Medical
Imaging
Intelligent
Reality
Lab

Dual Diffusion Implicit Bridges for Image-to-Image Translation

Xuan Su, Jiaming Song, Chenlin Meng, Stefano Ermon

Standford University, NVIDIA, CZ Biohub

Published as a conference paper at ICLR 2023

2023.04.12

Content

1. Introduction

- Transferring images from one domain to another while preserving the content representations is an important problem.

- Difficult to obtain paired images → unpaired translation methods are relevant (e.g. CycleGAN, Normalizing flows).

- First drawback: *adaptability* to alternative domains. → Multi-domain translation method is relevant (e.g. StarGAN).
  → Shared domain needs to be carefully chosen *a priori* (e.g. sketches v.s. photos).

- Second drawback: *lack of privacy protection* of the datasets → Access requirement to both datasets simultaneously and certain privacy-sensitive applications such as medical imaging.

1. Introduction

- For seeking to mitigate both problems, Dual Diffusion Implicit Bridges (DDIBs) are proposed.

- DDIBs is an image-to-image-translation method inspired by score/diffusion models.
  1. Not require paired training.
  2. Stay applicable in other pairs.

- DDIBs are based on the method known as denoising diffusion implicit models (DDIMs).

- Highlight two important theoretical properties:
  1. The *probability flow* (PF) ordinary differential equation (ODE) comprise the solution of a special Schrödinger Bridge Problem (SBP).
  2. DDIBs guarantee exact cycle consistency.

2. Preliminaries

2.1 Score-Based Generative Models (SGMs)

- Two representative models of score-based generative models:
  1. Score matching with Langevin dynamics (SMLD)
  2. Denoising diffusion probabilistic models (DDPMs)

- Both methods are contained within the framework of Stochastic Differential Equations (SDEs).

- **Stochastic Differential Equation (SDE) Representation.** A forward and a corresponding backward SDE describe general diffusion and the reversed, generative processes:

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

* w is the standard Wiener process, $\mathrm{f}(\mathrm{x}, t)$ is the drift coefficient, $g(t)$ is the diffusion coefficient, and $\nabla_{\mathrm{x}} \log p_t(\mathrm{x})$ is the score function.

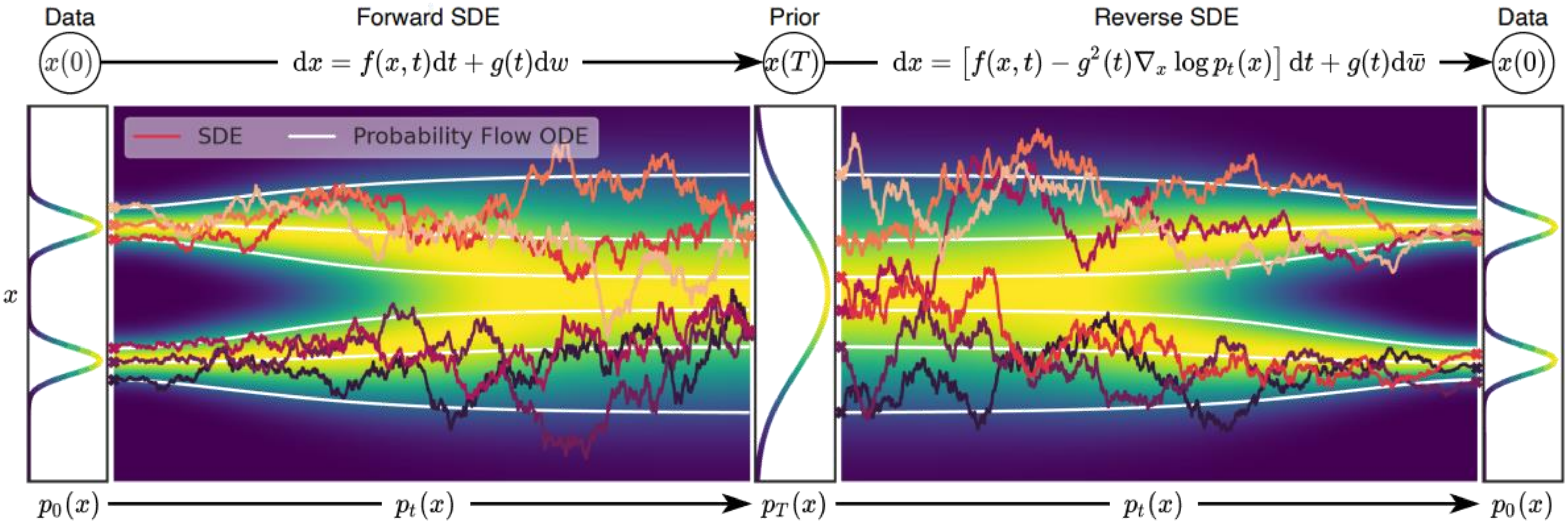
SMLD → A *Variance-Exploding* (VE) SDE: $\mathrm{dx} = \sqrt{\mathrm{d}[\sigma^2(t)] / \mathrm{d}t}\, \mathrm{dw}$.

DDPMs → A *Varinace-Preserving* (VP) SDE: $\mathrm{dx} = -[\beta(t) / 2]\mathrm{x}\, \mathrm{d}t + \sqrt{\beta(t)}\, \mathrm{dw}$

Notably, VP SDE = VE SDE through reparameterization.

2. Preliminaries

2.1 Score-Based Generative Models (SGMs)



Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations.

2. Preliminaries

2.1 Score-Based Generative Models (SGMs)

- **Probability Flow ODE.** Any diffusion process can be represented by a *deterministic* ODE = The *probability flow* (PF) ODE. PF ODEs enable *uniquely identifiable encodings* of data and are central to DDIBs. PF ODE holds the following form:

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

To solve the ODE, numerical ODE solvers are used:

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

In experiments, the ODE solver is implemented in DDIMs.

2. Preliminaries

2.2 Schrödinger Bridge Problem (SBP)

- DDIBs are Schrödinger Bridges between distributions.

- Schrödinger Bridge – as an entropy-regularized optimal transport problem – seeks two optimal policies that transform back-and-forth between two *arbitrary* distributions in a *finite* horizon.

  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 $D(p_0, p_1)$ that minimizes its KL-divergence to W: $P_{SBP} \coloneqq \operatorname{argmin}\{D_{KL}(P||W) \mid P \in D(p_0, p_1)\}$.

- The minimizer, P_{SBP}, is dubbed the *Schrödinger Bridge* between p_0 and p_1 over prior W.

2. Preliminaries

2.2 Schrödinger Bridge Problem (SBP)

- **Relation Between SBPs and SGMs.** Chen et al. (2021) establishes connections between SGMs and SBPs.
  The solution to the optimization can be expressed by the path measure of the forward, or equivalently backward, SDE:

$$\mathrm{dX}_t = [\mathrm{f} + g^2 \nabla_{\mathrm{x}} \log \Psi_t(\mathrm{x})]\, \mathrm{d}t + g(t)\, \mathrm{dw}, \quad \mathrm{dX}_t = \left[\mathrm{f} - g^2 \nabla_{\mathrm{x}} \log \widehat{\Psi}_t(\mathrm{x})\right] \mathrm{d}t + g(t)\, \mathrm{dw}$$

  * $\Psi, \widehat{\Psi}$ are the Schrödinger factors that satisfy density factorization: $p_t(\mathrm{x}) = \Psi_t(\mathrm{x})\widehat{\Psi}_t(\mathrm{x})$.
  * $\mathrm{z}_t = g(t)\nabla_{\mathrm{x}} \log \Psi_t(\mathrm{x}), \hat{\mathrm{z}}_t = g(t)\nabla_{\mathrm{x}} \log \widehat{\Psi}_t(\mathrm{x})$ → considered as the forward, backward policies

  The data log-likelihood objective for SBPs is computed and shown to be equal to that of SGMs with special choices of $\mathrm{z}_t, \hat{\mathrm{z}}_t$:

$$(\mathrm{z}_t, \hat{\mathrm{z}}_t) = \left(0, g(t)\nabla_{\mathrm{x}} \log p_t(\mathrm{x})\right)$$

Chen et al (2021). Tianrong Chen, Guan-Horng Liu, and Evangelos A Theodorou. Likelihood training of schrödinger bridge using forward-backward sdes theory. *arXiv preprint arXiv:2110.11291*, 2021.

2. Preliminaries

2.2 Schrödinger Bridge Problem (SBP)

- **Probability Flow ODE.** A deterministic PF ODE can be derived for SBPs and SGMs.

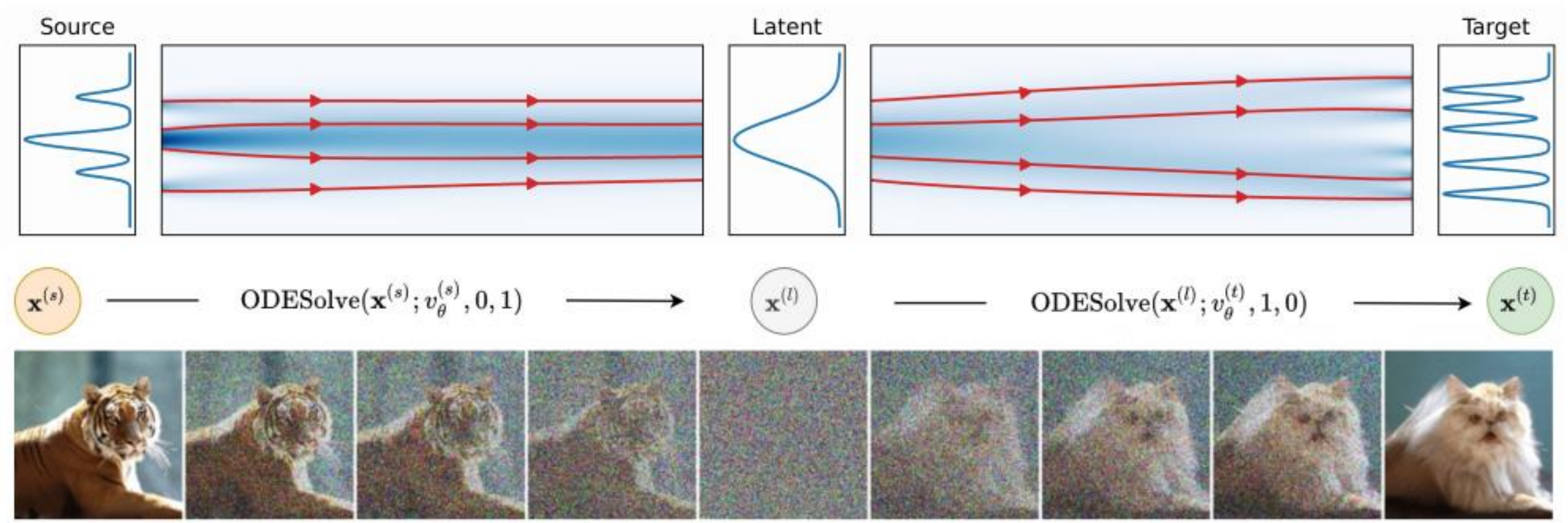
$$\mathrm{dx} = \left[\mathrm{f}(\mathrm{x}, t) + g(t)\mathrm{z} - \frac{1}{2} g(t)(\mathrm{z} + \hat{\mathrm{z}})\right] \mathrm{d}t, \quad \mathrm{dx} = \left[\mathrm{f}(\mathrm{x}, t) - \frac{1}{2} g(t)^2 \nabla_{\mathrm{x}} \log p_t(\mathrm{x})\right] \mathrm{d}t$$

In special choices of $(\mathrm{z}_t, \hat{\mathrm{z}}_t) = \left(0, g(t)\nabla_{\mathrm{x}} \log p_t(\mathrm{x})\right)$, the PF ODES for SGMs and SBPs are equivalent.

In summary:

1. SGMs are implicit optimal transport models, corresponding to SBPs with linear or degenerate drifts.
2. General SBPs additionally accept fully nonlinear diffusion.

3. Dual Diffusion Implicit Bridges



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 $\mathbf{x}^{(t)}$

3. Dual Diffusion Implicit Bridges

- Despite the simplicity of the method, DDIBs have several advantages.

- **Exact Cycle Consistency.** 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.

  Proposition 3.1 (DDIBs Enforce Exact Cycle Consistency)**.**

$$\text{source} \rightarrow \text{target: } \mathrm{x}^{(l)} = \mathrm{ODESolve}\left(\mathrm{x}^{(s)}; v_\theta^{(s)}, 0, 1\right); \quad \mathrm{x}^{(t)} = \mathrm{ODESolve}\left(\mathrm{x}^{(l)}; v_\theta^{(t)}, 1, 0\right);$$

$$\text{target} \rightarrow \text{source: } \mathrm{x}'^{(l)} = \mathrm{ODESolve}\left(\mathrm{x}^{(t)}; v_\theta^{(t)}, 0, 1\right); \quad \mathrm{x}'^{(s)} = \mathrm{ODESolve}\left(\mathrm{x}'^{(l)}; v_\theta^{(s)}, 1, 0\right)$$

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

  As PF ODEs are used, the cycle consistency property is guaranteed.

- **Data Privacy in Both Domains.** Only the source and target diffusion models are required, whose training processes do not depend on knowledge of the domain pair *a priori*. This process can be performed in a privacy sensitive manner.

3. Dual Diffusion Implicit Bridges

- **DDIBs are Two Concatenated Schrödinger Bridges.**
  DDIB process: the source data distribution → the latent space → the target distribution

  Question: What is the nature of such connections between distributions?
  Answer: an optimal transport perspective: these connections are special Schrödinger Bridges.

  Proposition 3.2 (PF ODE Equivalence)**.**

$$\mathrm{dx} = \left[\mathrm{f}(\mathrm{x}, t) + g(t)\mathrm{z} - \frac{1}{2} g(t)(\mathrm{z} + \hat{\mathrm{z}})\right] \mathrm{d}t, \quad \mathrm{dx} = \left[\mathrm{f}(\mathrm{x}, t) - \frac{1}{2} g(t)^2 \nabla_{\mathrm{X}} \log p_t(\mathrm{x})\right] \mathrm{d}t$$

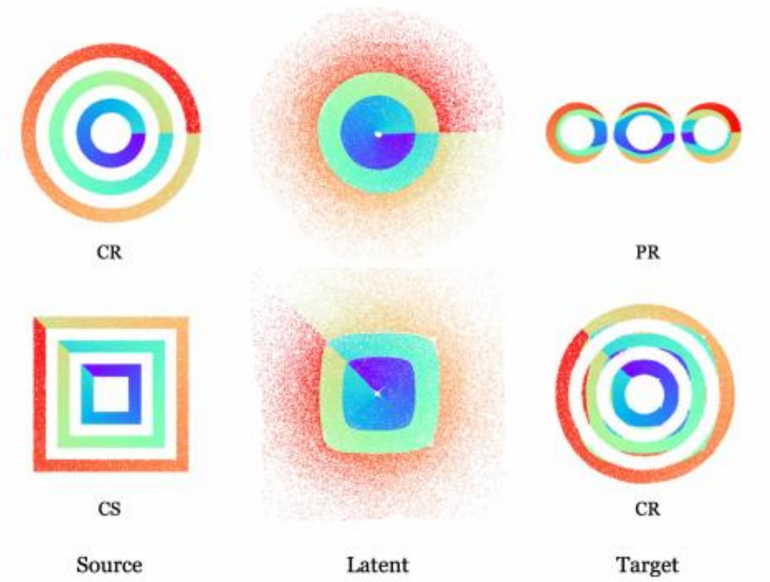
  These two PF ODEs for SBPs and SGMs are equivalent with forward, backward policies $(\mathrm{z}_t, \hat{\mathrm{z}}_t) = \left(0, g(t)\nabla_{\mathrm{X}} \log p_t(\mathrm{x})\right)$ and particularly in DDIMs.

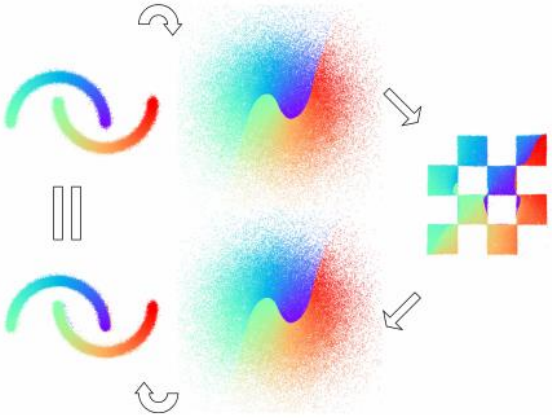
  Thus, DDIBs are intrinsically entropy-regularized optimal transport. The translation process can then be recognized as traversing through two concatenated Schrödinger Bridges, one forward and one reversed. The mapping is unique and minimizes an optimal transport objective.

4. Experiments

4.1 2D Synthetic Experiments

- Domain translation on synthetic datasets drawn from complex two-dimensional distributions, with various shapes and configurations.
- Totally six 2D datasets: Moons (M); Checkerboard (CB); Concentric Rings (CR); Concentric Squares (CS); Parallel Rings (PR); and Parallel Squares (PS). Normalized to zero mean, and identity covariance.
- **Cycle Consistency.** The cycle consistency property guaranteed by DDIBs.





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

4. Experiments

4.2 Example-Guided Color Transfer

- Comparison to Alternative OT Methods. Comparison by the pixel-wise MSEs between color-transferred images generated by DDIBs, and images produced by alternate methods (Earth Mover’s Distance; Sinkhorn distance; linear and Gaussian mapping estimation).

Reference Image | Source Image 1 | Target Image 1 | Source Image 2 | Target Image 2

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

4. Experiments

4.3 Quantitative Translation Evaluation

- Quantitatively, DDIBs deliver competitive results on paired domain tests compared to CycleGAN and AlignFlow.
- **Paired Domain Translation.** DDIBs are evaluated on benchmark paired dataset: Facades and Maps. One dataset contains real photos taken via a camera or a satellite; while the other comprises the corresponding segmentation images.

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

4. Experiments

4.4 Class-Conditional ImageNet Translation

- DDIBs are applied to translation among ImageNet classes using the pretrained diffusion models from Dhariwal & Nichol (2021). These models incorporate a technique known as *classifier guidance*. DDIBs are able to create faithful target image that maintain much of the original content such as animal poses, complexions and emotions.

- **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.

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

5. Conclusions

- Dual Diffusion Implicit Bridges (DDIBs) – a new, simplistic image translation method that stems from latest progresses in score-based diffusion models - are 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.

- DDIBs are limited in their application to color transfer. Rooted in optimal transport, DDIBs translation mimics the mass-moving process which may be problematic at times. Future work may remedy these issues, or extend DDIBs to applications with different dimensions in the source and target domains.