# DUAL DIFFUSION IMPLICIT BRIDGES FOR IMAGE-TO-IMAGE TRANSLATION

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Common image-to-image translation methods rely on joint training over data from both source and target domains. This prevents the training process from preserving privacy of domain data (*e.g.*, in a federated setting), and often means that a new model has to be trained for a new pair of domains. We present Dual Diffusion Implicit Bridges (DDIBs), an image translation method based on diffusion models, that circumvents training on domain pairs. Image translation with DDIBs relies on two diffusion models trained independently on each domain, and is a two-step process: DDIBs first obtain latent encodings for source images with the source diffusion model, and then decode such encodings using the target model to construct target images. Both steps are defined via an ODE, thus the process is cycle consistent only up to discretization errors of the ODE solvers. Theoretically, we interpret DDIBs as concatenation of source to latent, and latent to target Schrödinger Bridges, a form of entropy-regularized optimal transport, to explain the efficacy of the method. Experimentally, we apply DDIBs on both synthetic and high-resolution image datasets, to demonstrate their utility in a wide variety of translation tasks and their connections to existing optimal transport methods.

## 1 INTRODUCTION

Transferring images from one domain to another while preserving the content is an important problem in computer vision, with wide applications that span image in-painting (Zhao et al., 2020a), colorization (Xiao et al., 2019) and style transfer (Xu et al., 2021). In tasks such as style transfer, it is usually difficult to obtain paired images of realistic scenes and their corresponding artistic renditions, making unpaired translation methods particularly relevant. Common methods on unpaired translation are based on generative adversarial networks (GANs, Goodfellow et al. (2014); Zhu et al. (2017)) or normalizing flows (Grover et al., 2020). Training such models typically involves minimizing an adversarial loss between a specific pair of source and target datasets.

While capable of producing high-quality images, these methods suffer from a severe drawback in their *adaptability* to alternative domains. Concretely, training a model on a source-target pair requires specific tuning on this pair of datasets. Given a different pair, existing models cannot be easily adapted for translation. If we were to do pairwise translation among a set of domains, the total number of models needed is quadratic in the number of domains – an unacceptable computational cost in practice. One alternative is to find a shared domain that connects to each source / target domains as in StarGANs (Choi et al., 2018). However, the shared domain needs to be carefully chosen *a priori*; if the shared domain contains less information than the target domain (*e.g.* sketches v.s. photos), then it creates an unwanted information bottleneck between the source and target domains.

An additional disadvantage of existing models resides in their *lack of privacy protection* of the datasets: training a single model requires access to both datasets simultaneously. Such setting may be inconvenient or impossible, when data providers are reluctant about giving away their data; or for certain privacy-sensitive applications such as medical imaging. For example, quotidian hospital usage may require translation of patients' X-ray and MRI images taken from machines in other hospitals. Most existing translation methods will fail in such scenarios, as joint training across hospitals is difficult, if not impossible.

In this paper, we seek to mitigate both problems of existing image translation methods. We present Dual Diffusion Implicit Bridges (DDIBs), an image-to-image translation method inspired by recent

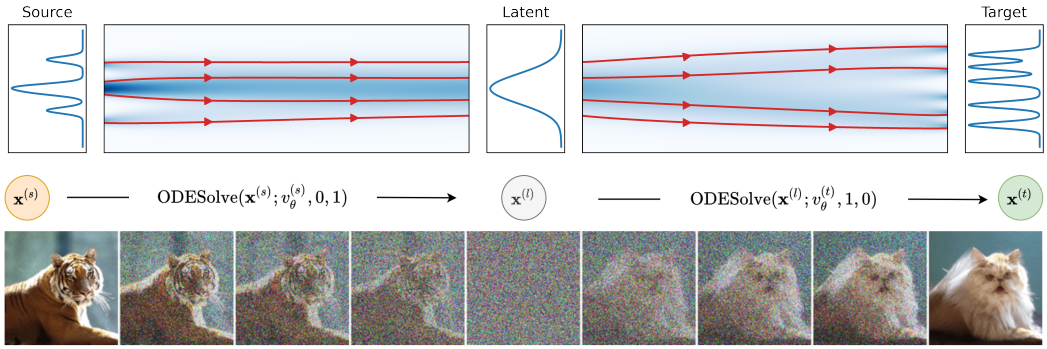



Figure 1: **Dual Diffusion Implicit Bridges**: DDIBs leverage two ODEs for image translation. Given a source image $\mathbf{x}^{(s)}$, the source ODE runs in the forward direction to convert it to the latent $\mathbf{x}^{(l)}$, while the target, reverse ODE then constructs the target image $\mathbf{x}^{(t)}$. (*Top*) Illustration of the DDIBs idea between two one-dimensional distributions. (*Bottom*) DDIBs from a tiger to a cat using a pretrained conditional diffusion model.

advances in diffusion models (Song et al., 2020a;b), that decouples paired training, and empowers the trained model for a particular domain to stay applicable in other pairs wherever the domain reappears. Since the training process now concentrates on one dataset at a time, DDIBs can also be applied in federated settings, and owners of domain data can effectively preserve their data privacy.

Specifically, DDIBs are developed based on the method known as denoising diffusion implicit models (DDIMs, Song et al. (2020a)). DDIMs invent a particular parameterization of the diffusion process, that creates a smooth, deterministic and reversible mapping between images and their latent representations. This mapping is captured using the solution to a so-called *probability flow* (PF) ordinary differential equation (ODE) that forms the cornerstone of DDIBs. Translation with DDIBs on a source-target pair requires two different PF ODEs: the source PF ODE converts input images to a latent space; while the target ODE then synthesizes images in the target domain.

Importantly, trained diffusion models are specific to the individual domains, and rely on no domain pairing information. Effectively, DDIBs make it possible to save a trained model of a certain domain for future use, when it arises as the source or target in a new pair. Pairwise translation with DDIBs requires only a *linear* number of diffusion models (which can be further reduced with conditional models (Dhariwal & Nichol, 2021)), and training does not require scanning both datasets concurrently. Unlike StarGANs that leverage a shared domain, in DDIBs, the latent space stores all information about the data.

Theoretically, we analyze the translation process in DDIBs. We borrow the theory from (Chen et al., 2021a) to show that, in essence, the probability flow ODEs in DDIBs comprise the solution of a special Schrödinger Bridge Problem (SBP) with linear or degenerate drift, between the data and the latent distributions. This justification of DDIBs from an optimal transport viewpoint that alternative translation methods lack serves as a strong theoretical advantage of our method. In addition, we show that DDIBs guarantee exact cycle consistency only up to discretization errors introduced in the ODE solvers.

Experimentally, we first present synthetic experiments on two-dimensional datasets to demonstrate DDIBs' cycle-consistency property. We then evaluate our method on a variety of image modalities, with qualitative and quantitative results: we validate its usage in example-guided color transfer, paired and unpaired translation, and conditional ImageNet translation. These results establish DDIBs as a scalable, theoretically rigorous addition to the family of unpaired image translation methods.

## 2 Preliminaries

### 2.1 Score-based Generative Models (SGMs)

While our actual implementation utilizes DDIMs, we first briefly introduce the broader family of models known as score-based generative models. Two representative models of this family are *score matching with Langevin dynamics* (SMLD) (Song & Ermon, 2019) and *denoising diffusion probabilistic models* (DDPMs) (Ho et al., 2020). Both methods are contained within the framework of Stochastic Differential Equations (SDEs) proposed in Song et al. (2020b).

**Stochastic Differential Equation (SDE) Representation** Song et al. (2020b); Anderson (1982) use a forward and a corresponding backward SDE to describe general diffusion and the reversed, generative processes:

$$\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\,\mathrm{d}t + g(t)\,\mathrm{d}\mathbf{w}\,, \quad \mathrm{d}\mathbf{x} = [\mathbf{f} - g^2 \nabla_{\mathbf{x}} \log p_t(\mathbf{x})]\,\mathrm{d}t + g(t)\,\mathrm{d}\mathbf{w} \tag{1}$$

where $\mathbf{w}$ is the standard Wiener process, $\mathbf{f}(\mathbf{x}, t)$ is the vector-valued drift coefficient, $g(t)$ is the scalar diffusion coefficient, and $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ is the score function of the noise perturbed data distribution (as defined by the forward SDE with initial condition $p_0(\mathbf{x})$ being the data distribution). At the endpoints $t = \{0, 1\}$, the forward Eq. (1) admits the data distribution $p_0$ and the easy-to-sample prior $p_1$ as the boundary distributions. Within this framework, the SMLD method can be described using a *Variance-Exploding* (VE) SDE with increasing noise scales $\sigma(t)$: $\mathrm{d}\mathbf{x} = \sqrt{\mathrm{d}[\sigma^2(t)]\,/\,\mathrm{d}t}\,\mathrm{d}\mathbf{w}$. In comparison, DDPMs are endowed with a *Variance-Preserving* (VP) SDE: $\mathrm{d}\mathbf{x} = -[\beta(t)/2]\mathbf{x}\,\mathrm{d}t + \sqrt{\beta(t)}\,\mathrm{d}\mathbf{w}$ with $\beta(t)$ being another noise sequence. Notably, the VP SDE can be reparameterized into an equivalent VE SDE (Song et al., 2020a).

**Probability Flow ODE** Any diffusion process can be represented by a *deterministic* ODE that carries the same marginal densities as the diffusion process throughout its evolution. This ODE is termed the *probability flow* (PF) ODE (Song et al., 2020b). PF ODEs enable *uniquely identifiable encodings* (Song et al., 2020b) of data, and are central to DDIBs as we solve these ODEs for forward and reverse conversion between data and their latents. For the forward SDE introduced in Eq. (1), the equivalent PF ODE holds the following form:

$$\mathrm{d}\mathbf{x} = \left[\mathbf{f}(\mathbf{x}, t) - \frac{1}{2} g(t)^2 \nabla_{\mathbf{x}} \log p_t(\mathbf{x})\right] \mathrm{d}t \tag{2}$$

which follows immediately from the SDEs given the score function. In practice, we use $\theta$-parameterized score networks $\mathbf{s}_{t,\theta} \approx \nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ to approximate the score function. We may then use numerical ODE solvers to construct $\mathbf{x}$ at different times. Empirically, it has been shown that SGMs have relatively low discretization errors when reconstructing $\mathbf{x}$ at $t = 0$ via ODE solvers (Song et al., 2020a). For conciseness, we use $v_\theta = \mathrm{d}\mathbf{x}\,/\,\mathrm{d}t$ to denote the $\theta$-parameterized velocity field (as defined from Eq. (2), where we replace $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ with $\mathbf{s}_{t,\theta}$), and use the symbol $\mathrm{ODESolve}$ to denote the mapping from $\mathbf{x}(t_0)$ to $\mathbf{x}(t_1)$:

$$\mathrm{ODESolve}(\mathbf{x}(t_0); v_\theta, t_0, t_1) = \mathbf{x}(t_0) + \int_{t_0}^{t_1} v_\theta(t, \mathbf{x}(t))\,\mathrm{d}t\,, \tag{3}$$

which allows us to abstract away the exact model (be it a score-based or a diffusion model), or the integrator used. In our experiments, we implement the ODE solver in DDIMs (Song et al., 2020a) (Appendix B); while we acknowledge other available ODE solvers that are usable within our framework, such as the DPM-solver (Lu et al., 2022), the Exponential Integrator (Zhang & Chen, 2022), and the second-order Heun solver (Karras et al., 2022).

### 2.2 Schrödinger Bridge Problem

Our analysis will show that DDIBs are Schrödinger Bridges (Chen et al., 2016; Léonard, 2013) between distributions. Let $\mathcal{D}(p_0, p_1)$ denote the set of distributions over the path space , with marginals $p_0$, $p_1$ at time $t = 0$, $t = 1$, respectively. Given a prior reference measure $W$, the well-known *Schrödinger Bridge Problem* (SBP) seeks the most likely evolution across time $t$ between the marginals $p_0$ and $p_1$:

**Problem 1** (Schrödinger Bridge Problem)**.** *With prescribed distributions $p_0, p_1$ and a reference measure $W$ as the prior, the SBP finds a distribution from $\mathcal{D}(p_0, p_1)$ that minimizes its KL-divergence to $W$: $P_{SBP} \coloneqq \arg\min\{D_{KL}(P\|W) \mid P \in \mathcal{D}(p_0, p_1)\}$.*

The minimizer, $P_{\text{SBP}}$, is referred to as the *Schrödinger Bridge* between $p_0$ and $p_1$ over prior $W$. The SBP has close connections to the Monge-Kantorovich (MK) optimal transport problem (Chen et al., 2021b). While the basic MK problem seeks the cost-minimizing plan to transport masses between two distributions, the SBP incorporates an additional entropy term (for details, see Page 61 of Peyré et al. (2019)) .

**Algorithm 1** High-level Pseudo-code for DDIBs

**Input:** data sample from source domain $\mathbf{x}^{(s)} \sim p_s(\mathbf{x})$, source model $v_\theta^{(s)}$, target model $v_\theta^{(t)}$.
**Output:** $\mathbf{x}^{(t)}$, the result in the target domain.
$\mathbf{x}^{(l)} = \mathrm{ODESolve}(\mathbf{x}^{(s)}; v_\theta^{(s)}, 0, 1)$ // obtain latent code from source domain data
$\mathbf{x}^{(t)} = \mathrm{ODESolve}(\mathbf{x}^{(l)}; v_\theta^{(t)}, 1, 0)$ // obtain target domain data from latent code
**return** $\mathbf{x}^{(t)}$

**Relationship Between SBPs and SGMs** Chen et al. (2021a) establishes connections between SGMs and SBPs. In summary, SGMs are implicit optimal transport models, corresponding to SBPs with linear or degenerate drifts. General SBPs additionally accept fully nonlinear diffusion. To formalize this observation, the authors first establish similar forward and backward SDEs for SBPs:

$$\mathrm{d}\mathbf{x} = [\mathbf{f} + g^2 \nabla_{\mathbf{x}} \log \Phi_t(\mathbf{x})]\,\mathrm{d}t + g(t)\,\mathrm{d}\mathbf{w}\,, \quad \mathrm{d}\mathbf{x} = [\mathbf{f} - g^2 \nabla_{\mathbf{x}} \log \hat{\Phi}_t(\mathbf{x})]\,\mathrm{d}t + g(t)\,\mathrm{d}\mathbf{w} \tag{4}$$

where $\Phi, \hat{\Phi}$ are the *Schrödinger factors* that satisfy density factorization: $p_t(\mathbf{x}) = \Phi_t(\mathbf{x})\hat{\Phi}_t(\mathbf{x})$. The vector-valued quantities $\mathbf{z}_t = g(t)\nabla_{\mathbf{x}} \log \Phi_t(\mathbf{x}), \hat{\mathbf{z}}_t = g(t)\nabla_{\mathbf{x}} \log \hat{\Phi}_t(\mathbf{x})$ fully characterize dynamics of the SBP, thus can be considered as the forward, backward "policies", analogous to policy-based methods described in Schulman et al. (2015); Pereira et al. (2019). To draw a link between SBPs and SGMs, the data log-likelihood objective for SBPs is computed and shown to be equal to that of SGMs with special choices of $\mathbf{z}_t, \hat{\mathbf{z}}_t$ (details in Chen et al. (2021a)). Importantly, likelihood equality occurs with the following policies:

$$(\mathbf{z}_t, \hat{\mathbf{z}}_t) = (0, g(t)\,\nabla_{\mathbf{x}} \log p_t(\mathbf{x})) \tag{5}$$

Such $(\mathbf{z}_t, \hat{\mathbf{z}}_t)$ are achievable when the end marginal is equal to the tractable Gaussian prior that we sample from; this is exactly the case in SGMs. The log-likelihood equivalence has profound implications, suggesting that SGMs are a special case of SBPs with degenerate forward policy $\mathbf{z}_t$ and a multiple of the score function as its backward $\hat{\mathbf{z}}_t$.

**Probability Flow ODE** In a similar vein to the SGM SDEs, a corresponding deterministic PF ODE can be derived for SBPs with identical marginal densities across $t \in [0, 1]$. The following PF ODE, which we shall show to be equivalent to that of SGMs, characterizes the probability flow of the optimal processes of the SBP defined in Eq. (4) (Chen et al., 2021a):

$$\mathrm{d}\mathbf{x} = \left[\mathbf{f}(\mathbf{x}, t) + g(t)\,\mathbf{z} - \frac{1}{2} g(t)(\mathbf{z} + \hat{\mathbf{z}})\right] \mathrm{d}t\,. \tag{6}$$

where $\mathbf{z}$ depends on $\mathbf{x}$.

## 3 Dual Diffusion Implicit Bridges

DDIBs leverage the connections between SGMs and SBPs to perform image-to-image translation, with two diffusion models trained separately on the two domains. DDIBs contain two steps, described in Alg. 1 and illustrated in Fig. 1. At the core of the algorithm is the ODE solver $\mathrm{ODESolve}$ from Eq. (3). Given a source model represented as a vector field, *i.e.*, $v_\theta^{(s)}$ defined from Eq. (2), DDIBs first apply $\mathrm{ODESolve}$ in the source domain to obtain the encoding $\mathbf{x}^{(s)}$ of the image at the end time $t = 1$; we refer to this as the *latent code* (associated with the diffusion model for the domain). Then, the source latent code is fed as the initial condition (target latent code at $t = 1$) to $\mathrm{ODESolve}$ with the target model $v_\theta^{(t)}$ to obtain the target image $\mathbf{x}^{(t)}$. As discussed earlier, we implement $\mathrm{ODESolve}$ with DDIMs (Song et al., 2020a), which are known to have reasonably small discretization errors. While recent developments in higher order ODE solvers (Zhang & Chen, 2022; Lu et al., 2022; Karras et al., 2022) that generalize DDIMs can also be used here, we leave this investigation in future work.

Despite the simplicity of the method, DDIBs have several advantages over prior methods, which we discuss below.

**Exact Cycle Consistency** A desirable property of image translation algorithms is the *cycle consistency* property: transforming a data point from the source domain to the target domain, and then back to source, will recover the original data point in the source domain. The following proposition validates the cycle consistency of DDIBs.

**Proposition 3.1** (DDIBs Enforce Exact Cycle Consistency)**.** *Given a sample from source domain* $\mathbf{x}^{(s)}$*, a trained source model* $v_\theta^{(s)}$*, and a target model* $v_\theta^{(t)}$*, define:*

$$\mathbf{x}^{(l)} = \text{ODESolve}(\mathbf{x}^{(s)}; v_\theta^{(s)}, 0, 1); \quad \mathbf{x}^{(t)} = \text{ODESolve}(\mathbf{x}^{(l)}; v_\theta^{(t)}, 1, 0), \tag{7}$$

$$\mathbf{x}'^{(l)} = \text{ODESolve}(\mathbf{x}^{(t)}; v_\theta^{(t)}, 0, 1); \quad \mathbf{x}'^{(s)} = \text{ODESolve}(\mathbf{x}'^{(l)}; v_\theta^{(s)}, 1, 0) \tag{8}$$

*Assume zero discretization error. Then,* $\mathbf{x}^{(s)} = \mathbf{x}'^{(s)}$*.*

Since PF ODEs are used, the cycle consistency property is guaranteed. We show empirically in Section 4.1 that, even with discretization error, DDIBs would incur very small cycle inconsistency in practice. In contrast, methods based on GANs would have to introduce specific training terms and minimize a cycle consistency loss, which usually forces the use of paired domain data.

**Data Privacy in Both Domains** In the DDIBs translation process, only the source and target diffusion models are required, whose training processes do not depend on knowledge of the domain pair *a priori*. In fact, this process can even be performed in a privacy sensitive manner (graphic illustration in Appendix A). Let Alice and Bob be the data owners of the source and target domains, respectively. Suppose Alice intends to translate images to the target domain. However, Alice does not want to share the data with Bob (and vice versa, Bob does not want to reveal their data either). Then, Alice can simply train a diffusion model with the source data, encode the data to the latent space, transmit the latent codes to Bob, and then ask Bob to run their trained diffusion model and send the results back. In this procedure, only the latent code and the target results are transmitted between the two data vendors, and both parties have naturally ensured that their data are not directly revealed.

**DDIBs are Two Concatenated Schrödinger Bridges** DDIBs link the source data distribution to the latent space, and then to the target distribution. What is the nature of such connections between distributions? We offer an answer from an optimal transport perspective: these connections are special *Schrödinger Bridges* between distributions. This, in turn, explicates the name of our method: dual diffusion implicit bridges are based on denoising *diffusion implicit* models (Song et al., 2020a), and contain *two* separate Schrödinger *Bridges* that connect the data and latent distributions. Specifically, as considered earlier, when conditions about the policies $\mathbf{z}_t, \hat{\mathbf{z}}_t$ in Eq. (5) and the density $p_1(\mathbf{x})$ being a Gaussian prior are met, the data likelihoods (at $t = 0$) for SGMs and SBPs are identical. Indeed, these conditions are fulfilled in SGMs and particularly in DDIMs. This verifies SGMs as special linear or degenerate SBPs. Forward and reverse solving the PF ODE for SGMs, as done in DDIBs, is equivalent to flowing through the optimal processes of particular SBPs, as shown in the following proposition:

**Proposition 3.2** (PF ODE Equivalence[1])**.** *Eq.* (2) *is equivalent to Eq.* (6) *with forward, backward policies* $(\mathbf{z}_t, \hat{\mathbf{z}}_t) = (0, g\nabla_{\mathbf{x}} \log p_t(\mathbf{x}))$ *as attained in SGMs and particularly in DDIMs.*

Thus, DDIBs have deep connections to entropy-regularized optimal transport: they are Schrödinger Bridges between distributions. The translation process can then be understood as traversing through two concatenated Schrödinger Bridges, one forward and one reversed. The mapping is unique and minimizes a notion of OT distance, which probably elucidates the superior performance of DDIBs. In contrast, if we train the source and target models separately with normalizing flow models that do not have such a connection, there are many viable, invertible mappings, and the resulting image translation algorithm may not necessarily have good performance. This is probably the reason why AlignFlow (Grover et al., 2020) still have to incorporate an adversarial loss even when cycle-consistency is guaranteed.

## 4 Experiments

We present a series of experiments to demonstrate the effectiveness of DDIBs. First, we describe synthetic experiments on two-dimensional datasets, to demonstrate DDIBs' cycle-consistent and

[1] Proof in Appendix D.

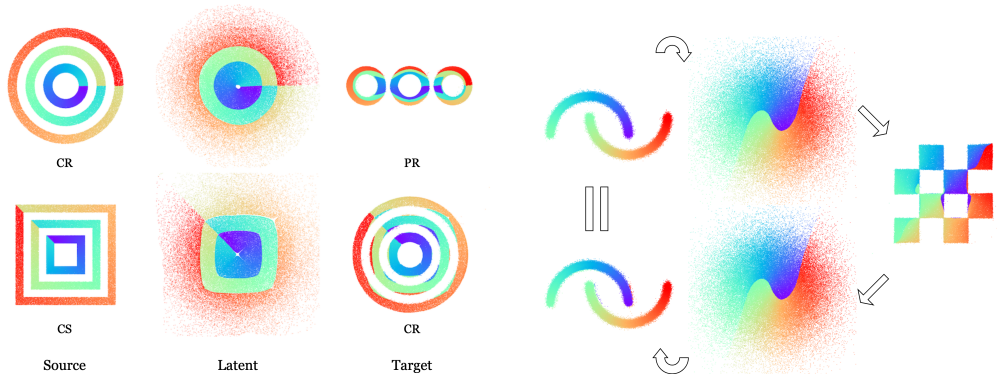
Table 1: Cycle consistency of DDIBs. Experiment legend, PR $\circlearrowleft$ PS, means that we translate from PR to PS and then back. The numbers are the averaged L2 distances between the original points and their coordinates after cycle translation. Data points are standardized to have unit variance.

| PR $\circlearrowleft$ PS | PS $\circlearrowleft$ CS | CR $\circlearrowleft$ PR | CR $\circlearrowleft$ CS | M $\circlearrowleft$ CB |
|---|---|---|---|---|
| 0.0143 | 0.0065 | 0.0106 | 0.0078 | 0.0122 |

optimal transport properties. Next, we validate DDIBs on a variety of image translation tasks, including color transfer, paired and unpaired translation, and conditional ImageNet translation[2].

### 4.1 2D Synthetic Experiments

We first perform domain translation on synthetic datasets drawn from complex two-dimensional distributions, with various shapes and configurations, in Fig. 2a. In total, we consider six 2D datasets: Moons (M); Checkerboards (CB); Concentric Rings (CR); Concentric Squares (CS); Parallel Rings (PR); and Parallel Squares (PS). The datasets are all normalized to have zero mean, and identity covariance. We assign colors to points based on the point identities (*i.e.*, if a point in the source domain is red, its corresponding point in the target domain is also colored red). Clearly, the transformation is *smooth* between columns. For example, on the top-right corner, red points in the CR dataset are mapped to similar coordinates, both in the latent and in the target domain.




(a) Smooth translation of synthetic datasets. (*Left*) The source datasets: CR and CS. (*Middle*) DDIBs' latent representation. (*Right*) Results of translation to the target domains.

(b) **Cycle consistency**: After translating the Moons dataset to Checkerboards and then back to Moons, DDIBs restore almost the exact same points as the original ones.

Figure 2: Smoothness and cycle consistency of DDIBs.

**Cycle Consistency** Fig. 2b illustrates the cycle consistency property guaranteed by DDIBs. It concerns 2D datasets: Moons, and Checkerboards. Starting from the Moons dataset, DDIBs first obtain the latents and construct the Checkerboards points. Next, DDIBs do translations in the reverse direction, transforming the points back to the latent and the Moons space. After this round trip, points are approximately mapped to their original positions. A similar, smooth color topology is observed in this experiment. Table 1 reports quantitative evaluation results on cycle-consistent translation among multiple datasets. As the datasets are normalized to unit standard deviation, the reported values are negligibly small and validate the cycle consistent property of DDIBs.

### 4.2 Example-Guided Color Transfer

DDIBs can be used on an interesting application: example-guided color transfer. This refers to the task of modifying the colors of an input image, conditioned on the color palette of a reference image. To use DDIBs for color transfer, we train one diffusion model per image, on its normalized RGB space. During translation, DDIBs obtain encodings of the original colors, and apply the diffusion model of the reference image to attain the desired color palette. Fig. 3 visualizes our color experiments.

[2] Code will be released upon publication of this paper

Table 2: Mean Squared Error (MSE) comparing color transfer results of DDIBs with common OT methods on two images. Each number represents the MSE between DDIBs and the corresponding OT method. MSE is computed pixel-wise after normalizing images to $[-1, 1]$.

| IMAGE | EMD | SINKHORN | LINEAR | GAUSSIAN |
|---|---|---|---|---|
| TARGET 1 | 0.0337 | 0.0281 | 0.0352 | 0.0370 |
| TARGET 2 | 0.0293 | 0.0326 | 0.0500 | 0.0751 |

Table 3: MSE comparing DDIBs and baselines on paired test sets. MSE is computed pixel-wise after normalizing images to $[-1, 1]$.

| DATASET | MODEL | A → B | B → A | DATASET | MODEL | A → B | B → A |
|---|---|---|---|---|---|---|---|
| FACADES | CYCLEGAN | 0.7129 | 0.3286 | MAPS | CYCLEGAN | 0.0245 | 0.0953 |
| | ALIGNFLOW | 0.5801 | **0.2512** | | ALIGNFLOW | 0.0209 | **0.0897** |
| | DDIBS | **0.5312** | 0.3946 | | DDIBS | **0.0194** | 0.1302 |

**Comparison to Alternative OT Methods** As DDIBs are related to regularized OT, we compare the pixel-wise MSEs between color-transferred images generated by DDIBs, and images produced by alternate methods, in Table 2. We compare DDIBs against four OT methods: Earth Mover's Distance; Sinkhorn distance (Cuturi, 2013); linear and Gaussian mapping estimation (Perrot et al., 2016). Results of DDIBs are very close to those of OT methods. Appendix E.2 includes full color translation results.

Figure 3: **Example-Guided Color Transfer**: Given the first image as the reference image, DDIBs modify the colors of two input images to similarly follow a snowy winter color palette.

### 4.3 QUANTITATIVE TRANSLATION EVALUATION

Quantitatively, we demonstrate that DDIBs achieve competitive results on paired domain tests. Such numerical evaluation is despite that DDIBs are formulated with a weaker setting: diffusion models are trained independently, on separate datasets. In comparison, methods such as CycleGAN and AlignFlow assume access to both datasets during training and jointly optimize for the translation loss.

**Paired Domain Translation** As in similar works, we evaluate DDIBs on benchmark paired datasets (Zhu et al., 2017): Facades and Maps. Both are image segmentation tasks. In the pairs of datasets, one dataset contains real photos taken via a camera or a satellite; while the other comprises the corresponding segmentation images. These datasets provide one-to-one image alignment, which allows quantitative evaluation via a distance metric such as mean-squared error (MSE) between generated samples and the corresponding ground truth. To facilitate the workings of DDIBs, we additionally employ a color conversion heuristic motivated by optimal transport on image colors (Appendix E.1). Table 3 reports the evaluation results. Surprisingly, DDIBs are able to produce segmentation images that surpass alternative methods in MSE terms; while reverse translations also achieve decent performance.

### 4.4 CLASS-CONDITIONAL IMAGENET TRANSLATION

In this experiment, we apply DDIBs to translation among ImageNet classes. To this end, we leverage the pretrained diffusion models from Dhariwal & Nichol (2021). The authors optimized performance of diffusion models, and end up with a "UNet" (Ho et al., 2020) architecture with particular width, attention and residual configurations. The models are learned on 1000 ImageNet classes, each with around 1000 training images, and at a variety of resolutions. Our experiments use the model with

resolution $256 \times 256$. Moreover, these models incorporate a technique known as *classifier guidance* (Dhariwal & Nichol, 2021), that leverage classifier gradients to steer the sampling process towards arbitrary class labels during image generation. The learned models combined with classifier guidance can be effectively considered as 1000 different models. Fig. 4a exhibits select translation samples, where the source images are from ImageNet validation sets. DDIBs are able to create faithful target images that maintain much of the original content such as animal poses, complexions and emotions, while accounting for differences in animal species.

**Multi-Domain Translation** Given conditional models on the individual domains, DDIBs can be applied to translate between arbitrary pairs of source-target domains, while requiring no additional fine-tuning or adaptation. Fig. 4b displays results of translating a common image of a roaring lion (with class label 291), to various other ImageNet classes. Interestingly, some animals roar, while others stick their tongues out. DDIBs successfully internalize characteristics of distinct animal species, and produce closest animal postures in OT distances to the original shouting lion.

(a) **Conditional ImageNet Translation**: Selected translation samples from various ImageNet classes such as 282: Tiger Cat, 283: Persian Cat, 294: Brown Bear and 295: Black Bear.

(b) **Multi-domain translation**: Given the center, source image from class label 291, DDIBs translate it to other animal species, entirely using only a pre-trained conditional diffusion model.

Figure 4: Translation among ImageNet classes.

## 5 Related Works

**Diffusion Models** With roots in thermodynamics (Sohl-Dickstein et al., 2015), diffusion models reverse the dynamics of a noising process to create data samples. These models excel in tasks ranging from stroke-based image editing and composition (Meng et al., 2021), image density estimation (Kingma et al., 2021), to linear-inverse based image restoration (Kawar et al., 2022). Seminal diffusion models are *denoising diffusion probabilistic models* (DDPMs, Ho et al. (2020)), that proposed Gaussian parameterizations in the ELBO objective; while ILVR (Choi et al., 2021) invents

a conditional method to guide DDPM generation towards reference images. *Denoising diffusion implicit models* (DDIMs, Song et al. (2020a)) improve the inference speed of DDPMs via a non-Markovian sampling process. DDIMs invent an ODE to describe its generative process, that sits at the core of DDIBs.

While GANs (Goodfellow et al., 2014) have been dominating deep generative leaderboards, a recent publication (Dhariwal & Nichol, 2021) optimizes the architecture of diffusion models to outperform GANs in image synthesis; whereas Nichol & Dhariwal (2021) improves diffusion models to win log-likelihood competitions. In our experiments, we leverage their industry-scale models to produce high-quality translation results.

**Diffusion Models for Unpaired Image Translation** While the widely-received GANs (Goodfellow et al., 2014; Zhu et al., 2017; Almahairi et al., 2018; Zhao et al., 2020b) achieve remarkable performance in unpaired image translation, an array of recent publications instead employ diffusion models. For instance, DiffusionCLIP (Kim & Ye, 2021) performs image manipulation by incorporating text guidance into DDPMs; and UNIT-DDPM (Sasaki et al., 2021) proposes a novel coupling between the domain pairs, to train joint DDPMs for translation. Unlike their joint training, DDIBs directly apply pretrained diffusion models and leverage the geometry of the shared space for image translation.

**Schrödinger Bridge for Generative Modeling** Solutions to the SBP (Section 2) are commonly used to design generative models. For instance, De Bortoli et al. (2021) proposes a novel approximation of the Iterative Proportional Fitting scheme to solve the SBP for image generation in continuous spaces, analogous to the Sinkhorn algorithm in discrete domain; while Wang et al. (2021) presents a new generative formulation via entropic interpolation with an SBP. Strong connections between SBPs and diffusion models have been established in Chen et al. (2021a) where diffusion models are shown to be special cases of SBPs. While the initial design of DDIBs was not to explicitly mimic a Schrödinger Bridge, our theoretical analysis reveals its true characterization as the solution to a SBP.

## 6 Conclusions

We present Dual Diffusion Implicit Bridges (DDIBs), a new, simplistic image translation method that stems from latest progresses in diffusion models, and is theoretically grounded as Schrödinger Bridges in the image space. DDIBs solve two key problems. First, DDIBs avoid optimization on a coupled loss specific to the given domain pair only. Second, DDIBs better safeguard dataset privacy as they no longer require presence of both datasets during training. Powerful pretrained diffusion models are then integrated into our DDIBs framework, to perform a comprehensive series of experiments that prove DDIBs' practical values in domain translation. Our method is limited in its application to color transfer, as one model is required for each image, which demands significant compute for mass experiments. Rooted in optimal transport, DDIBs translation mimics the mass-moving process which may be problematic at times (Appendix C). Future work may remedy these issues, or extend DDIBs to applications with different dimensions in the source and target domains.

## REFERENCES


Amjad Almahairi, Sai Rajeshwar, Alessandro Sordoni, Philip Bachman, and Aaron Courville. Augmented cyclegan: Learning many-to-many mappings from unpaired data. In *International Conference on Machine Learning*, pp. 195–204. PMLR, 2018.

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

Tianrong Chen, Guan-Horng Liu, and Evangelos A Theodorou. Likelihood training of schr\" odinger bridge using forward-backward sdes theory. *arXiv preprint arXiv:2110.11291*, 2021a.

Yongxin Chen, Tryphon T Georgiou, and Michele Pavon. On the relation between optimal transport and schrödinger bridges: A stochastic control viewpoint. *Journal of Optimization Theory and Applications*, 169(2):671–691, 2016.

Yongxin Chen, Tryphon T Georgiou, and Michele Pavon. Stochastic control liaisons: Richard sinkhorn meets gaspard monge on a schrodinger bridge. *SIAM Review*, 63(2):249–313, 2021b.

Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon. Ilvr: Conditioning method for denoising diffusion probabilistic models. *arXiv preprint arXiv:2108.02938*, 2021.

Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8789–8797, 2018.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26:2292–2300, 2013.

Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion schr\" odinger bridge with applications to score-based generative modeling. *arXiv preprint arXiv:2106.01357*, 2021.

Prafulla Dhariwal and Alex Nichol. Diffusion models beat gans on image synthesis. *arXiv preprint arXiv:2105.05233*, 2021.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

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

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *arXiv preprint arXiv:2006.11239*, 2020.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. *arXiv preprint arXiv:2206.00364*, 2022.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *arXiv preprint arXiv:2201.11793*, 2022.

Gwanghyun Kim and Jong Chul Ye. Diffusionclip: Text-guided image manipulation using diffusion models. *arXiv preprint arXiv:2110.02711*, 2021.

Diederik P Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. *arXiv preprint arXiv:2107.00630*, 2021.

Christian Léonard. A survey of the schr\" odinger problem and some of its connections with optimal transport. *arXiv preprint arXiv:1308.0215*, 2013.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps. *arXiv preprint arXiv:2206.00927*, 2022.

Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. Sdedit: Guided image synthesis and editing with stochastic differential equations. *arXiv preprint arXiv:2108.01073*, 2021.

Alex Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. *arXiv preprint arXiv:2102.09672*, 2021.

Marcus Pereira, Ziyi Wang, Ioannis Exarchos, and Evangelos A Theodorou. Neural network architectures for stochastic control using the nonlinear feynman-kac lemma. *arXiv preprint arXiv:1902.03986*, 2019.

Michaël Perrot, Nicolas Courty, Rémi Flamary, and Amaury Habrard. Mapping estimation for discrete optimal transport. *Advances in Neural Information Processing Systems*, 29:4197–4205, 2016.

Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

Hiroshi Sasaki, Chris G Willcocks, and Toby P Breckon. Unit-ddpm: Unpaired image translation with denoising diffusion probabilistic models. *arXiv preprint arXiv:2104.05358*, 2021.

John Schulman, Sergey Levine, Pieter Abbeel, Michael Jordan, and Philipp Moritz. Trust region policy optimization. In *International conference on machine learning*, pp. 1889–1897. PMLR, 2015.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pp. 2256–2265. PMLR, 2015.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020a.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *arXiv preprint arXiv:1907.05600*, 2019.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020b.

Gefei Wang, Yuling Jiao, Qian Xu, Yang Wang, and Can Yang. Deep generative learning via schr\"{o} dinger bridge. *arXiv preprint arXiv:2106.10410*, 2021.

Yuxuan Xiao, Aiwen Jiang, Changhong Liu, and Mingwen Wang. Single image colorization via modified cyclegan. In *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 3247–3251. IEEE, 2019.

Wenju Xu, Chengjiang Long, Ruisheng Wang, and Guanghui Wang. Drb-gan: A dynamic res-block generative adversarial network for artistic style transfer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 6383–6392, 2021.

Qinsheng Zhang and Yongxin Chen. Fast sampling of diffusion models with exponential integrator. *arXiv preprint arXiv:2204.13902*, 2022.

Lei Zhao, Qihang Mo, Sihuan Lin, Zhizhong Wang, Zhiwen Zuo, Haibo Chen, Wei Xing, and Dongming Lu. Uctgan: Diverse image inpainting based on unsupervised cross-space translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5741–5750, 2020a.

Yihao Zhao, Ruihai Wu, and Hao Dong. Unpaired image-to-image translation using adversarial consistency loss. In *European Conference on Computer Vision*, pp. 800–815. Springer, 2020b.

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