BRACE: ADVERSARIAL AND CONTRASTIVE ENDPOINT REGULARIZATION FOR SCHRÖDINGER-BRIDGE IMAGE ENHANCEMENT

Anonymous authors
Paper under double-blind review

ABSTRACT

Unpaired image enhancement can be viewed as transporting a degraded image distribution toward a clean one under a strong constraint of source-scene fidelity. Unlike broad domain translation, the output must undergo substantial appearance correction while remaining closely aligned with the input structure. Schrödinger-bridge methods provide a principled stochastic transport framework for this setting, but existing bridge-based objectives can be overly conservative, underconstrain target realism, and require multiple refinement steps at inference time. We propose Bridge Regularization with Adversarial and Contrastive Endpoints (BRACE), a Schrödinger-bridge method that strengthens endpoint supervision with explicit constraints for target realism and local structural consistency. The resulting translator preserves the bridge construction while producing more realistic and faithful outputs with very small inference budget. Across low-light enhancement, dehazing, and underwater enhancement benchmarks, BRACE consistently improves over a Schrödinger-bridge baseline and remains competitive with dedicated unpaired baselines. Our results indicate that explicit endpoint supervision enables effective bridge-based transport for image enhancement with few inference steps.

1 INTRODUCTION

Image restoration aims to recover a visually improved image from a degraded observation. In the supervised setting, learning is based on paired examples, where each degraded input is matched with a clean, pixel-aligned target. Such pairs simplify the problem, since the model can be trained by direct comparison to the desired output. However, in many practical settings, paired data are difficult or impossible to obtain. This challenge arises across several enhancement problems, including low-light correction, dehazing, and underwater restoration (Wei et al., 2018; Li et al., 2018; 2019). Restoration must then be treated as an unpaired image-to-image translation problem, where the goal is to transport the empirical distribution of degraded images toward that of clean images while preserving the content of each individual input (Isola et al., 2017; Liu et al., 2017). The latter requirement is especially restrictive for enhancement tasks, where the translated image must remain tightly aligned with the source scene despite a substantial change in appearance.

Existing unpaired restoration methods have largely followed two directions. GAN-based models (Zhu et al., 2017; Park et al., 2020) learn target-domain appearance through adversarial supervision and use additional constraints to preserve source content. More recently, optimal-transport and Schrödinger-bridge methods have formulated unpaired translation as stochastic transport between empirical image measures (De Bortoli et al., 2021; Shi et al., 2023; De Bortoli et al., 2024). In this view, unpaired translation is realized by learning a stochastic process whose endpoint marginals match the source and target distributions. This perspective is particularly appealing for restoration, since it offers a principled way to combine distributional transport with input-conditioned structural consistency.

At the same time, unpaired image enhancement exposes a gap that is not fully resolved by existing approaches. Unlike broad domain translation, enhancement requires large appearance changes while keeping the output tightly tied to the source scene. This makes the learning objective especially

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

delicate: adversarial criteria can encourage target-domain appearance but offer only weak control over input fidelity (Blau & Michaeli, 2018; Cohen et al., 2018), whereas bridge-based objectives are typically defined through endpoint regression (Shi et al., 2023; De Bortoli et al., 2024), which provides strong structural anchoring but only indirect control over perceptual realism. The missing ingredient is therefore not a different transport mechanism, but stronger supervision at the translated endpoint itself: the model should be pushed explicitly toward the target domain while remaining constrained by local evidence from the input image.

In this paper, we introduce Bridge Regularization with Adversarial and Contrastive Endpoints (BRACE), a Schrödinger-bridge method for unpaired image enhancement. The method strengthens endpoint supervision by combining bridge-based translation with explicit constraints for target-domain realism and local correspondence. In particular, it augments the bridge-based objective with adversarial supervision at the translated endpoint and a PatchNCE constraint against the source image. Our contributions are threefold. (i) We introduce a Schrödinger-bridge framework for unpaired image enhancement that explicitly regularizes translated endpoints for realism and local fidelity. (ii) We show that this endpoint supervision supports effective inference with very small inference-step budgets. (iii) We evaluate the method on unpaired enhancement benchmarks spanning low-light correction, dehazing, and underwater restoration, together with a standard translation benchmark, to study how stronger endpoint supervision affects bridge-based translation across these settings.

2 Related Work

Unpaired translation with adversarial models. GAN-based methods were the dominant early approach to unpaired image-to-image translation. CycleGAN (Zhu et al., 2017) preserves source content through cycle consistency, while CUT (Park et al., 2020) replaces cycle reconstruction with a patch-based contrastive objective. Variants of this paradigm were later applied to restoration settings with unpaired degraded and clean images (Jiang et al., 2021; Engin et al., 2018; Fabbri et al., 2018). These methods remain strong baselines, but content preservation is enforced only indirectly, and the trade-off between perceptual realism and structural fidelity can be difficult to control (Blau & Michaeli, 2018; Cohen et al., 2018).

Unpaired image enhancement. A parallel line of work targets specific degradations rather than generic translation. In low-light enhancement, methods such as EnlightenGAN (Jiang et al., 2021), GDP (Fei et al., 2023), and AGLLDiff (Lin et al., 2025) combine unpaired or weakly supervised training with task-specific inductive biases. Related formulations have also been developed for dehazing (Engin et al., 2018; Lan et al., 2025b;a) and underwater restoration (Fabbri et al., 2018; Schein et al., 2024). These approaches are often effective within a fixed degradation family, but they do not aim for a unified transport formulation across multiple enhancement tasks. Our focus is instead on a single bridge-based framework that can be applied across low-light enhancement, dehazing, and underwater restoration.

Schrödinger bridges for unpaired transport. Schrödinger-bridge methods offer a transport-based alternative to adversarial mappings. Starting from the classical path-space formulation (Léonard, 2013), diffusion-based solvers made bridge learning feasible through iterative proportional fitting (IPF)-style procedures (De Bortoli et al., 2021). Diffusion Schrödinger Bridge Matching introduced Iterative Markovian Fitting (IMF) and its implementation DSBM (Shi et al., 2023), where bridge learning reduces to Brownian bridge regression within an alternating projection scheme. Schrödinger Bridge Flow (SBFlow) (De Bortoli et al., 2024) recast this line as a flow of path measures, resulting in α-IMF and an online bidirectional translator that avoids sample caching and can be implemented with a single conditioned network. This family forms the technical basis of our method.

Bridge-based variants. Subsequent work has extended the Schrödinger-bridge framework in several directions. UNSB (Kim et al., 2024) casts unpaired translation as a sequence of adversarial Schrödinger-bridge problems, while ASBM (Gushchin et al., 2024) develops a discrete-time adversarial IMF solver that operates at very small inference budgets. Other works move toward restoration-specific formulations: DDSB (Yi et al., 2026) introduces degradation-aware transport with trajectory-consistency regularization, and DehazeSB (Lan et al., 2025b) adapts bridge-based

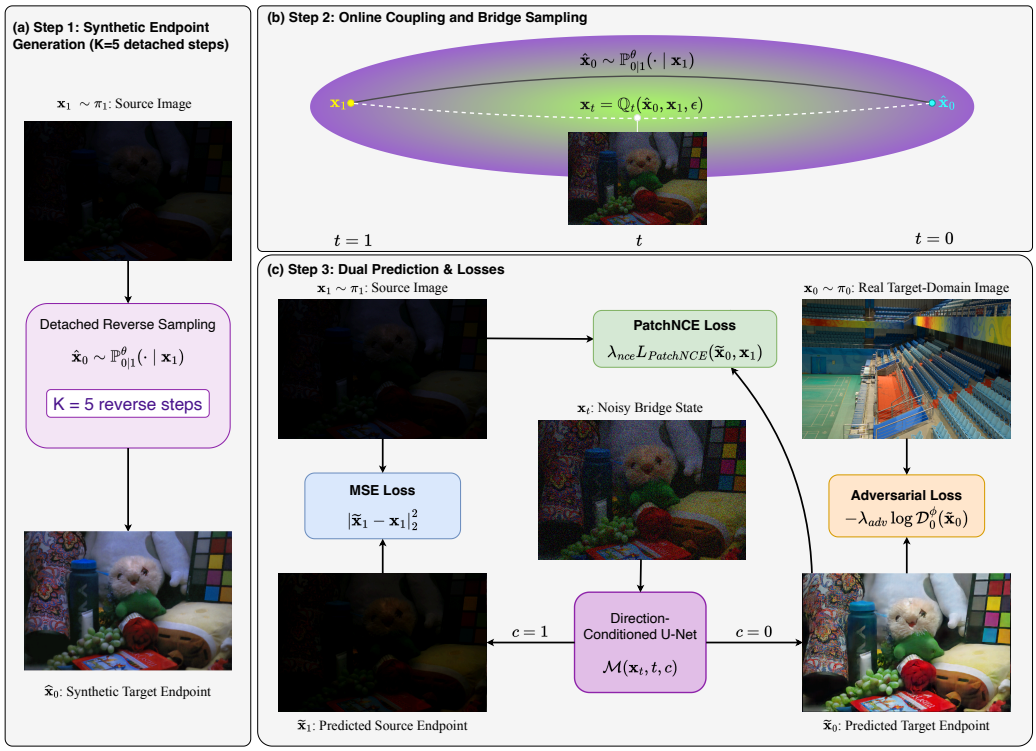


Figure 1: Overview of BRACE for the $1 \to 0$ training direction. A detached K-step sampler first generates a synthetic target endpoint $\hat{\mathbf{x}}_0$ from the source image $\mathbf{x}_1$. A noisy bridge state $\mathbf{x}_t$ is then sampled from the Brownian interpolant between $\mathbf{x}_1$ and $\hat{\mathbf{x}}_0$. From the same bridge state, a direction-conditioned U-Net predicts both endpoints. The inherited MSE term supervises the source-side prediction $\tilde{\mathbf{x}}_1$, while PatchNCE and adversarial supervision regularize the target-side prediction $\tilde{\mathbf{x}}_0$. A real sample from the target domain provides the discriminator signal. The opposite direction is defined symmetrically.

transport to dehazing. IPMF (Kholkin et al., 2026) instead revisits bidirectional IMF heuristics from a combined IPF/IMF perspective. Across these directions, a common limitation for image enhancement is that target-domain realism and source-conditioned local fidelity remain only indirectly controlled at the translated endpoint. Our work addresses this gap by strengthening endpoint supervision within a bridge-based translator.

3 Schrödinger Bridges for Unpaired Image Translation

Let π_0 and π_1 be two distributions defined in $\mathbb{R}^d$ and assume one can independently sample $\mathbf{x}_0 \sim \pi_0$ and $\mathbf{x}_1 \sim \pi_1$. Our goal is to find a coupling distribution Π^* on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals $\Pi^*_0 = \pi_0$ and $\Pi^*_1 = \pi_1$ such that if $\mathbf{x}_0 \sim \pi_0$ and $\hat{\mathbf{x}}_1 \sim \Pi^*(\cdot|\mathbf{x}_0)$, then $\hat{\mathbf{x}}_1 \sim \pi_1$, and it minimizes the Entropy regularized Optimal Transport (EOT) defined as

$$\Pi^* = \underset{\Pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)}{\arg\min} \int \frac{1}{2} ||\mathbf{x}_0 - \mathbf{x}_1||^2 \, \mathrm{d}\Pi(\mathbf{x}_0, \mathbf{x}_1) - \tau \mathcal{H}(\Pi), \quad \mathbf{x}_0 \sim \pi_0, \mathbf{x}_1 \sim \pi_1, \tag{1}$$

where τ is the entropy regularization coefficient, $\mathcal{H}(\Pi)$ is the entropy of Π, and $\mathcal{P}$ is the space of all coupling distributions in $\mathbb{R}^d \times \mathbb{R}^d$ that do not necessarily minimize EOT. Intuitively, we want to move one distribution into the other with minimal difference between corresponding samples[1].

While Equation 1 is the solution we are interested in, it can be achieved through different techniques and formulations. Among them is *the Schrödinger Bridge Problem*, which seeks for a continuous-

[1]Regularized by the entropy $\mathcal{H}(\Pi)$, but for small τ the intuition holds.

time stochastic process $(\mathbf{x}_t)_{t\in[0,1]}$ with endpoints $\mathbf{x}_0 \sim \pi_0$ and $\mathbf{x}_1 \sim \pi_1$. Formally, it can be defined as the following optimization problem

$$\mathbb{P}^* = \underset{\mathbb{P}\in\mathcal{P}([0,1]\times\mathbb{R}^d)}{\arg\min} \{\mathrm{KL}(\mathbb{P}||\mathbb{Q}); \mathbb{P}_0 = \pi_0, \mathbb{P}_1 = \pi_1\}. \tag{2}$$

$\mathbb{P}^*$ is the probability path measure induced by $\mathbf{x}_t$, which minimizes the KL-divergence between the reference path measure $\mathbb{Q}$, usually induced by the scaled Wiener process $\sqrt{\tau}\mathbf{w}_t$. $\mathcal{P}$ is the space of probability path measures defined in $\mathbb{R}^d$ space over the time interval $[0, 1]$. If the reference process is indeed $\sqrt{\tau}\mathbf{w}_t$ and both endpoints $\mathbf{x}_0, \mathbf{x}_1$ are known, then the transition kernels at any time t are given as

$$\mathbb{Q}_{t|0,1} = \mathcal{N}(\mathbf{x}_t; (1-t)\mathbf{x}_0 + t\mathbf{x}_1, \tau t(1-t)\mathbb{I}). \tag{3}$$

There are different approaches in the literature for obtaining $\mathbb{P}^*$, most of which are based on the Iterative Proportional Fitting (IPF). The IPF can be defined as a sequence of optimization problems that alternate between fitting forward ($1 \to 0$) and reverse ($0 \to 1$) directions, producing a sequence of path measures

$$\mathbb{P}^{2n+1} = \underset{\mathbb{P}\in\mathcal{P}([0,1]\times\mathbb{R}^d)}{\arg\min} \{\mathrm{KL}(\mathbb{P}||\mathbb{P}^{2n}); \mathbb{P}_1 = \pi_1\}, \tag{4}$$

$$\mathbb{P}^{2n+2} = \underset{\mathbb{P}\in\mathcal{P}([0,1]\times\mathbb{R}^d)}{\arg\min} \{\mathrm{KL}(\mathbb{P}||\mathbb{P}^{2n+1}); \mathbb{P}_0 = \pi_0\}, \tag{5}$$

with the initialization on $\mathbb{P}^0 = \mathbb{Q}$.

In this work, we show that the IPF procedure, although theoretically converging to $\mathbb{P}^*$, in practice achieves poor results on modern domain transfer benchmarks, and on some problems, such as low-light image enhancement, it collapses completely. We show that stabilising the learning procedure with adversarially trained discriminators on both domains π_0 and π_1 achieves much better results. Our formulation therefore combines the consistency induced by the IPF procedure with explicit adversarial supervision on the endpoint marginals.

4 BRACE

In this section, we introduce our BRACE algorithm, illustrated in Figure 1. First, we describe each individual element of our model; then, we explain the training procedure; lastly, we define the inference.

Model definition We define our generative model as a bridge process $\mathbf{x}_t$ under $\mathbb{P}^\theta$ parameterized with learnable parameters θ. The underlying dynamics of $\mathbf{x}_t$ can be expressed by the following Stochastic Differential Equation (SDE)

$$\mathrm{d}\mathbf{x}_t = \mathbf{v}_\theta(\mathbf{x}_t, t)\ \mathrm{d}t + \sqrt{\tau}\ \mathrm{d}\mathbf{w}_t, \qquad \mathbf{v}_\theta(\mathbf{x}_t, t) = \frac{\mathcal{M}_\theta(\mathbf{x}_t, t, c=1) - \mathbf{x}_t}{1-t}. \tag{6}$$

The definition of the drift $\mathbf{v}_\theta$ follows the Schrödinger Bridge Flow definition De Bortoli et al. (2024), and $\mathcal{M}_\theta$ is a neural network that based on the current state $\mathbf{x}_t$ at time t, and target time $c \in \{0, 1\}$, it estimates clean samples from either π_0 or π_1.

For training purposes, we define two discriminators, $\mathcal{D}_0^\phi$ and $\mathcal{D}_1^\psi$, which determine whether $\mathbf{x}_0$ and $\mathbf{x}_1$, respectively, are sampled from the data distributions π_0 and π_1. We follow the LSGAN Mao et al. (2017) formulation

$$\phi^* = \underset{\phi}{\arg\min}\, \mathbb{E}_{\mathbf{x}_0\sim\pi_0}\left[\left(\mathcal{D}_0^\phi(\mathbf{x}_0) - 1\right)^2\right] + \mathbb{E}_{\hat{\mathbf{x}}_0\sim\mathbb{P}^\theta_{0|t}}\left[\left(\mathcal{D}_0^\phi(\hat{\mathbf{x}}_0)\right)^2\right], \tag{7}$$

$$\psi^* = \underset{\psi}{\arg\min}\, \mathbb{E}_{\mathbf{x}_1\sim\pi_1}\left[\left(\mathcal{D}_1^\psi(\mathbf{x}_1) - 1\right)^2\right] + \mathbb{E}_{\hat{\mathbf{x}}_1\sim\mathbb{P}^\theta_{1|0}}\left[\left(\mathcal{D}_1^\psi(\hat{\mathbf{x}}_1)\right)^2\right]. \tag{8}$$

The consistency between the source and target should be assured by the IPF formulation; however, to strengthen this signal preserving ability, we also incorporated Patchwise Contrastive Loss from Park et al. (2020) (see Appendix C).

Training procedure At each training step, both endpoints are independently sampled from the data $\mathbf{x}_0, \mathbf{x}_1 \sim \pi_0 \otimes \pi_1$. Inspired by IPF, based on the index of the training batch (odd or even), we either optimize for the $0 \to 1$ direction or the $1 \to 0$. The entire training procedure is shown in Algorithm 1; here, for simplicity, we only describe optimizing for $0 \to 1$.

In this scenario, we draw $\mathbf{x}_0 \sim \pi_0$ and the opposite endpoint is sampled from $\hat{\mathbf{x}}_1 \sim \mathbb{P}^{\theta}_{1|0}(\cdot|\mathbf{x}_0)$ at $t = 1$, which corresponds to $\mathbf{x}_0$ using the current version of the modeled bridge process $\mathbb{P}^{\theta}$. We sample from $\mathbb{P}^{\theta}$ using any off-the-shelf sampler (see Appendix B and E) with K intermediate, uniformly distributed steps. The *autograd* is switched off during this generation.

Next, the intermediate sample is drawn $\mathbf{x}_t \sim \mathbb{Q}_{t|0,1}(\cdot|\mathbf{x}_0, \hat{\mathbf{x}}_1)$ using Equation 3 and a random timestep $t \sim \mathcal{U}(0, 1)$. That intermediate sample is then passed through the model $\mathcal{M}_{\theta}$ to obtain predictions of both endpoints $\tilde{\mathbf{x}}_0 = \mathcal{M}_{\theta}(\mathbf{x}_t, t, c = 0)$, and $\tilde{\mathbf{x}}_1 = \mathcal{M}_{\theta}(\mathbf{x}_t, t, c = 1)$.

To summarize, from the previous steps, we have $\{\mathbf{x}_0, \mathbf{x}_1, \hat{\mathbf{x}}_1, \mathbf{x}_t, \tilde{\mathbf{x}}_0, \tilde{\mathbf{x}}_1\}$, and intuitively, we have two goals: **(i)** the bridge should terminate at two distributions π_0 and π_1. Therefore all generated samples need to fulfill $\tilde{\mathbf{x}}_0 \sim \pi_0$, $\hat{\mathbf{x}}_1 \sim \pi_1$, $\tilde{\mathbf{x}}_1 \sim \pi_1$, and **(ii)** the bridge should preserve the scene details, by following EOT, which implies $\mathbf{x}_0 \approx \tilde{\mathbf{x}}_0$ and $\hat{\mathbf{x}}_1 \approx \tilde{\mathbf{x}}_1$.

To fulfill the above requirements, we propose the composition of the following objectives. First, by applying the standard IPF procedure with the following loss function

$$\mathcal{L}_{IPF} = ||\mathbf{x}_0 - \tilde{\mathbf{x}}_0||_2^2, \tag{9}$$

the method theoretically addresses both criteria; however, this ensures **(ii)** much more, as the training objective is the desired feature itself, and **(i)** is optimized only implicitly.

To enforce the **(i)** condition, we apply adversarial loss

$$\mathcal{L}_{Adv} = \left(\mathcal{D}_1^{\psi}(\tilde{\mathbf{x}}_1) - 1\right)^2, \tag{10}$$

which improves the alignment with the other domain but does not address **(ii)**.

The requirement **(ii)**, although IPF enforces it too, can be further improved with PatchNCE loss Park et al. (2020) with its details postponed to Appendix C. We refer to this loss as $\mathcal{L}_{PatchNCE}$. We use the above losses to construct the final optimization criterion for the model $\mathcal{M}_{\theta}$

$$\theta^* = \arg\min_{\theta} \mathcal{L}_{IPF} + \lambda_{Adv}\mathcal{L}_{Adv} + \lambda_{PatchNCE}\mathcal{L}_{PatchNCE}. \tag{11}$$

Lastly, to ensure that discriminators $\mathcal{D}_0^{\phi}, \mathcal{D}_1^{\psi}$ are aligned with data distributions π_0, π_1, they are trained adversarially

$$\mathcal{L}_{\mathcal{D}_1} = \left(\mathcal{D}_1^{\psi}(\hat{\mathbf{x}}_1)\right)^2 + \left(\mathcal{D}_1^{\psi}(\mathbf{x}_1) - 1\right)^2. \tag{12}$$

Here, we use $\hat{\mathbf{x}}_1$ rather than $\tilde{\mathbf{x}}_1$ as the generated sample for computing the discriminator loss because $\hat{\mathbf{x}}_1$ is sampled from the full process $\mathbb{P}_{1|0}$ and is detached from the autograd graph. Naturally, for the opposite $1 \to 0$ direction, the other discriminator would be used and optimized.

Inference During inference, given a random sample $\mathbf{x}_1 \sim \pi_1$, we transport it using the SDE from Equation 6 and an off-the-shelf sampler to obtain $\hat{\mathbf{x}}_0 \sim \mathbb{P}^{\theta}_{0|1}\pi_1$. See Algorithm 2 for more details. The resulting sample $\hat{\mathbf{x}}_0$ is expected to follow the target distribution π_0 while preserving relevant features and attributes of the source sample $\mathbf{x}_1$. For example, if π_1 and π_0 correspond to the distributions of dark and light images, respectively, the generated image may depict the same scene as the dark input image but under normal lighting conditions.

5 Experiments

5.1 Experimental Setup

We evaluate BRACE on unpaired image enhancement in three settings: low-light enhancement, dehazing, and underwater enhancement. For low-light enhancement we use LOL (Wei et al., 2018) and LOL-v2-Real (Chen et al., 2021).

Table 1: Quantitative comparison on LOL and LOL-Real. We split the methods into three groups: GAN-based, Diffusion-based, and SB-based. The best and second-best results are highlighted in **bold** and underline, respectively.

Model	NFE	LOL				LOL-Real			
		LPIPS ↓	PSNR ↑	SSIM ↑	FID ↓	LPIPS ↓	PSNR ↑	SSIM ↑	FID ↓
CycleGAN	1	0.207	19.86	**0.529**	105.13	0.326	<u>18.71</u>	0.421	150.86
CUT	1	0.230	<u>20.03</u>	<u>0.510</u>	149.25	0.357	15.89	0.228	193.93
EnlightenGAN	1	0.231	16.52	0.451	111.68	0.241	17.23	**0.455**	104.75
GDP	1000	0.338	15.91	0.380	107.61	0.355	14.54	0.315	95.92
AGLLDiff	10	<u>0.195</u>	19.20	0.502	<u>97.20</u>	**0.194**	**18.85**	<u>0.434</u>	**85.68**
ASBM	4	0.437	16.58	0.399	132.65	0.485	15.16	0.281	136.97
DDSB*	5	0.518	15.40	0.215	286.00	0.545	14.10	0.082	258.00
UNSB	4	0.276	19.05	0.417	113.31	0.320	17.99	0.297	162.35
SBFlow	1	0.424	9.37	0.067	112.57	0.298	18.25	0.393	101.88
BRACE (ours)	1	**0.193**	**20.19**	0.470	**71.18**	<u>0.238</u>	14.85	0.191	<u>87.44</u>

* We use our reimplementation due to unavailable official source code.

For dehazing we train on ITS and test on SOTS-Indoor (Li et al., 2018), and additionally use the RESIDE-6K split introduced by Song et al. (2022). For underwater enhancement we use EUVP (Islam et al., 2020) and UIEB (Li et al., 2019). All enhancement models are trained without paired supervision. When a dataset provides aligned image pairs, we discard the alignment during training and treat the degraded and clean images as two separate image collections. To check that the added endpoint losses remain compatible with unpaired translation, we also report AFHQ Cat → Wild (Choi et al., 2020) as a reference benchmark.

We compare against unpaired translation methods, Schrödinger-bridge methods, and task-specific enhancement models. The translation baselines are CycleGAN (Zhu et al., 2017) and CUT (Park et al., 2020). The bridge-based baselines are UNSB (Kim et al., 2024), ASBM (Gushchin et al., 2024), DDSB (Yi et al., 2026), and SBFlow (De Bortoli et al., 2024). Where a restoration domain has an established specialist, we include it as well: EnlightenGAN (Jiang et al., 2021), GDP (Fei et al., 2023), and AGLLDiff (Lin et al., 2025) for low-light enhancement, and DehazeSB (Lan et al., 2025b) for dehazing. This set places BRACE alongside generic unpaired translators, related bridge-based methods, and domain-specific restoration baselines.

On the enhancement datasets we report LPIPS (Zhang et al., 2018), PSNR, SSIM (Wang et al., 2004), and FID (Heusel et al., 2017). On AFHQ Cat → Wild, which has no paired targets, we report FID, MUSIQ (Ke et al., 2021), CLIP-IQA (Wang et al., 2023), and LPIPS to the input as a source-consistency score. Additional implementation details, hyperparameter settings, inference-step budgets, and per-dataset configurations are given in Appendix D and Appendix E.

5.2 Unpaired Image Enhancement

Low-light enhancement. Table 1 reports results on LOL and LOL-Real. On LOL, BRACE achieves the best LPIPS, PSNR, and FID, while remaining close to the strongest methods in SSIM. On LOL-Real, it remains competitive with the best task-specific models, reaching the second-best LPIPS and FID, while methods such as AGLLDiff and EnlightenGAN are stronger in PSNR or SSIM. The visual results (Figure 2) follow the same pattern. Methods driven mainly by adversarial alignment often brighten the image more aggressively, but can also introduce visible artifacts or unstable color reproduction, whereas more conservative bridge-based models often fail to correct illumination strongly enough. In contrast, BRACE tends to recover brightness and local detail more reliably while preserving scene geometry.

Dehazing. On SOTS-Indoor and RESIDE6k (Table 2), BRACE performs strongly on both synthetic dehazing benchmarks. It ranks second only to DehazeSB on SOTS-Indoor, and on RESIDE6k it achieves the best LPIPS together with competitive SSIM and FID. Dehazing reveals a slightly different failure mode from low-light enhancement: when supervision is too weak, the model tends to leave part of the haze veil in place, resulting in washed-out contrast and incomplete visibility

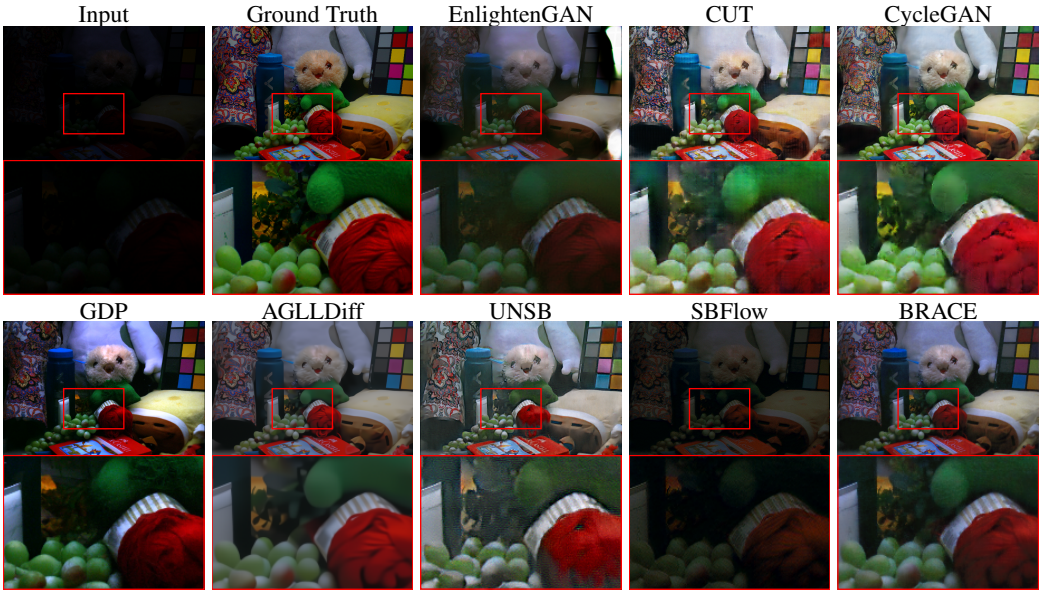


Figure 2: Qualitative comparison on the LOL test set. Red boxes indicate regions shown enlarged below each image. BRACE produces stronger illumination recovery and sharper local detail than SBFlow, while remaining visually competitive with strong unpaired enhancement baselines.

Table 2: Quantitative comparison on SOTS-Indoor and RESIDE6k datasets using four evaluation metrics: LPIPS, PSNR, SSIM, and FID. We group the methods into GAN-based and Schrödinger Bridge-based. The best and second-best results are highlighted in **bold** and underline, respectively.

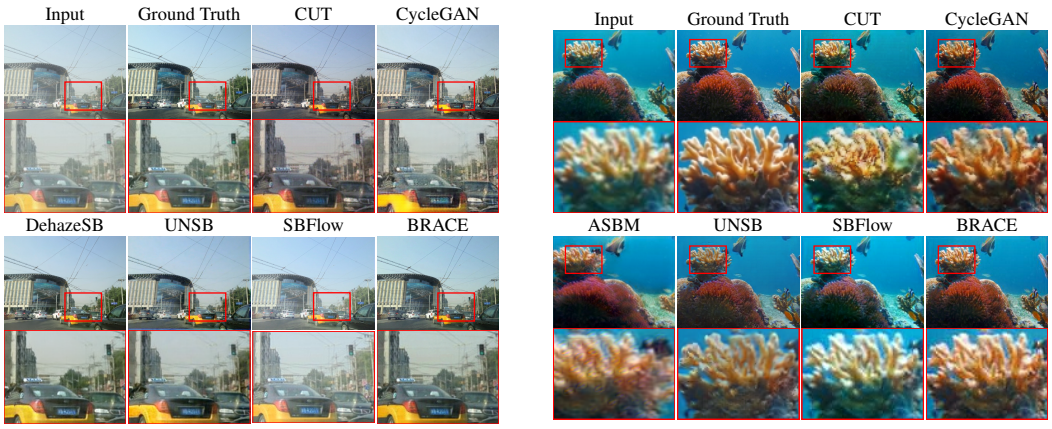
Model	NFE	SOTS-indoor				RESIDE6k			
		LPIPS ↓	PSNR ↑	SSIM ↑	FID ↓	LPIPS ↓	PSNR ↑	SSIM ↑	FID ↓
CycleGAN	1	0.123	19.87	0.673	34.04	0.140	21.48	0.708	30.84
CUT	1	0.134	18.90	0.652	34.76	0.112	20.60	0.704	25.44
DehazeSB	5	**0.082**	**23.61**	**0.767**	**19.10**	0.092	**22.67**	**0.758**	**19.19**
ASBM	4	0.280	13.01	0.296	49.37	0.247	13.80	0.330	46.52
DDSB*	5	0.275	18.24	0.539	60.00	0.353	18.11	0.530	62.50
UNSB	4	0.112	19.81	0.680	29.74	0.131	19.69	0.688	32.98
SBFlow	1	0.167	15.50	0.499	33.88	0.106	17.15	0.501	23.13
BRACE (ours)	1	0.085	19.59	0.686	20.28	**0.088**	20.17	0.725	20.46

* We use our reimplementation due to unavailable official source code.

recovery. This behavior is visible in several bridge-based baselines. By contrast, strongly adversarial methods may remove haze more aggressively but can distort colors or over-sharpen structures. BRACE generally produces a more balanced result, improving visibility without introducing equally strong visual side effects.

Underwater image enhancement. Table 3 reports results on EUVP and UIEB. On EUVP, BRACE achieves the best LPIPS, PSNR, and SSIM, together with the second-best FID. On UIEB, the ranking is less favorable: CycleGAN remains the strongest method overall, while BRACE is still clearly ahead of the other bridge-based baselines. This is consistent with the visual difficulty of underwater enhancement, where color cast removal, contrast recovery, and texture preservation must be handled simultaneously. In this setting, conservative transport often leaves the image too close to the degraded input, while stronger GAN-based corrections may overshoot the color restoration. BRACE improves the former behavior, but the remaining gap on UIEB suggests that underwater restoration remains the hardest domain for the current formulation.

Overall, the three enhancement tasks show a consistent pattern: stronger endpoint supervision improves the perceptual quality of bridge-based enhancement and usually improves distributional



(a) RESIDE6k dataset

(b) EUVP dataset

Figure 3: Qualitative comparison on the RESIDE6k and EUVP datasets. Red boxes indicate regions shown enlarged below each image.

Table 3: Quantitative comparison on the EUVP and UIEB datasets for underwater image enhancement using four evaluation metrics: LPIPS, PSNR, SSIM, and FID. We group the methods into GAN-based and Schrödinger Bridge-based. The best and second-best results are highlighted in **bold** and underline, respectively.

Model	NFE	EUVP				UIEB			
		LPIPS ↓	PSNR ↑	SSIM ↑	FID ↓	LPIPS ↓	PSNR ↑	SSIM ↑	FID ↓
CycleGAN	1	0.161	21.52	0.625	**34.60**	**0.121**	**21.49**	**0.699**	**34.43**
CUT	1	0.282	18.53	0.513	50.02	0.154	20.62	0.652	48.77
ASBM	4	0.265	20.78	0.624	68.04	0.263	18.11	0.576	75.81
DDSB*	5	0.380	19.47	0.494	71.00	0.327	17.77	0.442	111.50
UNSB	4	0.194	21.49	0.629	36.88	0.197	19.21	0.592	63.33
SBFlow	1	0.223	19.68	0.565	40.27	0.237	16.02	0.522	40.70
BRACE (ours)	1	**0.146**	**23.78**	**0.700**	35.47	0.151	18.68	0.642	39.10

* We use our reimplementation due to unavailable official source code.

alignment as well. The main gain is not that BRACE dominates every specialized baseline on every benchmark, but that a single bridge-based method remains competitive across several enhancement domains while avoiding the most common failure modes of both conservative bridge objectives and purely adversarial translation. This trend is also visible in the qualitative comparisons of Figure 3. Conservative bridge-based methods often leave residual degradation, such as incomplete dehazing or persistent color cast, whereas purely adversarial models may produce stronger corrections at the cost of less stable local structure. BRACE typically lies between these extremes, producing more decisive enhancement while preserving fine detail in enlarged regions. A further practical advantage is that these results are obtained with one inference step, whereas several competing bridge-based methods require substantially longer trajectories.

5.3 Compatibility and Ablation

Reference translation benchmark. We evaluate BRACE on AFHQ Cat → Wild as a reference benchmark for standard unpaired image translation. As shown in Table 4, BRACE substantially improves over SBFlow in FID, MUSIQ, and CLIP-IQA, indicating better target-domain alignment and perceptual quality. LPIPS to the input image is slightly worse than for SBFlow, suggesting a modest reduction in source-detail preservation. This is consistent with the role of the added endpoint supervision: it improves target realism, but can also encourage stronger departures from the input appearance. At the same time, UNSB and CUT remain stronger overall on this benchmark, with UNSB achieving the best results across all four metrics. We therefore treat AFHQ primarily as a compatibility check rather than as a main application benchmark. In this role, it shows that the

Table 4: Quantitative results on AFHQ Cat → Wild using four metrics: FID, MUSIQ, CLIP-IQA, and LPIPS with the input. The methods are grouped into GAN-based and SB-based. The best and second-best results are highlighted in **bold** and underline, respectively.

Model	NFE	FID ↓	MUSIQ ↑	CLIPIQA ↑	LPIPS ↓
CycleGAN	1	43.35	60.79	0.654	0.598
CUT	1	35.76	61.02	0.718	0.597
ASBM	4	134.07	61.48	0.562	0.478
UNSB	4	**25.49**	**64.89**	**0.727**	**0.360**
SBFlow	100	97.92	47.19	0.549	0.469
BRACE (ours)	100	66.74	60.50	0.691	0.493

Table 5: Ablation of the loss components of BRACE on LOL and SOTS-Indoor. We compare the full model with variants trained without the adversarial loss and without the PatchNCE loss. The best and second-best results are highlighted in **bold** and underline, respectively.

	LOL				SOTS-indoor			
Configuration	**LPIPS ↓**	**PSNR ↑**	**SSIM ↑**	**FID ↓**	**LPIPS ↓**	**PSNR ↑**	**SSIM ↑**	**FID ↓**
w/o Adv. Loss	0.475	8.43	0.054	146.66	0.138	14.88	0.422	25.60
w/o NCE Loss	0.213	18.82	0.455	85.90	0.086	**19.64**	0.678	21.83
BRACE	**0.193**	**20.19**	**0.470**	**71.18**	**0.085**	19.59	**0.686**	**20.28**

proposed endpoint supervision improves a bridge-based translator even outside the enhancement setting.

Loss ablation. We ablate the two additional loss components of BRACE on two representative enhancement tasks: low-light enhancement (LOL) and dehazing (SOTS-Indoor). Table 5 compares the full model with variants trained without the adversarial loss and without the PatchNCE loss. The same pattern appears in both domains. Removing the adversarial loss causes a severe drop across all metrics, showing that explicit target-domain supervision is the main factor missing from plain bridge-based endpoint regression. Removing PatchNCE leads to a smaller but still consistent degradation. The adversarial-only variant remains much stronger than the PatchNCE-only variant on both LOL and SOTS-Indoor, but it still falls short of the full model. Taken together, the ablation supports the design of BRACE: the adversarial term provides the main realism signal, while PatchNCE acts as a complementary constraint that improves local structure and perceptual fidelity.

6 Conclusion

We presented BRACE, a Schrödinger-bridge method for unpaired image enhancement that strengthens bridge-based translation with explicit supervision for target-domain realism and local correspondence. Across low-light enhancement, dehazing, and underwater restoration, the proposed supervision consistently improves the bridge-based baselines considered in the paper, while results on a reference translation benchmark show that the resulting translator remains effective beyond the enhancement setting. These findings support the central claim of the paper: when substantial appearance changes must be achieved under a strong constraint of source-scene fidelity, bridge-based translation benefits from stronger endpoint supervision than pointwise endpoint regression alone. A further practical outcome is that these gains can be obtained with very small inference-step budgets.

This study also has several limitations. Although BRACE consistently improves over the generic bridge-based baselines considered here, task-specific methods can still perform better on individual benchmarks. Performance also remains sensitive to dataset-dependent design choices, including the bridge noise scale, the inference budget, and the relative weighting of the endpoint losses. More broadly, our analysis is limited to a small set of enhancement domains and to a single Schrödinger-bridge formulation. Extending these ideas to broader restoration settings, developing more adaptive bridge parameterizations, and understanding more precisely why stronger endpoint supervision enables effective low-step inference are natural directions for future work.

AI USE STATEMENT

In this work, we used generative AI tools to aid and polish writing, including improving phrasing, grammar, and clarity, and for retrieval and discovery, including identifying potentially relevant related work. We did not use generative AI tools for research ideation, drafting sections of the paper from scratch, generating synthetic datasets, or proving mathematical claims. The remaining disclosure categories are not applicable to this work.

All AI-assisted writing was carefully reviewed and edited by the authors. Related-work suggestions obtained with AI assistance were checked through manual literature review. Any AI-assisted code edits were manually inspected and validated through the standard development and experimental workflow used in this project. We take full responsibility for the final content of this work, including all text, claims, code, and experimental results.

ETHICS STATEMENT

This work uses public benchmark datasets and does not involve human subjects or private data collection. Unpaired image enhancement can be beneficial in applications such as photography, robotics, remote sensing, and scientific imaging, but it can also be misused to alter visual evidence, generate misleading imagery, or support surveillance in low-visibility environments. Moreover, enhancement errors may hallucinate structure or suppress important details. We therefore present BRACE as a research method rather than a validated deployment system, and caution against its use in safety-critical or high-stakes applications.

REPRODUCIBILITY STATEMENT

We have made several efforts to support reproducibility. The main paper specifies the problem setting, model formulation, training objectives, evaluation metrics, and experimental protocol, while the appendix provides additional details on datasets, preprocessing, implementation choices, inference-step budgets, and dataset-specific hyperparameters. The method section describes the bridge construction, endpoint supervision terms, and sampling procedure used at training and inference time, and the experiments section reports the benchmark settings and baselines used in each evaluation. The full codebase, configuration files, and scripts required to reproduce the reported experiments will be released with the final version of the paper.

REFERENCES

Yochai Blau and Tomer Michaeli. The perception-distortion tradeoff. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6228–6237. IEEE, 2018.

Wei Chen, Wangjing Wang, Wenhan Yang, and Jiaying Liu. Lol-v2: A new benchmark for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. Stargan v2: Diverse image synthesis for multiple domains. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8185–8194. IEEE, 2020.

Joseph Paul Cohen, Margaux Luck, and Sina Honari. Distribution matching losses can hallucinate features in medical image translation. In *International conference on medical image computing and computer-assisted intervention*, pp. 529–536. Springer, 2018.

Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion schrödinger bridge with applications to score-based generative modeling. *Advances in neural information processing systems*, 34:17695–17709, 2021.

Valentin De Bortoli, Iryna Korshunova, Andriy Mnih, and Arnaud Doucet. Schrodinger bridge flow for unpaired data translation. *Advances in Neural Information Processing Systems*, 37:103384–103441, 2024.

Deniz Engin, Anil Genç, and Hazim Kemal Ekenel. Cycle-dehaze: Enhanced cyclegan for single image dehazing. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 938–9388. IEEE, 2018.

Cameron Fabbri, Md Jahidul Islam, and Junaed Sattar. Enhancing underwater imagery using generative adversarial networks. In *2018 IEEE international conference on robotics and automation (ICRA)*, pp. 7159–7165. IEEE, 2018.

Ben Fei, Zhaoyang Lyu, Liang Pan, Junzhe Zhang, Weidong Yang, Tianyue Luo, Bo Zhang, and Bo Dai. Generative diffusion prior for unified image restoration and enhancement. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9935–9946. IEEE, 2023.

Nikita Gushchin, Daniil Selikhanovych, Sergei Kholkin, Evgeny Burnaev, and Alexander Korotin. Adversarial schrödinger bridge matching. *Advances in Neural Information Processing Systems*, 37:89612–89651, 2024.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.

Md Jahidul Islam, Youya Xia, and Junaed Sattar. Fast underwater image enhancement for improved visual perception. *IEEE robotics and automation letters*, 5(2):3227–3234, 2020.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *2017 IEEE conference on computer vision and pattern recognition (CVPR)*, pp. 5967–5976. Ieee, 2017.

Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *IEEE transactions on image processing*, 30:2340–2349, 2021.

Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 5128–5137. IEEE, 2021.

Sergei Kholkin, Grigoriy Ksenofontov, David Li, Nikita Kornilov, Nikita Gushchin, Alexandra Suvorikova, Alexey Kroshnin, Evgeny Burnaev, and Aleksandr Korotin. Diffusion & adversarial schrödinger bridges via iterative proportional markovian fitting. In *International Conference on Learning Representations*, volume 2026, pp. 19192–19235, 2026.

Beomsu Kim, Gihyun Kwon, Kwanyoung Kim, and Jong Chul Ye. Unpaired image-to-image translation via neural schrödinger bridge. In *International Conference on Learning Representations*, volume 2024, pp. 19312–19331, 2024.

Yunwei Lan, Zhigao Cui, Chang Liu, Jialun Peng, Nian Wang, Xin Luo, and Dong Liu. Exploiting diffusion prior for real-world image dehazing with unpaired training. In *Proceedings of the AAAI conference on artificial intelligence*, volume 39, pp. 4455–4463, 2025a.

Yunwei Lan, Zhigao Cui, Xin Luo, Chang Liu, Nian Wang, Menglin Zhang, Yanzhao Su, and Dong Liu. When schrödinger bridge meets real-world image dehazing with unpaired training. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8756–8765. IEEE, 2025b.

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

Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE transactions on image processing*, 28 (1):492–505, 2018.

Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. An underwater image enhancement benchmark dataset and beyond. *IEEE transactions on image processing*, 29:4376–4389, 2019.

Yunlong Lin, Tian Ye, Sixiang Chen, Zhenqi Fu, Yingying Wang, Wenhao Chai, Zhaohu Xing, Wenxue Li, Lei Zhu, and Xinghao Ding. Aglldiff: Guiding diffusion models towards unsupervised training-free real-world low-light image enhancement. *Proceedings of the AAAI conference on artificial intelligence*, 39(5):5307–5315, 2025.

Ming-Yu Liu, Thomas Breuel, and Jan Kautz. Unsupervised image-to-image translation networks. *Advances in neural information processing systems*, 30, 2017.

Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pp. 2794–2802, 2017.

Taesung Park, Alexei A Efros, Richard Zhang, and Jun-Yan Zhu. Contrastive learning for unpaired image-to-image translation. In *European conference on computer vision*, pp. 319–345. Springer, 2020.

Tatiana Tais Schein, Gustavo Pereira De Almeira, Stephanie Loi Briao, Rodrigo Andrade De Bem, Felipe Gomes De Oliveira, and Paulo LJ Drews-Jr. Udbe: Unsupervised diffusion-based brightness enhancement in underwater images. In *2024 International Conference on Machine Learning and Applications (ICMLA)*, pp. 664–670. IEEE, 2024.

Yuyang Shi, Valentin De Bortoli, Andrew Campbell, and Arnaud Doucet. Diffusion schrödinger bridge matching. *Advances in neural information processing systems*, 36:62183–62223, 2023.

Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. *arXiv preprint arXiv:2204.03883*, 2022.

Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. *Proceedings of the AAAI conference on artificial intelligence*, 37(2):2555–2563, 2023.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018.

Jingjun Yi, Qi Bi, Hao Zheng, Huimin Huang, Yixian Shen, Haolan Zhan, Wei Ji, Yawen Huang, Yuexiang Li, Xian Wu, et al. Degradation-aware dynamic schrödinger bridge for unpaired image restoration. *Advances in Neural Information Processing Systems*, 38:2415–2439, 2026.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 586–595, 2018.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *2017 IEEE international conference on computer vision (ICCV)*, pp. 2242–2251. Ieee, 2017.

Appendix

A Detailed Training and Inference Procedures

This appendix gives the full training and inference procedures in algorithmic form. Algorithm 1 summarizes the bidirectional training loop. At each iteration, one translation direction is active, a synthetic endpoint is generated with K reverse steps, and the generator and the discriminator of the active target domain are updated. Algorithm 2 gives inference in the $1 \to 0$ direction. The continuous bridge coordinate $t \in [0, 1]$ is distinct from the reverse-step budgets: K denotes the number of reverse steps used to generate synthetic endpoints during training, while N denotes the number of reverse steps used at inference.

Algorithm 1 BRACE training

Input: datasets π_0 and π_1, reference path measure $\mathbb{Q}$, number of training steps N_{train}, bridge process $\mathbb{P}^\theta$ driven by model $\mathcal{M}$ with parameters θ, discriminators $\mathcal{D}_0, \mathcal{D}_1$ parameters ϕ, ψ, training-time simulation step budget K, loss weights $\lambda_{\text{adv}}, \lambda_{\text{nce}}$, any off-the-shelf sampler Sample
for $n \in \{1, \dots, N_{\text{train}}\}$ **do**
  Sample $(\mathbf{x}_0, \mathbf{x}_1) \sim (\pi_0 \otimes \pi_1)$
  if n is odd **then**
    $\hat{\mathbf{x}}_0 \leftarrow \mathrm{sg}(\mathrm{Sample}_K(\mathbb{P}^\theta_{0|1}, \mathbf{x}_1))$
    Sample $t \sim \mathcal{U}(0, 1)$
    $\mathbf{x}_t \leftarrow \mathbb{Q}_{t|0,1}(\hat{\mathbf{x}}_0, \mathbf{x}_1)$
    $\tilde{\mathbf{x}}_0 \leftarrow \mathcal{M}_\theta(\mathbf{x}_t, t, 0), \quad \tilde{\mathbf{x}}_1 \leftarrow \mathcal{M}_\theta(\mathbf{x}_t, t, 1)$
    $\mathcal{L}_G \leftarrow \|\tilde{\mathbf{x}}_1 - \mathbf{x}_1\|_2^2 + \lambda_{\text{adv}}(\mathcal{D}_0^\phi(\tilde{\mathbf{x}}_0) - 1)^2 + \lambda_{\text{nce}}\mathcal{L}_{\text{PatchNCE}}(\tilde{\mathbf{x}}_0, \mathbf{x}_1)$
    Update θ with one gradient step on $\mathcal{L}_G$
    $\mathcal{L}_{\mathcal{D}_0} \leftarrow (\mathcal{D}_0^\phi(\hat{\mathbf{x}}_0))^2 + (\mathcal{D}_0^\phi(\mathbf{x}_0) - 1)^2$.
    Update ϕ with one gradient step on $\mathcal{L}_{\mathcal{D}_0}$
  else
    $\hat{\mathbf{x}}_1 \leftarrow \mathrm{sg}(\mathrm{Sample}_K(\mathbb{P}^\theta_{1|0}, \mathbf{x}_0))$
    Sample $t \sim \mathcal{U}(0, 1)$
    $\mathbf{x}_t \leftarrow \mathbb{Q}_{t|0,1}(\mathbf{x}_0, \hat{\mathbf{x}}_1)$
    $\tilde{\mathbf{x}}_0 \leftarrow \mathcal{M}_\theta(\mathbf{x}_t, t, 0), \quad \tilde{\mathbf{x}}_1 \leftarrow \mathcal{M}_\theta(\mathbf{x}_t, t, 1)$
    $\mathcal{L}_G \leftarrow \|\tilde{\mathbf{x}}_0 - \mathbf{x}_0\|_2^2 + \lambda_{\text{adv}}(\mathcal{D}_1^\psi(\tilde{\mathbf{x}}_1) - 1)^2 + \lambda_{\text{nce}}\mathcal{L}_{\text{PatchNCE}}(\tilde{\mathbf{x}}_1, \mathbf{x}_0)$
    Update θ with one gradient step on $\mathcal{L}_G$
    $\mathcal{L}_{\mathcal{D}_1} \leftarrow (\mathcal{D}_1^\psi(\hat{\mathbf{x}}_1))^2 + (\mathcal{D}_1^\psi(\mathbf{x}_1) - 1)^2$.
    Update ψ with one gradient step on $\mathcal{L}_{\mathcal{D}_1}$
  end if
end for
Output: trained generator and discriminator parameters (θ, ϕ, ψ)

Algorithm 2 Inference in the $1 \to 0$ direction

Input: degraded input image $\mathbf{x}_1$, bridge process $\mathbb{P}^\theta$ driven by generator $\mathcal{M}$ with parameters θ, inference reverse-step budget N, reverse sampler $\mathcal{S}$
Set $1 = t_0 > t_1 > \cdots > t_N = 0$ and initialize $x^{(0)} \leftarrow x_1$
for $n \in \{0, \dots, N-1\}$ **do**
  $\tilde{x}_0^{(n)} \leftarrow \mathcal{M}_\theta(x^{(n)}, t_n, 0)$
  if $n < N - 1$ **then**
    $x^{(n+1)} \leftarrow \mathcal{S}(\mathbb{P}^\theta, x^{(n)}, \tilde{x}_0^{(n)}, t_n, t_{n+1})$
  end if
end for
Output: translated clean image $\tilde{x}_0^{(N-1)}$

B BROWNIAN REVERSE SAMPLERS

This section records the reverse sampling updates used in the Brownian setting. Let $t > s$ be two consecutive points on the reverse-time grid, let x_t denote the current bridge state, and let $\hat{x}_{\mathrm{tgt}}$ denote the target-endpoint prediction produced by the conditioned network. The bridge temperature is $\tau > 0$.

Ancestral sampler. For the Brownian bridge, the reverse transition is given by the conditional posterior

$$x_s \sim q(x_s \mid x_t, \hat{x}_{\mathrm{tgt}}), \tag{13}$$

with mean and variance

$$\mu_{t\to s} = \frac{s}{t} x_t + \left(1 - \frac{s}{t}\right) \hat{x}_{\mathrm{tgt}}, \qquad \sigma^2_{t\to s} = \tau^2 \frac{s(t-s)}{t}. \tag{14}$$

Accordingly, one reverse step is

$$x_s = \mu_{t\to s} + \tau \sqrt{\frac{s(t-s)}{t}}\, \epsilon, \qquad \epsilon \sim \mathcal{N}(0, I). \tag{15}$$

Euler–Maruyama sampler. Under the Brownian configuration, the reverse SDE uses drift

$$b(x_t, t; \hat{x}_{\mathrm{tgt}}) = \frac{x_t - \hat{x}_{\mathrm{tgt}}}{t}, \tag{16}$$

and constant diffusion coefficient τ. The Euler–Maruyama update from t to s is therefore

$$x_s = x_t + \frac{x_t - \hat{x}_{\mathrm{tgt}}}{t}(s - t) + \tau \sqrt{t - s}\, \epsilon, \qquad \epsilon \sim \mathcal{N}(0, I). \tag{17}$$

In both cases, training-time synthetic endpoint generation uses a reverse-step budget K, while inference uses a possibly different budget N. These reverse-step budgets should not be confused with the continuous bridge coordinate t.

C PATCH-LEVEL CONTRASTIVE OBJECTIVE

The patch-level contrastive term follows PatchNCE (Park et al., 2020) and is applied to intermediate features of the conditioned U-Net. In the active direction, let $\tilde{x}_{\mathrm{tgt}}$ denote the predicted target endpoint and x_{src} the source image. For each layer $l \in \mathcal{L}$, features are extracted as $\phi_l(\tilde{x}_{\mathrm{tgt}})$ and $\phi_l(x_{\mathrm{src}})$. A shared set of spatial indices I_l is sampled, and the corresponding patch features are projected by a layer-specific MLP h_l and ℓ_2-normalized:

$$q_i^{(l)} = \frac{h_l(\phi_l(\tilde{x}_{\mathrm{tgt}})_i)}{\|h_l(\phi_l(\tilde{x}_{\mathrm{tgt}})_i)\|_2}, \qquad k_i^{(l)} = \frac{h_l(\phi_l(x_{\mathrm{src}})_i)}{\|h_l(\phi_l(x_{\mathrm{src}})_i)\|_2}, \qquad i \in I_l. \tag{18}$$

Key features are detached. The positive pair is given by the same spatial index in query and key, and all other sampled key patches in the batch act as negatives. The layer-wise loss is

$$\mathcal{L}^{(l)}_{\mathrm{PatchNCE}} = -\frac{1}{|I_l|} \sum_{i \in I_l} \log \frac{\exp\Big(\langle q_i^{(l)}, k_i^{(l)} \rangle / \kappa\Big)}{\sum_j \exp\Big(\langle q_i^{(l)}, k_j^{(l)} \rangle / \kappa\Big)}, \tag{19}$$

and the final objective averages over layers,

$$\mathcal{L}_{\mathrm{PatchNCE}} = \frac{1}{|\mathcal{L}|} \sum_{l \in \mathcal{L}} \mathcal{L}^{(l)}_{\mathrm{PatchNCE}}. \tag{20}$$

Relative to CUT, the only substantive difference is that features are taken from the conditioned U-Net defining the Schrödinger-bridge translator.

D EXPERIMENTAL PROTOCOL

We evaluate all methods on seven benchmarks spanning low-light enhancement, dehazing, underwater image enhancement, and unpaired image translation: LOL and LOL-Real for low-light enhancement, AFHQ Cat $\rightarrow$ Wild for translation, the standard RESIDE indoor protocol, where models are trained on ITS and evaluated on SOTS-Indoor and RESIDE-6K for dehazing, and EUVP and UIEB for underwater enhancement. In every benchmark, training is performed in the unpaired setting, with source and target images sampled independently from their respective domains. For LOL and LOL-Real, images are processed at their native resolution and evaluation follows the standard paired protocol against the normal-light reference image. For ITS/SOTS-Indoor, images are likewise processed at their native 620×460 resolution. For RESIDE-6K, training images are used at 400×400, while evaluation is performed on the original test resolution. For EUVP and UIEB, images are center-cropped and resized to 256×256, and evaluation follows the standard paired protocol against the reference image. For AFHQ, all images are resized to 256×256. Evaluation on AFHQ is unpaired: FID is computed against the target wild domain, while LPIPS is computed with respect to the input image to measure source consistency under translation.

To ensure a fair comparison, all trainable baselines are run with the same batch size and the same number of generator optimization steps within each benchmark. For LOL and LOL-Real, we use batch size 2 and train for 100k generator steps. For AFHQ, ITS/SOTS-Indoor, RESIDE-6K, EUVP, and UIEB, we train for 150k generator steps, using batch size 8 for AFHQ, 1 for ITS/SOTS-Indoor, 2 for RESIDE-6K, and 4 for both EUVP and UIEB. In practice, these budgets were sufficient for all trainable methods to reach stable validation performance. Models are implemented from official code when available and otherwise from the published paper description. In particular, UNSB, SBFlow, DehazeSB, DDSB, ASBM, and BRACE are evaluated with the same 30M U-Net backbone, rather than their original architectures, in order to isolate the effect of the training objective from differences in model capacity. Checkpoints are selected using the same validation criterion across trainable methods: the best validation LPIPS model is used for low-light enhancement, dehazing, and underwater enhancement, whereas the best validation FID model is used for AFHQ. For SBFlow and BRACE, the inference-step budget is selected per dataset from a small candidate set of NFEs using validation performance and then fixed for test evaluation. When no official validation split is available, we hold out a small subset of training batches for model selection. Other iterative baselines use the step counts specified in their original papers or released implementations, as their discrete-time inference grids are tied to training and cannot be changed at test time without retraining. Methods that do not require training, such as GDP and AGLLDiff, are evaluated directly from their released implementations.

All reported results are obtained from a single run with random seed 42 on one NVIDIA A100 GPU with 40 GB of VRAM. For BRACE, the reported training runs required approximately 24 hours on LOL, 26 hours on LOL-Real, 32 hours on AFHQ Cat $\rightarrow$ Wild, 30 hours on ITS/SOTS-Indoor, 27 hours on RESIDE-6K, 25 hours on EUVP, and 20 hours on UIEB. Trainable baselines were run on the same hardware and under the same benchmark-specific step budgets, although wall-clock time varied across implementations. The full research project additionally involved preliminary tuning and unsuccessful pilot runs, which are not included in these reported training times. The quantitative metrics reported in the main paper are LPIPS, SSIM, PSNR, and FID for LOL, LOL-Real, ITS/SOTS-Indoor, RESIDE-6K, EUVP, and UIEB, and FID, MUSIQ, CLIP-IQA, and LPIPS for AFHQ. Additional implementation details and benchmark-specific hyperparameters are provided in the following section.

E IMPLEMENTATION DETAILS OF BRACE

E.1 ARCHITECTURE AND OPTIMIZATION

BRACE uses a single class-conditioned U-Net to predict either bridge endpoint, together with two domain-specific discriminators and PatchNCE projectors trained jointly with the generator. The network architecture itself is shared across all benchmarks, while the image resolution, batch size, training budget, and checkpoint selection criterion vary by dataset. The full architecture and optimization details are summarized in Tables 6, 7, and 8.

Table 6: Architecture of BRACE.

Module	Item	Value
Generator	Backbone	U-Net (Diffusers)
Generator	Widths	64-64-128-128-256-256
Generator	ResBlocks / scale	2
Generator	Attention	yes (head dim 64)
Generator	Time conditioning	sinusoidal
Generator	Domain conditioning	learned embedding (2 classes)
Discriminator	Type	2× multi-scale PatchGAN
Discriminator	Scales	2
Discriminator	Channels	64-128-256-512
Discriminator	Normalization	instance norm
PatchNCE	Feature layers	conv_in, down0, down1, down2, mid
PatchNCE	Patches / layer	256
PatchNCE	Projector	$d \to 256 \to 256$
PatchNCE	Temperature	0.07

Table 7: Optimization settings shared across all BRACE experiments.

Setting	Value
Generator optimizer	AdamW
Discriminator optimizer	Adam
Generator learning rate	1×10^{-4}
Discriminator learning rate	1×10^{-5}
Generator betas	$(0.9, 0.999)$
Discriminator betas	$(0.5, 0.999)$
LR schedule	cosine annealing
Min LR (gen. / disc.)	10^{-7} / 10^{-6}
EMA	off
Gradient clipping	off
Precision	fp32
Seed	42

For LOL and LOL-Real, images are processed at their native resolution; internally, the U-Net pads each input to a multiple of 2^6 and crops the output back to the original size. ITS/SOTS-Indoor is likewise processed at native resolution, whereas RESIDE6K uses 400×400 training images and is evaluated at the original test resolution. For EUVP and UIEB, images are center-cropped and resized to 256×256. For AFHQ, all images are resized from 512×512 to 256×256 for both training and evaluation.

E.2 Sampling and Loss Configuration

BRACE uses the same bridge-based training procedure across all benchmarks. A synthetic endpoint is first generated by the current translator, treated as stop-gradient, and then paired with a real sample from the opposite domain to define a Brownian bridge state. Training alternates the two translation directions across steps and updates only the discriminator of the active target domain. The shared method settings and benchmark-specific hyperparameters are summarized in Tables 9 and 10.

For low-light enhancement, dehazing, and underwater enhancement, inference with NFE$= 1$ corresponds to a single endpoint prediction at $t = 1$, without running a multi-step reverse chain. In all three restoration settings, however, the synthetic endpoint used during training is still generated with $K = 5$ Euler–Maruyama steps. For AFHQ Cat $\to$ Wild, inference uses a 100-step ancestral sampler, reflecting the stronger appearance change required by the translation setting.

Table 8: Benchmark-specific training settings for BRACE. Generator steps denote the number of generator updates.

Benchmark	Train Image size	Test Image size	Batch size	Generator steps	Checkpoint metric
LOL / LOL-Real	native	native	2	100k	LPIPS
AFHQ Cat → Wild	256×256	256×256	8	150k	FID
ITS-SOTS	native	native	1	150k	LPIPS
RESIDE6K	400×400	native	2	150k	LPIPS
EUVP	256×256	256×256	4	150k	LPIPS
UIEB	256×256	256×256	4	150k	LPIPS

Table 9: Shared sampling and loss settings of BRACE.

Item	Value
Bridge	Brownian
Bridge noise	$\sigma_t = \tau\sqrt{t(1-t)}$
Synthetic endpoint steps K	5
Direction schedule	alternating $0 \to 1$ / $1 \to 0$
Synthetic endpoint	stop-gradient
GAN loss	LSGAN
Active discriminators	one per domain
Discriminator update	active target domain only
Final output clipping	yes, to $[-1, 1]$

F Sensitivity to the Inference-Step Budget

Figures 4, 5, and 6 show a consistent pattern. First, both methods achieve their strongest results at very small inference budgets in almost all metrics, indicating that long reverse trajectories are not beneficial in this enhancement regime. Second, BRACE outperforms SBFlow at every tested budget on all three benchmarks, so its advantage is not tied to a single operating point. On SOTS-Indoor and EUVP, SBFlow degrades noticeably in LPIPS and FID as NFE increases, whereas BRACE degrades more gradually. On LOL, SBFlow fails to correct illumination at any tested budget (PSNR remains near 9 dB), so the gap there is dominated by the weakness of the underlying objective rather than by the sampler alone.

We also note that performance does not improve at the training-time synthetic-generation budget $K = 5$. A plausible explanation is that both methods are trained primarily through endpoint-level supervision: the bridge regression acts on single endpoint predictions, and in BRACE the adversarial and PatchNCE terms are also applied to direct endpoint estimates rather than to composed multi-step trajectories. As a result, longer rollouts are not explicitly optimized during training, and prediction errors can accumulate along the reverse chain.

G Assets, Sources, and Licenses

We use only existing public datasets, libraries, and baseline implementations.

- **LOL / LOL-Real**: We evaluate on the original LOw-Light (LOL) dataset Wei et al. (2018) and the LOL-v2 benchmark, including LOL-Real Chen et al. (2021). Both datasets are publicly available for academic research use; we cite the original sources and follow their publicly stated usage conditions.
- **ITS / SOTS-Indoor and RESIDE-6K**: Our dehazing experiments use public subsets and protocols derived from the RESIDE benchmark of Li et al. (2018). In particular, the ITS/SOTS-Indoor protocol follows the standard indoor RESIDE setting, while the RESIDE-6K split follows the protocol popularized by Song et al. (2022). These data are publicly released for research use through the corresponding benchmark resources, and we use them only for model training and evaluation within this work.

Table 10: Benchmark-specific sampling and loss hyperparameters for BRACE. When two datasets are grouped in one column, values are reported in the same order as in the column header.

Setting	LLIE (LOL / LOL-Real)	Dehazing (ITS-SOTS / RESIDE6K)	Underwater (EUVP / UIEB)	AFHQ Cat → Wild
Temperature τ	0.01/0.01	0.05/0.01	0.01/0.01	3.0
Sampler	Euler–Maruyama	Euler–Maruyama	Euler–Maruyama	Ancestral
Inference NFE	1/1	1/1	1/1	100
λ_{mse}	1.0/1.0	1.0/1.0	1.0/1.0	1.0
λ_{adv}	0.002/0.002	0.002/0.002	0.001/0.0005	0.055
λ_{nce}	0.0005/0.0005	0.0005/0.0005	0.0002/0.0004	0.0034

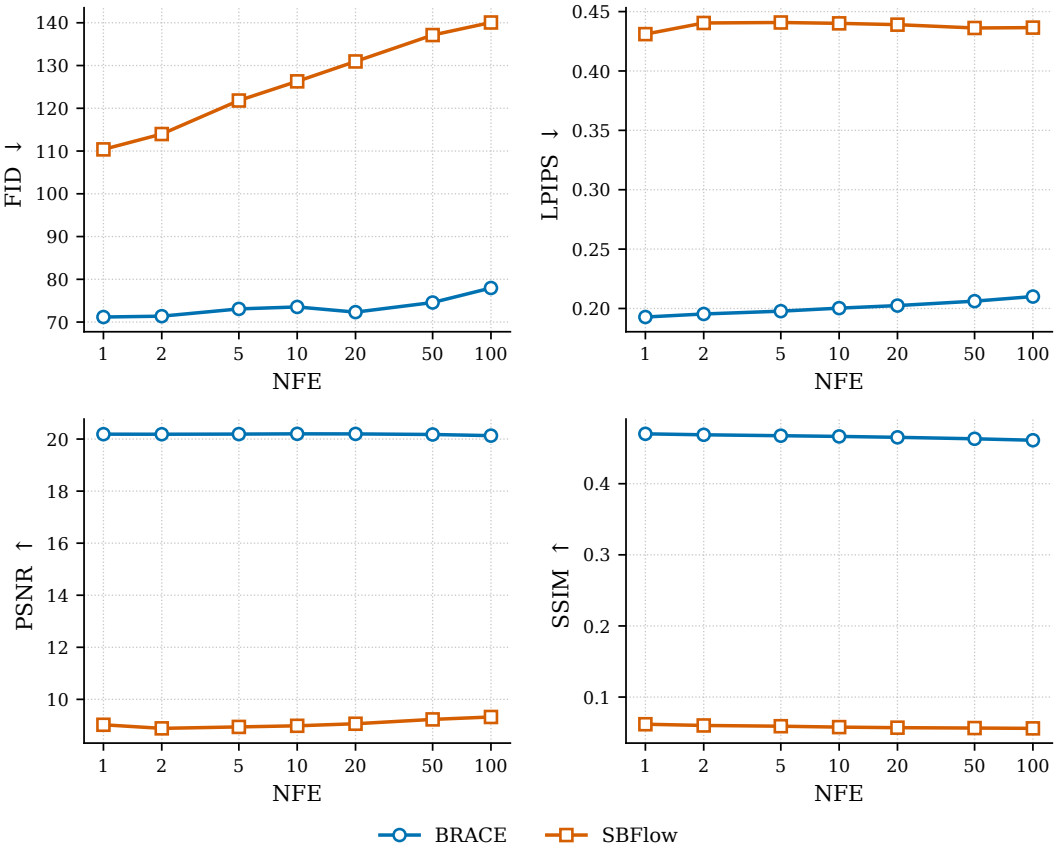


Figure 4: Sensitivity to the inference-step budget on LOL. We sweep the number of function evaluations (NFE) for BRACE and SBFlow while keeping training fixed. Both methods achieve their strongest results at very small inference budgets, but BRACE remains consistently better across the full NFE range. Increasing NFE particularly degrades SBFlow in FID and LPIPS, whereas BRACE remains comparatively stable.

- **EUVP**: We evaluate underwater enhancement on the EUVP dataset introduced by Islam et al. (2020). EUVP is publicly released for research use and contains both paired and unpaired underwater image collections. We use the dataset under its publicly stated academic-use conditions and cite the original paper and dataset source.

- **UIEB**: We also evaluate on the Underwater Image Enhancement Benchmark (UIEB) introduced by Li et al. (2019). UIEB is publicly available for research use and is widely used as a standard underwater enhancement benchmark. We use the published benchmark split and cite the original source.

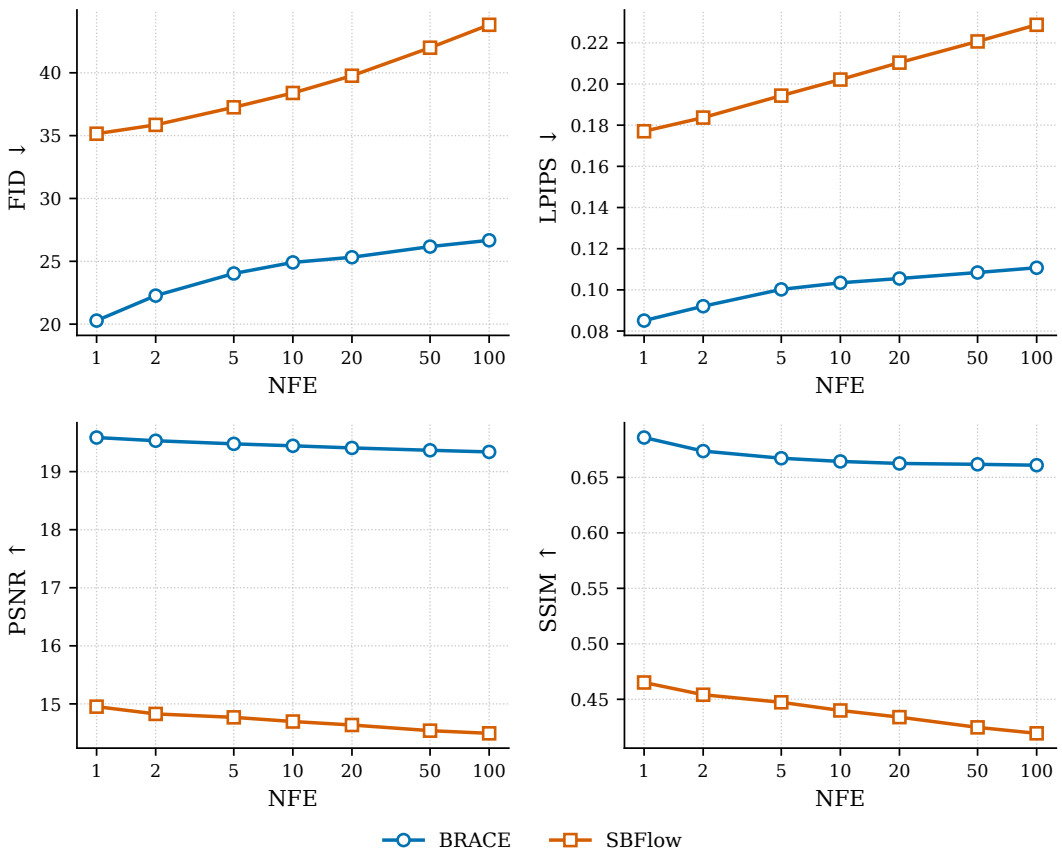


Figure 5: Sensitivity to the inference-step budget on SOTS-Indoor. As on LOL, the best results for both methods occur at small NFEs. Across all tested budgets, BRACE retains a clear advantage over SBFlow in perceptual, distortion-based, and distributional metrics, indicating that the stronger endpoint supervision improves the translator itself rather than only a specific sampling regime.

- **AFHQ**: We evaluate on the Animal Faces-HQ (AFHQ) dataset introduced by Choi et al. (2020). The dataset is publicly available and distributed under the CC BY-NC 4.0 license.
- **Baseline methods**: CycleGAN, CUT, EnlightenGAN, GDP, AgLLDiff, UNSB, DehazeSB, DDSB, ASBM and SBFlow are reimplemented by the authors. We cite the original papers for all baselines and, when available, consult official public repositories as implementation references. No third-party baseline code is redistributed as part of this work.
- **Core software libraries**: Our implementation relies on public open-source machine learning libraries, including PyTorch, Lightning, Diffusers, Transformers, Accelerate, TorchMetrics, and Hydra. We use these libraries under their respective published licenses as provided by their official distributions.

To the best of our knowledge, all assets were used in accordance with their intended research-use conditions and applicable licenses.

H Additional Visual Results

H.1 Low-Light Image Enhancement

We provide additional qualitative comparisons for low-light image enhancement on LOL and LOL-Real, Figure 7 and Figure 8. These examples illustrate differences in illumination recovery, color rendering, and local detail preservation across methods.

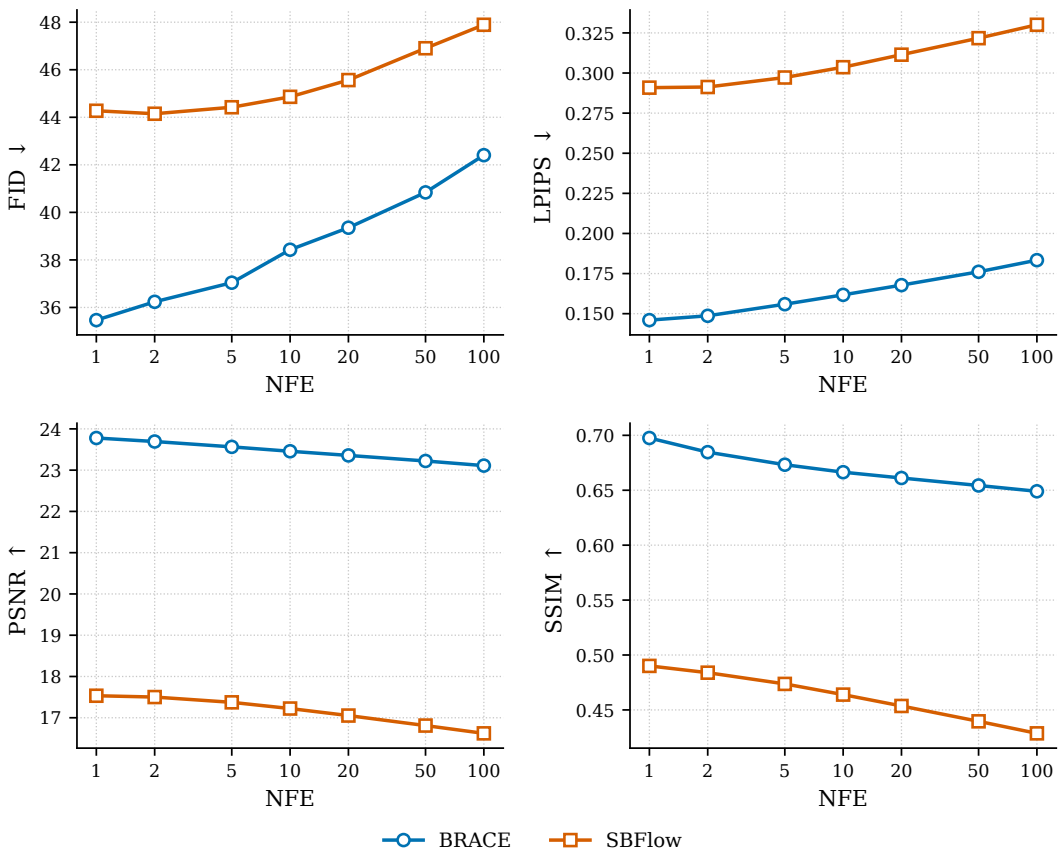


Figure 6: Sensitivity to the inference-step budget on EUVP. Both methods favor short inference trajectories in the underwater enhancement setting, but BRACE remains consistently stronger than SBFlow throughout the sweep. The gap is especially pronounced in PSNR, SSIM, and LPIPS, while FID also remains lower for BRACE at every tested NFE.

H.2 Dehazing

We provide additional qualitative comparisons for image dehazing on SOTS-Indoor and RESIDE6k in Figures 9, 10, and 11. These examples highlight differences in haze removal, contrast recovery, color restoration, and preservation of fine scene structure across methods.

H.3 Underwater Image Enhancement

We provide additional qualitative comparisons for underwater image enhancement on EUVP and UIEB in Figures 12 and 13. These examples illustrate differences in color correction, visibility enhancement, contrast recovery, and preservation of textures and object boundaries across methods.

H.4 Unpaired Image Translation

We provide additional qualitative comparisons for AFHQ Cat $\rightarrow$ Wild translation, Figures 14, 15 and 16. These examples highlight target-domain realism, texture synthesis, and preservation of source structure.

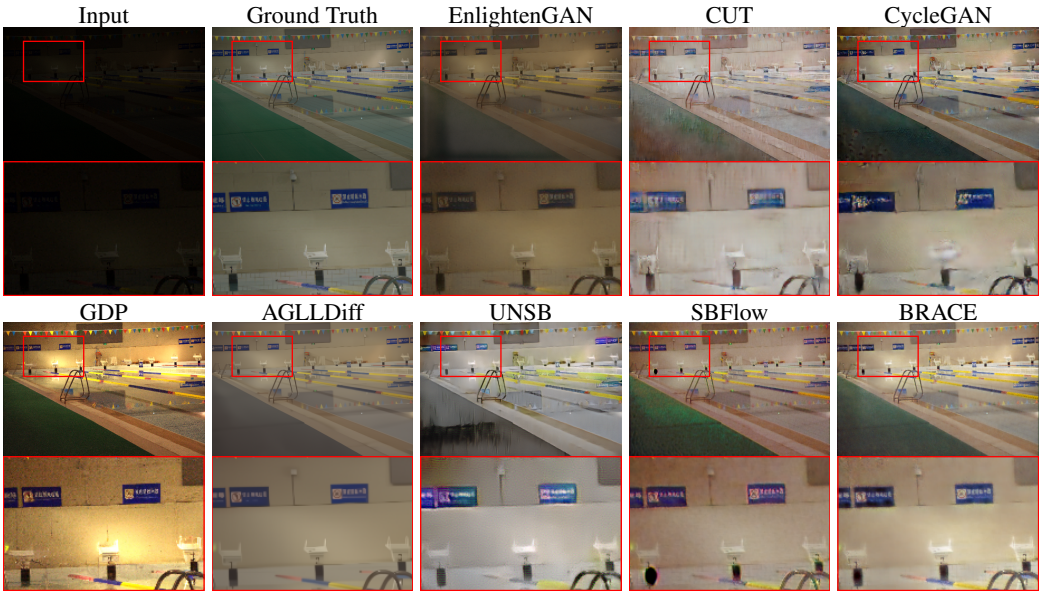


Figure 7: Qualitative comparison on the LOL Real test set. Red boxes indicate regions shown enlarged below each image.

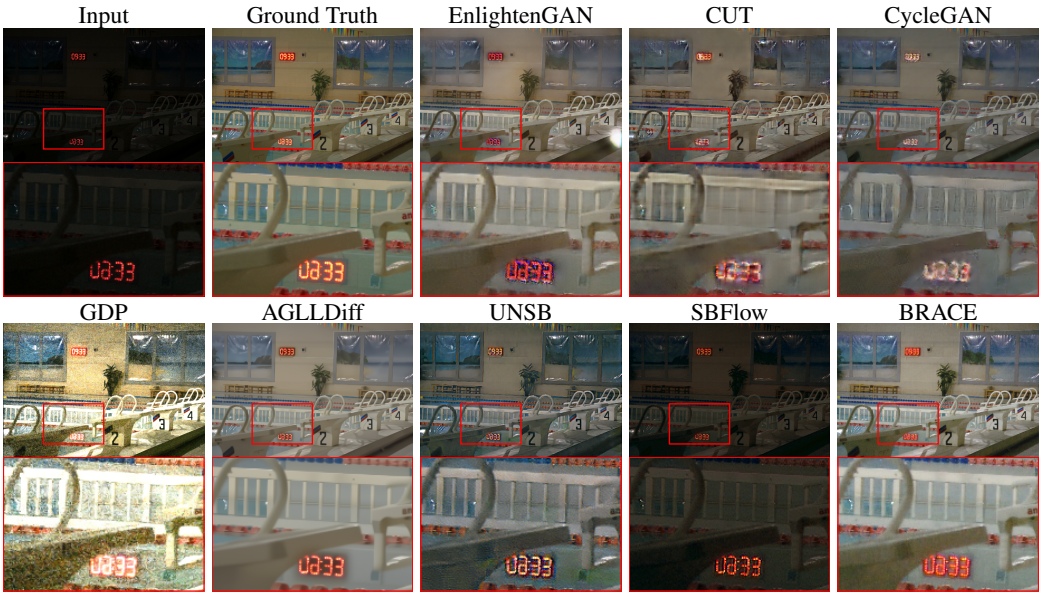


Figure 8: Additional qualitative comparison on LOL. Red boxes indicate regions shown enlarged below each image.

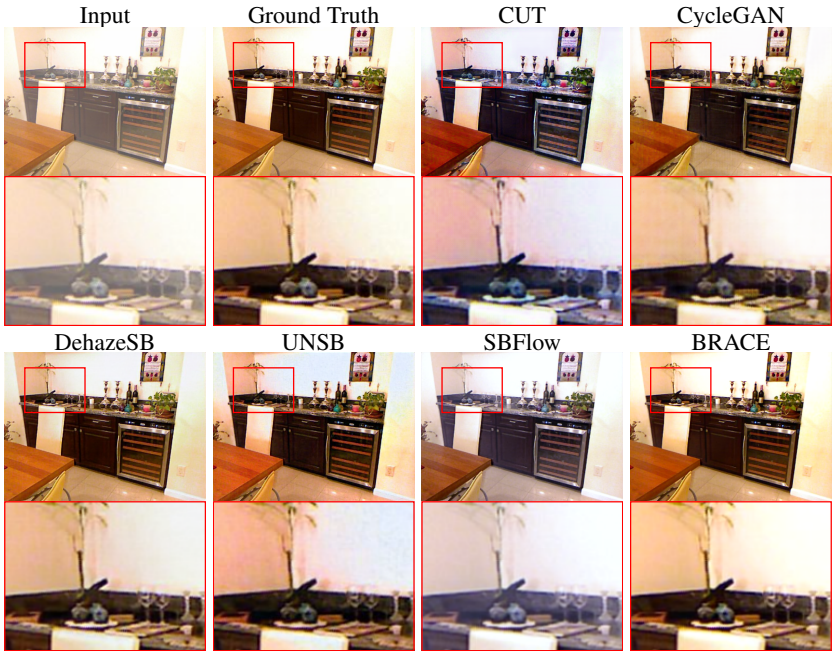


Figure 9: Qualitative comparison on the SOTS-Indoor dataset. Red boxes indicate regions shown enlarged below each image.

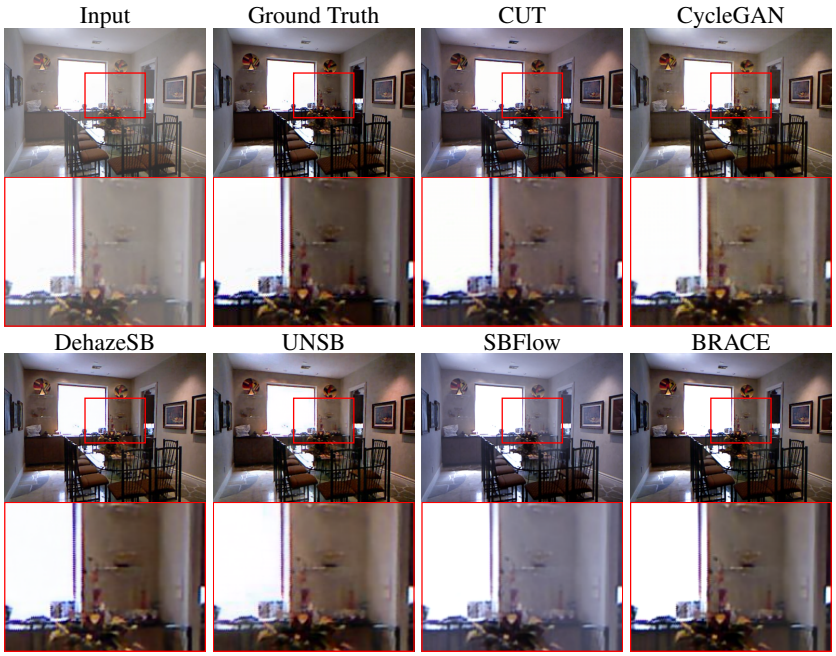


Figure 10: Qualitative comparison on the SOTS-Indoor dataset. Red boxes indicate regions shown enlarged below each image.

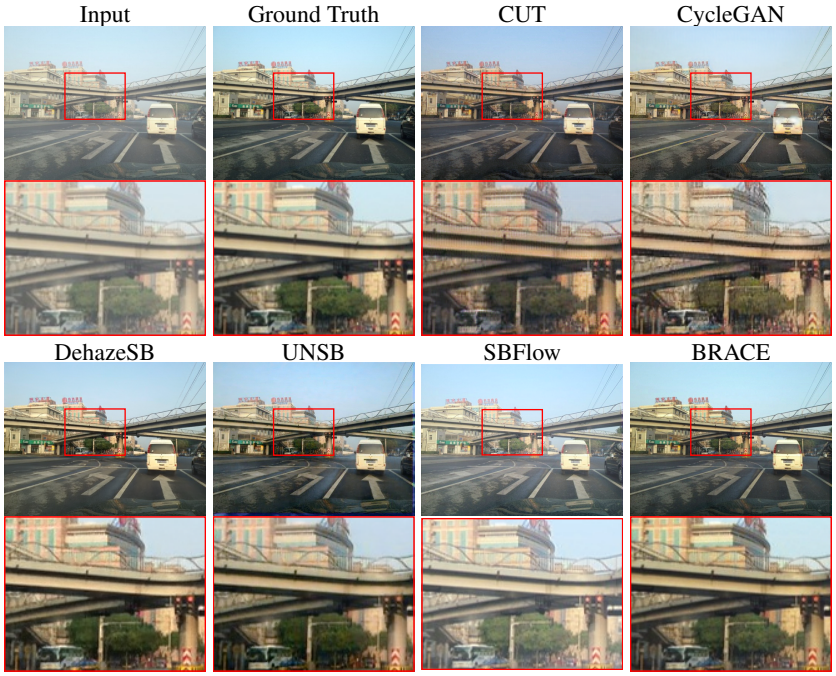


Figure 11: Qualitative comparison on the Reside6k dataset. Red boxes indicate regions shown enlarged below each image.

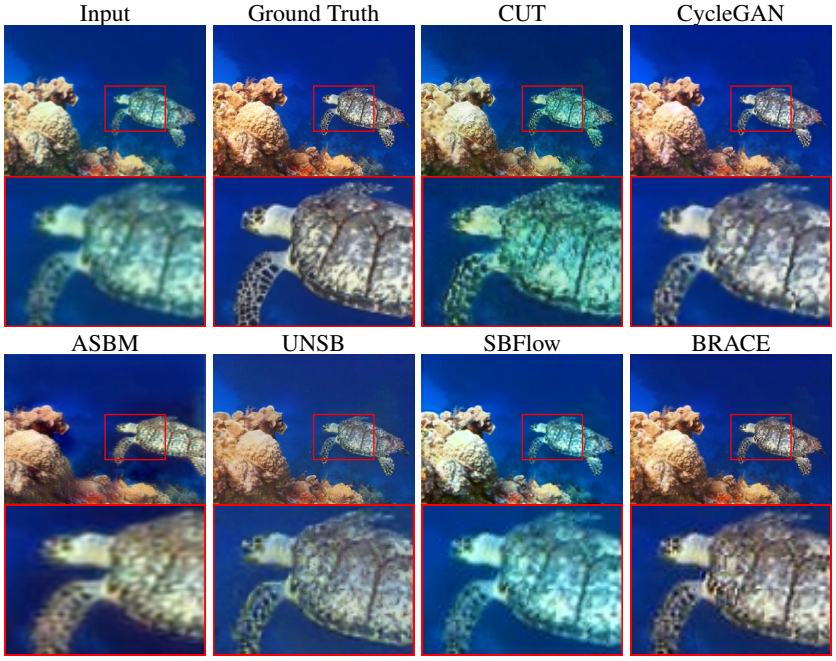


Figure 12: Qualitative comparison on the EUVP dataset. Red boxes indicate regions shown enlarged below each image.



Figure 13: Qualitative comparison on the UIEB dataset. Red boxes indicate regions shown enlarged below each image.

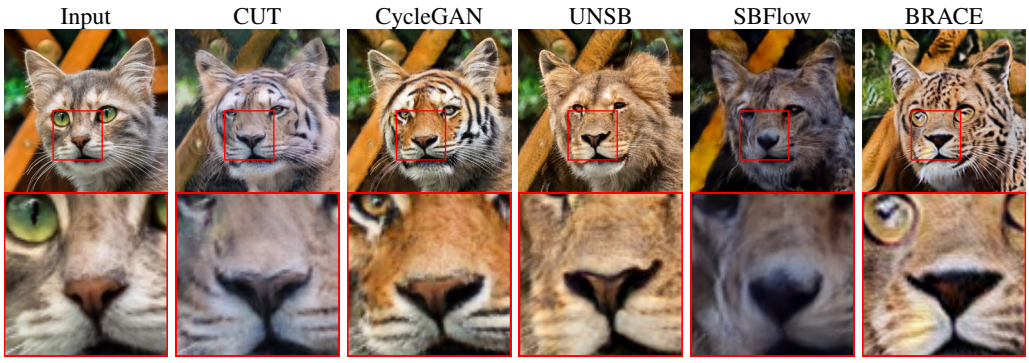


Figure 14: Qualitative comparison on AFHQ Cat → Wild. Red boxes indicate regions shown enlarged below each image. Compared with SBFlow, BRACE produces sharper target-domain appearance and more plausible local texture, while remaining visually competitive with strong unpaired translation baselines.

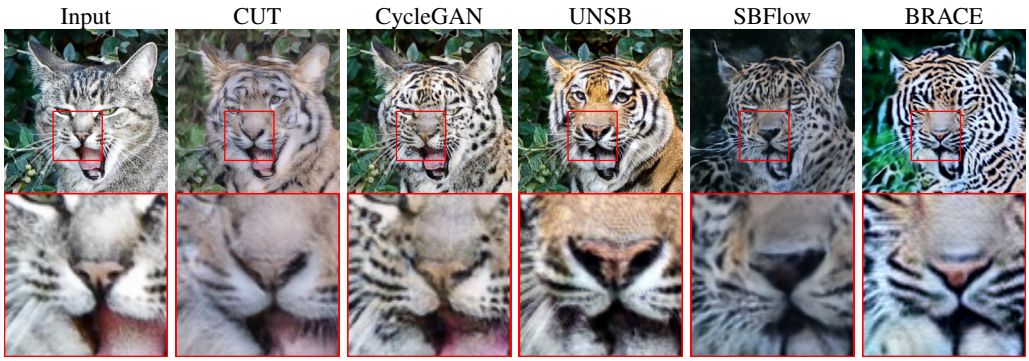


Figure 15: Additional qualitative comparison on AFHQ Cat → Wild. The examples highlight differences in target-domain appearance and structure preservation.

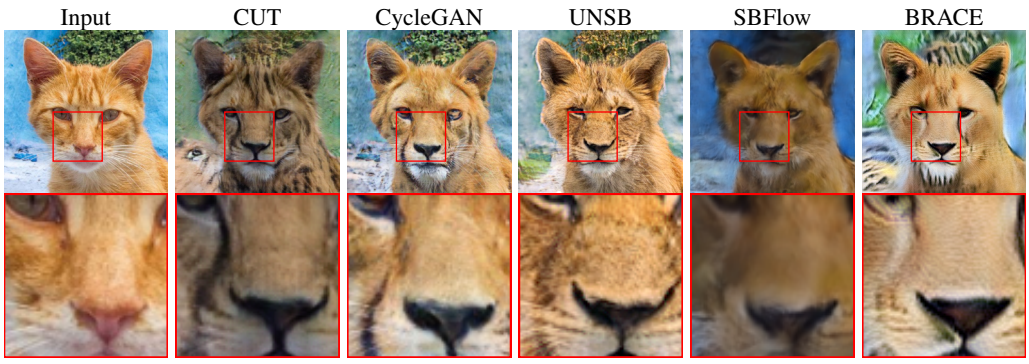


Figure 16: Additional qualitative comparison on AFHQ Cat → Wild. The examples highlight differences in target-domain appearance and structure preservation.