Anatomy-Conserving Unpaired CBCT-to-CT Translation via Schrödinger Bridge

Ke Shi[1*], Song Ouyang[1*], Gang Liu[2*], Yong Luo[1], Kehua Su[1†], Zhiwen Liang[3,2†], and Bo Du[1]

[1] School of Computer Science, National Engineering Research Center for Multimedia Software and Hubei Key Laboratory of Multimedia and Network Communication Engineering, Wuhan University, China
[2] Cancer Center, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology
[3] Electronic Information School, Wuhan University, China
{shike0201,ouyangsong,luoyong,skh,dubo}@whu.edu.cn
{liugang9910,lzw1091981}@163.com
*Equal contribution; †Corresponding authors

Abstract. Unpaired Cone-beam CT (CBCT)-to-CT translation is pivotal for radiotherapy planning, aiming to synergize CBCT's clinical practicality with CT's dosimetric precision. Existing methods, limited by scarce paired data and registration errors, struggle to preserve anatomical fidelity—a critical requirement to avoid incorrect diagnosis and inadequate treatments. Current CycleGAN-derived approaches risk structural distortions, while diffusion models oversmooth high-frequency details vital for dose calculation in the reverse diffusion. In this paper, we propose the Anatomy-Conserving Schrödinger Bridge (ACSB), a novel unpaired medical image translation framework leveraging entropy-regularized optimal transport to disentangle modality-specific artifacts from anatomy. We incorporate a carefully designed generator, Anatomy-Conserving vision transformer (AC-ViT) to integrate multi-scale anatomical priors via attention-guided feature fusion. We further adopt frequency-aware optimization targeting radiotherapy-critical spectral components. Extensive experiments on the dataset demonstrate the superiority of the proposed ACSB, showcasing excellent generalization over different anatomically distinct regions. Code: https://github.com/Lalala-iks/ACSB

Keywords: Unpaired CBCT-to-CT translation · Schrödinger Bridge · Anatomical fidelity.

1 Introduction

Image-guided radiotherapy (IGRT) has revolutionized precision cancer treatment by enabling real-time anatomical tracking during radiation delivery[6,14]. Central to IGRT workflows, cone-beam CT (CBCT) serves as the primary imaging modality for frequent scans due to its clinical advantages, including lower

radiation exposure and greater portability compared to traditional CT[26]. However, CBCT suffers from primarily lower image quality due to scatter artifacts and noise[8]. This degradation in image quality can lead to inaccurate diagnosis. Consequently, Synthetic CT (sCT) generation[1] techniques aim to bridge this gap by translating CBCT into CT-equivalent images, combining CBCT's accessibility with CT's dosimetric precision.

Deep learning has revolutionized medical image translation for sCT generation. Early supervised approaches (e.g., U-Net [21] variants) relied on pixel-wise losses but required strictly paired CBCT-CT, which are time-consuming and prone to errors in clinical practice. Generative Adversarial Networks (GANs) [9] introduced adversarial training to enhance image realism, while CycleGAN[27]-based methods enabled unpaired translation through cycle-consistency constraints. However, GANs still suffer from mode collapse and gradient instability. While diffusion models have shown impressive results in high-quality image generation through iterative denoising processes, they still struggle to fully recover anatomical details through learned reverse diffusion[10,23,7,11,20]. Despite these features, current methods are limited by a critical oversight: they focus primarily on global intensity matching rather than preserving radiation-sensitive anatomical features, which poses potential risks in radiotherapy applications.

In this paper, we propose ACSB, an Anatomy-conserving Schrödinger Bridge framework. We first establish a direct mapping for unpaired CBCT-to-CT translation using optimal transport theory, then incorporate an anatomy-conserving vision transformer to integrate multi-scale features, while a frequency-constraint component is added to prioritize anatomical details, improving the preservation of critical structures. Our main contributions are threefold: (1) We pioneer the adaptation of Schrödinger Bridge theory [24,5] to cross-modal medical translation, establishing an effective unified framework that simultaneously addresses unpaired learning and precise anatomical preservation; (2) We design a novel generator architecture that enforces anatomical perception by synergizing spectral grouping and local-global hierarchy; (3) We demonstrate unprecedented robustness across different anatomically distinct regions (chest, head and neck (H&N)).

2 Method

An overview of the proposed method is shown in Fig. 1. To go for unpaired CBCT-to-CT translation, we incorporates Schrödinger Bridge into our ACSB framework that directly connects two arbitrary distributions via entropy optimal transport and interpolation mechanism (IPM). Furthermore, our ACSB enforces anatomy-conserving by local-global hierarchy in our anatomy-conserving vision transformer (AC-ViT) generator. Below, we first present how Schrödinger Bridge handles unpaired distributions, then elaborate the IPM and AC-ViT, followed by detailed descriptions of training and generation.

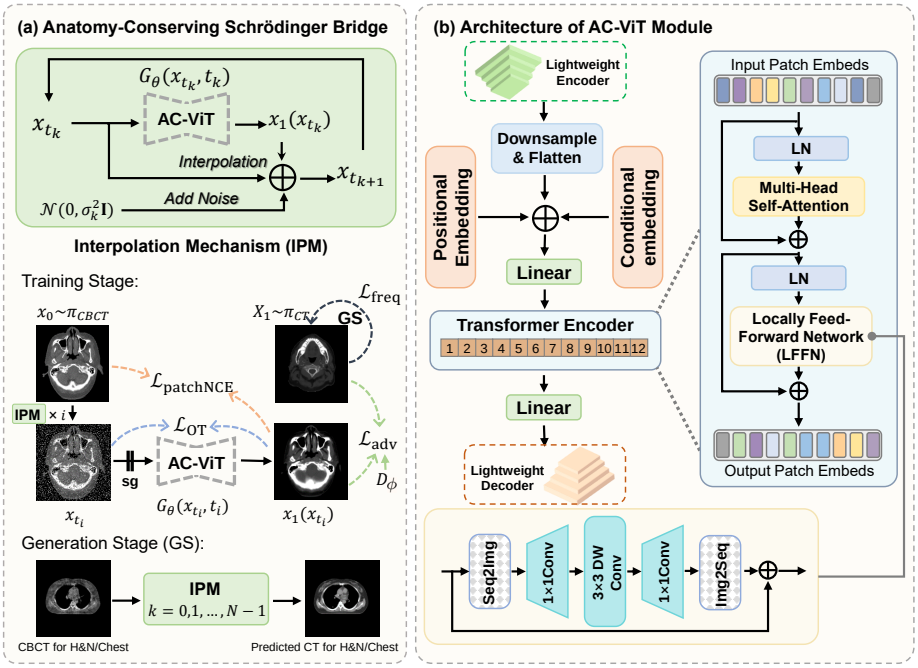

Fig. 1. Overview of the our ACSB framework. (a) The Interpolation Mechanism (IPM) applied in both training and generation stages, showing the progressive transformation from CBCT to CT; (b) The Anatomy-Conserving Vision Transformer (AC-ViT) architecture, featuring a local-global hierarchical design and multi-scale feature alignment for preserving anatomical details.

2.1 Unpaired Translation via Entropy Optimal Transport

Conventional diffusion models assume Gaussian noise priors, limiting their applicability to medical imaging where data distributions are non-Gaussian and often unpaired. Instead, we adopt the Schrödinger Bridge framework that directly connects π_{CBCT} and π_{CT} through the optimal transport plan $\mathbb{Q}^*$, minimizing both control energy and distributional divergence by [16]:

$$\min_{\mathbb{Q}} \mathbb{E}_{\mathbb{Q}} \left[\int_0^1 \frac{1}{2} \|u(t, x_t)\|^2 dt \right] + \lambda D_{\text{KL}}(\mathbb{Q} \parallel \mathbb{W}^\tau) \quad \text{s.t.} \quad x_0 \sim \pi_{\text{CBCT}}, x_1 \sim \pi_{\text{CT}}, \tag{1}$$

where $u(t, x_t)$ is parameterized by the proposed AC-ViT generator G_θ (see Section 2.3), and $\mathbb{W}^\tau$ is the Wiener measure with variance τ.

2.2 Interpolation Mechanism (IPM)

To achieve a smooth trajectory bridging π_{CBCT} and π_{CT}, we propose an Interpolation Mechanism (IPM) that recursively refines intermediate samples x_{t_k} by

blending the current state and the generator's prediction, plus a controlled noise injection. Specifically, given the generator output $x_1(x_{t_k}) = G_\theta(x_{t_k}, t_k)$, the next state $x_{t_{k+1}}$ is sampled via:

$$x_{t_{k+1}} \sim \mathcal{N}\Big(\alpha_{k+1}\, x_1(x_{t_k}) + \big(1 - \alpha_{k+1}\big)\, x_{t_k},\ \sigma_{k+1}^2\, \mathbf{I} \Big), \tag{2}$$

where

$$\alpha_{k+1} = \frac{t_{k+1} - t_k}{1 - t_k}, \quad \sigma_{k+1}^2 = \tau\, \alpha_{k+1}\, \big(1 - \alpha_{k+1}\big). \tag{3}$$

Here, α_{k+1} balances *anatomical structure preservation* from x_{t_k} and the predicted target domain features in $x_1(x_{t_k})$, while σ_{k+1}^2 controls the injected noise level via the diffusion parameter τ.

Notably, IPM is utilized *both* in the training stage (for random time-step refinement) and in the generation stage (for multi-step iterative sampling).

2.3 Anatomy-Conserving Vision Transformer (AC-ViT)

As illustrated in Fig. 1(b), our AC-ViT generator is designed to ensure multi-scale feature alignment while preserving critical anatomical details in CBCT-to-CT translation. In AC-ViT, the input CBCT x_{t_k} first passes through two convolution layers with downsampling. For embedding phase, instead of standard positional embeddings [25], we adopt spectral-aware positional embedding based on FFT [2] to suppress grid artifacts and preserve key diagnostic details. A dynamic conditional encoding is adopted to integrate transport dynamics (time step t and noise level σ) via channel-wise affine transformations and concatenation.

To refine both global anatomical consistency and local structural details, the combined token sequence is progressively refined through a stack of 12 Transformer Encoder blocks. Each block follows a sequential architecture to combine global attention and localized processing. First, the input tokens undergo Layer Normalization (LN), followed by Multi-Head Self-Attention (MHSA) to capture long-range spatial dependencies. The output is then combined with the original input via residual connections. Formally, for encoder input $f_k^{(i)}$ $(i = 1, 2, ..., 12)$:

$$u^{(i)} = \mathrm{MHSA}(\mathrm{LN}(f_k^{(i)})) + f_k^{(i)}. \tag{4}$$

Then, we adopt a Locally Feed-Forward Network (LFFN) to enhance local spatial coherence. Here, the input is temporarily reshaped into a 2D feature map (Seq2Img), processed by two 1×1 convolutions along with a depth-wise 3×3 convolution, then flattened back into a token sequence (Img2Seq). This localized processing stage is preceded by another LN and similarly integrated with residual connections, reinforcing feature stability across layers:

$$f_k^{(i+1)} = \mathrm{LFFN}(\mathrm{LN}(u^{(i)})) + u^{(i)}. \tag{5}$$

Lastly, a Linear and a symmetric lightweight decoder, mirroring the encoder's structure are adopted to reconstruct the $f_k^{(13)}$ to generate the final CT image.

2.4 ACSB Training & Generation Stage

Training Stage As illustrated in Fig. 1(a), we train G_θ (AC-ViT) and a patch-level discriminator [21] D_ϕ jointly, augmented by anatomical regularization and frequency-domain constraints. Specifically, we randomly choose a time-step t_i to optimize, use the IPM (Eq. (2)) for i iterations (starting from $x_0 \sim \pi_{\text{CBCT}}$) to obtain x_{t_i}. Then we get $x_1(x_{t_i}) = G_\theta(x_{t_i}, t_i)$ and sample $x_1 \sim \pi_{\text{CT}}$ meanwhile. We employ an Optimal Transport loss $\mathcal{L}_{\text{OT}}$ to map π_{CBCT} and π_{CT} as follows:

$$x_1(x_{t_i}) = G_\theta(x_{t_N}, t_N), \quad \mathcal{L}_{\text{OT}} = \mathbb{E}\|x_{t_i} - x_1(x_{t_i})\|^2. \tag{6}$$

The patch-based discriminator D_ϕ distinguishes genuine CT samples from synthesized ones by:

$$\mathcal{L}_{\text{adv}} = \mathbb{E}_{x_1}[\log D_\phi(x_1)] + \mathbb{E}[\log(1 - D_\phi(x_{t_i}))]. \tag{7}$$

To conserve anatomical information, we use NCE loss for regularization and frequency-aware loss for dual-domain constraints in frequency-level. We enforce semantic alignment between the input CBCT x_0 and the synthetic CT ($x_1(x_{t_i})$ through contrastive learning in patch-level by:

$$\mathcal{L}_{\text{patchNCE}} = -\mathbb{E}\left[\log \frac{\exp(\psi(x_0) \cdot \psi(x_1(x_{t_i}))}{\sum x_{t_i} \exp(\psi(x_0) \cdot \psi(x_1(x_{t_i}))}\right], \tag{8}$$

where $\psi(\cdot)$ extracts features from AC-ViT's multi-scale layers. In order to guide the model to focus more on the difficulty to sythesize frequency details in CT images, we further incorporating the focal frequency loss [15] into the indentity mapping when choosing $x_1 \sim \pi_{CT}$ as the input, thereby mitigating the issue of high-frequency detail loss in synthesized CT images by:

$$\mathcal{L}_{\text{freq}} = \|\mathcal{F}(x_1(x_{t_k})) - \mathcal{F}(x_1)\|_1 + \|\|\mathcal{F}(x_1(x_{t_k}))\|_1 - \|\mathcal{F}(x_1)\|_1\|_2, \tag{9}$$

where $\mathcal{F}$ denotes fast Fourier transform (FFT) [3], enforcing both phase and magnitude alignment. The full objective of ACSB becomes:

$$L_{\text{ACSB}} = \mathcal{L}_{\text{adv}} + \lambda_{\text{OT}}\mathcal{L}_{\text{OT}} + \lambda_{\text{patchNCE}}\mathcal{L}_{\text{patchNCE}} + \lambda_{\text{freq}}\mathcal{L}_{\text{freq}}, \tag{10}$$

where λ_{OT}, $\lambda_{\text{patchNCE}}$, λ_{freq} are trade-off hyperparameters.

Generation Stage During inference, as shown in Fig. 1(a), we run a multi-step iterative process over the time steps $\{t_k\}_{k=0}^N \subset [0, 1]$. We initialize $x_{t_0} = x_0 \sim \pi_{\text{CBCT}}$, and at each step: we first use the generator G_θ to produce an intermediate CT image $x_1(x_{t_k}) = G_\theta(x_{t_k}, t_k)$, and then apply the IPM to obtain the next state $x_{t_{k+1}}$ by interpolating between x_{t_k} and $x_1(x_{t_k})$, plus noise injection with variance $\sigma_{k+1}^2\mathbf{I}$. After N iterations, we arrive at $x_{t_N} \sim G_\theta(x_{t_N})$.

Table 1. Quantitative comparison on intra-region evaluation.

	H&N → H&N				Chest → Chest			
	SSIM↑	PSNR↑	MAE↓	RMSE↓	SSIM↑	PSNR↑	MAE↓	RMSE↓
Pix2Pix [13]	0.944	31.968	2.178	6.627	0.912	30.789	2.431	7.470
CycleGAN [27]	0.949	33.067	1.962	6.228	0.920	30.853	2.290	7.540
CUT [19]	0.951	32.997	1.978	6.051	0.931	31.381	2.054	7.089
SynDiff [18]	0.917	29.154	3.994	9.568	0.921	28.463	4.342	10.774
ResViT [4]	**0.963**	33.483	1.753	5.716	0.928	31.128	2.041	7.207
UNSB [16]	0.957	33.350	1.885	5.878	0.932	31.187	2.080	7.286
ACSB (Ours)	0.962	**33.791**	**1.655**	**5.569**	**0.933**	**31.457**	**2.018**	**7.021**

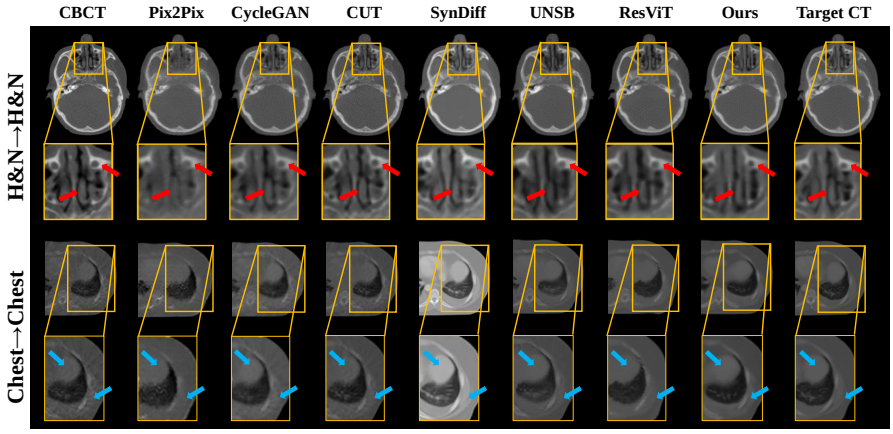

Fig. 2. Visual comparison for Intra-region evaluation by different models.

3 Experiments & Results

3.1 Datasets and Implementation

We evaluate our method on two expert-curated clinical datasets acquired from a tertiary referral center. Each dataset comprises strictly registered paired CBCT-CT 3D scans from distinct anatomical regions: H&N and chest. The 3D scans were segmented into 2D slices along the Z-axis. After quality screening, they were partitioned into a training set (2,232 pairs) and a test set (480 pairs).

To rigorously validate the anatomical conserving adaptability of our ACSB framework, we established a dual-phase evaluation protocol: (1) **Intra-region evaluation.** Models were trained and tested within the same anatomical region (e.g., H&N→ H&N). This setting allows direct measurement of the model's modality capability without interference from anatomical variations. (2) **Cross-region evaluation.** Models trained on one region (e.g., H&N→ Chest) were

Table 2. Quantitative comparison on cross-region evaluation.

	H&N → Chest				Chest → H&N			
	SSIM↑	PSNR↑	MAE↓	RMSE↓	SSIM↑	PSNR↑	MAE↓	RMSE↓
Pix2Pix [13]	0.882	28.147	3.724	10.096	0.898	26.825	4.891	11.735
CycleGAN [27]	0.904	27.108	4.723	12.114	0.945	28.749	3.845	9.875
CUT [19]	0.900	28.408	3.840	10.078	0.943	29.190	3.473	9.415
SynDiff [18]	0.854	22.42	9.528	20.284	0.814	17.848	20.921	33.318
ResViT [4]	0.901	27.963	3.503	10.430	0.897	26.125	4.273	12.988
UNSB [16]	0.897	26.369	5.184	13.080	0.947	29.133	3.030	9.386
ACSB (Ours)	**0.916**	**29.437**	**3.203**	**9.121**	**0.949**	**29.640**	**2.980**	**8.835**

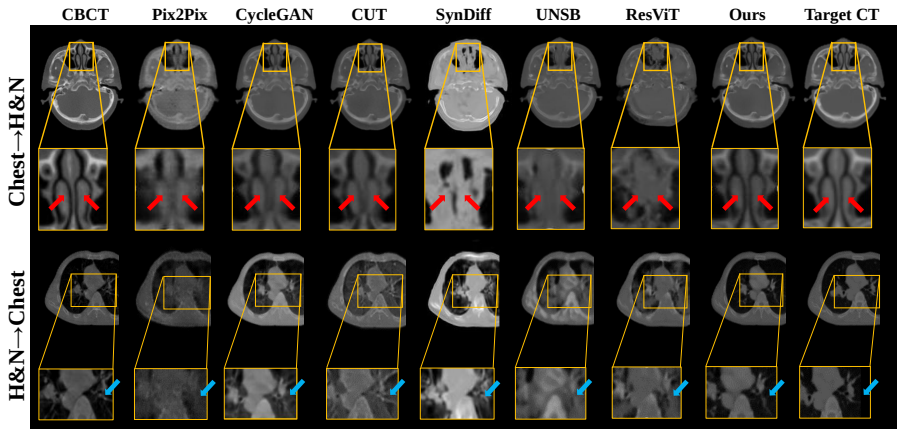

Fig. 3. Visual comparison for Cross-region evaluation by different models.

directly applied on another. This paradigm is to examine the model's ability to generalize across anatomical mismatches while preserving region-agnostic tissue characteristics.

The model is implemented in PyTorch (v2.4.1 with CUDA 12.1) and run on an NVIDIA RTX 3090 GPU (24 GB VRAM). During training, the paired CBCT-CT images are randomly shuffled, with random cropping applied to resize each image to 256×256 pixels. We use the Adam optimizer with an initial learning rate of 2e-4, which is fixed for the first 100 epochs and then linearly decays over the next 100 epochs, for a total of 200 epochs and a batch size of 4.

3.2 Comparison with SOTA

We compare our ACSB with six state-of-the-art synthesis methods spanning three paradigms: CNN[17]-based methods (Pix2Pix [13], CycleGAN [27], CUT

Table 3. Ablation study results for the times of downsampling and L_{freq}.

# of downsampling	L_{freq}	SSIM↑	PSNR↑	MAE↓	RMSE↓
4	×	0.909	29.051	2.862	9.242
2	×	0.961	33.689	1.717	5.592
2	✓	**0.962**	**33.791**	**1.655**	**5.569**

[19]), diffusion-based methods (SynDiff [18], UNSB [16]), and one transformer-based method (ResViT [4]). To ensure consistency, each network of these methods was retrained from publicly available implementations to achieve its best synthesis results. The Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Structural Similarity Index (SSIM) [22] and Peak Signal-to-Noise Ratio (PSNR) [12] are adopted for quantitive evaluation.

Intra-region Evaluation Table 1 presents quantitative results for all methods evaluated under intra-region conditions, clearly demonstrating that our method outperforms other methods. Visual comparisons in Fig. 2 further reveal that our method precisely transfers modality-specific characteristics (note that our method effectively suppresses the scanner-specific artifacts commonly observed in baseline methods as indicated by the blue arrows) while preserving critical anatomical features (red arrows).

Cross-region Evaluation As shown in Table 2, our method establishes new benchmarks in cross-region adaptation with siginficant improvements across all metrics. This breakthrough in generalization capability is visually in Fig. 3, where most of the competing methods exhibit 2 failure patterns: (1) structural disintegration in anatomy level (red arrows), and (2) pathological hallucination of non-existent tissue interfaces (blue arrows). These failures arise from inadequate cross-domain feature disentanglement—a challenge effectively addressed by our AC-ViT generator, demonstrating our method's unique capability to disentangle modality-specific information from anatomical features.

3.3 Ablation Study

We investigate the effectiveness of downsampling times, and frequency-aware loss L_{freq}, as shown in Table 3. It indicates that reducing the downsampling times significantly improved results, underscoring the advantage of retaining higher-resolution features. Furthermore, adding L_{freq} to this 2-downsampling setting yielded additional gains, suggesting that frequency-domain constraints effectively preserve fine structural information and enhance overall equality.

4 Conclusion

In this paper, we propose ACSB, a novel unpaired CBCT-to-CT translation framework that directly bridges arbitrary medical imaging domains via entropy-regularized optimal transport. Unlike methods requiring paired data or Gaussian priors, ACSB learns a stochastic trajectory between CBCT and CT distributions while preserving anatomical fidelity through AC-ViT and frequency-aware loss. The AC-ViT architecture captures multi-scale anatomical features, and the frequency-aware loss enforces alignment of high-frequency components critical for diagnostic details. Comprehensive experiments results on H&N and chest datasets demnostrate the effectiveness and excellent generalization capability.

Acknowledgments. This work is supported by the National Key Research and Development Program of China (2023YFC2705700), the National Natural Science Foundation of China (Grant No. U23A20318, 62272354 and 12005072), the Science and Technology Major Project of Hubei Province (Grant No. 2024BAB046), the Foundation for Innovative Research Groups of Hubei Province (Grant No. 2024AFA017) and the Open Research Fund of Hubeikey Laboratory of Precision Radiation Oncology: No.jzfs014.

Disclosure of Interests. The authors declare that there is no conflict of interest.

References

1. Altalib, A., McGregor, S., Li, C., Perelli, A.: Synthetic ct image generation from cbct: A systematic review. IEEE Transactions on Radiation and Plasma Medical Sciences pp. 1–1 (2025). https://doi.org/10.1109/TRPMS.2025.3533749
2. Brigham, E.O., Morrow, R.E.: The fast fourier transform. IEEE Spectrum **4**(12), 63–70 (1967). https://doi.org/10.1109/MSPEC.1967.5217220
3. Brigham, E.O., Morrow, R.E.: The fast fourier transform. IEEE spectrum **4**(12), 63–70 (1967)
4. Dalmaz, O., Yurt, M., Çukur, T.: Resvit: residual vision transformers for multimodal medical image synthesis. IEEE Transactions on Medical Imaging **41**(10), 2598–2614 (2022)
5. De Bortoli, V., Thornton, J., Heng, J., Doucet, A.: Diffusion schrödinger bridge with applications to score-based generative modeling. Advances in Neural Information Processing Systems **34**, 17695–17709 (2021)
6. De Crevoisier, R., Lafond, C., Mervoyer, A., Hulot, C., Jaksic, N., Bessières, I., Delpon, G.: Image-guided radiotherapy. Cancer/Radiothérapie **26**(1-2), 34–49 (2022)
7. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. Advances in neural information processing systems **34**, 8780–8794 (2021)
8. Gao, L., Xie, K., Sun, J., Lin, T., Sui, J., Yang, G., Ni, X.: Streaking artifact reduction for cbct-based synthetic ct generation in adaptive radiotherapy. Medical Physics p. 879–893 (2023). https://doi.org/10.1002/mp.16017
9. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks. Communications of the ACM **63**(11), 139–144 (2020)

10. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
11. Ho, J., Salimans, T.: Classifier-free diffusion guidance. arXiv preprint arXiv:2207.12598 (2022)
12. Hore, A., Ziou, D.: Image quality metrics: Psnr vs. ssim. In: 2010 20th international conference on pattern recognition. pp. 2366–2369. IEEE (2010)
13. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (July 2017)
14. Jaffray, D.A.: Image-guided radiotherapy: from current concept to future perspectives. Nature reviews Clinical oncology **9**(12), 688–699 (2012)
15. 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)
16. Kim, B., Kwon, G., Kim, K., Ye, J.C.: Unpaired image-to-image translation via neural schrödinger bridge. arXiv preprint arXiv:2305.15086 (2023)
17. O'shea, K., Nash, R.: An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458 (2015)
18. Özbey, M., Dalmaz, O., Dar, S.U., Bedel, H.A., Özturk, Ş., Güngör, A., Çukur, T.: Unsupervised medical image translation with adversarial diffusion models. IEEE Transactions on Medical Imaging **42**(12), 3524–3539 (2023)
19. Park, T., Efros, A.A., Zhang, R., Zhu, J.Y.: Contrastive learning for unpaired image-to-image translation. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16. pp. 319–345. Springer (2020)
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. Sara, U., Akter, M., Uddin, M.S.: Image quality assessment through fsim, ssim, mse and psnr—a comparative study. Journal of Computer and Communications **7**(3), 8–18 (2019)
23. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502 (2020)
24. Tang, Z., Hang, T., Gu, S., Chen, D., Guo, B.: Simplified diffusion schrödinger bridge. arXiv preprint arXiv:2403.14623 (2024)
25. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
26. Walter, C., Schmidt, J., Dula, K., Sculean, A.: Cone beam computed tomography (cbct) for diagnosis and treatment planning in periodontology: A systematic review. Quintessence International,Quintessence International (2016)
27. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)