Preoperative Simulation of Personalized Breast Reconstruction via Implant Prior Controllable Diffusion

Yizhi Wang[1,2,†], Tianze Yu[3,†], Hengjie Yu[2], Maoling Liu[5], Wenxuan Pan[2], Zhongxiang Ding[6], Peifen Fu[4,*], Yaochu Jin[2,*]

[1]Zhejiang University, Hangzhou 310003, China.
[2]School of Engineering, Westlake University, Hangzhou, Zhejiang 310030, China.
[3]School of Medicine, Zhejiang University, Hangzhou 310003, China.
[4]Department of Breast Surgery, The First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou 310003, China.
[5]Hangzhou Medical College, Hangzhou, Zhejiang 310013, China
[6]Department of Radiology, Affiliated Hangzhou First People's Hospital, Westlake University School of Medicine, Hangzhou, 310006, China.

fupeifen@zju.edu.cn, jinyaochu@westlake.edu.cn

Abstract. Breast reconstruction plays a crucial role in restoring body image and psychosocial well-being after mastectomy, yet preoperative planning relies heavily on surgeons' experience and lacks practical planning tools. In this work, we present a controllable generative framework to simulate postoperative breast reconstruction outcomes from preoperative chest CT. Our pipeline comprises three components: (1) a conditional 3D latent diffusion model with an inpainting scheme that enables high-fidelity synthesis within the surgical region; (2) an implant canonical template library and anatomy-aware alignment method to construct a physiologically plausible implant spatial prior for guiding the generation; (3) a dedicated control branch that injects the implant prior into the diffusion process to enforce the implant and breast geometry changes. Experiments show that our framework can generate realistic postoperative volumes, and spatial prior injection substantially improves the geometry accuracy, leading to superior performance over competing baselines in both fidelity and controllability. We further demonstrate that simulated outcomes vary coherently with the user-defined implant configurations. Overall, our work enables preoperative visualization under diverse surgical configurations and supports decision-making with a flexible, patient-specific planning tool for breast reconstruction. The source code is available at https://github.com/wangyz0915/DiffuBreast.

Keywords: Breast Reconstruction, Simulation and Visualization, Latent Diffusion Models, Controllable Generation.

1 Introduction

According to the global cancer statistics, breast cancer ranked as the second most frequently diagnosed cancer and the fourth leading cause of cancer-related mortality [1].

* Corresponding authors.
† Contributed equally to this work.

For many patients, mastectomy remains the primary surgical option [2]. To improve women's quality of life and reduce the psychological distress from mastectomy, an increasing number of breast reconstructions have been performed worldwide [3]. To restore the breast appearance disrupted by mastectomy, breast reconstruction is commonly performed to recreate a natural breast mound and match the contralateral breast [4]. In clinical practice, selection and placement of implants often rely on preoperative coarse measurement and intraoperative repeated adjustments, an experience-driven process that is procedure-intensive, lacks standardized criteria, and may prolong operative time and increase complication risk [5, 6]. Moreover, appropriate implant sizing and accurate placement are critical for achieving optimal postoperative breast morphology. However, constrained by the manufacturers' discrete, predefined shapes and volumes, available implants may not fit patient’s breast geometry. Consequently, postoperative asymmetry and suboptimal aesthetic outcomes remain common, highlighting the need for accurate preoperative simulation and decision support to enable patient-specific planning and expectation management.

Recent advances in AI methods have driven broader clinical applications, with growing interest in AI-assisted surgical planning [7, 8] and operative outcome simulation [9, 10]. However, methods for breast reconstruction remain limited. Existing studies provide cosmetic assessment of breast reconstruction outcomes using postoperative photographs [11–13] or CT scans [14], with limited utility for preoperative planning. Specifically, Chartier et al. proposed BreastGAN [15], and Zeng et al. trained a Stable Diffusion model to simulate breast augmentation outcomes from photographs [16], where the model is not readily shareable due to privacy concerns. More importantly, they generate outcomes without modeling key surgical variables (e.g., clinical attributes and implant parameters). In parallel, existing commercial platforms provide 3D body rendering of breast augmentation results for manufacturer-defined implant models, yet they prioritize visual plausibility over anatomical and morphological fidelity. Notably, unlike the bilateral enlargement, reconstruction aims to restore the mastectomy side to a natural appearance and achieve symmetry with the contralateral breast, making augmentation-oriented simulation methods less applicable to reconstruction. These gaps motivate the development of open, patient-specific tools that can accurately simulate reconstruction outcomes across diverse, user-defined surgical configurations, going beyond qualitative visualization toward reliable preoperative planning.

2 Methodology

2.1 Problem formulation

Given a patient’s preoperative chest CT volume $\mathbf{x}^{pre}$ and a binary mask $\mathbf{m}$ delineating the surgical breast region of interest (ROI), our goal is to synthesize the postoperative volume $\hat{\mathbf{x}}^{post}$ based on the clinical attributes $\mathbf{c}_c$ and implant configuration $\mathbf{c}_i$. The synthesized result should (i) be consistent with the anatomy outside the ROI, and (ii) realistically capture implant-induced changes in breast appearance within the ROI, approximating the true postoperative volume $\mathbf{x}^{post}$.

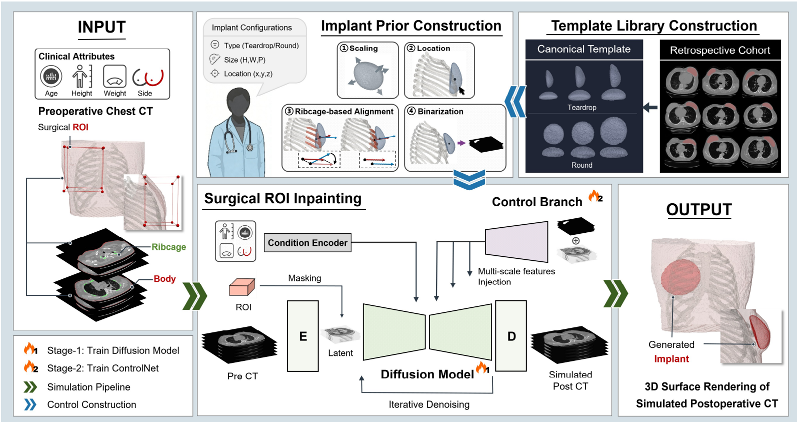

Fig. 1. Overview of the proposed breast reconstruction simulation framework. Inputs are the preoperative chest CT, ROI mask, and clinical attributes $\mathbf{c}_c$. Given an intended implant configuration, the corresponding implant template is retrieved from the library, then 1) scaled, 2) placed, and 3) reoriented to 4) form the implant mask. A diffusion inpainting model iteratively denoises in the latent space and generates the outcome within the ROI. The implant prior mask is injected via control branch, whose multi-scale features are added to the diffusion UNet decoder to enforce implant geometry during denoising. The output is a simulated postoperative CT with faithful anatomy and morphology, visualized via 3D surface rendering.

2.2 Latent Diffusion Inpainting for Breast ROI Generation

Diffusion models generate samples through progressive denoising from Gaussian noise, enabling high-fidelity and diverse synthesis [17]. In this study, we adopt the conditional Latent Diffusion Model (LDM) with the inpainting scheme to reduce the computational cost for 3D synthesis, while restricting generation within the surgical ROI [18]. Let $\mathbf{z}_0 = E(\mathbf{x}^{post})$ denote the encoded latent of the postoperative volume, and the ROI $\mathbf{m}$ be downsampled to the latent resolution. At diffusion timestep t, the ROI with surgical outcomes is perturbed, while context outside the ROI is unchanged, yielding the partially noised latent:

$$\mathbf{z}_t = \left(\sqrt{\bar{\alpha}_t}\,\mathbf{z}_0 + \sqrt{1-\bar{\alpha}_t}\,\boldsymbol{\epsilon}\right) \odot \mathbf{m} \;+\; \mathbf{z}_0 \odot (1-\mathbf{m}) \tag{1}$$

where $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I})$ is Gaussian noise, and $\bar{\alpha}_t$ controls the noise level. The denoising UNet $\boldsymbol{\epsilon}_\theta$ is trained with a region-restricted noise-prediction loss to predicts the injected noised within the ROI for every step t:

$$\mathcal{L}_{\text{Breast}} = \mathbb{E}_{\epsilon}\left[\frac{1}{|\mathbf{m}|} \left\| \left(\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\mathbf{z}_t, t, \mathbf{m}, \mathbf{c}_c)\right) \odot \mathbf{m} \right\|_2^2\right] \tag{2}$$

During inference, we encode the preoperative volume $\mathbf{x}^{pre}$ and mask the planned surgical region with $\boldsymbol{\epsilon}$. The model then iteratively denoises within the ROI conditioned on the clinical attribute $\mathbf{c}_c$ while preserving the contralateral breast outside $\mathbf{m}$. The final latent is decoded to obtain the simulated postoperative volume $\hat{\mathbf{x}}^{post} = D(\hat{\mathbf{z}}_0)$.

2.3 Implant Prior Construction for Spatial Control

In our cohort, implant is parameterized by type and mass: type follows manufacturer-defined families (e.g., teardrop vs. round), while the mass corresponds to geometric specifications (H: height, W: width, P: projection; cm). We initially condition these attributes together with $\mathbf{c}_c$ via cross-attention, yet this implicit conditioning provides only weak geometric supervision. We therefore model the implant configuration as the explicit 3D spatial prior, providing stronger guidance during synthesis.

Canonical implant template construction. We annotate implant masks from postoperative CTs to build a template library for each type-mass configuration. To overcome the variability from soft tissue confinement and implant rotation, we center each implant at its centroid and align the orientation by rotating PCA axes into a shared canonical coordinate frame. The aligned masks of the same type–mass are voxel-wise averaged to obtain the canonical implant template. Notably, we enable anisotropic scaling of the canonical template to continuously adjust (H, W, P), allowing synthesis beyond manufacturer catalog entries and facilitating flexible preoperative planning.

Patient-specific implant prior construction. Given a preoperative CT, with the specifies the implant type, size and placement center (x, y, z), we retrieve the corresponding template and estimate the in-vivo implant orientation from the PCA axes of ribcage mask of local placement region. We rotate the implant to align its projection axis (the PCA axis with the smallest eigenvalue) with the corresponding ribcage principal direction, while preserving axes orthogonality, yielding an anatomically plausible pose that conforms implant base to the chest wall. Accordingly, the implicit configuration $\mathbf{c}_i$ is converted into a spatial implant prior $\mathbf{I}_p$, providing control for synthesis.

2.4 Spatial Implant Controllable Generation

Building on the inpainting LDM (Sec. 2.2), we enforce the specific implant configuration by injecting the implant prior $\mathbf{I}_p$ via ControlNet [19, 20]. A trainable control branch g_ϕ (initialized from the UNet encoder) encodes $\mathbf{I}_p$ into multi-scale features that are added to the UNet mid and up blocks. The resulting noise predictor is:

$$\boldsymbol{\epsilon}_{\theta,\phi} = \boldsymbol{\epsilon}_\theta \left(\mathbf{z}_t, t, \mathbf{m}, \mathbf{c}_c, g_\phi(\mathbf{I}_p)\right) \tag{3}$$

Training follows the same region-restricted denoising objective (Eq. (2)), replacing $\boldsymbol{\epsilon}_\theta$ by $\boldsymbol{\epsilon}_{\theta,\phi}$. We freeze the pretrained backbone $\boldsymbol{\epsilon}_\theta$ and optimize only g_ϕ. To sharpen implant boundaries and reduce edge jitter, we additionally fine-tune the g_ϕ with implant consistency feedback. For early diffusion steps ($t \leq 500$), where denoising predominantly recovers low-level structural details, we use an intermediate denoised postoperative estimation and decode it into the image space:

$$\hat{\mathbf{x}}_{0,t}^{post} = D\left(\frac{\mathbf{z}_t - \sqrt{1-\bar{\alpha}_t}\,\epsilon_{\theta,\phi})}{\sqrt{\bar{\alpha}_t}} \odot \mathbf{m} \; + \; \mathbf{z}_0 \odot (1 - \mathbf{m})\right) \tag{4}$$

where $D(\cdot)$ denotes the VAE decoder. We feed $\hat{\mathbf{x}}_{0,t}^{post}$ into the frozen implant segmentation network, and minimize the Dice loss [21] between the predicted implant mask and the prior I_p, encouraging stable, accurate implant contours in the synthesized ROI.

3 Experiments

3.1 Experiments Settings

Dataset. We retrospectively collected chest CT scans of patients undergoing mastectomy with immediate breast reconstruction at The First Affiliated Hospital, Zhejiang University School of Medicine, for training and validation. To evaluate preoperative applicability and cross-institutional generalization, we additionally curated paired pre/postoperative CT cases from Hangzhou First People's Hospital, Affiliated Hospital of Westlake University School of Medicine, as an external test set. For each patient, we recorded clinical attributes (age, height, weight, surgical side) and implant configurations (type, mass). After excluding cases with severe implant deformation and displacement, 470/107/87 subjects were used for training/validation/testing with patient-level splits.

Preprocessing and implementation details. We annotate implant masks from CT scans, and train SwinUNETR [22] for automatic implant segmentation. Ribcage and body masks are obtained using TotalSegmentator [23]. For the LDM, we adopt a 3D UNet [24] with attention module at the deepest layer. CT volumes are resampled to $256 \times 256 \times 160$ and encoded using MedVAE [25]. Models are trained using AdamW with learning rate 2×10^{-6} scaled by batch size. We use a cosine noise schedule with 1000 timesteps and perform inference with DDIM [26] sampling using 200 steps.

Competing methods and protocols. We compare against a conditional LDM baseline that conditions on clinical attributes and implant-configuration via cross-attention, using either (i) BiomedCLIP [27] encoded text prompts or (ii) MLP encoded numeric attributes. We further benchmark representative spatially controllable diffusion variants: (iii) GLIGEN [28], (iv) Simple-ControlNet [29]; and (v) T2I-Adapter [30]. For fair comparison, the controllable baselines are adapted by replacing the 2D convolutional modules with 3D counterparts and flattening volumetric tokens for the attention modules, while sharing the same 3D LDM backbone, training data, with only their control branches fine-tuned. We report SSIM, PSNR, MAE, LPIPS, GMSD, and additionally assess controllability by implant Dice/IoU between real and synthesized CT volumes.

3.2 Results

Table 1. Quantitative comparison of synthesis quality across competing methods. Best results are highlighted in bold.

Model	LPIPS↓	SSIM↑	PSNR↑	MAE↓	GMSD ↓
Text-encoded	0.088	0.904	28.382	0.028	0.144
Num-encoded	0.089	0.904	27.590	0.030	0.146
GLIGEN	0.086	0.91	28.646	0.027	0.141
Simple-ControlNet	0.088	0.908	28.480	0.028	0.142
T2I-Adapter	0.082	0.911	28.772	0.027	0.140
Ours	**0.080**	**0.912**	**29.046**	**0.025**	**0.137**

Generative quality. As shown in Table 1, our method achieves the best synthesis quality across all metrics. Compared with implicit cross-attention conditioning (Text-/Num-encoded), methods with spatial control consistently improve fidelity, yielding lower perceptual/pixel-wise errors (LPIPS, MAE) and higher structural similarity (SSIM), even when only the control branch is fine-tuned. Since the dominant appearance changes in the ROI are driven by implant placement and geometry, the spatial prior provides the model with geometric guidance, yielding more consistent outputs. In particular, the lower GMSD indicates better agreement of the implant and breast contour boundaries within the ROI.

Table 2. Quantitative comparison of implant mask agreement between synthesized and reference postoperative CT volumes.

Model	Dice↑	IoU↑
Text-encoded	0.599	0.449
Num-encoded	0.544	0.395
GLIGEN	0.613	0.456
Simple-ControlNet	0.606	0.452
T2I-Adapter	0.722	0.600
Ours	**0.750**	**0.627**

Segmentation results. Table 2 reports implant Dice/IoU between the synthesized and reference postoperative CTs. The Text-/Num-encoded baselines inject implant configuration solely through cross-attention. Although visually plausible, the condition is weakly coupled to the generated implant geometry, resulting in low Dice and IoU. This limitation is particularly evident in challenging scenarios (e.g., the second case in Figure 2, where the implant within the ROI is larger than the native breast), where strong preoperative context in the inpainting can bias the generated ROI toward the unmasked context and follow the contralateral breast shape rather than the intended condition. These observations highlight the need for explicit, spatially grounded control.

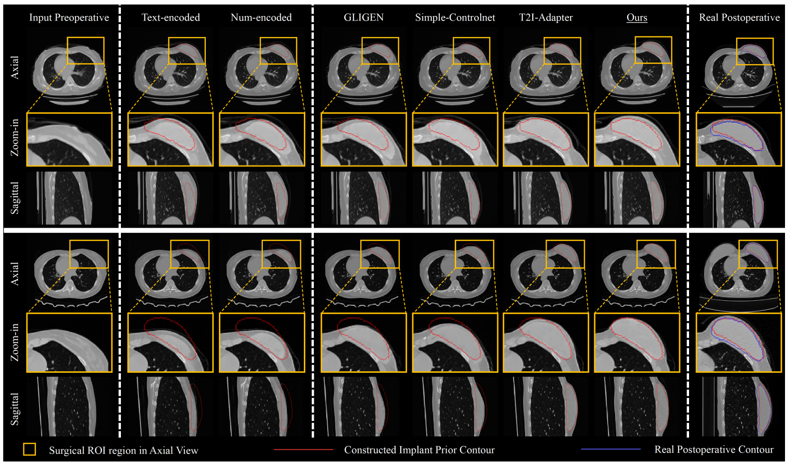

Fig. 2. Qualitative comparison of synthesis. Axial (top) and sagittal (bottom) views are shown for two patients: (a) unilateral reconstruction with an implant size close to the native breast; (b) bilateral reconstruction with larger implants. The postoperative implant contour is outlined in blue, and the implant spatial prior used for spatial control is outlined in red.

With the proposed spatial control signal, close agreement between the injected control (red) and the reference postoperative implant contour (blue) in Figure 2 suggests that this prior provides a reliable geometric cue. It substantially improves implant accuracy across controllable variants, with T2I-Adapter and our method achieving the best overlap. We attribute the inferior performance of GLIGEN and Simple-ControlNet to attention-based control injection, which is applied at the lowest resolution of UNet in our implementation for memory constraints and may be attenuated during decoding. In contrast, T2I-Adapter and our approach inject control features into convolutional feature maps across scales, enabling stronger enforcement of geometric constraints. Finally, with the segmentation-consistency fine-tuning, our method produces sharper implant boundaries and attains the highest Dice/IoU in Table 2.

Simulation results. Figure 3 visualizes controllable simulations generated from preoperative volumes. Row 1 shows the unilateral reconstruction case. Under the ground truth implant configuration (middle column), the synthesis closely matches the real postoperative implant contour and skin-surface profile. By iteratively reducing P and increasing H, our edited configuration (H: 13.3, W: 13, P: 3.3) produces a plausible alternative with a smoother breast contour transition and improved symmetry compared with the real postoperative outcome.
Row 2 shows a bilateral reconstruction case, where we vary implant mass from small to large. We simulate bilateral reconstruction by sequentially synthesizing the ROI of each breast. Consistent with prior findings, the injected spatial implant prior reduces over-reliance on the contralateral breast, enforces the specified configuration, and produces the expected change in breast prominence. Overall, our framework supports

patient-specific preoperative simulation under user-defined implant configurations, facilitating surgical planning and expectation management.

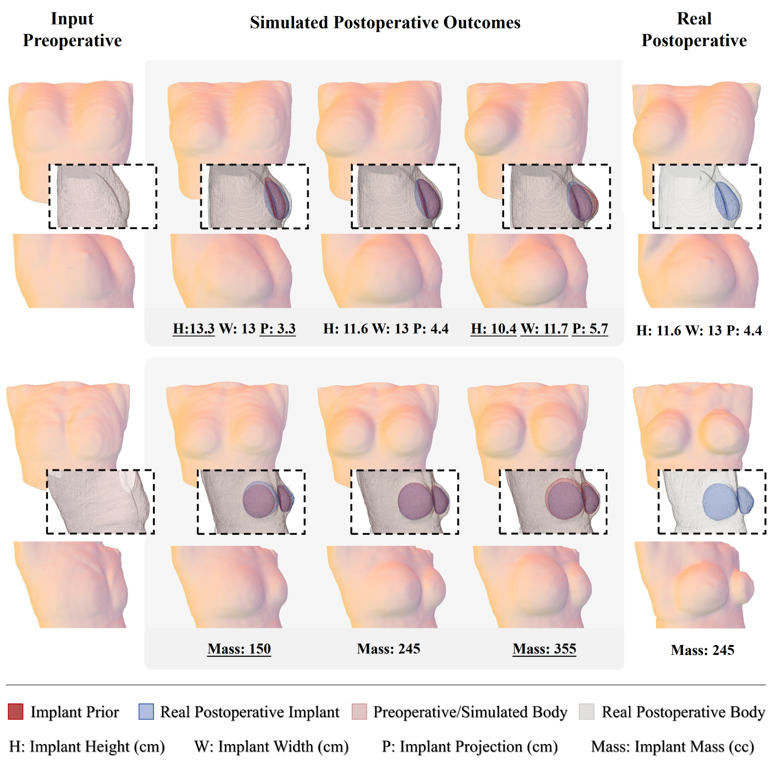

Fig. 3. Qualitative demonstration of controllable simulation using our framework (skin surface renderings derived from CT). Starting from the same preoperative input (left), we generate multiple postoperative simulations by editing the implant configuration. Insets show the injected implant prior (red) and the postoperative implant contour (blue). Row 1 presents a unilateral reconstruction case with varying implant size parameters (H, W, P), and Row 2 presents a bilateral reconstruction case with varying implant mass (corresponds to fixed H W P combinations defined by manufacturer). For each case, the result under the real surgical configuration is shown in the **middle simulated column**; edited configuration variables for the synthesized results are underlined.

4 Conclusion

In this study, we presented a controllable generative framework for patient-specific simulation of postmastectomy breast reconstruction outcomes from preoperative chest CT. By integrating latent diffusion based inpainting with an explicit, editable 3D implant spatial prior injected through a dedicated control branch, our method enables anatomically consistent synthesis while providing fine-grained control over implant geometry and placement. Our model consistently improves both fidelity and implant segmentation accuracy over competing baselines and generalizes robustly in cross-institution evaluation. A limitation is that the simulation remains most reliable when the postoperative implants do not deform much, and the pretrained rib segmentation model works accurately. Future work will try to model biomechanically implant–tissue interaction and deformation, explore parametric controls that map implant configurations directly to simulation outcomes to reduce reliance on spatial priors. Overall, the proposed framework offers a practical route to preoperative visualization and decision support for breast reconstruction planning.

Acknowledgments. We thank the clinician Minya Yao from the Department of Breast Surgery, The First Affiliated Hospital, School of Medicine, Zhejiang University, for her valuable assistance in reconstructive surgery. We also thank Kun Shu from Affiliated Hangzhou First People's Hospital, Westlake University School of Medicine, for facilitating communication and helpful discussions. This work is funded in part by an International Collaboration Fund for Creative Research of National Natural Science Foundation of China (NSFC ICFCRT) under the Grant no. W2441019.

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

References

1. Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R.L., Soerjomataram, I., Jemal, A.: Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians. 74, 229 – 263 (2024).
2. Rocco, N., Catanuto, G.F., Accardo, G., Velotti, N., Chiodini, P., Cinquini, M., Privitera, F., Rispoli, C., Nava, M.B.: Implants versus autologous tissue flaps for breast reconstruction following mastectomy. Cochrane Database Syst Rev. 10, CD013821 (2024).
3. Bolliger, M., Gambone, L., Haeusler, T., Mikula, F., Kampf, S., Fitzal, F.: Patient Satisfaction, Esthetic Outcome, and Quality of Life in Oncoplastic and Reconstructive Breast Surgery: A Single Center Experience. Breast Care (Basel). 19, 215–222 (2024).
4. Gilmour, A., Cutress, R., Gandhi, A., Harcourt, D., Little, K., Mansell, J., Murphy, J., Pennery, E., Tillett, R., Vidya, R., Martin, L.: Oncoplastic breast surgery: A guide to good practice. Eur J Surg Oncol. 47, 2272–2285 (2021).
5. Puckett, Y., Schaffner, A.D.: Implant-Based Breast Reconstruction. In: StatPearls. StatPearls Publishing, Treasure Island (FL) (2025).
6. Zhong, T., Fletcher, G.G., Brackstone, M., Frank, S.G., Hanrahan, R., Miragias, V., Stevens, C., Vesprini, D., Vito, A., Wright, F.C.: Postmastectomy Breast Reconstruction in Patients with Non-Metastatic Breast Cancer: A Systematic Review. Curr Oncol. 32, 231 (2025).
7. Xu, J., Wei, Y., Jiang, S., Zhou, H., Li, Y., Chen, X.: Intelligent surgical planning for automatic reconstruction of orbital blowout fracture using a prior adversarial generative network. Medical Image Analysis. 99, 103332 (2025).
8. Lee, J., Xu, X., Kim, D., Kuang, T., Deng, H.H., Song, X., Soubra, Y., Dharia, R., Liebschner, M.A.K., Gateno, J., Yan, P.: Facial Appearance Prediction with Conditional Multiscale Autoregressive Modeling for Orthognathic Surgical Planning.

9. Ji, Z., Li, N., Xue, P., Dong, Y., Ma, L.: 3D Dynamic Prediction of Missing Teeth in Diverse Patterns via Centroid-prompted Diffusion Model.
10. Kawata, N., Iwao, Y., Matsuura, Y., Higashide, T., Okamoto, T., Sekiguchi, Y., Nagayoshi, M., Takiguchi, Y., Suzuki, T., Haneishi, H.: Generation of short-term follow-up chest CT images using a latent diffusion model in COVID-19. Jpn J Radiol. 43, 622–633 (2025).
11. Park, S., Kim, Y.B., Chang, J.S., Choi, S.H., Chung, H., Lee, I.J., Byun, H.K.: Objective and Interpretable Breast Cosmesis Evaluation with Attention Guided Denoising Diffusion Anomaly Detection Model, http://arxiv.org/abs/2402.18362, (2024).
12. Park, J.K., Baek, S., Heo, C.Y., Jeong, J.H., Myung, Y.: A Novel, Deep Learning-Based, Automatic Photometric Analysis Software for Breast Aesthetic Scoring. Archives of Plastic Surgery. 51, 30–35 (2024).
13. Guo, C., Smith, T.L., Feng, Q., Benitez-Quiroz, F., Vicini, F., Arthur, D., White, J., Martinez, A.: A fully automatic framework for evaluating cosmetic results of breast conserving therapy. Machine Learning with Applications. 10, 100430 (2022).
14. Kim, D.-Y., Lee, S.J., Kim, E.-K., Kang, E., Heo, C.Y., Jeong, J.H., Myung, Y., Kim, I.A., Jang, B.-S.: Feasibility of anomaly score detected with deep learning in irradiated breast cancer patients with reconstruction. npj Digit. Med. 5, 1–7 (2022).
15. Chartier, C., Watt, A., Lin, O., Chandawarkar, A., Lee, J., Hall-Findlay, E.: BreastGAN: Artificial Intelligence-Enabled Breast Augmentation Simulation. Aesthetic Surgery Journal Open Forum. 4, ojab052 (2022).
16. Zeng, W., Schroeder, A., Moyer, K., Thompson, J.: Augmentation AI: Use of a Novel Deep Learning-Based Generative Artificial Intelligence System As a Tool for Breast Augmentation Simulation. Aesthetic Surgery Journal Open Forum. 7, ojaf018.019 (2025).
17. Ho, J., Jain, A., Abbeel, P.: Denoising Diffusion Probabilistic Models, http://arxiv.org/abs/2006.11239, (2020).
18. Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R., Gool, L.V.: RePaint: Inpainting using Denoising Diffusion Probabilistic Models, http://arxiv.org/abs/2201.09865, (2022).
19. Zhang, L., Rao, A., Agrawala, M.: Adding Conditional Control to Text-to-Image Diffusion Models. Presented at the Proceedings of the IEEE/CVF International Conference on Computer Vision (2023).
20. Li, M., Yang, T., Kuang, H., Wu, J., Wang, Z., Xiao, X., Chen, C.: ControlNet++: Improving Conditional Controls with Efficient Consistency Feedback, http://arxiv.org/abs/2404.07987, (2024).
21. Milletari, F., Navab, N., Ahmadi, S.-A.: V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation, http://arxiv.org/abs/1606.04797, (2016).
22. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H., Xu, D.: Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images, http://arxiv.org/abs/2201.01266, (2022).
23. Wasserthal, J., Breit, H.-C., Meyer, M.T., Pradella, M., Hinck, D., Sauter, A.W., Heye, T., Boll, D., Cyriac, J., Yang, S., Bach, M., Segeroth, M.: TotalSegmentator: robust segmentation of 104 anatomical structures in CT images. Radiology: Artificial Intelligence. 5, e230024 (2023).
24. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, http://arxiv.org/abs/1505.04597, (2015).
25. Varma, M., Kumar, A., Sluijs, R. van der, Ostmeier, S., Blankemeier, L., Chambon, P., Bluethgen, C., Prince, J., Langlotz, C., Chaudhari, A.: MedVAE: Efficient Automated Interpretation of Medical Images with Large-Scale Generalizable Autoencoders, http://arxiv.org/abs/2502.14753, (2025).

26. Song, J., Meng, C., Ermon, S.: Denoising Diffusion Implicit Models, http://arxiv.org/abs/2010.02502, (2022).
27. Zhang, S., Xu, Y., Usuyama, N., Xu, H., Bagga, J., Tinn, R., Preston, S., Rao, R., Wei, M., Valluri, N., Wong, C., Tupini, A., Wang, Y., Mazzola, M., Shukla, S., Liden, L., Gao, J., Crabtree, A., Piening, B., Bifulco, C., Lungren, M.P., Naumann, T., Wang, S., Poon, H.: BiomedCLIP: a multimodal biomedical foundation model pretrained from fifteen million scientific image-text pairs, http://arxiv.org/abs/2303.00915, (2025).
28. Li, Y., Liu, H., Wu, Q., Mu, F., Yang, J., Gao, J., Li, C., Lee, Y.J.: GLIGEN: Open-Set Grounded Text-to-Image Generation, http://arxiv.org/abs/2301.07093, (2023).
29. Feng, Z., Chen, L., Sun, Y., Liu, J., Feng, S.: Simplifying Control Mechanism in Text-to-Image Diffusion Models. Proceedings of the AAAI Conference on Artificial Intelligence. 39, 3013–3021 (2025).
30. Mou, C., Wang, X., Xie, L., Wu, Y., Zhang, J., Qi, Z., Shan, Y.: T2I-Adapter: Learning Adapters to Dig Out More Controllable Ability for Text-to-Image Diffusion Models. Proceedings of the AAAI Conference on Artificial Intelligence. 38, 4296–4304 (2024).