Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Self-consistent recursive diffusion bridge for medical image translation

Fuat Arslan [a,b,1], Bilal Kabas [a,b,1], Onat Dalmaz [a,b], Muzaffer Ozbey [a,b], Tolga Çukur [a,b,c,*]

[a] *Department of Electrical and Electronics Engineering, Bilkent University, Ankara 06800, Turkey*
[b] *National Magnetic Resonance Research Center (UMRAM), Bilkent University, Ankara 06800, Turkey*
[c] *Neuroscience Program, Bilkent University, Ankara 06800, Turkey*

ARTICLE INFO

Keywords:
Medical image translation
Synthesis
Generative
MRI
CT
Diffusion
Bridge

ABSTRACT

Denoising diffusion models (DDM) have gained recent traction in medical image translation given their high training stability and image fidelity. DDMs learn a multi-step denoising transformation that progressively maps random Gaussian-noise images provided as input onto target-modality images as output, while receiving indirect guidance from source-modality images via a separate static channel. This denoising transformation diverges significantly from the task-relevant source-to-target modality transformation, as source images are governed by a non-noise distribution. In turn, DDMs can suffer from suboptimal source-modality guidance and performance losses in medical image translation. Here, we propose a novel self-consistent recursive diffusion bridge (SelfRDB) that leverages direct source-modality guidance within its diffusion process for improved performance in medical image translation. Unlike DDMs, SelfRDB devises a novel forward process with the start-point taken as the target image, and the end-point defined based on the source image. Intermediate image samples across the process are expressed via a normal distribution whose mean is taken as a convex combination of start-end points, and whose variance is controlled by additive noise. Unlike regular diffusion bridges that prescribe zero noise variance at start-end points and high noise variance at mid-point of the process, we propose a novel noise scheduling with monotonically increasing variance towards the end-point in order to facilitate information transfer between the two modalities and boost robustness against measurement noise. To further enhance sampling accuracy in each reverse step, we propose a novel sampling procedure where the network recursively generates a transient-estimate of the target image until convergence onto a self-consistent solution. Comprehensive experiments in multi-contrast MRI and MRI-CT translation indicate that SelfRDB achieves state-of-the-art results in terms of image quality.

1. Introduction

Medical images acquired under multiple modalities capture complementary diagnostic information on bodily tissues (Iglesias et al., 2013; Lee et al., 2017), but running multi-modal protocols is burdening given associated economic and labor costs (Ye et al., 2013; Huynh et al., 2016; Jog et al., 2017; Joyce et al., 2017). A powerful approach to extend the scope of imaging-based assessments without elevating costs is medical image translation, wherein unacquired target modalities are predicted from acquired source modalities (Cordier et al., 2016; Wu et al., 2016; Zhao et al., 2017; Huang et al., 2018). Important clinical applications of translation include imputation of target modalities with a higher degree of diagnostically redundant information that are excluded from imaging protocols in order to lower redundancy and increase scan efficiency, imputation of invasive target modalities from non-invasive source modalities in order to avoid injection of harmful contrast agents or exposure to ionizing radiation (Lee et al., 2019), and imputation of missing target modalities in imaging protocols in order to improve protocol homogeneity across participants in retrospective imaging studies (Clark et al., 2019). That said, medical image translation is an inherently challenging problem as signal levels for a given tissue show nonlinear variations across separate modalities that are difficult to characterize analytically (Roy et al., 2013; Alexander et al., 2014; Huang et al., 2017). As such, learning-based methods that excel at solving nonlinear problems have recently become the de facto framework for medical image translation (Van Nguyen et al., 2015; Vemulapalli et al., 2015; Sevetlidis et al., 2016; Nie et al., 2016).

Learning-based methods commonly aim to capture a conditional prior for the distribution of target images given respective source images, albeit differ in the approach that they adopt in order to learn this prior (Bowles et al., 2016; Chartsias et al., 2018; Nie et al., 2018; Yang

[*] Corresponding author at: Department of Electrical and Electronics Engineering, Bilkent University, Ankara 06800, Turkey.
E-mail address: cukur@ee.bilkent.edu.tr (T. Çukur).

[1] F. Arslan and B. Kabas contributed equally to the study.

https://doi.org/10.1016/j.media.2025.103747
Received 19 September 2024; Received in revised form 2 July 2025; Accepted 22 July 2025
Available online 5 August 2025

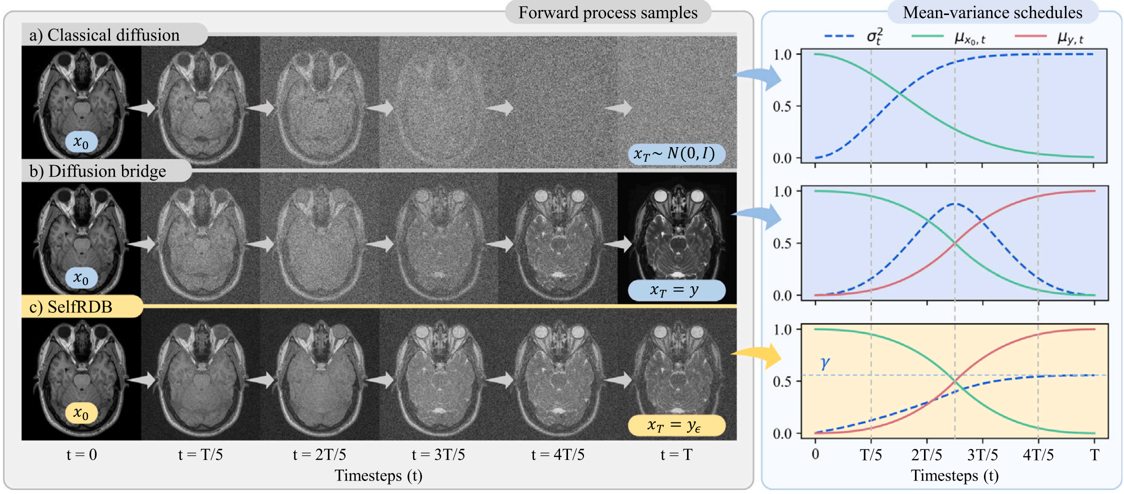

Fig. 1. Diffusion methods commonly take the target image as the start-point $\boldsymbol{x}_0$ of the diffusion process, albeit they can differ in expression of image samples in remaining timesteps. Illustrations of images across the forward process are depicted along with underlying schedules for the mean ($\mu_{x_0,t}$, $\mu_{y,t}$) and noise variance (σ_t^2). **(a) Classical diffusion:** DDMs use a white-Gaussian noise image as an asymptotic end-point $\boldsymbol{x}_T \sim N(0, I)$. Intermediate samples are obtained by adding increasing levels of random Gaussian noise onto the target image. **(b) Regular diffusion bridge:** Regular bridges use the source image as a finite end-point, $\boldsymbol{x}_T = \boldsymbol{y}$. Intermediate samples are taken as a convex combination of source–target images, corrupted with additive noise. Noise variance is zero at start- and end-points, and it peaks at the mid-point. **(c) Proposed:** SelfRDB is a novel diffusion bridge that uses a noise-added source image as the end-point, $\boldsymbol{x}_T = \boldsymbol{y}_\epsilon$. Intermediate samples still depend on a convex combination of source–target images, yet SelfRDB uniquely prescribes monotonically-increasing noise variance towards the end-point.

et al., 2018; Wei et al., 2019). Among previous methods, generative adversarial networks (GAN) have been widely adopted for their exceptional realism in synthesized target images (Yu et al., 2018; Armanious et al., 2019; Li et al., 2019; Dar et al., 2019b; Yu et al., 2019), and successfully reported in diverse tasks including multi-contrast MRI (Kim et al., 2021; Yurt et al., 2021; Hu et al., 2022; Xia et al., 2023; Han et al., 2023; Zhang et al., 2025) and MRI-CT translation (Jin et al., 2019; Dalmaz et al., 2022; Gu et al., 2023; Xin et al., 2024). Yet, GANs capture an implicit prior through a generator–discriminator interplay, so they are susceptible to training instabilities that often hamper image fidelity (Wang et al., 2020; Zhou et al., 2020). Instead, recent studies have employed denoising diffusion models (DDM) to capture an explicit prior with improved training stability (Özbey et al., 2023; Meng et al., 2022; Lyu and Wang, 2022; Wang et al., 2024). DDMs employ a forward process that gradually degrades the target image by repeated addition of Gaussian noise till an asymptotic end-point of a pure noise image is reached (Fig. 1a). To recover the target image, a reverse process is then operationalized via a recovery network that progressively denoises the random noise image while receiving the source image as a separate, static input (Ho et al., 2020). Since the forward process is completely agnostic to the source modality, the reverse process is devised to learn a denoising transformation from noise to target images under indirect guidance from the source modality (Özbey et al., 2023). The recovery network then compromises between the task-irrelevant denoising transformation and the task-relevant source-to-target image transformation at each reverse diffusion step, which can result in under- or over-emphasis of source-modality guidance (Song et al., 2021; Güngör et al., 2023). As such, DDMs can suffer from suboptimal translation performance due to the divergence between the denoising and source-to-target transformations (Liu et al., 2023a).

An emerging approach to enhance task relevance in diffusion-based priors employs diffusion bridges that can directly transform between two separate modalities (Delbracio and Milanfar, 2023; Chung et al., 2023). To do this, diffusion bridges define the start- and end-points of the forward process based on target and source images, respectively (Fig. 1b). As the imaging operator linking the two modalities is typically unknown, image samples in intermediate steps are derived from a normal distribution whose mean is a convex combination of start- and end-points (Liu et al., 2023a; Kim et al., 2024a). Initiating sampling on the source image, the reverse process progressively maps the source onto the target image. Few recent imaging studies have successfully employed diffusion bridges in the reconstruction of single-modal images from degraded measurements due to factors such as undersampling or low resolution (Mirza et al., 2023; Kim and Ye, 2024; Kim et al., 2024b). However, the potential of diffusion bridges in medical image translation remains largely unexplored, as existing methods face several key challenges. Regular diffusion bridges adopt a noise scheduling with zero variance at start-end points albeit high variance near the mid-point of the diffusion process (Su et al., 2023). Zero variance at the end-point results in *a hard-prior on the source modality* reflecting a deterministic Dirac-delta distribution centered on source images within the training set, hampering reliability against source-image variability due to measurement noise (Fig. 2a). Meanwhile, high variance at the mid-point can disrupt information transfer across the diffusion process between source and target modalities. Furthermore, diffusion bridges typically synthesize a *one-shot estimate of intermediate samples*, limiting sampling accuracy for generated images (Peng et al., 2022).

Here, we propose a novel self-consistent recursive diffusion bridge, SelfRDB, to improve performance in multi-modal medical image translation. Unlike regular diffusion bridges, SelfRDB leverages a novel noise scheduling in its forward process, with monotonically increasing variance towards the end-point that corresponds to a noise-added source image (Fig. 1c). As such, it captures *a soft-prior on the source modality* to attain improved robustness to measurement noise, while it facilitates information transfer between modalities by prescribing lower variance near the mid-point of the process (Fig. 2b). To avoid loss of tissue information at the noise-added end-point, SelfRDB's recovery network employs stationary guidance from the original source image in the reverse process. Finally, to improve sampling accuracy in each reverse step, SelfRDB leverages a novel *self-consistent recursive estimation procedure* for the target image, and uses this self-consistent estimate to synthesize intermediate samples with enhanced accuracy (Fig. 3). Comprehensive demonstrations are performed for multi-contrast MRI and MRI-CT translation. Our results clearly indicate the superiority of SelfRDB against competing GAN and diffusion models, including previous diffusion bridges. Code for SelfRDB is available at https://github.com/icon-lab/SelfRDB.

Contributions

- To our knowledge, SelfRDB is the first diffusion bridge for medical image translation between separate modalities in the literature.

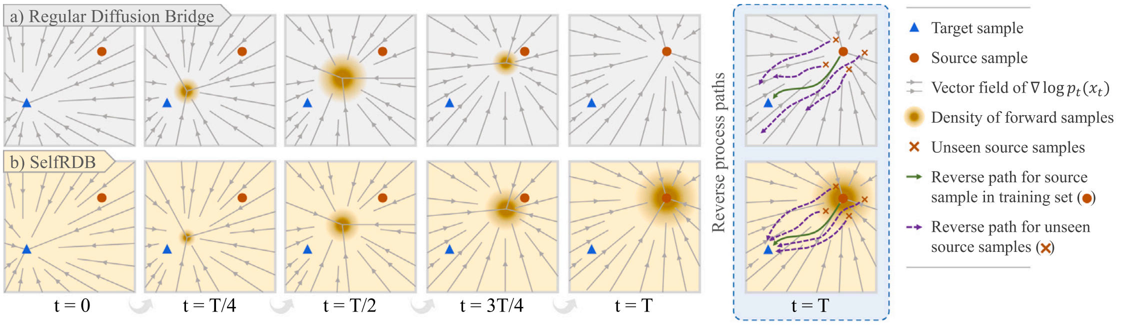
a) Regular Diffusion Bridge
b) SelfRDB
t = 0
t = T/4
t = T/2
t = 3T/4
t = T
Reverse process paths
t = T
Target sample
Source sample
Vector field of $\nabla \log p_t(x_t)$
Density of forward samples
Unseen source samples
Reverse path for source sample in training set (●)
Reverse path for unseen source samples (×)

Fig. 2. Diffusion models learn the score function of the data through a multi-step transformation between the start- and end-points of the underlying diffusion process. Image samples are typically corrupted with Gaussian noise that smooths the data distribution by masking some of the original image features. Smoothing enables more uniform coverage of the data space in order to boost reliability against noise-induced variability. **(a) Regular diffusion bridges** use zero noise variance at the end-point constraining them to a Dirac-delta distribution centered on the source images within the training set. This can compromise generalization performance to source images outside the training set (see purple-colored dashed paths). **(b) SelfRDB** instead uses monotonically-increasing variance towards the end-point, so it is trained on noise-added source images. This improves robustness against variability in measurement noise levels of source images between training and test sets (see purple-colored dashed paths).

- SelfRDB leverages a novel forward diffusion process that captures a soft-prior on the source modality to improve robustness against measurement noise and to facilitate information transfer between source–target modalities.
- SelfRDB leverages a novel self-consistent recursive estimation procedure to improve sampling accuracy in reverse diffusion steps.

2. Related work

2.1. Generative diffusion models

Diffusion models have emerged as a promising alternative to GANs for generative modeling in computer vision applications (Ho et al., 2020). In the conventional diffusion framework, DDMs create image samples by progressively denoising a pure noise sample through an iterative process, guided by a neural network trained to optimize a correlate of the data likelihood known as the score function (Song et al., 2020). The gradual stochastic sampling approach and the explicit likelihood formulation enable DDMs to deliver improved sample quality and diversity through more reliable network mappings (Dhariwal and Nichol, 2021). As such, recent studies have adopted DDMs for implementing various single-modality imaging tasks, including image reconstruction (Jalal et al., 2021; Chung and Ye, 2022; Güngör et al., 2023), unconditional image generation (Pinaya et al., 2022b), and anomaly detection (Wolleb et al., 2022; Pinaya et al., 2022a). Extending beyond these unimodal applications, here we focus on multi-modal translation tasks that involve mappings between separate imaging modalities.

2.2. Multi-modal medical image translation with DDMs

DDMs have recently been adopted in multi-modal medical image translation given their improved image fidelity (Özbey et al., 2023; Meng et al., 2022; Lyu and Wang, 2022; Wang et al., 2024). Employing a forward process where target images are corrupted with additive noise over a large number of steps, DDMs progressively map a random noise image onto the target under indirect guidance from the source image (Ho et al., 2020). This multi-step denoising transformation helps improve training stability over GANs (Wolterink et al., 2017; Dong et al., 2019). Unfortunately, the image mapping performed by the denoising transformation is weakly associated with the desired source-to-target image mapping for translation tasks, and the source-image guidance in DDMs is primarily implicit (Liu et al., 2023a). In turn, these limitations can compromise performance in DDM-based translation. To address these issues, here we introduce the first diffusion bridge for multi-modal medical image translation to our knowledge. Unlike DDMs that express intermediate samples as noise-added target images and use an end-point of Gaussian noise, SelfRDB expresses intermediate samples as a convex combination of source and target images corrupted with additive noise, and employs an end-point of a noise-added source image. Unlike DDMs that use one-shot sampling in each reverse step, SelfRDB employs self-consistent recursive estimation to improve sampling accuracy. Based on these unique advances, we provide the first demonstrations of multi-contrast MRI and MRI-CT translation based on diffusion bridges.

2.3. Image translation with diffusion bridges

Diffusion bridges are an emerging alternative to DDMs to improve flexibility in generative modeling tasks. Several computer vision studies (Daras et al., 2022; Delbracio and Milanfar, 2023; Chung et al., 2023) and a few recent imaging studies (Mirza et al., 2023; Kim and Ye, 2024) have devised diffusion bridges for single-modal reconstruction tasks, with the aim to recover an image from its linearly corrupted measurements (e.g., blurred, undersampled or low-resolution). Unlike single-modal diffusion bridges operating on a single imaging modality, SelfRDB performs a translation task to map between distinct source and target modalities whose relationship is uncharacterized. Several recent computer vision studies have also built diffusion bridges for multi-modal translation tasks (Su et al., 2023; Liu et al., 2023a; Kim et al., 2024a). However, regular diffusion bridges were commonly employed based on a noise schedule with zero variance at start-end points corresponding to target-source images, yet high variance near the mid-point. This scheduling can hamper generalization to source images in the test set due to native variability in the level of measurement noise on the source modality (Chung et al., 2023), and induce substantial losses in tissue information near the mid-point of the diffusion process during source-to-target mapping (Liu et al., 2023a). To address these limitations, SelfRDB uniquely leverages a monotonically-increasing noise variance towards the end-point. Furthermore, compared to previous single- and multi-modal bridges that use a one-shot sampling process in reverse steps, SelfRDB leverages a novel self-consistent recursive estimation procedure to improve accuracy in generation of intermediate image samples.

3. Theory and methods

3.1. Diffusion bridges

Diffusion bridges are a general framework to describe the evolution between two arbitrary probability distributions across a finite time interval $t \in [0, T]$ (Liu et al., 2023a). In the context of mapping a

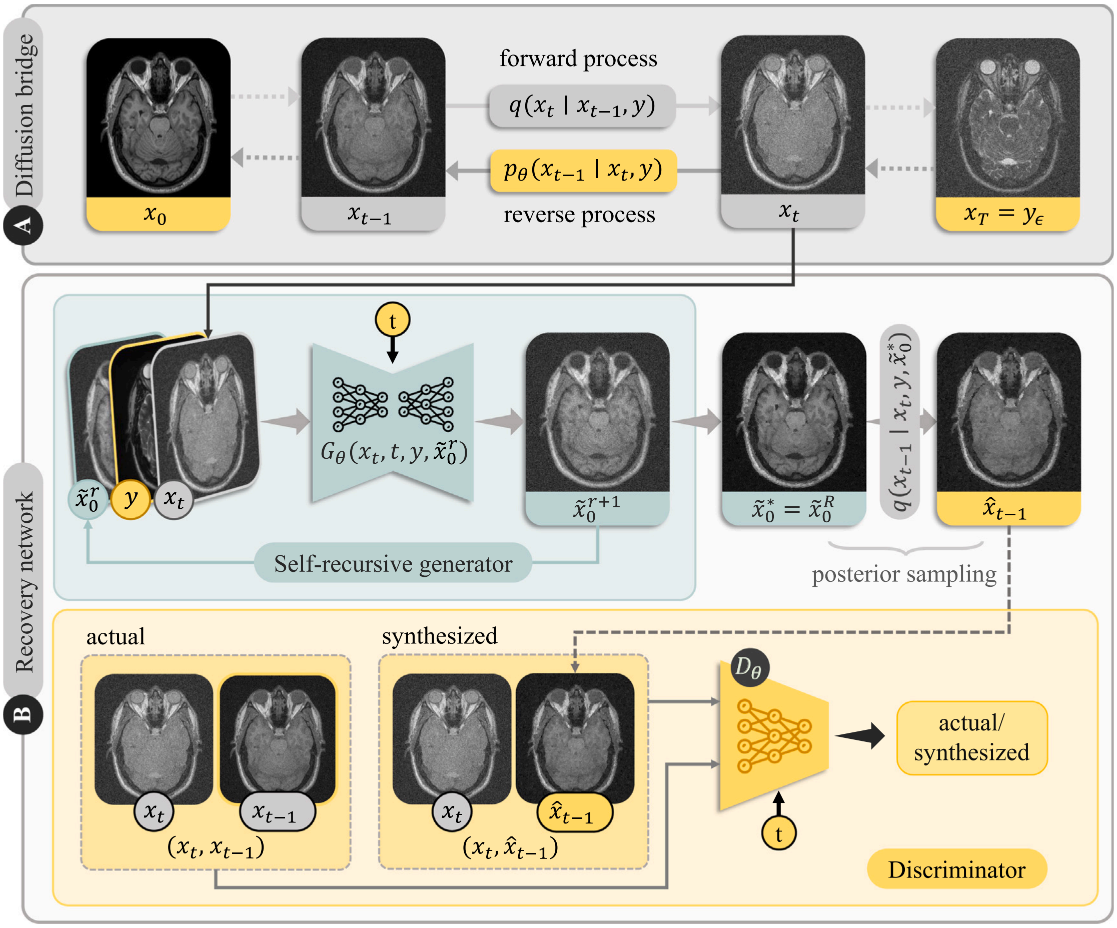

Fig. 3. SelfRDB casts a diffusion bridge between source and target images of an anatomy. (a) In the forward process, the start-point $\boldsymbol{x}_0$ is taken as the target image and the end-point $\boldsymbol{x}_T$ is taken as a noise-added version of the source image $\boldsymbol{y}_\epsilon$. Intermediate image samples are derived via the forward transition probability $q(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}, \boldsymbol{y})$, whose mean is a convex combination of target–source images, and whose variance is driven by noise. In the reverse process, sampling is initiated on $\boldsymbol{x}_T = \boldsymbol{y}_\epsilon$, and intermediate samples are derived via the reverse transition probability $p_\theta(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t, \boldsymbol{y}_\epsilon)$. (b) Reverse diffusion steps are operationalized via a recovery network $G_\theta(\boldsymbol{x}_t, t, \boldsymbol{y}, \tilde{\boldsymbol{x}}_0^r)$ that recursively generates a target-image estimate $\tilde{\boldsymbol{x}}_0^{r+1}$ at the current timestep, given the target-image estimate from the previous recursion $\tilde{\boldsymbol{x}}_0^r$ and the original source image $\boldsymbol{y}$. Recursions are stopped upon convergence onto a self-consistent solution $\tilde{\boldsymbol{x}}_0^* = G_\theta(\boldsymbol{x}_t, t, \boldsymbol{y}, \tilde{\boldsymbol{x}}_0^*)$, which is then used for posterior sampling of $\hat{\boldsymbol{x}}_{t-1}$ according to the normal distribution $q(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t, \boldsymbol{y}, \tilde{\boldsymbol{x}}_0^*)$. To improve posterior sampling, a discriminator subnetwork $D_\theta(\boldsymbol{x}_{t-1}$ or $\hat{\boldsymbol{x}}_{t-1}, t, \boldsymbol{x}_t)$ is used to perform adversarial learning on the recovered sample $\hat{\boldsymbol{x}}_{t-1}$.

source image $\boldsymbol{x}_T := \boldsymbol{y}$ onto a target image $\boldsymbol{x}_0$, the learning objective for diffusion bridges can be expressed as:

$$\min_{p \in \mathcal{P}_{[0,T]}} D_{\mathrm{KL}}(p \,\|\, q), \quad \text{s.t. } p_0 = p_{\text{target}}, p_T = p_{\text{source}}, \tag{1}$$

where $\mathcal{P}_{[0,T]}$ is the space of path measures with marginal densities for the target and source ($p_0 = p_{\text{target}}$ and $p_T = p_{\text{source}}$) taken as boundary conditions, and q is the reference path measure. Solution of (1) is the optimal path measure $p^* \in \mathcal{P}_{[0,T]}$ that can be described via the following forward-reverse stochastic differential equations (Chen et al., 2021):

$$\begin{aligned} \mathrm{d}\boldsymbol{x}_t &= [\boldsymbol{f} + g^2 \nabla \log \Psi(\boldsymbol{x}_t, t)]\mathrm{d}t + g\mathrm{d}\boldsymbol{w}_t, \ \boldsymbol{x}_0 \sim p_{\text{target}}, \\ \mathrm{d}\boldsymbol{x}_t &= [\boldsymbol{f} - g^2 \nabla \log \bar{\Psi}(\boldsymbol{x}_t, t)]\mathrm{d}t + g\mathrm{d}\bar{\boldsymbol{w}}_t, \ \boldsymbol{x}_T \sim p_{\text{source}}. \end{aligned} \tag{2}$$

Here, $\boldsymbol{f}$ is the drift coefficient, g is the diffusion coefficient, $\boldsymbol{w}_t, \bar{\boldsymbol{w}}_t$ are forward-reverse Wiener processes, and $\nabla \log \Psi(\boldsymbol{x}_t, t)$, $\nabla \log \bar{\Psi}(\boldsymbol{x}_t, t)$ are nonlinear forward-reverse drift terms related to the score function $\nabla \log p_t(\boldsymbol{x}_t)$ (Nelson, 1967). In contrast to DDMs based on linear drifts (Song et al., 2020), the nonlinear drifts in diffusion bridges enable the use of non-Gaussian p_{source}. A Dirac-delta distribution is assumed for the target modality, i.e., $p_0(\cdot) := \delta_x(\cdot)$, such that the marginal density at the start-point is taken as $p_0(\boldsymbol{x}_0) = 1$ given a (target, source) image pair $(\boldsymbol{x}_0, \boldsymbol{x}_T)$. This assumption ensures computational tractability by decoupling the constraints in Eq. (2) (Liu et al., 2023a).

In regular diffusion bridges, high-quality image pairs from target and source modalities are taken as start- and end-points of the diffusion process as in Eq. (1), with zero additive-noise variance at $t = 0$ and $t = T$ (Fig. 1b). This choice ensures optimal transport for training data, and enables the bridge to directly translate high-quality source images during inference (Liu et al., 2023a). However, it also constrains the bridge to capture *a hard-prior on the source modality*, since the end-point follows a Dirac-delta distribution based on source images in the training set, i.e., $p_T(\boldsymbol{x}_T) = 1$ given a training pair $(\boldsymbol{x}_0, \boldsymbol{x}_T)$. Combined with high noise variance near $t = T/2$, this distributional constraint can compromise generalization and information transfer from source-to-target images (Fig. 2a).

3.2. SelfRDB

SelfRDB is a novel diffusion bridge for medical image translation that maps the source image $\boldsymbol{y} \sim p_{\text{source}}$ of an anatomy onto the respective target image $\boldsymbol{x}_0 \sim p_{\text{target}}$. To do this, it leverages a novel forward process with a soft-prior on the source modality to improve noise robustness and facilitate information transfer, and a novel reverse process with self-consistent recursion to improve sampling accuracy.

3.2.1. Forward process with soft-prior on source modality

SelfRDB forms a diffusion bridge between $\boldsymbol{x}_0$ and $\boldsymbol{y}$ based on the following forward transition probability (Fig. 1):

$$q(\boldsymbol{x}_t \mid \boldsymbol{x}_0, \boldsymbol{y}) = \mathcal{N}(\boldsymbol{x}_t; \mu_{x_0,t}\boldsymbol{x}_0 + \mu_{y,t}\boldsymbol{y}, \sigma_t^2 \boldsymbol{I}), \tag{3}$$

where $\boldsymbol{x}_t$ is the intermediate image sample at timestep t, $\mathcal{N}$ denotes the Gaussian distribution, and $\boldsymbol{I}$ is the identity matrix. Accordingly, given an image pair $(\boldsymbol{x}_0,\ \boldsymbol{y})$, intermediate image samples are generated as follows:

$$\boldsymbol{x}_t = \mu_{x_0,t}\boldsymbol{x}_0 + \mu_{y,t}\boldsymbol{y} + \sigma_t \epsilon, \tag{4}$$

where $\epsilon \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ is a standard normal variable. Note that the mean of $\boldsymbol{x}_t$ is determined via a convex combination of target and source images with weights $\mu_{x_0,t}$, $\mu_{y,t}$ (Liu et al., 2023a):

$$\mu_{x_0,t} = \frac{\bar{s}_t^2}{\bar{s}_t^2 + s_t^2}, \quad \mu_{y,t} = \frac{s_t^2}{\bar{s}_t^2 + s_t^2}, \tag{5}$$

where $s_t^2 := \int_0^t g(\tau)d\tau$ and $\bar{s}_t^2 := \int_t^T g(\tau)d\tau$ are the time-accumulated diffusion coefficients in forward and reverse directions. To satisfy positivity and symmetry conditions with respect to the mid-point (Chen et al., 2023), here we propose to use the following diffusion coefficient:

$$g(t) \propto \frac{(T - |2t - T|)^2}{4T(T-1)^2}. \tag{6}$$

Meanwhile, the variance of $\boldsymbol{x}_t$ depends on the scale parameter σ_t^2. In regular diffusion bridges, σ_t^2 is defined to follow Dirac-delta constraints at start- and end-points (Liu et al., 2023a):

$$\sigma_t^2 = \frac{\bar{s}_t^2 s_t^2}{\bar{s}_t^2 + s_t^2} \text{ (regular bridge)}, \tag{7}$$

where σ_t^2 peaks at $t = T/2$ and is reduced to 0 at $t = T$ resulting in an end-point $\boldsymbol{x}_T = \boldsymbol{y} \sim \mathcal{N}(\boldsymbol{x}_T; \boldsymbol{y}, \boldsymbol{0})$. Regular bridges learn an exact mapping between a target image and its paired source image, resulting in a hard-prior on the source modality. This can compromise generalization during inference on a source image drawn from a low-density region of the data space poorly covered in the training set (Song and Ermon, 2019).

In contrast, SelfRDB adopts a novel noise variance schedule where σ_t^2 grows monotonically across t:

$$\sigma_t^2 = \gamma \frac{s_t}{\bar{s}_t^2 + s_t^2} \text{ (SelfRDB)}, \tag{8}$$

where γ is a scalar. Note that σ_t^2 ranges in $[0\ \gamma]$, so γ is tuned as a hyperparameter to control the level of noise corruption at the end-point. The above schedule elicits an end-point of a noise-added source image $\boldsymbol{x}_T = \boldsymbol{y}_\epsilon \sim \mathcal{N}(\boldsymbol{x}_T; \boldsymbol{y}, \sigma_T^2 \boldsymbol{I})$. Noise addition relaxes the Dirac-delta constraint on the data distribution at the end-point, and smooths the corresponding data space to enable more uniform coverage. In turn, SelfRDB learns a mapping between a target image and the neighborhood of its paired source image (Fig. 2). The resultant soft-prior serves to enhance the reliability of SelfRDB against noise-induced variability in source images, thereby boosting generalization.

3.2.2. Reverse process with self-consistent recursive estimation

SelfRDB casts a reverse process to progressively map the noise-added source image at the end-point $\boldsymbol{x}_T = \boldsymbol{y}_\epsilon$ back onto the target image $\boldsymbol{x}_0$ at the start-point (see Alg. 1). Since $\boldsymbol{y}_\epsilon$ is corrupted by additive noise, stationary guidance from the original source image $\boldsymbol{y}$ is also employed to avoid potential losses in tissue information. Starting sampling at $\boldsymbol{x}_T$, intermediate image samples are drawn based on a network operationalization of the reverse transition probability $p_\theta(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t, \boldsymbol{y}) := q(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t, \boldsymbol{y})$. An adversarial recovery network comprising a generator–discriminator pair $(G_\theta,\ D_\theta)$ with parameters θ is adopted here, as inspired by the recent success of adversarial diffusion models in image synthesis tasks (Xiao et al., 2022; Özbey et al., 2023).

Algorithm 1: Inference for SelfRDB

```
Input:
  $\boldsymbol{y}$: original source image, $\boldsymbol{y}_\epsilon$: noise-added source image
  $G_\theta(\boldsymbol{x}_t, t, \boldsymbol{y}, \tilde{\boldsymbol{x}}_0)$: recovery network
  $T$: number of diffusion steps
  $R$: number of recursions
Output:
  $\hat{\boldsymbol{x}}_0$: recovered target image
$\boldsymbol{x}_T = \boldsymbol{y}_\epsilon$    ▷ set end-point sample
for $t = T, \dots, 1$ do
    $\tilde{\boldsymbol{x}}_0^1 = \boldsymbol{0}$    ▷ initialize target-image estimate
    for $r = 1, \dots, R$ do
        $\tilde{\boldsymbol{x}}_0^{r+1} = G_\theta(\boldsymbol{x}_t, t, \boldsymbol{y}, \tilde{\boldsymbol{x}}_0^r)$    ▷ update estimate
    $\tilde{\boldsymbol{x}}_0^* = \tilde{\boldsymbol{x}}_0^R$    ▷ retrieve self-consistent estimate
    $\hat{\boldsymbol{x}}_{t-1} \sim q(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t, \boldsymbol{y}, \tilde{\boldsymbol{x}}_0^*)$    ▷ posterior sampling
return $\hat{\boldsymbol{x}}_0$
```

Note that reverse diffusion steps can be implemented by deriving an analytical expression for $p_\theta(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t, \boldsymbol{y})$ based on a reparametrization of the reverse transition probability as $q(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t, \boldsymbol{y}, \boldsymbol{x}_0)$ (Ho et al., 2020). Since the actual $\boldsymbol{x}_0$ is unknown at timestep t during reverse diffusion, the generator G_θ can be used to produce a target-image estimate $\tilde{\boldsymbol{x}}_0^*$ as a surrogate of $\boldsymbol{x}_0$, as in DDMs (Ho et al., 2020). However, during a given reverse step, common diffusion methods produce a one-shot image estimate of $\boldsymbol{x}_0$ by performing a single forward-pass through the recovery network as $\tilde{\boldsymbol{x}}_0^* = G_\theta(\boldsymbol{x}_t, t, \boldsymbol{y})$ (Özbey et al., 2023). This one-shot estimation procedure is susceptible to deviations in synthesized image samples (i.e., $\boldsymbol{x}_t$) from the true data distribution, especially when a moderate number of timesteps T are prescribed for the diffusion process. When these deviations accumulate across reverse diffusion steps, they can yield significant estimation errors in $\tilde{\boldsymbol{x}}_0^*$ (Peng et al., 2022).

To alleviate estimation errors, SelfRDB instead leverages a novel self-consistent recursive estimation procedure that performs multiple recursions via the recovery network. Specifically, individual recursions in SelfRDB's estimation procedure are expressed as a forward pass through G_θ:

$$\tilde{\boldsymbol{x}}_0^{r+1} = G_\theta(\boldsymbol{x}_t, t, \boldsymbol{y}, \tilde{\boldsymbol{x}}_0^r), \tag{9}$$

where $r \in \mathbb{Z}^+$ denotes the recursion index, and $\tilde{\boldsymbol{x}}_0^r$ is the target-image estimate at the rth recursion. Note that, unlike conventional recovery networks in previous diffusion methods that only receive as input $(\boldsymbol{x}_t, t, \boldsymbol{y})$, the recovery network in SelfRDB additionally receives $\tilde{\boldsymbol{x}}_0^r$ to enable recursions. Initially setting $\tilde{\boldsymbol{x}}_0^1 = \boldsymbol{0}$, recursions are continued until a self-consistent solution is obtained at the Rth recursion:

$$\tilde{\boldsymbol{x}}_0^R = G_\theta(\boldsymbol{x}_t, t, \boldsymbol{y}, \tilde{\boldsymbol{x}}_0^{R-1}), \text{ s.t. } \tilde{\boldsymbol{x}}_0^R \approx \tilde{\boldsymbol{x}}_0^{R-1}. \tag{10}$$

The target-image estimate for the current timestep is then taken as the self-consistent solution, i.e., $\tilde{\boldsymbol{x}}_0^* = \tilde{\boldsymbol{x}}_0^R$. This recursive estimation procedure gives a chance for the generator to correct intermittent estimation errors across recursions, thereby improving the accuracy of target-image estimates.

Once an accurate target-image estimate $\tilde{\boldsymbol{x}}_0^*$ is derived, the image sample at timestep $t-1$ can be drawn from the reparametrized posterior by taking $\tilde{\boldsymbol{x}}_0^*$ as a surrogate for $\boldsymbol{x}_0$:

$$\hat{\boldsymbol{x}}_{t-1} \sim q(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t, \boldsymbol{y}, \tilde{\boldsymbol{x}}_0^*) \tag{11}$$

Based on Bayes' rule and Markov property of the diffusion process (Ho et al., 2020), the posterior can be expressed as:

$$q(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t, \boldsymbol{y}, \tilde{\boldsymbol{x}}_0^*) = \frac{q(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}, \boldsymbol{y}) q(\boldsymbol{x}_{t-1} \mid \boldsymbol{y}, \tilde{\boldsymbol{x}}_0^*)}{q(\boldsymbol{x}_t \mid \boldsymbol{y}, \tilde{\boldsymbol{x}}_0^*)}. \tag{12}$$

Note that, in Eq. (12), the terms in the fractional expression can be computed based on the forward transition probability in Eq. (3). In turn, here we derive the posterior probability for the novel diffusion process in SelfRDB as a Gaussian distribution $\mathcal{N}(\boldsymbol{x}_{t-1}; \boldsymbol{m}, \boldsymbol{v})$ such that:

$$\begin{aligned}\boldsymbol{m} =& \frac{\sigma_{t-1}^2}{\sigma_t^2}\frac{\mu_{x_0,t}}{\mu_{x_0,t-1}}\boldsymbol{x}_t + (\mu_{y,t-1} - \mu_{y,t}\frac{\sigma_{t-1}^2}{\sigma_t^2}\frac{\mu_{x_0,t}}{\mu_{x_0,t-1}})\boldsymbol{y} \\ &+(1 - \mu_{y,t-1}\frac{\sigma_{t|t-1}^2}{\sigma_t^2})\tilde{\boldsymbol{x}}_0^*, \end{aligned} \tag{13}$$

$$\boldsymbol{v} = \sigma_{t|t-1}^2 \frac{\sigma_{t-1}^2}{\sigma_t^2}, \tag{14}$$

where $\sigma_{t|t-1}^2 = \sigma_t^2 - \sigma_{t-1}^2(\mu_{x_0,t}/\mu_{x_0,t-1})^2$.

The recovery network also employs the discriminator D_θ to distinguish the synthetic samples produced with the aid of the generator from the actual image samples drawn using the forward diffusion process. Conditioned on $\boldsymbol{x}_t$, D_θ predicts a logit of the input sample at timestep $t-1$:

$$c = D_\theta((\hat{\boldsymbol{x}}_{t-1} \text{ or } \boldsymbol{x}_{t-1}), t, \boldsymbol{x}_t). \tag{15}$$

3.2.3. Learning procedure

Given a training set of target-source image pairs $(\boldsymbol{x}_0, \boldsymbol{y})$, the forward process described in Eqs. (3)–(4) is used to generate corresponding intermediate samples $\boldsymbol{x}_t$ for $t \in [0\ T]$ that bridge between each image pair. Afterward, these intermediate samples are used to train the adversarial recovery network in SelfRDB. The generator aims to produce accurate target-image estimates $\tilde{\boldsymbol{x}}_0^*$ that subsequently elicit realistic intermediate image samples $\hat{\boldsymbol{x}}_{t-1}$. Thus, following self-consistent recursive estimation of $\tilde{\boldsymbol{x}}_0^*$, G_θ is trained using pixel-wise ℓ_1 and adversarial loss terms (Dar et al., 2019a):

$$\begin{aligned} L_{G_\theta} =& \mathbb{E}_{t,q(\boldsymbol{x}_t|\boldsymbol{x}_0,\boldsymbol{y}),p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t,\boldsymbol{y})}\big\{\lambda_1 \|\boldsymbol{x}_0 - \tilde{\boldsymbol{x}}_0^*\|_1 \\ &-log(D_\theta(\hat{\boldsymbol{x}}_{t-1}))\big\}, \end{aligned} \tag{16}$$

where $\mathbb{E}$ is expectation, λ_1 is the weight of the pixel-wise loss. Meanwhile, the discriminator primarily aims to distinguish between synthetic and actual intermediate image samples, so D_θ is trained an adversarial loss with a gradient penalty (Dar et al., 2019a):

$$\begin{aligned} L_{D_\theta} =& \mathbb{E}_{t,q(\boldsymbol{x}_t|\boldsymbol{x}_0,\boldsymbol{y})}\Big\{\mathbb{E}_{q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t,\boldsymbol{y})}\big\{-log(D_\theta(\boldsymbol{x}_{t-1}))\big\} \\ &+\mathbb{E}_{p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t,\boldsymbol{y})}\big\{-log(1 - D_\theta(\hat{\boldsymbol{x}}_{t-1}))\big\} \\ &+\lambda_2\mathbb{E}_{q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t,\boldsymbol{y})}\big\{\|\nabla \boldsymbol{x}_{t-1} D_\theta(\boldsymbol{x}_{t-1})\|_2^2\big\}\Big\}, \end{aligned} \tag{17}$$

where λ_2 is the relative weight of the gradient penalty (Özbey et al., 2023).

4. Experiments

4.1. Datasets

Experiments were conducted on two multi-contrast MRI datasets (IXI,[2] BRATS Menze et al., 2015) and a multi-modal MRI-CT dataset (Nyholm et al., 2018). In each dataset, a three-way split was performed to create training, validation and test sets without any subject overlap. Separate volumes of a subject were spatially registered via affine transformation (Jenkinson and Smith, 2001). Each volume was normalized to a mean intensity of 1, and voxel intensities were then normalized to a range of [−1, 1] across subjects. A consistent 256 × 256 cross-sectional image size was attained via zero-padding.

4.1.1. IXI dataset

T_1-, T_2-, PD-weighted brain images from 40 healthy subjects were analyzed. In each volume, 100 axial cross-sections with brain tissue were selected. (25, 5, 10) subjects were reserved for (training, validation, test) splits, respectively containing (2500, 500, 1000) cross-sections for each translation task.

4.1.2. BRATS dataset

T_1-, T_2-, Fluid Attenuation Inversion Recovery (FLAIR), and $\mathrm{T}_{1\mathrm{ce}}$-weighted brain images from 55 glioma patients were analyzed. In each volume, 100 axial cross-sections containing brain tissue were selected. (25, 10, 20) subjects were reserved for (training, validation, test) splits, respectively containing (2500, 1000, 2000) cross-sections for each translation task.

4.1.3. MRI-CT dataset

T_1-, T_2-weighted MRI, and CT images of the pelvis from 15 subjects were analyzed. In each volume, 90 axial cross-sections were selected. (9, 2, 4) subjects were reserved for (training, validation, test) splits, respectively containing (810, 180, 360) cross-sections for each translation task.

4.2. Competing methods

SelfRDB was demonstrated against state-of-the-art methods based on diffusion bridge, DDM and GAN models. All competing methods were trained via supervised learning on paired source and target modalities. For each method, hyperparameter selection was performed to maximize performance on the validation set. The selected parameters included number of epochs, learning rate, loss-term weights, and the number of diffusion steps (for diffusion-based methods). For a given method, a common set of parameters was selected that attained near-optimal validation performance across translation tasks.

4.2.1. SelfRDB

SelfRDB comprised generator and discriminator subnetworks. The generator was implemented with a residual UNet backbone with 12 residual stages equally split between encoding and decoding modules (Ronneberger et al., 2015). Each residual stage halved spatial resolution in the encoder, and doubled spatial resolution in the decoder module. Learnable time embeddings were computed via a multi-layer perceptron that received as input a 256-dimensional sinusoidal time encoding (Ho et al., 2020). The time embeddings were added onto feature maps in each generator stage. The discriminator was implemented with a convolutional backbone with 6 stages (Güngör et al., 2023). Each stage halved spatial resolution, and time embeddings were also added onto feature maps in each discriminator stage. Cross-validated hyperparameters were 50 epochs, 10^{-4} learning rate, $T = 10$, $\gamma = 2.2$, $\lambda_1 = 1$, $\lambda_2 = 1$. For recursive estimation of $\tilde{\boldsymbol{x}}_0^*$, convergence was assumed when the relative change in $\tilde{\boldsymbol{x}}_0^r$ between consecutive recursions fell below 1%, and this criteria was met for $R = 2$.

4.2.2. DDIB

A diffusion bridge model was considered with architecture, noise schedule and loss functions adopted from Su et al. (2023). The source modality was input as stationary guidance to reverse diffusion steps. Cross-validated hyperparameters were 50 epochs, 10^{-4} learning rate, $T = 1000$.

4.2.3. I^2 SB

A diffusion bridge model was considered with architecture, noise schedule and loss functions adopted from Liu et al. (2023a). The forward diffusion process mapped between source and target modalities. Cross-validated hyperparameters were 50 epochs, 10^{-4} learning rate, $T = 1000$.

[2] https://brain-development.org/ixi-dataset/.

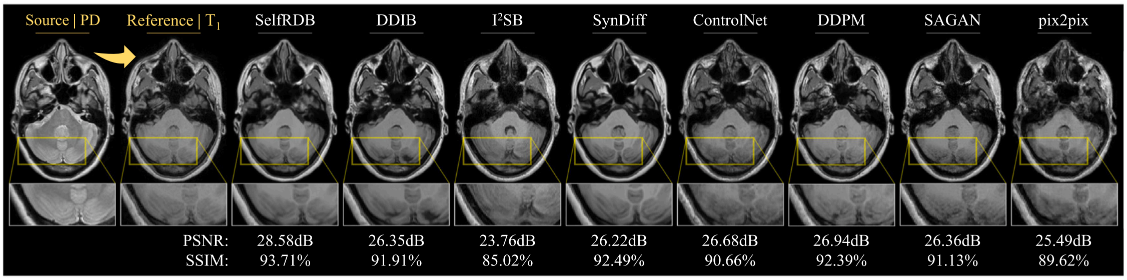

Fig. 4. Multi-contrast MRI translation for a representative PD→T_1 task in the IXI dataset. Synthesized target images for competing methods are shown along with the reference target image (i.e., ground truth) and the input source image. Zoom-in display windows are used to highlight differences in synthesis performance.

Table 1
Multi-contrast MRI translation in IXI. PSNR (dB) and SSIM (%) are listed as mean ± std across the test set, along with FID. Boldface marks the top-performing model in each task.

	T_2→T_1			T_1→T_2			PD→T_1			T_1→PD		
	PSNR ↑	SSIM ↑	FID ↓	PSNR ↑	SSIM ↑	FID ↓	PSNR ↑	SSIM ↑	FID ↓	PSNR ↑	SSIM ↑	FID ↓
pix2pix (Isola et al., 2017)	26.56 ±0.93	90.84 ±1.70	1.05	26.44 ±0.72	89.83 ±1.71	4.05	27.19 ±0.66	91.37 ±1.43	2.00	26.65 ±0.79	90.56 ±1.51	0.90
SAGAN (Zhang et al., 2019)	29.25 ±2.97	93.42 ±3.58	0.87	29.72 ±2.86	93.39 ±3.38	2.46	28.57 ±2.67	93.34 ±3.36	1.02	30.34 ±3.02	92.92 ±3.37	0.76
DDPM (Nichol and Dhariwal, 2021)	29.08 ±1.05	93.43 ±1.35	1.02	29.89 ±1.29	94.31 ±1.33	3.39	29.98 ±0.89	94.45 ±1.04	0.95	30.58 ±1.27	93.76 ±1.08	0.52
ControlNet (Pinaya et al., 2023)	28.59 ±1.03	92.55 ±1.30	0.67	28.09 ±1.00	89.41 ±1.51	3.03	27.97 ±0.77	92.13 ±1.28	1.37	29.26 ±1.10	91.36 ±1.27	1.01
SynDiff (Özbey et al., 2023)	30.13 ±1.38	94.60 ±1.23	0.55	30.19 ±1.45	94.24 ±1.36	2.67	29.74 ±1.34	94.81 ±1.12	1.16	30.89 ±1.42	94.20 ±1.04	0.57
I^2SB (Liu et al., 2023a)	21.07 ±0.47	47.06 ±1.85	2.57	21.98 ±0.55	77.61 ±1.90	10.37	21.61 ±0.42	77.95 ±1.81	2.63	24.88 ±0.80	79.44 ±1.94	1.21
DDIB (Su et al., 2023)	30.47 ±1.28	94.54 ±1.34	0.72	29.88 ±1.18	93.91 ±1.31	2.56	29.48 ±1.18	94.55 ±1.13	1.20	30.81 ±1.39	94.09 ±1.02	0.58
SelfRDB	**31.63** **±1.53**	**95.64** **±1.12**	**0.40**	**31.28** **±1.56**	**95.03** **±1.27**	**1.62**	**31.23** **±1.22**	**95.64** **±0.99**	**0.68**	**32.17** **±1.57**	**95.15** **±0.99**	**0.49**

4.2.4. SynDiff

A DDM model was considered with architecture, noise schedule and loss functions adopted from Özbey et al. (2023). Cross-validated hyperparameters were 50 epochs, 15×10^{-4} learning rate, $T = 1000$, $k = 250$ step size, adversarial loss weight of 1.

4.2.5. ControlNet

A latent DDM model was considered with architecture, noise schedule and loss functions adopted from the Generative Brain ControlNet[3] repository (Pinaya et al., 2022b, 2023). Cross-validated hyperparameters were 50 epochs, 2×10^{-4} learning rate and T = 1000.

4.2.6. DDPM

A DDM model was considered with architecture, noise schedule and loss functions adopted from Nichol and Dhariwal (2021). The source modality was input as stationary guidance to reverse diffusion steps. Cross-validated hyperparameters were 50 epochs, 10^{-4} learning rate, $T = 1000$.

4.2.7. SAGAN

A self-attention GAN (SAGAN) model was considered with architecture and loss functions adopted from Zhang et al. (2019). Cross-validated hyperparameters were 200 epochs, 2×10^{-4} learning rate, and adversarial loss weight of 0.1.

4.2.8. pix2pix

A GAN model was considered with architecture and loss functions adopted from Isola et al. (2017). Cross-validated hyperparameters were 200 epochs, 10^{-3} learning rate, and adversarial loss weight of 0.01.

4.3. Modeling procedures

Models were implemented via the PyTorch framework and executed on Nvidia RTX 4090 GPUs. For training, Adam optimizer was used with $\beta_1 = 0.5$, $\beta_2 = 0.9$. For evaluation, a single target image was synthesized from the respective source image for each cross section. Model performance was assessed via peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and Frechet inception distance (FID) metrics. PSNR and SSIM were reported as mean±standard deviation (std), whereas FID was reported as an aggregate measure across the test set. Prior to assessment, all images were normalized to a range of [0, 1]. The significance of performance differences in PSNR and SSIM was examined via non-parametric Wilcoxon signed-rank tests ($p < 0.05$).

5. Results

5.1. Multi-contrast MRI translation

We first demonstrated SelfRDB in multi-contrast MRI translation tasks. The proposed method was compared against diffusion bridge (DDIB, I^2SB), DDM (SynDiff, ControlNet, DDPM), and GAN models (pix2pix, SAGAN). Evaluations were first conducted on the IXI dataset that contains brain images from healthy subjects. Performance metrics in IXI are listed in Table 1. SelfRDB achieves the best performance

[3] https://github.com/Warvito/generative_brain_controlnet.

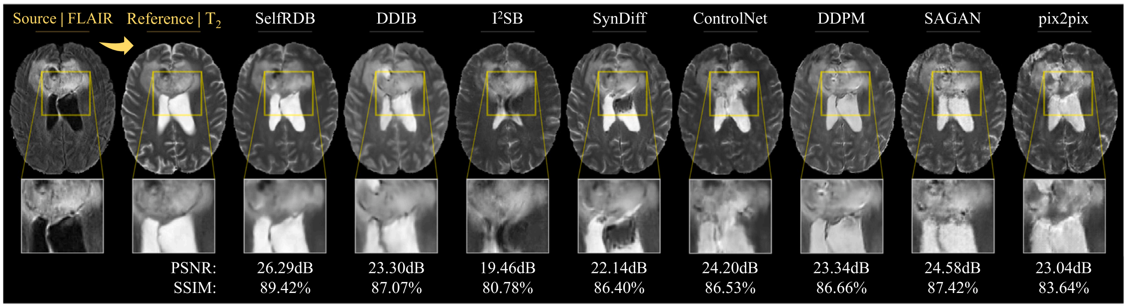

Fig. 5. Multi-contrast MRI translation for a representative FLAIR→T_2 task in the BRATS dataset. Synthesized target images for competing methods are shown along with the reference target image (i.e., ground truth) and the input source image.

Table 2
Multi-contrast MRI translation in BRATS. PSNR (dB) and SSIM (%) are listed as mean ± std across the test set, along with FID.

	T_2→T_1			T_1→T_2			FLAIR→T_2			T_2→FLAIR		
	PSNR ↑	SSIM ↑	FID ↓	PSNR ↑	SSIM ↑	FID ↓	PSNR ↑	SSIM ↑	FID ↓	PSNR ↑	SSIM ↑	FID ↓
pix2pix	26.97 ± 1.33	91.36 ± 2.50	0.56	26.59 ± 1.74	91.14 ± 2.64	4.40	25.25 ± 1.59	89.00 ± 2.95	4.98	26.85 ± 1.57	86.36 ± 3.11	1.50
SAGAN	27.78 ± 1.23	92.05 ± 2.30	0.46	26.82 ± 1.65	91.44 ± 2.58	3.86	26.08 ± 1.66	90.31 ± 2.98	4.24	27.08 ± 1.57	87.59 ± 2.93	1.40
DDPM	27.47 ± 1.28	92.24 ± 1.99	0.53	25.97 ± 2.09	90.24 ± 3.41	4.10	25.40 ± 1.76	89.79 ± 3.01	8.38	26.90 ± 1.84	87.86 ± 2.85	4.46
ControlNet	27.83 ± 1.24	92.10 ± 2.19	0.45	26.75 ± 1.78	91.52 ± 2.66	4.01	25.60 ± 1.77	89.56 ± 3.25	3.25	27.39 ± 1.52	88.44 ± 2.83	1.45
SynDiff	27.78 ± 1.72	93.05 ± 2.18	0.55	22.21 ± 1.52	87.93 ± 2.48	6.83	26.14 ± 1.91	91.01 ± 3.05	4.23	27.77 ± 1.77	89.62 ± 2.87	2.11
I^2SB	22.24 ± 2.18	79.87 ± 5.84	1.14	21.80 ± 2.33	80.75 ± 5.49	5.89	23.28 ± 2.33	84.38 ± 4.34	12.63	25.92 ± 2.00	83.51 ± 4.05	6.58
DDIB	27.77 ± 1.32	92.57 ± 2.17	0.67	25.44 ± 1.90	91.03 ± 2.89	4.38	25.52 ± 1.63	89.45 ± 2.90	11.80	24.51 ± 1.86	85.02 ± 3.06	2.91
SelfRDB	**28.85 ± 1.48**	**93.70 ± 1.87**	**0.43**	**27.58 ± 1.88**	**92.99 ± 2.44**	**3.55**	**26.85 ± 1.75**	**91.66 ± 2.72**	**3.15**	**27.98 ± 1.80**	**90.01 ± 2.70**	**1.38**

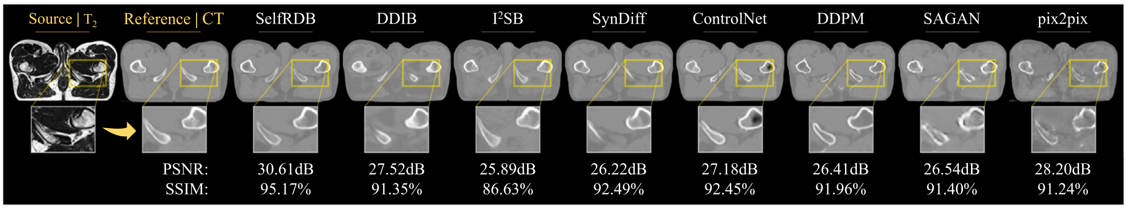

Fig. 6. Multi-modal MRI-CT translation for a representative T_2→CT task in the pelvic dataset. Synthesized target images for competing methods are shown along with the reference target image (i.e., ground truth) and the input source image.

metrics in each individual task, significantly outperforming baselines in PSNR/SSIM ($p < 0.05$). On average, SelfRDB outperforms diffusion bridges by 5.34 dB PSNR, 12.86% SSIM, 1.93 FID; DDMs by 2.05 dB PSNR, 2.09% SSIM, 0.61 FID; and GANs by 3.49 dB PSNR, 3.41% SSIM, 0.84 FID. Note that the relatively low performance of regular diffusion bridges suggest that their diffusion processes governed by the underlying noise schedules might not be suited to successfully translate between source and target MRI contrasts. Representative target images synthesized by competing methods are displayed in Fig. 4. Among diffusion-based baselines, DDIB and DDPM show susceptibility to hallucinations manifested as over-dark or over-bright signals that deviate from the actual tissue appearances in ground-truth images, I^2SB shows a degree of noise amplification and poor anatomical fidelity to the target modality, SynDiff can over-flatten tissue signals that yields loss of spatially-graded tissue features, and ControlNet shows a degree of spatial blur yielding suboptimal structure depiction. Among GAN-based baselines, pix2pix suffers from structural inaccuracies and residual noise-like artifacts, whereas SAGAN shows a degree of noise amplification and residual ringing artifacts that mask underlying tissue structure. In contrast, SelfRDB synthesizes target images with low artifact/noise levels and reliable depiction of fine structural features of brain tissues.

We then evaluated competing methods on the BRATS dataset that contains brain images from glioma patients. Performance metrics in BRATS are listed in Table 2. Again, we find that SelfRDB achieves the best performance metrics in each individual task, significantly outperforming baselines in PSNR/SSIM ($p < 0.05$). On average, SelfRDB outperforms diffusion bridges by 3.26 dB PSNR, 6.27% SSIM, 3.62 FID; DDMs by 1.38 dB PSNR, 1.81% SSIM, 1.24 FID; and GAN models by 1.14 dB PSNR, 2.18% SSIM, 0.55 FID. Representative target images synthesized by competing methods are displayed in Fig. 5. Corroborating the multi-contrast MRI translation results on the IXI dataset, DDIB suffers from hallucinatory features lead to inaccurate depiction of hyper- or hypo-intense tissue signals, and I^2SB shows poor anatomical consistency. Meanwhile, SynDiff shows regions of gross intensity errors near CSF tissue, likely due to leakage of signal intensity and image artifacts from the source modality, DDPM shows a degree of contrast loss and misses the hypo-intense signal region within the tumor lesion, and ControlNet shows a degree of structural inaccuracy and suboptimal contrast depiction. Among GAN-based baselines, SAGAN suffers from dark-pixel artifacts, noise amplification and structural inaccuracies, while pix2pix shows suboptimal contrast depiction and a degree of loss in structural details. In comparison, SelfRDB synthesizes target images with lower noise/artifact levels and more accurate anatomical depiction near both tumor lesions and healthy tissues in multi-contrast MRI scans.

Table 3
Multi-modal MRI-CT translation in the pelvic dataset. PSNR (dB) and SSIM (%) listed as mean ± std across the test set, along with FID.

	T_2→CT			T_1→CT		
	PSNR ↑	SSIM ↑	FID ↓	PSNR ↑	SSIM ↑	FID ↓
pix2pix	25.16 ± 2.08	87.33 ± 2.09	4.38	25.65 ± 2.22	84.18 ± 2.90	9.54
SAGAN	27.21 ± 1.89	90.58 ± 2.00	6.15	26.77 ± 2.20	91.00 ± 2.35	8.32
DDPM	26.88 ± 1.96	91.18 ± 2.03	4.57	26.39 ± 2.54	90.62 ± 5.04	10.57
ControlNet	27.95 ± 1.59	91.66 ± 1.44	6.36	27.39 ± 2.41	91.81 ± 2.44	11.79
SynDiff	26.54 ± 2.01	89.59 ± 2.61	8.30	27.41 ± 4.68	92.07 ± 5.32	11.04
I^{2}SB	26.54 ± 1.80	85.94 ± 2.47	3.71	25.21 ± 2.59	84.82 ± 7.40	13.35
DDIB	26.95 ± 1.42	90.05 ± 1.79	5.64	26.72 ± 2.94	91.74 ± 5.11	8.42
SelfRDB	**29.46 ± 2.15**	**93.62 ± 1.72**	**3.48**	**27.55 ± 3.32**	**92.29 ± 6.32**	**8.20**

Table 4
Challenging translation tasks. PSNR (dB) and SSIM (%) listed as mean ± std across the test set, along with FID.

	CT→T_1 (MRI-CT)			T_{1ce}→FLAIR (BRATS)		
	PSNR ↑	SSIM ↑	FID ↓	PSNR ↑	SSIM ↑	FID ↓
pix2pix	20.23 ± 1.06	71.17 ± 4.70	2.55	24.15 ± 2.07	85.23 ± 3.76	3.52
SAGAN	20.59 ± 1.28	81.74 ± 2.45	1.92	24.76 ± 1.94	85.70 ± 3.96	2.93
DDPM	20.58 ± 1.35	81.45 ± 2.81	1.64	24.65 ± 1.88	85.44 ± 3.94	2.92
ControlNet	20.47 ± 1.05	77.96 ± 2.47	3.33	24.31 ± 1.86	84.83 ± 4.06	2.98
SynDiff	17.09 ± 1.18	64.82 ± 4.32	2.51	24.40 ± 1.85	86.08 ± 4.87	3.73
I^{2}SB	16.29 ± 1.68	66.88 ± 4.11	5.80	24.38 ± 2.44	83.55 ± 5.26	3.00
DDIB	20.30 ± 1.15	82.09 ± 2.75	2.12	23.22 ± 1.84	83.21 ± 4.47	5.28
SelfRDB	**21.12 ± 1.49**	**82.50 ± 2.77**	**1.16**	**25.29 ± 2.24**	**87.46 ± 3.60**	**2.86**

5.2. Multi-modal MRI-CT translation

Next, we demonstrated SelfRDB in multi-modal MRI-CT translation tasks via comparisons against baselines on the pelvic MRI-CT dataset that contains healthy subjects. Performance metrics in the pelvic dataset are listed in Table 3. SelfRDB achieves the best performance metrics in each individual task, significantly outperforming baselines in PSNR/SSIM ($p < 0.05$). On average, SelfRDB outperforms diffusion bridges by 2.15 dB PSNR, 4.82% SSIM, 1.94 FID; DDMs by 1.41 dB PSNR, 1.80% SSIM, 2.93 FID; and GANs by 2.31 dB PSNR, 4.68% SSIM, 1.26 FID. Representative target images synthesized by competing methods are displayed in Fig. 6. Among diffusion-based baselines, DDIB and DDPM manifest hallucinatory bright features that resemble pelvic bone tissue, and DDIB also suffers from contrast losses across muscle tissue. Meanwhile, I^{2}SB and SynDiff show a degree of geometric distortion particularly evident in hyper-intense signal regions that cause deviation from the true structure of bone tissues, and ControlNet manifests some hallucinatory dark features that alter the tissue contrast. Among GAN-based baselines, SAGAN and pix2pix suffer from a degree of spatial blurring and loss of structural details. Compared against baselines, SelfRDB synthesizes target images with lower artifacts/noise, depicting pelvic bone and muscle tissues with a relatively high degree of anatomical accuracy.

5.3. Challenging translation tasks

Mapping between endogenous MRI contrasts and mapping MRI contrasts onto CT images are relatively well posed translation tasks, where the source modality carries substantial information regarding the target modality (Chartsias et al., 2018; Nie et al., 2018). Our results indicate that SelfRDB performs favorably against baselines and synthesizes high-fidelity target images in these tasks. Yet, there can be other imaging scenarios where the source modality carries weaker information related to the target modality, elevating the difficulty of the translation task. Accordingly, we performed demonstrations on two challenging translation tasks: predicting MRI from CT images, and predicting an endogenous MRI contrast from an exogenous MRI contrast. Table 4 lists performance metrics for competing methods. As expected, translation methods generally show relatively lower performance under elevated task difficulty. Yet, we find that SelfRDB still achieves the best performance metrics, significantly outperforming baselines in PSNR/SSIM ($p < 0.05$). On average, SelfRDB outperforms diffusion bridges by 2.16 dB PSNR, 6.05% SSIM, 2.04 FID; DDMs by 1.29 dB PSNR, 4.88% SSIM, 0.84 FID; and GAN models by 1.29 dB PSNR, 4.02% SSIM, 0.72 FID. These findings suggest that SelfRDB maintains its competitive performance against baselines even in tasks where the source modality carries relatively weaker information on the target modality.

5.4. Computational efficiency

A key consideration in medical image translation is the computational complexity of the translation models. Table 5 lists the training time per cross-section, inference time per cross section, and memory use of competing methods. As expected, GAN-based methods SAGAN and pix2pix that synthesize target images in a single forward pass through the generator network have notably low training/inference times. Among diffusion-based methods, SynDiff that includes diffusive and non-diffusive modules in its network architecture and I^{2}SB that employs a relatively more complex network architecture have the highest training times, whereas ControlNet that fine-tunes the encoder of a pre-trained latent DDM has the lowest training time overall. Meanwhile, DDIB, DDPM, and SelfRDB have moderate training times that are competitive with GAN methods. Note that the run time of diffusion methods during inference depend not only on the complexity of the network architecture but also on the number of sampling steps. SynDiff that employs 4 steps and SelfRDB that employs 10 steps offer more competitive run times to GAN-based methods, whereas remaining diffusion-based methods that use relatively large number of steps elicit prolonged inference times. In terms of memory use during inference, GAN-based and diffusion-based methods are generally comparable, except for I^{2}SB that yields significantly higher memory load as it employs a relatively large network architecture.

5.5. Ablation studies

We first conducted a set of ablation studies to examine the contribution of main design elements in SelfRDB to translation performance. For this purpose, variant models were formed by selectively ablating an individual design element from SelfRDB, while the remaining elements remained intact. To assess the importance of the soft prior on the source modality, we formed a variant model with a hard prior on the source

Table 5

Average training times per cross-section (sec), inference times per cross-section (sec) and memory load (gigabytes).

	pix2pix	SAGAN	DDPM	ControlNet	SynDiff	I^2SB	DDIB	SelfRDB
Training	0.033	0.081	0.134	0.039	1.824	0.790	0.065	0.170
Inference	0.003	0.007	43.484	76.198	0.135	55.203	19.575	0.372
Memory	0.64	1.34	1.52	2.06	2.15	10.26	2.10	2.22

Table 6

Performance of SelfRDB variants on representative medical image translation tasks. A variant ablated of a soft prior on the source modality, a variant ablated of self-consistent target-image estimates, a variant that received stationary guidance from the noise-added source image ($\boldsymbol{y}_\epsilon$) instead of the original source image ($\boldsymbol{y}$), and a variant that entirely ablated stationary source-image guidance were considered.

	T_1→T_2 (IXI)			T_2→T_1 (BRATS)			T_2→CT (MRI-CT)		
	PSNR ↑	SSIM ↑	FID ↓	PSNR ↑	SSIM ↑	FID ↓	PSNR ↑	SSIM ↑	FID
SelfRDB	**31.28 ± 1.56**	**95.03 ± 1.27**	**1.62**	**28.85 ± 1.48**	**93.70 ± 1.87**	**0.43**	**29.46 ± 2.15**	**93.62 ± 1.72**	**3.48**
w/o soft prior	30.64 ± 1.42	94.33 ± 1.27	1.77	26.98 ± 1.54	92.00 ± 2.20	0.52	28.54 ± 1.91	92.72 ± 1.73	4.18
w/o self-consistency	28.89 ± 1.15	93.63 ± 1.45	6.08	27.73 ± 1.44	91.62 ± 2.54	0.47	28.89 ± 1.97	92.76 ± 1.84	5.64
w $\boldsymbol{y}_\epsilon$-guidance	23.34 ± 0.62	81.13 ± 2.18	10.61	25.48 ± 1.23	86.90 ± 3.56	0.65	26.68 ± 2.06	89.74 ± 2.55	4.44
w/o source guidance	20.97 ± 0.68	75.95 ± 2.22	16.71	21.45 ± 1.82	78.72 ± 5.87	1.90	24.60 ± 1.33	83.40 ± 2.29	6.78

modality attained by adopting noise scheduling from common diffusion bridges, i.e., by prescribing zero noise variance at start- and end-points and high noise variance at mid-point of the diffusion process (Liu et al., 2023a). To assess the importance of self-consistent recursive estimation procedure in deriving target-image estimates, we formed a variant model that performed a one-shot estimation of $\boldsymbol{x}_0$ in each reverse step without performing any recursions (Ho et al., 2020). Lastly, we formed two variant models to assess the importance of stationary guidance from the original source image $\boldsymbol{y}$. In a first variant, the recovery network received stationary guidance from the noise-added source image $\boldsymbol{y}_\epsilon$ used to initiate the diffusion process at timestep T instead of $\boldsymbol{y}$. In a second variant, the recovery network did not receive any stationary source-image guidance. Table 6 lists performance metrics for SelfRDB and ablated variants on representative translation tasks. We find that SelfRDB achieves the best performance metrics in all translation tasks, significantly outperforming ablated variants in PSNR/SSIM ($p < 0.05$). The performance improvements that SelfRDB demonstrates over variants that systematically remove source-image guidance highlight the critical role of stationary anatomical guidance from clean source images in synthesizing high-quality target images. Taken together, these findings indicate that each proposed design element in SelfRDB makes an important contribution to its performance in multi-contrast MRI and multi-modal medical image translation.

The primary motivation in SelfRDB for leveraging a soft prior on the source modality is to improve robustness against perturbations induced by measurement noise on source images. Thus, we also performed ablation studies to assess reliability against varying noise levels in the source modality between training and test sets. SelfRDB with its soft prior was compared against a variant model with a hard prior built by adopting noise variance scheduling from common diffusion bridges (Liu et al., 2023a). To control the noise level in source images, zero-mean bivariate Gaussian white noise was added onto each cross-section at std. values ranging in [0.04 0.1] (Chung et al., 2022a). Fig. 7 plots performance of models trained on original images without perturbations when tested on images subjected to additive measurement noise at varying levels. Naturally, all models show a growing degree of performance loss under increasing noise perturbation, compared with their performance on the original source images without perturbations. Yet, the average performance losses across noise levels are 4.72 dB PSNR, 7.77% SSIM, 4.80 FID for the hard-prior variant, whereas a more modest 3.08 dB PSNR, 4.30% SSIM, 3.79 FID for SelfRDB. The relatively limited performance losses in SelfRDB suggest that the proposed soft prior on the source modality helps maintain a degree of reliability against measurement noise.

Lastly, we performed ablation studies to examine the influence of the number of recursions (R) used to produce target-image estimates in each reverse diffusion step. Fig. 8 displays performance metrics for SelfRDB variants based on varying $R \in [1\ 7]$. We find that $R = 2$ elicits a notable improvement in synthesized image quality over $R = 1$, albeit further increases in R do not yield major improvements. Thus, we deduce that the selected value of $R = 2$ offers a favorable trade-off between translation performance and computational complexity due to additional recursions.

6. Discussion

SelfRDB is a novel diffusion bridge for medical image translation that progressively transforms a source modality onto a target modality. Compared to GANs that are amenable to training instabilities, it is a diffusion-based method that builds an explicit prior to improve image fidelity. Compared to DDMs that are trained to learn a task-irrelevant noise-to-target (i.e., denoising) transformation, it directly learns a source-to-target transformation of high task relevance. Compared to regular diffusion bridges, it leverages enhanced noise scheduling and estimation procedures to boost sampling accuracy. Our demonstrations on multi-modal translation tasks clearly suggest that these unique technical attributes help SelfRDB to significantly improve performance over state-of-the-art baselines.

Several technical limitations could be addressed to further boost the performance and efficiency of SelfRDB in medical image translation. The first set of improvements concerns the reliability of translation models. SelfRDB draws intermediate image samples from a normal posterior probability similar to other diffusion-based methods, so it produces stochastic target images. Corroborating recent reports (Özbey et al., 2023), here we observed that multiple target images independently synthesized by SelfRDB show nominal variability (unreported). While this might be attributed to the diminishing noise variance towards the start-point of the diffusion process corresponding to the target modality, future research is warranted to evaluate the uncertainty of diffusion bridges in medical image translation. Note that, although SelfRDB is inherently a diffusion-based method, it employs an adversarial loss component that could induce susceptibility to training instabilities (Lan et al., 2020). Here, we did not observe any notable sign of instabilities such as mode collapse or poor convergence when inspecting training and validation performance. Yet, when needed, spectral normalization or feature matching techniques could be adopted to improve training stability (Lan et al., 2020).

A second set of improvements concerns the selection of source–target modalities for the translation models. Here, we primarily examined mappings between different endogenous MRI contrasts (e.g., T_1, T_2), and prediction of CT from MRI contrasts. SelfRDB showed reliable translation performance in these tasks, suggesting that tissue information required to synthesize target modalities is present to a high degree in source modalities. Yet, we found that all competing methods yielded relatively lower performance when mapping between exogenous MRI

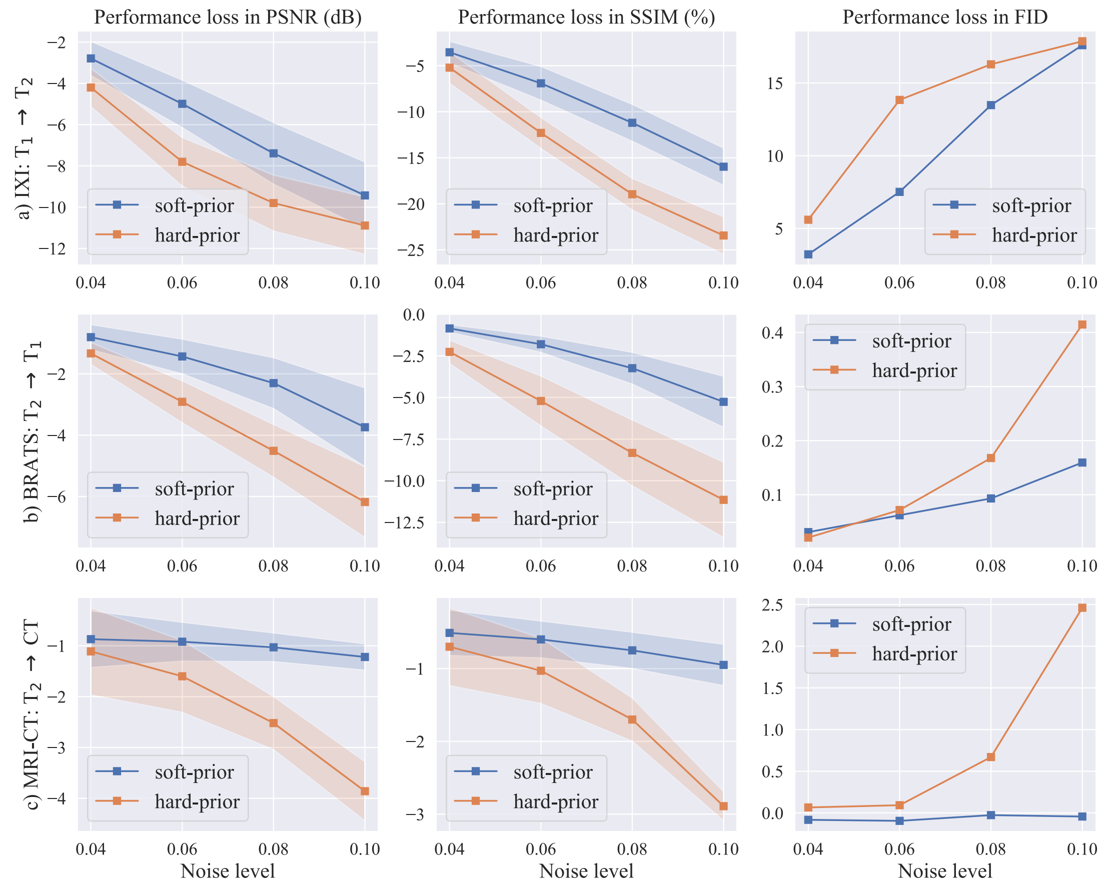

Fig. 7. Performance of soft-prior and hard-prior variants of SelfRDB in representative tasks. Models trained on original source images were tested on source images corrupted with varying levels of additive noise. Performance losses in terms of PSNR (left column), SSIM (middle column) and FID (right column) are plotted in reference to the performance levels based on original source images. Solid lines depict the average loss, and surrounding shaded regions depict the 65% confidence interval across the test set.

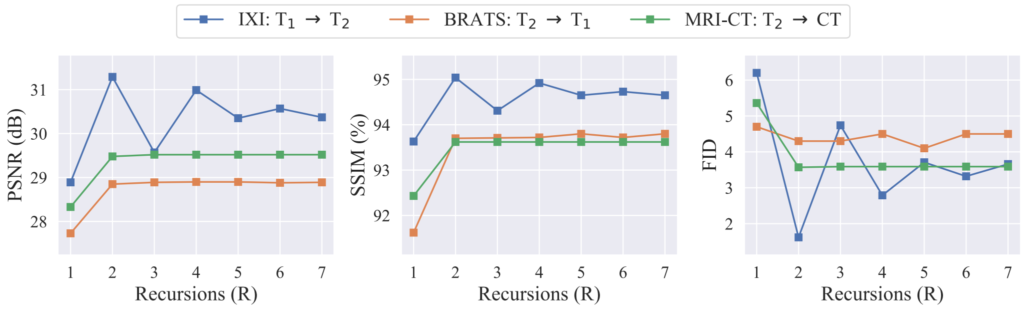

Fig. 8. Performance of SelfRDB variants that prescribe varying numbers of recursions to obtain target-image estimates in representative tasks. PSNR (left column), SSIM (middle column) and FID (right column; 10× values shown for BRATS) are plotted as a function of *R*.

contrasts based on injection of external contrast agents and endogenous MRI contrasts (Lee et al., 2019), and when predicting MR images with enhanced soft-tissue differentiation from CT images with primarily bone and soft-tissue differentiation (Özbey et al., 2023). These results suggest that when the information needed to synthesize the target modality is present to a weaker degree in the source modality, elevated difficulty of the translation task can hamper model performance. In such scenarios, translation performance might be improved by conducting many-to-one mappings that include additional source modalities to increase the degree of correlated tissue information with the target modality (Sharma and Hamarneh, 2020; Yurt et al., 2021), by employing test-time adaptation procedures or learned regularization terms to refine the synthetic target images (Nie et al., 2018; Ge et al., 2019; Elmas et al., 2023), or by acquiring a limited amount of target-modality data for improved guidance during target-image synthesis (Yurt et al., 2022). It remains important future work to assess the reliability of SelfRDB in a greater variety of challenging translation tasks, and the utility of the abovementioned approaches for mitigating performance losses in such conditions. This includes extending validation to additional datasets with varying distributions and characteristics to further corroborate the generalizability of the proposed method.

Lastly, a third set of improvements concerns the learning procedures, representational capacity and efficiency of translation models. Here, models were trained via supervised learning on paired source–target images within individual subjects. However, it may not always be feasible to curate paired training sets of sufficient size for adequate model training. In such cases, unsupervised model training on unpaired data can be performed by employing cycle-consistent (Özbey et al., 2023) or contrastive (Kim et al., 2024a) learning frameworks. Alternatively, models pre-trained for general medical image synthesis tasks can be adapted to conduct specific translation tasks via zero-shot or few-shot learning frameworks on compact training sets (Pinaya et al., 2022b, 2023; Güngör et al., 2023), which can also facilitate practical implementations. Here, we employed a recovery network based on a convolutional backbone. Recent studies on medical imaging tasks report that transformer backbones can elevate sensitivity to long-range interactions (Nezhad et al., 2025; Luo et al., 2021; Gungor et al., 2022) and enhance generalization performance to atypical anatomy (Korkmaz et al., 2022, 2023). Adoption of a backbone that allows enhanced contextual sensitivity in SelfRDB could thus improve the representation of long-range context during source-to-target mapping (Atli et al., 2024; Kabas et al., 2024). Future work could also expand comparative analyses to architectures that leverage adaptive normalization mechanisms for refined control over stylistic attributes, which may be valuable for modulating tissue contrast in medical image translation (Fetty et al., 2020; Kim et al., 2023; Liu et al., 2023b). Building on such insights, it may be worthwhile to explore incorporating similar mechanisms into the proposed architecture. Note that SelfRDB prescribes a lower number of timesteps in its diffusion process and thereby offers significantly higher efficiency than conventional DDMs. Yet, it still has longer inference times than GAN models that generate target images in a single forward pass. For efficiency improvements, acceleration approaches such as initiating sampling with an intermediate image derived from a secondary translation method (Chung et al., 2022b), or distillation of trained models onto fewer diffusion steps (Bedel and Çukur, 2024) could be considered.

7. Conclusion

In this study, we introduced a novel diffusion bridge, SelfRDB, for multi-modal medical image translation tasks. SelfRDB learns a task-relevant progressive transformation between source- and target-modality distributions. In reverse diffusion steps, it improves image fidelity via a self-consistent recursive estimation procedure and stationary guidance from the acquired source image. It further employs a monotonically-increasing scheduling for the noise variance towards the source image in order to facilitate information transfer between the modalities, and to build a soft prior on the source modality that enhances noise robustness. With these technical advances, SelfRDB achieves state-of-the-art image quality compared to leading GAN and diffusion methods, so it holds great promise for medical image translation applications.

CRediT authorship contribution statement

Fuat Arslan: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Bilal Kabas:** Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Onat Dalmaz:** Writing – original draft, Software, Investigation, Formal analysis, Data curation. **Muzaffer Ozbey:** Writing – original draft, Software, Investigation, Formal analysis, Data curation. **Tolga Çukur:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This study was supported in part by TUBA GEBIP 2015 and BAGEP 2017 fellowships awarded to T. Çukur.

References

Alexander, D.C., Zikic, D., Zhang, J., Zhang, H., Criminisi, A., 2014. Image quality transfer via random forest regression: Applications in diffusion MRI. In: Med Image Comput Comput Assist Interv. pp. 225–232.

Armanious, K., Jiang, C., Fischer, M., Küstner, T., Hepp, T., Nikolaou, K., Gatidis, S., Yang, B., 2019. MedGAN: Medical image translation using GANs. Comput. Med. Imaging Graph. 79, 101684.

Atli, O.F., Kabas, B., Arslan, F., Yurt, M., Dalmaz, O., Çukur, T., 2024. I2I-Mamba: Multi-modal medical image synthesis via selective state space modeling. arXiv:2405.14022.

Bedel, H.A., Çukur, T., 2024. DreaMR: Diffusion-driven counterfactual explanation for functional MRI. IEEE Trans. Med. Imaging http://dx.doi.org/10.1109/TMI.2024.3507008.

Bowles, C., Qin, C., Ledig, C., Guerrero, R., Gunn, R., Hammers, A., Sakka, E., Dickie, D., Hernández, M., Royle, N., Wardlaw, J., Rhodius-Meester, H., Tijms, B., Lemstra, A., Flier, W., Barkhof, F., Scheltens, P., Rueckert, D., 2016. Pseudo-healthy image synthesis for white matter lesion segmentation. In: Simul Synth Med Imaging. pp. 87–96.

Chartsias, A., Joyce, T., Giuffrida, M.V., Tsaftaris, S.A., 2018. Multimodal MR synthesis via modality-invariant latent representation. IEEE Trans. Med. Imaging 37 (3), 803–814.

Chen, Z., He, G., Zheng, K., Tan, X., Zhu, J., 2023. Schrodinger bridges beat diffusion models on text-to-speech synthesis. arXiv:2312.03491.

Chen, T., Liu, G.-H., Theodorou, E.A., 2021. Likelihood training of Schrodinger bridge using forward-backward SDEs theory. arXiv:2110.11291.

Chung, H., Kim, J., Ye, J.C., 2023. Direct diffusion bridge using data consistency for inverse problems. arXiv:2305.19809.

Chung, H., Lee, E.S., Ye, J.C., 2022a. MR image denoising and super-resolution using regularized reverse diffusion. arXiv:2203.12621.

Chung, H., Sim, B., Ye, J.C., 2022b. Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction. In: IEEE Conf Comput Vis Pattern Recognit. pp. 12413–12422.

Chung, H., Ye, J.C., 2022. Score-based diffusion models for accelerated MRI. Med. Image Anal. 80, 102479.

Clark, L.T., Watkins, L., Pina, I.L., Elmer, M., Akinboboye, O., Gorham, M., Jamerson, B., McCullough, C., Pierre, C., Polis, A.B., Puckrein, G., Regnante, J.M., 2019. Increasing diversity in clinical trials: Overcoming critical barriers. Curr. Probl. Cardiol. 44 (5), 148–172.

Cordier, N., Delingette, H., Le, M., Ayache, N., 2016. Extended modality propagation: Image synthesis of pathological cases. IEEE Trans. Med. Imaging 35, 2598–2608.

Dalmaz, O., Yurt, M., Çukur, T., 2022. ResViT: Residual vision transformers for multi-modal medical image synthesis. IEEE Trans. Med. Imaging 44 (10), 2598–2614.

Dar, S.U., Yurt, M., Karacan, L., Erdem, A., Erdem, E., 00C7ukur, T., 2019a. Image synthesis in multi-contrast MRI with conditional generative adversarial networks. IEEE Trans. Med. Imaging 38 (10), 2375–2388.

Dar, S.U., Yurt, M., Karacan, L., Erdem, A., Erdem, E., Cukur, T., 2019b. Image synthesis in multi-contrast MRI with conditional generative adversarial networks. IEEE Trans. Med. Imaging 38 (10), 2375–2388.

Daras, G., Delbracio, M., Talebi, H., Dimakis, A.G., Milanfar, P., 2022. Soft diffusion: Score matching for general corruptions. arXiv:2209.05442.

Delbracio, M., Milanfar, P., 2023. Inversion by direct iteration: An alternative to denoising diffusion for image restoration. Trans. Mach. Learn. Res..

Dhariwal, P., Nichol, A., 2021. Diffusion models beat GANs on image synthesis. In: Adv Neural Inf Process Syst. Vol. 34, pp. 8780–8794.

Dong, X., Wang, T., Lei, Y., Higgins, K., Liu, T., Curran, W., Mao, H., Nye, J., Yang, X., 2019. Synthetic CT generation from non-attenuation corrected PET images for whole-body PET imaging. Phys. Med. Biol. 64 (21), 215016.

Elmas, G., Dar, S.U., Korkmaz, Y., Ceyani, E., Susam, B., Özbey, M., Avestimehr, S., Çukur, T., 2023. Federated learning of generative image priors for MRI reconstruction. IEEE Trans. Med. Imaging 42 (7), 1996–2009.

Fetty, L., Bylund, M., Kuess, P., Heilemann, G., Nyholm, T., Georg, D., Lofstedt, T., 2020. Latent space manipulation for high-resolution medical image synthesis via the StyleGAN. Z. Med. Phys. 30 (4), 305–314.

Ge, Y., Wei, D., Xue, Z., Wang, Q., Zhou, X., Zhan, Y., Liao, S., 2019. Unpaired MR to CT synthesis with explicit structural constrained adversarial learning. In: Int Symp Biomed Imaging. pp. 1096–1099.

Gu, X., Zhang, Y., Zeng, W., Zhong, S., Wang, H., Liang, D., Li, Z., Hu, Z., 2023. Cross-modality image translation: CT image synthesis of MR brain images using multi generative network with perceptual supervision. Comput. Methods Programs Biomed. 237, 107571.

Gungor, A., Askin, B., Soydan, D.A., Saritas, E.U., Top, C.B., Çukur, T., 2022. TranSMS: Transformers for super-resolution calibration in magnetic particle imaging. IEEE Trans. Med. Imaging 41 (12), 3562–3574.

Güngör, A., Dar, S.U., Öztürk, Ş., Korkmaz, Y., Bedel, H.A., Elmas, G., Ozbey, M., Çukur, T., 2023. Adaptive diffusion priors for accelerated MRI reconstruction. Med. Image Anal. 102872.

Han, L., Zhang, T., Huang, Y., Dou, H., Wang, X., Gao, Y., Lu, C., Tan, T., Mann, R., 2023. An explainable deep framework: Towards task-specific fusion for multi-to-one MRI synthesis. Med. Image Comput. Comput. Assist. Interv. 14229, 45–55.

Ho, J., Jain, A., Abbeel, P., 2020. Denoising diffusion probabilistic models. In: Adv Neural Inf Process Syst. Vol. 33, pp. 6840–6851.

Hu, X., Shen, R., Luo, D., Tai, Y., Wang, C., Menze, B.H., 2022. AutoGAN-synthesizer: Neural architecture search for cross-modality MRI synthesis. In: Med Image Comput Comput Assist Interv. Vol. 13436, pp. 397–409.

Huang, Y., Shao, L., Frangi, A.F., 2017. Simultaneous super-resolution and cross-modality synthesis of 3D medical images using weakly-supervised joint convolutional sparse coding. Comput. Vis. Pattern Recognit. 5787–5796.

Huang, Y., Shao, L., Frangi, A.F., 2018. Cross-modality image synthesis via weakly coupled and geometry co-regularized joint dictionary learning. IEEE Trans. Med. Imaging 37 (3), 815–827.

Huynh, T., Gao, Y., Kang, J., Wang, L., Zhang, P., Lian, J., Shen, D., 2016. Estimating CT image from MRI data using structured random forest and auto-context model. IEEE Trans. Med. Imaging 35 (1), 174–183.

Iglesias, J.E., Konukoglu, E., Zikic, D., Glocker, B., Van Leemput, K., Fischl, B., 2013. Is synthesizing MRI contrast useful for inter-modality analysis? In: Med Image Comput Comput Assist Interv. pp. 631–638.

Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. Comput. Vis. Pattern Recognit. 1125–1134.

Jalal, A., Arvinte, M., Daras, G., Price, E., Dimakis, A.G., Tamir, J., 2021. Robust compressed sensing MRI with deep generative priors. In: Adv Neural Inf Process Syst. Vol. 34, pp. 14938–14954.

Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. Med. Image Anal. 5, 143–156.

Jin, C.-B., Kim, H., Liu, M., Jung, W., Joo, S., Park, E., Ahn, Y.S., Han, I.H., Lee, J.I., Cui, X., 2019. Deep CT to MR synthesis using paired and unpaired data. Sensors 19 (10), 2361.

Jog, A., Carass, A., Roy, S., Pham, D.L., Prince, J.L., 2017. Random forest regression for magnetic resonance image synthesis. Med. Image Anal. 35, 475–488.

Joyce, T., Chartsias, A., Tsaftaris, S.A., 2017. Robust multi-modal MR image synthesis. In: Med Image Comput Comput Assist Interv. pp. 347–355.

Kabas, B., Arslan, F., Nezhad, V.A., Ozturk, S., Saritas, E.U., Cukur, T., 2024. Physics-driven autoregressive state space models for medical image reconstruction. arXiv:2412.09331.

Kim, S., Chung, H., Park, S.H., Chung, E.-S., Yi, K., Ye, J.C., 2024b. Fundus image enhancement through direct diffusion bridges. IEEE J. Biomed. Heal. Inform. 1–12.

Kim, S., Jang, H., Hong, S., Hong, Y.S., Bae, W.C., Kim, S., Hwang, D., 2021. Fat-saturated image generation from multi-contrast MRIs using generative adversarial networks with Bloch equation-based autoencoder regularization. Med. Image Anal. 73 (102198).

Kim, B., Kwon, G., Kim, K., Ye, J.C., 2024a. Unpaired image-to-image translation via neural Schrödinger bridge. In: Int Conf Learn Rep.

Kim, H., Shin, Y., Hwang, D., 2023. DiMix: Disentangle-and-mix based domain generalizable medical image segmentation. In: Med Image Comput Comput Assist Interv. pp. 242–251.

Kim, J., Ye, J.C., 2024. HiCBridge: Resolution enhancement of hi-c data using direct diffusion bridge. URL: https://openreview.net/forum?id=RUvzlotXY0.

Korkmaz, Y., Cukur, T., Patel, V.M., 2023. Self-supervised MRI reconstruction with unrolled diffusion models. In: Med Image Comput Comput Assist Interv. pp. 491–501.

Korkmaz, Y., Dar, S.U.H., Yurt, M., Ozbey, M., Cukur, T., 2022. Unsupervised MRI reconstruction via zero-shot learned adversarial transformers. IEEE Trans. Med. Imaging 41 (7), 1747–1763.

Lan, H., Toga, A., Sepehrband, F., 2020. SC-GAN: 3D self-attention conditional GAN with spectral normalization for multi-modal neuroimaging synthesis. BioRxiv 2020.06.09.143297.

Lee, J., Carass, A., Jog, A., Zhao, C., Prince, J., 2017. Multi-atlas-based CT synthesis from conventional MRI with patch-based refinement for MRI-based radiotherapy planning. In: SPIE Med Imaging. Vol. 10133, p. 101331I.

Lee, D., Kim, J., Moon, W.-J., Ye, J.C., 2019. CollaGAN: Collaborative GAN for missing image data imputation. In: Comput Vis Pattern Recognit. pp. 2487–2496.

Li, H., Paetzold, J.C., Sekuboyina, A., Kofler, F., Zhang, J., Kirschke, J.S., Wiestler, B., Menze, B., 2019. DiamondGAN: Unified multi-modal generative adversarial networks for MRI sequences synthesis. In: Med Image Comput Comput Assist Interv. pp. 795–803.

Liu, S., Dowling, J.A., Engstrom, C., Greer, P.B., Crozier, S., Chandra, S.S., 2023b. Manipulating medical image translation with manifold disentanglement. In: Int Conf Dif Image Comput Tech Appl. pp. 332–339.

Liu, G.-H., Vahdat, A., Huang, D.-A., Theodorou, E.A., Nie, W., Anandkumar, A., 2023a. I^{2}SB: Image-to-image Schrödinger bridge. arXiv:2302.05872.

Luo, Y., Wang, Y., Zu, C., Zhan, B., Wu, X., Zhou, J., Shen, D., Zhou, L., 2021. 3D transformer-GAN for high-quality PET reconstruction. In: Med Image Comput Comput Assist Interv. pp. 276–285.

Lyu, Q., Wang, G., 2022. Conversion between CT and MRI images using diffusion and score-matching models. arXiv:2209.12104.

Meng, X., Gu, Y., Pan, Y., Wang, N., Xue, P., Lu, M., He, X., Zhan, Y., Shen, D., 2022. A novel unified conditional score-based generative framework for multi-modal medical image completion. arXiv:2207.03430.

Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.-A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, C., Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M.S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.-C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Van Leemput, K., 2015. The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. Med. Imaging 34 (10), 1993–2024.

Mirza, M.U., Dalmaz, O., Bedel, H.A., Elmas, G., Korkmaz, Y., Gungor, A., Dar, S.U., Çukur, T., 2023. Learning Fourier-constrained diffusion bridges for MRI reconstruction. arXiv:2308.01096.

Nelson, E., 1967. Dynamical Theories of Brownian Motion. Princeton University Press.

Nezhad, V.A., Elmas, G., Kabas, B., Arslan, F., Cukur, T., 2025. Generative autoregressive transformers for model-agnostic federated MRI reconstruction. arXiv:2502.04521.

Nichol, A.Q., Dhariwal, P., 2021. Improved denoising diffusion probabilistic models. In: Int Conf Mach Learn. pp. 8162–8171.

Nie, D., Cao, X., Gao, Y., Wang, L., Shen, D., 2016. Estimating CT image from MRI data using 3D fully convolutional networks. In: Deep Learn Data Label Med Appl. pp. 170–178.

Nie, D., Trullo, R., Lian, J., Wang, L., Petitjean, C., Ruan, S., Wang, Q., Shen, D., 2018. Medical image synthesis with deep convolutional adversarial networks. IEEE Trans. Biomed. Eng. 65 (12), 2720–2730.

Nyholm, T., Svensson, S., Andersson, S., Jonsson, J., Sohlin, M., Gustafsson, C., Kjellén, E., Söderström, K., Albertsson, P., Blomqvist, L., Zackrisson, B., Olsson, L.E., Gunnlaugsson, A., 2018. MR and CT data with multiobserver delineations of organs in the pelvic area—Part of the Gold Atlas project. Med. Phys. 45 (3), 1295–1300.

Özbey, M., Dalmaz, O., Dar, S.U., Bedel, H.A., Öztürk, Ş., Güngör, A., Çukur, T., 2023. Unsupervised medical image translation with adversarial diffusion models. IEEE Trans. Med. Imaging 42 (12), 3524–3539.

Peng, C., Guo, P., Zhou, S.K., Patel, V., Chellappa, R., 2022. Towards performant and reliable undersampled MR reconstruction via diffusion model sampling. arXiv:2203.04292.

Pinaya, W.H.L., Graham, M.S., Gray, R., Da Costa, P.F., Tudosiu, P.-D., Wright, P., Mah, Y.H., MacKinnon, A.D., Teo, J.T., Jager, R., Werring, D., Rees, G., Nachev, P., Ourselin, S., Cardoso, M.J., 2022a. Fast unsupervised brain anomaly detection and segmentation with diffusion models. In: Med Image Comput Comput Assist Interv. pp. 705–714.

Pinaya, W.H.L., Graham, M.S., Kerfoot, E., Tudosiu, P.-D., Dafflon, J., Fernandez, V., Sanchez, P., Wolleb, J., da Costa, P.F., Patel, A., Chung, H., Zhao, C., Peng, W., Liu, Z., Mei, X., Lucena, O., Ye, J.C., Tsaftaris, S.A., Dogra, P., Feng, A., Modat, M., Nachev, P., Ourselin, S., Cardoso, M.J., 2023. Generative AI for medical imaging: extending the MONAI framework. arXiv:2307.15208.

Pinaya, W.H.L., Tudosiu, P.-D., Dafflon, J., da Costa, P.F., Fernandez, V., Nachev, P., Ourselin, S., Cardoso, M.J., 2022b. Brain imaging generation with latent diffusion models. In: MICCAI Work Deep Gen Models. pp. 117–126.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: Med Image Comput Comput Assist Interv. Springer, pp. 234–241.

Roy, S., Jog, A., Carass, A., Prince, J.L., 2013. Atlas based intensity transformation of brain MR images. In: Multimodal Brain Image Anal. pp. 51–62.

Sevetlidis, V., Giuffrida, M.V., Tsaftaris, S.A., 2016. Whole image synthesis using a deep encoder-decoder network. In: Simul Synth Med Imaging. pp. 127–137.

Sharma, A., Hamarneh, G., 2020. Missing MRI pulse sequence synthesis using multi-modal generative adversarial network. IEEE Trans. Med. Imaging 39, 1170–1183.

Song, Y., Ermon, S., 2019. Generative modeling by estimating gradients of the data distribution. Adv. Neural Inf. Process. Syst. 32.

Song, Y., Shen, L., Xing, L., Ermon, S., 2021. Solving inverse problems in medical imaging with score-based generative models. arXiv:2111.08005.

Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B., 2020. Score-based generative modeling through stochastic differential equations. arXiv:2011.13456.

Su, X., Song, J., Meng, C., Ermon, S., 2023. Dual diffusion implicit bridges for image-to-image translation. arXiv:2203.08382.

Van Nguyen, H., Zhou, K., Vemulapalli, R., 2015. Cross-domain synthesis of medical images using efficient location-sensitive deep network. In: Med Image Comput Comput Assist Interv. pp. 677–684.

Vemulapalli, R., Van Nguyen, H., Zhou, S.K., 2015. Unsupervised cross-modal synthesis of subject-specific scans. In: Int Conf Comput Vis. pp. 630–638.

Wang, G., Gong, E., Banerjee, S., Martin, D., Tong, E., Choi, J., Chen, H., Wintermark, M., Pauly, J.M., Zaharchuk, G., 2020. Synthesize high-quality multi-contrast magnetic resonance imaging from multi-echo acquisition using multi-task deep generative model. IEEE Trans. Med. Imaging 39 (10), 3089–3099.

Wang, Z., Yang, Y., Chen, Y., Yuan, T., Sermesant, M., Delingette, H., Wu, O., 2024. Mutual information guided diffusion for zero-shot cross-modality medical image translation. IEEE Trans. Med. Imaging 1.

Wei, W., Poirion, E., Bodini, B., Durrleman, S., Colliot, O., Stankoff, B., Ayache, N., 2019. Fluid-attenuated inversion recovery MRI synthesis from multisequence MRI using three-dimensional fully convolutional networks for multiple sclerosis. J. Med. Imaging 6 (1), 014005.

Wolleb, J., Bieder, F., Sandkühler, R., Cattin, P.C., 2022. Diffusion models for medical anomaly detection. In: Med Image Comput Comput Assist Interv. Vol. 13438, pp. 35–45.

Wolterink, J.M., Dinkla, A.M., Savenije, M.H.F., Seevinck, P.R., van den Berg, C.A.T., Išgum, I., 2017. Deep MR to CT synthesis using unpaired data. In: Simul Synth Med Imaging. Cham, pp. 14–23.

Wu, Y., Yang, W., Lu, L., Lu, Z., Zhong, L., Huang, M., Feng, Y., Feng, Q., Chen, W., 2016. Prediction of CT substitutes from MR images based on local diffeomorphic mapping for brain PET attenuation correction. J. Nucl. Med. 57 (10), 1635–1641.

Xia, Y., Ravikumar, N., Lassila, T., Frangi, A.F., 2023. Virtual high-resolution MR angiography from non-angiographic multi-contrast MRIs: synthetic vascular model populations for in-silico trials. Med. Image Anal. 87 (102814).

Xiao, Z., Kreis, K., Vahdat, A., 2022. Tackling the generative learning trilemma with denoising diffusion GANs. In: Int Conf Learn Represent.

Xin, B., Young, T., Wainwright, C., Blake, T., Lebrat, L., Gaass, T., Benkert, T., Stemmer, A., Coman, D., Dowling, J., 2024. Deformation-aware GAN for medical image synthesis with substantially misaligned pairs. In: Med Imaging Deep Learn.

Yang, H., Sun, J., Carass, A., Zhao, C., Lee, J., Xu, Z., Prince, J., 2018. Unpaired brain MR-to-CT synthesis using a structure-constrained cycleGAN. arXiv:1809.04536.

Ye, D.H., Zikic, D., Glocker, B., Criminisi, A., Konukoglu, E., 2013. Modality propagation: Coherent synthesis of subject-specific scans with data-driven regularization. In: Med Image Comput Comput Assist Interv. pp. 606–613.

Yu, B., Zhou, L., Wang, L., Fripp, J., Bourgeat, P., 2018. 3D cGAN based cross-modality MR image synthesis for brain tumor segmentation. In: Int Symp Biomed Imaging. pp. 626–630.

Yu, B., Zhou, L., Wang, L., Shi, Y., Fripp, J., Bourgeat, P., 2019. Ea-GANs: Edge-aware generative adversarial networks for cross-modality MR image synthesis. IEEE Trans. Med. Imaging 38 (7), 1750–1762.

Yurt, M., Dalmaz, O., Dar, S.U.H., Özbey, M., Tınaz, B., Oğuz, K.K., Çukur, T., 2022. Semi-supervised learning of MRI synthesis without fully-sampled ground truths. IEEE Trans. Med. Imaging 41 (12), 3895–3906.

Yurt, M., Dar, S.U., Erdem, A., Erdem, E., Oguz, K.K., Çukur, T., 2021. MustGAN: multi-stream generative adversarial networks for MR image synthesis. Med. Image Anal. 70, 101944.

Zhang, H., Goodfellow, I., Metaxas, D., Odena, A., 2019. Self-attention generative adversarial networks. In: Int Conf Mach Learn. Vol. 97, pp. 7354–7363.

Zhang, Y., Peng, C., Wang, Q., Song, D., Li, K., Kevin Zhou, S., 2025. Unified multi-modal image synthesis for missing modality imputation. IEEE Trans. Med. Imaging 44 (1), 4–18.

Zhao, C., Carass, A., Lee, J., He, Y., Prince, J.L., 2017. Whole brain segmentation and labeling from CT using synthetic MR images. In: Mach Learn Med Imaging. pp. 291–298.

Zhou, T., Fu, H., Chen, G., Shen, J., Shao, L., 2020. Hi-net: Hybrid-fusion network for multi-modal MR image synthesis. IEEE Trans. Med. Imaging 39 (9), 2772–2781.