Exploring *Controllability*, *Structural Properties*, and *Safety* in Diffusion Models

TWsiam

Contributed Talks

Jia-Wei Liao (廖家緯)
Ph.D. Candidate in Computer Science,
National Taiwan University

Jia-Wei Liao

Pre-training: ***Math***

Fine-tuning: ***CS***

BS in Math

MS in Applied Math

PhD Candidate in CS

2020

2022

2023

2024

2025

USRP

DA Intern

Research Assistant

Research Intern

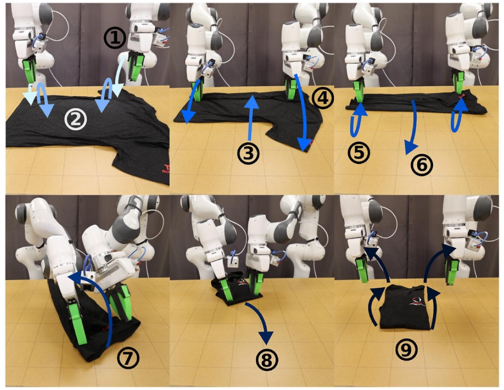
SWE Intern

Research Intern

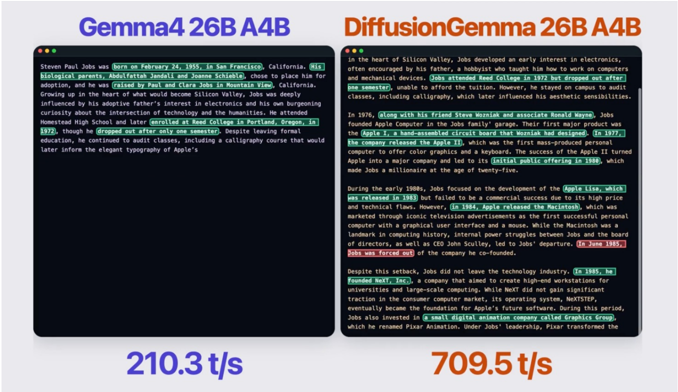
Learning the Distribution From the World

World Models
(Veo 3.1, 2025)

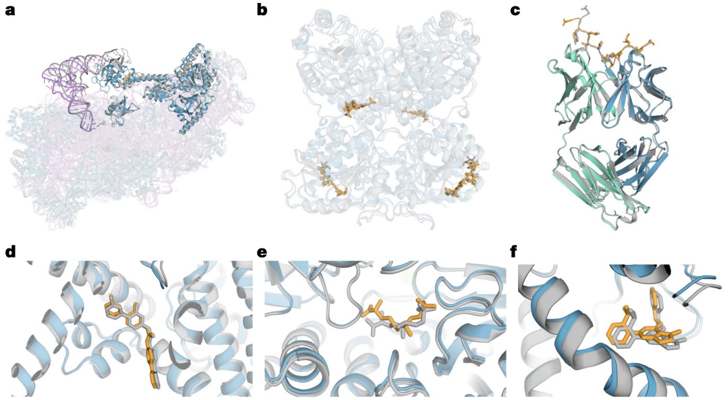
Robot Manipulation
(Diffusion Policy, 2023)



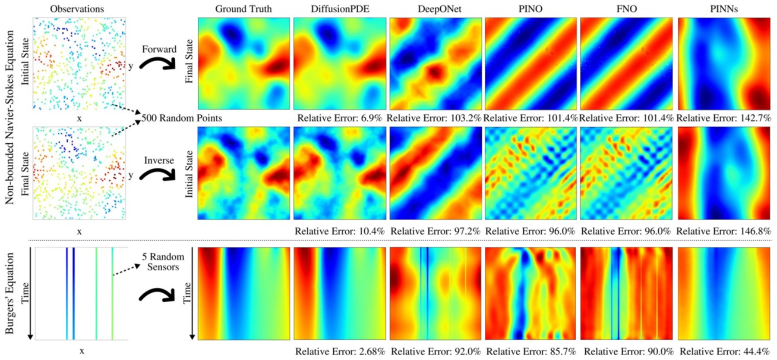
Large Language Model
(DiffusionGemma, 2026)



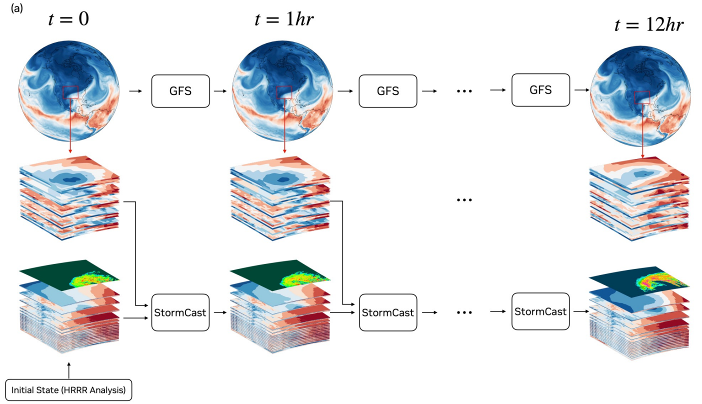
Biomolecular
(AlphaFold3, 2024)



Generative PDE Solver
(DiffusionPDE, 2024)



Weather Forecasting
(StormCast, 2024)



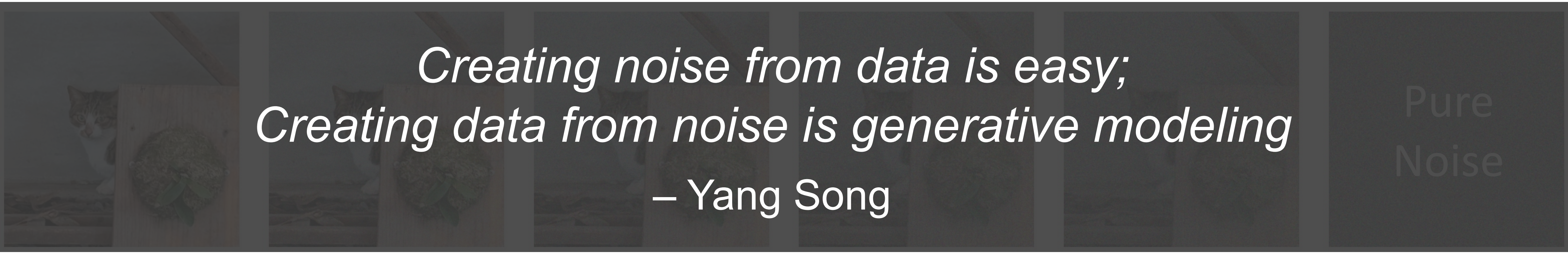
What is Diffusion Model?

Forward Process: Add noise step by step, from data to pure noise

Reverse Process: Generate data from pure noise by denoising

What is Diffusion Model?

Forward Process: Add noise step by step, from data to pure noise



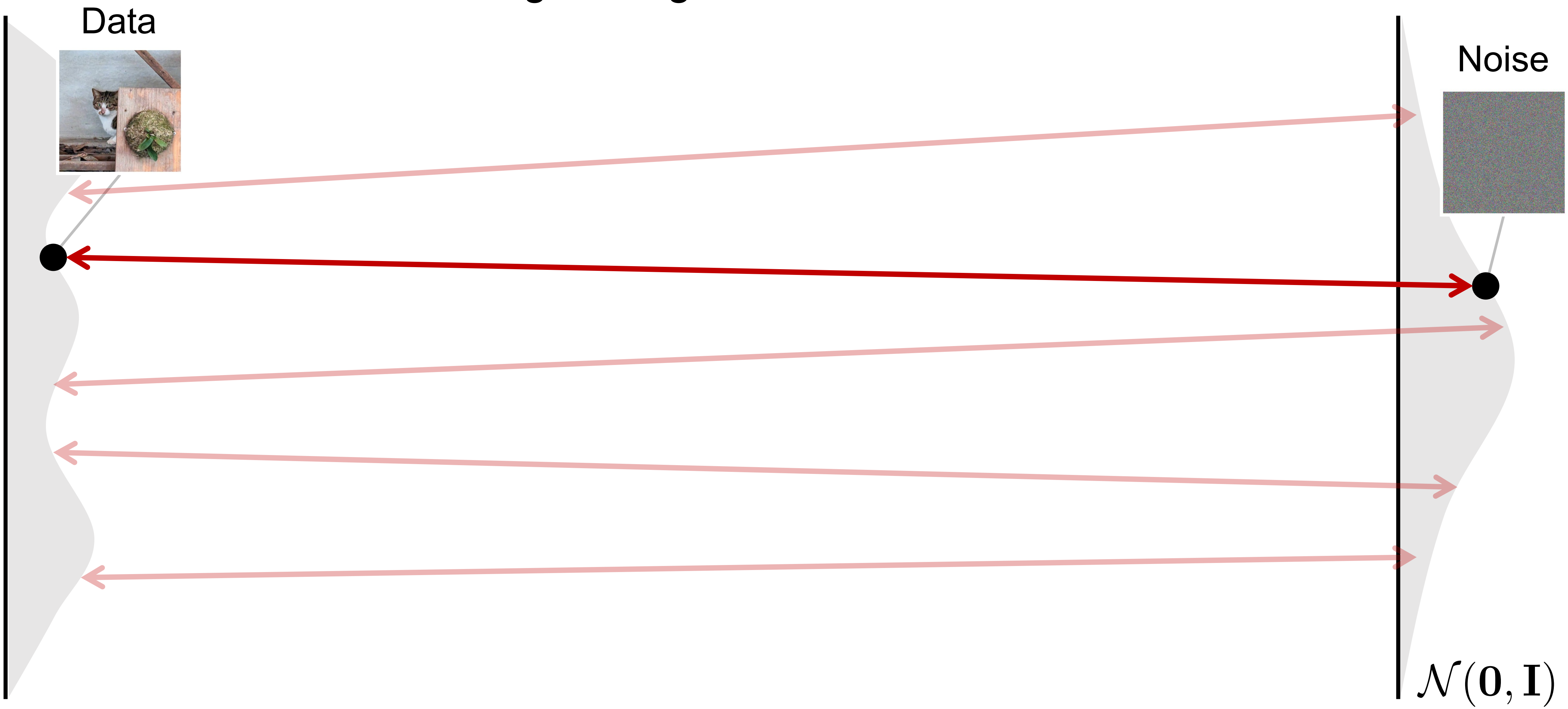
Creating noise from data is easy;
Creating data from noise is generative modeling

– Yang Song

Reverse Process: Generate data from pure noise by denoising

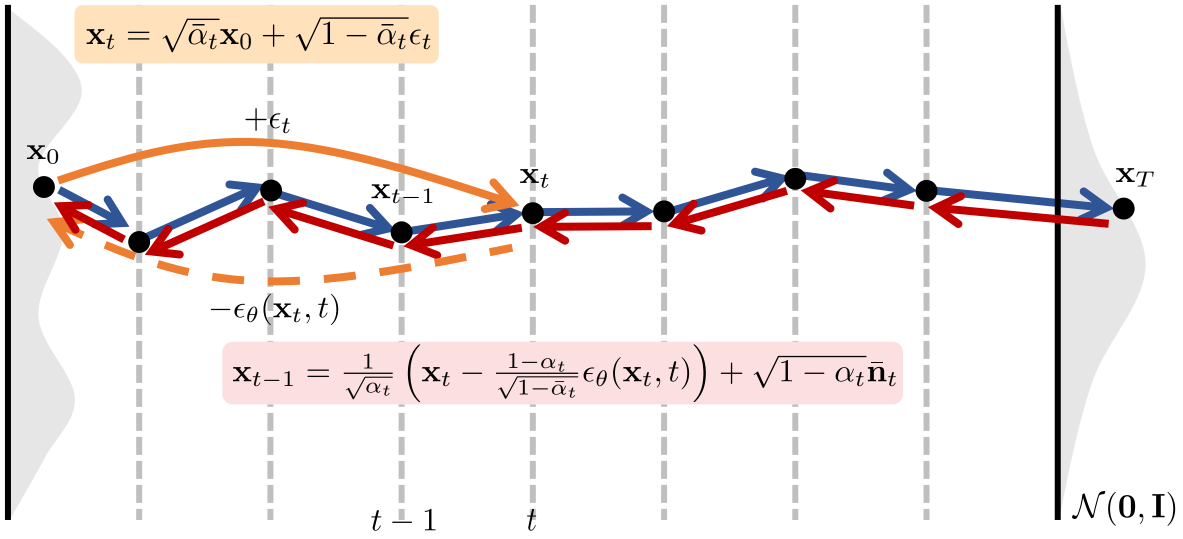
The Goal of Diffusion Models

Building a bridge between noise and data



Diffusion and Flow Matching Models

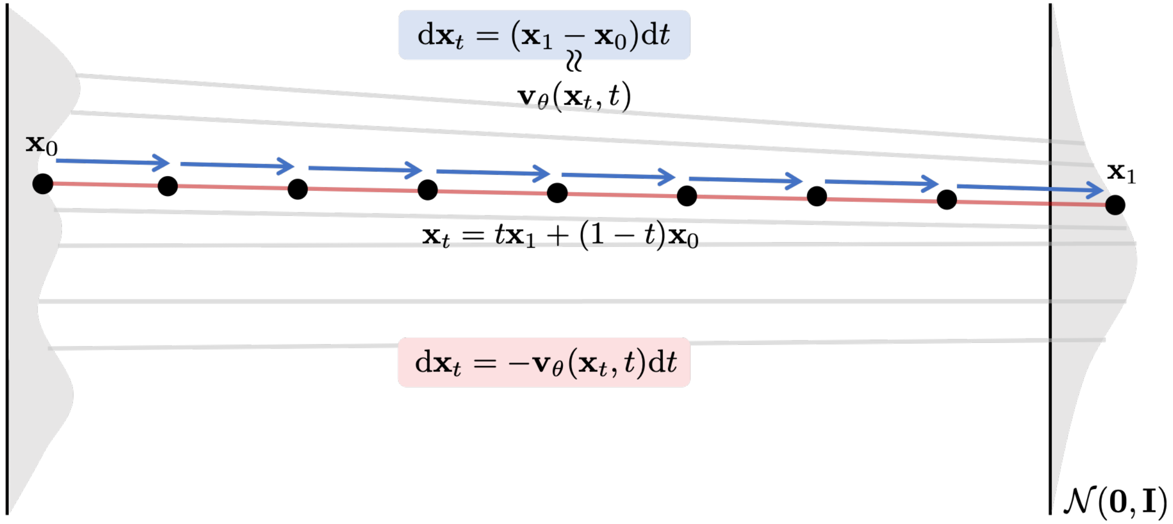
Diffusion Models



$$\mathcal{L}_{\text{DDPM}}(\theta) = \mathbb{E}_{t\sim U(1,T), \boldsymbol{x}_0\sim p_0, \epsilon_t \sim \mathcal{N}(\mathbf{0},\mathbf{I})} \|\epsilon_\theta(\boldsymbol{x}_t, t) - \epsilon_t\|_2^2$$

$$\boldsymbol{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar\alpha_t}}\epsilon_\theta(\boldsymbol{x}_t, t)\right) + \sqrt{1-\alpha_t}\boldsymbol{n}_t$$

Flow Matching Models

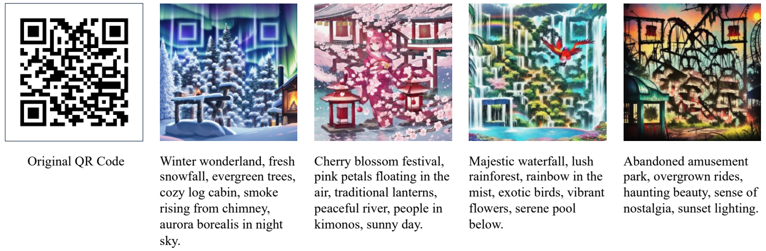


$$\mathcal{L}_{\text{FM}}(\theta) = \mathbb{E}_{t\sim U(0,1), \boldsymbol{x}_0\sim p_0, \boldsymbol{x}_1 \sim \mathcal{N}(\mathbf{0},\mathbf{I})} \|\boldsymbol{v}_\theta(\boldsymbol{x}_t, t) - (\boldsymbol{x}_1 - \boldsymbol{x}_0)\|_2^2$$

$$\boldsymbol{x}_1 = \boldsymbol{x}_0 + \int_0^1 \boldsymbol{v}_\theta(\boldsymbol{x}_t, t)\mathrm{d}t \quad \text{(Euler Method)}$$

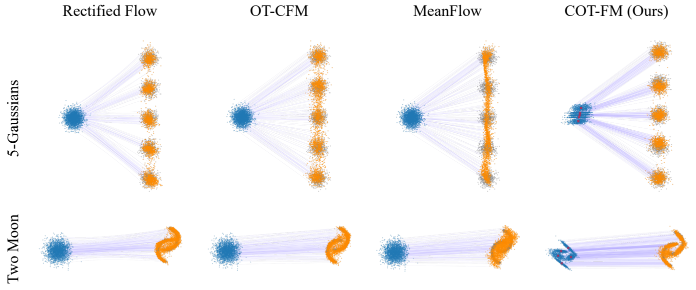
DiffQRCoder: Diffusion-based Aesthetic QR Code Generation with Scanning Robustness Guided Iterative Refinement

WACV 2025



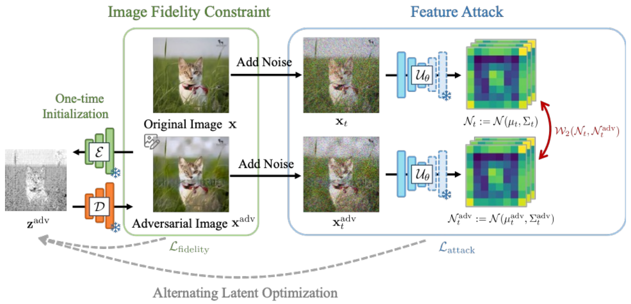
COT-FM: Cluster-wise Optimal Transport Flow Matching

CVPR 2026



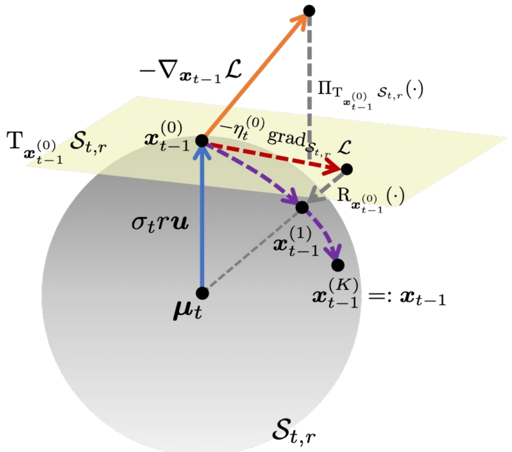
Pixel Is Not A Barrier: An Effective Evasion Attack for Pixel-Domain Diffusion Models

AAAI 2025



DiffRGD: An Inference-Time Diffusion Guidance Through Riemannian Gradient Descent

Submitted to ECCV 2026

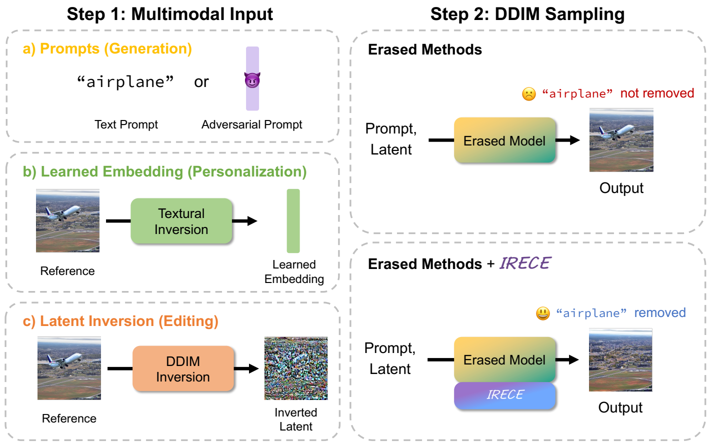


CNP-Flow: Unified Temporal Flow Matching via Conditional Noise Predictor

Submitted to NeurIPS 2026

M-ErasureBench: A Comprehensive Multimodal Evaluation Benchmark for Concept Erasure in Diffusion Models

WACV 2026



DiffQRCoder: Diffusion-based Aesthetic QR Code Generation with Scanning Robustness Guided Iterative Refinement

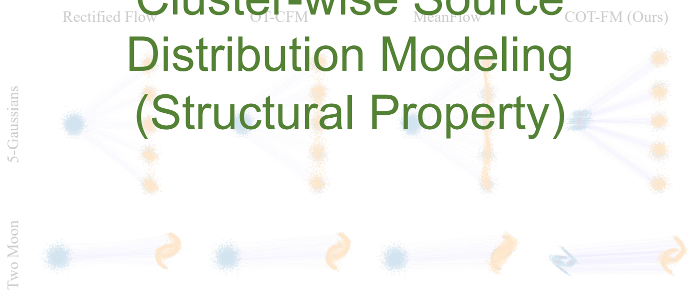
WACV 2025

Applications of Diffusion Guidance (Controllability)

COT-FM: Cluster-wise Optimal Transport Flow Matching

CVPR 2026

Cluster-wise Source Distribution Modeling (Structural Property)



Pixel Is Not A Barrier: An Effective Evasion Attack for Pixel-Domain Diffusion Models

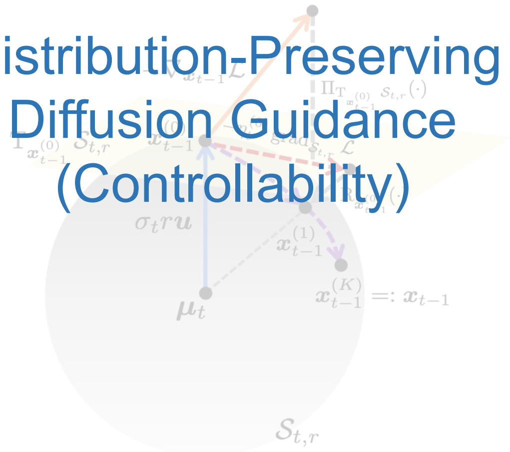
AAAI 2025

Adversarial Attack (Safety)

DiffRGD: An Inference-Time Diffusion Guidance Through Riemannian Gradient Descent

Submitted to ECCV 2026

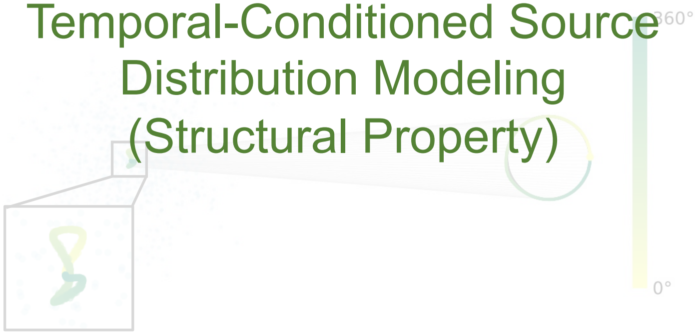
Distribution-Preserving Diffusion Guidance (Controllability)



CNP-Flow: Unified Temporal Flow Matching via Conditional Noise Predictor

Submitted to NeurIPS 2026

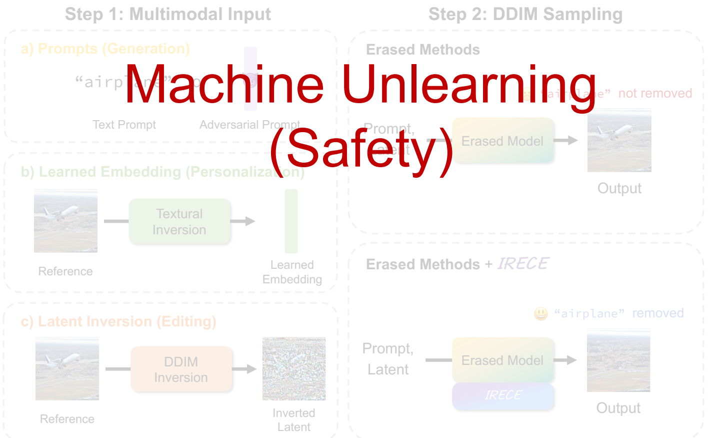
Temporal-Conditioned Source Distribution Modeling (Structural Property)



M-ErasureBench: A Comprehensive Multimodal Evaluation Benchmark for Concept Erasure in Diffusion Models

WACV 2026

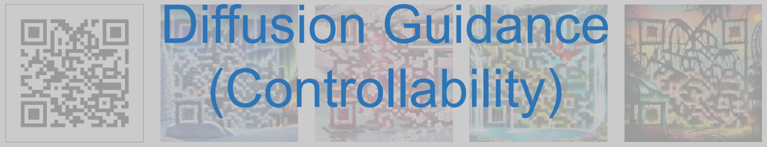
Machine Unlearning (Safety)



DiffQRCoder: Diffusion-based Aesthetic QR Code Generation with Scanning Robustness Guided Iterative Refinement

WACV 2025

Applications of Diffusion Guidance (Controllability)



Original QR Code

Winter wonderland, fresh snowfall, evergreen trees, cozy log cabin, smoke rising from chimney, aurora borealis in night sky.

Cherry blossom festival, pink petals floating in the air, traditional lanterns, peaceful river, people in kimonos, sunny day.

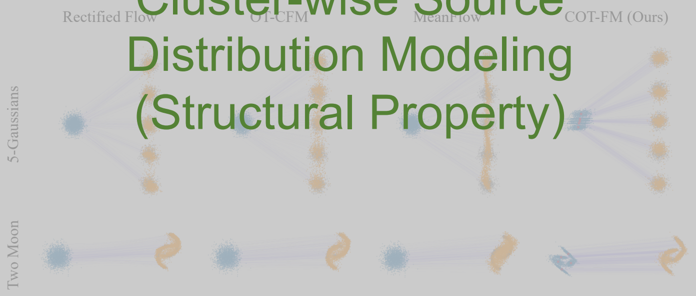
Majestic waterfall, lush rainforest, rainbow in the mist, exotic birds, vibrant flowers, serene pool below.

Abandoned amusement park, overgrown rides, haunting beauty, sense of nostalgia, sunset lighting.

COT-FM: Cluster-wise Optimal Transport Flow Matching

CVPR 2026

Cluster-wise Source Distribution Modeling (Structural Property)



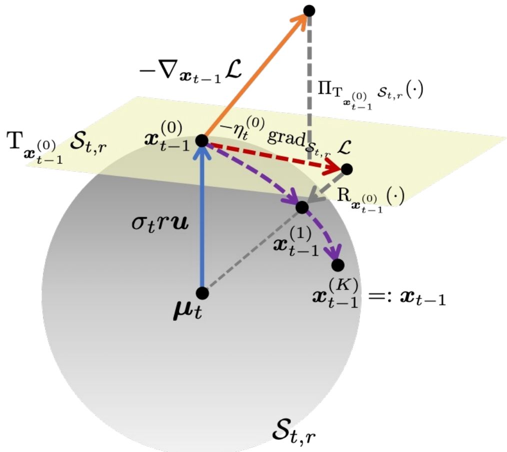
Pixel Is Not A Barrier: An Effective Evasion Attack for Pixel-Domain Diffusion Models

AAAI 2025

Adversarial Attack (Safety)

DiffRGD: An Inference-Time Diffusion Guidance Through Riemannian Gradient Descent

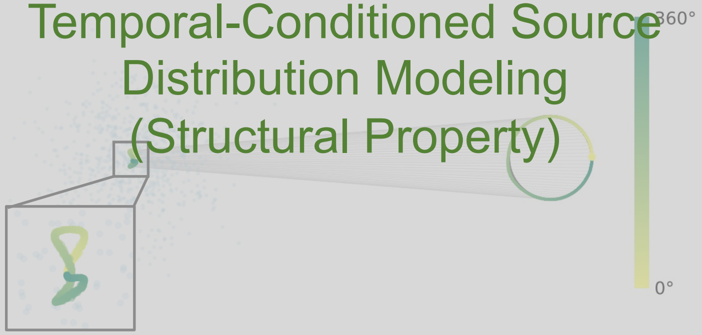
Submitted to ECCV 2026



CNP-Flow: Unified Temporal Flow Matching via Conditional Noise Predictor

Submitted to NeurIPS 2026

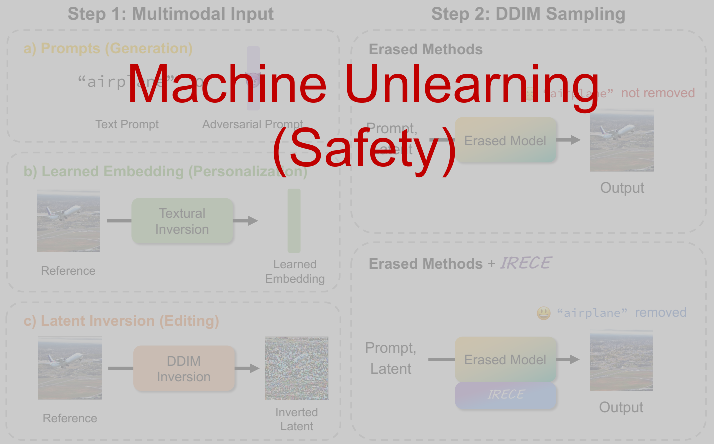
Temporal-Conditioned Source Distribution Modeling (Structural Property)



M-ErasureBench: A Comprehensive Multimodal Evaluation Benchmark for Concept Erasure in Diffusion Models

WACV 2026

Machine Unlearning (Safety)



Inference-Time Diffusion Controlling

Image Inverse Problem

$$\boldsymbol{y} = \mathcal{A}\boldsymbol{x} + \boldsymbol{\eta}$$

Operator

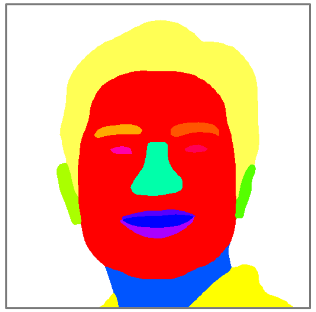
Noise

Conditional Generation
(Segmentation Mask)

ψ : Face Segmentation Model

Measurement $\boldsymbol{y}$

Reconstruction $\hat{\boldsymbol{x}}$



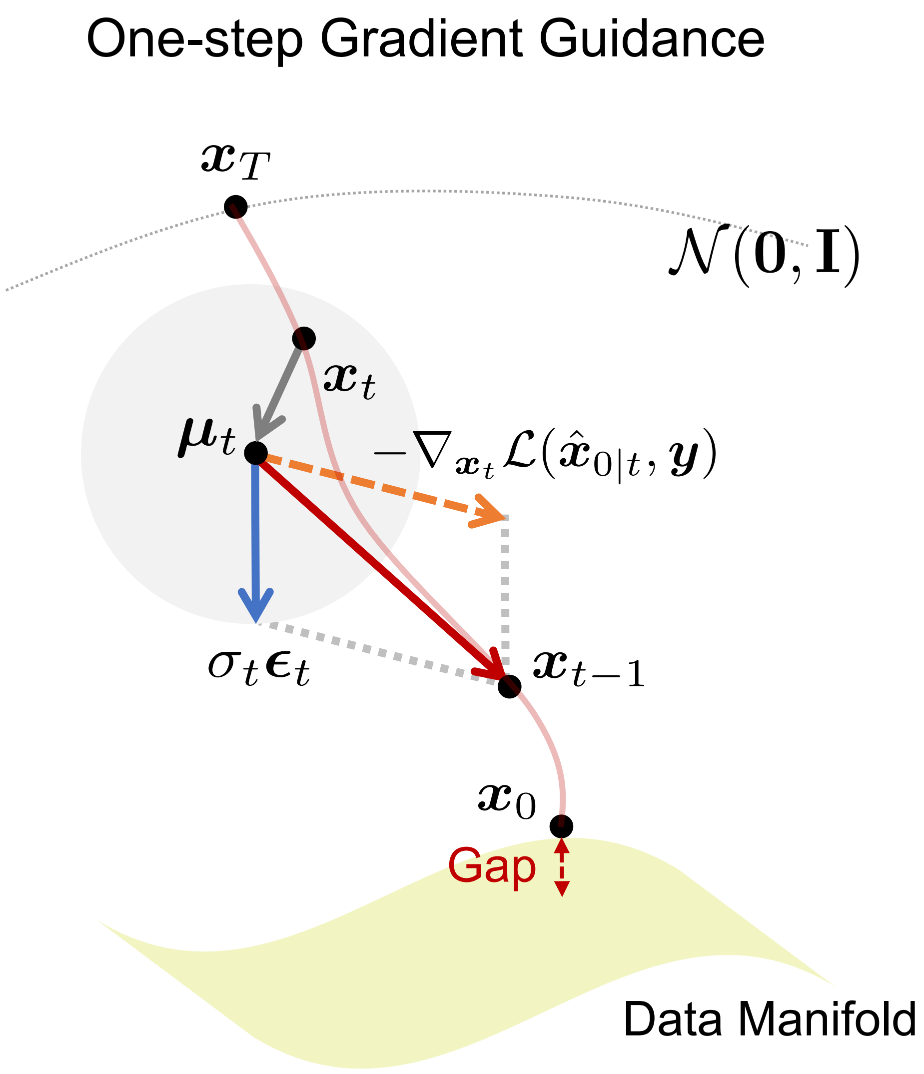
Condition $\boldsymbol{y}$

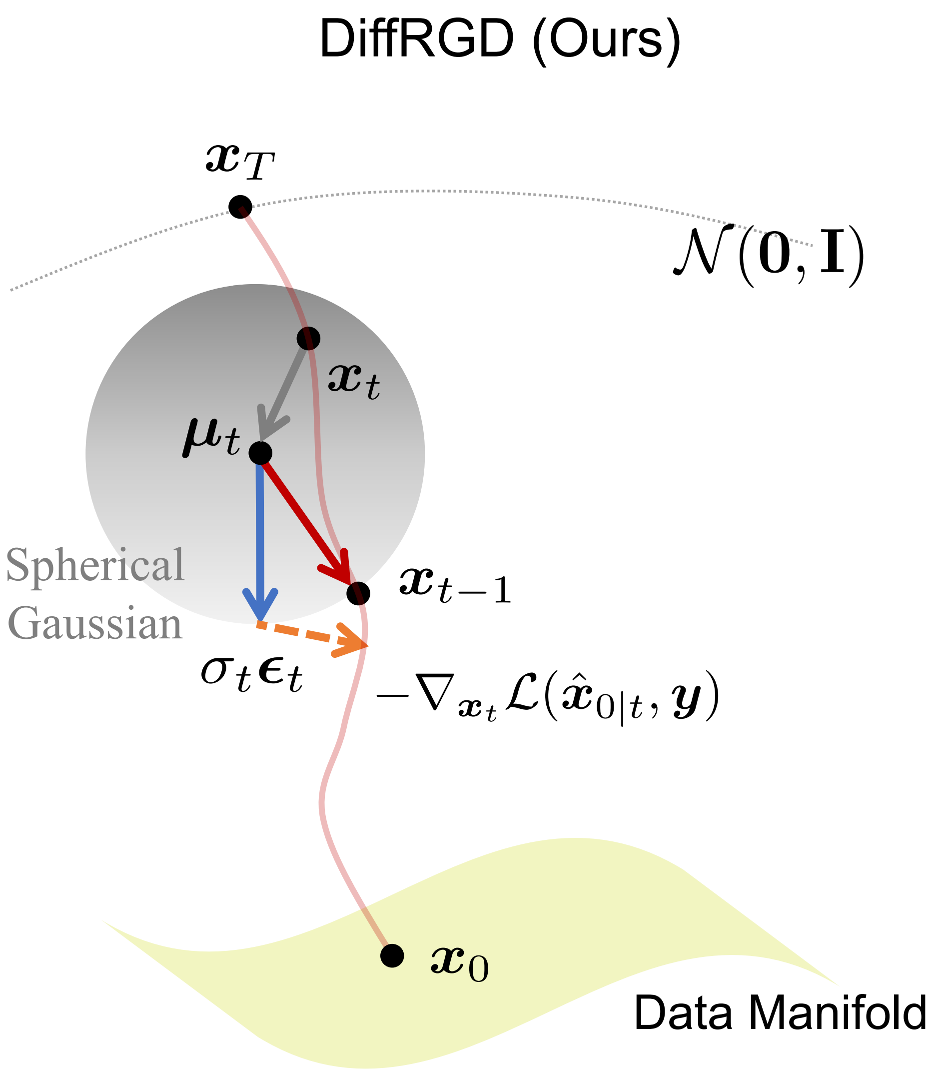
Generation $\hat{\boldsymbol{x}}$

$$\mathcal{L}_{\text{inv}}(\hat{\boldsymbol{x}}, \boldsymbol{y}) = \|\mathcal{A}\hat{\boldsymbol{x}} - \boldsymbol{y}\|_2^2$$

$$\mathcal{L}_{\text{seg}}(\hat{\boldsymbol{x}}, \boldsymbol{y}) = \|\psi(\hat{\boldsymbol{x}}) - \boldsymbol{y}\|_2^2$$

Distribution Preserving Diffusion Guidance



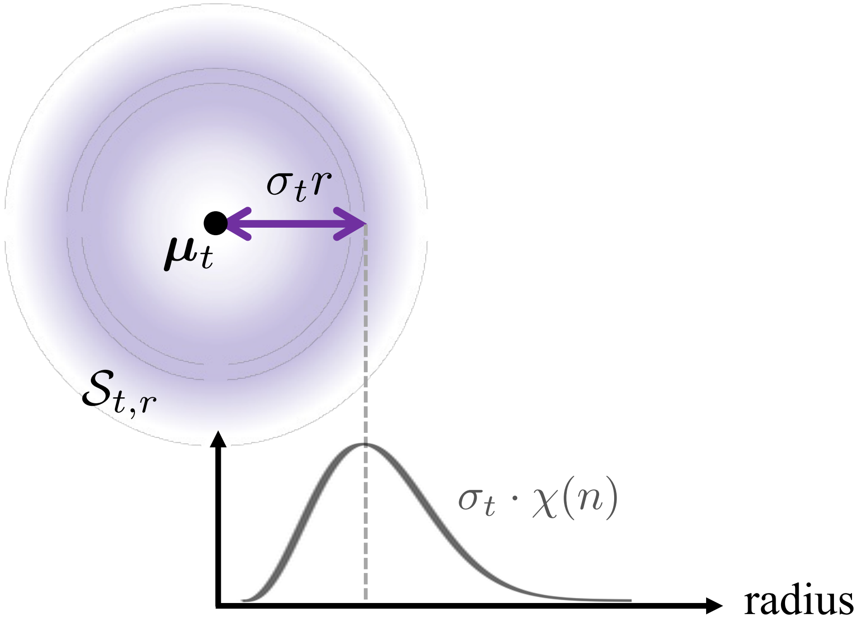


Distribution Preserving Optimization

Constructing the geometric structure

Polar Decomposition

$$\boldsymbol{x}_t \sim \mathcal{N}(\boldsymbol{\mu}_t, \sigma_t^2 \mathbf{I}_n) \iff \boldsymbol{x}_t = \boldsymbol{\mu}_t + \sigma_t r \boldsymbol{u} \qquad r \sim \chi(n), \quad \boldsymbol{u} \sim \mathrm{Unif}(\mathbb{S}^{n-1})$$



Constraint Optimization:

$$\boldsymbol{x}_{t-1} = \underset{\boldsymbol{x} \in \mathcal{S}_{t,r}}{\operatorname{argmin}} \, \mathcal{L}(\hat{\boldsymbol{x}}_0(\boldsymbol{x}, t-1), \boldsymbol{y})$$

$$\mathcal{S}_{t,r} := \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x} - \boldsymbol{\mu}_t\|_2 = \sigma_t r\}$$

Riemannian Gradient Descent

Tangent Space

$$\mathrm{T}_{\boldsymbol{x}}\,\mathcal{S}_{t,r} := \left\{\boldsymbol{v} \in \mathbb{R}^n \mid (\boldsymbol{x} - \boldsymbol{\mu}_t)^\top \boldsymbol{v} = 0\right\}$$

Projection Operator

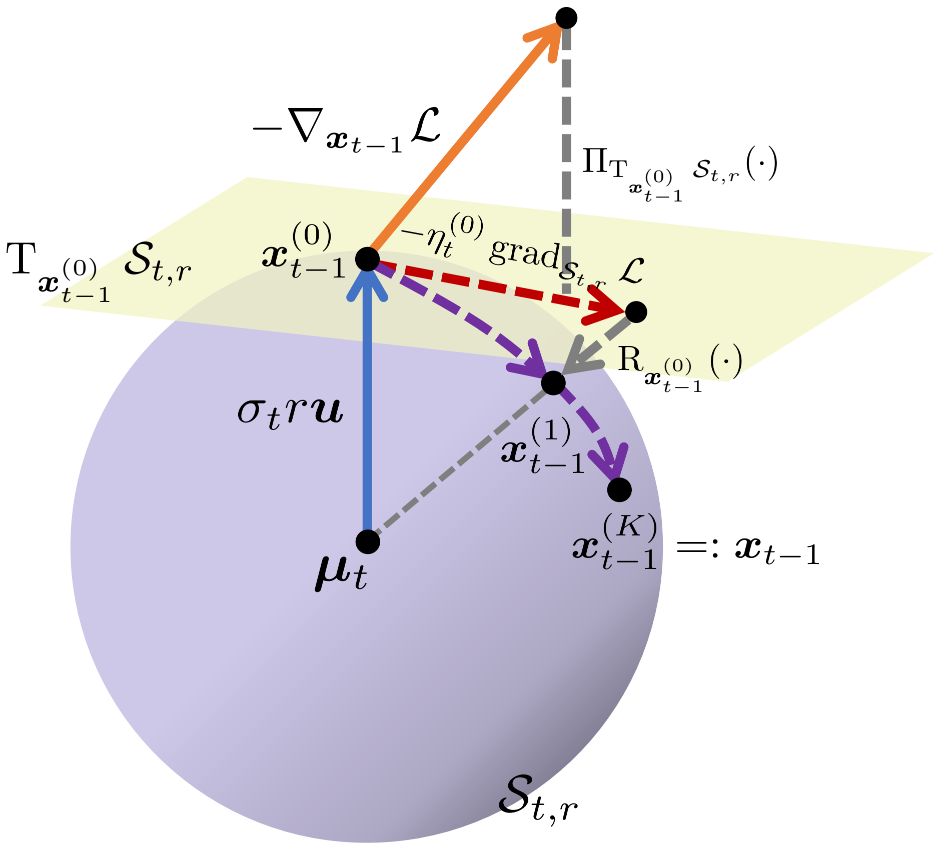
$$\Pi_{\mathrm{T}_{\boldsymbol{x}}\,\mathcal{S}_{t,r}}(\boldsymbol{v}) := \left(\mathbf{I}_n - \frac{(\boldsymbol{x} - \boldsymbol{\mu}_t)(\boldsymbol{x} - \boldsymbol{\mu}_t)^\top}{\|\boldsymbol{x} - \boldsymbol{\mu}_t\|_2^2}\right)\boldsymbol{v}$$

Retraction

$$\mathrm{R}_{\boldsymbol{x}}(\boldsymbol{v}) := \boldsymbol{\mu}_t + \sigma_t r \cdot \frac{\boldsymbol{x} - \boldsymbol{\mu}_t + \boldsymbol{v}}{\|\boldsymbol{x} - \boldsymbol{\mu}_t + \boldsymbol{v}\|_2}$$

Riemannian Gradient

$$\mathrm{grad}_{\mathcal{S}_{t,r}}\,\mathcal{L} := \Pi_{\mathrm{T}_{\boldsymbol{x}_{t-1}^{(k)}}\,\mathcal{S}_{t,r}}\left(\nabla_{\boldsymbol{x}_{t-1}^{(k)}}\mathcal{L}(\hat{\boldsymbol{x}}_0(\boldsymbol{x}_{t-1}^{(k)}, t-1), \boldsymbol{y})\right)$$

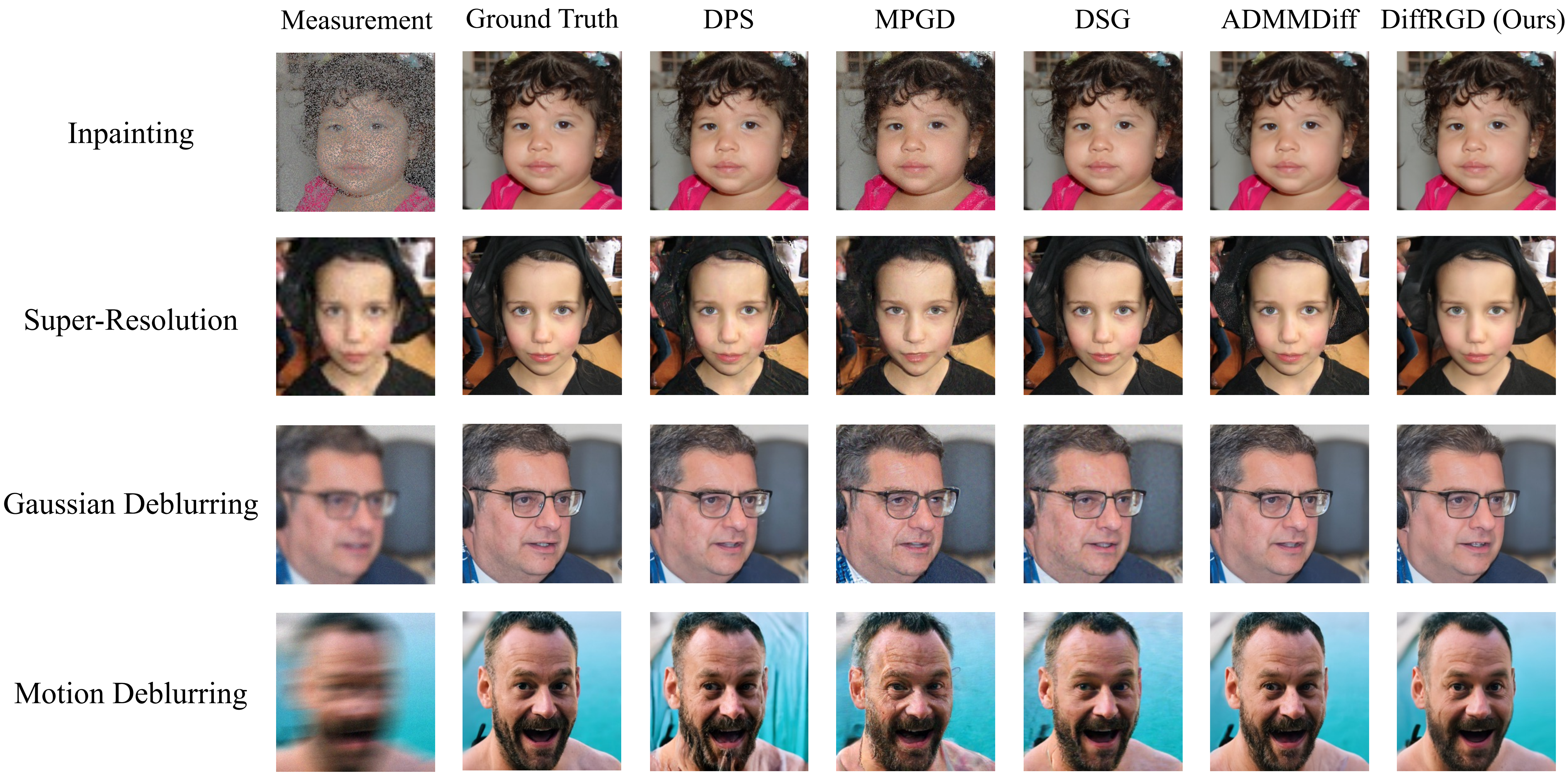


$$\boldsymbol{x}_{t-1}^{(k+1)} = \mathrm{R}_{\boldsymbol{x}_{t-1}^{(k)}}\left(-\eta_t^{(k)}\,\mathrm{grad}_{\mathcal{S}_{t,r}}\,\mathcal{L}\right)$$

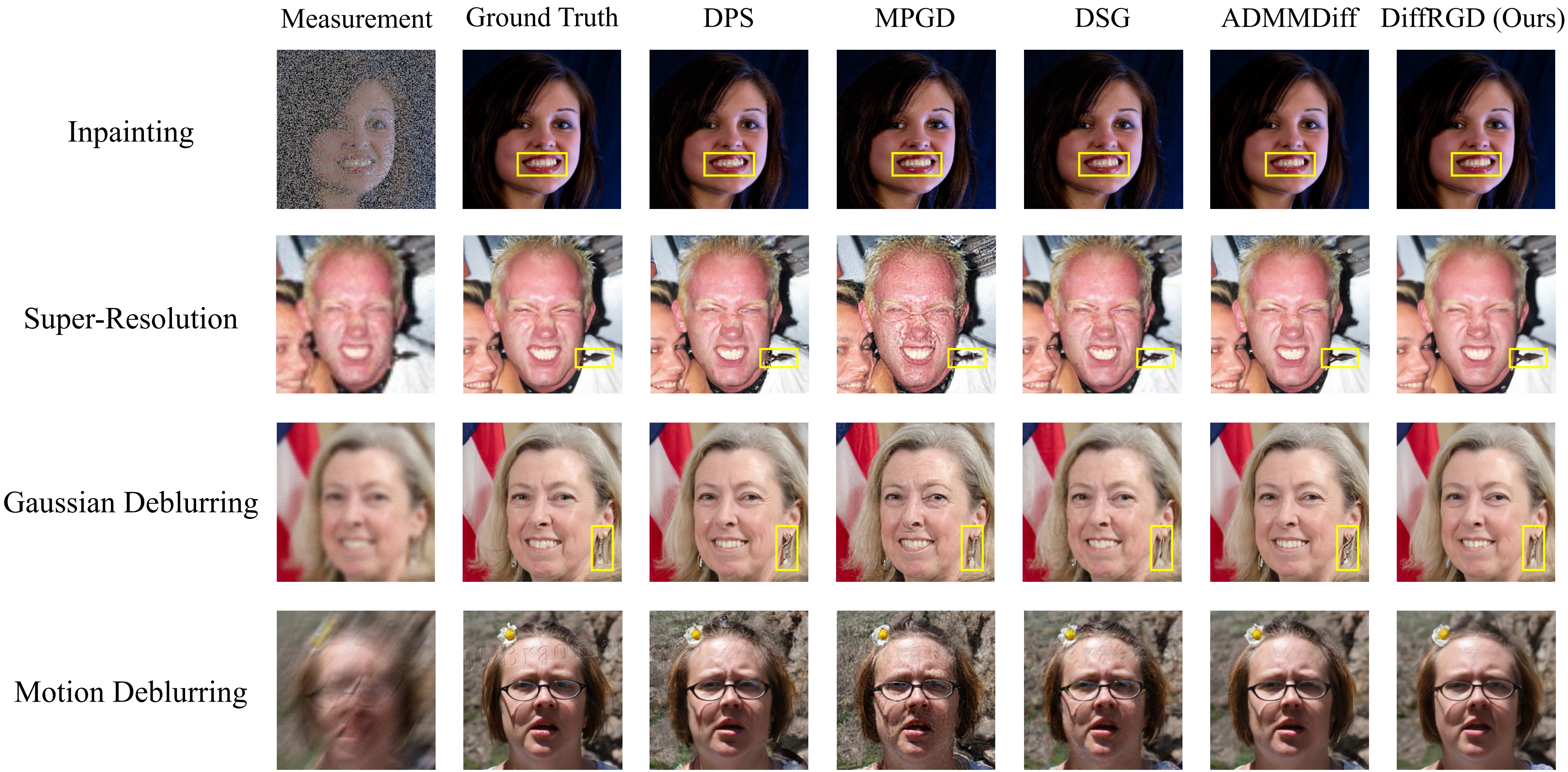
Experiments: Image Inverse Problems

Tasks	Methods	Venue	PSNR ↑	SSIM ↑	LPIPS ↓	FID ↓
Inpainting	DPS [5]	ICLR 2023	30.44	0.863	0.153	42.68
	MPGD [13]	ICLR 2024	27.51	0.724	0.256	68.24
	DSG [51]	ICML 2024	31.03	0.866	0.144	36.30
	ADMMDiff [56]	CVPR 2025	<u>32.38</u>	<u>0.899</u>	<u>0.119</u>	<u>29.52</u>
	DiffRGD (Ours)	-	**34.04***	**0.926***	**0.096***	**21.88***
Super-Resolution 4×	DPS [5]	ICLR 2023	26.03	0.727	0.260	80.05
	MPGD [13]	ICLR 2024	24.40	0.614	0.354	101.50
	DSG [51]	ICML 2024	<u>26.71</u>	<u>0.737</u>	<u>0.256</u>	<u>74.67</u>
	ADMMDiff [56]	CVPR 2025	26.48	0.712	0.297	96.69
	DiffRGD (Ours)	-	**27.77***	**0.783***	**0.220***	**63.94***
Gaussian Deblurring	DPS [5]	ICLR 2023	25.88	0.721	0.237	69.38
	MPGD [13]	ICLR 2024	24.07	0.576	0.328	95.12
	DSG [51]	ICML 2024	**27.45**	0.751	0.259	<u>75.85</u>
	ADMMDiff [56]	CVPR 2025	26.57	**0.757**	<u>0.226</u>	79.30
	DiffRGD (Ours)	-	<u>26.80</u>	**0.757**	**0.218***	**63.83***
Motion Deblurring	DPS [5]	ICLR 2023	24.47	0.685	0.271	80.75
	MPGD [13]	ICLR 2024	23.15	0.569	0.357	106.99
	DSG [51]	ICML 2024	<u>26.80</u>	0.709	0.290	87.96
	ADMMDiff [56]	CVPR 2025	**27.26**	**0.778**	**0.222**	72.92
	DiffRGD (Ours)	-	25.84	<u>0.736</u>	<u>0.250</u>	**72.26**

Experiments: Image Inverse Problems

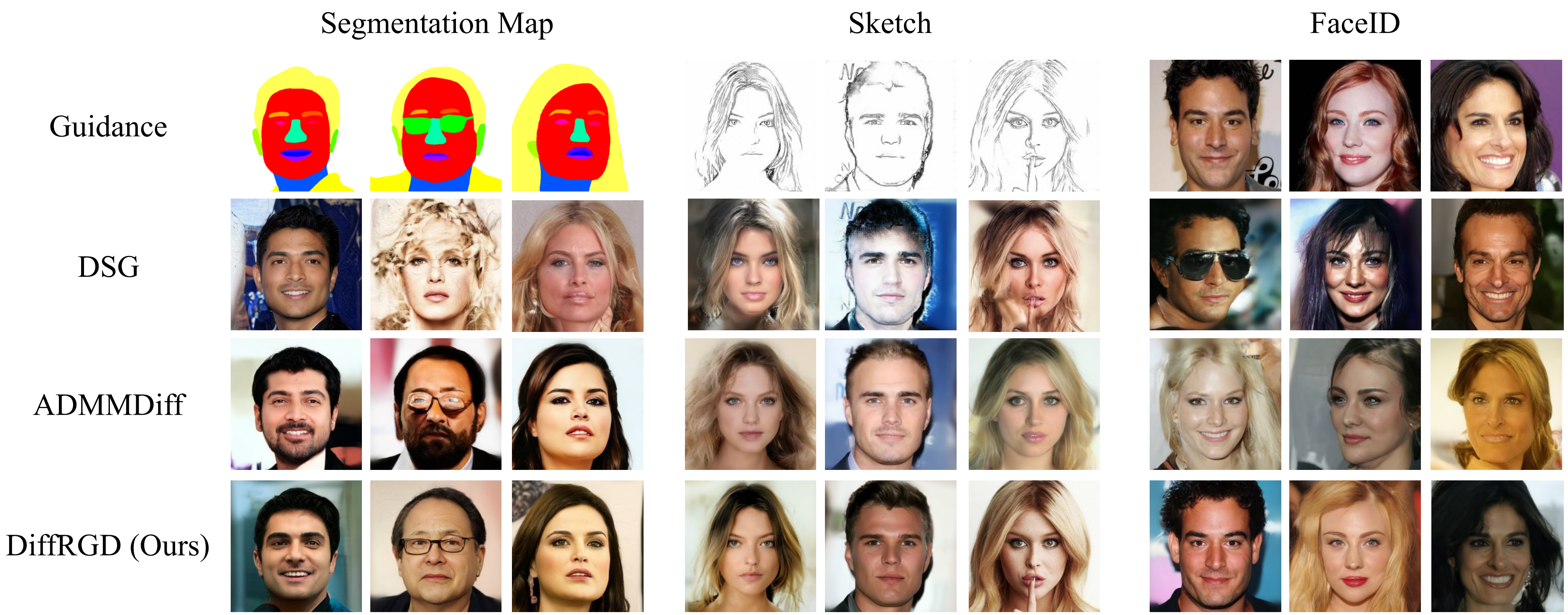


Experiments: Image Inverse Problems

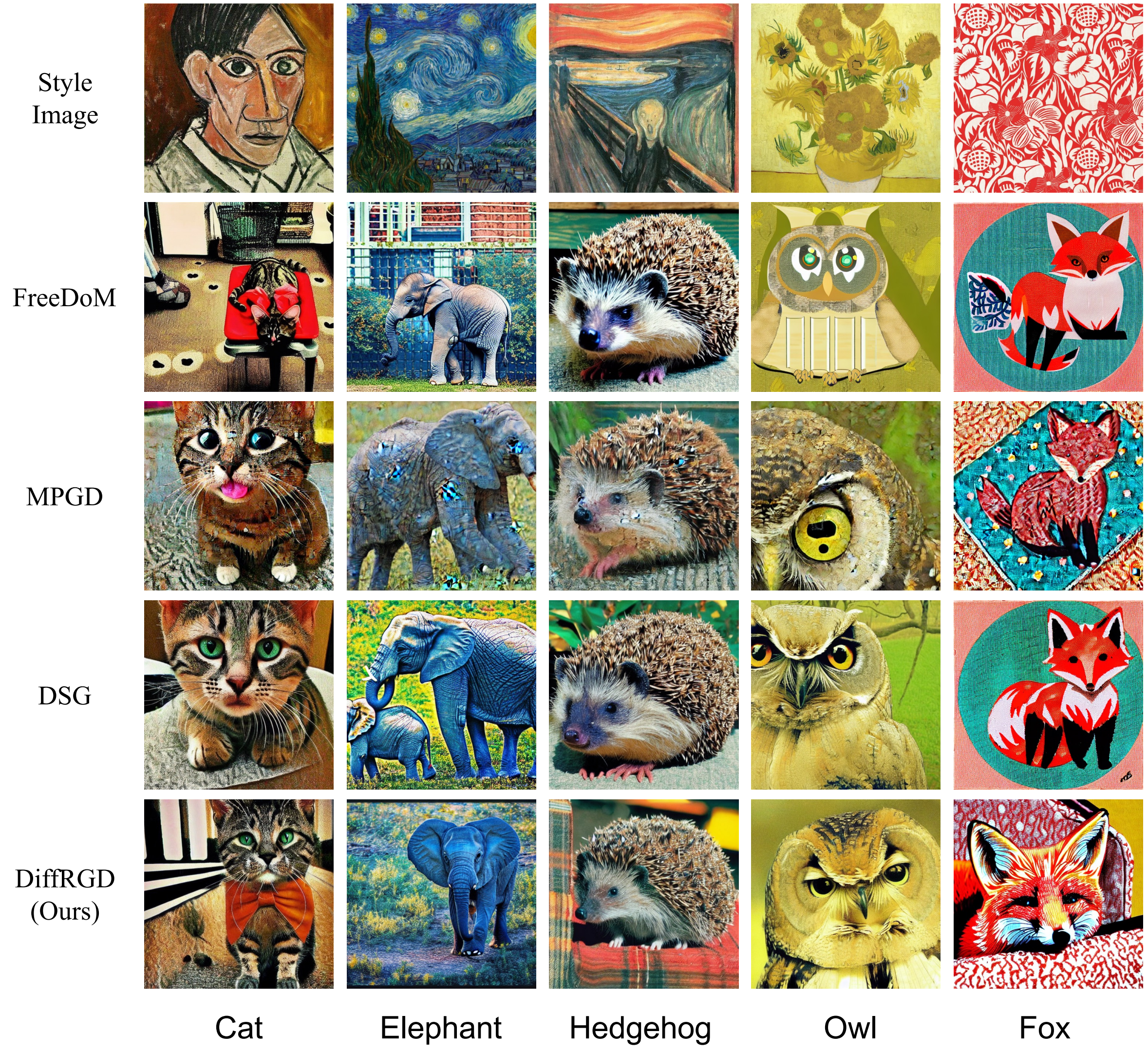


Experiments: Conditional Generation

Methods	Segmentation Map			Sketch			FaceID		
	mIoU ↑	FID ↓	KID ↓	Sketch-ℓ_2 ↓	FID ↓	KID ↓	FaceID-ℓ_2 ↓	FID ↓	KID ↓
FreeDoM [53]	0.622	156.02	0.057	30.85	101.90	0.017	0.557	127.05	0.032
DSG [51]	0.750	117.48	<u>0.028</u>	<u>21.36</u>	107.00	0.024	<u>0.340</u>	<u>95.27</u>	<u>0.014</u>
ADMMDiff [56]	<u>0.758</u>	<u>101.86</u>	0.031	30.82	<u>97.52</u>	<u>0.014</u>	0.346	100.81	<u>0.014</u>
DiffRGD (Ours)	**0.804***	**96.10***	**0.026**	**19.48***	**87.82***	**0.012**	**0.303***	**93.80**	**0.011**



Experiments: Style-Guided Generation



DiffQRCoder: Diffusion-based Aesthetic QR Code Generation with Scanning Robustness Guided Iterative Refinement

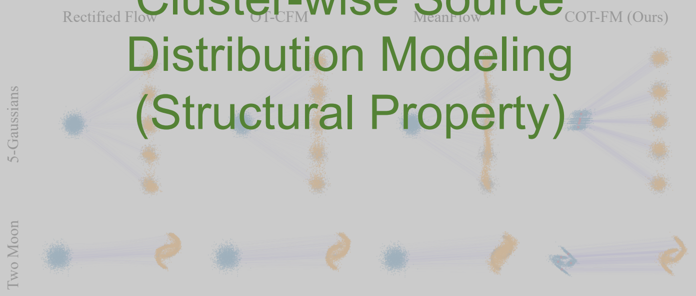
WACV 2025

Applications of Diffusion Guidance (Controllability)

COT-FM: Cluster-wise Optimal Transport Flow Matching

CVPR 2026

Cluster-wise Source Distribution Modeling (Structural Property)



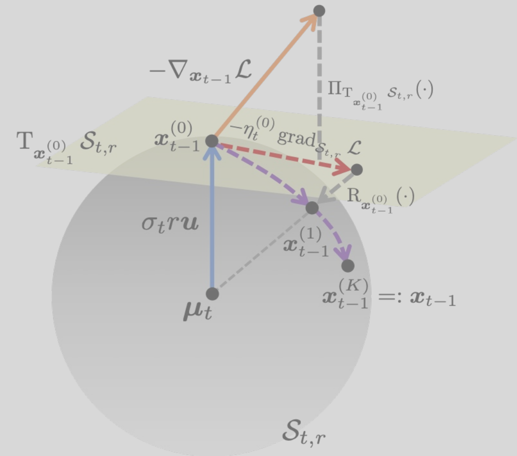
Pixel Is Not A Barrier: An Effective Evasion Attack for Pixel-Domain Diffusion Models

AAAI 2025

Adversarial Attack (Safety)

DiffRGD: An Inference-Time Diffusion Guidance Through Riemannian Gradient Descent

Submitted to ECCV 2026



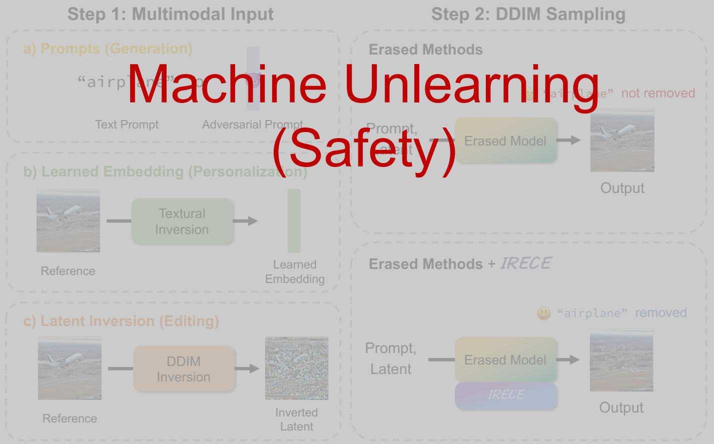
CNP-Flow: Unified Temporal Flow Matching via Conditional Noise Predictor

Submitted to NeurIPS 2026

M-ErasureBench: A Comprehensive Multimodal Evaluation Benchmark for Concept Erasure in Diffusion Models

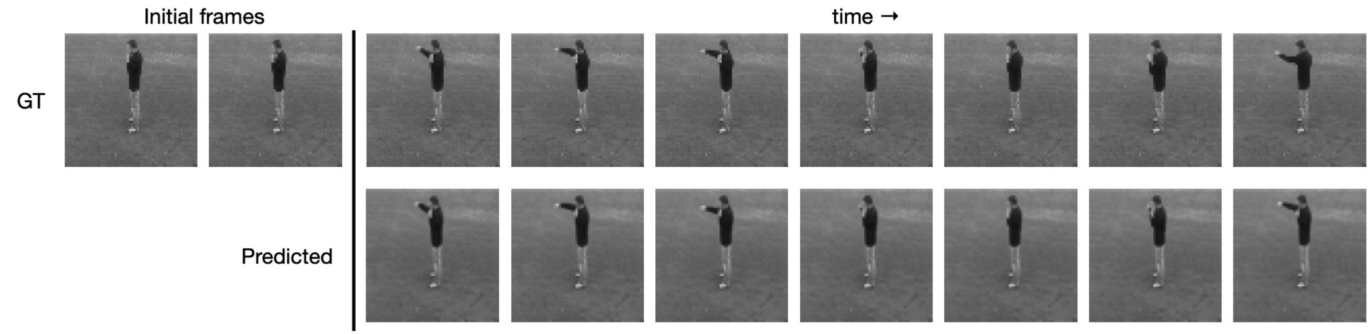
WACV 2026

Machine Unlearning (Safety)

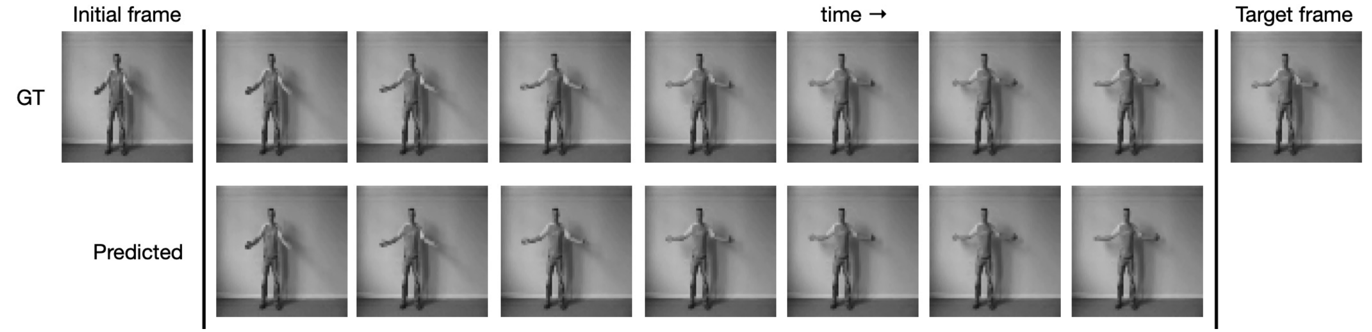


Temporal Generation

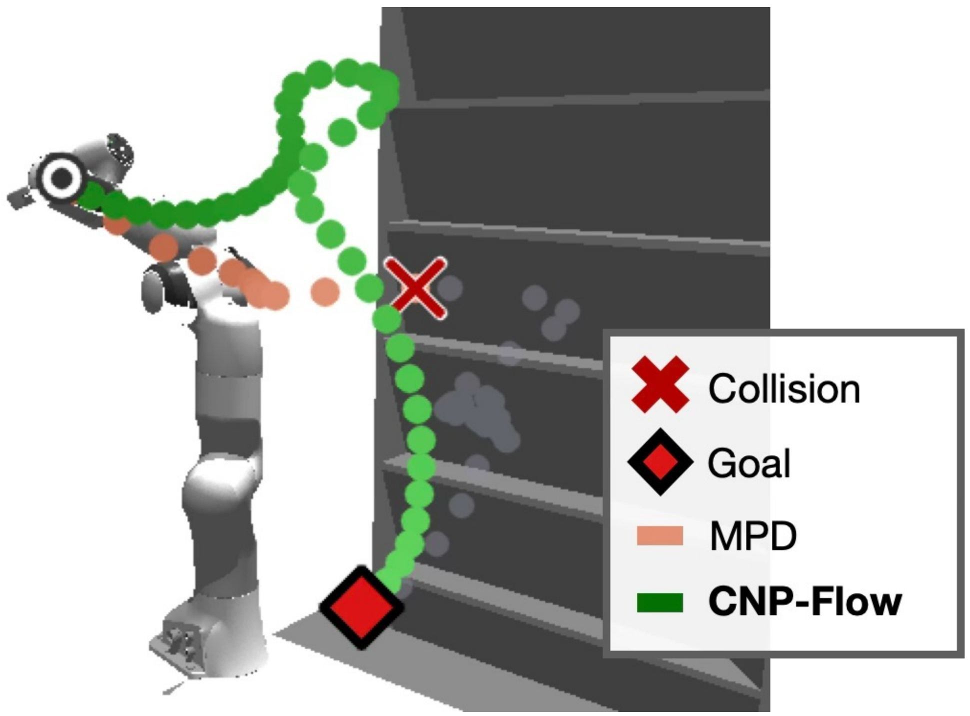
Video Prediction



Video Interpolation



Motion Planning

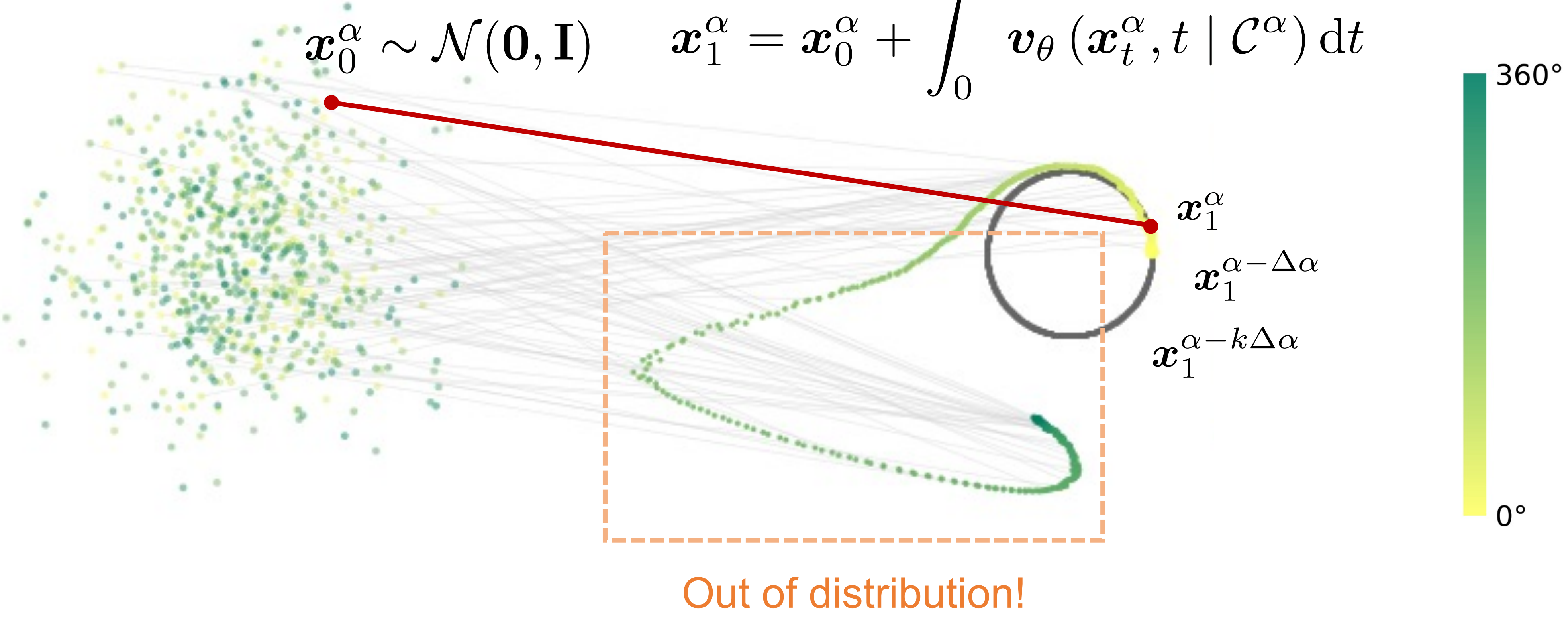


2D Toy Example: Temporal Phase Generation

Standard FM randomly pair Gaussian noise samples with temporal data, resulting in chaotic velocity fields that hinder the learning of smooth temporal dynamics

$$\mathcal{C}^{\alpha} = [\boldsymbol{x}_1^{\alpha-\Delta\alpha}, \boldsymbol{x}_1^{\alpha-k\Delta\alpha}]$$

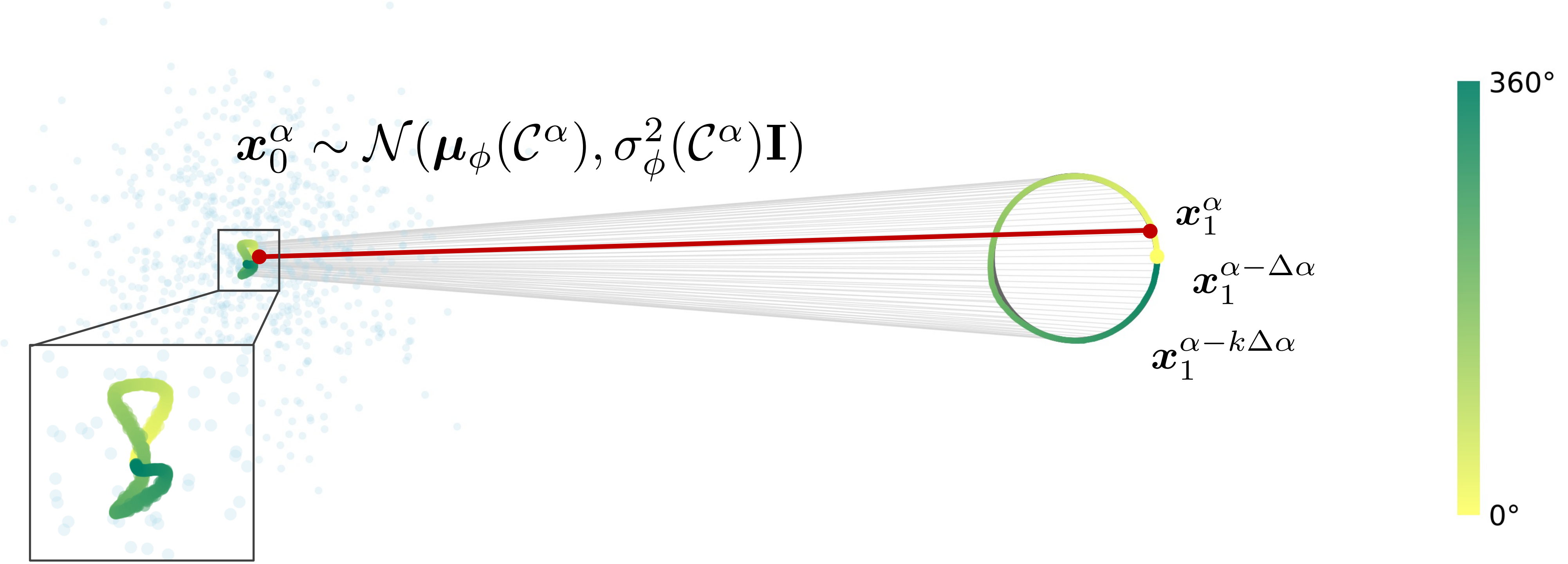
$$\boldsymbol{x}_0^{\alpha} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \qquad \boldsymbol{x}_1^{\alpha} = \boldsymbol{x}_0^{\alpha} + \int_0^1 \boldsymbol{v}_{\theta}\left(\boldsymbol{x}_t^{\alpha}, t \mid \mathcal{C}^{\alpha}\right) \mathrm{d}t$$



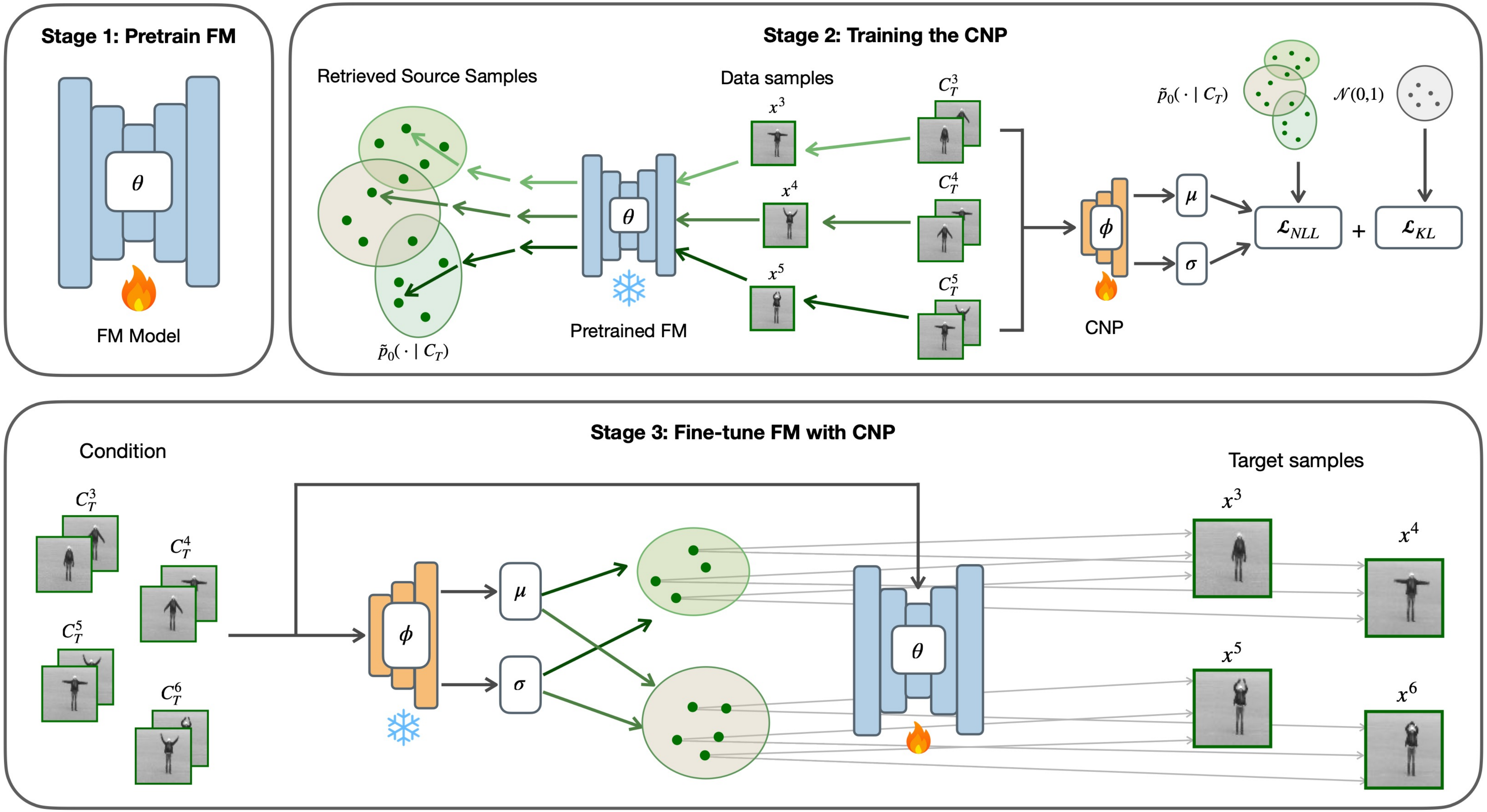
2D Toy Example: Temporal Phase Generation

Goal: Learn the Conditional Noise Predictor (CNP) to

$$\mathcal{C}^{\alpha} = [\boldsymbol{x}_1^{\alpha-\Delta\alpha}, \boldsymbol{x}_1^{\alpha-k\Delta\alpha}]$$

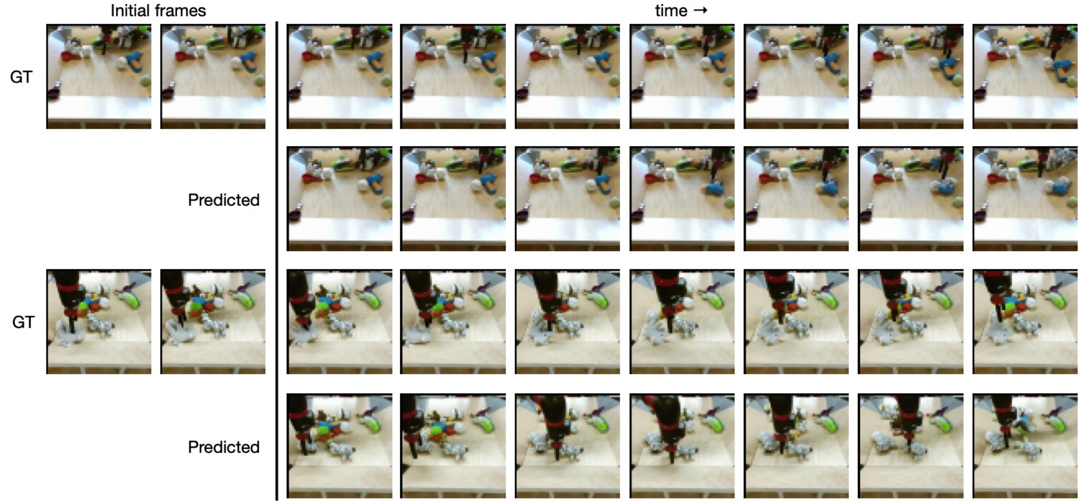


Training the CNP-Flow

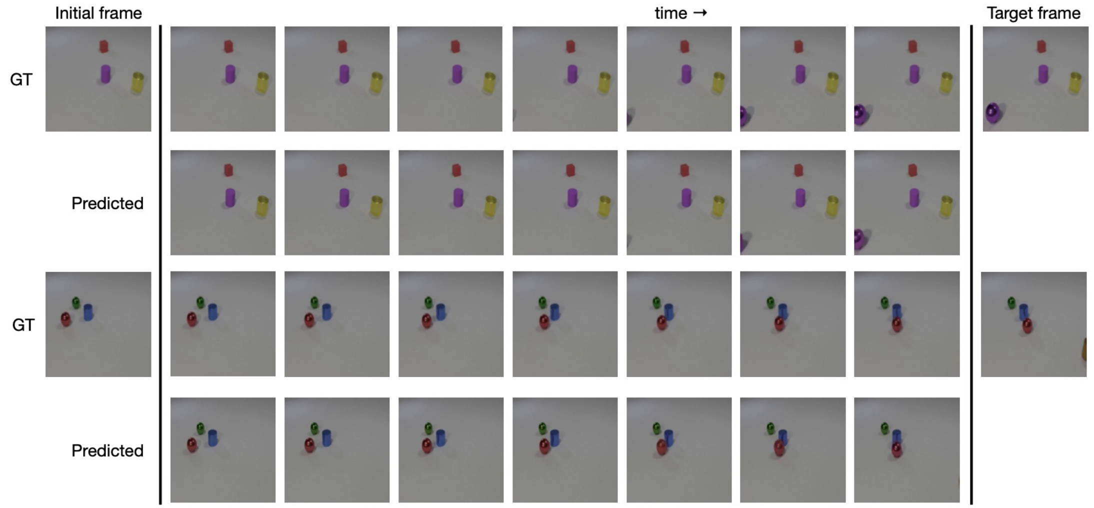


Experiments

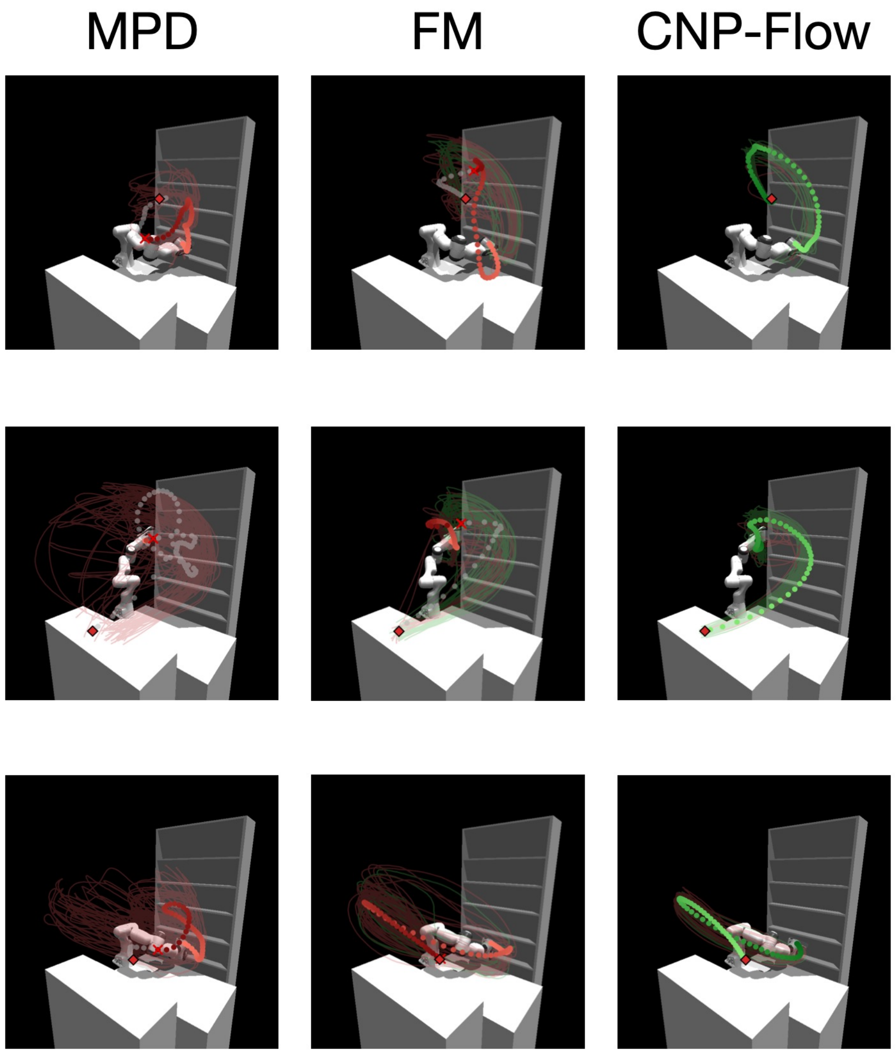
Video Prediction



Video Interpolation



Motion Planning



Takeaways

My Research = Controllability x Structural Properties x Safety

- Creating data from noise is generative modeling
- Diffusion guidance should preserve the data distribution
- Modeling *source distributions* improves the generation task

Collaborators


Chun-Yen Shih


Ju-Hsun Weng


Winston Wang

Tzu-Sian Wang


Ernie Chu


Li-Xuan Peng


Chien-Sheng Chiang


Kuan-Hsun Tu


Hsuan-Chi Lu


Cheng-Fu Chou


Tsung-Wei Ke


Jun-Cheng Chen


Mei-Heng Yueh


Min Sun

Thank you!

JWのAI

@jwliao1209 • 670 subscribers • 80 videos

More about this channel **...more**

Subscribe