Magnetic Resonance Unpaired Image Translation with Pseudometric Schrödinger Bridges

Shuwen Wei[1*] **Samuel W. Remedios**[1] **Zhangxing Bian**[1] **Shimeng Wang**[1]
Jinwei Zhang[2] **Junyu Chen**[3] **Yihao Liu**[4] **Lianrui Zuo**[4]
Muhammad Faizyab Ali Chaudhary[1] **Blake E. Dewey**[3] **Aaron Carass**[1]
Jerry L. Prince[1]
[1]Johns Hopkins University [2]University of Massachusetts Amherst
[3]Johns Hopkins School of Medicine [4]Vanderbilt University

Abstract

Unpaired image translation is a critical preprocessing strategy for synthesizing missing sequences and harmonizing contrast variability in magnetic resonance imaging (MRI). However, in medical image processing, distribution-level translation is insufficient. It is vitally important that the anatomy is not distorted, corrupted, or lost after translation. Previous methods, like diffusion Schrödinger bridge matching (DSBM), achieve unpaired image translation, but do not preserve anatomy. We introduce the pseudometric Schrödinger bridge (PMSB), which trains two independent volume-preserving, isometry-regularized diffeomorphisms to map data from each domain to a latent anatomical pseudometric space. Then, DSBM is used within that space, explicitly minimizing the transport gap while preserving anatomy. Extensive experiments on the multi-contrast OASIS-3 dataset demonstrate that PMSB establishes a new state-of-the-art in structural consistency for cross-contrast synthesis. PMSB outperforms other leading methods and drastically reduces anatomical hallucinations, notably yielding a mean PSNR of 26.70 dB, a mean SSIM of 0.86, and a mean LNCC of 0.24.

1 Introduction

Magnetic resonance (MR) imaging (MRI) is a cornerstone of non-invasive neuroscience and clinical research (Westbrook et al., 2011; Brown et al., 2014; Prince and Links, 2015). However, the lack of acquisition standards across multi-center cohorts often leads to missing sequences, while variability in field strengths, pulse sequences, and reconstruction algorithms causes severe contrast inconsistencies (Biberacher et al., 2016). This heterogeneity severely limits both uniform downstream analysis and the generalizability of medical image analysis methods (Guan and Liu, 2022). Image translation has emerged as a vital preprocessing strategy to solve this by synthesizing missing sequences and harmonizing contrast variability, ensuring a complete and standardized dataset (Mirzaalian et al., 2016; Dewey et al., 2019; Wrobel et al., 2020; Zuo et al., 2020, 2021, 2022, 2023). However, because these synthesized images are often used for sensitive clinical measurements, preserving strict anatomical fidelity during image translation is important.

Because paired scans are scarce (Tian et al., 2022), unpaired image translation is preferred for its scalability (Zhu et al., 2017; Liu et al., 2017; Huang et al., 2018; Park et al., 2020; Sasaki et al., 2021), despite the risk of anatomical distortion. While diffusion Schrödinger bridge matching (DSBM) (Bortoli et al., 2021; Shi et al., 2023) provides state-of-the-art generative performance in distribution-level

*Corresponding author: swei14@jhu.edu code: https://github.com/Shuwen-Wei/pmsb

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

image translation, its implicit Euclidean transport cost ignores anatomical fidelity, causing structural shifts in medical images (Wei et al., 2025). The latent metric Schrödinger bridge (LMSB) (Wei et al., 2025) mitigates this by applying DSBM within an anatomy-guided latent metric space. However, a strict metric space forces images with identical anatomy but varying contrasts to map to distinct latent points, leaving an unnecessary transport gap.

In this paper, we propose the pseudometric Schrödinger bridge (PMSB) to explicitly resolve this limitation. By generalizing the latent metric space into a pseudometric space, PMSB theoretically allows disparate contrasts of identical anatomy to map to the exact same latent coordinate, effectively reducing the transport gap and improving structural preservation. Our key contributions are fourfold: **1)** We formulate a rigorous set of geometrical assumptions, including deterministic coupling, diffeomorphism, volume preservation, and approximate isometry, to provide a theoretical foundation for constructing latent random variables and constraining the function space; **2)** We provide a new theoretical perspective for deriving the learning objective and prove its validity under the geometrical assumptions; **3)** We mathematically generalize the latent metric space of LMSB into a pseudometric space via pullback metrics (i.e., replacing a shared mapping with independent mappings), establishing a mathematically superior space for unpaired image translation; and **4)** We demonstrate through experiments on the multi-contrast OASIS-3 dataset that PMSB outperforms current state-of-the-art unpaired methods, achieving unprecedented structural consistency and anatomical preservation.

2 Related Work

Coupling. Coupling is a mathematical technique that relates two separate probability distributions within a shared space (Villani, 2008), providing a natural formulation for image translation (Wei et al., 2025). However, a general mathematical coupling, such as an independent coupling, does not inherently preserve anatomical information, which is a strict requirement for reliable medical image analysis. Therefore, finding a proper coupling for medical image translation requires introducing explicit structural constraints or geometric assumptions into the underlying formulation. The following sections review how existing methods incorporate these constraints and assumptions.

Generative Adversarial Networks. A generative adversarial network (GAN) (Goodfellow et al., 2020) finds a deterministic coupling by adversarial training, which uses distribution-level supervision but does not explicitly preserve anatomy (Hiasa et al., 2018; Yang et al., 2020; Liu et al., 2021; Cackowski et al., 2023). To preserve anatomy, GAN-based methods incorporate auxiliary consistency losses (e.g., cycle-consistency loss (Hiasa et al., 2018) proposed in CycleGAN (Zhu et al., 2017) and DualGAN (Yi et al., 2017), segmentation loss (Cackowski et al., 2023), or both (Yang et al., 2020)).

Diffusion Models. Diffusion models (Anderson, 1982; Sohl-Dickstein et al., 2015; Ho et al., 2020; Song et al., 2021b,a) learn to map noise to data distributions through gradual denoising, effectively establishing a stochastic coupling between domains. While highly effective for generation, ensuring structural fidelity in unpaired translation remains challenging. Recent methods tackle this by combining diffusion with other paradigms or explicit constraints: the dual diffusion implicit bridge (DDIB) (Su et al., 2023; Wei et al., 2026) trades generation quality for improved structural fidelity, SynDiff (Özbey et al., 2023) couples diffusive and non-diffusive modules via cycle-consistency to enhance translation quality, and ContourDiff (Chen et al., 2025b) directly enforces spatial constraints by leveraging domain-invariant anatomical contour representations during diffusion.

Optimal Transport. Optimal transport (OT) (Monge, 1781; Kantorovich, 1942; Villani, 2008) finds an optimal coupling minimizing a transport cost, which requires careful designed to preserve anatomy. Exact solutions for OT plan between discrete measures are computationally expensive, though entropy-regularized OT can be solved efficiently via the Sinkhorn algorithm (Cuturi, 2013; Sinkhorn, 1967). Deep learning approaches like Wasserstein GANs (Arjovsky et al., 2017) solve OT via Kantorovich duality (Kantorovich, 1940) but are difficult to train. Specifically, enforcing the required bounded Lipschitz continuity for Euclidean costs relies on approximations (e.g., spectral normalization (Miyato et al., 2018), weight clipping (Arjovsky et al., 2017), or gradient penalties (Gulrajani et al., 2017)), which hinder generalization to arbitrary transport costs, motivating alternative dynamic formulations.

Schrödinger Bridge. The Schrödinger bridge (SB) (Schrödinger, 1931, 1932; Léonard, 2014) is such a dynamic formulation, which finds an optimal coupling with respect to a reference path measure that must be carefully designed to preserve anatomy. SB is equivalent to an entropy-regularized OT (Föllmer, 1988; Léonard, 2014). Specifically, a non-drift diffusion reference measure

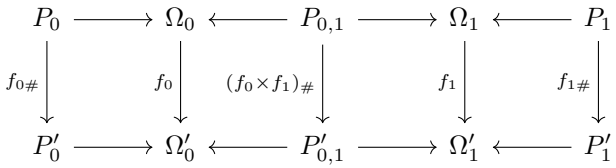

Figure 1: Diagram of the maps f_0 and f_1 alongside their induced pushforwards on probability measures. Although the true coupling $P_{0,1}$ is not necessarily the OT plan between P_0 and P_1, PMSB seeks to learn f_0 and f_1 such that the resulting pushforward coupling $P'_{0,1}$ is the exact OT plan between P'_0 and P'_1 under a squared Euclidean cost.

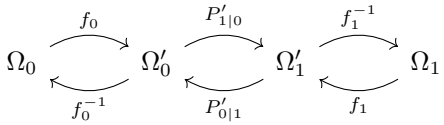

Figure 2: Bidirectional image translation using the pushforward measure $P'_{0,1}$. Translation from the domain Ω_0 to the domain Ω_1 is achieved by composing the forward mappings from left to right, whereas the reverse translation from Ω_1 to Ω_0 follows the inverse path from right to left.

in SB corresponds to a squared Euclidean transport cost in OT, with the entropy term weighted by the total diffusion amount. While traditionally solved via computationally expensive iterative projection fitting (IPF) (Kullback, 1968; Rüschendorf and Thomsen, 1993; Rüschendorf, 1995), modern diffusion-based approaches (Bortoli et al., 2021) solve SBs more efficiently. For instance, DSBM (Shi et al., 2023) uses iterative Markovian fitting (IMF) (Shi et al., 2023; Peluchetti, 2023) for state-of-the-art image translation, while unpaired neural Schrödinger bridge (UNSB) (Kim et al., 2024) scales SBs to high-resolution unpaired data via sequential adversarial learning. However, directly applying SBs to medical images causes anatomical shifts since the implicit Euclidean cost ignores anatomical fidelity (Wei et al., 2025). To overcome this, the latent metric Schrödinger bridge (LMSB) (Wei et al., 2025) mitigates such anatomical shifts by applying DSBM within an intermediary, anatomy-guided latent metric space. However, its reliance on a strict metric space leaves an unnecessary residual transport gap that our proposed method resolves.

3 Pseudometric Schrödinger Bridge

3.1 Translation with Pushforward Measures

Let $(\Omega_i, \mathcal{F}_i, P_i)$ for $i \in \{0, 1\}$ be two probability spaces with associated random variables $X_i \sim P_i$, where $\Omega_i \subseteq \mathbb{R}^d$ are measurable sets, $\mathcal{F}_i \subseteq 2^{\Omega_i}$ are their respective σ-algebras, and $P_i : \mathcal{F}_i \to [0, 1]$ are probability measures. A coupling between these two spaces is defined as a new probability space $(\Omega_{0,1}, \mathcal{F}_{0,1}, P_{0,1})$, where $\Omega_{0,1} = \Omega_0 \times \Omega_1$, $\mathcal{F}_{0,1}$ is the product σ-algebra, and $P_{0,1} : \mathcal{F}_{0,1} \to [0, 1]$ such that the marginals satisfy $\mathrm{proj}_{0\#} P_{0,1} = P_0$ and $\mathrm{proj}_{1\#} P_{0,1} = P_1$. Here, $\mathrm{proj}_i : \Omega_{0,1} \to \Omega_i$ are the projection maps, and the subscript $\#$ denotes the pushforward of the measure.

A true coupling $P_{0,1}$ linking medical images of identical anatomy across different contrasts is not necessarily the OT plan under a squared Euclidean transport cost $c(x_0, x_1) = \|x_1 - x_0\|^2$. This discrepancy highlights the fundamental issue of directly applying DSBM to medical image translation: medical images with identical underlying anatomy but different contrasts do not necessarily have small Euclidean distances in the raw image space.

The core intuition of PMSB is to learn intermediate maps $f_0 : \Omega_0 \to \Omega'_0$ and $f_1 : \Omega_1 \to \Omega'_1$ such that the pushforward coupling $P'_{0,1} = (f_0 \times f_1)_\# P_{0,1}$ becomes the exact OT plan between the pushforward marginals $P'_0 = f_{0\#} P_0$ and $P'_1 = f_{1\#} P_1$. Formally, we aim for:

$$P'_{0,1} = \underset{P' \in \Gamma(P'_0, P'_1)}{\arg\min} \int_{\Omega'_{0,1}} \|x'_1 - x'_0\|^2 P'(dx'_0, dx'_1), \tag{1}$$

where $\Gamma(P'_0, P'_1)$ is the set of joint laws with marginals P'_0 and P'_1, respectively. A diagram illustrating these mappings is shown in Fig. 1.

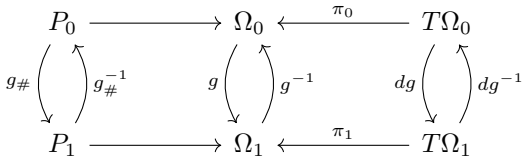

Figure 3: Diagram of a diffeomorphism g between differentiable manifolds Ω_0 and Ω_1, alongside its induced pushforward $g_\#$ on probability measures and the differential dg on tangent bundles. For $i \in \{0, 1\}$, the bundle projection π_i maps the tangent bundle $T\Omega_i$ to its base space Ω_i.

By applying DSBM between P_0' and P_1', we obtain the optimal coupling to an entropy-regularized OT problem, efficiently approximating the solution to Eq. 1. Once f_0, f_1, and $P_{0,1}'$ are obtained, bidirectional image translation follows naturally as shown in Fig. 2. For the forward translation, we map a source image $x_0 \in \Omega_0$ to the latent space $x_0' = f_0(x_0)$, sample $x_1' \in \Omega_1'$ from $P_{0,1}'(\cdot|x_0')$, and apply the inverse mapping $x_1 = f_1^{-1}(x_1')$. The reverse translation symmetrically computes $x_1' = f_1(x_1)$, samples $x_0' \sim P_{0|1}'(\cdot \mid x_1')$, and reconstructs $x_0 = f_0^{-1}(x_0')$.

3.2 Geometrical Assumptions

In medical image translation, Ω_0 and Ω_1 represent images describing an identical set of anatomies but with different contrasts. To ensure that the learned mappings preserve the necessary anatomical structures without distorting probability distributions, we must establish a set of rigorous geometrical assumptions that govern the true coupling $P_{0,1} \in \Gamma(P_0, P_1)$.

Assumption 3.1. ***(Deterministic Coupling)*** *The true coupling $P_{0,1} \in \Gamma(P_0, P_1)$ is a deterministic coupling, meaning $\exists g : \Omega_0 \to \Omega_1$ such that $P_{0,1} = (\mathrm{id}, g)_\# P_0$, where $\mathrm{id} : \Omega_0 \to \Omega_0$ is the identity map, and $(\mathrm{id}, g) : \Omega_0 \to \Omega_0 \times \Omega_1$.*

Because the true coupling $P_{0,1}$ should yield a high measure for pairs (x_0, x_1) sharing the same anatomy, the underlying function g must preserve this anatomy. Furthermore, we assume that distinct elements in Ω_0 correspond to distinct anatomies, and all anatomies in Ω_1 exist in Ω_0. This excludes non-injective mappings (which would map distinct source anatomies to a single target point) and non-surjective mappings (where target points lack a source preimage). This bijective nature, combined with the reasonable assumption of differentiability for both g and g^{-1}, motivates our second assumption:

Assumption 3.2. ***(Diffeomorphism)*** *The map g is a diffeomorphism between two differentiable manifolds Ω_0 and Ω_1 equipped with the standard topology and a maximal differentiable atlas.*

The map g induces a pushforward measure $P_1 = g_\# P_0$ (Fig. 3, probability measure mappings). For probability density functions, $p_1 = (p_0 \circ g^{-1})|\det(dg^{-1})|$. If $|\det(dg)| \neq 1$, the pullback g^* scales the volume forms, effectively causing anatomies in Ω_1 to become more concentrated or dispersed than in Ω_0. This is undesirable, because the true anatomical distribution should remain independent of the imaging contrast. If a specific contrast forces anatomies to artificially concentrate in one region, it degrades the capability to distinguish between different anatomical structures. Therefore, we enforce:

Assumption 3.3. ***(Volume Preservation)*** *The map g induces local linear maps $dg : T\Omega_0 \to T\Omega_1$ that preserve volume, meaning $|\det(dg)| = 1$, where dg refers to its matrix representation.*

However, solely preserving volume is insufficient. There are infinitely many volume-preserving maps where dg severely stretches or shrinks tangent vectors along different directions (Fig. 3, tangent bundle mappings). Under such mappings, metric tensor fields are not preserved under the pullback g^*, meaning anatomical distances are distorted (e.g., two distinct anatomies that are far apart in Ω_0 might falsely appear close in Ω_1). While strictly preserving the Euclidean metric within the same dimension restricts g to rigid transformations which are overly restrictive, we instead use local isometry as a soft structural regularizer to encourage g to be a local isometry of the Euclidean metric space:

Assumption 3.4. ***(Approximate Isometry)*** *The map g induces local linear maps dg that are as close to being isometries of the Euclidean metric space as possible, meaning $dg^\mathsf{T} dg \approx \mathrm{id}$, where id is the identity matrix and dg refers to its matrix representation.*

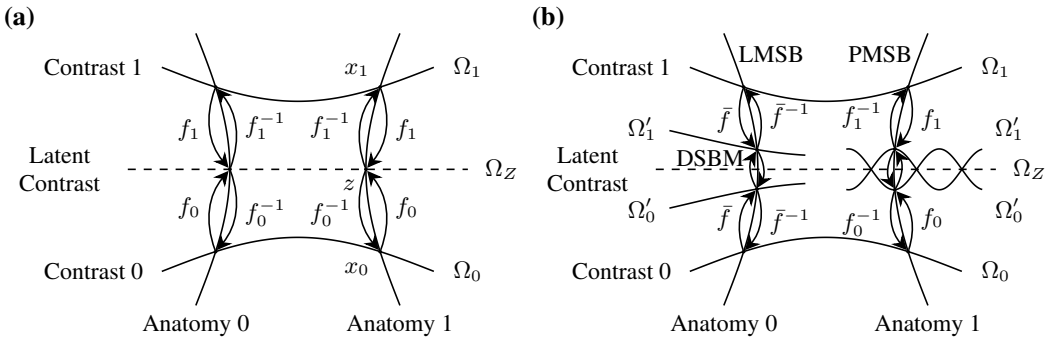

Figure 4: **(a) Construction of latent variable.** Images $x_0 \in \Omega_0$ and $x_1 \in \Omega_1$ share anatomy (vertical curves) but differ in contrast (horizontal curves). Ideally, paired samples map to an identical latent z. In practice, using segmentation maps as a proxy z omits fine anatomical details, so the exact mapping $f_0(x_0) = f_1(x_1) = z$ cannot be achieved. **(b) Comparison of LMSB and PMSB.** LMSB (left) restricts $f_0 = f_1 = \bar{f}$, inducing a strict metric space. PMSB (right) uses independent f_0 and f_1, inducing a pseudometric space that allows for closer domain overlap. Because practical ranges Ω_0' and Ω_1' do not perfectly align with Ω_Z, both methods apply DSBM between them to find the optimal coupling $P_{0,1}'$ for image translation (Fig. 2).

3.3 Learning Objectives

We demonstrate (see Appendix A.1) that satisfying the OT goal in Eq. 1 is mathematically equivalent to formulating a new OT problem over the original domains:

$$P_{0,1} = \underset{P \in \Gamma(P_0, P_1)}{\arg\min} \int_{\Omega_{0,1}} \|f_1(x_1) - f_0(x_0)\|^2 P(dx_0, dx_1). \tag{2}$$

Finding analytical forms for f_0 and f_1 to perfectly satisfy Eq. 2 is difficult. However, under Assumption 3.1, we can encourage $P_{0,1}$ to be the OT plan of Eq. 2 by learning f_0 and f_1 through minimizing the following paired objective (proof in Appendix A.2):

$$\underset{f_0, f_1}{\arg\min} \int_{\Omega_{0,1}} \|f_1(x_1) - f_0(x_0)\|^2 P_{0,1}(dx_0, dx_1). \tag{3}$$

Because paired data from $P_{0,1}$ is generally unavailable in practice, we must find an alternative way to optimize this. By introducing a latent random variable Z taking values $z \in \Omega_Z$ and applying the triangle inequality, we can construct a strict upper bound for the paired objective. As detailed in Appendix A.3, minimizing this upper bound allows us to effectively minimize the original paired objective without ever requiring paired samples. Crucially, this upper bound mathematically decouples the joint formulation into two independent, unpaired learning objectives:

$$\underset{f_i}{\arg\min} \int_{\Omega_{i,Z}} \|f_i(x_i) - z\|^2 P_{i,Z}(dx_i, dz), \quad \text{for } i \in \{0, 1\}. \tag{4}$$

3.4 Constructing Latent Variable

To ensure that the sum of our two unpaired objectives in Eq. 4 remains a strict, mathematically valid upper bound to the paired objective in Eq. 3, we require the conditional measures to be invariant across domains: $P_{Z|0}(\cdot|x_0) = P_{Z|1}(\cdot|x_1)$ for any true pair $(x_0, x_1) \sim P_{0,1}$. As proven in Appendix A.4, this identity is a natural consequence of our geometrical assumptions. Under a deterministic coupling (Assumption 3.1), $x_1 = g(x_0)$, meaning x_0 and x_1 represent the exact same underlying anatomy viewed under different contrasts. Because g is a diffeomorphism (Assumption 3.2), it is a bijection, making the σ-algebras generated by the two domains identical. Therefore, conditioning on x_0 provides the exact same information as conditioning on x_1.

A sufficient condition to satisfy this conditional identity is to construct z such that it perfectly isolates this shared anatomy, stripping away domain-specific contrast. Visually, z represents the invariant

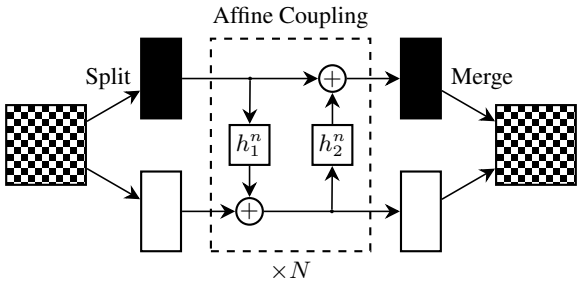

Figure 5: Network architecture for constructing diffeomorphisms. The input is partitioned using a checkerboard decomposition, processed sequentially through N affine coupling layers parameterized by differentiable functions h_1^n and h_2^n, and finally merged to produce the output.

anatomy rendered in a domain-agnostic latent contrast (Fig. 4 (a)). Without loss of generality, we can simplify this target by structuring $P_{i,Z}$ as a deterministic coupling. Since the optimal solution $f_i(x_i)$ for the regression in Eq. 4 computes the conditional mean of z, it is mathematically equivalent to assuming z follows a Dirac measure centered at this mean. Consequently, for any paired sample $(x_0, x_1) \sim P_{0,1}$, the target latent variable z must be identically the same coordinate in Ω_Z.

To strictly satisfy Assumptions 3.2–3.4, the latent space Ω_Z must be a continuous manifold diffeomorphic to Ω_0 and Ω_1, preserving both probability density and relative anatomical distances. Because constructing this ideal manifold directly from complex medical images is practically intractable, we approximate Ω_Z using discrete segmentation maps. While these identical labels provide an effective, domain-invariant structural anchor, this discrete proxy inherently omits fine-grained anatomical details. Consequently, a perfect spatial mapping $f_0(x_0) = f_1(x_1) = z$ cannot be realized for all microscopic features. This limitation naturally motivates the need to strictly constrain the function space. Intuitively, if f_0 and f_1 are rigidly restricted to be volume-preserving and an approximate local isometry, they cannot arbitrarily distort, collapse, or hallucinate local information.

3.5 Constraining Function Space

To enforce these necessary geometric constraints, the assumptions established for the true mapping g must logically extend to f_0 and f_1. To guarantee that f_0 and f_1 are volume-preserving diffeomorphisms (Assumptions 3.2 and 3.3), we implement them using an invertible affine coupling architecture inspired by normalizing flows (Rezende and Mohamed, 2015; Dinh et al., 2015, 2017; Kingma and Dhariwal, 2018; Stimper et al., 2023). As depicted in Fig. 5, the network is composed of N reversible layers characterized by differentiable blocks h_1^n and h_2^n. By construction, the determinant of the local linear map dh for the overall network h is strictly 1:

$$\det(dh) = \prod_{n=1}^{N} \det \begin{pmatrix} \mathrm{id} & dh_2^n \\ 0 & \mathrm{id} \end{pmatrix} \det \begin{pmatrix} \mathrm{id} & 0 \\ dh_1^n & \mathrm{id} \end{pmatrix} = 1. \tag{5}$$

To enforce approximate isometry (Assumption 3.4), we inject Gaussian noise $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$ into the input images in Eq. 4, which yields the following:

$$\arg\min_{f_i} \int_{\Omega_{i,Z,E}} \|f_i(x_i + \epsilon) - z\|^2 P_{i,Z,E}(dx_i, dz, d\epsilon), \quad \text{for } i \in \{0, 1\}, \tag{6}$$

where $P_{i,Z,E} = P_{i,Z} \otimes P_E$. We prove in Appendix A.5 that this noise injection effectively acts as a trace penalty on $df_i^\mathsf{T} df_i$, encouraging df_i to be isometries of the Euclidean metric space. Ultimately, while the constructed latent target z drives macroscopic alignment, these strict volume-preserving and approximate local isometry constraints force the mappings to independently preserve anatomical details, effectively compensating for the lack of resolution in discrete segmentation maps.

3.6 Implication of Pullback Metrics

PMSB deliberately relies on independent mappings, f_0 and f_1. To demonstrate why this separation minimizes the transport gap more effectively than shared-mapping frameworks like LMSB, we

analyze their induced pullback metrics, revealing profound topological differences in the resulting latent space. Assume the domains are disjoint ($\Omega_0 \cap \Omega_1 = \emptyset$). Let $f : \Omega_0 \cup \Omega_1 \to \Omega_0' \cup \Omega_1'$ be a piecewise mapping defined as:

$$f(x) = \begin{cases} f_0(x), & \text{if } x \in \Omega_0, \\ f_1(x), & \text{if } x \in \Omega_1. \end{cases} \tag{7}$$

The LMSB approach explicitly restricts $f_0 = f_1 = \bar{f}$ as illustrated in Fig. 4 (b). Given that $(\Omega_0' \cup \Omega_1', d')$ is a metric space equipped with metric d', the pullback metric $d = f^* d'$ guarantees the formation of a strict metric space over $(\Omega_0 \cup \Omega_1, d)$ (proven in Appendix A.6). Under this strict metric, images with identical anatomy but different contrasts remain distinguishable and non-overlapping.

In contrast, PMSB permits independent mappings f_0 and f_1. Because $f_0 \neq f_1$, the resulting pullback metric $d = f^* d'$ relaxes the identity of indiscernibles, thereby inducing a pseudometric space (Appendix A.6). This fundamental theoretical distinction means that $x_0 \in \Omega_0$ and $x_1 \in \Omega_1$ with identical anatomy can explicitly map to the exact same latent coordinate ($f_0(x_0) = f_1(x_1)$) as illustrated in Fig. 4 (b).

By relaxing the metric space into a pseudometric space, PMSB dramatically reduces the transportation gap between Ω_0' and Ω_1'. Because LMSB's restrictive search space ($f_0 = f_1$) is merely a strict subset of PMSB's search space, PMSB inherently achieves a lower bound for the transport objective in Eq. 3, mathematically validating its superior performance in reducing the domain gaps.

4 Experiments

We evaluate PMSB on two MRI translation settings constructed from the publicly available OASIS-3 dataset (LaMontagne et al., 2019). In both settings, Ω_0 is the domain of T_1-weighted (T_1-w) slices, and Ω_1 is the domain of either T_2-weighted (T_2-w) or fluid-attenuated inversion recovery (FLAIR) slices. We strictly select at most one session per subject to ensure all scans are unique.

Dataset. For T_1-w $\leftrightarrow$ T_2-w, we use 402 paired sessions from a 3 T Siemens TrioTim scanner. Both modalities share the same native acquisition matrix ($176 \times 256 \times 256$ voxels, 1 mm isotropic spacing), representing a translation across contrast. For T_1-w $\leftrightarrow$ FLAIR, we use 402 paired sessions from a 3 T Siemens Biograph_mMR scanner. Native T_1-w volumes are $176 \times 240 \times 256$ voxels ($\sim$1.055 mm in-plane, $\sim$1.20 mm through-plane), while FLAIR volumes are highly anisotropic at $256 \times 256 \times 35$ voxels ($\sim$0.86 mm in-plane, $\sim$5 mm slice spacing), representing a translation across both contrast and resolution. Each 402-subject cohort is partitioned into disjoint sets of 334 training, 34 validation, and 34 testing subjects. To construct an unpaired setting during training and validation, subjects in these splits are divided equally between the two domains. For testing, all 34 subjects contribute to both domains. Extracting 30 axial slices from the center of the brain per subject yields 5,010 unpaired training slices, 510 unpaired validation slices, and 1,020 paired testing slices per domain.

Preprocessing. To preserve native domain distributions, we omit the bias-field correction. Raw T_1-w volumes are rigidly aligned to the MNI template ($192 \times 224 \times 192$ voxels, 1 mm isotropic) using ANTs (Avants et al., 2009). T_2-w and FLAIR volumes are then co-registered and resampled onto this MNI-space T_1-w grid. The extracted axial slices are of size 224×192 (height $\times$ width). Finally, slice intensities are truncated at zero (removing negatives), clipped at the 99.9^{th} percentile, and linearly scaled to $[0, 1]$.

Training Diffeomorphism. To train PMSB, we first train the diffeomorphism functions f_0 and f_1 via two independent networks (Fig. 5). We construct the training targets applying SynthSeg (Billot et al., 2023) independently to each modality (i.e., T_1-w, T_2-w, and FLAIR), collapsing its 33 labels into four groups: background, cerebrospinal fluid (CSF), grey matter (GM), and white matter (WM). To preserve unsegmented background structures (e.g., muscle, fat, skull), we retain the original image intensities in the background of the training targets, while assigning unique constant values to the CSF, GM, and WM regions (Fig. 6, Targets). Although SynthSeg does not capture all fine anatomical details, these structures remain well-preserved in the mapped outputs (Fig. 6, Outputs). This preservation is encouraged by the function space constraints (Assumptions 3.2–3.4), the anatomical benefits of which align with findings in LMSB (Wei et al., 2025).

Training PMSB. Once f_0 and f_1 are trained, we train a DSBM between the mapped domains Ω_0' and Ω_1' (Fig. 4 (b), PMSB). Operating in this latent space is substantially more effective than

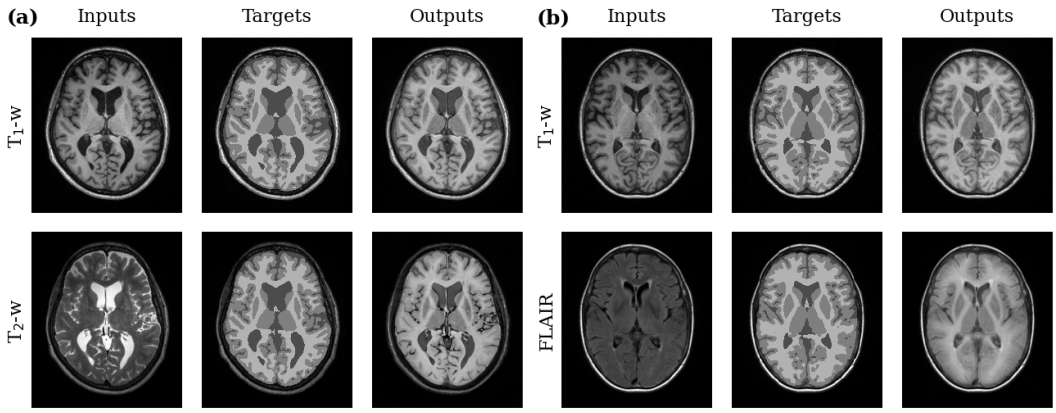

Figure 6: Training the diffeomorphism functions f_0 and f_1 for **(a) T_1-w and T_2-w**, and **(b) T_1-w and FLAIR**. Inputs are original MRI slices, and targets are constructed segmentation maps. The mapped outputs are intermediate latent representations, which exhibit a reduced transportation gap while successfully preserving fine anatomical details absent from the targets.

mapping directly between Ω_0 and Ω_1 because the domain gap is significantly reduced. Following the original DSBM implementation (Shi et al., 2023), we use Iterative Markovian Fitting (IMF) (Shi et al., 2023; Peluchetti, 2023) to solve the Schrödinger Bridge problem over 30 IMF steps. We initialize with independent coupling, running 5,000 forward and backward iterations in the first step. Each subsequent step uses 1,000 forward and backward iterations, storing simulation trajectories every 1,000 iterations. The diffusion process employs 100 steps and a constant diffusion coefficient ($\sigma^2 = 0.01$).

Comparison Methods. We compare PMSB with diverse state-of-the-art unpaired image translation methods: SynDiff (Özbey et al., 2023), which couples diffusive and non-diffusive modules via cycle-consistency; ContourDiff (Chen et al., 2025b), which enforces spatial constraints using domain-invariant contours; and UNSB (Kim et al., 2024), which scales SB to high-resolution data using sequential adversarial learning. We also evaluate DSBM (Shi et al., 2023), a state-of-the-art SB solver for natural images using IMF, and LMSB (Wei et al., 2025), which applies DSBM within an anatomy-guided latent metric space. Because DSBM is the core mechanism for LMSB and PMSB, we ensure a fair comparison by using identical training parameters.

Results. Quantitative evaluations of the translation methods are presented in Table 1. To evaluate structural alignment, we use the peak signal-to-noise ratio (PSNR) and the structural similarity index measure (SSIM), alongside a squared local normalized cross-correlation (LNCC), a standard learning objective in image registration (Chen et al., 2025a). These three structural metrics are calculated after applying a brain mask obtained by HD-BET (Isensee et al., 2019). The data show that PMSB consistently outperforms the compared baselines in structural preservation, achieving the highest PSNR, SSIM, and LNCC in almost all tasks. While DSBM achieves the lowest overall Fréchet inception distance (FID), LMSB and PMSB exhibit slightly worse FID scores because they do not translate the two domains directly. However, this yields superior structure preservation which is more important for medical images than FID. This trade-off aligns with findings previously reported in LMSB (Wei et al., 2025). This quantitative observation is heavily supported by the qualitative comparison in Figure 7. Visual inspections reveal that PMSB generates images closest to the ground truth (GT), successfully preserving complex anatomical structures and fine details. In contrast, methods like UNSB and DSBM suffer from severe structural distortions and hallucinate brain tissues. This degradation is particularly evident in the T_1-w $\rightarrow$ T_2-w and T_2-w $\rightarrow$ T_1-w translations, because their domain gap is substantially larger than the gap between T_1-w and FLAIR. Although SynDiff and ContourDiff maintain better overall structural integrity than UNSB and DSBM, they still fall short of the structural fidelity achieved by PMSB.

Sampling Strategies. Note that in Table 1 and Figure 7, we use the ordinary differential equation (ODE) sampler for DSBM, LMSB, and PMSB. A systematic ablation comparing the stochastic differential equation (SDE) and ODE samplers is provided in Appendix B.1. This comparison

Table 1: Quantitative comparison ($N = 1,020$) of translation methods across different contrasts. Performance is evaluated using a domain consistency metric (FID) alongside three structural consistency metrics (PSNR, SSIM, and LNCC). SSIM and LNCC are percentages. For DSBM, LMSB, and PMSB, the ODE sampler is used. Bold number indicates the best result within each task and metric.

Metric	Task	SynDiff	ContourDiff	UNSB	DSBM	LMSB	PMSB
PSNR ↑	T_1-w → T_2-w	21.84 ± 1.11	22.14 ± 1.03	19.04 ± 1.25	16.17 ± 1.02	18.79 ± 1.15	**25.03** ± 1.34
	T_2-w → T_1-w	24.69 ± 1.04	25.49 ± 1.37	21.45 ± 1.14	18.42 ± 1.08	20.98 ± 1.42	**27.20** ± 2.04
	T_1-w → FLAIR	29.19 ± 1.34	26.57 ± 1.06	29.73 ± 1.41	**29.81** ± 1.78	28.99 ± 2.54	29.40 ± 1.79
	FLAIR → T_1-w	23.08 ± 1.02	23.67 ± 1.63	24.20 ± 0.90	24.59 ± 1.38	24.38 ± 2.37	**25.18** ± 1.24
	Overall	24.70 ± 3.21	24.47 ± 1.96	23.60 ± 4.60	22.25 ± 6.17	23.29 ± 4.44	**26.70** ± 2.05
SSIM ↑	T_1-w → T_2-w	77.35 ± 2.37	82.38 ± 2.11	65.60 ± 3.57	53.54 ± 4.02	61.14 ± 3.93	**88.71** ± 2.14
	T_2-w → T_1-w	82.27 ± 1.65	85.93 ± 1.74	71.21 ± 2.42	53.82 ± 4.13	65.11 ± 4.17	**90.88** ± 1.66
	T_1-w → FLAIR	83.46 ± 2.33	79.57 ± 2.95	84.76 ± 1.81	85.22 ± 2.01	**85.40** ± 2.41	**85.40** ± 2.05
	FLAIR → T_1-w	75.02 ± 1.90	78.38 ± 2.75	78.92 ± 3.85	77.85 ± 4.58	78.66 ± 4.73	**79.22** ± 3.98
	Overall	79.53 ± 4.00	81.57 ± 3.36	75.12 ± 8.43	67.61 ± 16.36	72.58 ± 11.37	**86.05** ± 5.08
LNCC ↑	T_1-w → T_2-w	15.69 ± 1.45	22.01 ± 2.06	10.74 ± 1.00	13.62 ± 1.37	12.81 ± 1.06	**27.87** ± 2.45
	T_2-w → T_1-w	20.98 ± 1.99	25.38 ± 2.26	13.43 ± 1.11	14.13 ± 1.44	14.39 ± 1.06	**30.83** ± 2.57
	T_1-w → FLAIR	13.93 ± 1.76	16.61 ± 2.50	16.44 ± 2.34	16.79 ± 2.50	**17.50** ± 2.63	17.38 ± 2.69
	FLAIR → T_1-w	14.82 ± 1.76	18.07 ± 2.91	19.54 ± 3.02	19.00 ± 2.97	**19.75** ± 3.12	19.73 ± 3.05
	Overall	16.35 ± 3.16	20.52 ± 3.96	15.04 ± 3.80	15.89 ± 2.50	16.11 ± 3.11	**23.95** ± 6.42
FID ↓	T_1-w → T_2-w	55.79	32.39	18.98	**18.71**	34.69	33.23
	T_2-w → T_1-w	32.65	33.54	19.94	**17.42**	21.96	21.42
	T_1-w → FLAIR	49.73	54.27	30.49	**16.39**	20.45	19.65
	FLAIR → T_1-w	57.76	46.17	25.65	**16.32**	21.35	20.69
	Overall	48.98 ± 11.41	41.59 ± 10.51	23.77 ± 5.36	**17.21** ± 1.12	24.61 ± 6.75	23.75 ± 6.36

highlights a fundamental trade-off: SDE sampling yields better FID, whereas ODE sampling achieves better structural fidelity that is paramount for medical image translation.

Sensitivity to Spatial Registration. To ensure that PMSB's translation capabilities do not strictly depend on pre-solved spatial alignment, we assessed the model's performance without rigid MNI template registration in Appendix B.2. PMSB maintains highly comparable structural consistency across unregistered intra-subject slices. This confirms that while MNI alignment is a standard neuroimaging preprocessing step, it is not a prerequisite for PMSB, supporting the method's viability in general unpaired settings.

Segmentation-Target Robustness. Because PMSB uses segmentation maps as a practical proxy for the latent anatomical target z, we evaluated its sensitivity to the quality of these auxiliary targets in Appendix B.3. By artificially degrading the SynthSeg targets via morphological dilation, we observed only marginal decreases in structural metrics. Constrained by the volume-preserving and approximate isometry regularizations, PMSB successfully preserves the true anatomical details, demonstrating strong robustness to moderate auxiliary segmentation errors.

Latent-Domain Gap Analysis. A core theoretical advantage of PMSB over LMSB is the relaxation of the shared-mapping constraint into independent, domain-specific mappings. To empirically validate that this pseudometric formulation more effectively reduces the transport gap, we analyzed the inter-domain latent distances in Appendix C. The results confirm that PMSB achieves a smaller latent gap than LMSB. Notably, the absolute reduction is most significant in tasks with large initial image-space domain gaps, such as T_1-w ↔ T_2-w, which explains the asymmetric performance gains in Table 1 and validates our theoretical claims.

Generalizability to Pathological Anatomy. To demonstrate generalizability beyond the aging and Alzheimer's cohorts in OASIS-3 (LaMontagne et al., 2019), we expanded our evaluation to the BraTS20 dataset (Menze et al., 2014; Bakas et al., 2017, 2018) in Appendix D, which features severe tumor-induced anatomical deformations. PMSB consistently outperformed state-of-the-art baselines in structural metrics across both translation directions on this dataset, demonstrating robust generalizability to domains with significant pathological abnormalities.

Computational Costs. We report the training and per-slice inference time of all evaluated methods in Appendix E. While the multi-stage training of the diffeomorphisms and the DSBM requires a substantial initial investment, PMSB efficiently models both translation directions within a single

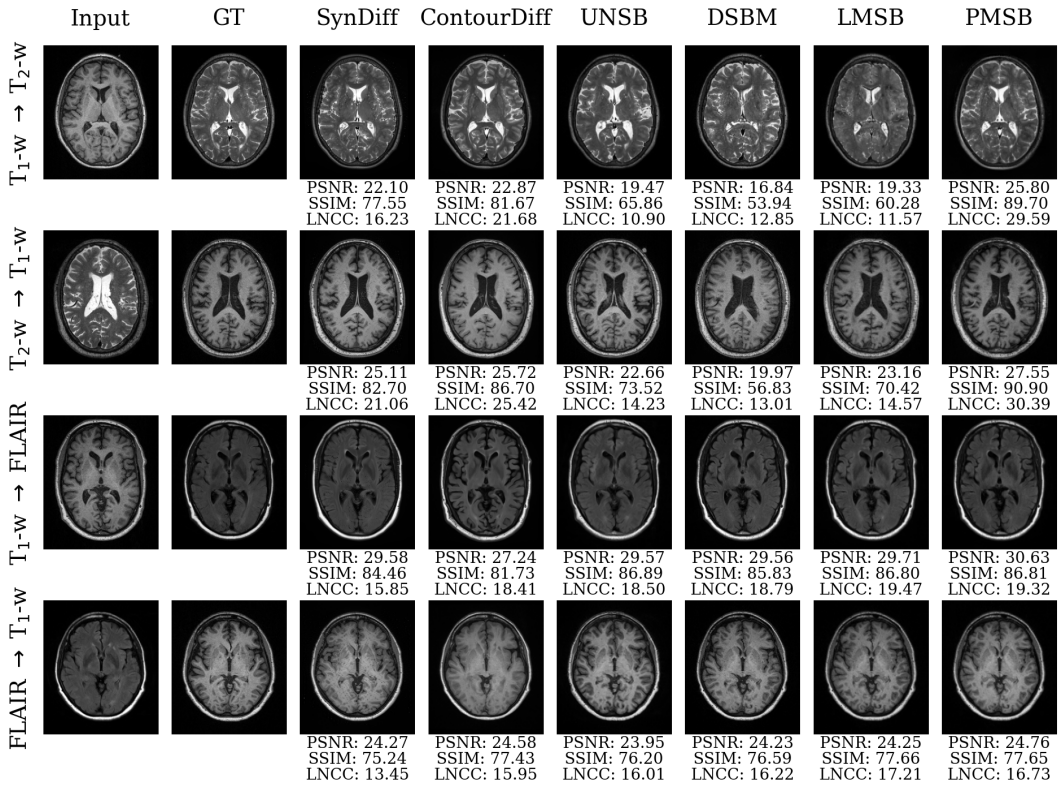

Figure 7: Qualitative image translation comparison across contrasts. Rows (top-bottom): T_1-w $\rightarrow$ T_2-w, T_2-w $\rightarrow$ T_1-w, T_1-w $\rightarrow$ FLAIR, and FLAIR $\rightarrow$ T_1-w. Columns (left-right): Input, Ground truth (GT), and outputs from SynDiff, ContourDiff, UNSB, DSBM, LMSB, and PMSB (with individual PSNR, SSIM, LNCC scores below each).

training phase, unlike baselines such as ContourDiff and UNSB which require separate models per direction.

5 Conclusion

In this paper, we introduced the pseudometric Schrödinger bridge (PMSB) framework, which advances unpaired MR image translation by generalizing the latent metric space into a latent pseudometric space. By using independent, domain-specific mappings, PMSB theoretically allows disparate contrasts of identical anatomy to map to the exact same latent coordinate, explicitly minimizing the transport gap. We grounded this approach in a rigorous theoretical foundation, proving the validity of our unpaired learning objectives under a set of geometrical assumptions.

Despite achieving state-of-the-art structural consistency, practical limitations remain before clinical deployment. Foremost, PMSB currently operates on 2D slices. Scaling the diffeomorphic mappings and diffusion process from 2D slices to full 3D volumetric MRI data is a critical next step to prevent out-of-plane inconsistencies. Furthermore, while PMSB demonstrably mitigates anatomical shifts, its ultimate clinical utility must be rigorously validated through downstream tasks like segmentation.

Finally, this work opens several theoretical avenues for future research. While our rigid geometric constraints successfully compensate for the discrete, segmentation-based proxy of the latent variable z, developing theoretically optimal strategies for directly constructing a continuous, domain-invariant latent manifold remains an important goal. Additionally, alternative invertible architectures should be explored to natively avoid the checkerboard artifacts introduced by the current spatial decomposition, removing the reliance on post-processing (bilinear upsampling and average pooling). Lastly, while we mathematically guarantee that PMSB inherently bounds and reduces anatomical shift, theoretically quantifying the exact magnitude of this remaining shift stands as an open and compelling problem.

Acknowledgments and Disclosure of Funding

This work is supported by the NIH through the NIBIB grant R01EB036013 (PI: J.L. Prince) and a Johns Hopkins University Discovery Grant (PI: A. Carass).

References

Anderson, B. D. O. (1982). Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326.

Arjovsky, M., Chintala, S., and Bottou, L. (2017). Wasserstein generative adversarial networks. In *34nd International Conference on Machine Learning (ICML 2017)*, pages 214–223. PMLR.

Avants, B. B., Tustison, N., Song, G., et al. (2009). Advanced normalization tools (ANTS). *Insight j*, 2(365):1–35.

Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J. S., Freymann, J. B., Farahani, K., and Davatzikos, C. (2017). Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific Data*, 4(1):170117.

Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., Shinohara, R. T., Berger, C., Ha, S. M., Rozycki, M., et al. (2018). Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. *arXiv preprint arXiv:1811.02629*.

Biberacher, V., Schmidt, P., Keshavan, A., Boucard, C. C., Righart, R., Sämann, P., Preibisch, C., Fröbel, D., Aly, L., Hemmer, B., et al. (2016). Intra-and interscanner variability of magnetic resonance imaging based volumetry in multiple sclerosis. *NeuroImage*, 142:188–197.

Billot, B., Greve, D. N., Puonti, O., Thielscher, A., Van Leemput, K., Fischl, B., Dalca, A. V., Iglesias, J. E., et al. (2023). SynthSeg: Segmentation of brain MRI scans of any contrast and resolution without retraining. *Medical Image Analysis*, 86:102789.

Bortoli, V. D., Thornton, J., Heng, J., and Doucet, A. (2021). Diffusion Schrödinger bridge with applications to score-based generative modeling. In *Advances in Neural Information Processing Systems*, volume 34, pages 17695–17709.

Brown, R. W., Cheng, Y.-C. N., Haacke, E. M., Thompson, M. R., and Venkatesan, R. (2014). *Magnetic Resonance Imaging: Physical Principles and Sequence Design*. John Wiley & Sons.

Cackowski, S., Barbier, E. L., Dojat, M., and Christen, T. (2023). ImUnity: A generalizable VAE-GAN solution for multicenter MR image harmonization. *Medical Image Analysis*, 88(102799).

Chen, J., Liu, Y., Wei, S., Bian, Z., Subramanian, S., Carass, A., Prince, J. L., and Du, Y. (2025a). A survey on deep learning in medical image registration: New technologies, uncertainty, evaluation metrics, and beyond. *Medical Image Analysis*, 100:103385.

Chen, Y., Konz, N., Gu, H., Dong, H., Chen, Y., Li, L., Lee, J., and Mazurowski, M. A. (2025b). ContourDiff: Unpaired Medical Image Translation with Structural Consistency. *Machine Learning for Biomedical Imaging*, 3:711–727.

Cuturi, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, volume 26.

Dewey, B. E., Zhao, C., Reinhold, J. C., Carass, A., Fitzgerald, K. C., Sotirchos, E. S., Saidha, S., Oh, J., Pham, D. L., Calabresi, P. A., et al. (2019). DeepHarmony: A deep learning approach to contrast harmonization across scanner changes. *Magnetic Resonance Imaging*, 64:160–170.

Dinh, L., Krueger, D., and Bengio, Y. (2015). NICE: Non-linear independent components estimation. In *The 3rd International Conference on Learning Representations (ICLR) 2015*.

Dinh, L., Sohl-Dickstein, J., and Bengio, S. (2017). Density estimation using Real NVP. In *The 5rd International Conference on Learning Representations (ICLR) 2017*.

Föllmer, H. (1988). Random fields and diffusion processes. In *École d'Été de Probabilités de Saint-Flour XV–XVII, 1985–87. Lecture Notes in Mathematics*, volume 1362, pages 101–203.

Goodfellow, I. et al. (2020). Generative Adversarial Networks. *Communications of the ACM*, 63(11):139–144.

Guan, H. and Liu, M. (2022). Domain Adaptation for Medical Image Analysis: A Survey. *IEEE Transactions on Biomedical Engineering*, 69(3):1173–1185.

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. C. (2017). Improved training of Wasserstein GANs. In *Advances in Neural Information Processing Systems*, volume 30.

Hiasa, Y., Otake, Y., Takao, M., Matsuoka, T., Takashima, K., Carass, A., Prince, J. L., Sugano, N., and Sato, Y. (2018). Cross-modality image synthesis from unpaired data using CycleGAN. In *Workshop on Simulation and Synthesis in Medical Imaging (SASHIMI) held in conjunction with the 21st International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2018)*, volume 11037 of *Lecture Notes in Computer Science*, pages 31–41. Springer Berlin Heidelberg.

Ho, J., Jain, A., and Abbeel, P. (2020). Denoising Diffusion Probabilistic Models. In *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851.

Huang, X., Liu, M.-Y., Belongie, S., and Kautz, J. (2018). Multimodal unsupervised image-to-image translation. In *European Conference on Computer Vision*, pages 179–196. Springer.

Isensee, F., Schell, M., Pflueger, I., Brugnara, G., Bonekamp, D., Neuberger, U., Wick, A., Schlemmer, H.-P., Heiland, S., Wick, W., Bendszus, M., Maier-Hein, K. H., and Kickingereder, P. (2019). Automated brain extraction of multisequence MRI using artificial neural networks. *Human Brain Mapping*, 40(17):4952–4964.

Kantorovich, L. V. (1940). On an efficient method of solving some classes of extremal problems. *Doklady Akademii Nauk SSSR*, 28:212–215.

Kantorovich, L. V. (1942). On the Translocation of Masses. *Doklady Akademii Nauk SSSR*, 37(7–8):227–230.

Kim, B., Kwon, G., Kim, K., and Ye, J. C. (2024). Unpaired Image-to-Image Translation via Neural Schrödinger Bridge. In *The 12th International Conference on Learning Representations (ICLR) 2024*.

Kingma, D. P. and Dhariwal, P. (2018). Glow: Generative flow with invertible 1×1 convolutions. In *Advances in Neural Information Processing Systems*, volume 31.

Kullback, S. (1968). Probability densities with given marginals. *The Annals of Mathematical Statistics*, 39(4):1236–1243.

LaMontagne, P. J., Benzinger, T. L. S., Morris, J. C., Keefe, S., Hornbeck, R., Xiong, C., Grant, E., Hassenstab, J., Moulder, K., Vlassenko, A. G., et al. (2019). OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease. *medrxiv*, pages 2019–12.

Léonard, C. (2014). A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete and Continuous Dynamical Systems*, 34(4):1533–1574.

Liu, M.-Y., Breuel, T., and Kautz, J. (2017). Unsupervised image-to-image translation networks. In *Advances in Neural Information Processing Systems*, volume 30.

Liu, X., Xing, F., Prince, J. L., Carass, A., Stone, M., El Fakhri, G., and Woo, J. (2021). Dual-Cycle Constrained Bijective Vae-Gan For Tagged-To-Cine Magnetic Resonance Image Synthesis. In *18th International Symposium on Biomedical Imaging (ISBI 2021)*, pages 1448–1452.

Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al. (2014). The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Medical Imaging*, 34(10):1993–2024.

Mirzaalian, H., Ning, L., Savadjiev, P., Pasternak, O., Bouix, S., Michailovich, O., Grant, G., Marx, C. E., Morey, R. A., Flashman, L. A., et al. (2016). Inter-site and inter-scanner diffusion MRI data harmonization. *NeuroImage*, 135:311–323.

Miyato, T. et al. (2018). Spectral Normalization for Generative Adversarial Networks. In *The 6rd International Conference on Learning Representations (ICLR) 2018*.

Monge, G. (1781). Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences de Paris*, pages 666–704.

Özbey, M., Dalmaz, O., Dar, S. U., Bedel, H. A., Özturk, Ş., Güngör, A., and Cukur, T. (2023). Unsupervised medical image translation with adversarial diffusion models. *IEEE Transactions on Medical Imaging*, 42(12):3524–3539.

Park, T., Efros, A. A., Zhang, R., and Zhu, J.-Y. (2020). Contrastive learning for unpaired image-to-image translation. In *European Conference on Computer Vision*, pages 319–345.

Peluchetti, S. (2023). Diffusion bridge mixture transports, Schrödinger bridge problems and generative modeling. *Journal of Machine Learning Research*, 24(374):1–51.

Prince, J. L. and Links, J. M. (2015). *Medical Imaging Signals and Systems*. Pearson.

Rezende, D. and Mohamed, S. (2015). Variational inference with normalizing flows. In *32nd International Conference on Machine Learning (ICML 2015)*, pages 1530–1538.

Rüschendorf, L. (1995). Convergence of the iterative proportional fitting procedure. *The Annals of Statistics*, pages 1160–1174.

Rüschendorf, L. and Thomsen, W. (1993). Note on the Schrödinger equation and I-projections. *Statistics & Probability Letters*, 17(5):369–375.

Sasaki, H., Willcocks, C. G., and Breckon, T. P. (2021). UNIT-DDPM: Unpaired image translation with denoising diffusion probabilistic models. *arXiv preprint arXiv:2104.05358*.

Schrödinger, E. (1931). *Über die Umkehrung der Naturgesetze*. Sitzungsberichte der Preussischen Akademie der Wissenschaften. Physikalisch-mathematische Klasse. Verlag der Akademie der Wissenschaften in Kommission bei Walter De Gruyter u. Company.

Schrödinger, E. (1932). Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. *Annales de l'institut Henri Poincaré*, 2(4):269–310.

Shi, Y., De Bortoli, V., Campbell, A., and Doucet, A. (2023). Diffusion Schrödinger Bridge Matching. In *Advances in Neural Information Processing Systems*, volume 36, pages 62183–62223.

Sinkhorn, R. (1967). Diagonal equivalence to matrices with prescribed row and column sums. *The American Mathematical Monthly*, 74(4):402–405.

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., and Ganguli, S. (2015). Deep unsupervised learning using nonequilibrium thermodynamics. In *32nd International Conference on Machine Learning (ICML 2015)*, pages 2256–2265.

Song, J., Meng, C., and Ermon, S. (2021a). Denoising Diffusion Implicit Models. In *The 9th International Conference on Learning Representations (ICLR) 2021*.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. (2021b). Score-Based Generative Modeling through Stochastic Differential Equations. In *The 9th International Conference on Learning Representations (ICLR) 2021*.

Stimper, V., Liu, D., Campbell, A., Berenz, V., Ryll, L., Schölkopf, B., and Hernández-Lobato, J. M. (2023). normflows: A PyTorch Package for Normalizing Flows. *Journal of Open Source Software*, 8(86):5361.

Su, X., Song, J., Meng, C., and Ermon, S. (2023). Dual Diffusion Implicit Bridges for Image-to-Image Translation. In *The 11th International Conference on Learning Representations (ICLR) 2023*.

Tian, D., Zeng, Z., Sun, X., Tong, Q., Li, H., He, H., Gao, J.-H., He, Y., and Xia, M. (2022). A deep learning-based multisite neuroimage harmonization framework established with a traveling-subject dataset. *NeuroImage*, 257:119297.

Villani, C. (2008). *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg.

Wei, S., Remedios, S. W., Dewey, B. E., Bian, Z., Wang, S., Chen, J., Jedynak, B. M., Saidha, S., Calabresi, P. A., Carass, A., and Prince, J. L. (2025). Optical Coherence Tomography Harmonization with Anatomy-Guided Latent Metric Schrödinger Bridges. In *Advances in Neural Information Processing Systems*, volume 38, pages 5716–5744.

Wei, S., Remedios, S. W., Dewey, B. E., Bian, Z., Wang, S., Chen, J., Liu, Y., Jedynak, B., Liu, T. Y. A., Saidha, S., Calabresi, P. A., Prince, J. L., and Carass, A. (2026). Optical coherence tomography harmonization via dual diffusion implicit bridges. In *Medical Imaging 2026: Image Processing*, volume 13925, page 139250G.

Westbrook, C., Roth, C. K., and Talbot, J. (2011). *MRI in Practice*. John Wiley & Sons.

Wrobel, J., Martin, M., Bakshi, R., Calabresi, P., Elliot, M., Roalf, D., Gur, R., Gur, R., Henry, R., Nair, G., Oh, J., Papinutto, N., Pelletier, D., Reich, D., Rooney, W., Satterthwaite, T., Stern, W., Prabhakaran, K., Sicotte, N., Shinohara, R., and Goldsmith, J. (2020). Intensity warping for multisite MRI harmonization. *NeuroImage*, 223:117242.

Yang, H., Sun, J., Carass, A., Zhao, C., Lee, J., Prince, J. L., and Xu, Z. (2020). Unsupervised MR-to-CT Synthesis Using Structure-Constrained CycleGAN. *IEEE Transactions on Medical Imaging*, 39(12):4249–4261.

Yi, Z., Zhang, H., Tan, P., and Gong, M. (2017). DualGAN: Unsupervised Dual Learning for Image-to-Image Translation. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2868–2876.

Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2242–2251.

Zuo, L., Dewey, B. E., Carass, A., He, Y., Shao, M., Reinhold, J. C., and Prince, J. L. (2020). Synthesizing realistic brain MR images with noise control. In *Workshop on Simulation and Synthesis in Medical Imaging (SASHIMI) held in conjunction with the 23st International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2020)*, pages 21–31. Springer.

Zuo, L., Dewey, B. E., Liu, Y., He, Y., Newsome, S. D., Mowry, E. M., Resnick, S. M., Prince, J. L., and Carass, A. (2021). Unsupervised MR harmonization by learning disentangled representations using information bottleneck theory. *NeuroImage*, 243:118569.

Zuo, L., Liu, Y., Xue, Y., Dewey, B. E., Remedios, S. W., Hays, S. P., Bilgel, M., Mowry, E. M., Newsome, S. D., Calabresi, P. A., et al. (2023). HACA3: A unified approach for multi-site MR image harmonization. *Computerized Medical Imaging and Graphics*, 109:102285.

Zuo, L., Liu, Y., Xue, Y., Han, S., Bilgel, M., Resnick, S. M., Prince, J. L., and Carass, A. (2022). Disentangling a single MR modality. In *Workshop on Data Augmentation, Labelling, and Imperfections (DALI) held in conjunction with the 25st International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2022)*, pages 54–63. Springer.

Appendices

A Proofs **16**

A.1 Proof of Equivalent Optimal Transport 16

A.2 Proof of Paired Learning Objective 16

A.3 Proof of Unpaired Learning Objective 17

A.4 Proof of Latent Variable Conditional Measure 17

A.5 Proof of Isometry Regularization 18

A.6 Proof of Metric and Pseudometric Spaces 19

B Ablation Study **20**

B.1 SDE and ODE Samplers 20

B.2 Sensitivity to MNI Registration 20

B.3 Segmentation-Target Robustness 21

C Latent-Domain Gap Analysis **21**

D Generalizability to Pathological Anatomy **22**

E Computational Costs **22**

A Proofs

A.1 Proof of Equivalent Optimal Transport

Our goal is to find the maps f_0 and f_1 such that the pushforward coupling $P'_{0,1}$ is the OT plan between the two pushforward probability spaces P'_0 and P'_1 with the squared Euclidean transport cost, see Fig. 1 for the notations. We can write it as:

$$P'_{0,1} = \underset{P' \in \Gamma(P'_0, P'_1)}{\arg\min} \int_{\Omega'_{0,1}} \|x'_1 - x'_0\|^2 P'(dx'_0, dx'_1), \tag{8}$$

which is the goal in Eq. 1.

Because $P'_{0,1} = (f_0 \times f_1)_{\#} P_{0,1}$, $P'_0 = f_{0\#} P_0$, and $P'_1 = f_{1\#} P_1$, we can rewrite Eq. 8 as:

$$P_{0,1} = \underset{P \in \Gamma(P_0, P_1)}{\arg\min} \int_{\Omega'_{0,1}} \|x'_1 - x'_0\|^2 (f_0 \times f_1)_{\#} P(dx'_0, dx'_1), \tag{9}$$

where we substitute $P' = (f_0 \times f_1)_{\#} P$ and the optimization searches over the joint law space of $\Gamma(P_0, P_1)$ instead of $\Gamma(P'_0, P'_1)$. We can further rewrite Eq. 9 by using the change of variable formula as:

$$P_{0,1} = \underset{P \in \Gamma(P_0, P_1)}{\arg\min} \int_{\Omega_{0,1}} \|f_1(x_1) - f_0(x_0)\|^2 P(dx_0, dx_1), \tag{10}$$

which is the equivalent OT problem in Eq. 2.

A.2 Proof of Paired Learning Objective

To learn the maps f_0 and f_1, we assume that the optimal coupling $P_{0,1}$ in Eq. 10 is a deterministic coupling, i.e., $P_{0,1} \in \Gamma_D(P_0, P_1)$ where $\Gamma_D(P_0, P_1) \subset \Gamma(P_0, P_1)$ is the subset of all deterministic couplings with marginals being P_0 and P_1. For any deterministic coupling, we have Lemma A.1.

Lemma A.1. *$\forall$ deterministic coupling $P_{0,1} \in \Gamma_D(P_0, P_1)$, $\exists f_0, f_1$ such that $(f_0, f_1, P_{0,1})$ jointly minimize the following objective:*

$$\underset{f_0, f_1, P \in \Gamma(P_0, P_1)}{\arg\min} \int_{\Omega_{0,1}} \|f_1(x_1) - f_0(x_0)\|^2 P(dx_0, dx_1), \tag{11}$$

and the minimal objective value strictly equals 0.

Proof. If $P_{0,1} \in \Gamma_D(P_0, P_1)$ is a deterministic coupling, then $\exists g : \Omega_0 \to \Omega_1$ such that $P_{0,1} = (\mathrm{id}, g)_{\#} P_0$, where $\mathrm{id} : \Omega_0 \to \Omega_0$ is the identity function. We substitute $P = P_{0,1}$ in the objective function in Eq. 11, and the objective value becomes:

$$\int_{\Omega_{0,1}} \|f_1(x_1) - f_0(x_0)\|^2 (\mathrm{id}, g)_{\#} P_0(dx_0, dx_1), \tag{12}$$

which further simplifies to the following by using the change of variable formula:

$$\int_{\Omega_0} \|(f_1 \circ g)(x_0) - f_0(x_0)\|^2 P_0(dx_0). \tag{13}$$

If we choose f_0 and f_1 such that $f_1 \circ g = f_0$, then the objective value becomes 0. Because the objective function in Eq. 11 is always greater than or equal 0, such $(f_0, f_1, P_{0,1})$ jointly minimizes this objective function. A trivial solution that exists is $f_0 = g$ and $f_1 = \mathrm{id}$. Therefore, $\exists f_0, f_1$ such that $(f_0, f_1, P_{0,1})$ jointly minimize the objective in Eq. 11. □

Note that although we prove the existence of f_0 and f_1, they are not necessarily unique. Using Lemma A.1, we have Proposition A.1.

Proposition A.1. *$\forall$ deterministic coupling $P_{0,1} \in \Gamma_D(P_0, P_1)$, $\exists f_0, f_1$ such that $P_{0,1}$ is the OT plan of Eq. 10, and f_0, f_1 minimize the following objective:*

$$\underset{f_0, f_1}{\arg\min} \int_{\Omega_{0,1}} \|f_1(x_1) - f_0(x_0)\|^2 P_{0,1}(dx_0, dx_1). \tag{14}$$

Proof. If $P_{0,1} \in \Gamma_D(P_0, P_1)$ is a deterministic coupling, according to Lemma A.1, $\exists f_0, f_1$ such that $(f_0, f_1, P_{0,1})$ jointly minimize the objective in Eq. 11. Therefore, with the existing f_0 and f_1 fixed, $P_{0,1}$ is the OT plan of Eq. 10. Moreover, with $P_{0,1}$ fixed, the existing f_0 and f_1 minimize the objective in Eq. 14. □

Therefore, based on Proposition A.1, we prove the existence of f_0 and f_1 such that $P_{0,1}$ is the OT plan of Eq. 10, and we can learn the maps f_0 and f_1 by minimizing the objective in Eq. 14, which is the learning objective in Eq. 3.

A.3 Proof of Unpaired Learning Objective

The learning objective in Eq. 14 can be rewritten as the following objective by introducing a new latent random variable z:

$$\int_{\Omega_{0,1,Z}} \|f_1(x_1) - f_0(x_0)\|^2 P_{0,1,Z}(dx_0, dx_1, dz), \tag{15}$$

where $z \in \Omega_Z$ is the new latent random variable, $\Omega_{0,1,Z} = \Omega_0 \times \Omega_1 \times \Omega_Z$, and $P_{0,1,Z}$ is the probability measure on $\Omega_{0,1,Z}$ with the marginal being $P_{0,1}$. We note that $\|f_1(x_1) - f_0(x_0)\|$ is bounded by:

$$\|f_1(x_1) - f_0(x_0)\| \leq \|f_1(x_1) - z\| + \|f_0(x_0) - z\|, \tag{16}$$

where we use the triangle inequality. Taking the square on both sides in Eq. 16, we have:

$$\|f_1(x_1) - f_0(x_0)\|^2 \leq 2\|f_1(x_1) - z\|^2 + 2\|f_0(x_0) - z\|^2, \tag{17}$$

where we use the fact that $\forall a, b \in \mathbb{R} : 2ab \leq a^2 + b^2$. Therefore, the learning objective in Eq. 15 is upper bounded by:

$$\int_{\Omega_{0,Z}} 2\|f_0(x_0) - z\|^2 P_{0,Z}(dx_0, dz) + \int_{\Omega_{1,Z}} 2\|f_1(x_1) - z\|^2 P_{1,Z}(dx_1, dz), \tag{18}$$

where we use the inequality in Eq. 17, split two terms, and integrate over irrelevant variables x_0 or x_1. Here, $\Omega_{0,Z} = \Omega_0 \times \Omega_Z$, $\Omega_{1,Z} = \Omega_1 \times \Omega_Z$, and $P_{0,Z}$ and $P_{1,Z}$ are probability measures on $\Omega_{0,Z}$ and $\Omega_{1,Z}$, respectively.

Therefore, the learning objective in Eq. 14 can be alternatively optimized from its upper bound, with the new learning objective:

$$\arg\min_{f_0, f_1} \int_{\Omega_{0,Z}} \|f_0(x_0) - z\|^2 P_{0,Z}(dx_0, dz) + \int_{\Omega_{1,Z}} \|f_1(x_1) - z\|^2 P_{1,Z}(dx_1, dz), \tag{19}$$

where we drop the proportional constant 2 in front. We note that both f_0 and f_1 only depend on one of the two terms in Eq. 19, and thus the optimization can be split into two independent optimizations:

$$\arg\min_{f_0} \int_{\Omega_{0,Z}} \|f_0(x_0) - z\|^2 P_{0,Z}(dx_0, dz), \tag{20}$$

and

$$\arg\min_{f_1} \int_{\Omega_{1,Z}} \|f_1(x_1) - z\|^2 P_{1,Z}(dx_1, dz), \tag{21}$$

which are the two unpaired learning objectives in Eq. 4. However, for the sum of these independent unpaired objectives to remain a strict, mathematically valid upper bound to the original paired objective, we must ensure that the underlying conditional measures of the latent variable are invariant across domains, a property we prove in the following section.

A.4 Proof of Latent Variable Conditional Measure

To establish the consistency of the unpaired objectives with the paired objective, we must ensure that the conditional probability measures satisfy $P_{Z|0}(\cdot|x_0) = P_{Z|1}(\cdot|x_1)$ for $P_{0,1}$-almost every pair (x_0, x_1).

Using the disintegration theorem, the conditional probability measure $P_{Z|0}$ can be expressed by marginalizing out x_1:

$$P_{Z|0}(dz|x_0) = \int_{\Omega_1} P_{Z|0,1}(dz|x_0, x_1) P_{1|0}(dx_1|x_0). \tag{22}$$

Under Assumption 3.1, the true coupling $P_{0,1}$ is deterministic, meaning $x_1 = g(x_0)$ almost everywhere. Consequently, the conditional measure $P_{1|0}(\cdot|x_0)$ collapses to the Dirac measure $\delta_{g(x_0)}(\cdot)$.

Furthermore, because g is a diffeomorphism (Assumption 3.2), it is a bijection. Therefore, the σ-algebras generated by the random variables X_0 and X_1 are identical (i.e., $\sigma(X_0) = \sigma(X_1) = \sigma(X_0, X_1)$). Conditioning on both x_0 and x_1 provides no additional information over conditioning on x_1 alone, yielding:

$$P_{Z|0,1}(dz|x_0, x_1) = P_{Z|1}(dz|x_1). \tag{23}$$

Substituting Eq. 23 and the Dirac measure into Eq. 22, the integral simplifies to evaluating the integrand exactly at $x_1 = g(x_0)$:

$$P_{Z|0}(dz|x_0) = \int_{\Omega_1} P_{Z|1}(dz|x_1) \delta_{g(x_0)}(dx_1) = P_{Z|1}(dz|g(x_0)). \tag{24}$$

Since $x_1 = g(x_0)$ for any pair (x_0, x_1) drawn from the true coupling $P_{0,1}$, we obtain the required identity:

$$P_{Z|0}(dz|x_0) = P_{Z|1}(dz|x_1). \tag{25}$$

This mathematically shows that sampling the latent variable z conditioned on x_0 should recover the exact same conditional distribution as sampling it conditioned on x_1 for any true pair $(x_0, x_1) \sim P_{0,1}$. By fulfilling this essential requirement of cross-domain invariance, we guarantee that optimizing the independent unpaired objectives remains a strict, valid proxy for minimizing the true paired objective.

A.5 Proof of Isometry Regularization

To study the new unpaired learning objectives in Eq. 6, we note that when the noise ϵ is small, we can approximate $f_i(x_i + \epsilon)$ through the first-order Taylor expansion:

$$f_i(x_i + \epsilon) = f_i(x_i) + df_i \epsilon, \tag{26}$$

where df_i is the induced linear map, i.e., the Jacobian of the function f_i evaluated at x_i. Then, we can expand the integrand in Eq. 6 by:

$$\|f_i(x_i + \epsilon) - z\|^2 = \|f_i(x_i) - z\|^2 + 2(f_i(x_i) - z)^\mathsf{T} df_i \epsilon + \epsilon^\mathsf{T} df_i^\mathsf{T} df_i \epsilon, \tag{27}$$

which contains three terms. We substitute Eq. 27 into the objective function in Eq. 6 and we obtain:

$$\begin{aligned} \int_{\Omega_{i,Z,E}} \|f_i(x_i + \epsilon) - z\|^2 P_{i,Z,E}(dx_i, dz, d\epsilon) = &\int_{\Omega_{i,Z,E}} \|f_i(x_i) - z\|^2 P_{i,Z,E}(dx_i, dz, d\epsilon) \\ &+ \int_{\Omega_{i,Z,E}} 2(f_i(x_i) - z)^\mathsf{T} df_i \epsilon P_{i,Z,E}(dx_i, dz, d\epsilon) \\ &+ \int_{\Omega_{i,Z,E}} \epsilon^\mathsf{T} df_i^\mathsf{T} df_i \epsilon P_{i,Z,E}(dx_i, dz, d\epsilon), \end{aligned} \tag{28}$$

which contains three integrations. In the first integration, we can integrate over ϵ because the integrand does not depend on ϵ. In the second integration, we notice that ϵ is an independent Gaussian random variable with a zero mean, and thus the integration equals 0. In the third integration, we can integrate over z because the integrand does not depend on z. Therefore, we can simplify Eq. 28 as:

$$\begin{aligned} \int_{\Omega_{i,Z,E}} \|f_i(x_i + \epsilon) - z\|^2 P_{i,Z,E}(dx_i, dz, d\epsilon) = &\int_{\Omega_{i,Z}} \|f_i(x_i) - z\|^2 P_{i,Z}(dx_i, dz) \\ &+ \int_{\Omega_{i,E}} \epsilon^\mathsf{T} df_i^\mathsf{T} df_i \epsilon P_{i,E}(dx_i, d\epsilon), \end{aligned} \tag{29}$$

where two terms remain and the first term is exactly the unpaired learning objective in Eq. 4. We also notice that the second term can be further simplified as:

$$\int_{\Omega_{i,E}} \epsilon^\mathsf{T} df_i^\mathsf{T} df_i \epsilon P_{i,E}(dx_i, d\epsilon) = \sigma^2 \int_{\Omega_i} \mathrm{tr}(df_i^\mathsf{T} df_i) P_i(dx_i), \tag{30}$$

where tr refers to the trace of a matrix, and σ^2 is the variance of each entry of the Gaussian random variable ϵ. Therefore, the new unpaired learning objective in Eq. 6 becomes:

$$\arg\min_{f_i} \int_{\Omega_{i,Z}} \|f_i(x_i) - z\|^2 P_{i,Z}(dx_i, dz) + \sigma^2 \int_{\Omega_i} \mathrm{tr}(df_i^\mathsf{T} df_i) P_i(dx_i), \tag{31}$$

which is equivalent to adding a regularization term weighted by σ^2 to the original unpaired learning objective in Eq. 4.

To find out what the regularization term does in Eq. 31, we note that $|\det(df_i)| = 1$ because f_i is a volume-preserving map that is constructed by the neural network structure in Fig. 5 as proved by Eq. 5. Therefore, $\det(df_i^\mathsf{T} df_i) = |\det(df_i)|^2 = 1$. Moreover, $df_i^\mathsf{T} df_i$ is a real symmetric and semi-positive definite matrix, which has real non-negative eigenvalues. Let its eigenvalues be λ_j for $j \in \{1, 2, \ldots, d\}$, we have:

$$\prod_{j=1}^{d} \lambda_j = 1, \tag{32}$$

where we use $\det(df_i^\mathsf{T} df_i) = \prod_{j=1}^{d} \lambda_j$. Using the arithmetic mean-geometric mean inequality, we have:

$$\mathrm{tr}(df_i^\mathsf{T} df_i) = \sum_{j=1}^{d} \lambda_j \geq d \left(\prod_{j=1}^{d} \lambda_j \right)^{\frac{1}{d}} = d, \tag{33}$$

which shows that the trace of $df_i^\mathsf{T} df_i$ is always greater than or equal to d. The equality holds if and only if $\lambda_1 = \lambda_2 = \cdots = \lambda_d = 1$, which means that the minimum value is reached if and only if the following condition is met:

$$df_i^\mathsf{T} df_i = \mathrm{id}, \tag{34}$$

because a real symmetric matrix with all eigenvalues equal to 1 must be the identity matrix. Therefore, the regularization term in Eq. 31 encourages df_i to be a local isometry of the Euclidean metric space as close as possible, and minimizing the new unpaired learning objectives in Eq. 6 achieves the isometry regularization in Assumption 3.4.

A.6 Proof of Metric and Pseudometric Spaces

Let (Ω', d') be a metric space defining the latent domain, where $\Omega' = \Omega'_0 \cup \Omega'_1$. A metric d' fundamentally satisfies non-negativity, symmetry, the triangle inequality, and the identity of indiscernibles ($d'(x', y') = 0 \iff x' = y'$).

We evaluate the pullback metric $d = f^* d'$ under the piecewise mapping $f : \Omega_0 \cup \Omega_1 \to \Omega'$:

$$f(x) = \begin{cases} f_0(x), & \text{if } x \in \Omega_0, \\ f_1(x), & \text{if } x \in \Omega_1. \end{cases} \tag{35}$$

For LMSB, the formulation strictly restricts $f_0 = f_1 = \bar{f}$. Consequently, $f(x) = \bar{f}(x)$ for all $x \in \Omega_0 \cup \Omega_1$. Because $\bar{f}$ is explicitly structured as a diffeomorphism, it is globally bijective. Thus, if $d(x_0, x_1) = d'(\bar{f}(x_0), \bar{f}(x_1)) = 0$, the metric properties of d' force $\bar{f}(x_0) = \bar{f}(x_1)$. Because $\bar{f}$ is strictly injective, $\bar{f}(x_0) = \bar{f}(x_1) \implies x_0 = x_1$. Therefore, the pullback metric $f^* d'$ preserves the identity of indiscernibles and successfully forms a strict metric space over $\Omega_0 \cup \Omega_1$.

For PMSB, f_0 and f_1 are distinct functions optimized independently. We must analyze whether $d(x_0, x_1) = 0$ inherently implies $x_0 = x_1$. Consider two elements $x_0 \in \Omega_0$ and $x_1 \in \Omega_1$, we have $x_0 \neq x_1$ by definition of the separate domains. We evaluate their distance in the latent space: $d(x_0, x_1) = d'(f_0(x_0), f_1(x_1))$. Because $f_0 \neq f_1$, the optimization process allows $f_0(x_0)$ to perfectly align with $f_1(x_1)$. In this scenario, we have $d'(f_0(x_0), f_1(x_1)) = 0$ while $x_0 \neq x_1$. Because the identity of indiscernibles is violated ($d(x_0, x_1) = 0 \implies x_0 = x_1$), the resulting pullback metric $f^* d'$ forms a pseudometric space. This mathematically confirms PMSB's capacity to transport distinct images to overlapping latent coordinates.

Table 2: Ablation study ($N = 1,020$) comparing SDE and ODE samplers for DSBM, LMSB, and PMSB models. Translation performance is evaluated using a domain consistency metric (FID) alongside three structural consistency metrics (PSNR, SSIM, and LNCC). SSIM and LNCC are percentages. Bold number indicates the best result within each task and metric across all configurations.

Metric	Task	DSBM		LMSB		PMSB	
		SDE	ODE	SDE	ODE	SDE	ODE
PSNR ↑	T_1-w → T_2-w	16.18 ± 1.03	16.17 ± 1.02	18.82 ± 1.14	18.79 ± 1.15	24.85 ± 1.31	**25.03** ± 1.34
	T_2-w → T_1-w	18.45 ± 1.08	18.42 ± 1.08	21.03 ± 1.41	20.98 ± 1.42	27.00 ± 1.96	**27.20** ± 2.04
	T_1-w → FLAIR	29.78 ± 1.76	**29.81** ± 1.78	29.01 ± 2.49	28.99 ± 2.54	29.39 ± 1.79	29.40 ± 1.79
	FLAIR → T_1-w	24.63 ± 1.36	24.59 ± 1.38	24.41 ± 2.33	24.38 ± 2.37	**25.20** ± 1.20	25.18 ± 1.24
	Overall	22.26 ± 6.16	22.25 ± 6.17	23.32 ± 4.44	23.29 ± 4.44	26.61 ± 2.08	**26.70** ± 2.05
SSIM ↑	T_1-w → T_2-w	54.54 ± 3.98	53.54 ± 4.02	62.16 ± 3.84	61.14 ± 3.93	87.55 ± 2.07	**88.71** ± 2.14
	T_2-w → T_1-w	54.92 ± 4.06	53.82 ± 4.13	66.07 ± 3.96	65.11 ± 4.17	89.84 ± 1.56	**90.88** ± 1.66
	T_1-w → FLAIR	85.33 ± 1.96	85.22 ± 2.01	**85.55** ± 2.34	85.40 ± 2.41	85.52 ± 2.01	85.40 ± 2.05
	FLAIR → T_1-w	78.35 ± 4.30	77.85 ± 4.58	79.07 ± 4.40	78.66 ± 4.73	**79.64** ± 3.64	79.22 ± 3.98
	Overall	68.29 ± 15.91	67.61 ± 16.36	73.21 ± 10.95	72.58 ± 11.37	85.64 ± 4.37	**86.05** ± 5.08
LNCC ↑	T_1-w → T_2-w	12.79 ± 1.26	13.62 ± 1.37	12.29 ± 0.95	12.81 ± 1.06	26.30 ± 2.23	**27.87** ± 2.45
	T_2-w → T_1-w	13.53 ± 1.33	14.13 ± 1.44	14.18 ± 1.08	14.39 ± 1.06	29.49 ± 2.39	**30.83** ± 2.57
	T_1-w → FLAIR	16.50 ± 2.47	16.79 ± 2.50	17.24 ± 2.63	**17.50** ± 2.63	17.14 ± 2.67	17.38 ± 2.69
	FLAIR → T_1-w	18.82 ± 3.03	19.00 ± 2.97	19.54 ± 3.18	**19.75** ± 3.12	19.58 ± 3.06	19.73 ± 3.05
	Overall	15.41 ± 2.78	15.89 ± 2.50	15.81 ± 3.22	16.11 ± 3.11	23.13 ± 5.74	**23.95** ± 6.42
FID ↓	T_1-w → T_2-w	**18.50**	18.71	34.09	34.69	32.85	33.23
	T_2-w → T_1-w	**17.20**	17.42	20.62	21.96	20.72	21.42
	T_1-w → FLAIR	**15.39**	16.39	19.63	20.45	18.92	19.65
	FLAIR → T_1-w	**13.45**	16.32	18.24	21.35	17.02	20.69
	Overall	**16.14** ± 2.20	17.21 ± 1.12	23.14 ± 7.36	24.61 ± 6.75	22.38 ± 7.14	23.75 ± 6.36

B Ablation Study

B.1 SDE and ODE Samplers

As DSBM, LMSB, and PMSB all belong to the family of diffusion models, they can be sampled using either a stochastic differential equation (SDE) or an ordinary differential equation (ODE) formulation. SDE preserves the full path measure of the stochastic process through injected noise during generation, whereas ODE is a deterministic process that solely matches the marginal densities along the path.

To evaluate the impact of these sampling strategies on medical image translation, we conduct a systematic ablation study, with quantitative results detailed in Table 2. The results reveal a clear and consistent trade-off between domain consistency and structural fidelity across all three models. The SDE sampler generally achieves lower (better) FID scores compared to the ODE sampler. This indicates that the stochastic noise injected during SDE sampling helps the model better capture the target data distribution, leading to images with higher perceptual realism and domain consistency.

Conversely, the ODE sampler consistently yields slightly higher PSNR, SSIM, and LNCC scores across most translation tasks. Because ODE is deterministic and removes the stochastic variations inherent in SDE, it enforces stricter pixel-level structural correspondence and alignment between the source and the translated images. In medical imaging applications, where preserving exact anatomical structures and avoiding hallucinated features is paramount, structural fidelity typically takes precedence over marginal improvements in perceptual distribution. Therefore, we use the ODE sampler for the main results presented in Table 1, as it provides the most reliable and structurally faithful translations.

B.2 Sensitivity to MNI Registration

We assess the dependence of PMSB on template registration by retraining the model for T_1-w $\leftrightarrow$ T_2-w translation completely independent of MNI template alignment. For a fair comparison, we use the exact same 402 paired sessions and subject split as the registered baseline. However, because different native head orientations produce different extracted axial slices, the two conditions do not operate on identical images.

Table 3: PMSB translation performance with and without MNI registration on OASIS-3. The two conditions use the same subjects and split but different extracted slices. SSIM and LNCC are percentages. Bold number indicates the better value within each task and metric.

Metric	Task	With MNI	No MNI
PSNR ↑	T_1-w → T_2-w	25.03 ± 1.34	**25.11** ± 1.42
	T_2-w → T_1-w	**27.20** ± 2.04	24.71 ± 2.37
SSIM ↑	T_1-w → T_2-w	88.71 ± 2.14	**89.63** ± 1.96
	T_2-w → T_1-w	**90.88** ± 1.66	87.23 ± 2.15
LNCC ↑	T_1-w → T_2-w	27.87 ± 2.45	**28.74** ± 3.01
	T_2-w → T_1-w	**30.83** ± 2.57	27.65 ± 2.78
FID ↓	T_1-w → T_2-w	33.23	**19.07**
	T_2-w → T_1-w	**21.42**	42.34

Table 4: PMSB sensitivity to white-matter label dilation on OASIS-3 ($N = 1{,}020$ paired test slices). SSIM and LNCC are percentages. Bold number indicates the best result within each task and metric.

Metric	Task	Original	WM, $r = 1$	WM, $r = 2$
PSNR ↑	T_1-w → T_2-w	**25.03** ± 1.34	24.74 ± 1.33	24.32 ± 1.19
	T_2-w → T_1-w	**27.20** ± 2.04	26.96 ± 1.95	26.77 ± 1.72
SSIM ↑	T_1-w → T_2-w	**88.71** ± 2.14	88.31 ± 1.99	87.74 ± 2.01
	T_2-w → T_1-w	**90.88** ± 1.66	90.40 ± 1.73	90.13 ± 1.63
LNCC ↑	T_1-w → T_2-w	**27.87** ± 2.45	27.37 ± 2.42	26.88 ± 2.37
	T_2-w → T_1-w	**30.83** ± 2.57	30.33 ± 2.53	30.03 ± 2.47
FID ↓	T_1-w → T_2-w	33.23	**25.35**	27.20
	T_2-w → T_1-w	**21.42**	22.33	24.36

Table 3 demonstrates that PMSB maintains highly comparable structural and perceptual scores across these unregistered intra-subject slices. Translation from T_1-w → T_2-w yields structural and perceptual metrics slightly better than the MNI-registered baseline. While the reverse translation from T_2-w → T_1-w exhibits a moderate drop in structural and perceptual scores. This confirms that while MNI alignment is a standard neuroimaging preprocessing step, it is not a strict prerequisite for PMSB, supporting the method's viability in general unpaired settings.

B.3 Segmentation-Target Robustness

We assess the sensitivity of PMSB to segmentation quality by artificially degrading the auxiliary segmentation targets. Specifically, we dilate the SynthSeg white-matter (WM) label with radii $r = 1$ and $r = 2$, and retrain both the diffeomorphisms and the bridge on the same OASIS-3 training and validation sets. Table 4 reports the translation performance on the same 1,020 paired OASIS-3 test slices.

Structural metrics exhibit only modest decreases despite the boundary corruption introduced by the morphological dilation. This robustness stems from the explicit volume-preserving and approximate isometry constraints applied to the mapping functions f_0 and f_1, which force the network to preserve local tissue geometries rather than strictly memorize the target labels. Consequently, PMSB successfully preserves fine anatomical details that are absent from the corrupted SynthSeg targets (as shown visually in Fig. 6), demonstrating strong robustness to moderate auxiliary segmentation errors.

C Latent-Domain Gap Analysis

A core theoretical advantage of PMSB is the relaxation of the shared-mapping constraint into independent, domain-specific mappings, which induces a pseudometric space capable of overlapping identically-structured cross-contrast anatomies. To validate this, we measure the mean squared

Table 5: Mean squared distance (MSD $\times 10^3$) between paired OASIS-3 test images in original and mapped spaces ($N = 1{,}020$).

Metric	Task	Original MR	LMSB latent	PMSB latent
MSD $\times 10^3$	T_1-w $\leftrightarrow$ T_2-w	52.22 ± 11.87	16.72 ± 2.73	13.31 ± 2.79
	T_1-w $\leftrightarrow$ FLAIR	15.58 ± 3.79	8.28 ± 1.24	8.00 ± 1.23

Table 6: BraTS20 translation performance ($N = 1{,}089$ paired test slices). SSIM and LNCC are percentages. Bold number indicates the best result within each task and metric.

Metric	Task	SynDiff	DSBM	PMSB
PSNR $\uparrow$	T_1-w $\rightarrow$ T_2-w	19.79 ± 1.33	16.19 ± 1.35	**21.12** ± 1.55
	T_2-w $\rightarrow$ T_1-w	20.22 ± 0.83	15.33 ± 1.12	**21.12** ± 1.20
SSIM $\uparrow$	T_1-w $\rightarrow$ T_2-w	75.94 ± 2.38	56.08 ± 4.82	**83.33** ± 2.91
	T_2-w $\rightarrow$ T_1-w	77.31 ± 1.67	51.93 ± 5.39	**83.85** ± 2.15
LNCC $\uparrow$	T_1-w $\rightarrow$ T_2-w	13.12 ± 2.54	19.29 ± 2.52	**21.56** ± 2.93
	T_2-w $\rightarrow$ T_1-w	16.16 ± 2.44	19.15 ± 2.58	**23.47** ± 2.75
FID $\downarrow$	T_1-w $\rightarrow$ T_2-w	**36.74**	49.55	55.74
	T_2-w $\rightarrow$ T_1-w	70.04	72.03	**53.56**

distance between paired test images initially in the raw image space, and subsequently after mapping to the latent spaces of LMSB and PMSB.

Table 5 reports these inter-domain distances, scaled by 10^3, for the 1,020 OASIS-3 test pairs. The results confirm that PMSB achieves a consistently smaller latent gap than LMSB across both T_1-w $\leftrightarrow$ T_2-w and T_1-w $\leftrightarrow$ FLAIR translation tasks. Furthermore, the absolute reduction is most significant in tasks with large initial image-space domain gaps. For the T_1-w $\leftrightarrow$ T_2-w translation, where LMSB leaves a substantial residual latent gap of 16.72 ± 2.73, PMSB significantly closes the distance to 13.31 ± 2.79. In comparison, for the T_1-w $\leftrightarrow$ FLAIR task where LMSB already achieves tight alignment, the absolute gains appear to be smaller. These measurements empirically validate our theoretical claims regarding pseudometric mapping and explain the asymmetric performance gains observed in Table 1.

D Generalizability to Pathological Anatomy

To demonstrate that PMSB generalizes beyond normal neuroanatomy and healthy aging variants, we evaluate T_1-w $\leftrightarrow$ T_2-w translation performance on the BraTS20 dataset, which features severe tumor-induced anatomical deformations. The 369 available subjects are partitioned into disjoint sets of 304 training, 32 validation, and 33 testing subjects. By extracting 33 axial slices per subject, we obtain 5,016 unpaired training slices, 528 validation slices, and 1,089 paired test slices. For PMSB, the latent segmentation targets are constructed by integrating explicit tumor labels with the base SynthSeg segmentation maps.

Table 6 compares PMSB against SynDiff and DSBM on this pathological cohort. PMSB consistently outperforms both baselines, with the highest PSNR, SSIM, and LNCC across both translation directions. This structural superiority demonstrates that the pseudometric mapping framework successfully generalizes to domains containing significant topological abnormalities, preserving complex pathological anatomy as effectively as normal anatomy.

E Computational Costs

We report the computational costs of all evaluated methods in Table 7, specifically the training and per-slice inference time. Each method was trained and evaluated using a single GPU equipped with 48 GB of memory (hardware included four NVIDIA A40 and two NVIDIA RTX A6000). Both the training and the inference time include processing overhead, such as disk I/O.

Table 7: Training and per-slice inference time across methods.

Metric	**SynDiff**	**ContourDiff**	**UNSB**	**DSBM**	**LMSB**	**PMSB**
Sampler	Default	Default	Default	SDE / ODE	SDE / ODE	SDE / ODE
Train (h)	55.8	33.0	75.7	39.5	47.8	56.0
Infer (s/slice)	0.24	1.42	3.15	4.95 / 12.87	4.90 / 12.51	4.89 / 12.41

While the multi-stage training of the diffeomorphisms and the DSBM requires a substantial training time of 56.0 hours, PMSB efficiently encapsulates bidirectional translation capabilities within a single training phase. For a strict comparison, baselines such as ContourDiff and UNSB require independent training for each translation direction, which doubles their reported training time when bidirectional mapping is required. For DSBM, LMSB, and PMSB, inference time is reported for both the SDE and ODE samplers.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and Section 1 summarize theory and experiments, which Sections 3–4 instantiate.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Section 5 summarize the remaining gaps in both practical usability and theoretical limits.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: Section 3 presents Assumptions 3.1–3.4 while Appendices A.1–A.6 collect formal lemmas, propositions, and proofs.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Section 3 describes the theory and the algorithm. Section 4 presents dataset, preprocessing, and network training parameters.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: The paper uses public datasets and the code is available.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

Answer: [Yes]

Justification: Section 4 lists training, validation, and testing set split sizes and network training parameters.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: Tables 1, 2 report mean, standard deviation, and number of samples in the evaluation.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Appendix E reports computational costs.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: We preserve anonymity.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: There is no societal impact of the work performed.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.

- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: Owners of assets including data, preprocessing code, comparison models throughout the paper are properly credited as citations.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not release new assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.