Improved Tumor Segmentation using Selective Synthetic Augmentation for Enhanced Surgical Planning in Breast MRI

Miguel Luna[1], John Baek[1], Won Hwa Kim[1,2], Wan Gyu Son[2], Kwang Min Lee[2], Hye Jung Kim[2], and Jaeil Kim[1,3]

[1] BeamWorks Inc., Daegu, Republic of Korea
[2] Department of Radiology, School of Medicine, Kyungpook National University, Kyungpook National University Chilgok Hospital, Daegu, Republic of Korea
[3] School of Computer Science and Engineering, Kyungpook National University, Daegu, Republic of Korea
miguel@beamworks.co.kr

Abstract. Breast-conserving surgery (BCS) is the preferred treatment for early-stage breast cancer, offering survival rates comparable to mastectomy while preserving breast aesthetics. Accurate tumor segmentation is essential for surgical planning, yet segmentation models often exhibit biases toward specific tumor sizes, particularly underperforming on smaller tumors. To address this, we propose a novel approach that uses generative models to improve segmentation across tumor sizes. Specifically, we adapt the Stable Diffusion model and apply a Denoising Diffusion Probabilistic Model (DDPM) inversion approach to generate synthetic tumors of controlled sizes within real breast MRIs, helping to balance tumor size distribution in the training data. By augmenting the dataset with 10–20% synthetic tumor images, our method significantly improves segmentation accuracy for small tumors without compromising performance for larger tumors. This enhancement allows for more precise tumor assessment, leading to better-informed surgical decisions and potentially reducing unnecessary mastectomies.

Keywords: Breast Cancer Segmentation · Stable Diffusion · DDPM Inversion · Synthetic Data Augmentation in MRI.

1 Introduction

Breast cancer is the second most diagnosed cancer worldwide [23]. A key decision in its management is choosing between breast-conserving surgery (BCS) and mastectomy. For decades, radical mastectomy was the standard treatment [22], but randomized controlled trials (RCTs) in the 1980s demonstrated that BCS combined with radiation therapy (RT) provides equivalent survival rates and local control, leading to a paradigm shift [5]. Today, BCS plus RT is used in approximately 70% of early-stage cases [5].

A primary advantage of BCS lies in preserving the shape of the breast, improving psychological well-being [5]. BCS is preferred for younger patients with

small, unifocal tumors in favorable locations [7]. Successful BCS requires achieving clear margins at a microscopic level minimizing the risk of loco-regional disease, while preserving healthy breast tissue [3]. Obtaining clear pathological margins (typically $>$ 1mm) free of cancerous cells surrounding the tumor is essential [2]. The presence of compromised margins is associated with poorer oncological outcomes, which makes precise identification of the tumor boundary a key factor during surgical planing [2].

Magnetic Resonance Imaging (MRI) plays a vital role in surgical planning by accurately delineating tumor extent, local spread to the skin and nipple, and lymph node involvement [23]. However, its high sensitivity can lead to overestimation, potentially resulting in excessive tissue excision and higher mastectomy rates [26]. While automated segmentation models can provide tumor measurements, they are often biased toward specific tumor sizes [9], under-performing on the smaller lesions that are critical for surgical eligibility.

To address this challenge, we leverage generative models, which are effective at augmenting datasets to enhance model performance [19], particularly in long-tailed scenarios by addressing challenges posed by underrepresented classes [24, 6]. In medical imaging, they offer a solution to data scarcity and annotation challenges [13]. Acknowledging that generating entire synthetic images can introduce unwanted biases [1], our framework instead inserts synthetic tumors into real, healthy MRI slices. We achieve this by adapting a pre-trained Stable Diffusion model [20] and using Denoising Diffusion Probabilistic Models (DDPMs) inversion [12] to generate tumors of controlled sizes. This targeted augmentation strategy is used to train a SwinUNETR segmentation model [8], enhancing its ability to accurately segment tumors across the size spectrum while preserving anatomical realism.

Our key contributions are as follows:

- We introduce a novel framework leveraging Stable Diffusion and DDPM inversion for controlled synthetic tumor generation in real breast MRIs.
- We address tumor size imbalance by generating synthetic tumors of varying sizes, improving segmentation robustness across the size spectrum, especially for small tumors due to their relevance in surgical planning.
- Our robust segmentation model enables more accurate tumor size assessment, potentially improving surgical planning and reducing unnecessary mastectomies.

2 Methodology

This section outlines our method for improving breast MRI segmentation with synthetic tumor augmentation. Our pipeline includes three key components: Stable Diffusion for tumor generation, DDPM inversion for controlled insertion, and SwinUNETR for segmentation.

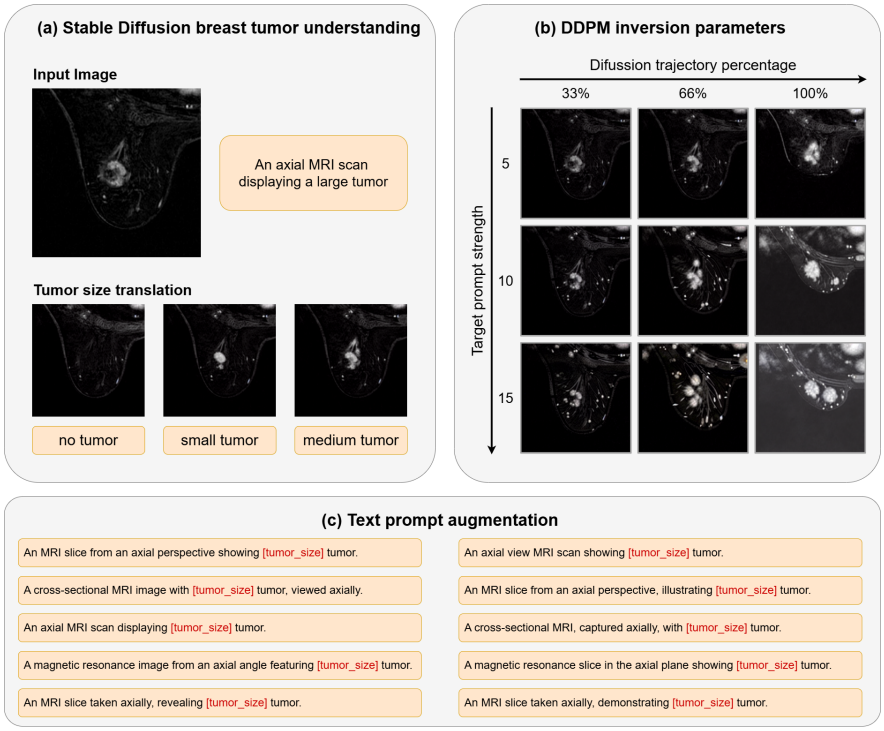

Fig. 1: Overview of synthetic tumor generation and augmentation strategies. (a) Stable Diffusion ability to modify tumor presence in MRI images. (b) Effects of inversion trajectory and text prompt strength, showing increased sensitivity to long trajectories and loss of realism with strongly weighted prompts. (c) A diverse set of text prompts used to describe MRI images, enhancing variability in synthetic data generation.

2.1 Stable Diffusion adaptation for breast MRI

Stable Diffusion [20] is a latent diffusion model (LDM) that generates high-quality images by iteratively denoising a Gaussian noise input through a U-Net-based architecture. By working in the latent space of a Variational Autoencoder (VAE) [14] instead of pixel space, the model significantly reduces computational cost and improves image quality. The generation process is guided by text prompts via a CLIP (Contrastive Language-Image Pretraining) text encoder [18]. However, while Stable Diffusion is effective for natural images, struggles with breast MRI due to its absence in the training data, requiring domain adaptation.

Fine-tuning strategy Stable Diffusion operates on 2D image space and utilizes descriptive text prompts to guide image generation. In our adaptation dataset,

each image was resized to 512×512 pixels, and a descriptive prompt specifying "2D axial MRI view" (as illustrated in Fig. 1c) was assigned to each image. The tumor size is determined according to the pixel area of the segmentation mask within each 2D slice. Specifically, we categorized tumors as: tiny (less than 200 pixels), small (200-399 pixels), medium (400-699 pixels), large (700-1599 pixels), and very large (1600 pixels or more). For images of healthy breast, the prompt explicitly stated "no tumor". This size-controlled prompting strategy effectively guides the model to generate tumors of the desired dimensions, as demonstrated in Fig. 1a.

To efficiently fine-tune the pre-trained Stable Diffusion model, we employ Low-Rank Adaptation (LoRA) [11]. LoRA enables efficient fine-tuning for medical imaging by training only a small set of parameters instead of the existing model weights, reducing computational cost. Specifically, we only fine-tune the diffusion model in latent space, while freezing the pre-trained VAE encoder/decoder and CLIP text encoder. This strategy utilizes the pre-trained model broad understanding of general image features while refining the diffusion process specifically for the breast MRI domain.

Addressing censorship issues A significant challenge we encountered was the built-in safety mechanisms of generative models, which flag terms like "breast" as NSFW (Not Safe For Work), preventing their use in text prompts. To circumvent this, we rephrased our prompts using generic terminology like 'MRI image' and avoided explicit anatomical references, as illustrated in Fig. 1c. We also cropped images to single-breast views to avoid triggering automated content filtering on both breasts. These adaptations allowed us to successfully fine-tune Stable Diffusion for breast MRI synthetic tumor generation without triggering the censorship mechanisms.

2.2 DDPM inversion for tumor insertion

Denoising Diffusion Probabilistic Models (DDPMs) [10] generate images through an iterative denoising process. The forward diffusion process gradually adds Gaussian noise to an image x_0 over T time steps, resulting in a noisy image x_T. This process can be defined as:

$$x_t = \sqrt{\alpha_t} x_{t-1} + \sqrt{1-\alpha_t}\epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, I), \tag{1}$$

where α_t controls the noise schedule. The reverse diffusion process then iteratively denoises x_T using a learned model to approximate the original image x_0.

While DDIM (Denoising Diffusion Implicit Models)[25] enables faster sampling, its strict one-to-one mapping between image and noise limits flexibility for localized editing. Following Huberman et al.[12], we use DDPM inversion to extract noise vectors that can reconstruct the original image when reversed. This stochastic process allows for diverse synthetic tumor insertions into real

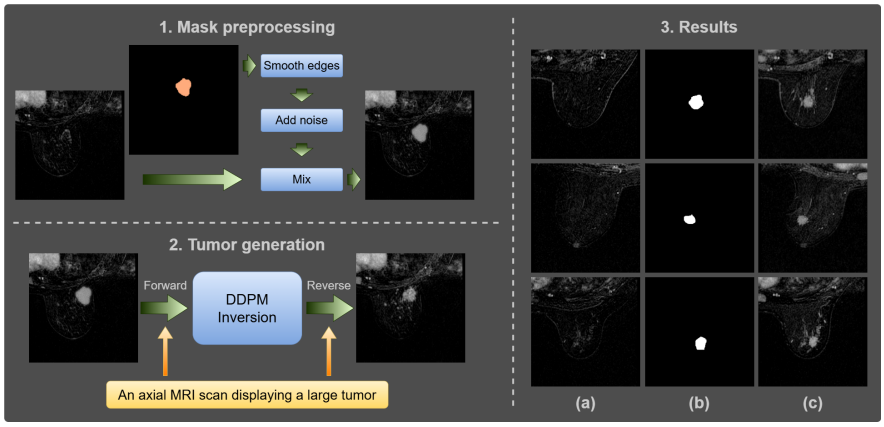

Fig. 2: Tumor insertion into healthy breast MRIs using DDPM inversion. (1) Tumor mask insertion into a healthy breast image as a noisy mask, (2) Artificial mask transformation to realistic tumor via DDPM inversion, and (3) Example results: (a) real healthy breast MRI, (b) tumor mask, (c) generated image with a synthetic tumor.

MRI slices, even with the same mask. Empirically, we found that a reverse diffusion trajectory between 75% and 85% balances realism and editability. Shorter trajectories (e.g., 50%) limit editing flexibility, while longer ones (approaching 100%) risk mode collapse, losing key image features. Additionally, text prompt strength has a notable impact on realism. Based on expert feedback, we found that values of 5 or lower preserve anatomical integrity, while higher values often result in unrealistic anatomical and visual artifacts (Fig. 1b).

To expand the segmentation dataset, we apply DDPM inversion to insert tumors into healthy breast MRI images (Fig. 2). Specifically, we smooth mask boundaries using a Gaussian kernel to ensure seamless tumor-tissue transitions, and replace tumor mask pixels with Gaussian noise (mean = 0.3, std = 0.3) for normalized images (0–1 range). This method leverages the inversion process to transform artificial masks into realistic-looking tumors, guided by text prompts indicating expected tumor sizes.

2.3 SwinUNETR for tumor segmentation

SwinUNETR [8] combines the strengths of U-Net architectures [21] and Swin Transformers [15]. Its hierarchical Swin Transformer backbone with shifted window attention efficiently captures multi-scale features, which are essential for accurate medical image segmentation. The U-Net skip connections allow for the fusion of these multi-scale features, combining local and global context. We utilize SwinUNETR for tumor segmentation due to its strong performance in medical imaging, ability to model long-range dependencies, and efficient feature extraction.

2.4 Segmentation model training with synthetic data

We train a SwinUNETR-based segmentation model using a supervised approach on a manually annotated breast MRI dataset. To address the previously identified performance gap for smaller tumors, we generate synthetic tumor-augmented images using our fine-tuned Stable Diffusion model via DDPM inversion as described in Fig. 2. Following, we retrain the segmentation model in two stages. First, we train on a mix of real and synthetic images to improve feature representation. Then, we fine-tune using only real images to enhance generalization to clinical cases. Both stages employ Dice loss [16] to ensure robust segmentation across all tumor sizes. This two-stage strategy balances synthetic data diversity with real-world accuracy.

3 Experiments

3.1 Dataset

This retrospective study used T1-subtraction MR images from 130 breast cancer patients who underwent surgery at Kyungpook National University Chilgok Hospital between January 2015 and December 2018. The study was approved by the Institutional Review Board (approval number, KNUCH 2022-03-006-002). Clinical tumor size, location, and other tumor characteristics were obtained from radiology reports. Two radiology technicians manually annotated indexed breast lesions on 2D axial slices, guided by these reports. To ensure quality, two radiologists with 10 and 20 years of experience independently reviewed and validated all segmentations.

Annotations focused exclusively on the indexed tumor identified in the radiology report, ensuring segmentation performance assessment was limited to surgically treated lesions. Other untreated masses that may have had varying degrees of diagnostic uncertainty or clinical significance were ignored to reduce ambiguity in our evaluation.

For experiments, the dataset was randomly split into training (50 samples), validation (15 samples), and testing (65 samples) sets.

3.2 Implementation details

Our experiments were conducted using PyTorch 2.2, integrating models from various open-source frameworks. For synthetic tumor generation, we used Stable Diffusion v1.4 from the Hugging Face diffusers library [17]. Fine-tuning was performed with Low-Rank Adaptation (LoRA) [11], using a rank of 4, a learning rate of 10^{-4}, and a batch size of 1 for 60000 steps on a TITAN RTX GPU with 24GB of memory. The DDPM inversion process followed the implementation of Huberman et al.[12]. For tumor segmentation, we employed SwinUNETR, implemented in the MONAI framework [4], training the model on 2D slices of size 512×512 with a batch size of 16.

Table 1: Dice Scores (%) by Tumor Size and Synthetic Data Percentage.

Synthetic Data (%)	$\leq$ 2 cm		$>$ 2 cm, $\leq$ 5 cm		$>$ 5 cm		All Sizes	
	Dice	95% CI	Dice	95% CI	Dice	95% CI	Dice	95% CI
0	62.6	[49.6, 74.3]	68.7	[56.5, 79.7]	75.3	[67.4, 81.3]	67.5	[60.0, 74.5]
2	68.3	[57.1, 78.2]	70.4	[59.4, 80.0]	72.7	[66.5, 78.7]	70.0	[63.4, 75.9]
5	63.3	[50.4, 74.9]	69.6	[58.0, 79.8]	73.3	[66.5, 79.5]	67.8	[60.5, 74.3]
10	**72.8**	[63.4, 80.9]	**72.5**	[61.7, 81.8]	**75.8**	[70.2, 80.9]	**73.2**	[67.1, 78.6]
20	**73.9**	[65.1, 81.5]	**73.1**	[61.9, 82.5]	**73.1**	[67.4, 78.4]	**73.4**	[67.4, 78.7]
30	66.9	[54.3, 78.2]	68.3	[55.7, 79.6]	74.9	[67.6, 81.2]	68.9	[61.6, 75.7]
40	69.9	[57.8, 80.3]	71.1	[59.4, 81.2]	73.1	[66.1, 79.2]	71.0	[64.0, 77.3]
50	67.5	[55.5, 78.1]	72.6	[62.1, 81.8]	71.4	[64.1, 78.2]	70.3	[63.5, 76.4]
60	65.7	[53.1, 77.1]	69.5	[57.7, 80.0]	76.1	[69.5, 81.5]	69.2	[62.0, 75.8]

3.3 Evaluation protocol

We assessed segmentation performance by computing the Dice score between the predicted 3D tumor mask and the corresponding ground truth annotation. To obtain a 3D assessment, we stacked all predicted and annotated slices into volumetric representations before evaluation. To quantify uncertainty, we calculated 95% confidence intervals (CI) using non-parametric bootstrap clustering analysis.

Given the importance of tumor size in surgical planning, we further analyzed segmentation performance based on tumor size. Tumors were categorized into three groups according to their largest cross-sectional diameter in 3D space, as recorded in the radiology report: smaller than 2 cm, between 2 cm and 5 cm, and larger than 5 cm.

3.4 Results

Table 1 illustrates the impact of synthetic data augmentation on Dice scores for tumor segmentation, categorized by tumor size. Without synthetic data (0%), the model achieved its highest Dice score for tumors larger than 5 cm (75.3; 95%CI: [67.4, 81.3]), while its performance dropped significantly for tumors smaller than 2 cm (62.6; 95%CI: [49.6, 74.3]). This baseline result confirms the model difficulty with smaller tumors, which are particularly relevant for breast-conserving surgery (BCS).

The introduction of synthetic data had varying effects depending on tumor size. For smaller tumors (less than 2 cm), we observed a marked improvement in Dice scores with the addition of synthetic data. The most significant gains occurred at 10% and 20% synthetic data, where Dice scores reached 72.8 (95%CI: [63.4, 80.9]) and 73.9 (95%CI: [65.1, 81.5]), respectively. This suggests that synthetic augmentation is particularly beneficial for improving the segmentation of smaller, more challenging tumors. In contrast, for medium-sized (2–5 cm) and

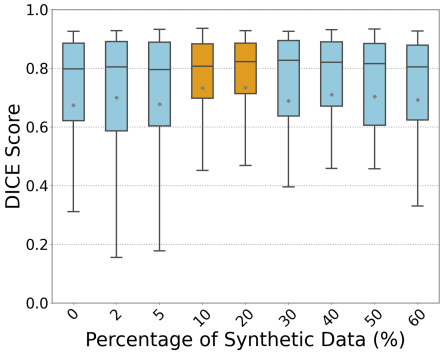

Fig. 3: Distribution of Dice scores with 10% and 20% synthetic data, showing improved segmentation robustness.

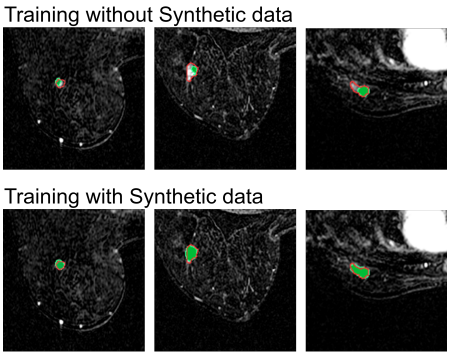

Fig. 4: Segmentation examples of tumors under 2 cm. Red: ground truth; Green: prediction.

larger tumors (>5 cm), Dice scores remained relatively stable across different levels of synthetic data.

Fig.3 shows the Dice score distributions across all tumor sizes for varying amounts of synthetic data. Smaller variations in scores indicate greater model robustness. In each box plot, the central line represents the median, while the mean is indicated by a dot within the box. Models trained with 10% and 20% synthetic data exhibited the least variation, suggesting that this range may be optimal. As shown in Fig.4, segmentation of small tumors improved. Importantly, performance on other tumor sizes was not adversely affected. However, increasing the proportion of synthetic data beyond 20% did not yield further improvements and instead led to a decline in performance. This may be due to the model overfitting to synthetic data patterns, reducing accuracy on real small tumors.

3.5 Limitations

Our study has several limitations. First, the experiments were conducted on a small dataset from a single institution, which may limit generalizability. Tumor descriptions were simplified to 2D visible size, omitting details like tumor type (mass vs. non-mass), focality (unifocal, multifocal, or multicentric), cancer subtype, and distance to the skin or nipple, all important for surgical assessment. Patient-specific factors such as breast size, BMI, tissue density, and other conditions were also not considered. Lastly, tumor size in text prompts was based solely on segmentation masks, but alternative estimation methods are needed for better scalability.

4 Conclusion

This study demonstrated that augmenting breast cancer MRI datasets with 10–20% synthetic images—generated using Stable Diffusion and DDPM inver-

sion—significantly improves the segmentation of small breast tumors, without compromising performance on larger tumors. These improvements may lead to more accurate tumor measurements, supporting surgeons in making better-informed decisions between BCS and mastectomy, while preserving oncological safety. Future work should focus on validating these findings in larger, multi-institutional datasets, incorporating detailed tumor characteristics relevant to surgical planning, accounting for patient-specific anatomical factors, and developing scalable methods for tumor size estimation beyond segmentation masks.

Acknowledgments. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT)(2022R1A2C2009415), the Korea Medical Device Development Fund grant funded by the Korea government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of Health & Welfare, the Ministry of Food and Drug Safety) (Project Number: 1711197554, RS-2023-00227526), the 'Digital Healthcare Medical Device Verification Support Project' of the Ministry of Health and Welfare, Daegu Metropolitan City, and the Korea Health Industry Development Institute (Project Number: B0080716000969), and Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT)(NO. RS-2023-00223446, Development of object-oriented synthetic data generation and evaluation methods).

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

References

1. Babul, K.A.R., Sathish, R., Pattanaik, M.: Synthetic simplicity: Unveiling bias in medical data augmentation. In: Data Engineering in Medical Imaging: Second MICCAI Workshop, DEMI 2024, Held in Conjunction with MICCAI 2024, Marrakesh, Morocco, October 10, 2024, Proceedings. vol. 15265, p. 64. Springer Nature (2025)
2. Bundred, J.R., Michael, S., Stuart, B., Cutress, R.I., Beckmann, K., Holleczek, B., Dahlstrom, J.E., Gath, J., Dodwell, D., Bundred, N.J.: Margin status and survival outcomes after breast cancer conservation surgery: prospectively registered systematic review and meta-analysis. bmj **378** (2022)
3. Cardoso, F., Kyriakides, S., Ohno, S., Penault-Llorca, F., Poortmans, P., Rubio, I., Zackrisson, S., Senkus, E.: Early breast cancer: Esmo clinical practice guidelines for diagnosis, treatment and follow-up. Annals of oncology **30**(8), 1194–1220 (2019)
4. Cardoso, M.J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., Murrey, B., Myronenko, A., Zhao, C., Yang, D., et al.: Monai: An open-source framework for deep learning in healthcare. arXiv preprint arXiv:2211.02701 (2022)
5. Christiansen, P., Mele, M., Bodilsen, A., Rocco, N., Zachariae, R.: Breast-conserving surgery or mastectomy?: impact on survival. Annals of Surgery Open **3**(4), e205 (2022)
6. Elberg, R., Parra, D., Petrache, M.: Long tail image generation through feature space augmentation and iterated learning. arXiv preprint arXiv:2405.01705 (2024)
7. Fajdic, J., Djurovic, D., Gotovac, N., Hrgovic, Z.: Criteria and procedures for breast conserving surgery. Acta Informatica Medica **21**(1), 16 (2013)

8. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In: International MICCAI brainlesion workshop. pp. 272–284. Springer (2021)
9. Hiraman, A., Viriri, S., Gwetu, M.: Lung tumor segmentation: a review of the state of the art. Frontiers in Computer Science **6**, 1423693 (2024)
10. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems (2020)
11. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al.: Lora: Low-rank adaptation of large language models. ICLR **1**(2), 3 (2022)
12. Huberman-Spiegelglas, I., Kulikov, V., Michaeli, T.: An edit friendly ddpm noise space: Inversion and manipulations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12469–12478 (2024)
13. Jung, H.K., Kim, K., Park, J.E., Kim, N.: Image-based generative artificial intelligence in radiology: comprehensive updates. Korean Journal of Radiology **25**(11), 959 (2024)
14. Kingma, D.P., Welling, M., et al.: Auto-encoding variational bayes (2013)
15. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 10012–10022 (2021)
16. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV). pp. 565–571. Ieee (2016)
17. von Platen, P., Patil, S., Lozhkov, A., Cuenca, P., Lambert, N., Rasul, K., Davaadorj, M., Nair, D., Paul, S., Berman, W., Xu, Y., Liu, S., Wolf, T.: Diffusers: State-of-the-art diffusion models. https://github.com/huggingface/diffusers (2022)
18. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: International conference on machine learning. pp. 8748–8763. PmLR (2021)
19. Rahat, F., Hossain, M.S., Ahmed, M.R., Jha, S.K., Ewetz, R.: Data augmentation for image classification using generative ai. arXiv preprint arXiv:2409.00547 (2024)
20. 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. pp. 10684–10695 (2022)
21. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. pp. 234–241. Springer (2015)
22. Sakorafas, G.H.: The origins of radical mastectomy. AORN journal **88**(4), 605–608 (2008)
23. Sharma, S., Vicenty-Latorre, F.G., Elsherif, S., Sharma, S.: Role of mri in breast cancer staging: A case-based review. Cureus **13**(12) (2021)
24. Shin, J., Kang, M., Park, J.: Fill-up: Balancing long-tailed data with generative models. arXiv preprint arXiv:2306.07200 (2023)
25. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. International Conference on Learning Representations (2020)
26. Thompson, J.L., Wright, G.P.: The role of breast mri in newly diagnosed breast cancer: an evidence-based review. The American Journal of Surgery **221**(3), 525–528 (2021)