Be Tangential to Manifold:
Discovering Riemannian Metric for Diffusion Models

[1] HOKKAIDO UNIVERSITY [2] CyberAgent AI Lab

Shinnosuke Saito[1], Takashi Matsubara[2,1]

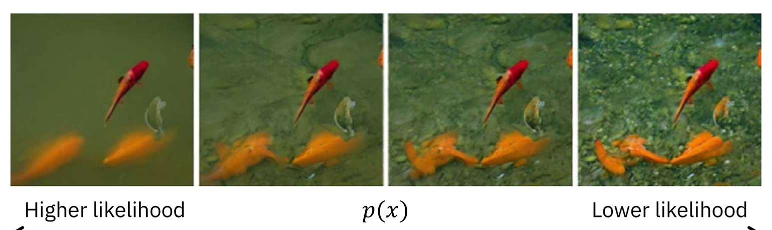
ECCV 2026 | MALMÖ | SEPT 8-12

Paper

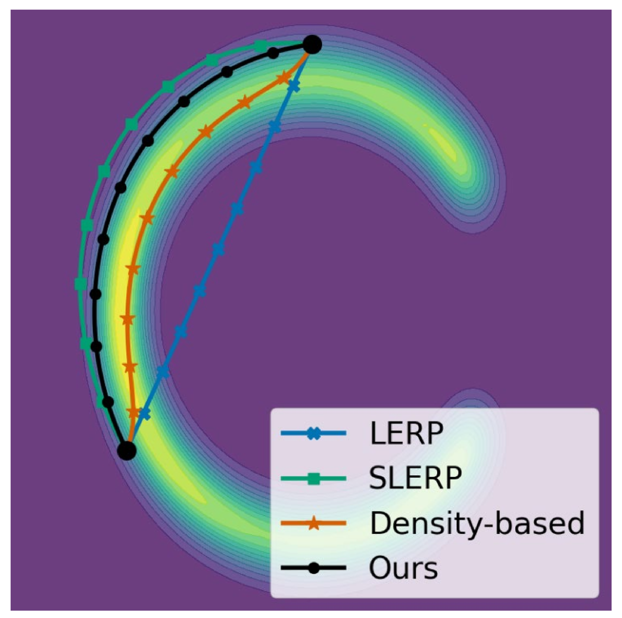
Motivation

- **Problem:** Diffusion models implicitly learn the data manifold. For operations such as *interpolation* and *guidance* to respect this structure, we need to characterize its intrinsic geometry.
- **What is missing from density-based metrics?**[1]

High-density samples are often over-smoothed and lack details[2].



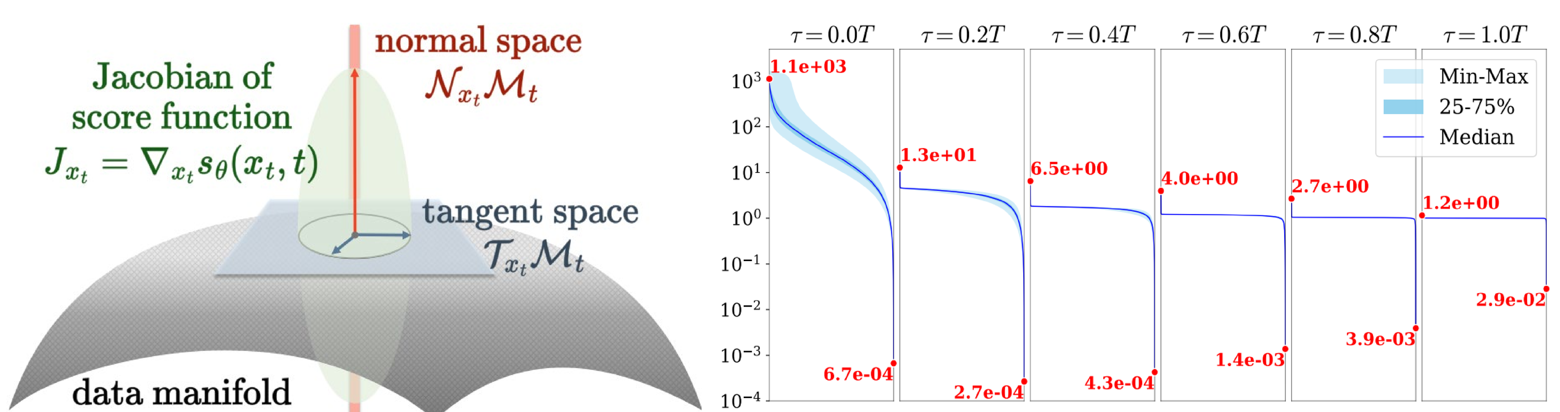
Density describes where samples concentrate, but not how the data manifold is locally oriented.



Can we define a Riemannian metric that reflects the intrinsic geometry of the data manifold?

Score Jacobian Reveals Geometry

- **Prior work:** The Jacobian of score function $J_{x_t} \coloneqq \nabla_{x_t} s_\theta(x_t, t)$ exhibits tangent-normal spectral separation[3,4].



Jacobian spectrum shows tangent–normal separation.

- **Insight:** Measure a local direction by its induced score change.

Proposed Riemannian Metric

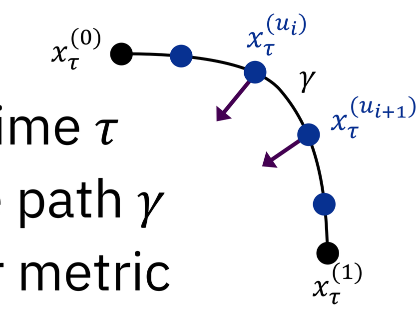
- **Definition:** For $x_t \in \mathbb{R}^D$ and tangent vectors $v, w \in T_{x_t}\mathbb{R}^D$, let $J_{x_t} \coloneqq \nabla_{x_t} s_\theta(x_t, t)$ denote the Jacobian of the score function.

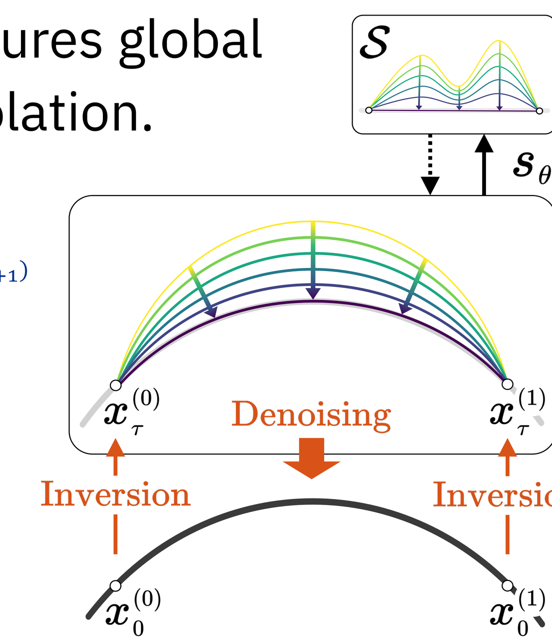
$$g_{x_t}(v, w) \coloneqq \langle J_{x_t} v, J_{x_t} w \rangle = v^\top G_{x_t} w, \quad G_{x_t} \coloneqq J_{x_t}^\top J_{x_t}$$

  - Pullback of the Euclidean metric through the score function
  - For unit directions: Tangent → low cost / Normal → high cost
  - Training-free / No architecture modification

Geodesic Interpolation

- **Goal:** Evaluate whether our metric captures global manifold geometry via geodesic interpolation.
- **Method:**
  1. invert endpoints to time τ
  2. discretize & initialize path γ
  3. optimize γ under our metric



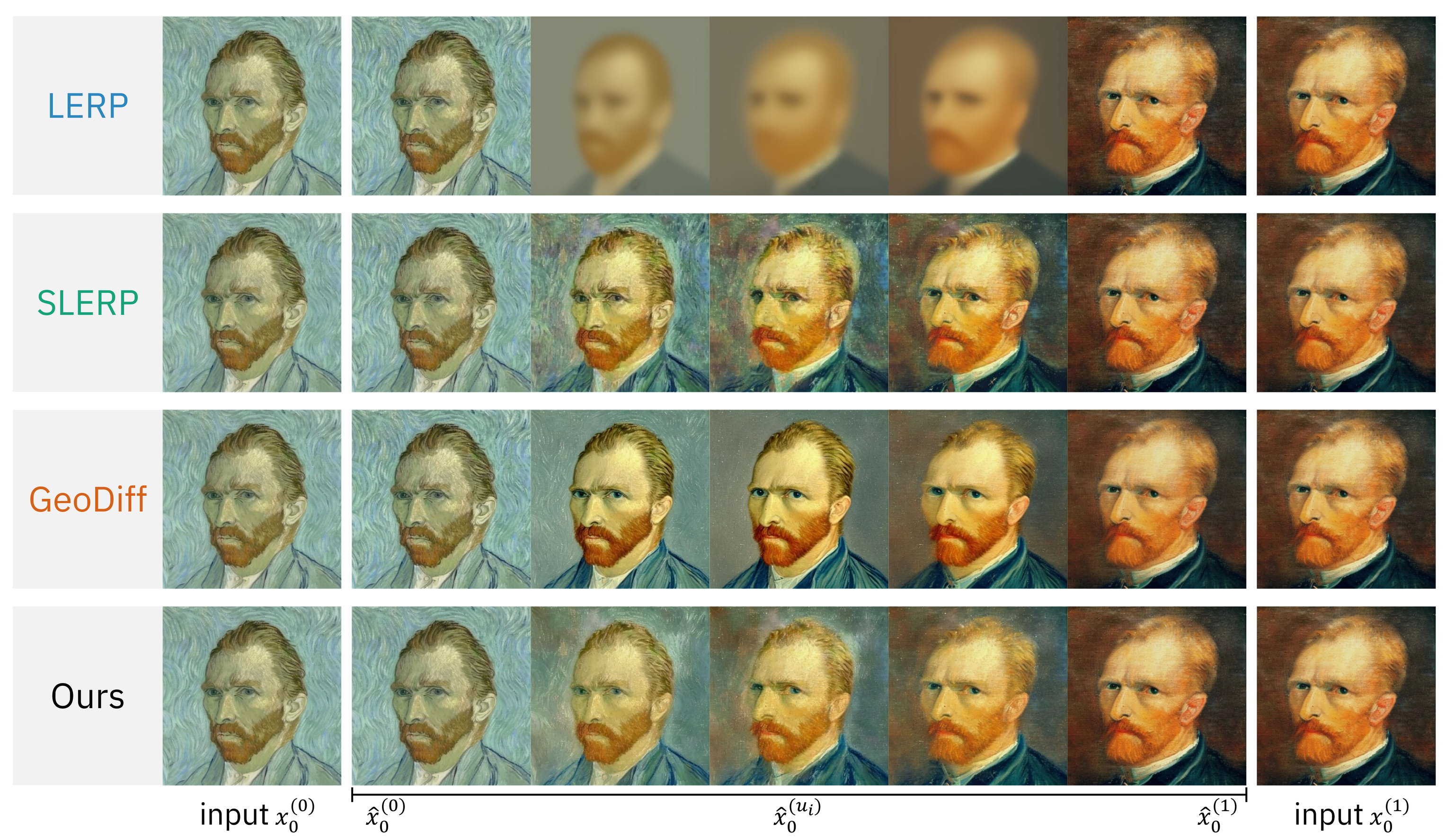


$$\gamma^* = \arg\min_{\gamma} \frac{1}{2\Delta u} \sum_{i=0}^{N-1} \left\| s_\theta(x_\tau^{(u_{i+1})}, \tau) - s_\theta(x_\tau^{(u_i)}, \tau) \right\|_2^2$$

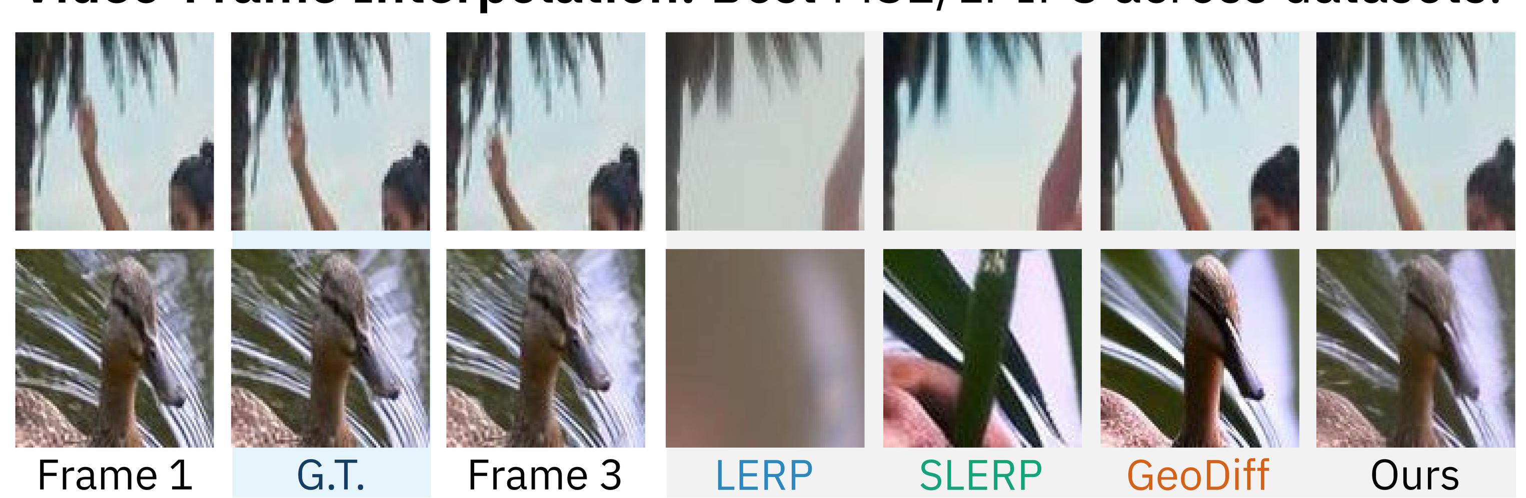
minimize score changes along the path

  4. denoise the optimized path

- **Image Interpolation:** Ours preserves fine details, while Geodiff tends to oversmooth; quantitative results in the paper.



- **Video-Frame Interpolation:** Best MSE/LPIPS across datasets.

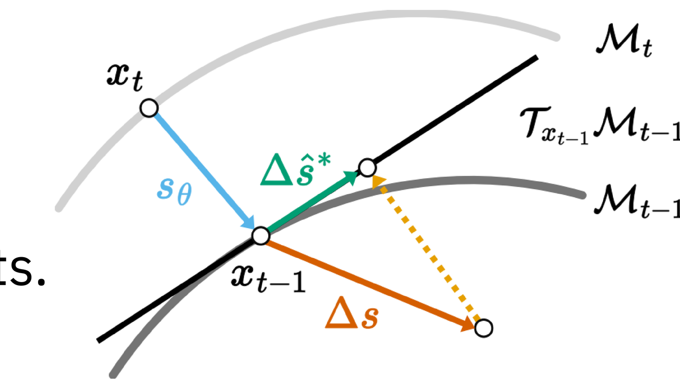


Model: Stable Diffusion 2.1-base

Method	MSE ↓ ($\times 10^{-3}$) DAVIS	Human	RE10K	LPIPS ↓ DAVIS	Human	RE10K
LERP	<u>12.135</u>	4.566	6.299	0.590	0.379	0.377
SLERP	15.440	6.080	6.128	0.487	0.320	0.301
NAO	108.211	99.867	121.680	0.679	0.668	0.664
NoiseDiff	46.881	41.994	28.867	0.561	0.552	0.482
GeoDiff	13.253	<u>3.363</u>	<u>5.941</u>	<u>0.334</u>	<u>0.184</u>	<u>0.229</u>
FIM-based	30.172	11.638	12.679	0.535	0.388	0.373
Ours	**8.777**	**2.018**	**2.771**	**0.318**	**0.170**	**0.178**

Metric-based Guidance Correction

- **Goal:** Evaluate whether our metric captures local manifold geometry via guidance correction. CFG may push samples off-manifold, causing artifacts.



- **Method:** Correct CFG guidance Δs.

$$\Delta\hat{s}^* = \operatorname{argmin}_{\Delta\hat{s}} \underbrace{d_E\big(x_{t-1}(\Delta s), x_{t-1}(\Delta\hat{s})\big)^2}_{\text{Euclidean distance: preserve original guidance}} + \lambda \underbrace{d_g\big(x_{t-1}(0), x_{t-1}(\Delta\hat{s})\big)^2}_{\text{Riemannian distance: stay close under our metric}}$$

$(I + \lambda G_{x_t})\Delta\hat{s}^* = \Delta s$ } Solved by 1 Conjugate Gradient step from Δs

Normal components are attenuated more than tangent ones.

- **Text-to-Image Generation:** Ours achieves Better FID across all tested CFG scales while maintaining CLIP Score.

"A man is cooking in a crowded kitchen"
"A person is in the grass holding a frisbee"
"A small white bird with a long beak on the branch"
"A plane flies over water with two islands nearby"
"Scissors and a pile of postage stamps topped by one of President Nixon"

Model: Stable Diffusion 2.1-base, dataset: MS-COCO 2014 val.

Method	$w = 5.0$ FID ↓	CLIP ↑	$w = 7.5$ FID ↓	CLIP ↑	$w = 12.5$ FID ↓	CLIP ↑
CFG	<u>11.69</u>	0.313	14.29	0.314	<u>17.28</u>	0.315
CFG++	11.87	0.313	<u>13.98</u>	0.314	17.76	0.315
Ours	**11.53**	0.313	**13.81**	0.314	**16.04**	0.315

Guidance distillation w/ LCM reduces FID at no extra inference cost.

Method	FID ↓	CLIP ↑
CFG	17.91	0.306
Ours	**16.98**	0.306

Limitations & Future Work

- Computational Cost: Geodesic interpolation requires iterative optimization. Guidance correction adds 4 score evaluations per step.

Interpolation runtime

Method	Time (s)
FIM-based	178.29
GeoDiff	216.79
Ours	392.92

- Broader Scope: Extend to other tasks & flow-matching models.

References

[1] Yu et al., CVPR, 2025.
[2] Karczewski et al., ICML, 2025.
[3] Stanczuk et al., ICML, 2024.
[4] Ventura et al., ICLR, 2025.