SFBD-OMNI: Bridge Models for Lossy Measurement Restoration with Limited Clean Samples

Haoye Lu[1,2], Yaoliang Yu[1,2], Darren Lo[1]
[1]School of Computer Science, University of Waterloo [2]Vector Institute ✉ {haoye.lu, yaoliang.yu, dlslo} @uwaterloo.ca

ICLR 2026

Summary

We propose SFBD-OMNI, a bridge model framework that recovers clean data distributions from corrupted observations under arbitrary measurement processes, with theoretical convergence guarantees.

- Frames distribution restoration as a **one-sided entropic optimal transport** problem
- **Test criterion for recoverability**: small clean sample set enables recovery when identifiability fails
- **Online variant** for end-to-end training with guaranteed convergence
- Handles **arbitrary corruption**: Gaussian noise, pixel masking, grayscale, blur, Poisson, compressive sensing

Motivation

- Stable Diffusion trained on LAION (12M+ pairs) — hard to avoid copyrighted/private content
- Diffusion models **memorize training samples** (Daras et al., 2024)
- In medical imaging, cleaner scans require higher radiation doses
- In astronomy, clean images need long exposures under ideal

KL Ambient Projection (KLAP)

Given corrupted distribution $q = \mathcal{T}_r p_{\text{data}}$, recover p_{data}:

$$p^* = \arg\min_p \; D_{\text{KL}}(q \,\|\, \mathcal{T}_r p)$$

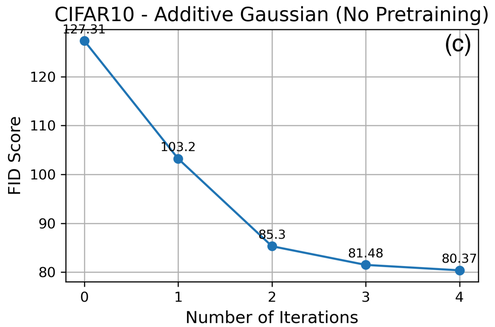
- **Identifiable** ($\mathcal{T}_r$ injective): additive Gaussian noise, pixel masking
- **Non-identifiable** ($\mathcal{T}_r$ not injective): grayscale, Gaussian blur

When non-identifiable, augment with prior h from small clean set:

$$p^*_\lambda = \arg\min_p \; D_{\text{KL}}(q \,\|\, \mathcal{T}_r p) + \lambda \, D_{\text{KL}}(h \,\|\, p)$$

Identifiability $\neq$ recoverability: sample complexity is $\Theta((\log N)^{-2})$ — pretraining on clean data is necessary.

Identifiability ≠ Recoverability



FID without clean samples. Even with identifiable corruption, FID saturates at 80.37 vs. 10.81 with just 50 clean samples.

Preliminary

Diffusion models learn distributions via forward SDE $dx_t = dw_t$ $(x_0 \sim \mu)$ and reverse:

$$dx_t = -\nabla \log p_t(x_t)\, dt + d\bar{w}_t$$

SFBD operates at finite τ: denoise corrupted samples via backward SDE, then retrain. Limited to Gaussian corruption.

Bridge models generalize to arbitrary distributions. Given $(x, y) \sim \pi$, minimize CDM loss:

$$\mathcal{L}_{\text{CDM}}(\theta, \pi) = \mathbb{E}_{t,\,(x,y),\,x_t} \| f(x_t; x, y, t) - f_\theta(x_t; y, t) \|^2$$

Sample posterior $\pi_{0|1}(x|y)$ by integrating backward: $dx_t = f_\theta(x_t; y, t)\, dt + g(t)\, d\bar{w}_t$

One-Sided Entropic OT Formulation

Applying the Donsker-Varadhan principle to the KL divergence:

$$\min_p \min_{u_y} \; \mathbb{E}_{y\sim q}\big[D_{\text{KL}}(u_y \| p) - \mathbb{E}_{x\sim u_y}[\log r(y|x)]\big] + \lambda D_{\text{KL}}(h \| p)$$

- Equivalent to one-sided entropic OT with cost $c(x, y) = -\log r(y|x)$
- Only y-marginal constrained to q; x-marginal optimized
- Yields **alternating minimization** with closed-form updates for both subproblems

SFBD-OMNI Algorithm

Alternating minimization with closed-form subproblems. Bridge models learn posterior $u_y^k(x)$ via flow matching.

$$u_y^k(x) = \frac{p^k(x)\, r(y|x)}{\mathcal{T}_r p^k(y)}, \qquad p^{k+1}(x) = \tfrac{1}{1+\lambda}\tilde{p}^{k+1}(x) + \tfrac{\lambda}{1+\lambda} h(x)$$

Algorithm 1: SFBD-OMNI

Input: $\mathcal{E}_{\text{clean}}$, $\mathcal{E}_{\text{noisy}}$, $\mathcal{L}_{\text{CDM}}$

// Pretrain on clean samples

$\theta \leftarrow \arg\min \mathcal{L}_{\text{CDM}}(\theta,\ h_{\mathcal{E}_{\text{clean}}}(x)\, r(y|x))$

$\mathcal{E} \leftarrow \{x \sim u_\theta(x|y) : y \in \mathcal{E}_{\text{noisy}}\}$

// Iteratively optimize

for $k = 1, \ldots, K$ **do**

$\theta \leftarrow \arg\min \mathcal{L}_{\text{CDM}}(\theta,\ p(x)\, r(y|x))$

with $p = \frac{1}{1+\lambda} p_{\mathcal{E}} + \frac{\lambda}{1+\lambda} h_{\mathcal{E}_{\text{clean}}}$

$\mathcal{E} \leftarrow \{x \sim u_\theta(x|y) : y \in \mathcal{E}_{\text{noisy}}\}$

Output: u_θ

Algorithm 2: Online SFBD-OMNI

Input: $\mathcal{E}_{\text{clean}}$, $\mathcal{E}_{\text{noisy}}$, $\mathcal{L}_{\text{CDM}}$, γ

// Pretrain on clean samples

$\theta \leftarrow \arg\min \mathcal{L}_{\text{CDM}}(\theta,\ h_{\mathcal{E}_{\text{clean}}}(x)\, r(y|x))$

$\mathcal{E} \leftarrow \{x \sim u_\theta(x|y) : y \in \mathcal{E}_{\text{noisy}}\}$

// Iteratively optimize (online)

for $k = 1, \ldots, K$ **do**

$\theta \leftarrow \arg\min \mathcal{L}_{\text{CDM}}(\theta,\ p(x)\, r(y|x))$

with $p = \frac{1}{1+\lambda} p_{\mathcal{E}} + \frac{\lambda}{1+\lambda} h_{\mathcal{E}_{\text{clean}}}$

Replace ratio γ of $\mathcal{E}$ with $x \sim u_\theta(x|y)$

Output: u_θ

Convergence Guarantee

Proposition (Convergence to the Optimum)

For $\gamma \in (0, 1]$ and arbitrary p^0: $p^k \to p^*_\lambda$ as $k \to \infty$. When $\lambda \to 0$:

$$D_{\text{KL}}(h^\dagger \| p^{k+1}) \le D_{\text{KL}}(h^\dagger \| p^k), \qquad \min_{1\le k\le K} D_{\text{KL}}(q \| \mathcal{T}_r p^k) \le \frac{D_{\text{KL}}(h^\dagger \| p^0)}{\gamma K}$$

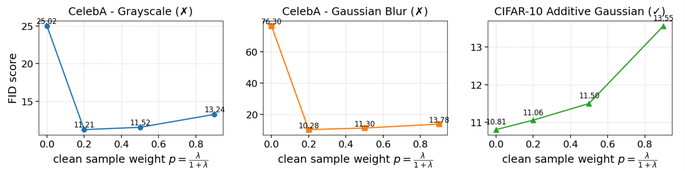
- Monotonically decreasing KL divergence at each iteration
- Smaller γ = cheaper per-step (no optimizer reset); total time may decrease

Performance Comparison (FID ↓)

Method	Pixel Mask	Add. Gauss.	Grayscale (C10)	Gauss. Blur	Grayscale (CelebA)
SURE-Score	220.01	132.61	109.04	191.96	219.81
AmbientDiff	28.88	--	--	--	--
EMDiffusion	**21.08**	86.47	115.11	91.89	59.04
SFBD	--	13.53	--	--	--
SFBD-OMNI	21.31	**10.81**	32.61	11.60	11.85
Online SFBD-OMNI	22.43	11.06	**31.32**	**10.28**	**11.21**

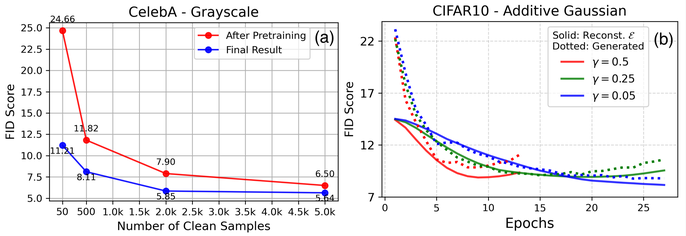
All methods pretrained on 50 clean images. Identifiable: Pixel Mask, Additive Gaussian. Non-identifiable: Grayscale, Gaussian Blur.

Effect of Clean Sample Weight



FID under different clean sample weights $p = \lambda/(1+\lambda)$. When identifiability holds, regularization is unnecessary. When it fails, clean samples as a soft prior are essential.

Ablation Studies



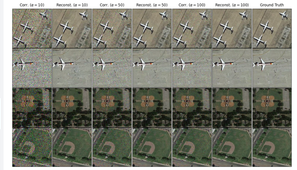
(a) Number of clean samples. More clean samples improve FID with diminishing returns.

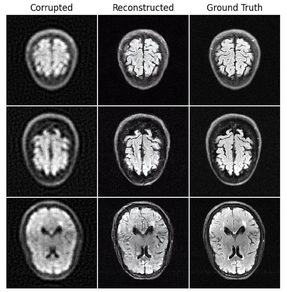
(b) Update ratio γ. Smaller γ yields more stable training.

Generated Samples

CIFAR-10 (Gaussian) **CelebA** (Gaussian) **CelebA** (Grayscale) **CIFAR-10** (Mask)

High-Resolution Results





Satellite (Poisson corruption) **MRI** (Compressive sensing)

Conclusion

- **SFBD-OMNI:** principled framework for distribution recovery from corrupted data, no adversarial training
- Effective with small clean sample set even when identifiability fails
- Online variant: efficient end-to-end training with convergence guarantees

Code: github.com/watml/SFBD-omni **Contact:** haoye.lu@uwaterloo.ca