆Corresponding author: H. D. Cheng hengda.cheng@usu.edu
[1]Kuan Huang, and H. D. Cheng are with the Department of Computer Science, Utah State University, Logan, Utah, USA, 84341
[2]Yingtao Zhang is with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China
[3]Ping Xing is with the Ultrasound Department, the First Affiliated Hospital of Harbin Medical University, Harbin, China

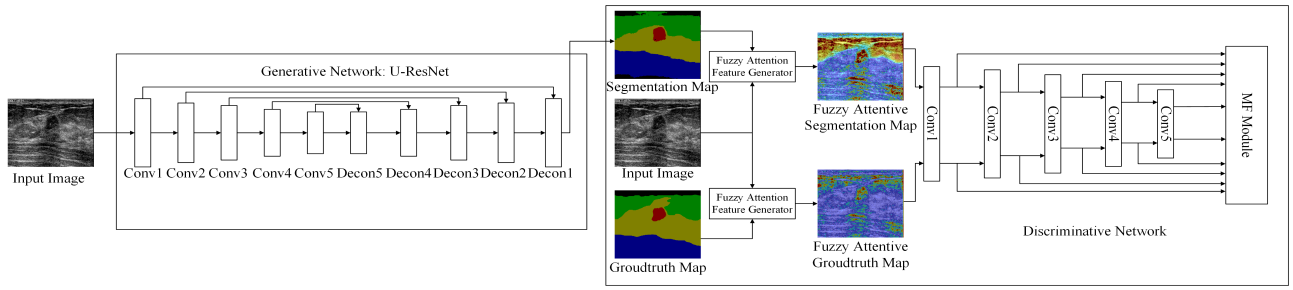


Fig. 1. An overview of the proposed MSF-GAN.

network with ResNet-101 as its backbone) as its G-net to generate pixel-wise segmentation results, denoted as segmentation maps. All input BUS images are first resized to $128 \times 128$ and then fed into G-net. A segmentation map of size $128 \times 128 \times C$ is generated for an input BUS image, where $C$ represents the total number of categories. Each pixel contains $C$ values in the segmentation map and each element represents the probability to the corresponding category. Then, we use a fuzzy attentive feature generator that takes an original BUS image and its groundtruth map as inputs to compute a fuzzy attentive groundtruth map. Similarly, we compute a fuzzy attentive segmentation map by an original BUS image and its segmentation map. The fuzzy attentive feature generator will be introduced in subsection II-B in detail. The D-net is composed of five convolutional layers with kernels of size $4 \times 4$, stride 2, padding 1, and ReLU activation function. It takes a fuzzy attentive groundtruth map and a fuzzy attentive segmentation map as two inputs and calculates a multi-scale $L_1$ loss on their uncertainty maps, which will be introduced in subsection II-C. The objective of G-net is to generate accurate segmentation maps and the objective of D-net is to distinguish the uncertainty of the segmentation maps and groundtruth maps. For an input BUS image, if the uncertainty map of the segmentation map is very close to the uncertainty map of the groundtruth map, then it is hard for D-net to discriminate them. In contrast, if the uncertain map of the segmentation map is not close to the uncertain map of the groundtruth map, it means there still exists uncertainty in the segmentation map. The goal is to make G-net generate very accurate segmentation maps which contain similar uncertainty maps to the groundtruth maps. In this study, we enhance the discriminating ability of the D-net by using a fuzzy attentive feature generator and a multi-scale $L_1$ loss calculated on uncertainty maps and therefore force G-net to generate more accurate segmentation maps that are very close to the groundtruth maps.

### B. Fuzzy Attentive Feature Generator

The target for the fuzzy attentive feature generator is to transform the input of the D-net to the fuzzy domain. Fig. 2 illustrates the proposed fuzzy attentive feature generator. It takes a pair of an original BUS image and its segmentation map generated by the G-net, or a pair of an original BUS image and its groundtruth map as inputs. Specifically, for an original image, its segmentation map and its groundtruth map are individually transformed into the fuzzy domain by a convolutional operator with a kernel size of $1 \times 1$ and Sigmoid function as activation function. The operation of fuzzification can be represented by:

$$F_x = \text{Conv1} \times 1(x) \tag{1}$$

where $x$ can be an original BUS image of size $128 \times 128$, a segmentation map generated by G-net of size $128 \times 128 \times C$, or a groundtruth map of size $128 \times 128 \times C$. After fuzzification, $x$ is transformed into $F_x$ of size $128 \times 128 \times C$. Then, we respectively perform a fuzzy *AND* operator on a pair of the fuzzified original image (denoted as $F_o$) and fuzzified segmentation map (denoted as $F_{pre}$), and on a pair of $F_o$ and the fuzzified groundtruth map (denoted as $F_{gt}$) to generate a fuzzy attentive groundtruth map $FA_{gt}$ and a fuzzy attentive segmentation map $FA_{pre}$. This operation can be represented by:

$$FA_{pre} = \min(F_o, F_{pre}) \tag{2}$$

$$FA_{gt} = \min(F_o, F_{gt}) \tag{3}$$

where min is the *AND* operator in fuzzy logic that performs a pixel-wise minimization operation on its two inputs. $FA_{pre}$ and $FA_{gt}$ are of size $128 \times 128 \times C$. Different from reference [12] that directly uses groundtruth map masked images and segmentation map masked images as the inputs of D-net, we first generate three types of fuzzified maps and then compute two fuzzy attentive maps and use them as the inputs of D-net. We can train D-net better by using the fuzzy attentive maps to extract multi-scale features and calculate a multi-scale $L_1$ loss on uncertainty maps extracted from these fuzzy attentive maps because through a non-linear transformation of the fuzzification and fuzzy *AND* operator in fuzzy feature generator, the fuzzy features are more discriminable than the non-fuzzy features and we can also measure uncertainty on fuzzy features.

### C. Multi-Scale Fuzzy Entropy Module

In D-net, five convolutional layers are used to extract multi-scale features on the input fuzzy attentive groundtruth map $FA_{gt}$ and fuzzy attentive segmentation map $FA_{pre}$. These features are then fed into the proposed MSF module to calculate a multi-scale $L_1$ loss on uncertainty maps, which are calculated via $FA_{gt}$ and $FA_{pre}$. By training a powerful D-net to better discriminate the uncertainty map of $FA_{gt}$

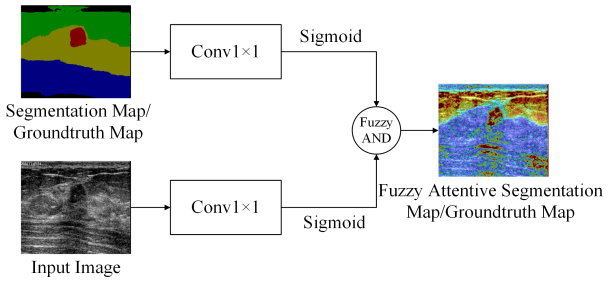


Fig. 2. Illustration of the proposed fuzzy attentive feature generator.

and that of $FA_{pre}$, G-net is forced to generate more accurate segmentation maps.

As shown in Fig. 1, D-net takes a fuzzy attentive groundtruth map $FA_{gt}$ and a fuzzy attentive segmentation map $FA_{pre}$ as the inputs, then employs five convolutional layers to extract multi-scale features. Let $L$ denote the total number of convolutional layers in D-net (here $L = 5$). Let $f^l(FA_{pre})$ and $f^l(FA_{gt})$ denote the feature map extracted by the $l$-th layer of D-net, respectively. Then, it performs a $1 \times 1$ convolution with ReLU activation function on $f^l(FA_{pre})$ and $f^l(FA_{gt})$ respectively to transform their channel number to $C$ to calculate fuzzy entropy. The transformed feature maps are denoted as:

$$T^l_{pre} = \text{conv1} \times 1(f^l(FA_{pre})) \tag{4}$$

$$T^l_{gt} = \text{conv1} \times 1(f^l(FA_{gt})) \tag{5}$$

Then, it calculates the fuzzy entropy on $T^l_{pre}$ and $T^l_{gt}$ to represent their uncertainty maps, respectively:

$$E^l_{pre}(i,j) = -\frac{1}{\log C}\sum_{c=1}^{C} T^l_{pre}(i,j,c) \cdot \log T^l_{pre}(i,j,c) \tag{6}$$

$$E^l_{gt}(i,j) = -\frac{1}{\log C}\sum_{c=1}^{C} T^l_{gt}(i,j,c) \cdot \log T^l_{gt}(i,j,c) \tag{7}$$

where $T^l_{pre}(i,j,c)$ and $T^l_{gt}(i,j,c)$ represent the values of the $i$-th row, $j$-th column and $c$-th channel of $T^l_{pre}$ and $T^l_{gt}$, respectively. It then computes a multi-scale $L_1$ loss on the uncertainty maps $E^l_{pre}(i,j)$ and $E^l_{gt}(i,j)$ by:

$$\min_{\theta_G}\max_{\theta_D} \mathcal{L}(\theta_G, \theta_D) = \frac{1}{N}\sum_{n=1}^{N} \ell_{mae}(E^{l,n}_{pre}, E^{l,n}_{gt}) \tag{8}$$

where $\theta_G$ and $\theta_D$ denote the parameters of G-net and D-net, respectively; $N$ denotes the total number of training images; $E^{l,n}_{pre}$ and $E^{l,n}_{gt}$ denote the uncertainty map extracted by the $l$-th layer on the $n$-th training image, respectively. $\ell_{mae}$ is the Mean Absolute Error (MAE) ($L_1$ loss), defined as:

$$\ell_{mae}(E^l_{pre}, E^l_{gt}) = \frac{1}{L}\sum_{l=1}^{L} \left\| E^l_{pre} - E^l_{gt} \right\|_1 \tag{9}$$

The loss $\mathcal{L}$ in Eq. (8) can capture a rich contextual relationship between pixels by using the multi-scale uncertainty maps $E^l_{pre}$ and $E^l_{gt}$ generated by different convolutional layers. During the training of MSF-GAN, we minimize $\mathcal{L}$ with respect to the parameters $\theta_G$ of G-net, while maximizing it with respect to the parameters $\theta_D$ of D-net. The objective of G-net is to generate accurate segmentation maps that contain similar uncertainty to groundtruth maps so that $\mathcal{L}$ is minimized. The uncertainty is represented by fuzzy entropy. In contrast, the objective of D-net is to distinguish the uncertainty of segmentation maps from the uncertainty of groundtruth maps and therefore force G-net to generate accurate segmentation maps. When D-net is powerful enough, it can distinguish these two kinds of uncertainty maps very well so that $\mathcal{L}$ is maximized. To implement this strategy, we train G-net and D-net in an alternating scheme: first, fix G-net and train D-net to maximize $\mathcal{L}$, and then fix D-net and train G-net to minimize $\mathcal{L}$. During the training procedure, both G-net and D-net are becoming more and more powerful. By using fuzzy attentive feature maps and the multi-scale $L_1$ loss computed from these fuzzy attentive feature maps, the discriminating ability of D-net is further enhanced compared with [12]. Therefore, our more powerful D-net can better guide G-net to generate more accurate segmentation maps close to groundtruth maps.

## III. Experiment Results and Discussion

We evaluate the performance of the proposed MSF-GAN on three datasets: a multi-layer dataset, Dataset B [3] and Dataset BUSI [13]. Dataset B and dataset BUSI are two public BUS datasets where groundtruths annotations only separate tumors and background. Dataset B has 163 images and Dataset BUSI has 780 images. The multi-layer dataset is a private dataset consisting of 325 images. The groundtruth annotations include four breast anatomical layers (fat layer, mammary layer, muscle layer, background layer) and tumors. The privacy of the patient is well protected. The experimental procedures involving human subjects described in this paper were approved by the Institutional Review Board. In total, there are 1268 images used for evaluation.

To ensure a fair comparison, we set these parameters to be the same for all compared methods. All experiments are conducted on Ubuntu 18.04 system with Intel(R) Xeon(R) CPU E5-2620 2.00 GHz and two NVIDIA GeForce 1080Ti graphics cards. An Adam optimizer with learning rate = 0.0001, $\beta_1 = 0.9$, and $\beta_2 = 0.99$ is used for training. The batch size is set as 12, and the number of training epochs is set as 80. The initial weights are initialized randomly. Input images are augmented by horizontal flip, horizontal shift, vertical shift, rotation, zooming, and shear mapping before fed into networks. We employ 10-fold cross-validation to evaluate the performance of MSF-GAN and six compared methods.

We further compare the segmentation performance of MSF-GAN and six state-of-the-art deep neural network-based methods on above mentioned three BUS datasets. The six compared methods are: U-Net with ResNet-101 as its backbone (denoted as U-ResNet), U-Net with VGG-16 as its backbone (denoted as U-VGG), FCN-8s [6], SegAN [12], PSPNet [14], and Deeplabv3+ [15]. We use five metrics for the evaluation. They are: True Positive Ratio (TPR), False

Positive Ratio (FPR), Intersection over Union (IoU), Dice's Coefficient (DSC), and Area Error Ratio (AER).

TABLE I
RESULTS OF MULTI-LAYER SEGMENTATION (IOU (%))

| | Class1 | Class2 | Class3 | Class4 | Class5 | Mean |
|---|---|---|---|---|---|---|
| U-ResNet | 81.50 | 73.41 | 72.07 | 74.47 | 75.29 | 75.35 |
| U-VGG | 70.34 | 66.72 | 66.17 | 65.91 | 74.66 | 68.76 |
| FCN-8s | 82.57 | 75.47 | 75.53 | 78.59 | 74.42 | 77.32 |
| SegAN | 81.68 | 75.89 | 72.53 | 81.69 | 77.23 | 77.80 |
| PSPNet | 82.07 | 74.40 | 74.49 | 77.36 | 74.75 | 76.61 |
| Deeplab | 78.91 | 68.71 | 67.33 | 73.94 | 69.04 | 71.58 |
| MSF-GAN | **83.11** | **77.05** | **73.11** | **81.98** | **78.50** | **78.75** |

* Class1: fat layer, Class2: mammary layer, Class3: muscle layer, Class4: background, Class5: tumor. Bold values are the best results for the corresponding classes.

Table I compares the performance of MSF-GAN and six compared methods on the multi-layer dataset in terms of IoU. On this dataset, MSF-GAN achieves the best segmentation results for all classes in terms of IoU. Specifically, it achieves the highest mean IoU value of 78.75% among five classes including fat layer, mammary layer, muscle layer, background and tumor. It should be noticed that the proposed MSF-GAN outperforms SegAN, which is also a GAN-based network using a multi-scale $L_1$ loss, for all classes in terms of IoU. As shown in Table I, the proposed fuzzy attentive feature generator and multi-scale $L_1$ loss calculated on multi-scale uncertainty maps are efficient to enhance the discriminating ability of D-net, and force G-net to generate more accurate segmentation results.

TABLE II
RESULTS ON PUBLIC DATASETS (%)

| Datasets | Methods | TPR | FPR | IoU | DSC | AER |
|---|---|---|---|---|---|---|
| Dataset B [3] | U-ResNet | 83.58 | **34.40** | 71.43 | 79.45 | **50.82** |
| | U-VGG | 79.30 | 45.84 | 68.16 | 76.40 | 66.54 |
| | FCN-8s | 82.72 | 41.14 | 67.50 | 76.87 | 58.42 |
| | SegAN | 81.13 | 49.96 | 70.11 | 78.05 | 68.83 |
| | PSPNet | 81.08 | 40.42 | 69.77 | 78.24 | 59.34 |
| | Deeplab | 63.68 | 36.06 | 52.93 | 61.91 | 72.38 |
| | MSF-GAN | **84.57** | 40.31 | **73.30** | **81.58** | 55.73 |
| Dataset BUSI [13] | U-ResNet | **79.40** | 46.02 | 69.26 | 77.90 | 66.62 |
| | U-VGG | 78.66 | 41.98 | 68.77 | 77.37 | 63.32 |
| | FCN-8s | 74.23 | 46.69 | 63.16 | 73.03 | 72.63 |
| | SegAN | 76.23 | 25.95 | 69.21 | 77.83 | 49.71 |
| | PSPNet | 77.11 | 46.65 | 65.21 | 74.75 | 69.54 |
| | Deeplab | 59.88 | 39.39 | 49.65 | 59.39 | 79.52 |
| | MSF-GAN | 78.34 | **20.98** | **71.12** | **79.99** | **42.64** |

Table II compares the performance of MSF-GAN and six state-of-the-art methods on Dataset B and Dataset BUSI in terms of TPR, FPR, IoU, DSC and AER. MSF-GAN has the highest TPR value of 84.57%, the highest IoU value of 73.30% and the highest DSC value of 81.58% on Dataset B. MSF-GAN achieves the best performance in terms of IoU, FPR, DSC, and AER and a comparable TPR value on Dataset BUSI. Specifically, it improves the second-best method by 2.69%, 2.68%, 19.15%, and 14.22% for IoU, DSC, FPR and AER, respectively.

## IV. CONCLUSIONS

We propose a novel MSF-GAN method for BUS image segmentation consisting of a generative network and a discriminative network. MSF-GAN employs a fuzzy attentive feature generator to extract fuzzy attentive feature maps respectively from segmentation maps generated by the generative network and from groundtruth maps, and then uses an MSF module to extract multi-scale uncertainty maps from these fuzzy attentive feature maps to calculate a multi-scale $L_1$ loss that can capture the rich contextual relationship among pixels. By using the fuzzy attentive feature generator and the multi-scale $L_1$ loss calculated on uncertainty maps, the discriminating ability of the discriminative network is enhanced and can better guide the generative network to generate more accurate segmentation results. The proposed MSF-GAN outperforms six state-of-the-art deep neural network-based methods in terms of TPR, FPR, IoU, DSC, and AER on three BUS datasets.

## REFERENCES


[1] Q. Huang, X. Huang, L. Liu, Y. Lin, X. Long, and X. Li, "A case-oriented web-based training system for breast cancer diagnosis," *Computer methods and programs in biomedicine*, vol. 156, pp. 73–83, 2018.

[2] A. E. Ilesanmi, U. Chaumrattanakul, and S. S. Makhanov, "Methods for the segmentation and classification of breast ultrasound images: a review," *Journal of Ultrasound*, pp. 1–16, 2021.

[3] M. H. Yap, G. Pons, J. Marti, S. Ganau, M. Sentis, R. Zwiggelaar, A. K. Davison, and R. Marti, "Automated breast ultrasound lesions detection using convolutional neural networks," *IEEE journal of biomedical and health informatics*, vol. 22, no. 4, pp. 1218–1226, 2017.

[4] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[6] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[7] B. Shareef, M. Xian, and A. Vakanski, "Stan: Small tumor-aware network for breast ultrasound image segmentation," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2020, pp. 1–5.

[8] B. Lei, S. Huang, R. Li, C. Bian, H. Li, Y.-H. Chou, and J.-Z. Cheng, "Segmentation of breast anatomy for automated whole breast ultrasound images with boundary regularized convolutional encoder–decoder network," *Neurocomputing*, vol. 321, pp. 178–186, 2018.

[9] B. Lei, S. Huang, H. Li, R. Li, C. Bian, Y.-H. Chou, J. Qin, P. Zhou, X. Gong, and J.-Z. Cheng, "Self-co-attention neural network for anatomy segmentation in whole breast ultrasound," *Medical image analysis*, vol. 64, p. 101753, 2020.

[10] K. Huang, H.-D. Cheng, Y. Zhang, B. Zhang, P. Xing, and C. Ning, "Medical knowledge constrained semantic breast ultrasound image segmentation," in *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, 2018, pp. 1193–1198.

[11] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *arXiv preprint arXiv:1406.2661*, 2014.

[12] Y. Xue, T. Xu, H. Zhang, L. R. Long, and X. Huang, "Segan: Adversarial network with multi-scale $l_1$ loss for medical image segmentation," *Neuroinformatics*, vol. 16, no. 3, pp. 383–392, 2018.

[13] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data in brief*, vol. 28, p. 104863, 2020.

[14] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2881–2890.

[15] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.