

# Clinical Prior-Guided Tumor Generation for Breast Ultrasound with Cross Domain Adaptation

Haoyu Pan<sup>1,2</sup>, Junyang Mo<sup>1</sup>, Hongxin Lin<sup>3</sup>, Chu Zhang<sup>1</sup>, Zijian Wu<sup>1</sup>,  
Yi Wang<sup>1,2,✉</sup>, and Qingqing Zheng<sup>4,✉</sup>

<sup>1</sup> School of Biomedical Engineering, Shenzhen University Medical School, Shenzhen University, Shenzhen, China

<sup>2</sup> Smart Medical Imaging, Learning and Engineering (SMILE) Lab

<sup>3</sup> Nuclear Medicine Department, the Seventh Affiliated Hospital, Sun Yat-Sen University, Shenzhen, China

<sup>4</sup> Shenzhen University of Advanced Technology, Shenzhen, China  
onewang@szu.edu.cn, zhengqingqing@suat-sz.edu.cn

**Abstract.** Computer-aided diagnosis (CAD) has become an essential solution for breast ultrasound (BUS) image analysis; however, the development of CAD systems is hindered by high-quality data scarcity and annotation challenges. We propose a novel clinical prior-guided tumor generation method that allows precise control over tumor characteristics, such as size, shape, and texture, using clinical knowledge from textual descriptions and structural masks. Additionally, our method enables cross-domain data generation, enhancing the adaptability of the synthetic data across different imaging conditions. Experiments on three public BUS datasets demonstrate the favorable generation quality and effective cross-domain adaptation of our method. Moreover, the improved accuracy in downstream classification and segmentation tasks further show the clinical utility and practical effectiveness of our synthetic images in supporting breast cancer diagnosis. *The code is available at <https://github.com/Violetphy/Clinical-Prior-Tumor-Generation>.*

**Keywords:** Breast ultrasound · Tumor generation · Diffusion model · Cross domain adaptation · Computer-aided diagnosis.

## 1 Introduction

Breast cancer is a prevalent and severe disease, and early diagnosis is crucial for improving patient outcomes [5]. Breast ultrasound (BUS) has proven to be an essential non-invasive tool for screening and diagnosing breast cancer [9]. The performance of existing computer-aided diagnosis (CAD) systems for BUS is often limited by the scarcity of real annotated data due to time-consuming annotation processes and strict privacy regulations [17].

---

✉Corresponding authors: Yi Wang and Qingqing Zheng.

Deep learning-based generative models for BUS synthesis have emerged as promising solutions, as they can augment limited real annotated datasets and enhance training of various diagnostic models including tumor detection [26], segmentation [4], and classification [29]. However, the randomness inherent in unconditional generative models often leads to synthesized images that fail to accurately reflect clinically relevant characteristics [3]. Conditional generative models guided by text that contains category information of tumors (benign/malignant), while useful for classification tasks, still require post-processing annotations for segmentation tasks [7]. In addition, shape-guided conditional models, though may be utilized in different downstream tasks, typically rely on existing masks to generate more data, often producing results using the same mask repeatedly which leads to insufficient tumor diversity [6]. Therefore, these methods still rely on manual annotation or existing labeled data, limiting their full potential for autonomous data generation in medical applications. Additionally, US characteristics vary largely across datasets due to differences in equipment, scanning protocols, and patient demographics [16]. This variability creates a cross-domain gap, impeding the adaptability of generative models trained on a single dataset.

To overcome aforementioned limitations, we propose the very first clinical-prior guided tumor generation method for BUS, which is expected to enhance the performance of computer-aided breast cancer diagnosis through synthetic data. Our main contributions can be summarized as follows:

- **Clinical prior-guided tumor generation framework:** We introduce a diffusion-based controllable generative framework, which leverages clinical knowledge to constrain the generation of both structural and textural attributes, ensuring that the synthetic images accurately reflect clinically representative tumor characteristics.
- **Annotation-free in generating new data with clinical compliance:** In accordance with breast imaging reporting and data system (BI-RADS) [18], we automatically generate masks to represent the tumor shape, and design clinical text to describe echogenicity and boundary characteristics. This annotation-free method enables to generate new data with labels.
- **Cross-domain adaptation strategy:** We apply a low-rank adaptation (LoRA) fine-tuning strategy for efficient model adaptation across different datasets, ensuring that generated images possess domain-specific characteristics, which is crucial for real-world clinical applicability.
- **Enhancing the downstream tasks:** Extensive experiments on three public BUS datasets show the efficacy of the proposed tumor generation method, and also validate the applicability of our method in practical downstream classification and segmentation tasks.

## 2 Method

### 2.1 Framework Overview

We propose a clinical prior-guided tumor generation framework for BUS, illustrated in Fig. 1. The framework first trains a diffusion-based generative model

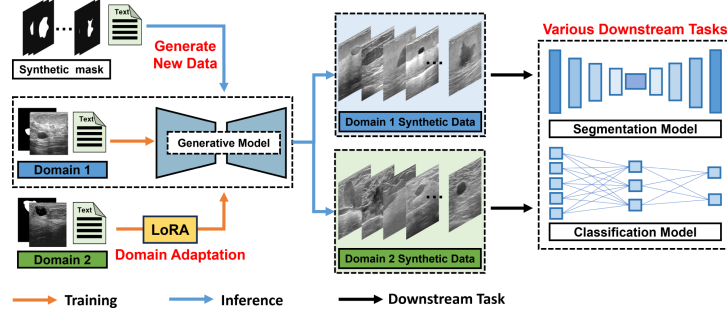


Fig. 1. Overview of the proposed tumor generation framework for breast ultrasound.

with real US images, masks, and clinical text to leverage clinical knowledge. Moreover, a LoRA-based fine-tuning strategy enables cross-domain adaptation with fewer parameters, allowing the model to generalize different clinical datasets. Finally, we utilize synthetic masks and clinical text to create diverse new data, benefiting for various downstream tasks.

**Joint Training Procedure** The training process for the generative model is shown in Fig. 2. The generative process is built on the stable diffusion framework [22] with a ControlNet [28] for structural guidance and a pretrained text encoder for textural refinement. Given an input image  $x_0$ , we encode it using the variational autoencoder (VAE) [14] to obtain latent representations characterized by mean  $\mu$  and log-variance  $\omega^2$ . The latent variable  $z_0$  is sampled via the reparameterization trick:

$$z_0 = \mu + \omega \odot \epsilon, \quad \epsilon \sim \mathcal{N}(0, I). \quad (1)$$

The latent variable  $z_0$  serves as the starting point for the forward diffusion process, where noise is iteratively added over  $T$  timesteps to produce noisy latent representations  $z_t$  with

$$z_t = \sqrt{\bar{\alpha}_t} z_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon, \quad (2)$$

where  $\bar{\alpha}_t$  is the cumulative product of noise schedule parameters. UNet [23] is employed as the denoising model  $\epsilon_\theta$ . It is trained to reverse the diffusion process by predicting and removing noise from  $z_t$  in each timestep. This denoising is conditioned on both text and structural inputs. Specifically, clinical descriptions are processed through a pretrained text encoder  $\tau_\phi$  to generate text embedding  $e$ ; and the mask  $m$  is processed by the encoder of ControlNet  $\xi_\theta$ , generating the structural features subsequently incorporated into the denoising networks as a residual feature. Thus, the denoising network is represented as  $\epsilon_\theta(z_t, t, e, \xi_\theta)$ .

**Loss Function** The training objective comprises two components, namely, mean squared error (MSE) between the predicted and true noise, ensuring effective denoising, and Kullback-Leibler (KL) divergence loss from the VAE encoding

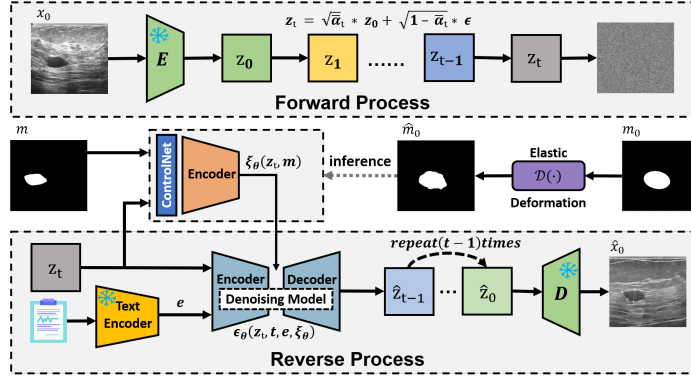


Fig. 2. Training process for the generative model.

to maintain a structured latent space:

$$L_{\text{MSE}} = \mathbb{E} [\|\epsilon - \epsilon_\theta(z_t, t, e, \xi_\theta)\|^2], \quad (3)$$

$$L_{\text{KL}} = -\frac{1}{2} \sum_{i=1}^d (1 + \log \omega_i^2 - \mu_i^2 - \omega_i^2), \quad (4)$$

where  $d$  is the dimensionality of the latent space. The total loss is:

$$L_{\text{total}} = L_{\text{MSE}} + \alpha L_{\text{KL}}, \quad (5)$$

where  $\alpha$  is a weighting factor.

## 2.2 Shape and Textual Constraints

Benign and malignant tumors display distinct characteristics in US images [9]. According to BI-RADS, benign tumors are typically regular, elliptical shapes that are wider than tall, with smooth edges and uniform echo patterns. Malignant tumors often appear taller than wide, with irregular shapes, complex internal structures, and diffuse boundaries due to tissue invasion and necrosis [24].

**Elastic Deformation Ellipse Model for Shape Constraint** To define the basic tumor shape, we model each tumor using an ellipse with semi-major axis  $a \in \mathbb{N}_+$  and semi-minor axis  $b \in \mathbb{N}_+$ , centered at  $(x'_0, y'_0)$ :

$$\left(\frac{x - x'_0}{a}\right)^2 + \left(\frac{y - y'_0}{b}\right)^2 = 1, \begin{cases} a \leq x'_0 \leq W - a, \\ b \leq y'_0 \leq H - b. \end{cases} \quad (6)$$

The center position  $(x'_0, y'_0)$  is constrained based on the image width  $W$  and height  $H$ . Benign tumors are modeled with  $a > b$  to reflect width greater than

height, while malignant tumors use  $a < b$  to represent height greater than width. The size of each tumor is varied by adjusting  $a$  and  $b$  within clinical ranges, allowing generation of small, medium, and large tumors. After creating the initial ellipse mask, we apply an elastic deformation [12] function  $\mathcal{D}(\cdot)$  defined as:

$$\hat{m}_0 = \mathcal{D}(m_0, \sigma), \quad (7)$$

where  $m_0$  is the initial binary ellipse mask, and  $\sigma$  controls the deformation level. The larger  $\sigma$  is, the greater the degree of elastic deformation is. Fig. 3(b) shows the generation of tumor masks for both benign and malignant tumors.

**Clinical Text Prompt for Textural Refinement** Textual descriptions are crafted based on clinical prior knowledge of benign and malignant breast tumors, which informs the generative model’s output by conditioning it to reflect appropriate echogenicity and boundary clarity. Based on BI-RADS, the echogenicity can be categorized mainly as homogeneous or heterogeneous, and the boundary as clear or unclear. Thus, we utilize a text template “*breast ultrasound image of **benign/malignant** tumor with **homogeneous/heterogeneous** echogenicity and **clear/unclear** boundary.*” to condition different tumors.

### 2.3 Cross-Domain Adaptation with LoRA

LoRA is applied to the UNet module. By introducing low-rank matrices  $A$  and  $B$  in the selected layers of the UNet, LoRA enables efficient adaptation while minimizing the risk of overfitting. The fine-tuning of the UNet parameters  $\theta \in \mathbb{R}^{d \times d}$  is represented as:

$$\theta' = \theta + AB, \quad (8)$$

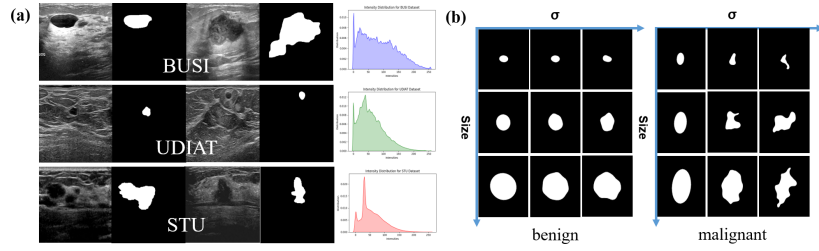
where  $A \in \mathbb{R}^{d \times r}$  and  $B \in \mathbb{R}^{r \times d}$  with  $r \ll d$ . During LoRA fine-tuning process, only the low-rank matrices  $A$  and  $B$  in the UNet are updated according to the loss function, while other parameters remain fixed.

## 3 Experiments and Results

### 3.1 Datasets and Experimental Settings

We evaluated our method on three public BUS datasets. **BUSI dataset** [1] (647 images: 437 benign/210 malignant) was randomly split into training (396), validation (54), and test (197) sets for generation and downstream tasks. **UDIAT dataset** [27] (163 images: 110 benign/53 malignant) was divided into training (130) and test (33) sets for cross-domain adaptation and downstream tasks. **STU dataset** [30] (42 mask-only images) was used exclusively as an external test set for segmentation evaluation. Fig. 3(a) shows BUS images, corresponding tumor masks and intensity distribution plots from different datasets.

During the training process, we employed a pre-trained stable diffusion (SD) v1-5 [22] for generative model and ControlNet [28] with the input image resized



**Fig. 3.** (a) BUS images, corresponding tumor masks and intensity distribution plots from different datasets. (b) Benign and malignant mask generation of our method.

**Table 1.** Generation performance comparison of different methods on BUSI dataset. Ours-*t* and Ours-*m* denote our method with only *text* or *mask* guidance.

Method	GAN	CGAN	DCGAN	DDPM	MT-DDPM	SGD	Ours- <i>t</i>	Ours- <i>m</i>	Ours
FID ↓	14.370	17.988	20.728	11.614	8.986	8.278	8.623	8.563	<b>6.858</b>
KID ↓	0.058	0.080	0.096	0.047	0.029	0.026	0.030	0.029	<b>0.022</b>

to  $512 \times 512$ . The weighting factor  $\alpha$  was set to 0.001. Using gradient accumulation with batch size of 64, the model was trained for 200 epochs with AdamW optimizer. The training process employed a warm-up strategy for the first 500 steps, gradually increasing the learning rate to initial value of  $1e-4$ , followed by a constant learning rate scheduler. Note that for fairness, we reimplemented all baselines with mask conditioning.

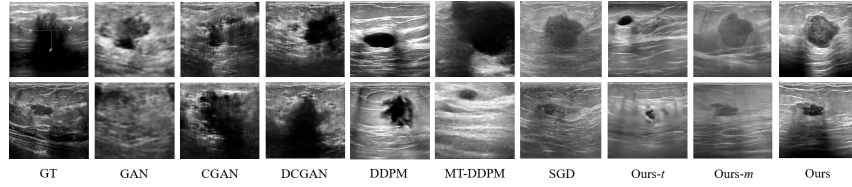
Fréchet Inception Distance (FID) [10] and Kernel Inception Distance (KID) [2] were used to evaluate generation performance. The Area Under the Receiver Operating Characteristic Curve (AUC) and the Dice Similarity Coefficient (DSC) were utilized to evaluate the classification and segmentation tasks, respectively.

### 3.2 Comparison on Tumor Generation

Table 1 lists the generation performance comparison of different methods. Fig. 4 visualizes the generated BUS images. It can be observed that GAN-based methods (GAN [8], CGAN [19], DCGAN [21]) suffer from anatomically implausible shapes and low image quality. While diffusion-based methods (DDPM [11], MT-DDPM [20], SGD [15]) achieve improved realistic background, but lack tumor specificity (e.g., indistinct echogenicity). Our method, which incorporates clinical prior guidance, outperforms these comparison methods. Specifically, by incorporating both structural masks and clinical text descriptions, our method offers precise control over tumor shape and texture, generating high-quality, clinically relevant images that closely match the ground truth.

### 3.3 Adaptable Cross-Domain Generation

To validate the efficacy of our LoRA-based adaptation, we compared it with full fine-tuning (859.52M trainable parameters) on the UDIAT dataset. As shown



**Fig. 4.** Visual comparison of BUS images generated by different methods.

**Table 2.** Impact of LoRA and different ranks of LoRA on UDIAT dataset.

Model	Rank	FID ↓	KID ↓	Parameters
fully fine-tune	-	8.088	0.024	859.52M
without LoRA	-	10.022	0.030	-
	2	7.921	0.022	0.40M
with LoRA	4	7.599	0.020	0.80M
	8	<b>7.316</b>	<b>0.019</b>	1.59M

in Table 2, LoRA with rank  $r = 8$  achieves comparable FID (7.316 *vs.* 8.088 for full fine-tuning) and KID (0.019 *vs.* 0.024) scores while updating only 1.59M parameters – 1.8% of the full fine-tuning parameter count. This demonstrates that LoRA effectively adapts domain-specific features with minimal computational overhead. Notably, the baseline model (trained on BUSI) without adaptation lacks of generalization (FID=10.022), highlighting the necessity of domain transfer mechanisms for clinical deployment across different imaging systems.

### 3.4 Data Augmentation for Downstream Tasks

We systematically evaluated the impact of synthetic data augmentation on classification and segmentation tasks using EfficientNet-B0 [25] and nnUNet [13].

Table 3 compares different real-to-synthetic data ratios in downstream tasks. The results indicate that incorporating synthetic images consistently improves the model’s performance compared to training on real images alone. Regardless of the classification or segmentation task, or whether training on BUSI or UDIAT dataset, the optimal performance (both in-domain and cross-domain) is observed within a real-to-synthetic ratio range of 1:1 to 1:4. This highlights the critical balance between real and synthetic data, as excessive synthetic proportions degrade performance due to synthetic-data-induced overfitting. It is suggested that by optimizing the ratio according to specific task and model architecture, we can maximize the advantages of synthetic data augmentation, ultimately contributing to more robust and generalizable processing.

We further compared ordinary augmentation method (integrating flip, brightness, rotation, and noise) and SGD [15] in downstream tasks, as shown in Table 4. Note that only our method and SGD [15] can simultaneously provide both classification and segmentation labels, while other methods [8,19,21,11,20] cannot

**Table 3.** Classification (AUC) & segmentation (DSC, %) performance on the test sets of BUSI, UDIAT, and STU with varying real-to-synthetic ratios.

		Classification		Segmentation		
Train	Real:Synthetic	BUSI	UDIAT	BUSI	UDIAT	STU
BUSI	Real only	0.799±0.004	0.639±0.005	74.1±0.6	81.1±0.5	88.8±0.3
	1:1	<b>0.838±0.003</b>	0.600±0.006	<b>82.5±0.7</b>	84.4±0.4	<b>89.1±0.3</b>
	1:2	0.825±0.005	0.572±0.004	77.0±0.8	<b>86.7±0.5</b>	88.6±0.1
	1:4	0.823±0.005	<b>0.733±0.008</b>	77.7±0.6	85.9±0.7	88.0±0.4
	1:8	0.827±0.004	0.601±0.006	76.4±0.6	86.0±0.5	87.6±0.3
	Synthetic only	0.764±0.004	0.539±0.007	64.2±0.6	69.2±0.5	85.4±0.2
UDIAT	Real only	0.446±0.004	0.711±0.011	50.5±0.9	88.6±0.5	89.2±0.3
	1:1	0.623±0.003	0.733±0.010	60.4±0.9	87.3±0.4	<b>90.0±0.2</b>
	1:2	0.548±0.002	<b>0.772±0.011</b>	69.0±0.9	<b>89.6±0.5</b>	90.0±0.5
	1:4	<b>0.649±0.007</b>	0.767±0.008	<b>69.5±1.1</b>	88.0±0.4	89.6±0.3
	1:8	0.630±0.003	0.739±0.011	69.3±0.9	88.1±0.5	89.3±0.3
	Synthetic only	0.580±0.005	0.733±0.011	61.6±1.0	78.3±0.6	87.5±0.4

**Table 4.** Comparison in classification (AUC) & segmentation (DSC, %) performance on the test sets of BUSI, UDIAT, and STU with different methods. \* indicates a significant difference  $p < 0.05$  compared to ours.

		Classification		Segmentation		
Train	Method	BUSI	UDIAT	BUSI	UDIAT	STU
BUSI	Ordinary Aug	0.825±0.004*	0.636±0.003*	76.9±0.5*	82.8±0.6*	88.9±0.2*
	SGD [15]	0.830±0.002*	0.724±0.006*	81.4±0.2*	84.9±0.3*	89.0±0.4
	Ours	<b>0.838±0.003</b>	<b>0.733±0.008</b>	<b>82.5±0.7</b>	<b>86.7±0.5</b>	<b>89.1±0.3</b>
UDIAT	Ordinary Aug	0.462±0.012*	0.724±0.003*	51.8±0.5*	87.9±0.3*	89.1±0.2*
	SGD [15]	0.633±0.006*	0.759±0.014*	68.3±0.7*	88.9±0.5*	89.8±0.3
	Ours	<b>0.649±0.007</b>	<b>0.772±0.011</b>	<b>69.5±1.1</b>	<b>89.6±0.5</b>	<b>90.0±0.2</b>

generate images with segmentation labels. The comparison results demonstrate that our method outperforms ordinary augmentation and SGD [15].

## 4 Conclusion

The data scarcity in the medical domain, especially the lack of real patients' images with tumors, hinders the construction of efficient and accurate diagnostic models. To address this, we propose a novel clinical prior-guided tumor generation method for BUS. By leveraging both textual descriptions and structural masks, our approach enables precise control over tumor characteristics. Additionally, we address the challenge of cross-domain image generation by incorporating a low-rank adaptation strategy, which allows the model to efficiently adapt across different imaging domains. Our method demonstrates its effectiveness through extensive experimentation on three BUS datasets, showing that it can generate high-quality, clinically relevant tumor images with realistic structural and textural features. Furthermore, we highlight the practical benefits of



our approach by demonstrating improvements in the performance of segmentation and classification models trained with synthetic images, thus validating its utility for downstream medical imaging tasks.

While our method prioritizes fidelity to BI-RADS standards and anatomical masks, biases may arise from incomplete modeling of low-level details or under-represented edge cases in training. Future work will explicitly compare latent representations of real vs. synthetic data to identify and mitigate such gaps. On the other hand, by embedding clinical priors, the generated tumors reflect diagnostically relevant features, but their clinical utility ultimately depends on expert validation. We will propose collaborations with clinicians to evaluate the diagnostic equivalence of synthetic data in downstream tasks in the future.

**Acknowledgments.** This work was supported in part by the National Natural Science Foundation of China under Grants 62471306 and 62206270, in part by the Shenzhen Medical Research Fund (D2402010), in part by the Guangdong-Hong Kong Joint Funding for Technology and Innovation (2023A0505010021), and in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2023A1515010644.

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## References

1. Al-Dhabyani, W., Gomaa, M., Khaled, H., Fahmy, A.: Dataset of breast ultrasound images. *Data in brief* **28**, 104863 (2020)
2. Bińkowski, M., Sutherland, D.J., Arbel, M., Gretton, A.: Demystifying MMD GANs. *arXiv preprint arXiv:1801.01401* (2018)
3. van Breugel, B., Liu, T., Oglic, D., van der Schaar, M.: Synthetic data in biomedicine via generative artificial intelligence. *Nature Reviews Bioengineering* **2**(12), 991–1004 (2024)
4. Chen, G., Li, L., Dai, Y., Zhang, J., Yap, M.H.: AAU-Net: An adaptive attention U-net for breast lesions segmentation in ultrasound images. *IEEE Transactions on Medical Imaging* **42**(5), 1289–1300 (2023)
5. Duggan, C., Trapani, D., Ilbawi, A.M., Fidarova, E., Laversanne, M., Curigliano, G., Bray, F., Anderson, B.O.: National health system characteristics, breast cancer stage at diagnosis, and breast cancer mortality: a population-based analysis. *The Lancet Oncology* **22**(11), 1632–1642 (2021)
6. Freiche, B., El-Khoury, A., Nasiri-Sarvi, A., Hosseini, M.S., Garcia, D., Basarab, A., Boily, M., Rivaz, H.: Ultrasound image generation using latent diffusion models. *arXiv preprint arXiv:2502.08580* (2025)
7. Fujioka, T., Mori, M., Kubota, K., Kikuchi, Y., Katsuta, L., Adachi, M., Oda, G., Nakagawa, T., Kitazume, Y., Tateishi, U.: Breast ultrasound image synthesis using deep convolutional generative adversarial networks. *Diagnostics* **9**(4), 176 (2019)
8. 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* **27** (2014)
9. Guo, R., Lu, G., Qin, B., Fei, B.: Ultrasound imaging technologies for breast cancer detection and management: a review. *Ultrasound in Medicine & Biology* **44**(1), 37–70 (2018)

10. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: GANs trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in Neural Information Processing Systems* **30** (2017)
11. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems* **33**, 6840–6851 (2020)
12. Hu, Q., Chen, Y., Xiao, J., Sun, S., Chen, J., Yuille, A.L., Zhou, Z.: Label-free liver tumor segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 7422–7432 (2023)
13. Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P.F., Kohl, S., Wasserthal, J., Koehler, G., Norajitra, T., Wirkert, S., et al.: nnU-net: Self-adapting framework for u-net-based medical image segmentation. *arXiv preprint arXiv:1809.10486* (2018)
14. Kingma, D.P.: Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013)
15. Konz, N., Chen, Y., Dong, H., Mazurowski, M.A.: Anatomically-controllable medical image generation with segmentation-guided diffusion models. In: *Medical Image Computing and Computer Assisted Intervention (MICCAI)*. pp. 88–98. Springer (2024)
16. Li, H., Wang, Y., Wan, R., Wang, S., Li, T.Q., Kot, A.: Domain generalization for medical imaging classification with linear-dependency regularization. *Advances in Neural Information Processing Systems* **33**, 3118–3129 (2020)
17. Maack, L., Holstein, L., Schlaefel, A.: GANs for generation of synthetic ultrasound images from small datasets. *Current Directions in Biomedical Engineering* **8**(1), 17–20 (2022)
18. Magny, S.J., Shikhman, R., Keppke, A.L.: Breast imaging reporting and data system. In: *StatPearls [Internet]*. StatPearls publishing (2023)
19. Mirza, M.: Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784* (2014)
20. Pan, S., Wang, T., Qiu, R.L., Axente, M., Chang, C.W., Peng, J., Patel, A.B., Shelton, J., Patel, S.A., Roper, J., et al.: 2D medical image synthesis using transformer-based denoising diffusion probabilistic model. *Physics in Medicine & Biology* **68**(10), 105004 (2023)
21. Radford, A.: Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434* (2015)
22. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 10684–10695 (2022)
23. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer Assisted Intervention (MICCAI)*. pp. 234–241. Springer (2015)
24. Shankar, P.M., Dumane, V., Reid, J.M., Genis, V., Forsberg, F., Piccoli, C.W., Goldberg, B.B.: Classification of ultrasonic b-mode images of breast masses using nakagami distribution. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* **48**(2), 569–580 (2001)
25. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: *International Conference on Machine Learning*. pp. 6105–6114 (2019)
26. Wang, Y., Wang, N., Xu, M., Yu, J., Qin, C., Luo, X., Yang, X., Wang, T., Li, A., Ni, D.: Deeply-supervised networks with threshold loss for cancer detection in automated breast ultrasound. *IEEE Transactions on Medical Imaging* **39**(4), 866–876 (2020)

27. Yap, M.H., Pons, G., Marti, J., Ganau, S., Sentis, M., Zwiggelaar, R., Davison, A.K., Marti, R.: Automated breast ultrasound lesions detection using convolutional neural networks. *IEEE Journal of Biomedical and Health Informatics* **22**(4), 1218–1226 (2017)
28. Zhang, L., Rao, A., Agrawala, M.: Adding conditional control to text-to-image diffusion models. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. pp. 3836–3847 (2023)
29. Zhou, Y., Chen, H., Li, Y., Liu, Q., Xu, X., Wang, S., Yap, P.T., Shen, D.: Multi-task learning for segmentation and classification of tumors in 3D automated breast ultrasound images. *Medical Image Analysis* **70**, 101918 (2021)
30. Zhuang, Z., Li, N., Joseph Raj, A.N., Mahesh, V.G., Qiu, S.: An RDAU-NET model for lesion segmentation in breast ultrasound images. *PloS one* **14**(8), e0221535 (2019)