Inverse Bridge Matching Distillation

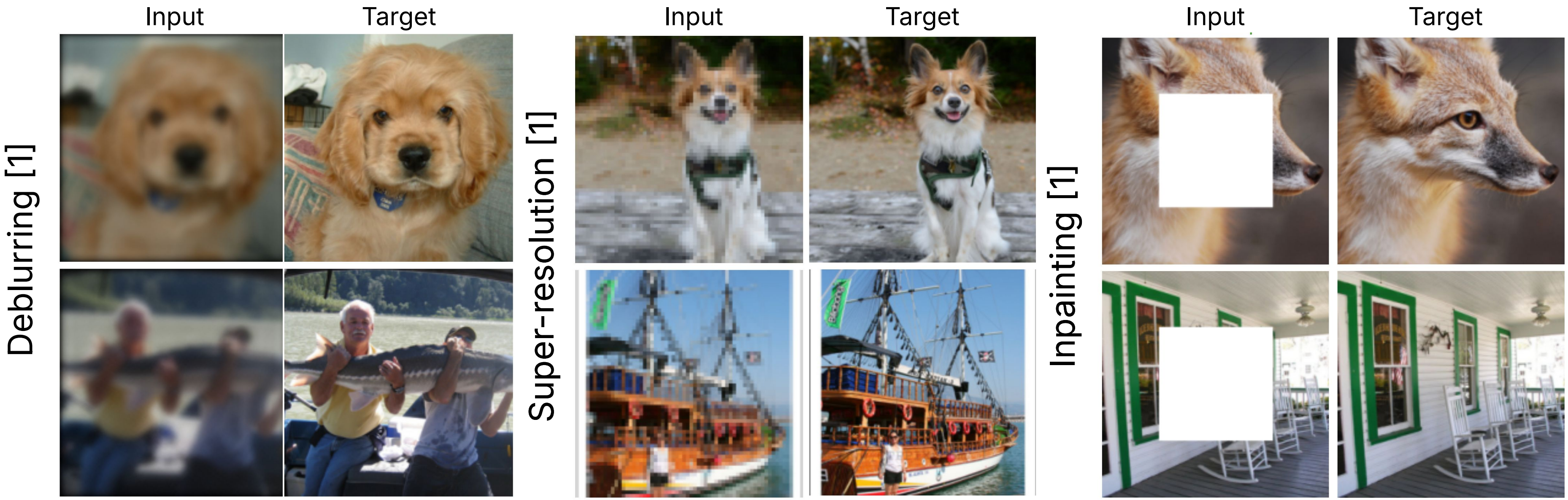
Nikita Gushchin[1,4] David Li[1,5,*] Daniil Selikhanovych[1,2,3,*] Evgeny Burnaev[1,4]
Dmitry Baranchuk[2] Alexander Korotin[1,4]

[1] Skolkovo Institute of Science and Technology
[2] Yandex Research
[3] HSE University
[4] Artificial Intelligence Research Institute
[5] Moscow Institute of Physics and Technology
[*] Equal contribution

International Conference On Machine Learning

Paired image-to-image translation problem

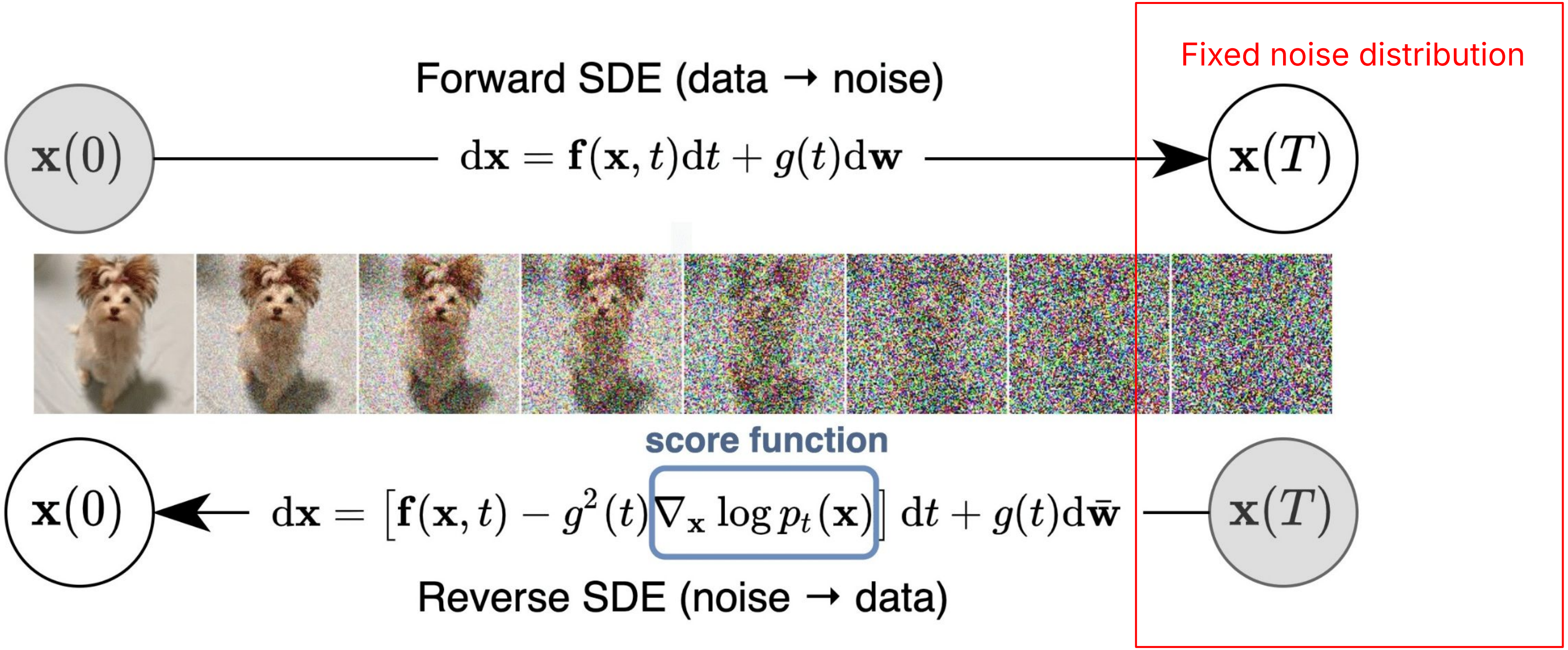
Paired image-to-image translation problem: given a paired dataset of corrupted/clean images — or of two distinct domains — the task is to construct a model that learns to transform inputs into aligned target images.



The question: Diffusion models are known for their superior image generation quality. Can we adapt them to solve image-to-image translation tasks?

Recap on diffusion models

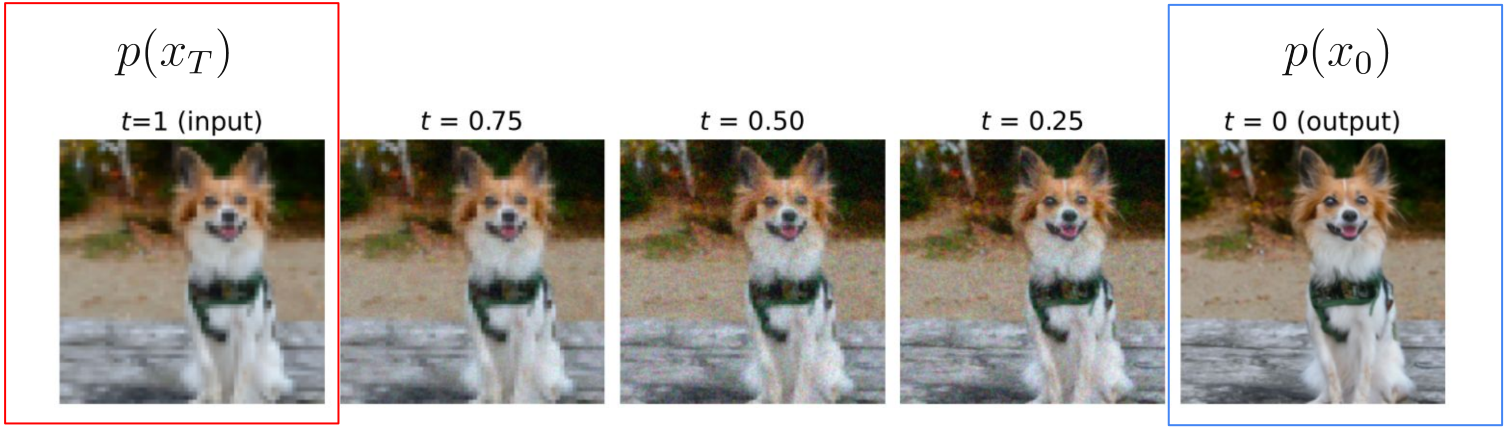
General idea of diffusion models: Define a forward SDE that gradually transforms data into noise. Then, learn the reverse process by estimating the score function using score matching.



The original question remains: Diffusion models are known for their superior image generation quality. Can we adapt them to image-to-image translation tasks?

What exactly do we want to achieve?

Motivation for diffusion bridge models: Build a diffusion process that directly transforms the input distribution $p(x_T)$ into the target distribution $p(x_0)$.



Generation process using the constructed diffusion:

$$dx_t = f_{\text{Bridge}}(x_t, t)dt + g(t)d\bar{w}_t, x_T \sim p(x_T).$$

Unlike standard diffusion which starts from noise, here we start from structured input.

Diffusion bridge

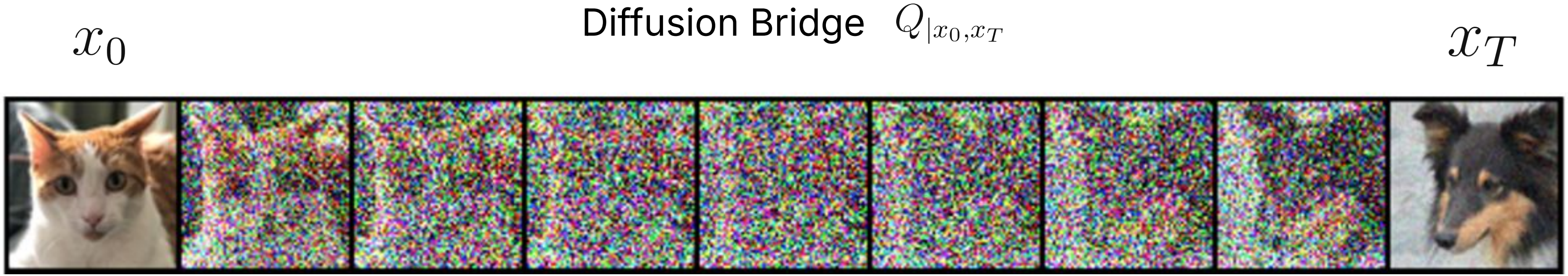
Let Q be a forward diffusion process (the “Prior”) over [0, T] in $\mathbb{R}^D$:

$$Prior\ Q: \quad dx_t = f(x_t, t)dt + g(t)dw_t,$$

Where $f(x_t, t) : \mathbb{R}^D \times [0, T] \to \mathbb{R}^D$ is a drift function and $g(t) : [0, T] \to \mathbb{R}^D$ is a noise schedule.

Diffusion bridge is the conditional stochastic process $Q_{|x_0, x_T}$, where the trajectory is fixed at given x_0 and x_T. It can be derived using the Doob h-transform [2, 3].

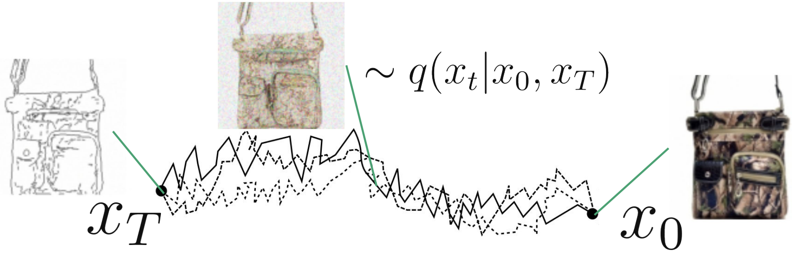
Example. For $f(x_t, t) = 0$ and $g(t) = \sigma$ diffusion bridge is the interpolation combined with noising and denoising:



Bridge Matching: General Approach to Build Diffusion Bridge Models

1. Build a mixture of diffusion bridges using paired data coupling $p(x_0, x_T)$ and a diffusion bridge $Q_{|x_0,x_T}$:

$\sim p(x_0, x_T)$



Diffusion Bridge $Q_{|x_0,x_T}$

2. Learn a diffusion that approximates this mixture by solving the Bridge Matching [1, 4, 5, 6] problem — a generalization of Flow Matching.

$$dx_t = \{f(x_t, t) - g^2(t)v^*(x_t, t)\}dt + g(t)d\bar{w}_t, x_T \sim p(x_T)$$

where the drift $v^*(x_t, t)$ is learned via solving Bridge Matching problem:

$$\min_\phi \mathbb{E}_{x_0,t,x_t}\left[\|v_\phi(x_t, t) - \boxed{\nabla_{x_t} \log q(x_t|x_0)}\|^2\right],$$

Known analytically for simple Q, such as linear SDEs.

$$(x_0, x_T) \sim p(x_0, x_T),\ t \sim U([0, T]),\ x_t \sim q(x_t|x_0, x_T).$$

Diffusion bridge models

The diffusion bridge model is provided by the reverse-time SDE:

$$dx_t = \{f(x_t, t) - g^2(t)v^*(x_t, t)\}dt + g(t)d\bar{w}_t, x_T \sim p(x_T)$$

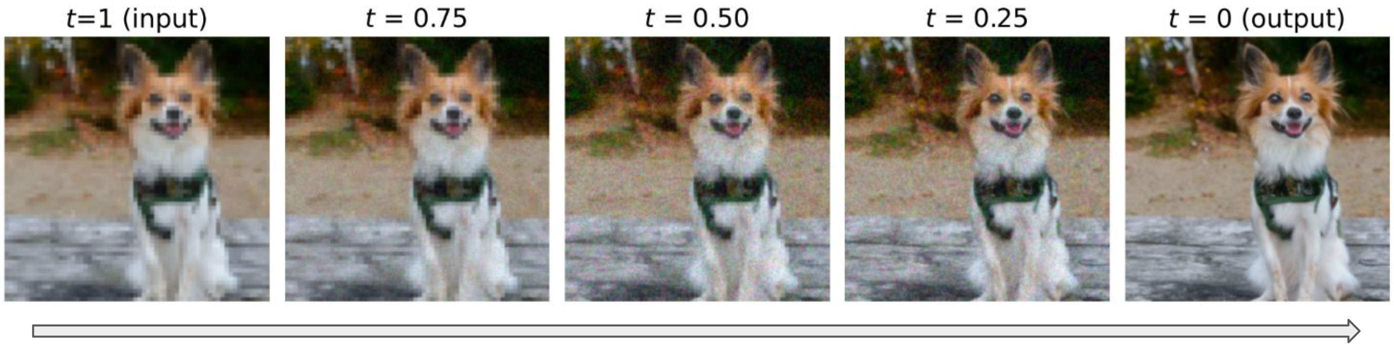
The drift $v^*(x_t, t)$ is learned via solving Bridge Matching problem:

$$\min_{\phi} \mathbb{E}_{x_0,t,x_t}\left[\|v_\phi(x_t, t) - \boxed{\nabla_{x_t} \log q(x_t|x_0)}\|^2\right],$$

Known analytically for simple Q, such as linear SDEs.

$$(x_0, x_T) \sim p(x_0, x_T),\ t \sim U([0, T]),\ x_t \sim q(x_t|x_0, x_T).$$

After learning one can simulate the reverse time SDE using numerical solvers (10-1000 steps) to transform input (corrupted) image to the clean image:



Generation process

Conditional and unconditional diffusion bridge models [7].

There are two types of diffusion bridge models (DBMs) which differ only by introducing additional conditioning on the final point x_T.

Unconditional DBMs:

The goal is to learn SDE drift $v(x_t, t)$ by optimizing:

$$\min_{\phi} \mathbb{E}_{x_0,t,x_t}\left[\|v_\phi(x_t,t) - \nabla_{x_t}\log q(x_t|x_0)\|^2\right],$$
$$(x_0, x_T) \sim p(x_0, x_T),\ t \sim U([0,T]),\ x_t \sim q(x_t|x_0, x_T).$$

Examples: I2SB [1], Stochastic Interpolants [8].

Conditional DBMs:

The goal is to learn SDE drift $v(x_t, t, x_T)$by optimizing:

$$\min_{\phi} \mathbb{E}_{x_0,t,x_t,x_T}\left[\|v_\phi(x_t,t,x_T) - \nabla_{x_t}\log q(x_t|x_0)\|^2\right]$$
$$(x_0, x_T) \sim p(x_0, x_T), \text{and } x_t \sim q(x_t|x_0, x_T).$$

Examples: DDBM [2], GOUB [9], ResShift [10].

Theoretical aspects: Conditional DBMs can be represented as classical diffusion models with a modified forward process. This allows adaptation of acceleration techniques developed for classical DMs. In contrast, unconditional DBMs cannot be expressed this way.

Bridge Matching in practice and x_0 reparametrization

In practice, the Prior process Q is chosen such, that both transitional density $q(x_t|x_0) = \mathcal{N}(x_t|\alpha_t x_0, \sigma^2 I)$ and bridge density $q(x_t|x_0, x_T)$ are Gaussian. In this case $\nabla_{x_t} \log q(x_t|x_0) = -\frac{x_t - \alpha_t x_0}{\sigma_t^2}$.

Hence, one can use change of variables: $v(x_t, t, x_T) = -\frac{x_t - \alpha_t \widehat{x}_0(x_t, t, x_T)}{\sigma_t^2}$

Original problem:

$$\min_{\phi} \mathbb{E}_{x_0, t, x_t, x_T} \left[\| v_\phi(x_t, t, x_T) - \nabla_{x_t} \log q(x_t|x_0) \|^2 \right]$$

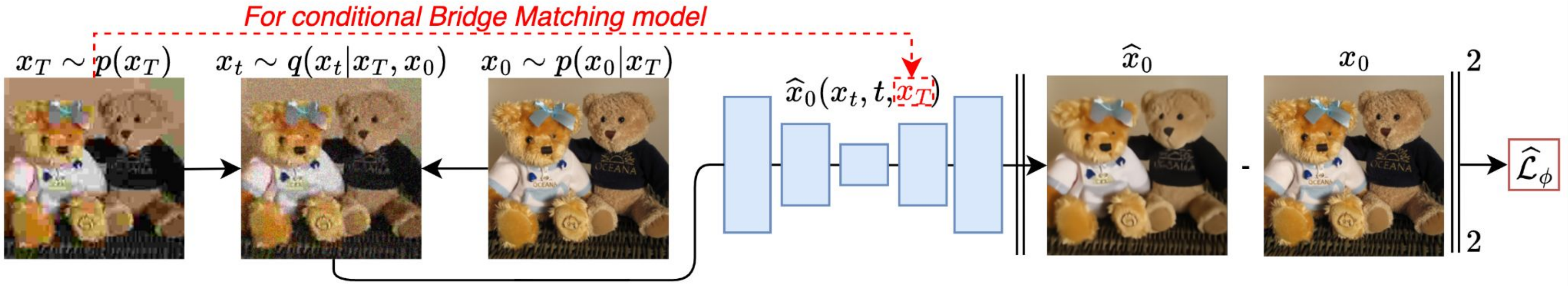
$$(x_0, x_T) \sim p(x_0, x_T), \text{and } x_t \sim q(x_t|x_0, x_T).$$

Reparameterized problem:

$$\min_{\phi} \mathbb{E}_{x_0, t, x_t, x_T} \left[\lambda(t) \| \widehat{x}_0^{\phi}(x_t, t, x_T) - x_0 \|^2 \right], \quad (7)$$

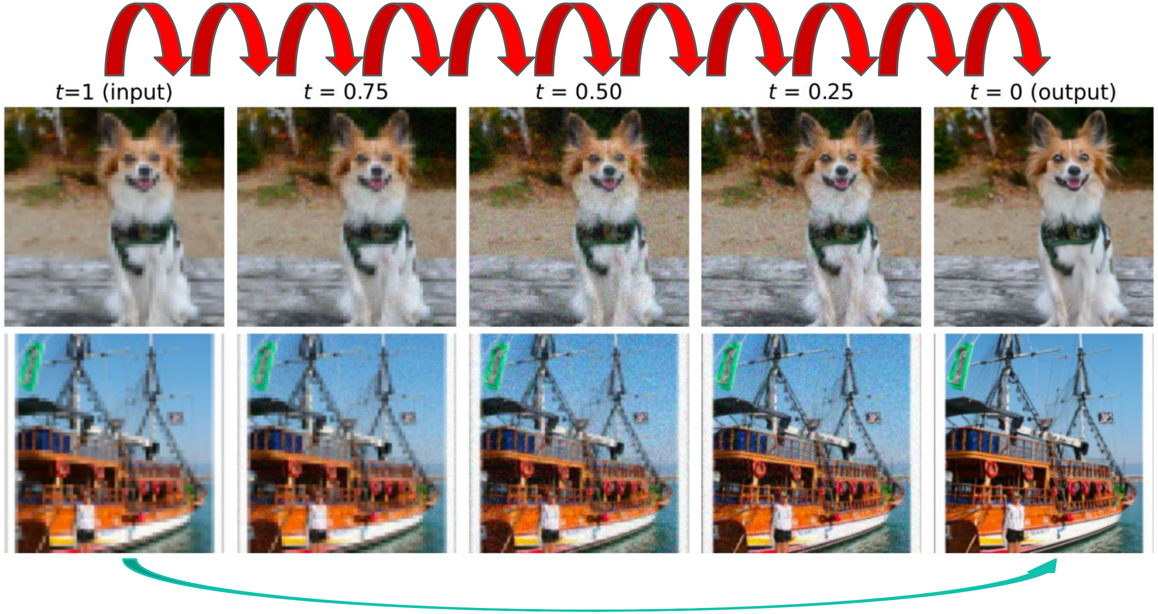
$$(x_0, x_T) \sim p(x_0, x_T),\ t \sim U([0, T]),\ x_t \sim q(x_t|x_0, x_T),$$

Learning pipeline for the reparameterized Diffusion Bridge model (Bridge Matching model)



Distillation of diffusion bridge models

Long inference problem: Diffusion bridge models produce high-quality translations, but — like classical diffusion models — they require 10 to 1000 of steps to simulate the reverse SDE.



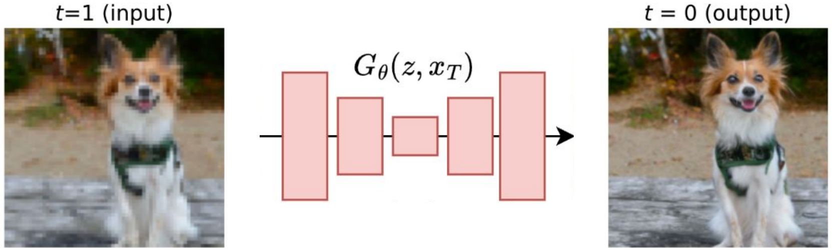
The goal of distillation is to train a new 1-step or few-step generator that mimics the full diffusion bridge model.

Inverse Bridge Matching Distillation [11]

The core idea is to learn a one-step generator G_θ of clean images from corrupted such that Diffusion Bridge Model for generated data coupling $p_\theta(x_0, x_T)$ matches the teacher Diffusion Bridge Model learned on the ground-truth data:

1. We parameterize stochastic map $p_\theta(x_0|x_T)$ by the one step stochastic generator $G_\theta(x_T, z)$,

$$z \sim N(0, I), x_T \sim p(x_T).$$



2. We learn the generator $G_\theta(x_T, z)$ in such way, that diffusion model for the generated data x_0^θ matches the pre-trained teacher data x_0^*.

Teacher model

Model for Gθ

$$\min_\theta \mathbb{E}_{x_t, t, x_0, x_T} \left[\lambda(t) \| \boxed{\widehat{x}_0^*(x_t, t, x_T)} - \boxed{\widehat{x}_0^\theta(x_t, t, x_T)} \|^2\right],$$

$$\text{s.t.} \quad \boxed{\widehat{x}_0^\theta = \arg\min_{\widehat{x}_0} \mathbb{E}_{x_t, t, x_0, x_T} \left[\lambda(t) \| \widehat{x}_0(x_t, t, x_T) - x_0 \|^2\right],}$$

$$\boxed{(x_0, x_T) \sim p_\theta(x_0, x_T)}, \; t \sim U([0, T]), \; x_t \sim q(x_t | x_0, x_T),$$

Data from G_θ

Intractable gradients problem since the objective includes term with argminimum.

Tractable objective for the inverse bridge matching problem.

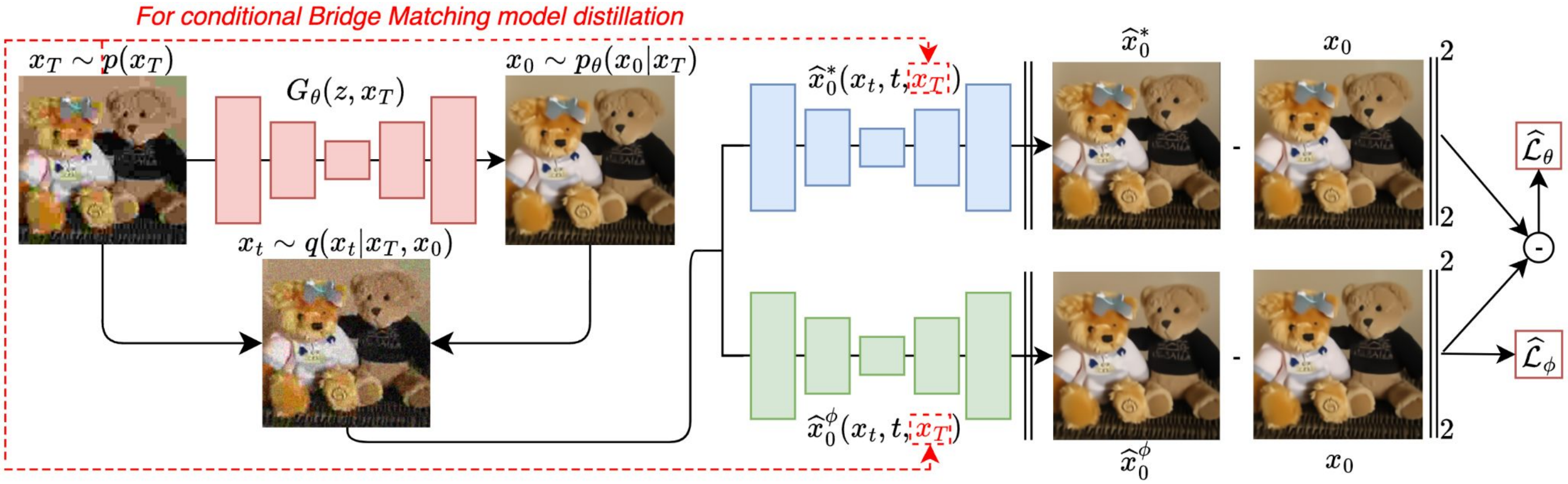
Solution: we show that the original constrained problem can be reformulated in the unconstrained one and used in practise:

$$\min_{\theta} \Big[\mathbb{E}_{x_t,t,x_0,x_T} \big[\lambda(t) \| \boxed{\widehat{x}_0^*(x_t,t,x_T)} - x_0 \|^2 \big] - \quad \text{Teacher model}$$

$$\min_{\phi} \mathbb{E}_{x_t,t,x_0,x_T} \big[\lambda(t) \| \boxed{\widehat{x}_0^{\phi}(x_t,t,x_T)} - x_0 \|^2 \big] \Big], \quad \text{Model for G}\theta$$

$$(x_0,x_T) \sim \underbrace{\boxed{p_\theta(x_0,x_T)}}_{\text{Data from } G_\theta},\; t \sim U([0,T]),\; x_t \sim q(x_t|x_0,x_T).$$

Learning pipeline for the inverse bridge matching distillation



The results of Inverse Bridge Matching Distillation (IBMD, ours).

Unconditional model,
1-step IBMD (ours)
beats **1000-step** teacher model.

Unconditional model,
1-step IBMD (ours)
accelerate inference up to **100 times.**

Conditional model,
IBMD (ours)
accelerate inference up to **5 times.**

Conditional model,
IBMD (ours)
accelerate inference in **50 times.**

We show the applicability of our IBMD distillation method in both *unconditional* and *conditional* settings.

We evaluate IBMD on diverse tasks, including super-resolution, JPEG restoration, inpaiting, and sketch-to-image translation.

In all settings, IBMD achieves significant inference speedups (up to 100×) and sometimes improves generation quality over the teacher.

IBMD (ours) outperforms previous approached for DBMs acceleration

4× super-resolution (bicubic)	ImageNet (256 × 256)		
	NFE	FID ↓	CA ↑
DDRM (Kawar et al., 2022)	20	21.3	63.2
DDNM (Wang et al., 2023)	100	13.6	65.5
ΠGDM (Song et al., 2023)	100	3.6	72.1
ADM (Dhariwal & Nichol, 2021)	1000	14.8	66.7
CDSB (Shi et al., 2022)	50	13.6	61.0
I^2SB (Liu et al., 2023a)	1000	2.8	70.7
IBMD-I^2SB (**Ours**)	1	**2.5**	**72.4**

4× super-resolution (pool)	ImageNet (256 × 256)		
	NFE	FID ↓	CA ↑
DDRM (Kawar et al., 2022)	20	14.8	64.6
DDNM (Wang et al., 2023)	100	9.9	67.1
ΠGDM (Song et al., 2023)	100	3.8	72.3
ADM (Dhariwal & Nichol, 2021)	1000	3.1	73.4
CDSB (Shi et al., 2022)	50	13.0	61.3
I^2SB (Liu et al., 2023a)	1000	2.7	71.0
IBMD-I^2SB (**Ours**)	1	**2.6**	**72.7**

Table 2. Results on the image JPEG restoration task with QF=5. Baseline results are taken from I^2SB (Liu et al., 2023a).

JPEG restoration, QF= 5.	ImageNet (256 × 256)		
	NFE	FID ↓	CA ↑
DDRM (Kawar et al., 2022)	20	28.2	53.9
ΠGDM (Song et al., 2023)	100	8.6	64.1
Palette (Saharia et al., 2022)	1000	8.3	64.2
CDSB (Shi et al., 2022)	50	38.7	45.7
I^2SB (Liu et al., 2023a)	1000	**4.6**	67.9
I^2SB (Liu et al., 2023a)	100	5.4	67.5
IBMD-I^2SB (**Ours**)	1	5.3	67.2

Table 4. Results on the image JPEG restoration task with QF=10. Baseline results are taken from I^2SB (Liu et al., 2023a).

JPEG restoration, QF= 10.	ImageNet (256 × 256)		
	NFE	FID ↓	CA ↑
DDRM (Kawar et al., 2022)	20	16.7	64.7
ΠGDM (Song et al., 2023)	100	6.0	71.0
Palette (Saharia et al., 2022)	1000	5.4	70.7
CDSB (Shi et al., 2022)	50	18.6	60.0
I^2SB (Liu et al., 2023a)	1000	**3.6**	72.1
I^2SB (Liu et al., 2023a)	100	4.4	71.6
IBMD-I^2SB (**Ours**)	1	3.8	72.4

Inpainting, Center (128 × 128)	ImageNet (256 × 256)		
	NFE	FID ↓	CA ↑
DDRM (Kawar et al., 2022)	20	24.4	62.1
ΠGDM (Song et al., 2023)	100	7.3	**72.6**
DDNM (Wang et al., 2022)	100	15.1	55.9
Palette (Saharia et al., 2022)	1000	6.1	63.0
I^2SB (Liu et al., 2023a)	10	5.4	65.97
DBIM (Zheng et al., 2024)	50	3.92	72.4
DBIM (Zheng et al., 2024)	100	**3.88**	**72.6**
CBD (He et al., 2024)	4	5.34	69.6
CBT (He et al., 2024)		4.77	70.3
IBMD-I^2SB (**Ours**)		5.1	70.3
IBMD-DDBM (**Ours**)		**4.03**	**72.2**
CBD (He et al., 2024)	2	5.65	69.6
CBT (He et al., 2024)		5.34	69.8
IBMD-I^2SB (**Ours**)		5.3	65.7
IBMD-DDBM (**Ours**)		**4.23**	**72.3**
IBMD-I^2SB (**Ours**)	1	6.7	65.0
IBMD-DDBM (**Ours**)		**5.87**	**70.6**

Our method (IBMD) **outperforms** previous acceleration methods based on consistency distillation (CBD/CBT) and more advanced sampling techniques (DBIM).

References

1. Liu, Guan-Horng, et al. "I $^ 2$ SB: Image-to-Image Schrödinger Bridge." *International Conference on Machine Learning*. PMLR, 2023.
2. Zhou, Linqi, et al. "Denoising Diffusion Bridge Models." *The Twelfth International Conference on Learning Representations*.
3. Doob, Joseph L., and J. I. Doob. *Classical potential theory and its probabilistic counterpart*. Vol. 262. New York: Springer, 1984.
4. Peluchetti, Stefano. "Non-denoising forward-time diffusions." *arXiv preprint arXiv:2312.14589* (2023).
5. Liu, X., Gong, C., & Liu, Q. (2022). Flow straight and fast: Learning to generate and transfer data with rectified flow. *arXiv preprint arXiv:2209.03003*.
6. Shi, Yuyang, et al. "Diffusion schrödinger bridge matching." Advances in Neural Information Processing Systems 36 (2023): 62183-62223.
7. De Bortoli, Valentin, et al. "Augmented bridge matching." *arXiv preprint arXiv:2311.06978* (2023).
8. Albergo, Michael Samuel, et al. "Stochastic Interpolants with Data-Dependent Couplings." *International Conference on Machine Learning*. PMLR, 2024.
9. Yue, Conghan, et al. "Image Restoration Through Generalized Ornstein-Uhlenbeck Bridge." *Forty-first International Conference on Machine Learning*.
10. Yue, Z., Wang, J., & Loy, C. C. (2023). Resshift: Efficient diffusion model for image super-resolution by residual shifting. *Advances in Neural Information Processing Systems*, *36*, 13294-13307.
11. Gushchin, Nikita, et al. "Inverse Bridge Matching Distillation." *arXiv preprint arXiv:2502.01362* (2025).