LEARNING SHARED BIDIRECTIONAL LATENT DYNAMICS FOR TRANSLATION BETWEEN CT AND MRI

Anonymous authors
Paper under double-blind review

ABSTRACT

Bidirectional image translation is an important problem in computer vision because many related visual domains require mappings in both directions. A representative case is translation between CT and MRI, which provide complementary anatomical and tissue information but may not always be simultaneously available. Existing methods commonly train separate models for translation from CT to MRI and from MRI to CT, treating the two directions as independent tasks and leaving their shared structure insufficiently explored. To address this limitation, we formulate the two translation directions as latent state evolution processes governed by shared dynamics. Modality specific VAEs encode CT and MRI into latent representations, and a shared velocity model transports the source representation toward the target. Since the same dynamics model must represent two different translation directions, we introduce latent anchors that combine the source state with direction information and guide velocity prediction. The anchor is used as a reference during latent evolution rather than as the predicted target itself. An image space refiner further corrects residual errors after decoding. Experiments on two public paired CT and MRI datasets demonstrate consistent performance across both translation directions and MRI contrasts, with the proposed framework achieving the highest mean PSNR and SSIM among the compared methods. Ablation studies further show that direct parameter sharing introduces a performance gap, while explicit direction modeling, particularly the proposed anchor formulation, largely recovers this loss.

1 INTRODUCTION

Cross modal image translation aims to transform images between imaging modalities while preserving corresponding anatomical structures Chen et al. (2025). In medical imaging, distinct acquisition mechanisms provide complementary yet heterogeneous observations of the same anatomy Hermessi et al. (2021); Huang et al. (2020). CT and MRI, for example, exhibit strong anatomical correspondence but differ substantially in tissue contrast, intensity distribution, and texture representation Ni et al. (2026); Liu (2024); Chen et al. (2025). In acute ischemic stroke assessment, imaging strategies based on CT and MRI can affect lesion identification and clinical decision making Rapillo et al. (2024). When one modality is unavailable or incomplete, generating the target modality from the available one may provide complementary imaging information for subsequent medical image analysis Jang et al. (2026); Feng et al. (2022); Huijben et al. (2024), creating demands for both translation from CT to MRI and translation from MRI to CT.

Many translation methods between CT and MRI treat translation from CT to MRI and translation from MRI to CT as independent source to target mappings and train separate models for the two directions Ni et al. (2026); Liu (2024); Peng et al. (2023); Chen et al. (2025). As illustrated in Fig. 1, separate modeling uses different models for the two directions, whereas direct parameter sharing uses one parameter set for both. Unified models have also been shown to support bidirectional image translation Xue et al. (2025). From a latent representation perspective, both directions can be modeled as evolution from a source modality state to a target modality state using continuous time dynamical systems and generative models based on velocity fields Lipman et al. (2023); Liu et al. (2022); Chen et al. (2018). Accordingly, we propose a shared latent dynamics paradigm for bidirectional translation between CT and MRI, using the same dynamics parameters to model latent state evolution in both directions. Modality specific variational autoencoders learn the latent

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

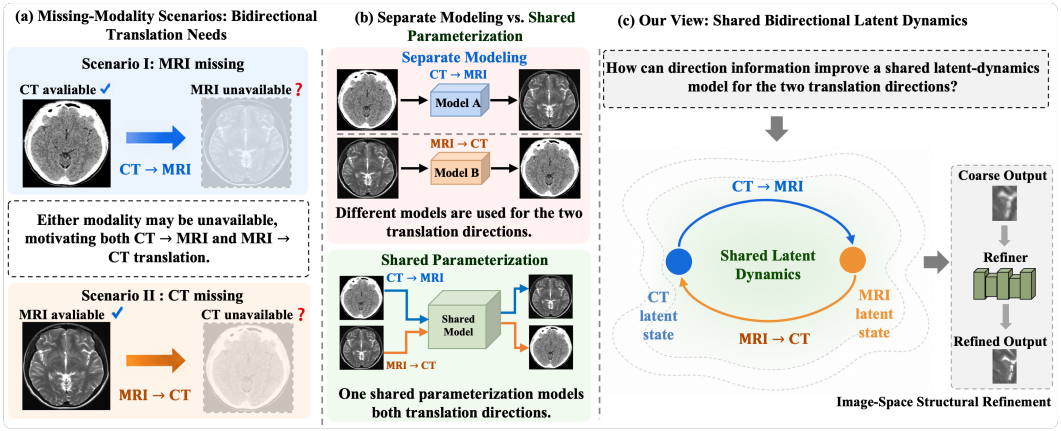

Figure 1: Motivation for unified bidirectional translation between CT and MRI. Different missing modality scenarios create demands for translation from CT to MRI and from MRI to CT. We model both directions as conditional state evolution processes within shared latent dynamics and study how direction information improves bidirectional modeling.

representations of CT and MRI Kingma & Welling (2013), while a time dependent velocity field and finite step integration progressively transport the source latent state toward the target latent state. Because the same core dynamics parameters model both conditional translation processes, we further study how translation direction information can be incorporated into shared latent dynamics. We design multiple direction modeling schemes and propose direction aware latent anchors, which jointly encode the translation direction and source latent state into explicit references for shared velocity field prediction. We also introduce image space refinement to improve local boundaries and fine grained structures.

Our main contributions are as follows:

- We formulate bidirectional translation between CT and MRI as two latent state evolution processes governed by one shared dynamics model, allowing both directions to be learned within a unified framework rather than with separate translation models.
- We study how translation direction should be represented within shared dynamics and propose Anchor BiFlow, which uses a latent anchor conditioned on the source state and translation direction to guide the shared velocity field without directly predicting the target latent representation.
- We validate the proposed framework on two public paired CT and MRI datasets. It achieves the highest mean PSNR and SSIM among the compared methods across all evaluated translation settings, while ablation studies show that explicit direction modeling, particularly Anchor BiFlow, largely recovers the performance loss caused by direct parameter sharing.

2 Related Work

2.1 Medical Image Translation Across Modalities

Medical image translation across modalities aims to synthesize one imaging modality from another while preserving anatomical correspondence Chen et al. (2025). Existing studies have considered both paired and unpaired settings, using adversarial learning, cycle consistency, shared latent representations, contrastive learning, and diffusion based generative models Tomar et al. (2021); Amini Amirkolaee et al. (2022); de Bel et al. (2021); Eslami et al. (2020); Luo et al. (2022); Nie et al. (2018); Isola et al. (2017); Zhu et al. (2017); Liu et al. (2017); Huang et al. (2018); Park et al. (2020). For CT and MRI translation, prior work has studied both translation from CT to MRI Ni et al. (2026); Liu (2024); Choo et al. (2024); Feng et al. (2022) and translation from MRI to CT Peng et al. (2023); Huijben et al. (2024) using adversarial, diffusion, and Brownian bridge formulations Wei et al. (2025); Jang et al. (2026); Özbey et al. (2023); Choo et al. (2024); Li et al. (2023). Most

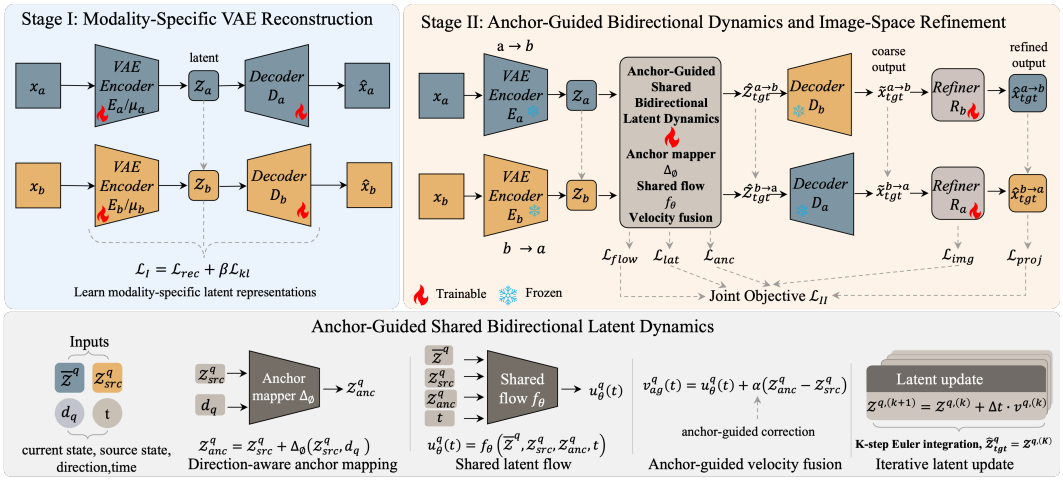

Figure 2: Overall framework. Stage I learns modality specific VAE latents. Stage II freezes the VAEs and trains shared bidirectional latent dynamics with anchor guidance and image space refinement. The bottom panel shows velocity prediction guided by the anchor and the latent update over K steps.

of these methods are designed for a specified source to target direction and therefore learn the two directions separately. Bidirectional translation has also been explored through reversible architectures, shared invertible representations, Brownian bridge diffusion, and latent mappings, including RevGAN, AlignFlow, BiBBDM, and VAE based approaches van der Ouderaa & Worrall (2019); Grover et al. (2020); Xue et al. (2025); Liang et al. (2025). In contrast, we focus on modeling both translation directions as latent state evolution processes governed by one shared dynamics model and on representing their directional differences within the shared dynamics.

2.2 Latent Dynamics Modeling for Bidirectional Translation

Diffusion models describe generation as progressive or stochastic state evolution Ho et al. (2020); Song et al. (2021b), and latent diffusion performs this process in a compressed latent space Rombach et al. (2022). Conditional diffusion and bridge formulations have been developed for image to image translation, including Brownian bridge, denoising diffusion bridge, and Schrödinger bridge models Saharia et al. (2022); Li et al. (2023); Wang et al. (2024); Zhou et al. (2024); Liu et al. (2023). In medical imaging, recent studies have explored adversarial diffusion, recursive diffusion bridges, slice consistent Brownian bridge models, and three dimensional diffusion bridge translation for CT and MRI synthesis Özbey et al. (2023); Choo et al. (2024); Arslan et al. (2025); Zhang et al. (2025). Flow matching and rectified flow instead learn time dependent velocity fields for continuous state transport Lipman et al. (2023); Liu et al. (2022), following a formulation related to Neural ODEs Chen et al. (2018). Inspired by this view, we formulate translation from CT to MRI and translation from MRI to CT as two conditional state evolution processes within shared latent dynamics. We further investigate how translation direction information should be incorporated into the shared dynamics to balance the two directions.

3 Methods

3.1 Problem Formulation and Shared Latent Dynamics

Given paired samples (x_a, x_b) from modality domains $\mathcal{X}_a$ and $\mathcal{X}_b$, bidirectional translation across modalities learns two ordered mappings while preserving anatomical correspondence. As shown in Fig. 2, we first train modality specific variational autoencoders and use their posterior means as deterministic latent representations:

$$z_a = \mu_a(x_a), \qquad z_b = \mu_b(x_b). \tag{1}$$

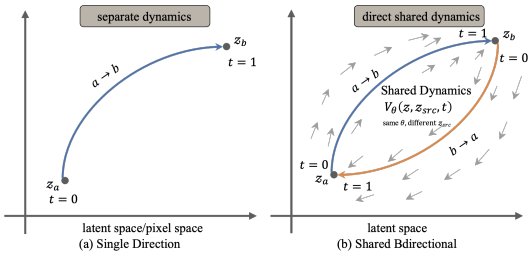

Figure 3: Separate dynamics use independently parameterized velocity fields for the two directions, whereas direct shared dynamics use one source conditioned field $v_\theta^{\mathrm{sh}}(z(t), z_{\mathrm{src}}, t)$.

The Stage I objective for modality specific VAE reconstruction is

$$\mathcal{L}_{\mathrm{I}} = \mathcal{L}_{\mathrm{vae}} = \sum_{m \in \{a,b\}} \left(\mathcal{L}_{\mathrm{rec}}^m + \beta \mathcal{L}_{\mathrm{kl}}^m\right), \tag{2}$$

where $\mathcal{L}_{\mathrm{rec}}^m$ and $\mathcal{L}_{\mathrm{kl}}^m$ denote the reconstruction and KL divergence terms for modality m, respectively. After Stage I, the modality specific encoders and decoders are frozen for latent dynamics learning.

For either direction, let z_{src} and z_{tgt} denote the ordered source and target latent representations, with $(z_{\mathrm{src}}, z_{\mathrm{tgt}}) \in \{(z_a, z_b), (z_b, z_a)\}$. To model both ordered processes with one dynamics parameterization, we define the direct shared latent dynamics as

$$\dot{z}(t) = v_\theta^{\mathrm{sh}}\big(z(t), z_{\mathrm{src}}, t\big), \qquad z(0) = z_{\mathrm{src}}, \tag{3}$$

where $t \in [0, 1]$ and the same dynamics parameters θ are used for both translation directions. Numerical integration maps z_{src} to the predicted target latent endpoint $\widehat{z}_{\mathrm{tgt}}$.

As illustrated in Fig. 3, the direct shared model contains no explicit direction variable. Since each source latent representation is produced by its corresponding modality specific encoder, z_{src} provides an implicit cue about the ordered source to target process. This direct shared formulation provides the basic bidirectional model, and we next study how translation direction can be modeled within shared latent dynamics.

3.2 Direction Modeling in Shared Latent Dynamics

The direct shared dynamics in Eq. 3 uses one source conditioned velocity field for the two ordered mappings. Let $q \in \{a \to b, b \to a\}$ denote an ordered translation process, with source and target latent states $(z_{\mathrm{src}}^q, z_{\mathrm{tgt}}^q)$. For flow supervision, we construct

$$\begin{aligned} z_t^q &= (1-t) z_{\mathrm{src}}^q + t z_{\mathrm{tgt}}^q, \\ v_q^* &= z_{\mathrm{tgt}}^q - z_{\mathrm{src}}^q. \end{aligned} \tag{4}$$

We use $\bar{z}^q$ to denote the latent state at which the velocity is evaluated. During flow supervision, $\bar{z}^q = z_t^q$; during numerical integration, $\bar{z}^q = z^{q,(k)}$.

To analyze the role of direction modeling, let $s^q = (\bar{z}^q, z_{\mathrm{src}}^q, t)$ denote the source conditioned input of the velocity predictor. Independent unidirectional modeling assigns separate predictors to the two ordered processes, whereas Source Conditioned BiFlow uses one parameterization for both directions. Under the same base velocity family, their optimal risks satisfy

$$\begin{aligned} \mathcal{R}_{\mathrm{uni}}^\star &= \inf_{\theta_{ab}, \theta_{ba}} \sum_q \mathbb{E}\left[\left\| h_{\theta_q}(s^q) - v_q^* \right\|_2^2\right], \\ \mathcal{R}_{\mathrm{sc}}^\star &= \inf_{\theta} \sum_q \mathbb{E}\left[\left\| h_{\theta}(s^q) - v_q^* \right\|_2^2\right], \\ \mathcal{R}_{\mathrm{uni}}^\star &\le \mathcal{R}_{\mathrm{sc}}^\star, \end{aligned} \tag{5}$$

where $\sum_q$ is over $q \in \{a \to b, b \to a\}$, and θ_q denotes θ_{ab} or θ_{ba} according to the ordered process q. The inequality holds because the independent model can choose the same predictor for

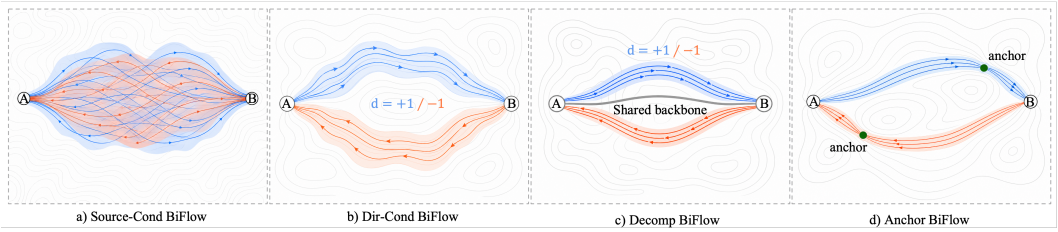

Figure 4: Direction modeling variants in shared bidirectional latent dynamics: (a) Source Conditioned BiFlow (**Source-Cond BiFlow**), (b) Direction Conditioned BiFlow (**Dir-Cond BiFlow**), (c) Direction Decomposed BiFlow (**Decomp BiFlow**), and (d) Anchor BiFlow.

both directions as a special case. Thus, direct shared dynamics impose a stronger parameterization constraint: although z^q_{src} provides an implicit direction cue, one source conditioned velocity field must fit the target velocities of both ordered processes.

For shared dynamics with direction modeling, we use $d_q = +1$ for $q = a \to b$ and $d_q = -1$ for $q = b \to a$. Let $\mathcal{H}_{\mathrm{sc}}$ denote the source conditioned shared family induced by $h_\theta(s^q)$, and let $\mathcal{H}\mathrm{dm}$ denote a shared velocity family with explicit direction modeling that contains $\mathcal{H}\mathrm{sc}$ as a special case. Its optimal risk is

$$\begin{aligned} \mathcal{R}^\star_{\mathrm{dm}} &= \inf_{g \in \mathcal{H}_{\mathrm{dm}}} \sum_q \mathbb{E}\left[\left\| g(s^q, d_q) - v^*_q \right\|_2^2 \right], \\ \mathcal{R}^\star_{\mathrm{dm}} &\le \mathcal{R}^\star_{\mathrm{sc}}, \qquad \mathcal{H}_{\mathrm{sc}} \subseteq \mathcal{H}_{\mathrm{dm}}. \end{aligned} \tag{6}$$

This motivates us to study how direction information should enter shared latent dynamics so as to reduce the approximation constraint associated with direction while retaining a unified bidirectional model.

As shown in Fig. 4, we instantiate the above direction modeling analysis with four BiFlow variants. These variants examine progressively more explicit ways of representing translation direction in shared latent dynamics.

Source Conditioned BiFlow serves as the direct shared formulation. It uses the source latent state as the only cue associated with direction, allowing the shared velocity field to infer the ordered process from the source representation:

$$v^q_{\mathrm{sc}}(t) = v^{\mathrm{sc}}_\theta\big(\bar{z}^q, z^q_{\mathrm{src}}, t\big). \tag{7}$$

This formulation keeps both directions within one compact velocity parameterization. Its direction representation, however, remains implicit in z^q_{src}, so the same source conditioned mapping must infer the ordered process while fitting the target velocities of both directions.

Direction Conditioned BiFlow makes the ordered process explicit by feeding the direction code into the velocity predictor:

$$v^q_{\mathrm{dc}}(t) = v^{\mathrm{dc}}_\theta\big(\bar{z}^q, z^q_{\mathrm{src}}, t, d_q\big). \tag{8}$$

This design gives the shared model direct access to the translation direction with minimal architectural change. Since the direction code is still used as an input condition, the change of the velocity field associated with direction is not explicitly separated.

Direction Decomposed BiFlow therefore moves direction modeling from input conditioning to velocity decomposition. It represents the velocity as a shared component plus a correction associated with direction:

$$\begin{aligned} v^q_{\mathrm{dd}}(t) = v^{\mathrm{com}}_\theta\big(\bar{z}^q, z^q_{\mathrm{src}}, t\big) \\ + c_\psi\big(\bar{z}^q, z^q_{\mathrm{src}}, t, d_q\big), \end{aligned} \tag{9}$$

where v^{com}_θ models the shared state evolution component, and c_ψ provides the correction associated with direction. This formulation retains a common velocity component while allowing velocity adjustment associated with direction. The correction term is still defined at the velocity level and does not provide an explicit latent reference associated with the source state and translation direction.

Algorithm 1 Anchor Guided Shared Bidirectional Latent Evolution

Require: Source latent state z^q_{src}, direction code $d_q \in \{+1, -1\}$, anchor offset mapper Δ_ϕ, shared flow predictor f_θ, anchor weight α, integration step number K
Ensure: Predicted target latent state $\widehat{z}^q_{\mathrm{tgt}}$

$z^q_{\mathrm{anc}} \leftarrow z^q_{\mathrm{src}} + \Delta_\phi(z^q_{\mathrm{src}}, d_q)$
$z^{q,(0)} \leftarrow z^q_{\mathrm{src}}, \quad \Delta t \leftarrow 1/K$
for $k = 0, 1, \ldots, K-1$ **do**
  $t_k \leftarrow k\Delta t$
  $\bar{z}^q \leftarrow z^{q,(k)}$
  $u^{q,(k)}_\theta \leftarrow f_\theta(\bar{z}^q, z^q_{\mathrm{src}}, z^q_{\mathrm{anc}}, t_k)$
  $v^{q,(k)}_{\mathrm{ag}} \leftarrow u^{q,(k)}_\theta + \alpha\,(z^q_{\mathrm{anc}} - z^q_{\mathrm{src}})$
  $z^{q,(k+1)} \leftarrow z^{q,(k)} + \Delta t\, v^{q,(k)}_{\mathrm{ag}}$
end for
$\widehat{z}^q_{\mathrm{tgt}} \leftarrow z^{q,(K)}$
return $\widehat{z}^q_{\mathrm{tgt}}$

To provide such a reference, the proposed Anchor BiFlow represents direction at the latent reference level. It constructs a source dependent anchor with direction awareness:

$$\begin{aligned} z^q_{\mathrm{anc}} &= z^q_{\mathrm{src}} + \Delta_\phi(z^q_{\mathrm{src}}, d_q), \\ u^q_\theta(t) &= f_\theta\big(\bar{z}^q, z^q_{\mathrm{src}}, z^q_{\mathrm{anc}}, t\big), \\ v^q_{\mathrm{ag}}(t) &= u^q_\theta(t) + \alpha\big(z^q_{\mathrm{anc}} - z^q_{\mathrm{src}}\big), \end{aligned} \tag{10}$$

where z^q_{anc} provides an explicit latent reference jointly determined by the source state and translation direction, and α controls the correction induced by the anchor. The anchor is supervised using the target latent state and serves as a directional reference conditioned on the source state for shared velocity prediction. Unlike direct target latent prediction or latent residual learning, the anchor is not used as the final translated representation. Instead, it provides a reference that explicitly encodes the translation direction and guides the shared velocity field during latent evolution. The predicted target latent endpoint is obtained through K-step integration of the velocity field guided by the anchor. Algorithm 1 summarizes how the proposed model uses the anchor to guide latent evolution. Other BiFlow variants follow the same Euler procedure after replacing $v^{q,(k)}_{\mathrm{ag}}$ with the corresponding selected velocity. The transported endpoint is then passed to the target modality decoder.

3.3 Image Space Refinement and Training Objective

The predicted endpoint $\widehat{z}^q_{\mathrm{tgt}}$ is decoded by the target modality decoder:

$$\widetilde{x}^q_{\mathrm{tgt}} = D_{\mathrm{tgt}}\big(\widehat{z}^q_{\mathrm{tgt}}\big). \tag{11}$$

Because the latent dynamics operate on compressed VAE codes, the decoded output is treated as a coarse target modality approximation in our framework. We reduce the remaining image domain residuals with a target domain refiner:

$$\widehat{x}^q_{\mathrm{tgt}} = R_{\mathrm{tgt}}\big(\widetilde{x}^q_{\mathrm{tgt}}\big), \tag{12}$$

where D_{tgt} and R_{tgt} are selected according to the target modality.

For the proposed Anchor BiFlow, the latent dynamics are supervised by the flow matching loss and the endpoint latent loss, where $v^q(t)$ denotes the predicted anchor guided velocity:

$$\begin{aligned} \mathcal{L}_{\mathrm{flow}} &= \mathbb{E}_{q,t}\left[\left\|v^q(t) - v^*_q\right\|_2^2\right], \\ \mathcal{L}_{\mathrm{lat}} &= \mathbb{E}_q\left[\left\|\widehat{z}^q_{\mathrm{tgt}} - z^q_{\mathrm{tgt}}\right\|_1\right]. \end{aligned} \tag{13}$$

The anchor is supervised as a latent reference:

$$\mathcal{L}_{\mathrm{anc}} = \mathbb{E}_q\left[\left\|z^q_{\mathrm{anc}} - z^q_{\mathrm{tgt}}\right\|_1\right]. \tag{14}$$

Table 1: Quantitative comparison on SynthRAD2023 and PCM2025.

Method	Training	Params (M)	FLOPs (G)	SynthRAD2023				PCM2025 T1				PCM2025 T2			
				CT→MRI		MRI→CT		CT→MRI		MRI→CT		CT→MRI		MRI→CT	
				PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
ResCycleGAN	Separate	11.37×2	97.49	17.59	0.7113	17.20	0.7330	14.33	0.3321	10.71	0.2622	12.40	0.2800	9.13	0.2570
Pix2pix		54.41×2	11.90	17.27	0.6960	17.31	0.7326	16.91	0.4624	12.74	0.2913	15.69	0.3504	12.16	0.2755
MedGAN		31.04×2	96.22	17.94	0.6757	18.59	0.6971	16.53	0.2895	10.12	0.2221	12.52	0.1990	11.04	0.1153
ARGAN		45.25×2	59.79	18.11	0.7337	18.36	0.7738	<u>19.31</u>	<u>0.5049</u>	<u>13.34</u>	0.3119	<u>16.17</u>	<u>0.3945</u>	<u>12.83</u>	0.3116
HisGAN		4.13×2	**7.27**	14.32	0.6425	19.43	<u>0.8044</u>	16.69	0.4189	11.66	0.2399	14.39	0.2935	11.75	0.2519
DDIM		5.21×2	97.99×50	16.53	0.6158	10.98	0.1054	11.71	0.0622	7.29	0.0103	12.04	0.2023	8.17	0.1211
BBDM		27.69×2	316.11×50	17.50	0.7160	17.75	0.7599	16.37	0.4215	10.81	<u>0.3419</u>	14.26	0.3773	12.03	<u>0.3568</u>
LDM		6.79×2	7.06×50	16.79	0.7033	17.44	0.7576	13.42	0.1836	8.95	0.1494	14.37	0.1797	8.18	0.1260
Rectified Flow		5.21×2	97.99×50	<u>19.13</u>	<u>0.7566</u>	<u>19.67</u>	0.7827	16.12	0.3514	11.82	0.2909	15.46	0.3314	10.62	0.2903
BiBBDM	Joint	27.69	316.19×50	17.60	0.7166	17.65	0.7510	16.82	0.4258	12.51	0.3418	14.55	0.3757	12.30	0.3448
Proposed		**4.15**	16.36×5	**19.47**	**0.7794**	**19.69**	**0.8068**	**19.43**	**0.5092**	**14.94**	**0.4023**	**17.60**	**0.4907**	**15.16**	**0.4105**

Note: Params and FLOPs are computed for the complete inference pipeline of each method, including all components required to produce the translated image. Separate training uses one model for each translation direction, whereas joint training uses one model for both directions. Thus, $\times 2$ in Params denotes the combined parameter count of the two models specific to the two directions. In FLOPs, $\times 50$ or $\times 5$ denotes the cost of each step multiplied by the number of inference steps for one translation.

The refined image is optimized by image reconstruction and target VAE projection consistency:

$$\begin{aligned} \mathcal{L}_{\text{img}} &= \mathbb{E}_q \left[\left\| \widehat{x}^q_{\text{tgt}} - x^q_{\text{tgt}} \right\|_1 + 0.5 \ \text{MSE}\left(\widehat{x}^q_{\text{tgt}}, x^q_{\text{tgt}} \right) \right], \\ x^q_{\text{proj}} &= D_{\text{tgt}} \left(\mu_{\text{tgt}}\left(\widehat{x}^q_{\text{tgt}} \right) \right), \\ \mathcal{L}_{\text{proj}} &= \mathbb{E}_q \left[\left\| x^q_{\text{proj}} - \widehat{x}^q_{\text{tgt}} \right\|_1 \right]. \end{aligned} \tag{15}$$

The Stage II objective is

$$\mathcal{L}_{\text{II}} = \lambda_{\text{flow}}\mathcal{L}_{\text{flow}} + \lambda_{\text{lat}}\mathcal{L}_{\text{lat}} + \lambda_{\text{anc}}\mathcal{L}_{\text{anc}} + \lambda_{\text{img}}\mathcal{L}_{\text{img}} + \lambda_{\text{proj}}\mathcal{L}_{\text{proj}}. \tag{16}$$

4 Experiments

4.1 Experimental Settings

Datasets. Experiments were conducted on two publicly available paired CT and MRI datasets, SynthRAD2023 Huijben et al. (2024) and PCM2025 Siam et al. (2025). We used an 8:2 subject level split for training and testing in each dataset to avoid subject overlap. For PCM2025, T1 weighted and T2 weighted MRI pairs were evaluated separately. All images were resized to 256×256 and normalized to $[0, 1]$.

Compared methods. We compared the proposed method with ResCycleGAN de Bel et al. (2021), Pix2pix Eslami et al. (2020), MedGAN Nie et al. (2018), ARGAN Luo et al. (2022), HisGAN Peng et al. (2023), DDIM Song et al. (2021a), BBDM Li et al. (2023), LDM Rombach et al. (2022), Rectified Flow Liu et al. (2022), and BiBBDM Xue et al. (2025). All baselines except BiBBDM were trained separately for each direction, whereas BiBBDM and the proposed method were trained jointly for bidirectional translation. Diffusion and flow baselines used 50 sampling or integration steps during inference. All methods were evaluated on the same subject level splits, preprocessing pipeline, and paired target images.

Implementation details. We first trained modality specific VAEs for CT and MRI reconstruction, and then froze their encoders and decoders when optimizing the shared bidirectional latent dynamics and image space refiner. The proposed latent dynamics used five step Euler integration. AdamW was used with a weight decay of 1×10^{-4}, and the batch size was 16. We report PSNR and SSIM between the translated images and paired target images. The results reported in Table 1 correspond to the complete model with the image space refiner.

4.2 Experimental Results.

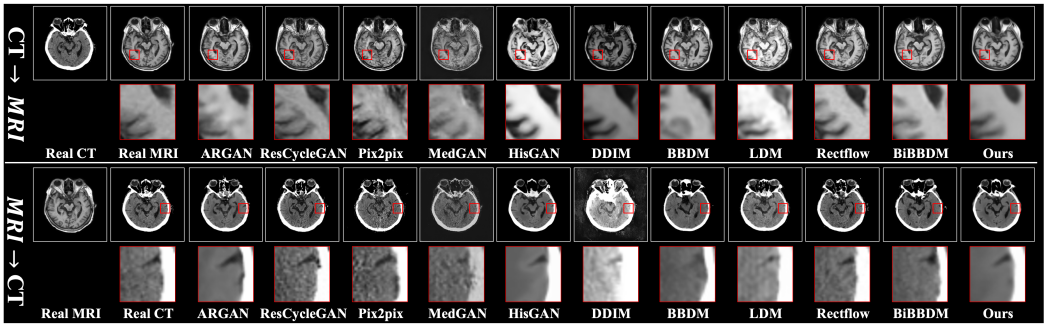

Figure 5: Qualitative comparison on SynthRAD2023 for translation from CT to MRI and from MRI to CT. The enlarged regions highlight representative anatomical boundaries and local tissue transitions. Compared with the competing methods, the proposed method produces fewer boundary discontinuities and local structural artifacts while maintaining better correspondence with the paired target images.

Table 2: Ablation study on SynthRAD2023.

Method	Training	Refiner	CT→MRI		MRI→CT	
			PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
Uni Flow	Separate		19.44 ±0.02	0.7603 ±0.003	19.17 ±0.07	0.7818 ±0.000
Source-Cond BiFlow	Joint		19.36 ±0.05	0.7515 ±0.002	19.11 ±0.07	0.7775 ±0.001
Dir-Cond BiFlow	Joint		19.30 ±0.04	0.7532 ±0.001	19.06 ±0.03	0.7747 ±0.001
Decomp BiFlow	Joint		19.30 ±0.06	0.7595 ±0.001	19.17 ±0.08	0.7799 ±0.002
Anchor BiFlow	Joint		19.44 ±0.02	0.7634 ±0.002	19.26 ±0.02	0.7831 ±0.002
Decomp BiFlow	Joint	✓	19.35 ±0.07	0.7745 ±0.001	18.98 ±0.05	0.7885 ±0.001
Anchor BiFlow	Joint	✓	**19.47** ±0.05	**0.7794** ±0.004	**19.69** ±0.08	**0.8068** ±0.006

Table 3: Decoder robustness analysis on SynthRAD2023 using Anchor BiFlow.

Method	CT→MRI		MRI→CT	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
Anchor BiFlow	19.44 ±0.02	0.7634 ±0.002	19.26 ±0.02	0.7831 ±0.002
+ Perturb	**19.62** ±0.01	0.7656 ±0.001	19.23 ±0.08	0.7862 ±0.001
+ Consistency	19.55 ±0.04	0.7673 ±0.000	**19.31** ±0.02	0.7883 ±0.000
+ Full	19.53 ±0.03	**0.7680** ±0.001	19.30 ±0.05	**0.7884** ±0.000

Quantitative and Qualitative Results To determine whether bidirectional translation can be unified without sacrificing either direction, we compare separate and joint methods across two datasets, two MRI contrasts, translation quality, model size, and inference cost. Table 1 shows that the proposed method achieves the highest mean result among the compared methods in all 12 metric and task combinations. The proposed method therefore maintains stronger overall performance across directions and MRI contrasts with 4.15M parameters and five inference steps. The enlarged ROIs in Fig. 5 further show that these quantitative gains are accompanied by fewer boundary discontinuities and local structural artifacts, although slight smoothing remains in fine regions.

Ablation Studies. To determine whether the performance loss from direct sharing can be recovered through direction modeling, we compare **Source-Cond BiFlow**, **Dir-Cond BiFlow**, **Decomp BiFlow**, and **Anchor BiFlow** against separately trained **Uni Flow**, and further evaluate the image space refiner. Table 2 shows that direct sharing introduces a small performance gap relative to **Uni Flow**, while adding an explicit direction label alone provides little improvement. **Decomp BiFlow** improves structural similarity in both directions and partially recovers this gap, whereas **Anchor BiFlow** matches or exceeds **Uni Flow** in all four metrics before refinement, showing that the improvement mainly comes from the latent dynamics with explicit direction modeling. The image space refiner further improves the final translation quality. Fig. 6 shows clearer local boundaries after refinement, although slight smoothing remains in fine regions.

Direction Sharing and Latent Evolution Analysis. Because a single shared dynamics model must represent two ordered translation processes, we examine whether direction modeling can improve target fitting without aggravating directional imbalance or velocity conflict. We therefore evaluate **Endpoint**, **Imbalance**, and **Conflict**, which measure endpoint discrepancy, discrepancy

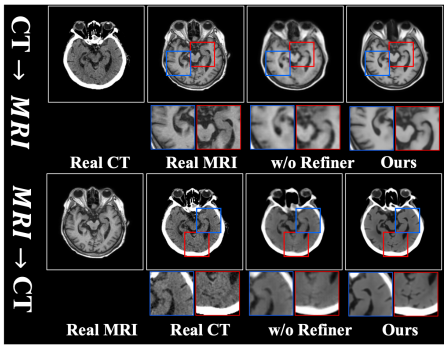

Figure 6: Effect of the image space refiner.

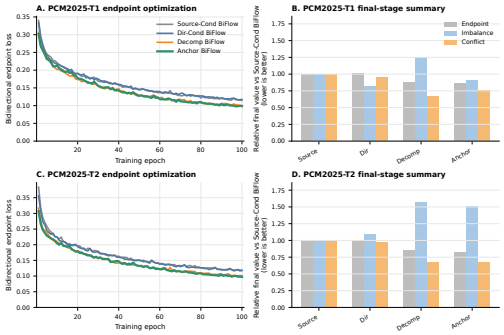

Figure 7: Analysis of direction sharing strategies on PCM2025 T1 and T2.

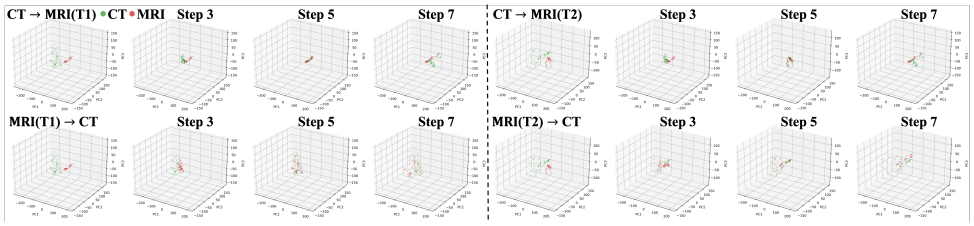

Figure 8: Latent evolution visualization on PCM2025.

between the two directions, and inconsistency with the supervised flow target, respectively. As shown in Fig. 7, **Dir-Cond BiFlow** provides little improvement over **Source-Cond BiFlow**, while **Decomp BiFlow** reduces **Endpoint** and **Conflict** at the cost of substantially increased **Imbalance**. **Anchor BiFlow** achieves the lowest final **Endpoint** on both T1 and T2 and reduces the imbalance introduced by **Decomp BiFlow**, providing the best overall tradeoff among the three criteria. Fig. 8 further shows a gradual evolution from the source latent region toward the target region, supporting continuous bidirectional latent transport.

Robustness Analysis of Latent Reconstruction. Fig. 8 shows that the transported representation reaches the target latent region but may not coincide exactly with the paired target code, which can make reconstruction sensitive to local latent deviations Camuto et al. (2021); Barrett et al. (2022); Khan & Storkey (2023). To improve decoder robustness, we retain **Anchor BiFlow** and introduce three VAE training strategies: **Perturb** reconstructs the clean image from locally perturbed latent codes, **Consistency** constrains decoder outputs within a local latent neighborhood, and **Full** combines both objectives. As shown in Table 3, these strategies consistently improve SSIM in both directions. **Consistency** improves all four mean metrics, while **Full** achieves the highest SSIM, indicating that local decoder robustness can compensate for residual endpoint deviations while keeping the architecture for latent transport and optimization settings unchanged.

5 Conclusion

We presented a shared bidirectional latent dynamics framework for translation between CT and MRI. The experiments reveal that direct parameter sharing introduces a performance gap relative to independently trained models, and that direction identity alone is insufficient to recover this loss. The proposed anchor formulation addresses the directional conflict in shared latent dynamics by providing an explicit reference conditioned on the source state rather than directly predicting the target latent endpoint. The image space refiner further improves both image fidelity and structural similarity, and the observed latent trajectories support the interpretation of translation as progressive evolution from the source modality region toward the target region. These findings show that successful unified bidirectional translation depends not only on parameter sharing but also on how directional variation is represented within the shared dynamics.

AI USE STATEMENT

Generative AI tools were used for language editing, LaTeX assistance, and consistency checks of the manuscript and reported tables. They were not used to generate experimental data or execute experiments. The authors independently verified all AI-assisted revisions against the source results and literature and take full responsibility for the paper.

ETHICS STATEMENT

We have ensured that all data used in this research comply with ethical guidelines and proper citations. The study follows responsible AI practices and avoids any data manipulation or misrepresentation.

REPRODUCIBILITY STATEMENT

The experiments in this work are fully reproducible. All datasets and code used for model training and evaluation are publicly available. Detailed information on the experimental setup, model configurations, and hyperparameters is provided to enable independent reproduction of results.

REFERENCES

Hamed Amini Amirkolaee, Dmitry Olegovich Bokov, and Himanshu Sharma. Development of a gan architecture based on integrating global and local information for paired and unpaired medical image translation. *Expert Systems with Applications*, 203:117421, 2022. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2022.117421.

Fuat Arslan, Bilal Kabas, Onat Dalmaz, Muzaffer Ozbey, and Tolga Cukur. Self-consistent recursive diffusion bridge for medical image translation. *Medical Image Analysis*, 106:103747, 2025. doi: 10.1016/j.media.2025.103747.

Ben Barrett, Alexander Camuto, Matthew Willetts, and Tom Rainforth. Certifiably robust variational autoencoders. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pp. 3663–3683. PMLR, 2022.

Alexander Camuto, Matthew Willetts, Stephen Roberts, Chris Holmes, and Tom Rainforth. Towards a theoretical understanding of the robustness of variational autoencoders. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pp. 3565–3573. PMLR, 2021.

Junxin Chen, Zhiheng Ye, Renlong Zhang, Hao Li, Bo Fang, Li bo Zhang, and Wei Wang. Medical image translation with deep learning: Advances, datasets and perspectives. *Medical Image Analysis*, 103:103605, 2025. doi: https://doi.org/10.1016/j.media.2025.103605.

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/69386f6bb1dfed68692a24c8686939b9-Paper.pdf.

Kyobin Choo, Youngjun Jun, Mijin Yun, and Seong Jae Hwang. Slice-Consistent 3D Volumetric Brain CT-to-MRI Translation with 2D Brownian Bridge Diffusion Model . In *proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, volume LNCS 15007. Springer Nature Switzerland, October 2024.

Thomas de Bel, John-Melle Bokhorst, Jeroen van der Laak, and Geert Litjens. Residual cyclegan for robust domain transformation of histopathological tissue slides. *Medical Image Analysis*, 70: 102004, 2021. doi: https://doi.org/10.1016/j.media.2021.102004.

Mohammad Eslami, Solale Tabarestani, Shadi Albarqouni, Ehsan Adeli, Nassir Navab, and Malek Adjouadi. Image-to-images translation for multi-task organ segmentation and bone suppression in chest x-ray radiography. *IEEE Transactions on Medical Imaging*, 39(7):2553–2565, 2020. doi: 10.1109/TMI.2020.2974159.

Eryan Feng, Pinle Qin, Rui Chai, Jianchao Zeng, Qi Wang, Yanfeng Meng, and Peng Wang. MRI generated from CT for acute ischemic stroke combining radiomics and generative adversarial networks. *IEEE J Biomed Health Inform*, 26(12):6047–6057, December 2022.

Aditya Grover, Christopher Chute, Rui Shu, Zhangjie Cao, and Stefano Ermon. Alignflow: Cycle consistent learning from multiple domains via normalizing flows. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 4028–4035, 2020. doi: 10.1609/aaai.v34i04.5820.

Haithem Hermessi, Olfa Mourali, and Ezzeddine Zagrouba. Multimodal medical image fusion review: Theoretical background and recent advances. *Signal Processing*, 183:108036, 2021. doi: https://doi.org/10.1016/j.sigpro.2021.108036.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc.

Bing Huang, Feng Yang, Mengxiao Yin, Xiaoying Mo, and Cheng Zhong. A review of multimodal medical image fusion techniques. *Computational and Mathematical Methods in Medicine*, 2020 (1):8279342, 2020. doi: https://doi.org/10.1155/2020/8279342.

Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal unsupervised image-to-image translation. In *Proceedings of the European conference on computer vision (ECCV)*, pp. 172–189, 2018.

Evi M.C. Huijben, Maarten L. Terpstra, Arthur Jr. Galapon, Suraj Pai, Adrian Thummerer, Peter Koopmans, Manya Afonso, Maureen van Eijnatten, Oliver Gurney-Champion, Zeli Chen, Yiwen Zhang, Kaiyi Zheng, Chuanpu Li, Haowen Pang, Chuyang Ye, Runqi Wang, Tao Song, Fuxin Fan, Jingna Qiu, Yixing Huang, Juhyung Ha, Jong Sung Park, Alexandra Alain-Beaudoin, Silvain Bériault, Pengxin Yu, Hongbin Guo, Zhanyao Huang, Gengwan Li, Xueru Zhang, Yubo Fan, Han Liu, Bowen Xin, Aaron Nicolson, Lujia Zhong, Zhiwei Deng, Gustav Müller-Franzes, Firas Khader, Xia Li, Ye Zhang, Cédric Hémon, Valentin Boussot, Zhihao Zhang, Long Wang, Lu Bai, Shaobin Wang, Derk Mus, Bram Kooiman, Chelsea A.H. Sargeant, Edward G.A. Henderson, Satoshi Kondo, Satoshi Kasai, Reza Karimzadeh, Bulat Ibragimov, Thomas Helfer, Jessica Dafflon, Zijie Chen, Enpei Wang, Zoltan Perko, and Matteo Maspero. Generating synthetic computed tomography for radiotherapy: Synthrad2023 challenge report. *Medical Image Analysis*, 97: 103276, 2024. doi: https://doi.org/10.1016/j.media.2024.103276.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1125–1134, 2017.

Han Jang, Na Yeon Han, Jangho Kwon, Hyunseok Seo, Beom Jin Park, and Kihwan Choi. Cyclic conditional diffusion models for ct-to-mr synthetic image segmentation with misaligned image pairs. *Expert Systems with Applications*, 304:130631, 2026. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2025.130631.

Asif Khan and Amos Storkey. Adversarial robustness of vaes through the lens of local geometry. In *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pp. 8954–8967. PMLR, 2023.

Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *International Conference on Learning Representations*, 2013. URL `https://api.semanticscholar.org/CorpusID:216078090`.

Bo Li, Kaitao Xue, Bin Liu, and Yu-Kun Lai. Bbdm: Image-to-image translation with brownian bridge diffusion models. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1952–1961, 2023. doi: 10.1109/CVPR52729.2023.00194.

Zhiwen Liang, Mengjie Cheng, Jinhui Ma, Ying Hu, Song Li, and Xin Tian. Multimodal medical image-to-image translation via variational autoencoder latent space mapping. *Med Phys*, 52(7): e17912, May 2025.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *11th International Conference on Learning Representations, ICLR 2023*, 2023.

Guan-Horng Liu, Arash Vahdat, De-An Huang, Evangelos A. Theodorou, Weili Nie, and Anima Anandkumar. I^2sb: Image-to-image schrödinger bridge. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pp. 22042–22062. PMLR, 2023.

Haoyang Liu. Brain ct to mri medical image transformation based on u-net. 2824(1):012002, aug 2024. doi: 10.1088/1742-6596/2824/1/012002.

Ming-Yu Liu, Thomas Breuel, and Jan Kautz. Unsupervised image-to-image translation networks. *Advances in neural information processing systems*, 30, 2017.

Xingchao Liu, Chengyue Gong, and qiang liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *NeurIPS 2022 Workshop on Score-Based Methods*, 2022. URL https://openreview.net/forum?id=gWxpdtQpiYV.

Yanmei Luo, Luping Zhou, Bo Zhan, Yuchen Fei, Jiliu Zhou, Yan Wang, and Dinggang Shen. Adaptive rectification based adversarial network with spectrum constraint for high-quality pet image synthesis. *Medical Image Analysis*, 77:102335, 2022. ISSN 1361-8415. doi: https://doi.org/10.1016/j.media.2021.102335.

Yusen Ni, Ji Ma, and Jinjin Chen. Medical volume ct-to-mri translation with multi-dimensional diffusion architecture. *Biomedical Signal Processing and Control*, 112:108627, 2026. doi: https://doi.org/10.1016/j.bspc.2025.108627.

Dong Nie, Roger Trullo, Jun Lian, Li Wang, Caroline Petitjean, Su Ruan, Qian Wang, and Dinggang Shen. Medical image synthesis with deep convolutional adversarial networks. *IEEE Transactions on Biomedical Engineering*, 65(12):2720–2730, 2018. doi: 10.1109/TBME.2018.2814538.

Taesung Park, Alexei A. Efros, Richard Zhang, and Jun-Yan Zhu. Contrastive learning for unpaired image-to-image translation. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm (eds.), *Computer Vision – ECCV 2020*, pp. 319–345, Cham, 2020. Springer International Publishing.

Yanjun Peng, Jindong Sun, Yande Ren, Dapeng Li, and Yanfei Guo. A histogram-driven generative adversarial network for brain mri to ct synthesis. *Knowledge-Based Systems*, 277:110802, 2023. ISSN 0950-7051. doi: https://doi.org/10.1016/j.knosys.2023.110802.

Costanza Maria Rapillo, Vincent Dunet, Silvia Pistocchi, Alexander Salerno, Valentina Darioli, Bruno Bartolini, et al. Moving from CT to MRI paradigm in acute ischemic stroke: Feasibility, effects on stroke diagnosis and long-term outcomes. *Stroke*, 55(5):1329–1338, 2024. doi: 10.1161/STROKEAHA.123.045154.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Bjorn Ommer. High-Resolution Image Synthesis with Latent Diffusion Models . In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10674–10685, Los Alamitos, CA, USA, June 2022. IEEE Computer Society. doi: 10.1109/CVPR52688.2022.01042.

Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 Conference Proceedings*, SIGGRAPH '22, New York, NY, USA, 2022. Association for Computing Machinery. doi: 10.1145/3528233.3530757.

Zakaria Shams Siam, Md Younus Akon, Israt Jahan Munmun, Abdullah Al-Amin, Md Abdus Salam, and Ishtiak Al Mamoon. A paired CT and MRI dataset for advanced medical imaging applications. *Data Brief*, 61:111768, June 2025.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021a. URL https://openreview.net/forum?id=St1giarCHLP.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021b. URL https://openreview.net/forum?id=PxTIG12RRHS.

Devavrat Tomar, Manana Lortkipanidze, Guillaume Vray, Behzad Bozorgtabar, and Jean-Philippe Thiran. Self-Attentive spatial adaptive normalization for Cross-Modality domain adaptation. *IEEE Trans Med Imaging*, 40(10):2926–2938, September 2021.

Tycho F. A. van der Ouderaa and Daniel E. Worrall. Reversible gans for memory-efficient image-to-image translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4715–4723, 2019.

Peiyong Wang, Bohan Xiao, Qisheng He, Carri Glide-Hurst, and Ming Dong. Score-based image-to-image brownian bridge. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 10765–10773, 2024. doi: 10.1145/3664647.3680999.

Jinbao Wei, Yuhang Chen, Zhijie Wang, Gang Yang, Shimin Tao, Jian Gao, Aiping Liu, and Xun Chen. Rethinking diffusion bridge model with dual alignments for medical image synthesis. In *Proceedings of the 33rd ACM International Conference on Multimedia*, MM '25, pp. 1052–1061, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400720352. doi: 10.1145/3746027.3754923.

Kaitao Xue, Bo Li, Ziyi Liu, Zhifen He, Bin Liu, Congxuan Zhang, and Yu-Kun Lai. Bibbdm: Bidirectional image translation with brownian bridge diffusion models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(11):10546–10559, 2025. doi: 10.1109/TPAMI.2025.3597667.

Shaorong Zhang, Tamoghna Chattopadhyay, Sophia I. Thomopoulos, Jose-Luis Ambite, Paul M. Thompson, and Greg Ver Steeg. Diffusion bridge models for 3d medical image translation. In *2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1–7, 2025. doi: 10.1109/EMBC58623.2025.11252839.

Linqi Zhou, Aaron Lou, Samar Khanna, and Stefano Ermon. Denoising diffusion bridge models. In *International Conference on Learning Representations*, 2024.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pp. 2223–2232, 2017.

Muzaffer Özbey, Onat Dalmaz, Salman U. H. Dar, Hasan A. Bedel, Şaban Özturk, Alper Güngör, and Tolga Çukur. Unsupervised medical image translation with adversarial diffusion models. *IEEE Transactions on Medical Imaging*, 42(12):3524–3539, 2023. doi: 10.1109/TMI.2023.3290149.

A Additional Method Analysis

A.1 Direction Dependence and Hypothesis Relations

Within the two modality setting, z^q_{src} encodes both source modality identity and sample specific anatomical and appearance information. For an ordered translation process $q \in \{a \to b, b \to a\}$, the flow matching target is

$$v^*_q = z^q_{\text{tgt}} - z^q_{\text{src}}. \tag{17}$$

For a paired latent representation (z_a, z_b), the two directional targets are

$$v^*_{a\to b} = z_b - z_a, \qquad v^*_{b\to a} = z_a - z_b = -v^*_{a\to b}. \tag{18}$$

The shared predictor must therefore represent two oppositely ordered transport processes while accounting for sample specific anatomical variation. The four BiFlow variants differ in how this directional dependence is represented. **Source-Cond BiFlow** represents direction implicitly through z^q_{src}. **Dir-Cond BiFlow** introduces d_q as an explicit input. **Decomp BiFlow** separates a shared velocity component from a correction associated with direction. **Anchor BiFlow** combines z^q_{src} and d_q through Δ_ϕ to construct a latent reference that participates in velocity prediction throughout numerical integration.

To formalize the effect of parameter sharing, let

$$s^q = (\bar{z}^q, z^q_{\mathrm{src}}, t)\,, \qquad v^*_q = z^q_{\mathrm{tgt}} - z^q_{\mathrm{src}}, \tag{19}$$

and let $\mathcal{H} = \{h_\theta\}$ denote the base family of source conditioned velocity predictors.

Independent and directly shared families. Independent unidirectional modeling assigns one predictor to each ordered process:

$$\mathcal{H}_{\mathrm{uni}} = \mathcal{H} \times \mathcal{H}. \tag{20}$$

Its optimal flow matching risk is

$$\mathcal{R}^\star_{\mathrm{uni}} = \inf_{(h_{ab},h_{ba})\in\mathcal{H}_{\mathrm{uni}}} \sum_q \mathbb{E}\left[\left\|h_q(s^q) - v^*_q\right\|_2^2\right]. \tag{21}$$

Source-Cond BiFlow ties the two ordered processes to one predictor:

$$\mathcal{H}_{\mathrm{sc}} = \{(h_\theta, h_\theta) \mid h_\theta \in \mathcal{H}\}\,. \tag{22}$$

Its optimal risk is

$$\mathcal{R}^\star_{\mathrm{sc}} = \inf_{(h,h)\in\mathcal{H}_{\mathrm{sc}}} \sum_q \mathbb{E}\left[\left\|h(s^q) - v^*_q\right\|_2^2\right]. \tag{23}$$

Since every tied predictor pair is also an element of $\mathcal{H} \times \mathcal{H}$,

$$\mathcal{H}_{\mathrm{sc}} \subseteq \mathcal{H}_{\mathrm{uni}}, \qquad \mathcal{R}^\star_{\mathrm{uni}} \le \mathcal{R}^\star_{\mathrm{sc}}. \tag{24}$$

Thus, direct parameter sharing imposes a stronger constraint than independent modeling.

Shared families with explicit direction modeling. The three explicit direction modeling variants all contain the directly shared formulation as a special case. For **Dir-Cond BiFlow**, ignoring d_q recovers the source conditioned predictor. For **Decomp BiFlow**, setting the correction c_ψ to zero recovers the shared component alone. For **Anchor BiFlow**, setting the anchor offset and the correction induced by the anchor to zero, while making the velocity predictor independent of the anchor input, recovers the source conditioned formulation.

Let $\mathcal{H}_{\mathrm{dc}}$, $\mathcal{H}_{\mathrm{dd}}$, and $\mathcal{H}_{\mathrm{ag}}$ denote the velocity families induced by these three formulations. Their relations with **Source-Cond BiFlow** are

$$\mathcal{H}_{\mathrm{sc}} \subseteq \mathcal{H}_r, \qquad r \in \{\mathrm{dc}, \mathrm{dd}, \mathrm{ag}\}. \tag{25}$$

Under the same flow matching objective and data distribution,

$$\mathcal{R}^\star_r \le \mathcal{R}^\star_{\mathrm{sc}}, \qquad r \in \{\mathrm{dc}, \mathrm{dd}, \mathrm{ag}\}. \tag{26}$$

These relations show that explicit direction modeling expands the directly shared family while retaining one jointly trained bidirectional framework. The practical behavior of these formulations is evaluated through the ablation study and the latent space diagnostics.

A.2 Latent Space Robustness Training for Modality Specific VAEs

Anchor BiFlow produces a transported endpoint $\widehat{z}_{\mathrm{tgt}}$ in the target latent space. Residual deviations from the paired target representation z_{tgt} are propagated through the target modality decoder and can affect image reconstruction. Recent studies have related VAE reconstruction robustness to the local geometry and sensitivity of encoder and decoder mappings Camuto et al. (2021); Barrett et al. (2022); Khan & Storkey (2023). We therefore study local decoder robustness through VAE training with latent perturbations.

The robustness strategies modify only the Stage I training objectives of the modality specific VAEs. The VAE architecture, latent dimensionality, training data, **Anchor BiFlow** architecture, and latent transport optimization settings remain unchanged. After each VAE variant is pretrained, its encoder and decoder are frozen, and **Anchor BiFlow** is optimized using the same transport objectives. For these experiments, Stage II optimizes only **Anchor BiFlow**, without the image space refiners or the associated refinement losses. The reported results are therefore obtained from directly decoded coarse outputs.

For modality $m \in \{a, b\}$, let

$$z_m = \mu_m(x_m) \tag{27}$$

denote the posterior mean latent representation of image x_m. A locally perturbed latent representation is constructed as

$$\widetilde{z}_m = z_m + \epsilon_m, \qquad \epsilon_m \sim \mathcal{N}\left(0, \sigma_m^2 I\right), \tag{28}$$

where σ_m controls the perturbation magnitude. The corresponding clean and perturbed reconstructions are

$$\widehat{x}_m = D_m(z_m), \qquad \widetilde{x}_m = D_m(\widetilde{z}_m). \tag{29}$$

The perturbation is applied only during VAE pretraining and is not used during latent transport or inference.

All variants use the baseline VAE objective

$$\mathcal{L}_{\text{base}} = \sum_{m \in \{a,b\}} \left[\ell_{\text{rec}}\left(D_m(z_m), x_m\right) + \beta \mathcal{L}_{\text{kl}}^m\right], \tag{30}$$

where ℓ_{rec} denotes the reconstruction loss used by the baseline VAE, $\mathcal{L}_{\text{kl}}^m$ denotes the KL divergence term for modality m, and β controls its contribution.

Perturbation reconstruction. The **Perturb** strategy requires the decoder to reconstruct the clean image from a locally perturbed latent representation:

$$\mathcal{L}_{\text{pert}} = \sum_{m \in \{a,b\}} \mathbb{E}_{\epsilon_m}\left[\ell_{\text{rec}}\left(D_m(z_m + \epsilon_m), x_m\right)\right]. \tag{31}$$

The corresponding VAE objective is

$$\mathcal{L}_{\text{VAE}}^{\text{pert}} = \mathcal{L}_{\text{base}} + \lambda_{\text{pert}} \mathcal{L}_{\text{pert}}, \tag{32}$$

where λ_{pert} controls the perturbation reconstruction term.

Local output consistency. The **Consistency** strategy constrains the decoder outputs before and after latent perturbation to remain consistent:

$$\mathcal{L}_{\text{cons}} = \sum_{m \in \{a,b\}} \mathbb{E}_{\epsilon_m}\left[\|D_m(z_m + \epsilon_m) - D_m(z_m)\|_1\right]. \tag{33}$$

The corresponding VAE objective is

$$\mathcal{L}_{\text{VAE}}^{\text{cons}} = \mathcal{L}_{\text{base}} + \lambda_{\text{cons}} \mathcal{L}_{\text{cons}}, \tag{34}$$

where λ_{cons} controls the consistency term. The clean reconstruction term in $\mathcal{L}_{\text{base}}$ provides image level supervision, while $\mathcal{L}_{\text{cons}}$ promotes local decoder smoothness around each latent representation.

Combined robustness training. The **Full** strategy jointly applies perturbation reconstruction and local output consistency:

$$\begin{aligned} \mathcal{L}_{\text{VAE}}^{\text{full}} = \mathcal{L}_{\text{base}} &+ \lambda_{\text{pert}} \mathcal{L}_{\text{pert}} \\ &+ \lambda_{\text{cons}} \mathcal{L}_{\text{cons}}. \end{aligned} \tag{35}$$

Within each training iteration, the same perturbed latent representation $\widetilde{z}_m$ is used to calculate both robustness terms. The perturbation reconstruction term encourages image recoverability around the clean latent representation, whereas the consistency term reduces variations in decoder outputs within the same local latent neighborhood.

As reported in **Table 3 of the main paper**, all three strategies improve SSIM in both translation directions. **Consistency** improves all four mean metrics, while **Full** achieves the highest SSIM in both directions, supporting the use of local decoder robustness for transported endpoints with residual latent deviations.

B Implementation and Training Details

We adopt a two stage training strategy. In Stage I, the CT VAE and MR VAE are pretrained separately. Their encoders and decoders are then frozen throughout Stage II, in which **Anchor BiFlow** and the two image space refiners are jointly optimized. All images are resized to 256×256, the batch size is 16, and the latent channel dimension is 8. The latent dynamics uses five Euler integration steps with $\alpha = 1.0$ for anchor guidance.

AdamW is used with a weight decay of 1×10^{-4}. The learning rate is 2×10^{-4} for VAE pretraining, the shared flow, and the anchor mapper, while the image space refiners use 1×10^{-4} during Stage II. The Stage I VAE objective combines image reconstruction and KL regularization, with a linear warmup schedule for the KL weight up to 1×10^{-4}. During Stage II, the velocity supervision, latent endpoint alignment, and anchor alignment losses are equally weighted. The refined image reconstruction and frozen VAE projection reconstruction losses are also assigned a weight of 1.0. The refined reconstruction loss is implemented as $\ell_1 + 0.5\,\mathrm{MSE}$.

For the latent robustness experiments, we use Gaussian perturbations with $\sigma_a = \sigma_b = 0.05$ and set $\lambda_{\mathrm{pert}} = 0.5$ and $\lambda_{\mathrm{cons}} = 0.1$.

C Additional Experiments

C.1 Optimization Diagnostics for Direction Modeling

We characterize shared bidirectional dynamics using three complementary latent space diagnostics: **Endpoint Error**, **Directional Imbalance**, and **Conflict**. Conflict is reported in the diagnostic figures as **Target Alignment Error**. These quantities measure endpoint discrepancy after numerical integration, the relative difference between the endpoint errors of the two directions, and the angular inconsistency between predicted velocities and supervised source to target displacements, respectively. These diagnostics are intended to characterize optimization behavior rather than generalization performance. They are therefore computed on the training pairs using the model parameters at the end of each epoch and are not included in the optimization objective.

Let a pair of spatially aligned images be encoded as

$$z_a = \mu_a(x_a), \qquad z_b = \mu_b(x_b). \tag{36}$$

For the two translation directions, numerical integration produces

$$\widehat{z}_b = \Phi_\Theta^{a\to b}(z_a), \qquad \widehat{z}_a = \Phi_\Theta^{b\to a}(z_b), \tag{37}$$

where Φ_Θ^q denotes the complete latent transport operator for direction q. This notation covers all evaluated variants and includes their direction code, velocity correction, or latent anchor when applicable.

Endpoint Error. The directional endpoint errors are

$$E_{a\to b} = \left\|\widehat{z}_b - z_b\right\|_1, \qquad E_{b\to a} = \left\|\widehat{z}_a - z_a\right\|_1. \tag{38}$$

The bidirectional endpoint error is

$$\mathcal{E}_{\mathrm{end}} = \frac{1}{2}\left(E_{a\to b} + E_{b\to a}\right). \tag{39}$$

The reported value is averaged over all training pairs. A lower value indicates that the integrated trajectories terminate closer to the paired target latent representations.

Directional Imbalance. To measure whether the two directions are optimized comparably, we define

$$\mathcal{E}_{\mathrm{imb}} = \frac{|E_{a\to b} - E_{b\to a}|}{E_{a\to b} + E_{b\to a} + \varepsilon}, \tag{40}$$

where ε is a small constant for numerical stability. A value close to zero indicates similar endpoint errors in the two directions, whereas a larger value indicates a greater relative discrepancy.

Conflict measured by Target Alignment Error. **Endpoint Error** describes discrepancy after complete trajectory integration, whereas **Conflict** measures local angular inconsistency between the predicted velocity and the paired source to target displacement. We evaluate the velocity field at the midpoint of the paired latent representations:

$$z_{\mathrm{mid}} = \frac{1}{2}\left(z_a + z_b\right), \qquad t_{\mathrm{mid}} = 0.5. \tag{41}$$

The predicted velocities are

$$\begin{aligned} v_{a\to b} &= v_\Theta^{a\to b}\left(z_{\mathrm{mid}}, z_a, t_{\mathrm{mid}}\right), \\ v_{b\to a} &= v_\Theta^{b\to a}\left(z_{\mathrm{mid}}, z_b, t_{\mathrm{mid}}\right), \end{aligned} \tag{42}$$

where v_Θ^q denotes the complete velocity predictor for direction q. The supervised displacement directions are

$$\Delta_{a\to b} = z_b - z_a, \qquad \Delta_{b\to a} = z_a - z_b. \tag{43}$$

Using

$$\cos(u, v) = \frac{\langle u, v\rangle}{\|u\|_2 \|v\|_2 + \varepsilon}, \tag{44}$$

the target alignment error is

$$\begin{aligned} \mathcal{E}_{\mathrm{align}} = 1 - \frac{1}{2}\Big[&\cos\left(v_{a\to b}, \Delta_{a\to b}\right) \\ &+ \cos\left(v_{b\to a}, \Delta_{b\to a}\right)\Big]. \end{aligned} \tag{45}$$

The reported value is averaged over all training pairs. A lower value means that the predicted velocities are more closely aligned with the correct source to target displacements.

Evaluation protocol. All three diagnostics are computed at the end of each training epoch and averaged over the complete training set. The ℓ_1 distances in **Endpoint Error** are averaged over all latent channels and spatial positions. The curves report the epoch values without additional smoothing. For the final stage summary in panel (d), each diagnostic is averaged over the last W epochs and normalized by the corresponding value of **Source-Cond BiFlow**:

$$R_s^r = \frac{\frac{1}{W}\sum_{e=E-W+1}^{E} \mathcal{E}_{s,e}^r}{\frac{1}{W}\sum_{e=E-W+1}^{E} \mathcal{E}_{s,e}^{\mathrm{sc}} + \varepsilon}, \qquad s \in \{\mathrm{end}, \mathrm{imb}, \mathrm{align}\}, \tag{46}$$

where $W = 10$ defines a short final training window used to reduce epoch level fluctuations in the summary statistics, r denotes a direction modeling variant, and ε is a small constant used for numerical stability.

Optimization behavior on PCM2025 T1.

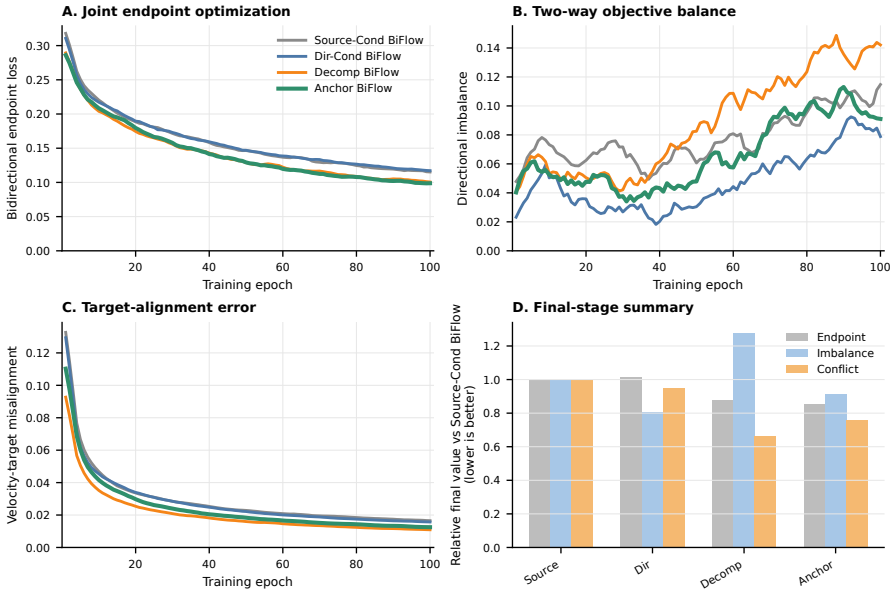

Figure 9: Optimization diagnostics of the four direction modeling variants on PCM2025 T1. Panels (a)–(c) show **Endpoint Error**, **Directional Imbalance**, and **Target Alignment Error (Conflict)** during training. Panel (d) summarizes the final stage results normalized by **Source-Cond BiFlow**. Lower values are better.

On PCM2025 T1, **Source-Cond BiFlow** and **Dir-Cond BiFlow** follow similar endpoint trajectories, while **Decomp BiFlow** and **Anchor BiFlow** reach lower final endpoint errors. **Anchor BiFlow** obtains the lowest final endpoint error. **Decomp BiFlow** achieves the lowest final target alignment error but shows the strongest increase in directional imbalance. These results indicate that endpoint fitting, local velocity alignment, and directional balance capture distinct aspects of the shared optimization process.

Optimization behavior on PCM2025 T2.

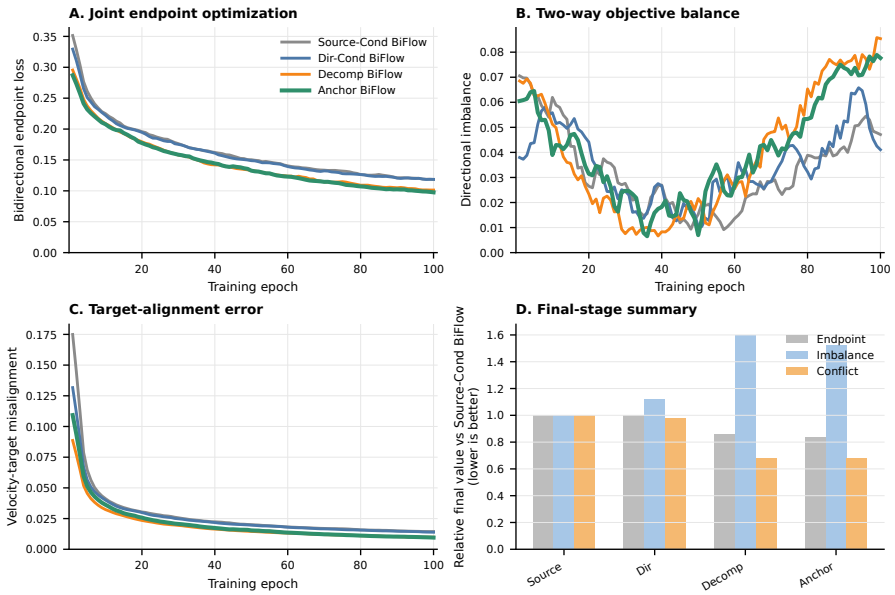

Figure 10: Optimization diagnostics of the four direction modeling variants on PCM2025 T2. Panels (a)–(c) show **Endpoint Error**, **Directional Imbalance**, and **Target Alignment Error (Conflict)** during training. Panel (d) summarizes the final stage results normalized by **Source-Cond BiFlow**. Lower values are better.

PCM2025 T2 shows a similar ordering for endpoint error. **Anchor BiFlow** again reaches the lowest final endpoint level, while **Source-Cond BiFlow** and **Dir-Cond BiFlow** remain close. Directional imbalance decreases during the early stage and rises later for all variants, even while endpoint errors continue to improve. For target alignment, **Anchor BiFlow** and **Decomp BiFlow** outperform the two conditioning based variants during the later stage, with **Anchor BiFlow** obtaining the lowest final value.

Across both MRI contrasts, **Anchor BiFlow** consistently provides the lowest final endpoint error, while a scalar direction code alone offers little consistent improvement over **Source-Cond BiFlow**. **Decomp BiFlow** and **Anchor BiFlow** generally improve local velocity alignment, although their imbalance behavior differs across the two settings. This comparison supports the use of an explicit directional structure for shared latent transport.

C.2 Inference Efficiency

Table 4: Inference efficiency comparison. Latency is reported as mean ± standard deviation. Separate methods train one model for each translation direction, whereas Joint methods learn both directions within one jointly trained framework.

Method	Training	Params (M)	FLOPs (G)	Steps	CT→MRI (ms)	MRI→CT (ms)	Avg. (ms)	FPS
ResCycleGAN	Separate	11.37×2	97.49	1	3.05 ± 0.04	3.06 ± 0.04	3.05 ± 0.04	327.41
Pix2pix		54.41×2	11.90	1	0.96 ± 0.04	0.96 ± 0.04	0.96 ± 0.04	1040.51
MedGAN		31.04×2	96.22	1	2.39 ± 0.04	2.40 ± 0.04	2.40 ± 0.04	417.42
ARGAN		45.25×2	59.79	1	1.82 ± 0.04	1.82 ± 0.04	1.82 ± 0.04	550.54
HisGAN		4.13×2	7.27	1	2.98 ± 0.14	2.99 ± 0.14	2.99 ± 0.14	334.86
DDIM		5.21×2	97.99×50	50	206.93 ± 0.11	206.93 ± 0.11	206.93 ± 0.11	4.83
BBDM		27.69×2	316.11×50	50	477.11 ± 0.90	478.32 ± 0.46	477.72 ± 0.68	2.09
LDM		6.79×2	7.06×50	50	56.51 ± 0.12	56.51 ± 0.12	56.51 ± 0.12	17.70
Rectified Flow		5.21×2	97.99×50	50	202.02 ± 0.11	202.02 ± 0.11	202.02 ± 0.11	4.95
BiBBDM	Joint	27.69	316.19×50	50	478.05 ± 0.33	478.42 ± 0.25	478.24 ± 0.29	2.09
Proposed		4.15	16.36×5	5	6.23 ± 0.16	6.30 ± 0.15	6.26 ± 0.16	159.64

Note: For Separate methods, $\times 2$ in Params denotes the combined parameter count of the two models trained for the two translation directions. In FLOPs, $\times 50$ or $\times 5$ denotes the computational cost of each step multiplied by the number of inference steps for one translation. FPS is calculated from the average latency.

Table 4 reports model size, computational cost, and measured inference latency for all compared methods. All latency measurements were conducted on an NVIDIA GeForce RTX 5090 D GPU with a batch size of 1 and an input resolution of 256×256. Each model was first executed for 10 warm up runs to stabilize GPU execution, followed by 50 timed inference runs. We report the mean latency and standard deviation over the timed runs.

The proposed framework contains 4.15M parameters and performs five flow evaluations for either translation direction, with 16.36G FLOPs for each evaluation. It achieves an average latency of 6.26 ± 0.16 ms and 159.64 FPS. One step GAN methods provide lower absolute latency, whereas the proposed framework remains substantially faster than the iterative diffusion and flow baselines while learning both translation directions within one jointly trained system.

C.3 Extended Latent Evolution and Decoded Translation Trajectories

To visualize the learned shared latent dynamics beyond the standard inference horizon, we extend the five step integration used in the main experiments to ten steps on PCM2025 T1 and PCM2025 T2. The standard procedure uses $\Delta t = 0.2$ over $t \in [0, 1]$. For visualization only, we retain the same Euler step size and continue integration from Step 6 to Step 10, giving an extrapolated trajectory over $t \in (1, 2]$. Steps 1 to 5 correspond to the reported inference protocol, while Steps 6 to 10 are used only for trajectory analysis. For each setting, the initial distributions and all intermediate latent states are projected using one PCA basis with identical axis ranges across steps.

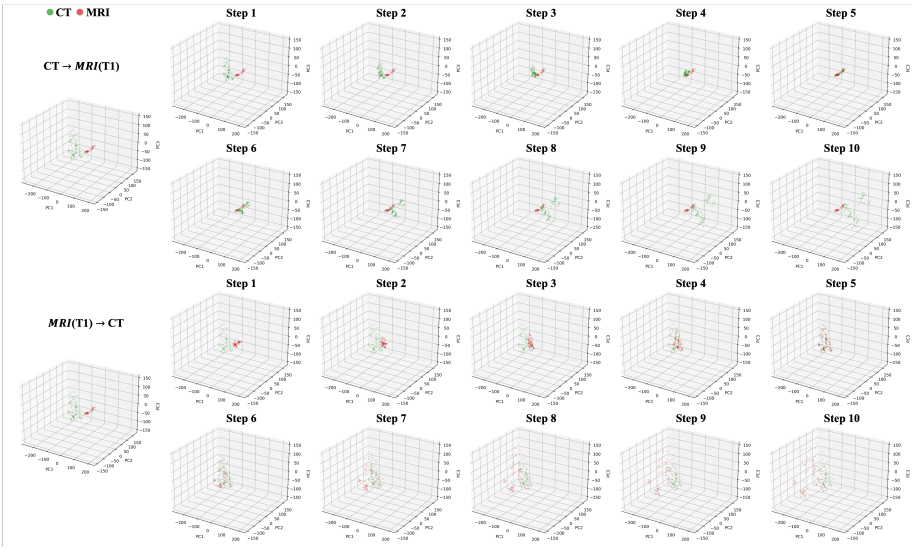

Figure 11: Extended latent evolution on PCM2025 T1.

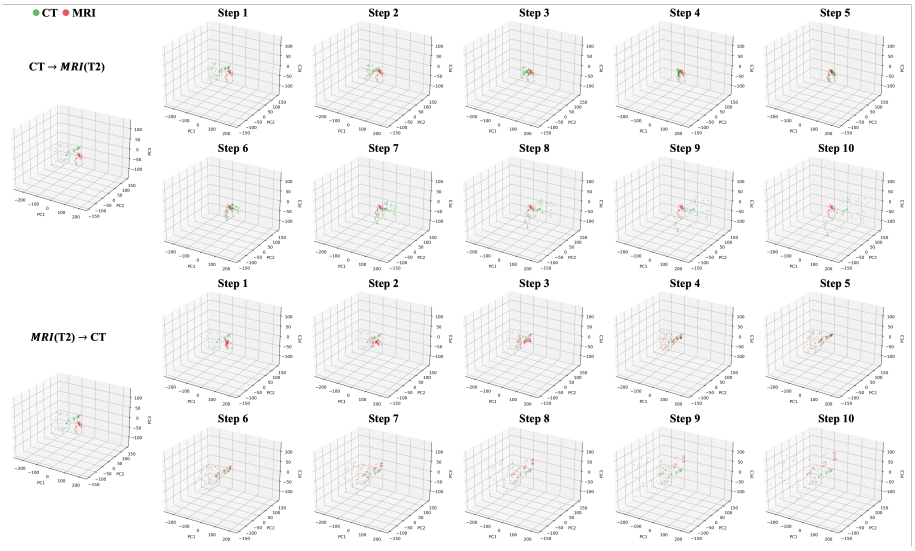

Figure 12: Extended latent evolution on PCM2025 T2.

Figures 11 and 12 extend **Fig. 8 in the main paper**. In both translation directions, the first five steps progressively move source modality latent samples toward the target modality distribution. Continued integration beyond Step 5 carries the samples past the target region, showing that the learned velocity field continues to define a coherent direction beyond the standard terminal point. The two directions exhibit different motion patterns, but the same qualitative behavior is observed for both T1 weighted and T2 weighted MRI.

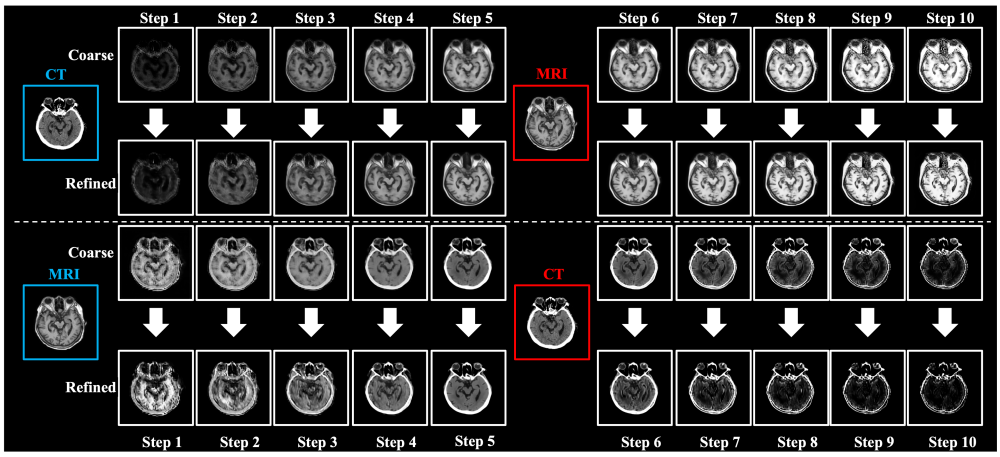

Figure 13: Decoded translation results along the extended latent trajectory of **Anchor BiFlow**. For each direction, intermediate transported latent states are decoded from Step 1 to Step 10. Each step shows both the directly decoded coarse output and the corresponding result after image space refinement.

Figure 13 shows the corresponding decoded images. During Steps 1 to 5, the translated images progressively approach the target modality. Continued integration beyond the standard inference horizon produces increasingly clear overshooting effects, including exaggerated intensity, structural distortion, blur, and reduced anatomical fidelity. The image space refiner suppresses local noise and decoding artifacts and improves boundary and texture details, but it does not reverse an under transported or overshot latent state. These observations support a functional separation in which **Anchor BiFlow** governs the global direction and progress of modality translation, while the image space refiner performs local residual correction around the decoded output.