# SuRe-Flow: Surface-to-CT Synthesis via Retrieval-Augmented Rectified Flow for CT-Free Superficial Radiotherapy

Cong Liu[1,2,3], Xinye Ni[2,3], and Kai Xie[2,3⋆]

[1] Faculty of Business Information, Shanghai Business School, Shanghai 201499, China
[2] Radiation Oncology Center, Affiliated Changzhou No. 2 People's Hospital of Nanjing Medical University, Changzhou, Jiangsu 213003, China
[3] Center of Medical Physics, Nanjing Medical University, Changzhou, Jiangsu 213003, China
tjcongliu@gmail.com; nxy@njmu.edu.cn; xkai1990@mail.ustc.edu.cn

**Abstract. Background**: Superficial radiotherapy still relies on CT simulation for dose calculation, adding resource burden and imaging radiation for benign or shallow disease. Optical surface scanning is non-ionizing, but surface-to-CT synthesis is ill-posed because one surface can match multiple internal anatomies. **Methods**: We propose SuRe-Flow (Surface Retrieval Flow), a retrieval-augmented rectified-flow framework for superficial synthetic CT generation. A contrastive surface encoder retrieves topologically matched bulk anatomical priors to reduce 1-to-$N$ ambiguity; ControlNet injects these priors into a pretrained RF backbone, and a dose-aware frequency loss preserves tissue-interface detail for electron dosimetry. **Results**: Trained on 20,022 patients and evaluated on 1,000 patients, SuRe-Flow outperformed adversarial and rectified-flow baselines, achieving 5.0 HU MAE in the superficial therapeutic zone (0–20 mm). Treatment-planning-system dosimetry yielded a 96.7% 3D Gamma Index passing rate (2%/2mm). **Conclusion**: SuRe-Flow enables dosimetrically accurate surface-conditioned CT synthesis, supporting the feasibility of CT-free superficial radiotherapy.

**Keywords:** Image Synthesis · Rectified Flow · Superficial Radiotherapy

## 1 Introduction

Superficial radiotherapy for keloids, hypertrophic scars, and non-melanoma skin cancers still relies on CT to map electron density for heterogeneity correction. Although these conditions are superficial and require relatively low doses (typically 12–20 Gy), this workflow burdens clinical resources and exposes a largely young patient population to ionizing radiation solely to acquire density maps.

We investigate the feasibility of a CT-free workflow using generative models, where synthetic CT volumes are generated directly from optical surface scans.

⋆ Corresponding author.

This approach is grounded in the physics of electron beam radiotherapy that has a finite therapeutic range, approximated by $R_{90} \approx E/3.2$ cm [3,4]. Consequently, the dose distribution is determined almost entirely by the superficial anatomy, making deep internal anatomy irrelevant for dosimetry. The generative task is therefore not to reproduce deep organs, but to accurately resolve the shallow scattering interfaces that determine surface dose deposition.

The recoverability of internal anatomy from external topology is widely documented in statistical shape modeling, where linear regression successfully predicts deep skeletal positioning from sparse skin landmarks [7,15]. While this deterministic coupling suggests DL models can map surface curvature to internal density, early attempts using cGANs [2] were limited by slice-based 2D inconsistencies and adversarial training instability. Consequently, modern cross-modality synthesis has shifted toward Rectified Flow (RF) [12]. By modeling complex, high-dimensional distributions via stable likelihood maximization, RF provides a robust, volumetrically consistent way for generative modeling.

Despite the stability of RF, utilizing it as a foundational prior via a standard ControlNet architecture encounters an ill-posed bottleneck. Mapping an external surface to a 3D internal density field is a fundamentally ill-posed 1-to-$N$ inverse problem, as a single surface is physically consistent with multiple internal tissue configurations. In terms of flow dynamics, the straight-line ODE trajectories originating from the same surface condition conflict because they point toward different valid internal anatomies. The standard RF MSE objective forces the network to average these conflicting velocity vectors. We empirically observed that this averaging results in degraded organ contrast, a finding consistent with recent reports in other ill-posed translation tasks [10,13,11].

To overcome this mode-averaging limitation, we propose *SuRe-Flow* (Surface Retrieval Flow) that transitions the generative task from unconstrained generation to a retrieval-augmented one. SuRe-Flow addresses the fundamental ambiguity through three contributions. First, we encode surfaces into a contrastive latent space, allowing the retrieval of a structurally isomorphic patient prior from a large-scale database. Second, we propose a Retrieval-Conditioned Flow that injects this prior, narrowing the solution space and disentangling the intersecting ODE trajectories. Finally, we introduce a Dose-Aware Frequency Penalty that explicitly penalizes image smoothness, ensuring the generation of the sharp tissue interfaces necessary for accurate dose calculation.

## 2 Method

### 2.1 Problem Formulation and Framework Overview

Let $x \in \Omega \subset \mathbb{R}^{D\times H\times W}$ denote the CT density field and $c \in \partial\Omega$ denote the corresponding surface. The objective is to learn the inverse mapping $p(x|c)$. This mapping is ill-posed, yielding $N$ valid internal configurations for a single surface $c$. In standard conditional RF, this 1-to-$N$ ambiguity creates a trajectory conflict, resulting in velocity vector averaging and degraded tissue contrast.

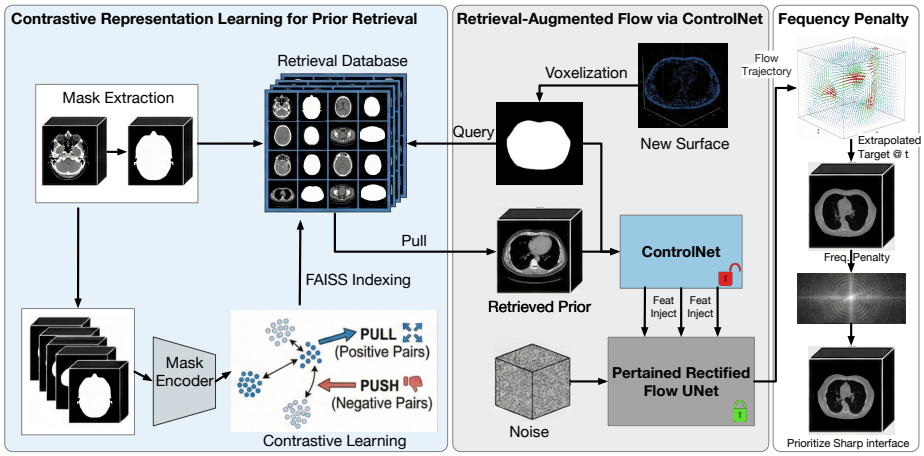


**Fig. 1.** Overview of the SuRe-Flow architecture. The Contrastive Representation Learning module embeds masks to construct a database for retrieving priors. The Retrieval-Augmented Flow via ControlNet module injects the retrieved prior to condition a pre-trained RF. To enforce sharp interfaces, a frequency penalty is computed on the extrapolated target during trajectory optimization.

SuRe-Flow formulates surface-to-CT synthesis as retrieval-augmented generation to disentangle the conflicting trajectories. As shown in Fig. 1, point-cloud surfaces and CT-derived contours are both voxelized into the same occupancy-mask input. First, 3D body masks are embedded via contrastive representation learning to construct an offline database comprising mask embeddings and their corresponding CT latents. The retrieval-augmented flow module then queries this database using the target voxelized surface to extract a reference prior. A trainable ControlNet subsequently integrates this prior to condition a frozen, pre-trained RF model. Finally, a standard RF loss is coupled with a dose-aware focal frequency penalty to enforce sharp tissue interfaces.

### 2.2 Contrastive Representation Learning for Anatomical Retrieval

Retrieving structurally isomorphic priors via voxel-wise metrics (e.g., Dice similarity) across a large-scale database is computationally prohibitive. To enable efficient retrieval, we train an encoder $\mathcal{F}_{\text{msk}}$, based on SegResEncoder [14], to map the surface into a compressed embedding space. The topological surface is first voxelized into a binary occupancy mask $m \in \{0,1\}^{D\times H\times W}$ to facilitate 3D convolutions. A projection head then maps the encoder's 512-D global average-pooled embeddings into a 256-D space for contrastive training.

This contrastive representation is learned by minimizing the Normalized Temperature-scaled Cross Entropy loss [1]. Positive mask pairs $(\tilde{m}_i, \tilde{m}_j)$ are generated via stochastic augmentations, including random affine transformations, scaling, and elastic deformations. Negative samples comprise all other masks

within the same batch. The loss is computed as:

$$\mathcal{L}_{cont} = -\log \frac{\exp(\text{sim}(h_i, h_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{I}_{[k \neq i]} \exp(\text{sim}(h_i, h_k)/\tau)} \tag{1}$$

where $h$ denotes the 256-D projection, $\text{sim}(\cdot, \cdot)$ represents cosine similarity, $\tau$ is the temperature scaling parameter, and $N$ is the batch size.

Following contrastive training, we construct an approximate nearest neighbor index of the 512-D $L_2$-normalized embeddings for all training masks using the FAISS library [9]; validation/test masks are excluded. For each target, the top-5 nearest anatomical neighbors are precomputed. During training, we regularize the generation by randomly sampling one reference prior from this top-5 neighborhood. During inference, the top-1 retrieved neighbor conditions the model.

### 2.3 Retrieval-Augmented Flow via ControlNet

To provide a robust anatomical foundation for the ill-posed surface-to-CT translation, we leverage a large-scale latent RF model (MAISI-V2 [19]) pre-trained on over 107k CT volumes. Freezing this generative backbone $\mathcal{F}_{\text{RF}}$ preserves generalized CT representations without catastrophic forgetting.

We adapt the frozen $\mathcal{F}_{\text{RF}}$ via a trainable ControlNet $\mathcal{F}_{\text{Ctrl}}$ that integrates the mask condition $c = \mathcal{F}_{\text{msk}}(m)$ and the anatomical prior $z_{\text{ref}} = \mathcal{E}_{\text{VAE}}(x_{\text{ref}})$, where $x_{\text{ref}}$ is the retrieved reference CT and $\mathcal{E}_{\text{VAE}}$ is MAISI's VAE encoder [5]. We deliberately project $c$ to match the spatial dimensions of $z_{\text{ref}}$ ($\mathbb{R}^{4 \times \frac{D}{4} \times \frac{H}{4} \times \frac{W}{4}}$) to facilitate multi-scale conditioning and reduce GPU memory overhead.

Specifically, $z_t$ and $c$ are concatenated $[z_t, c]$ and fed into $\mathcal{F}_{\text{Ctrl}}$ to anchor the generation to the target surface, while $z_{\text{ref}}$ is injected into intermediate blocks of $\mathcal{F}_{\text{Ctrl}}$ via spatial cross-attention. The queries $Q$ are derived from $\mathcal{F}_{\text{Ctrl}}$ hidden spatial features, while keys $K$ and values $V$ are projected from $z_{\text{ref}}$. This allows the network to query the retrieved internal anatomy where the surface condition is ambiguous. The target mask constrains geometry; the retrieved latent supplies only an anatomical prior. The ControlNet outputs are then added to the corresponding hidden states of $\mathcal{F}_{\text{RF}}$ through zero-initialized 3D convolutions.

The ControlNet is optimized using the velocity matching objective:

$$\mathcal{L}_{\text{RF}} = \int_0^1 \mathbb{E}_{z_0, z_1, t} \left[ ||v_\theta(z_t, c, z_{\text{ref}}, t) - (z_1 - z_0)||_2^2 \right] dt \tag{2}$$

where $z_0 \sim \mathcal{N}(0, I)$ and $z_1 = \mathcal{E}_{\text{VAE}}(x)$ is the target CT latent. By explicitly conditioning the predicted vector field on $z_{\text{ref}}$, this formulation narrows the solution space, disentangling the intersecting ODE trajectories that arise from the 1-to-$N$ ambiguity.

### 2.4 Dose-Aware Frequency Penalty

In superficial electron beam therapy, accurate dosimetry depends strictly on steep electron density gradients $\nabla \rho_e$ at tissue interfaces. To explicitly enhance

the preservation of these high-frequency structural transitions during generation, we introduce a dose-aware frequency penalty.

At any continuous timestep $t \in [0, 1]$, utilizing a first-order Euler step, we extrapolate the predicted target latent $\hat{z}_1$ directly from the current state $z_t$ [6,16]: $\hat{z}_1 = z_t + (1 - t)v_\theta(z_t, c, z_{\text{ref}}, t)$. We then transform both the predicted target $\hat{z}_1$ and the ground-truth target $z_1$ into the frequency domain via a Fast Fourier Transform $\mathcal{F}(\cdot)$. To force the network to prioritize sharp structural boundaries, we apply a Focal Frequency Loss [8], which utilizes a dynamic spectrum weighting matrix $W(u, v, w)$ to down-weight low-frequency global structures and explicitly penalize discrepancies in the high-frequency domain:

$$\mathcal{L}_{\text{freq}} = \mathbb{E}_{z_0, z_1, t}\left[||W(u, v, w) \odot (\mathcal{F}(\hat{z}_1) - \mathcal{F}(z_1))||_2^2\right] \tag{3}$$

where $\odot$ is the Hadamard product. According to Parseval's Theorem, minimizing this spectral distance guarantees a reduction of high-frequency structural variance in spatial domain. The final objective for ControlNet training is a weighted sum of the RF loss (2) and the frequency penalty (3): $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{RF}} + \lambda\mathcal{L}_{\text{freq}}$.

## 3 Experiments and Results

### 3.1 Datasets and Preprocessing

We utilized the CADS dataset [18], comprising 22,022 CT volumes aggregated from 100+ imaging centers across 16 countries. The dataset was partitioned into 20,022 training, 1,000 validation, and 1,000 testing volumes. All volumes were resampled to a $1 \times 1 \times 3$ mm resolution. Voxel intensities were clipped to $[-1000, 1000]$ HU and min-max normalized to $[0, 1]$. External body contours were extracted using TotalSegmentator [17].

### 3.2 Anatomical Prior Retrieval Analysis

We evaluated the retrieval module using DSC, ASSD, and HD95 across the 20,022-patient retrieval database (Table 1). The Top-1 neighbor yielded a mean DSC of $0.9442 \pm 0.05$ and an ASSD of $3.12 \pm 0.84$ mm. Top-3 and 5 neighbors maintained comparable alignment, validating the local neighborhood sampling during training. Conversely, bottom-ranked matches (Bottom-1 to 5) exhibited substantial deviation, confirming the retrieval module's discriminative capacity.

Qualitative evaluations (Fig. 2) are consistent with these quantitative metrics. The visualizations show stable external contours and internal structural proportions between the target queries and corresponding Top-5 neighbors.

### 3.3 Ablation Study on Anatomical Depth

To isolate module contributions, we conducted a surface-stratified ablation study. Euclidean distance transforms generated inward from the external surface defined three depth zones: Zone 1 (0–20 mm; critical therapeutic range), Zone 2 (20–50 mm; extended boundary), and Zone 3 (50–100 mm; near-field anatomy).

**Table 1.** Quantitative Evaluation of Anatomical Prior Retrieval

| Retrieval Rank | DSC ↑ | ASSD (mm) ↓ | HD95 (mm) ↓ |
|---|---|---|---|
| Top-1 | $\mathbf{0.9442 \pm 0.05}$ | $\mathbf{3.12 \pm 0.84}$ | $\mathbf{9.19 \pm 6.99}$ |
| Top-3 | $0.9275 \pm 0.04$ | $3.68 \pm 1.05$ | $11.86 \pm 8.24$ |
| Top-5 | $0.9077 \pm 0.04$ | $4.15 \pm 1.16$ | $12.81 \pm 9.06$ |
| Bottom-5 | $0.8373 \pm 0.03$ | $8.80 \pm 3.10$ | $19.22 \pm 15.20$ |
| Bottom-3 | $0.8335 \pm 0.04$ | $9.45 \pm 3.45$ | $23.38 \pm 16.10$ |
| Bottom-1 | $0.8017 \pm 0.05$ | $9.90 \pm 2.95$ | $25.33 \pm 17.80$ |

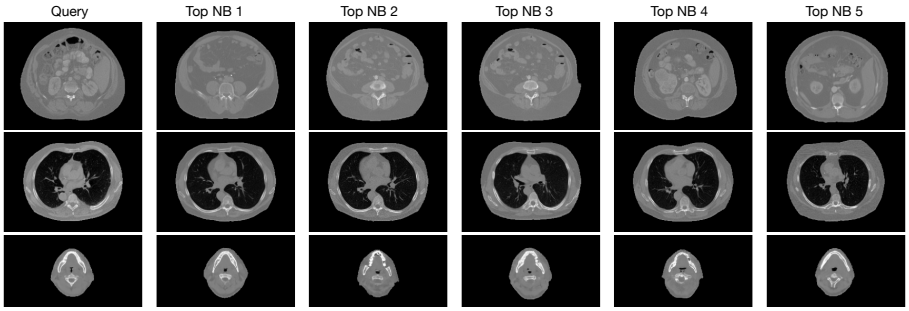


**Fig. 2.** Qualitative results of anatomical prior retrieval. The retrieved neighbors exhibit similar external contours and internal anatomical proportions to the target query.

We evaluated three configurations: a ControlNet-adapted Baseline RF (surface conditioning only), Baseline + Retrieval Prior, and SuRe-Flow (Baseline + Retrieval + Frequency Penalty). As Table 2 demonstrates, the Baseline RF struggles with the ill-posed mapping. Integrating the retrieval prior ($z_{\text{ref}}$) anchors the generation, reducing MAE in the critical Zone 1 from 30.2 to 5.3 HU, with corresponding error reductions in Zones 2 and 3. Finally, the dose-aware frequency penalty yields consistent PSNR improvements across all depths, ensuring the superficial structural fidelity required for accurate Zone 1 dosimetry.

### 3.4 Comparison with State-of-the-Art

We benchmarked SuRe-Flow against three generative baselines: a 3D Regression model, a 3D GAN adapted from Crowe et al. [2], and a frozen MAISI-v2 CT generation model [19] with surface conditioning ControlNet. To validate clinical motivation, we report image metrics alongside the superficial Zone 1 MAE. Furthermore, we simulate electron beam therapy within Monaco TPS, reporting the 3D Gamma Index passing rate (2%/2mm) to quantify dosimetric viability.

As shown in Table 3, the Regression and GAN baselines yield high Zone 1 errors, resulting in low dosimetric passing rates. The MAISI-v2 RF achieves improved metrics, but its lack of an anatomical prior and high-frequency constraints results in a Zone 1 MAE of 30.2 HU and a sub-clinical Gamma passing

**Table 2.** Depth-Stratified Ablation Study on Generative Performance

| Depth Zone | Model Variant | PSNR ↑ | MAE (HU) ↓ |
|---|---|---|---|
| **Zone 1** (0–20 mm) | Baseline RF | $30.57 \pm 1.24$ | $30.2 \pm 7.1$ |
| | + Retrieval Prior | $35.73 \pm 1.10$ | $5.3 \pm 1.4$ |
| | **SuRe-Flow (Full)** | $\mathbf{36.83 \pm 0.95}$ | $\mathbf{5.0 \pm 1.1}$ |
| **Zone 2** (20–50 mm) | Baseline RF | $19.35 \pm 2.15$ | $114.5 \pm 18.3$ |
| | + Retrieval Prior | $20.28 \pm 1.85$ | $102.7 \pm 15.6$ |
| | **SuRe-Flow (Full)** | $\mathbf{21.33 \pm 1.42}$ | $\mathbf{94.1 \pm 12.4}$ |
| **Zone 3** (50–100 mm) | Baseline RF | $19.77 \pm 2.41$ | $121.8 \pm 22.1$ |
| | + Retrieval Prior | $19.94 \pm 2.10$ | $107.3 \pm 18.5$ |
| | **SuRe-Flow (Full)** | $\mathbf{20.98 \pm 1.76}$ | $\mathbf{101.6 \pm 14.8}$ |

rate. SuRe-Flow integrates both modules, reducing the superficial error to 5.0 HU, which corresponds to a 96.7% dosimetric passing rate.

Qualitative comparisons (Fig. 3) reflect the quantitative results. The Regression model produces structural blurring, and GAN introduces adversarial artifacts. The MAISI-v2 RF generates plausible global anatomy but exhibits spatial ambiguity and smoothed boundaries at tissue interfaces. In contrast, SuRe-Flow reconstructs sharp density gradients that closely align with the ground truth.

**Table 3.** Quantitative Comparison with State-of-the-Art Models

| Method | Overall | | | Surface | Dosimetry |
|---|---|---|---|---|---|
| | MAE ↓ | PSNR ↑ | FID ↓ | Zone 1 MAE ↓ | Gamma ↑ |
| 3D Regression | 105.19 ± 25.32 | 20.51 ± 1.30 | 11.46 | 78.5 ± 28.5 | 72.3% |
| 3D GAN[2] | 72.70 ± 17.24 | 20.80 ± 1.66 | 5.43 | 69.6 ± 19.4 | 76.8% |
| MAISI-v2 RF [19] | 36.72 ± 13.10 | 22.17 ± 2.96 | 2.47 | 30.2 ± 7.1 | 89.4% |
| **SuRe-Flow** | **32.14 ± 12.86** | **25.68 ± 2.95** | **1.56** | **5.0 ± 1.1** | **96.7%** |

## 4 Conclusion

We proposed SuRe-Flow, a retrieval-augmented rectified-flow model for superficial radiotherapy, and evaluated it on a 1,000-patient test set. The retrieval module identified matched anatomical priors with a Top-1 DSC of 0.94, demonstrating anatomical alignment. Depth-stratified ablation confirmed integrating this prior with a dose-aware frequency penalty reduced deep spatial ambiguity and superficial boundary errors. SuRe-Flow outperformed existing state-of-the-art models. It reduced the MAE in the 0–20 mm superficial zone to 5.0 HU and yielded a 96.7% 3D Gamma Index passing rate (2%/2mm), establishing

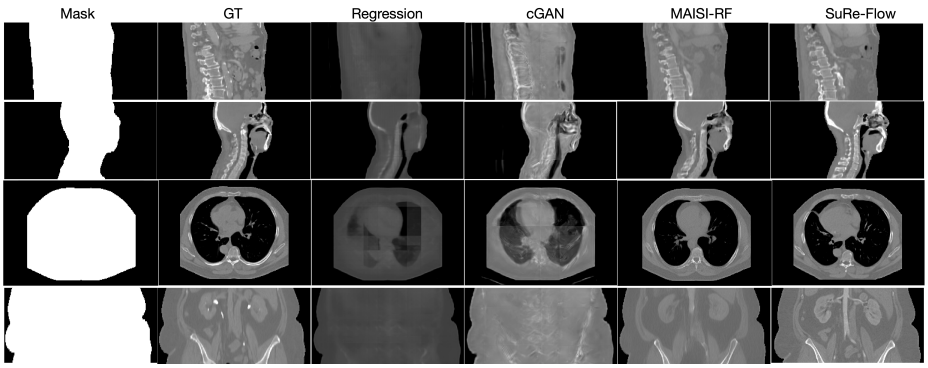


**Fig. 3.** Qualitative comparison of synthetic CT generation across baseline models and SuRe-Flow. SuRe-Flow preserves fine structural details and sharp boundary gradients.

the feasibility of a CT-free workflow for superficial treatments. Future work will validate noisy/occluded optical scans and investigate 4D anatomical tracking .

**Acknowledgments.** Supported by Shanghai Natural Science Foundation (25ZR1401273).

**Disclosure of Interests.** The authors declare no competing interests.

## References

1. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: Proceedings of the 37th International Conference on Machine Learning. vol. 119, pp. 1597–1607 (2020)
2. Douglass, M., Gorayski, P., Patel, S., Santos, A.: Synthetic cranial MRI from 3D optical surface scans using deep learning for radiation therapy treatment planning. Physical and Engineering Sciences in Medicine **46**(1), 367–375 (2023). https://doi.org/10.1007/s13246-023-01229-4
3. Gerbi, B.J., Antolak, J.A., Deibel, F.C., Followill, D.S., Herman, M.G., Higgins, P.D., Hogstrom, K.R., Huq, M.S., Khan, F.M., Mihailidis, D.N., Yorke, E.D.: Recommendations for clinical electron beam dosimetry: Supplement to the recommendations of task group 25. Medical Physics **36**(7), 3239–3279 (2009). https://doi.org/10.1118/1.3125820
4. Gibbons, J.P.: Khan's The Physics of Radiation Therapy. Lippincott Williams & Wilkins, 6 edn. (2019)
5. Guo, P., Zhao, C., Yang, D., Xu, Z., Nath, V., Tang, Y., Simon, B., Belue, M., Harmon, S., Turkbey, B., et al.: MAISI: Medical AI for synthetic imaging. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 4430–4441. IEEE (2025). https://doi.org/10.1109/WACV61041.2025.00435
6. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: Advances in Neural Information Processing Systems. vol. 33, pp. 6840–6851 (2020)

7. Iyer, K., Morris, A., Zenger, B., Karanth, K., Orkild, B.A., Korshak, O., Elhabian, S.: Statistical shape modeling of biventricular anatomy with shared boundaries. In: International Workshop on Statistical Atlases and Computational Models of the Heart. pp. 302–316. Springer (2022). https://doi.org/10.1007/978-3-031-23443-9_28
8. Jiang, L., Dai, B., Wu, W., Loy, C.C.: Focal frequency loss for image reconstruction and synthesis. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 13919–13929 (2021). https://doi.org/10.1109/ICCV48922.2021.01366
9. Johnson, J., Douze, M., Jégou, H.: Billion-scale similarity search with GPUs. IEEE Transactions on Big Data **7**(3), 535–547 (2019). https://doi.org/10.1109/TBDATA.2019.2921572
10. Li, Y., Yang, S., Liu, J.: Language-based image colorization: A benchmark and beyond. arXiv preprint arXiv:2503.14974 (2025)
11. Liu, X., Zhou, Z., Xu, Z., Cao, J., Chen, Z., Zhang, Y.: FideDiff: Efficient diffusion model for high-fidelity image motion deblurring. arXiv preprint arXiv:2510.01641 (2025)
12. Liu, X., Gong, C., Liu, Q.: Flow straight and fast: Learning to generate and transfer data with rectified flow. arXiv preprint arXiv:2209.03003 (2022)
13. Motorcu, H., Cetin, M.: PG-ControlNet: A physics-guided ControlNet for generative spatially varying image deblurring. arXiv preprint arXiv:2511.21043 (2025)
14. Myronenko, A.: 3D MRI brain tumor segmentation using autoencoder regularization. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. pp. 311–320. Springer (2019). https://doi.org/10.1007/978-3-030-11726-9_28
15. Nolte, D., Ko, S.T., Bull, A.M., Kedgley, A.E.: Reconstruction of the lower limb bones from digitised anatomical landmarks using statistical shape modelling. Gait & Posture **77**, 269–275 (2020). https://doi.org/10.1016/j.gaitpost.2020.02.010
16. Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., Chen, M.: Hierarchical text-conditional image generation with CLIP latents. arXiv preprint arXiv:2204.06125 (2022)
17. Wasserthal, J., Breit, H.C., Meyer, M.T., Pradella, M., Hinck, D., Sauter, A.W., Heye, T., Boll, D.T., Cyriac, J., Yang, S., et al.: TotalSegmentator: Robust segmentation of 104 anatomic structures in CT images. Radiology: Artificial Intelligence **5**(5), e230024 (2023). https://doi.org/10.1148/ryai.230024
18. Xu, M., Amiranashvili, T., Navarro, F., Fritsak, M., Hamamci, I.E., Shit, S., Wittmann, B., Er, S., Christ, S.M., de la Rosa, E., Deseoe, J., Graf, R., Möller, H., Sekuboyina, A., Peeken, J.C., Becker, S., Baldini, G., Haubold, J., Nensa, F., Hosch, R., Mirajkar, N., Khalid, S., Zachow, S., Weber, M.A., Langs, G., Wasserthal, J., Ozdemir, M.K., Fedorov, A., Kikinis, R., Tanadini-Lang, S., Kirschke, J.S., Combs, S.E., Menze, B.: CADS: A comprehensive anatomical dataset and segmentation for whole-body anatomy in computed tomography. arXiv preprint arXiv:2507.22953 (2025)
19. Zhao, C., Guo, P., Yang, D., Tang, Y., He, Y., Simon, B., Belue, M., Harmon, S., Turkbey, B., Xu, D.: MAISI-v2: Accelerated 3D high-resolution medical image synthesis with rectified flow and region-specific contrastive loss. arXiv preprint arXiv:2508.05772 (2025)