VarDeFlow: Variational Deformation Learning with Flow Matching for Multi-Organ Cross-Modality Medical Image Synthesis

Peeyush Kumar Singh[1], Inam Ul Haq Gulzar[1], Sneha Singh[1], Aditya Nigam[1], and Pankaj Gupta[2]

[1] School of Computing and Electrical Engineering, Indian Institute of Technology, Mandi, India
[2] Post-Graduate Institute of Medical Education and Research (PGIMER), Chandigarh, India
s24008@students.iitmandi.ac.in

Abstract. Cross-modality medical image synthesis remains fundamentally challenging due to residual spatial inconsistencies from complex non-rigid organ deformations and localized anatomical variability that persist beyond registration. Existing approaches either rely on explicit pre-registration procedures or integrate registration-guided training modules that are not retained during inference, thereby limiting their ability to model residual misalignments and anatomy-specific deformation dynamics. We propose a unified framework that combines deformation learning with flow matching for anatomically consistent multi-organ image synthesis. We introduce a variational deformation network that learns a source-conditioned prior distribution over plausible anatomy-specific deformation fields, capturing the statistics of residual misalignment without requiring paired target inputs at inference, unlike existing registration-guided methods. This integration achieves efficient and stable image generation with substantially fewer integration steps compared to diffusion-based counterparts while preserving anatomical fidelity. The proposed framework is trained and evaluated on a large-scale dataset of 873 patients across five anatomical regions: abdomen, brain, pelvis, head-and-neck, and thorax. Extensive quantitative and qualitative evaluations demonstrate consistent performance in synthesis accuracy across all anatomical regions, while maintaining computational efficiency. Ablation studies further validate the efficacy of this approach, highlighting the individual contributions of learned deformations and flow matching over registration-guided and diffusion-based baselines.

Keywords: Cross-modality synthesis · Deformation learning · Flow matching · Image registration · Multi-organ imaging.

1 Introduction

Medical image synthesis across different imaging modalities has emerged as a critical task in clinical practice and research, enabling the generation of missing

modalities without additional patient scanning [15,23]. Cross-modality synthesis, particularly between computed tomography (CT) to magnetic resonance (MR) imaging, addresses challenges in radiation exposure reduction, scan time optimization, and multi-modal data augmentation for downstream tasks such as segmentation and classification [8,26]. Despite significant progress in deep learning-based synthesis methods [3,24,25], spatial misalignment between source and target images remains a fundamental challenge. While classical deformable registration techniques [1,10] can provide coarse alignment, they rely on intensity-based similarity metrics that are inherently limited for cross-modality registration [16]. This results in persistent residual misalignments, particularly at organ boundaries and regions with fundamentally different tissue contrast properties between CT and MR as shown in Fig. 1. Table 1 reports NCC and Mattes MI across five anatomies following classical Elastix-based registration. Consistently low NCC and Mattes MI scores across all anatomies confirm that residual misalignment persists after registration, with the abdomen exhibiting the lowest Mattes MI, consistent with its higher susceptibility to non-rigid deformations.

Recent approaches have attempted to address this limitation through registration guided synthesis frameworks [7,11,12,27]. However, these methods either employ separate pre-registration pipelines that add computational overhead, or train registration modules that must be discarded at inference time as the registration task requires both moving and fixed references while synthesis task assumes that target is unavailable. Furthermore, most existing synthesis methods are evaluated on very few anatomical regions with limited data, limiting their generalizability and applicability in clinical workflows.

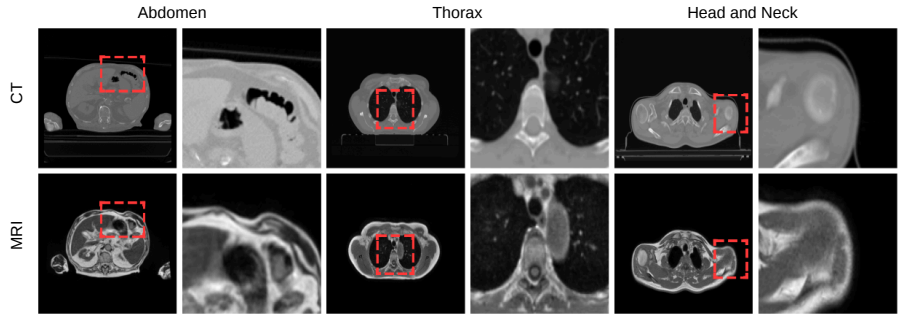

Fig. 1. Residual misalignment after classical Elastix-based registration. Red bounding boxes highlight soft-tissue interfaces where CT and MR exhibit persistent misalignment across Abdomen, Thorax, and Head and Neck anatomies, with zoomed regions revealing structural inconsistencies that challenge cross-modality synthesis.

Advances in generative modeling have introduced diffusion probabilistic models [6,20] and flow matching [13,14] as powerful alternatives to GANs for image

synthesis. While diffusion models have shown impressive results in medical imaging [5,9], their reliance on iterative denoising requires 50-100 sampling steps even with advanced solvers, compared to 1000 steps for standard implementations, limiting practical clinical deployment. Flow matching offers a compelling alternative by learning continuous normalizing flows with optimal transport, achieving superior sample quality with significantly fewer integration steps [13, 18].

Table 1. Registration quality metrics (Mean ± Std) across anatomies after classical Elastix-based registration. Higher NCC and Mattes MI indicate better alignment.

Metric	Brain	Pelvis	Head & Neck	Thorax	Abdomen
NCC	0.398 ± 0.09	0.465 ± 0.11	0.495 ± 0.10	0.580 ± 0.09	0.445 ± 0.09
Mattes MI	0.560 ± 0.04	0.544 ± 0.13	0.464 ± 0.13	0.496 ± 0.13	0.415 ± 0.11

In this work, we propose a unified framework that combines flow matching with learned deformation modeling for multi-organ medical image synthesis. Our approach makes three key contributions: (1) We introduce a variational deformation network that learns a source-conditioned prior over anatomically plausible deformation fields in a data-driven manner, requiring only source images at inference without paired target modalities. (2) We integrate this with a flow matching synthesis model that achieves efficient, high-quality volume generation with reduced sampling requirements compared to diffusion-based alternatives. (3) We train and evaluate on a large-scale dataset of 873 patients across five diverse anatomical regions (abdomen, brain, pelvis, head-and-neck, and thorax), conducting extensive experiments that validate the generalizability of our framework across anatomies with varying deformation characteristics and tissue contrast properties with comprehensive ablation studies confirming the individual contributions of learned deformations and flow matching efficiency.

2 Methodology

Our framework integrates learned deformation modeling with flow matching for efficient multi-organ medical image synthesis. Given a source volume $\mathbf{x}_{\text{source}} \in \mathbb{R}^{D\times H\times W}$, we aim to synthesize a corresponding target volume $\mathbf{x}_{\text{target}} \in \mathbb{R}^{D\times H\times W}$ that is both anatomically accurate and spatially aligned. We decompose this problem into two coupled tasks: (1) learning a variational deformation field $\boldsymbol{\phi}$ that captures anatomy-specific spatial variations, and (2) learning a flow matching model v_θ that maps source images to target images along optimal transport paths. The variational deformation network G_ψ parameterizes a distribution over plausible deformation fields conditioned on the source, while the synthesis network learns continuous normalizing flows that transport samples to the target distribution. These components are jointly trained end-to-end, enabling the model to simultaneously learn alignment correction and cross-modality transla-

tion patterns across multiple anatomical regions. An overview of the proposed framework is illustrated in Fig. 2.

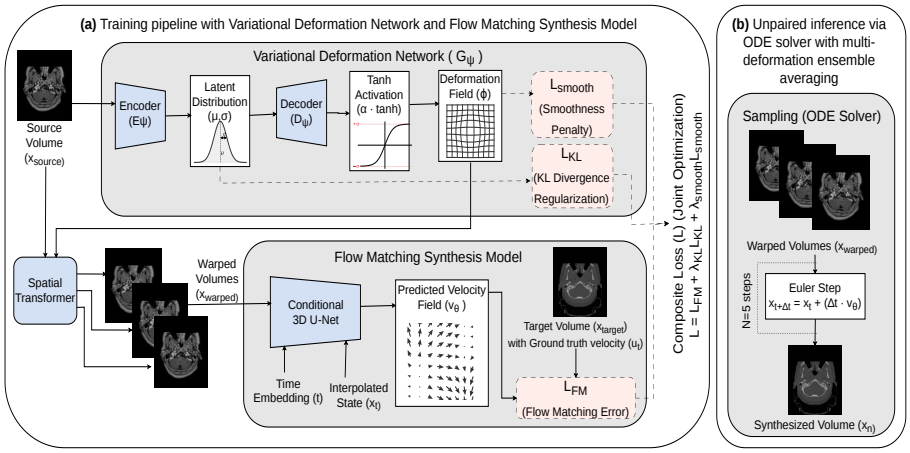

Fig. 2. Overview of the proposed VarDeFlow framework. **(a)** The source volume is warped by N stochastic deformation fields sampled from a Variational Deformation Network G_ψ, regularized via KL divergence and smoothness penalties without paired registration supervision. Each warped volume serves as an informed prior for a conditional 3D U-Net trained via flow matching to predict v_θ. **(b)** At inference, N warped volumes are Euler-integrated in 5 steps and ensemble-averaged.

2.1 Variational Deformation Network

Spatial misalignment persists even after classical registration due to non-rigid organ deformations and local tissue variations. We introduce a variational deformation network G_ψ that learns anatomy-specific spatial transformations in a data-driven manner. Since the target is unavailable at inference, G_ψ learns a distribution over anatomically plausible deformations conditioned on source appearance, effectively marginalizing over likely spatial variations for a given anatomy.

Architecture. G_ψ follows a VAE architecture with encoder E_ψ, stochastic latent bottleneck, and decoder D_ψ. The encoder extracts spatial features $\mathbf{h} = E_\psi(\mathbf{x}_{\text{source}}) \in \mathbb{R}^{d_z}$ via strided 3D convolutions, reducing spatial dimensions to a compact latent code that captures global anatomical structure.

Latent Distribution. The encoded features parameterize a multivariate Gaussian, from which we sample via the reparameterization trick:

$$\boldsymbol{\mu} = f_\mu(\mathbf{h}), \quad \log \boldsymbol{\sigma}^2 = f_\sigma(\mathbf{h}) \tag{1}$$

$$\mathbf{z} = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I}) \tag{2}$$

where f_μ, f_σ are linear projection layers and $\mathbf{z}$ is the sampled latent code. This stochastic sampling produces diverse yet anatomically plausible deformation fields.

Deformation Field Generation. The decoder D_ψ upsamples $\mathbf{z}$ via transposed 3D convolutions to a dense displacement field $\boldsymbol{\phi} \in \mathbb{R}^{3\times D\times H\times W}$, where each voxel (d, h, w) has displacement $(\Delta d, \Delta h, \Delta w)^\top$. Displacement magnitude is constrained via: $\boldsymbol{\phi} = \alpha \cdot \tanh(D_\psi(\mathbf{z}))$, where α is a learned parameter preventing unrealistic tissue folding.

Spatial Transformation. Deformation Field $\boldsymbol{\phi}$ is applied via spatial transformer by constructing a warped sampling grid $\mathbf{G}_{\text{warped}} = \mathbf{G}_{\text{identity}} + \boldsymbol{\phi}$, where $\mathbf{G}_{\text{identity}}$ contains normalized coordinates $[-1, 1]^3$. The warped source is obtained through trilinear interpolation:

$$\mathbf{x}_{\text{warped}} = \mathcal{T}(\mathbf{x}_{\text{source}}, \mathbf{G}_{\text{warped}}) \tag{3}$$

which serves as the aligned input to the flow matching synthesis model.

2.2 Flow Matching Synthesis Model

Given $\mathbf{x}_{\text{warped}}$, we employ flow matching to learn efficient mappings to the target distribution. Unlike diffusion models requiring 50–100 denoising steps, flow matching learns continuous normalizing flows via optimal transport with significantly fewer steps.

Optimal Transport Formulation. Flow matching constructs a time-dependent vector field $v_\theta(\mathbf{x}_t, t, \mathbf{x}_{\text{warped}})$ defining a trajectory from p_0 to p_1. Rather than Gaussian noise, we employ an *informed prior* $\mathbf{x}_0 = \mathbf{x}_{\text{warped}}$: since MR and CT share anatomy and differ only in contrast, this reduces the velocity field to a contrast transformation, enabling synthesis in few steps. The interpolated state at time $t \in [0, 1]$ is:

$$\mathbf{x}_t = (1 - t)\, \mathbf{x}_{\text{warped}} + t\, \mathbf{x}_{\text{target}} \tag{4}$$

Conditional Velocity Field. The velocity field v_θ is parameterized by a 3D U-Net conditioned on $\mathbf{x}_t$, time t, and $\mathbf{x}_{\text{warped}}$ via channel-wise concatenation:

$$v_\theta(\mathbf{x}_t, t, \mathbf{x}_{\text{warped}}) = \text{U-Net}([\mathbf{x}_{\text{warped}}, \mathbf{x}_t], t) \tag{5}$$

Time t is embedded via sinusoidal encodings and injected through adaptive group normalization.

Training Objective. The ground truth velocity is the difference between target and informed prior $\mathbf{u}_t = \mathbf{x}_{\text{target}} - \mathbf{x}_{\text{warped}}$, giving the flow matching loss:

$$\mathcal{L}_{\text{FM}} = \mathbb{E}_{t,\mathbf{x}_0,\mathbf{x}_1}\left[\|v_\theta(\mathbf{x}_t, t, \mathbf{x}_{\text{warped}}) - \mathbf{u}_t\|_2^2\right] \tag{6}$$

Both networks are jointly trained with a composite loss:

$$\mathcal{L} = \mathcal{L}_{\text{FM}} + \lambda_{\text{KL}}\mathcal{L}_{\text{KL}} + \lambda_{\text{smooth}}\mathcal{L}_{\text{smooth}} \tag{7}$$

where $\mathcal{L}_{\text{KL}} = -\frac{1}{2}\sum(1 + \log\boldsymbol{\sigma}^2 - \boldsymbol{\mu}^2 - \boldsymbol{\sigma}^2)$ regularises the deformation latent space and $\mathcal{L}_{\text{smooth}} = \|\nabla\boldsymbol{\phi}\|^2$ prevents tissue folding, with λ_{KL}=0.001 and λ_{smooth}=0.01, established empirically on the validation set.

Sampling Procedure. At inference, starting from $\mathbf{x}_0 = \mathbf{x}_{\text{warped}}$, we integrate the learned ODE using the Euler method over N=5 uniform steps ($\Delta t = 1/N$):

$$\mathbf{x}_{t+\Delta t} = \mathbf{x}_t + \Delta t \cdot v_\theta(\mathbf{x}_t, t, \mathbf{x}_{\text{warped}}) \tag{8}$$

The near-linear transport path of the informed prior makes N=5 sufficient, compared to 10–50 steps with a Gaussian prior and 50–100 for diffusion-based methods.

3 Experiments

Dataset Description. We evaluate on 873 patients from the SynthRAD2023 and SynthRAD2025 Grand Challenge [21, 22], spanning five anatomical regions: abdomen (175), brain (180), pelvis (180), head-and-neck (156), and thorax (182), each with paired CT/MR volumes and anatomical masks. Volumes are resampled to isotropic 1mm^3; CT clipped to $[-1024, 3000]$ HU and MR globally min-max normalized to $[0, 1]$. A stratified 70/10/20% train/val/test split is used per anatomy, with random 96^3 patch training and full-resolution sliding window inference with Gaussian blending in MONAI [2]. All baselines including ResViT and Diffusion models are trained on identical Elastix pre-registered data, receiving the same pre-registered inputs as RegGAN and RegConDIS. All baselines were retrained under identical protocols for fair comparison.

Implementation Details. Implemented in PyTorch [17] on an NVIDIA H200 GPU. The deformation encoder uses five 3D convolutional blocks $[32, 32, 64, 64, 128]$ with stride-2 downsampling to a 128-dim latent code; the decoder mirrors this with transposed convolutions back to 96^3. The synthesis U-Net uses channels $[128, 256, 512, 512]$ with 2 residual blocks per level and attention at resolution 12^3, taking $[\mathbf{x}_{\text{warped}}, \mathbf{x}_t]$ as input and predicting a 1-channel velocity field. Training uses 1M iterations, batch size 4, AdamW (lr=2×10^{-4}, weight decay 10^{-4}), with N=3 deformation ensemble samples and N=5 Euler steps at inference. Results are averaged over five runs.

4 Results

4.1 Quantitative and Qualitative Evaluation

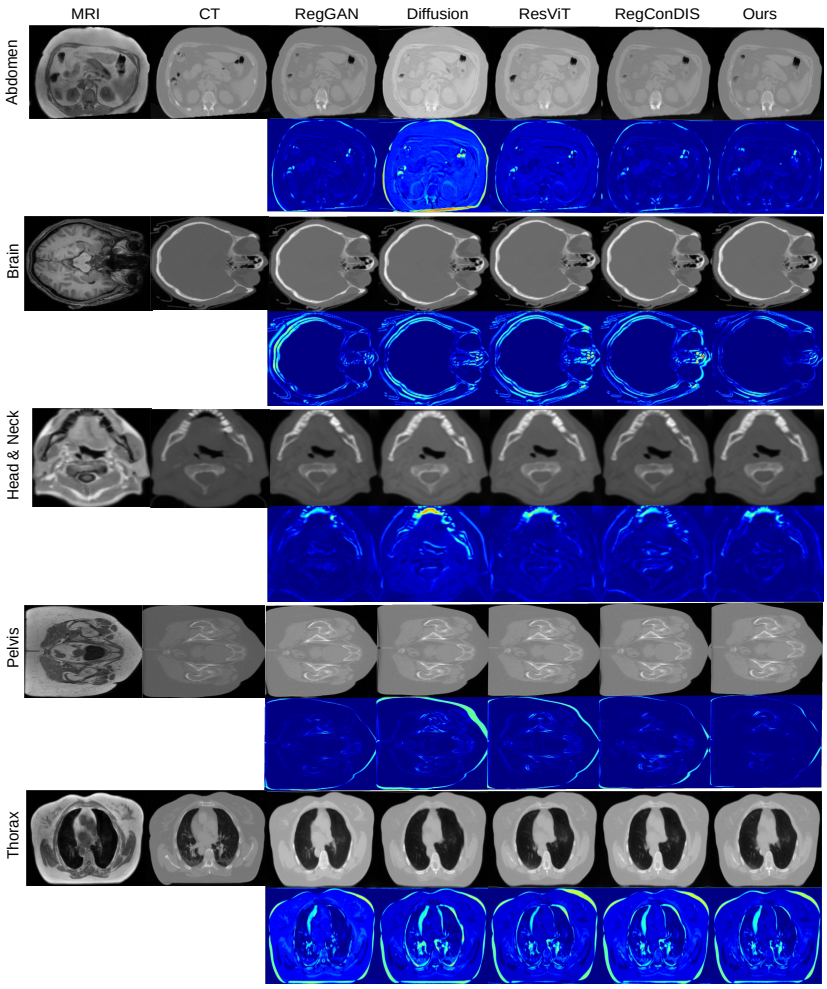

Fig. 3. Representative synthesis results across five anatomical regions. Error maps (blue: low, yellow: high) highlight visually lower reconstruction error for our method.

We evaluate synthesis quality using PSNR, SSIM, and MAE within anatomical mask regions. Table 2 presents per-anatomy results. Brain and pelvis yield the highest quality owing to rigid structural constraints, while abdomen and thorax show slightly lower metrics due to respiratory motion and soft tissue deformability. For context, nnU-Net and SwinUNETR (top SynthRAD leaderboard entries)

Table 3. Ablation study on MRI-to-CT synthesis (on pelvis translation task).

Method	MAE (HU) ↓	PSNR (dB) ↑	SSIM ↑	Steps
DDPM (100 DDIM steps)	50.76	29.28	0.864	100
Flow Matching w/o deformation	50.22	29.76	0.867	5
Flow Matching w/ deformation (N=1)	49.57	29.88	0.885	5
Flow Matching w/ deformation (N=3)	47.76	30.51	0.902	5
Flow Matching w/ deformation (N=5)	47.72	30.49	0.901	5

trained under identical protocols yield MAE/PSNR/SSIM of 60.26/28.74/0.884 and 61.33/28.44/0.878 (brain), 49.96/29.33/0.881 and 50.23/29.11/0.876 (pelvis); our method outperforms both ($p<0.05$). Thorax SSIM clusters at 0.80–0.82 across all methods due to near-zero MR signal in air-filled lung parenchyma. Figure 3 shows representative synthesis results across all anatomies.

Table 2. Quantitative evaluation across five anatomical regions for MRI-to-CT synthesis. All metrics computed within anatomical masks. Best in **bold**. Second best underlined. All improvements over the second-best method (RegConDIS) are statistically significant (Wilcoxon signed-rank test, $p<0.05$).

Anatomy	Method	MAE (HU) ↓	PSNR (dB) ↑	SSIM ↑
Abdomen	RegGAN [11]	106.13 ± 14.91	24.88 ± 1.39	0.820 ± 0.06
	ResViT [4]	107.14 ± 15.23	24.59 ± 1.44	0.818 ± 0.07
	Diffusion [5]	108.38 ± 15.41	24.07 ± 1.81	0.811 ± 0.09
	RegConDIS [12]	105.41 ± 12.12	26.08 ± 1.54	0.823 ± 0.07
	Ours	**103.11 ± 11.13**	**26.54 ± 1.94**	**0.836 ± 0.07**
Brain	RegGAN [11]	61.35 ± 3.93	28.87 ± 1.14	0.881 ± 0.02
	ResViT [4]	62.58 ± 4.08	28.34 ± 1.28	0.877 ± 0.02
	Diffusion [5]	62.65 ± 4.14	28.15 ± 1.34	0.873 ± 0.02
	RegConDIS [12]	59.17 ± 3.15	29.03 ± 1.12	0.889 ± 0.02
	Ours	**58.48 ± 3.45**	**29.37 ± 1.08**	**0.896 ± 0.01**
Pelvis	RegGAN [11]	51.45 ± 3.23	29.42 ± 0.78	0.861 ± 0.02
	ResViT [4]	51.22 ± 3.16	29.87 ± 0.69	0.845 ± 0.03
	Diffusion [5]	50.76 ± 3.93	29.28 ± 0.75	0.864 ± 0.03
	RegConDIS [12]	49.51 ± 4.12	29.88 ± 0.77	0.889 ± 0.02
	Ours	**47.76 ± 3.91**	**30.51 ± 1.12**	**0.902 ± 0.03**
Head&Neck	RegGAN [11]	91.84 ± 14.43	26.34 ± 2.32	0.855 ± 0.03
	ResViT [4]	92.28 ± 14.66	26.11 ± 2.61	0.848 ± 0.04
	Diffusion [5]	91.44 ± 14.18	26.90 ± 2.55	0.857 ± 0.03
	RegConDIS [12]	90.51 ± 15.33	27.18 ± 2.65	0.864 ± 0.02
	Ours	**88.12 ± 14.03**	**28.33 ± 2.11**	**0.882 ± 0.03**
Thorax	RegGAN [11]	116.12 ± 18.80	24.81 ± 2.23	0.809 ± 0.07
	ResViT [4]	117.01 ± 19.18	24.65 ± 2.52	0.803 ± 0.08
	Diffusion [5]	118.89 ± 20.02	24.21 ± 2.76	0.801 ± 0.10
	RegConDIS [12]	114.11 ± 18.22	25.58 ± 2.01	0.814 ± 0.09
	Ours	**113.65 ± 18.13**	**26.23 ± 1.89**	**0.822 ± 0.09**

4.2 Ablation Studies.

Effect of Learned Deformations and Flow Matching. We compare our full model against two baselines. First, we ablate learned deformations by setting $N = 0$ during both training and inference; results show that learned deformations improve translation. Second, we compare our flow matching approach against a diffusion-based baseline (DDPM [6] with 100 DDIM [19] sampling steps). Flow matching achieves comparable quality while requiring significantly fewer sampling steps (5 vs. 100).
Deformation Field Analysis. The learned scalar α (Sec. 2.1) converges to 0.03–0.05 across anatomies, constraining maximum displacement to 2.9–4.8 mm for 96^3 patches at $1\,\mathrm{mm}^3$ resolution consistent with the residual misalignment scale in Table 1. Abdomen yields the highest α ($\approx$0.05), reflecting greater non-rigid deformability, while brain yields the lowest ($\approx$0.03), consistent with its rigid structure.

5 Conclusion

We presented VarDeFlow, a unified framework for multi-organ medical image synthesis combining variational deformation learning with flow matching to address residual spatial misalignment without requiring paired registration inputs at inference. Extensive experiments across five anatomical regions validate our design, with ablation studies confirming the individual contributions of learned deformations and flow-based synthesis. Limitations include reduced performance in high soft-tissue deformability regions and patch-based training that may miss long-range anatomical dependencies. Future work may explore bidirectional synthesis, segmentation-mask conditioning for low-contrast regions, downstream anatomical plausibility assessment via DSC using pre-trained segmentation models and extensive clinical validation with radiologists.

Acknowledgements. This work was supported by ICMR (Grant ID: FIW-2024-01-00000151) Project No: IITM/ICMR/SS/537, IIT Mandi.

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

References

1. Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C.: A reproducible evaluation of ants similarity metric performance in brain image registration. Neuroimage **54**(3), 2033–2044 (2011)
2. 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)

3. Chartsias, A., Joyce, T., Giuffrida, M.V., Tsaftaris, S.A.: Multimodal mr synthesis via modality-invariant latent representation. IEEE transactions on medical imaging **37**(3), 803–814 (2017)
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. Dorjsembe, Z., Pao, H.K., Odonchimed, S., Xiao, F.: Conditional diffusion models for semantic 3d brain mri synthesis. IEEE Journal of Biomedical and Health Informatics **28**(7), 4084–4093 (2024)
6. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
7. Honkamaa, J., Khan, U., Koivukoski, S., Valkonen, M., Latonen, L., Ruusuvuori, P., Marttinen, P.: Deformation equivariant cross-modality image synthesis with paired non-aligned training data. Medical Image Analysis **90**, 102940 (2023)
8. Huo, Y., Xu, Z., Moon, H., Bao, S., Assad, A., Moyo, T.K., Savona, M.R., Abramson, R.G., Landman, B.A.: Synseg-net: Synthetic segmentation without target modality ground truth. IEEE transactions on medical imaging **38**(4), 1016–1025 (2018)
9. Kazerouni, A., Aghdam, E.K., Heidari, M., Azad, R., Fayyaz, M., Hacihaliloglu, I., Merhof, D.: Diffusion models for medical image analysis: A comprehensive survey. arXiv preprint arXiv:2211.07804 (2022)
10. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.: Elastix: a toolbox for intensity-based medical image registration. IEEE transactions on medical imaging **29**(1), 196–205 (2009)
11. Kong, L., Lian, C., Huang, D., Hu, Y., Zhou, Q., et al.: Breaking the dilemma of medical image-to-image translation. Advances in Neural Information Processing Systems **34**, 1964–1978 (2021)
12. Li, C., Chen, Z., Zhang, Y., Zhong, L., Yang, W.: Boosting medical image synthesis via registration-guided consistency and disentanglement learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 78–88. Springer (2025)
13. Lipman, Y., Chen, R.T., Ben-Hamu, H., Nickel, M., Le, M.: Flow matching for generative modeling. arXiv preprint arXiv:2210.02747 (2022)
14. 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)
15. Nie, D., Trullo, R., Lian, J., Petitjean, C., Ruan, S., Wang, Q., Shen, D.: Medical image synthesis with context-aware generative adversarial networks. In: International conference on medical image computing and computer-assisted intervention. pp. 417–425. Springer (2017)
16. Oliveira, F.P., Tavares, J.M.R.: Medical image registration: a review. Computer methods in biomechanics and biomedical engineering **17**(2), 73–93 (2014)
17. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems **32** (2019)
18. Pooladian, A.A., Ben-Hamu, H., Domingo-Enrich, C., Amos, B., Lipman, Y., Chen, R.T.: Multisample flow matching: Straightening flows with minibatch couplings. arXiv preprint arXiv:2304.14772 (2023)
19. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502 (2020)

20. Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. arXiv preprint arXiv:2011.13456 (2020)
21. Thummerer, A., van der Bijl, E., Galapon, A.J., Kamp, F., Savenije, M., Muijs, C., Aluwini, S., Steenbakkers, R.J., Beuel, S., Intven, M.P., et al.: Synthrad2025 grand challenge dataset: Generating synthetic cts for radiotherapy from head to abdomen. Medical physics **52**(7), e17981 (2025)
22. Thummerer, A., Van der Bijl, E., Galapon Jr, A., Verhoeff, J.J., Langendijk, J.A., Both, S., van den Berg, C.N.A., Maspero, M.: Synthrad2023 grand challenge dataset: Generating synthetic ct for radiotherapy. Medical physics **50**(7), 4664–4674 (2023)
23. Wolterink, J.M., Dinkla, A.M., Savenije, M.H., Seevinck, P.R., Van Den Berg, C.A., Išgum, I.: Deep mr to ct synthesis using unpaired data. In: International workshop on simulation and synthesis in medical imaging. pp. 14–23. Springer (2017)
24. Yi, X., Walia, E., Babyn, P.: Generative adversarial network in medical imaging: A review. Medical image analysis **58**, 101552 (2019)
25. Yu, B., Zhou, L., Wang, L., Shi, Y., Fripp, J., Bourgeat, P.: Ea-gans: edge-aware generative adversarial networks for cross-modality mr image synthesis. IEEE transactions on medical imaging **38**(7), 1750–1762 (2019)
26. Zhao, A., Balakrishnan, G., Durand, F., Guttag, J.V., Dalca, A.V.: Data augmentation using learned transformations for one-shot medical image segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8543–8553 (2019)
27. Zhong, L., Chen, Z., Shu, H., Zheng, Y., Zhang, Y., Wu, Y., Feng, Q., Li, Y., Yang, W.: Qacl: Quartet attention aware closed-loop learning for abdominal mr-to-ct synthesis via simultaneous registration. Medical image analysis **83**, 102692 (2023)