ALMOST NEARLY PERFECT

Bayesian inversion with physics-informed deep generative models

APPLICATIONS TO COMPUTATIONAL IMAGING

PROF. MARCELO PEREYRA

Heriot-Watt University & Maxwell Institute for Mathematical Sciences

2026

OUTLINE

01 BACKGROUND

02 LATINO-PRO: LAtent consisTency INverse sOlver with PRompt Optimization

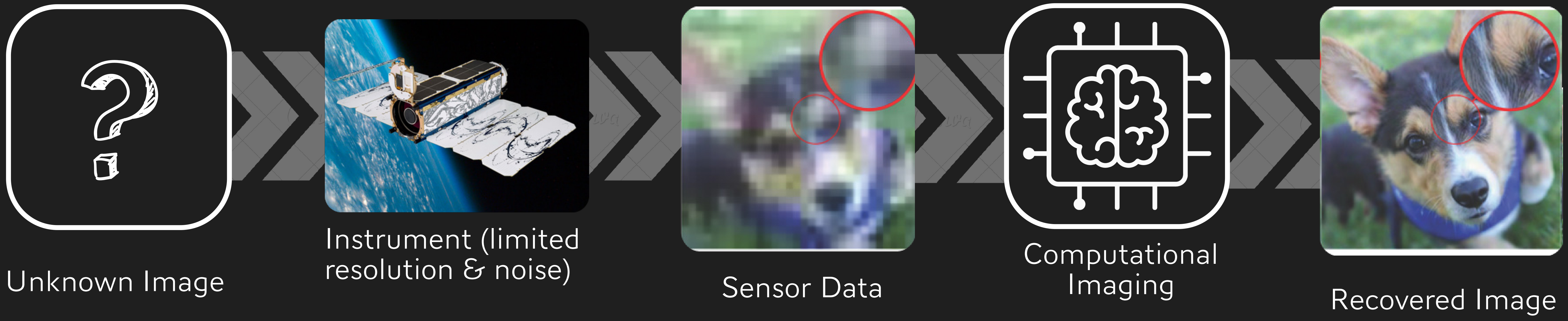
03 Learning few-step posterior samplers by unfolding and distillation

04 Conclusion

BACKGROUND

Problem statement:

Image data not useful in raw form (limited resolution, noise, blur..)



Research vision:

Smart computational imaging instruments through integrated physical & gen. AI models, Bayesian statistics and fast stochastic algorithms.

BAYESIAN STATISTICAL FRAMEWORK

We seek to perform inference on $x^\star \in \mathbb{R}^d$

from some data $y = Ax^\star + w$

We model $x^\star$ as a realisation of a r.v. $\mathbb{x}$

We model y as a realisation of a r.v. $(\mathbb{y}|\mathbb{x} = x^\star)$

We base our inferences about $\mathbb{x}$ on the posterior distribution

Physical model → $p(y|x)$; Statistical image model → $p(x)$

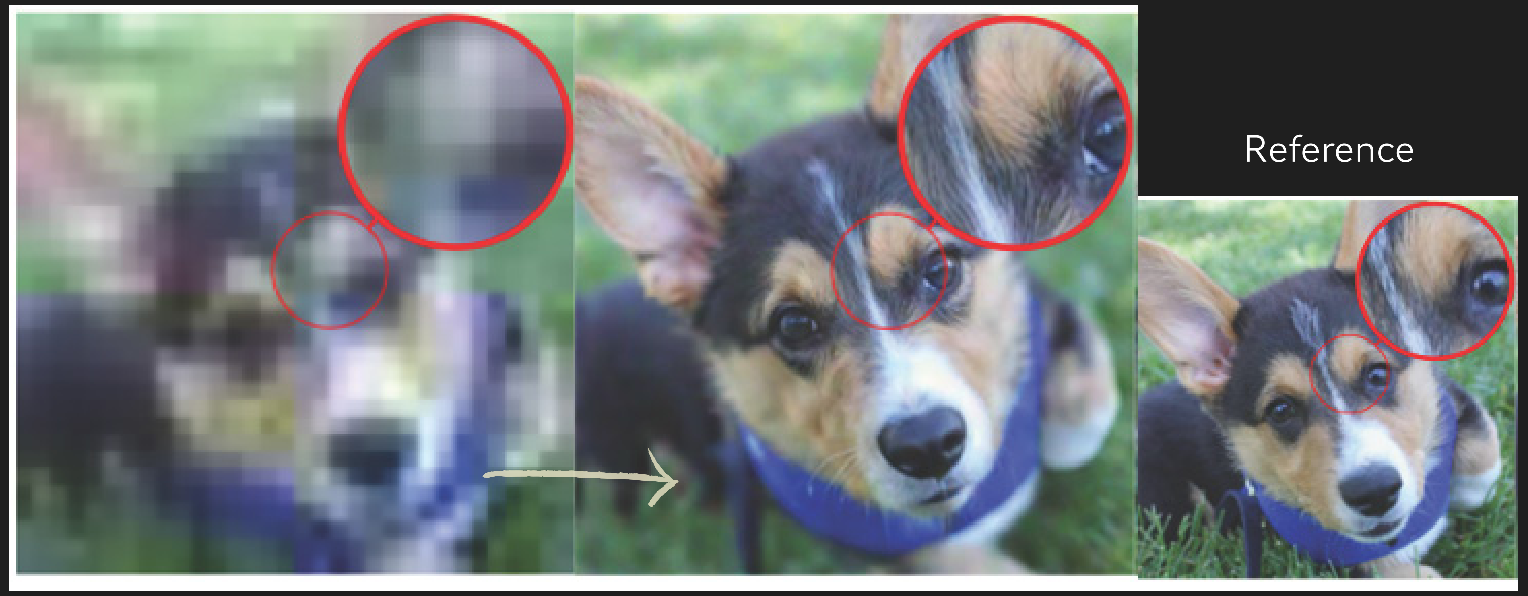
$$p(x|y) = \frac{p(y|x)p(x)}{\int_{\mathbb{R}^d} p(y|\tilde{x})p(\tilde{x})\mathrm{d}\tilde{x}}$$

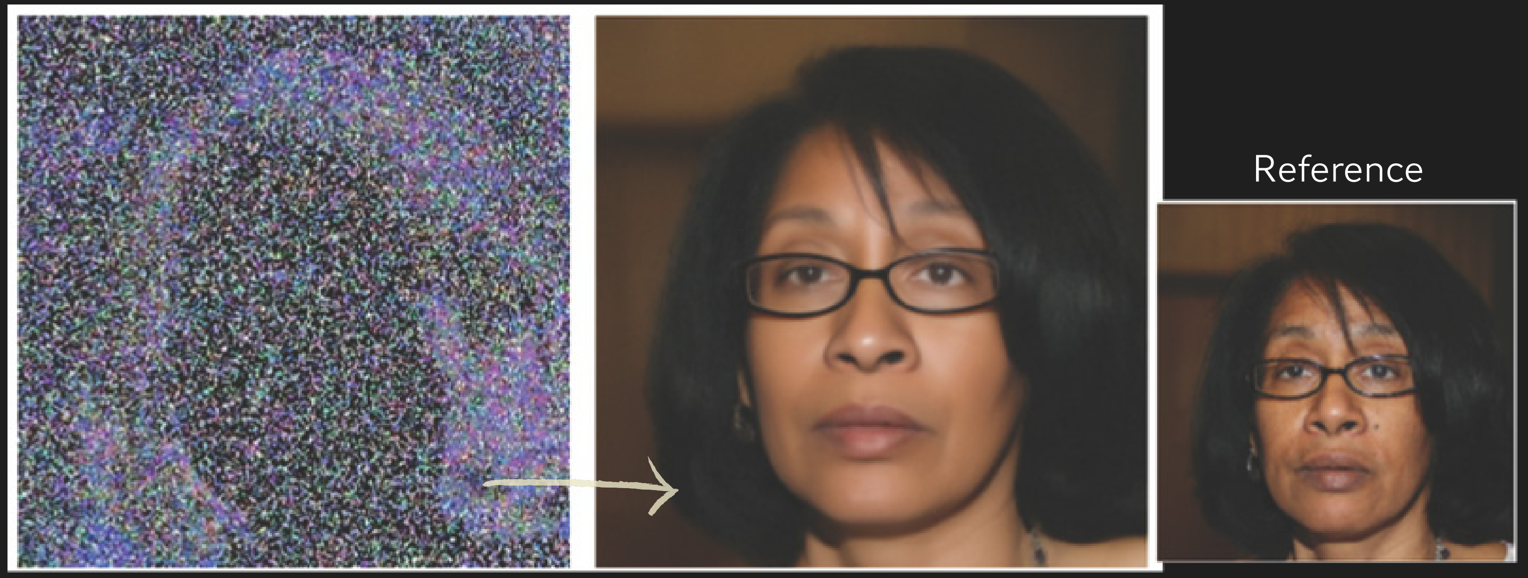
TODAY'S TALK

How modifying the mathematics of VLMs allows prompting w. physical model & measurement, while self-adjusting text prompt, transparently and with few NFEs.

Statistical image model

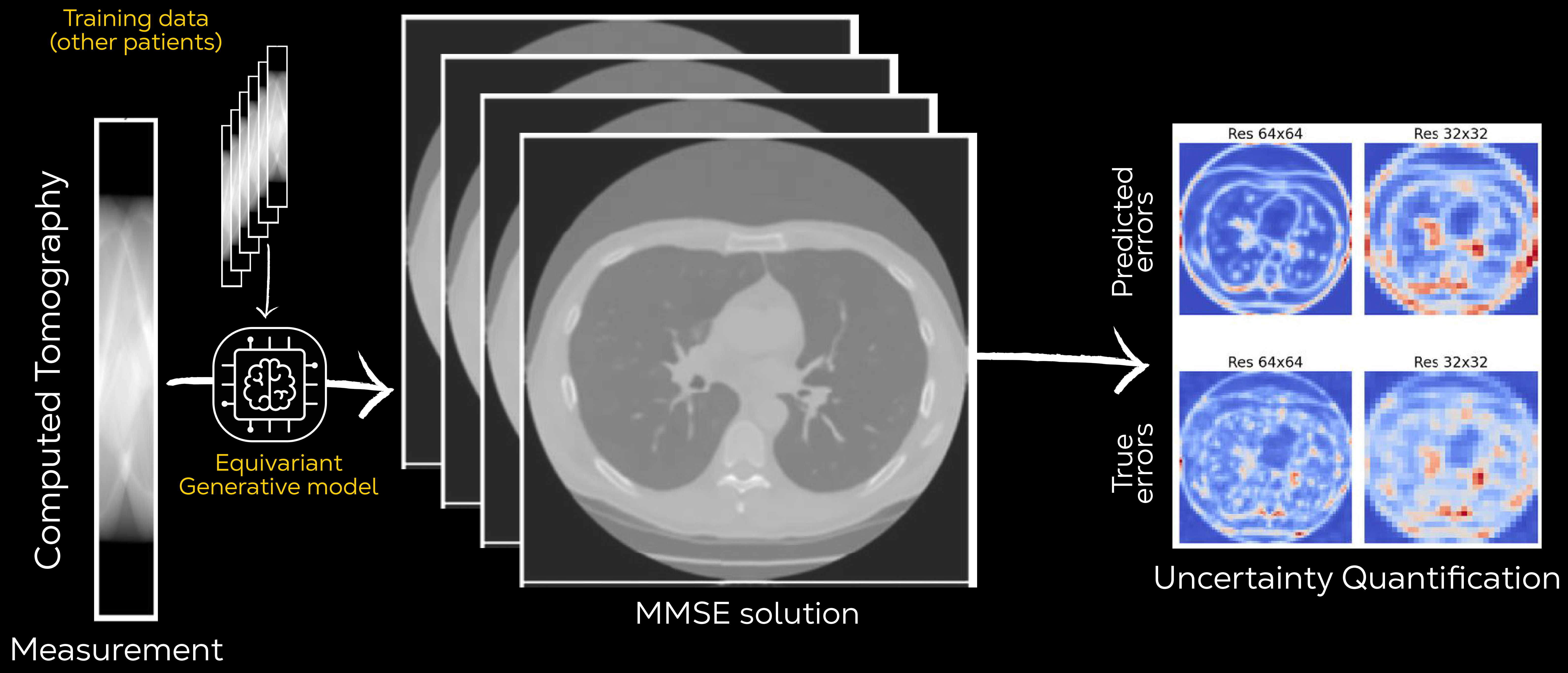
Sample from text2image generative model (Midjourney)



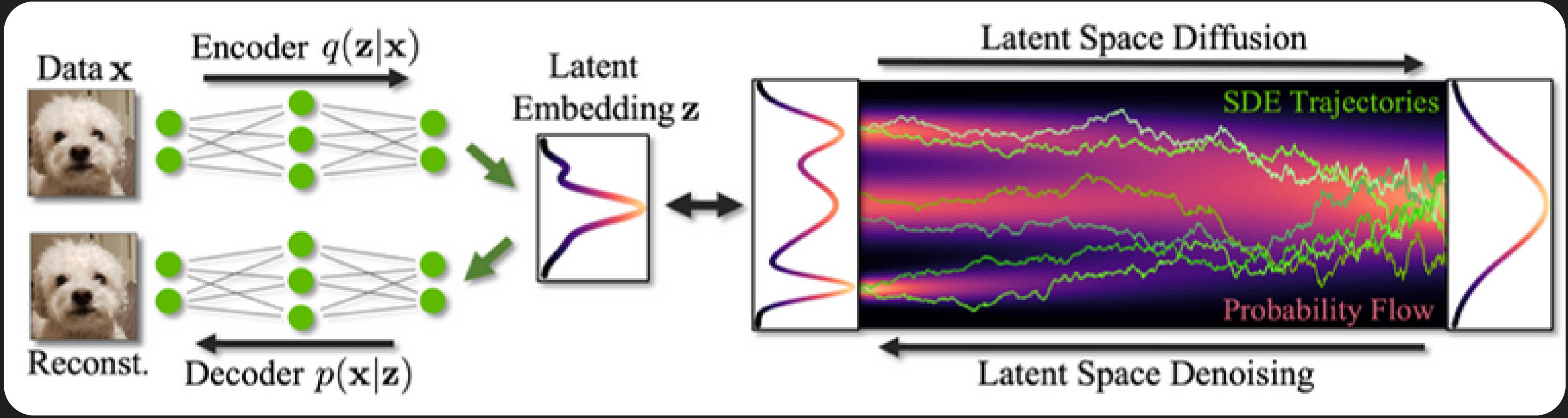


NOT TODAY

Self-supervised Gen-AI-based imaging: Leverage symmetries and invariance properties to learn empirical Bayesian models directly from the measurements - no reference/clean/labelled data required!



LATENT DIFFUSION MODELS



Kreis et al. NeurIPS 2023, LDMs Tutorial

$$d\mathbf{z}_t = -\frac{\beta_t}{2}\mathbf{z}_t dt + \sqrt{\beta_t} d\mathbf{w},$$

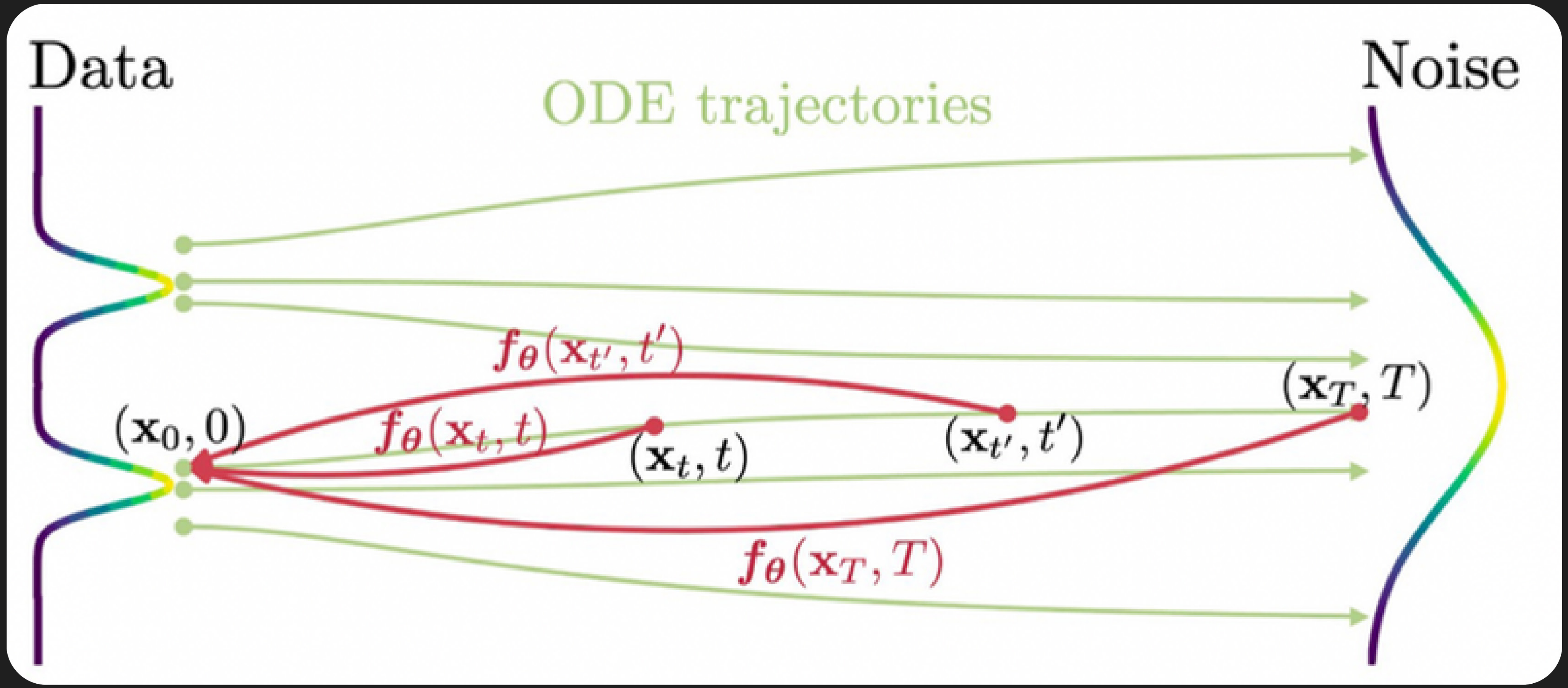
$$d\mathbf{z}_t = \left[-\frac{\beta_t}{2}\mathbf{z}_t - \beta_t \nabla_{\mathbf{z}_t} \log p_t(\mathbf{z}_t)\right] dt + \sqrt{\beta_t} d\bar{\mathbf{w}},$$

$$\mathcal{E} : \mathbb{R}^n \mapsto \mathbb{R}^d, \quad \mathcal{D} : \mathbb{R}^d \mapsto \mathbb{R}^n, \quad \mathbf{x} \approx \mathcal{D}(\mathcal{E}(\mathbf{x})),$$

arXiv:2405.14867

ACCELERATION

Probability Flow ODE, Distillation & Consistency Models (CMs)



arXiv:2303.01469

CMs are distilled diffusion models trained to transport any point on the ODE trajectory back to time 0. They are fast one-step samplers.

OUTLINE

01 BACKGROUND

02 LATINO-PRO: LAtent consisTency INverse sOlver with PRompt Optimization

03 Learning few-step posterior samplers by unfolding and distillation

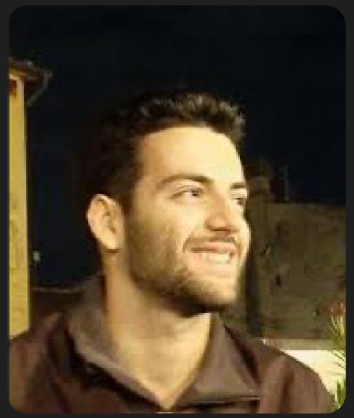
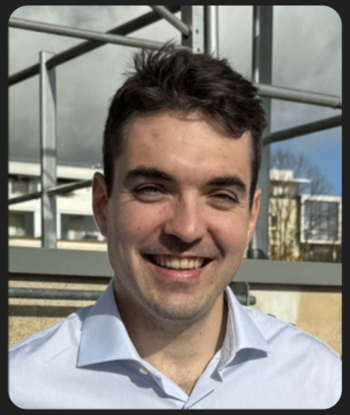
04 Conclusion

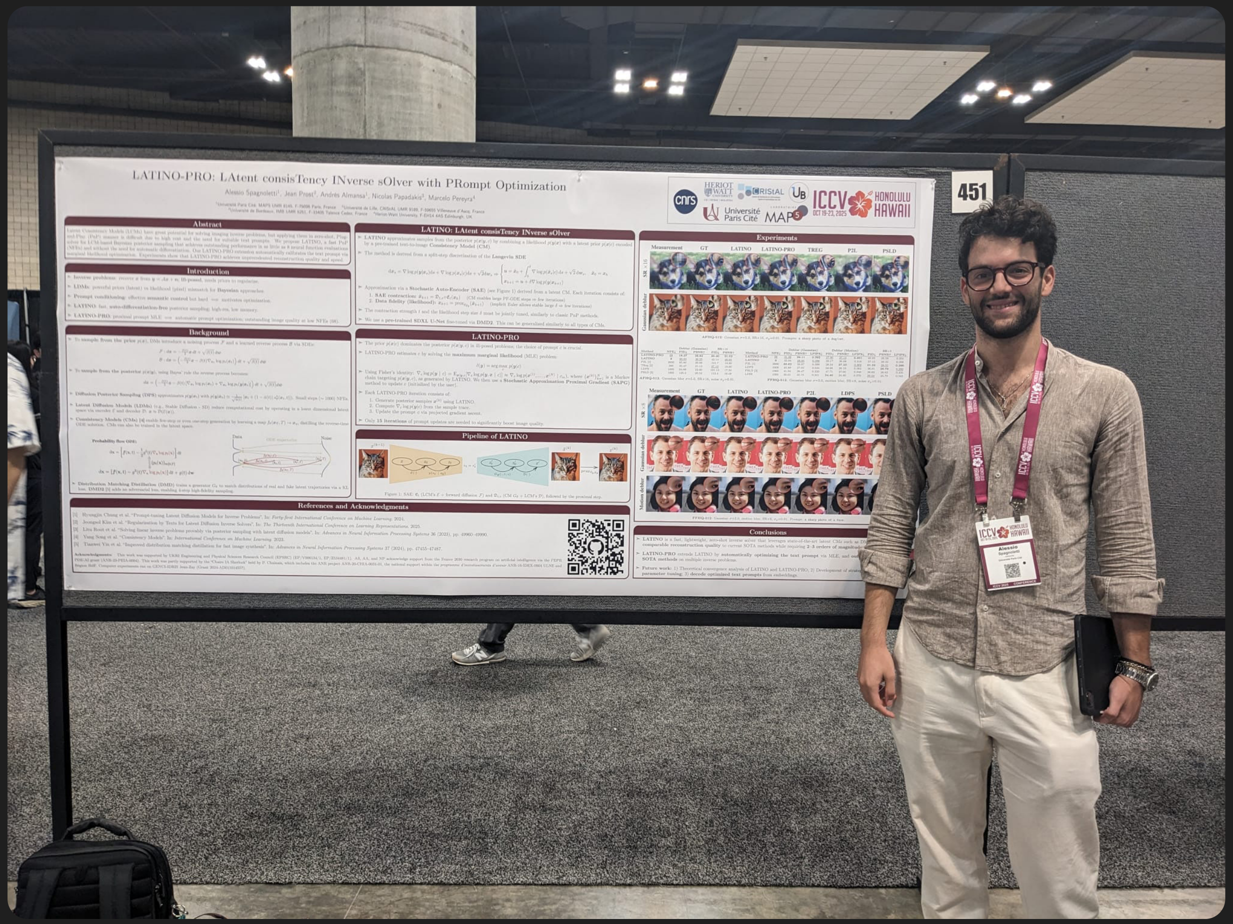
LATINO-PRO

LAtent consisTency INverse sOlver with PRompt Optimization

ICCV 2025

Joint work with Alessio Spagnoletti, Jean Prost, Andrés Almansa, & Nicolas Papadakis



LVTINO: LAtent Video consisTency INverse sOlver for High Definition Video Restoration

ICLR 2026

BAYESIAN STATISTICAL FRAMEWORK

We seek to perform inference on $x^\star \in \mathbb{R}^d$

from some data $y = Ax^\star + w$

We model $x^\star$ as a realisation of a r.v. x

We model y as a realisation of a r.v. $(\mathrm{y}|\mathrm{x} = x^\star)$

We base our inferences about x on the posterior distribution

$$p(x|y) = \frac{p(y|x)p(x)}{\int_{\mathbb{R}^d} p(y|\tilde{x})p(\tilde{x})\mathrm{d}\tilde{x}}$$

LANGEVIN SAMPLING

We sample the posterior by using the following diffusion process

text prompt

$$\mathrm{d}\boldsymbol{x}_s = \nabla \log p(\boldsymbol{y}|\boldsymbol{x}_s)\mathrm{d}s + \nabla \log p(\boldsymbol{x}_s|c)\mathrm{d}s + \sqrt{2}\mathrm{d}\boldsymbol{w}_s$$

data likelihood — image prior

Converges exponentially fast to $p(x|y)$ as time **s** increases.

Modular and explainable - clear data fidelity and regularization terms.

The *drift* is time-homogeneous, no need to approximate likelihoods.

Challenge: how to embed VLM priors (eg SDXL) within a Langevin process?

CORE IDEA 1

We propose to discretize the Langevin process as follows

$$\boldsymbol{u} = \boldsymbol{x}_k + \int_0^{\delta} \nabla \log p(\tilde{\boldsymbol{x}}_s|c)\mathrm{d}s + \sqrt{2}\mathrm{d}\boldsymbol{w}_s\,,\ \tilde{\boldsymbol{x}}_0 = \boldsymbol{x}_k\,,$$

$$\boldsymbol{x}_{k+1} = \boldsymbol{u} + \delta\nabla \log p(\boldsymbol{y}|\boldsymbol{x}_{k+1})\,,$$

image prior

data likelihood

The top line corresponds to a Langevin process targeting the prior.

The bottom line is an implicit or *proximal* step, exactly solvable.

Insight: can replace top line with other Markov kernels that contract random variables towards the prior.

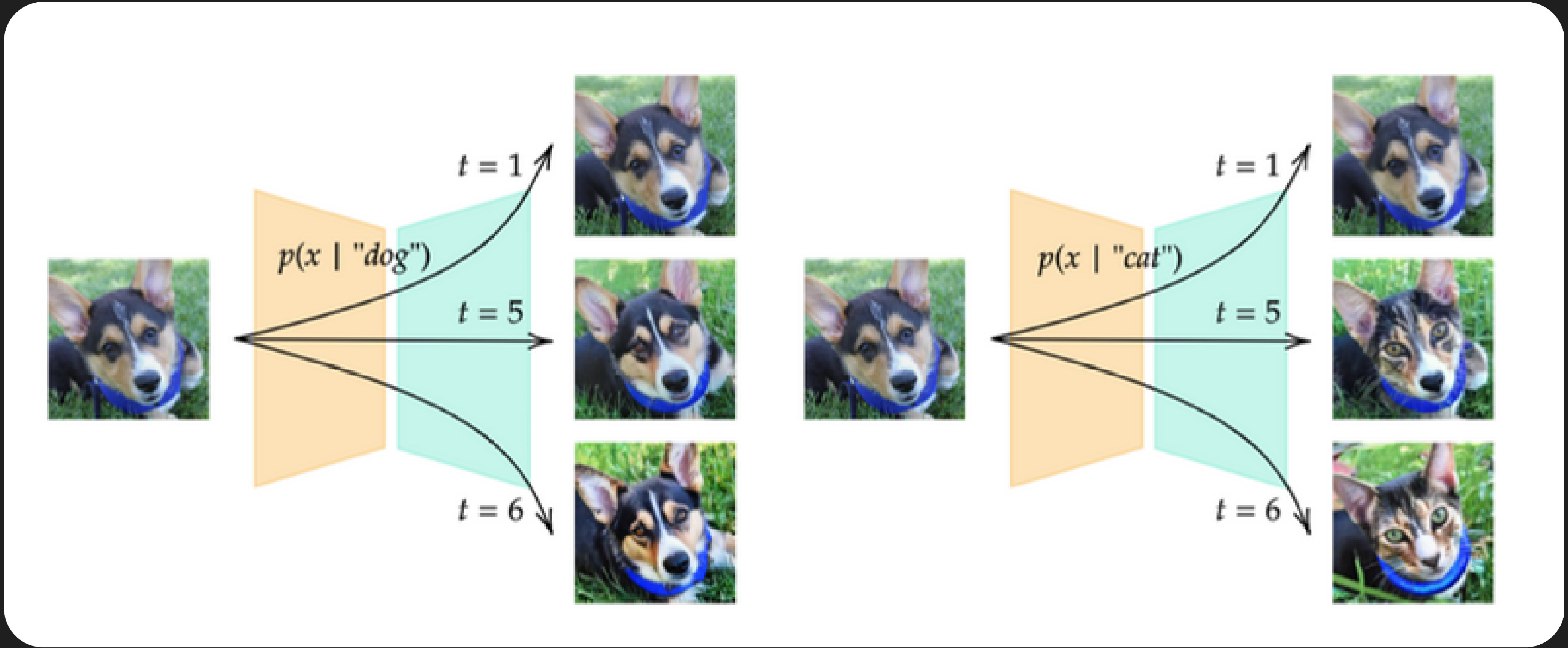
CORE IDEA 2

Auto-encode (distilled) DMs to contract random variables towards their internal generative model

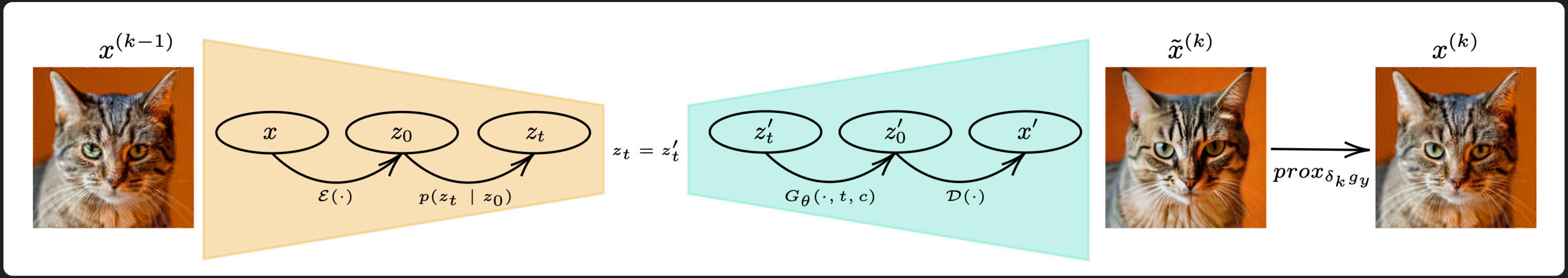
$$\mathfrak{E}_t : \quad \boldsymbol{z}_t | \boldsymbol{x} \sim \mathcal{N}(\sqrt{\alpha_t}\mathcal{E}(\boldsymbol{x}), (1-\alpha_t)\mathrm{Id}_d)$$

$$\mathfrak{D}_{t,c} : \quad \boldsymbol{x}' = \mathcal{D}(G_\theta(\boldsymbol{z}'_t, t, c))$$

probability flow model



LATINO



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

$\quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathrm{Id})$

$\quad \boldsymbol{z}_{t_k}^{(k)} \leftarrow \sqrt{\alpha_{t_k}}\mathcal{E}(\boldsymbol{x}^{(k-1)}) + \sqrt{1-\alpha_{t_k}}\boldsymbol{\epsilon}$ $\quad \triangleright$ Encode ← Exponential contraction in W2 and TV.

$\quad \boldsymbol{u}^{(k)} \leftarrow \mathcal{D}(G_\theta(\boldsymbol{z}_{t_k}^{(k)}, t_k, c))$ $\quad \triangleright$ Decode ← Expansive, but Lipschitz. Polynomial OK.

$\quad \boldsymbol{x}^{(k)} \leftarrow \mathrm{prox}_{\delta_k g_y}(\boldsymbol{u}^{(k)})$ $\quad \triangleright\ g_{\boldsymbol{y}} : \boldsymbol{x} \mapsto -\log p(\boldsymbol{y}|\boldsymbol{x})$ ← Maximally-monotone operator.

end for

CORE IDEA 3

Self-calibrate text prompt by maximum marginal likelihood optimization

for $m = 1, \ldots, M$ **do**

  for $k = 1, \ldots, N_m$ **do** ▷ LATINO

    $\epsilon \sim \mathcal{N}(0, \mathrm{Id})$

    $\boldsymbol{z}_{t_k}^{(k)} \leftarrow \sqrt{\alpha_{t_k}}\mathcal{E}(\boldsymbol{x}^{(k-1)}) + \sqrt{1-\alpha_{t_k}}\epsilon$

    $\boldsymbol{u}^{(k)} \leftarrow \mathcal{D}(G_\theta(\boldsymbol{z}_{t_k}^{(k)}, t_k, c_m))$

    $\boldsymbol{x}^{(k)} \leftarrow \mathrm{prox}_{\delta_k g_y}(\boldsymbol{u}^{(k)})$

  end for

  $h(c_m) \leftarrow \nabla_c \log p(\boldsymbol{z}_{t_1}^{(1)}, \ldots, \boldsymbol{z}_{t_{N_m}}^{(N_m)} | c_m)$

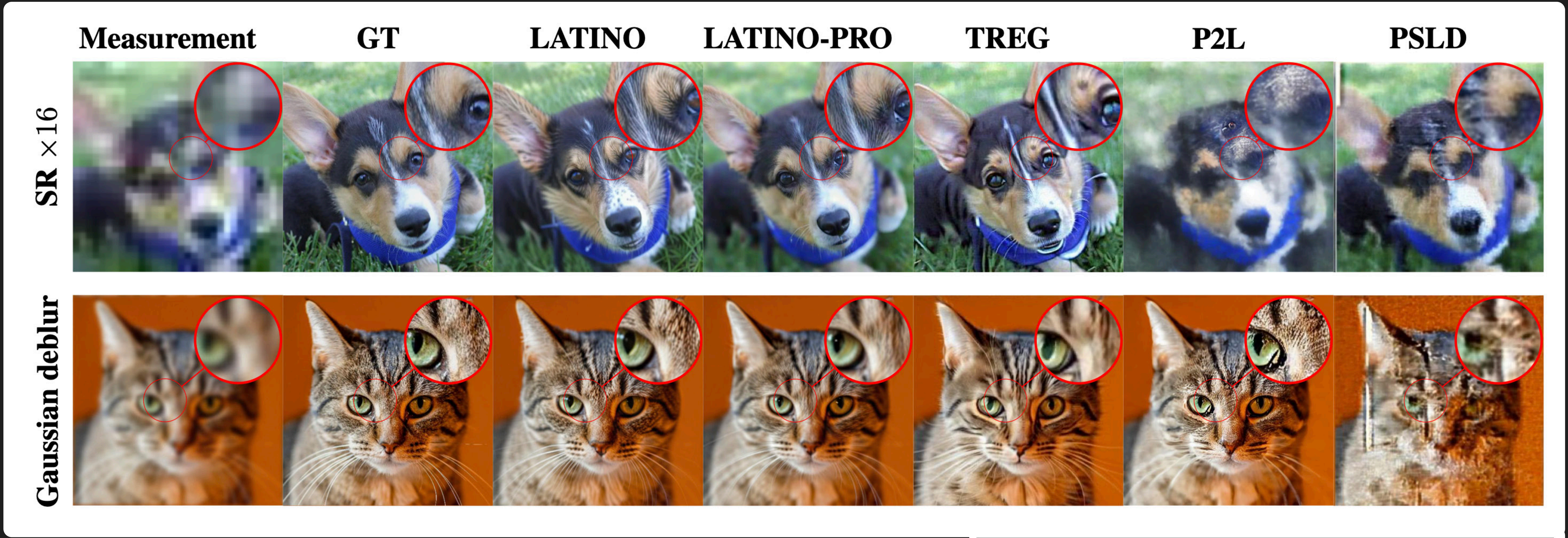
  $c_{m+1} = \Pi_C\left[c_m + \gamma_m h(c_m)\right]$ ▷ SAPG

  $\boldsymbol{x}^{(0)} \leftarrow \boldsymbol{x}^{(N_m)}$ ▷ Carry state forward

stochastic approx. proximal gradient step to compute

$$\hat{c}(\boldsymbol{y}) = \arg\max_{c \in \mathbb{R}^k} p(\boldsymbol{y} \mid c)$$

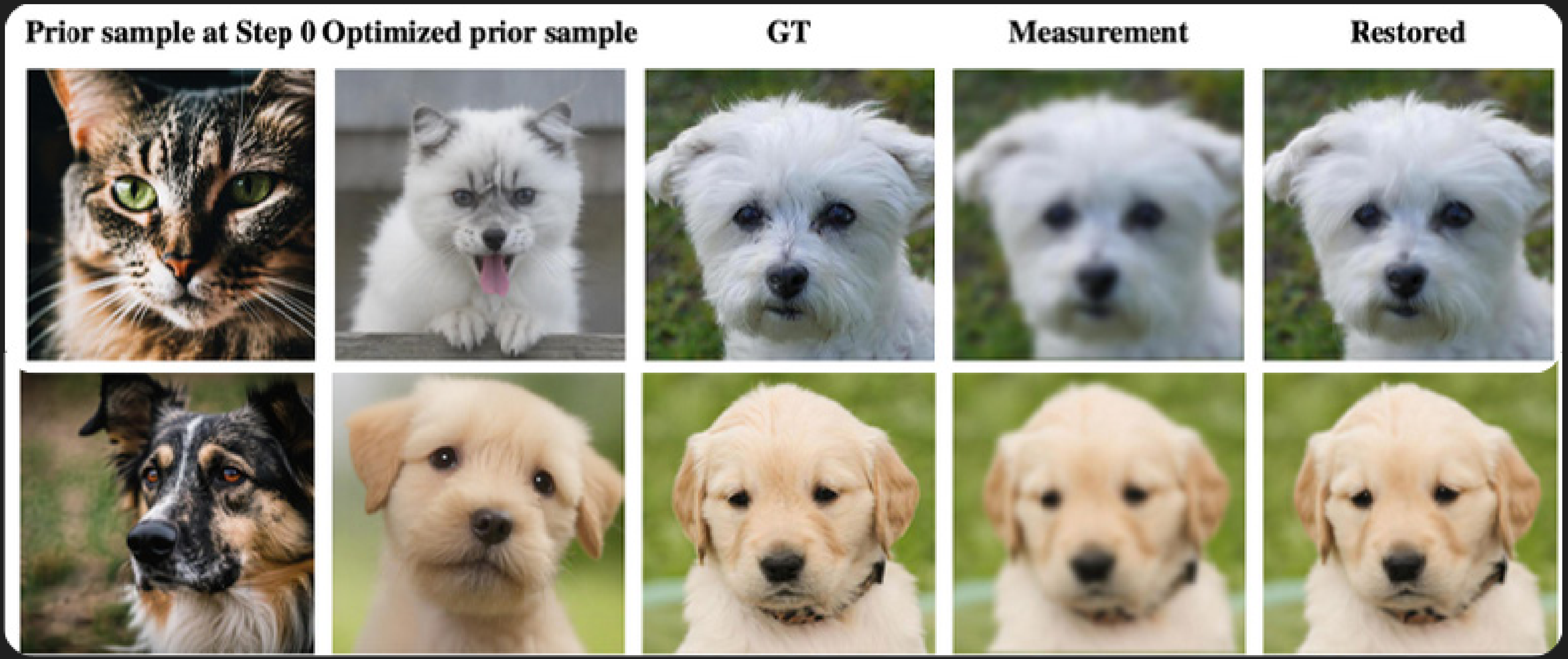
SOME RESULTS



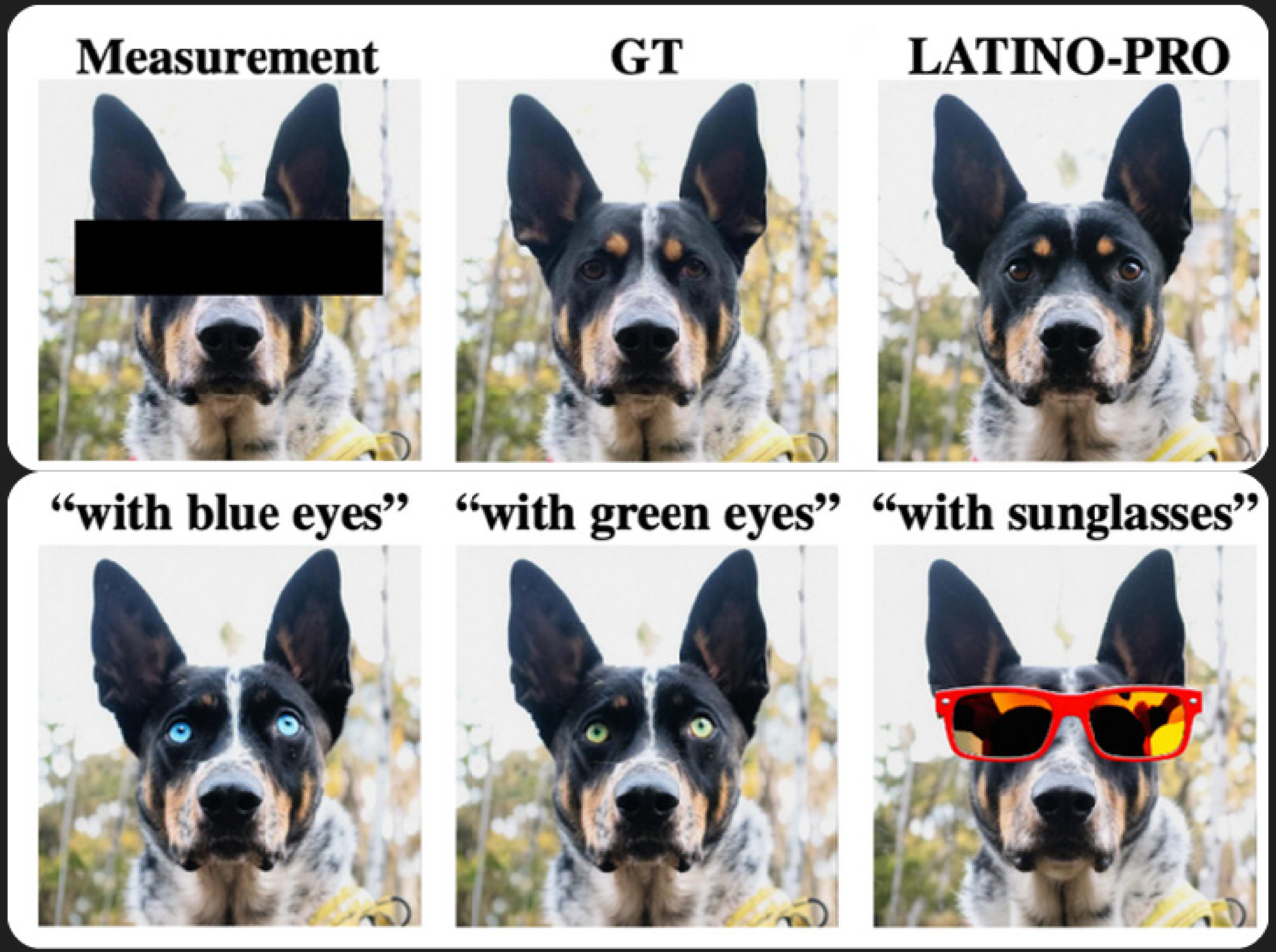
P2L: Chung et al. ICML 2024, TREG: Kim et al. ICLR 2025,
PSLD: Rout et al. NeurIPS 2023.

		Deblur (Gaussian)		SR×16	
Method	**NFE↓**	**FID↓**	**PSNR↑**	**FID↓**	**PSNR↑**
LATINO-PRO	68	**18.37**	**26.82**	**30.40**	**21.52**
LATINO	**8**	20.03	26.25	42.14	20.05
P2L [7]	2000	85.80	20.96	121.7	19.99
TReg [21]	200	35.47	21.13	37.13	19.60
LDPS	1000	64.88	22.60	101.13	17.34
PSLD [45]	1000	125.5	20.52	113.4	16.48

VISUALIZATION OF PROMPT OPTIMIZATION

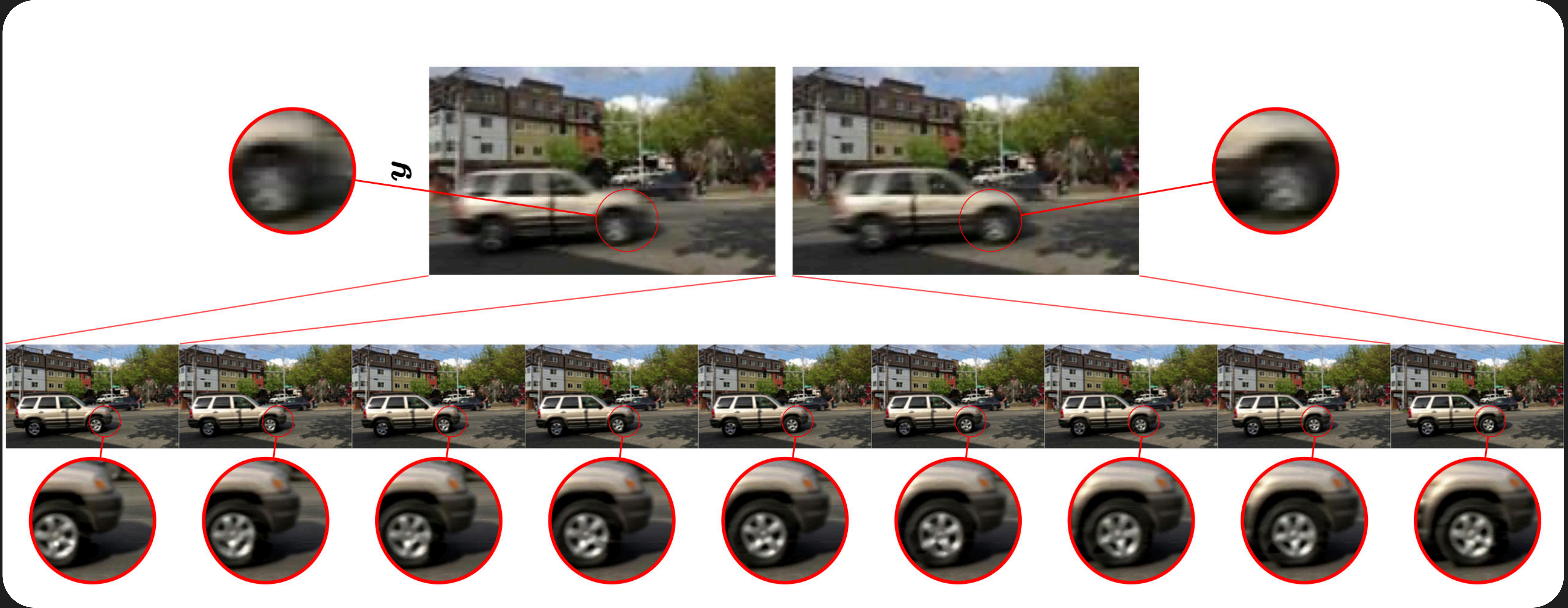


Samples from the prior before/after 4 SAPG steps



Samples with "semantically constrained" SAPG steps.

EXTENSION TO VIDEO RESTORATION

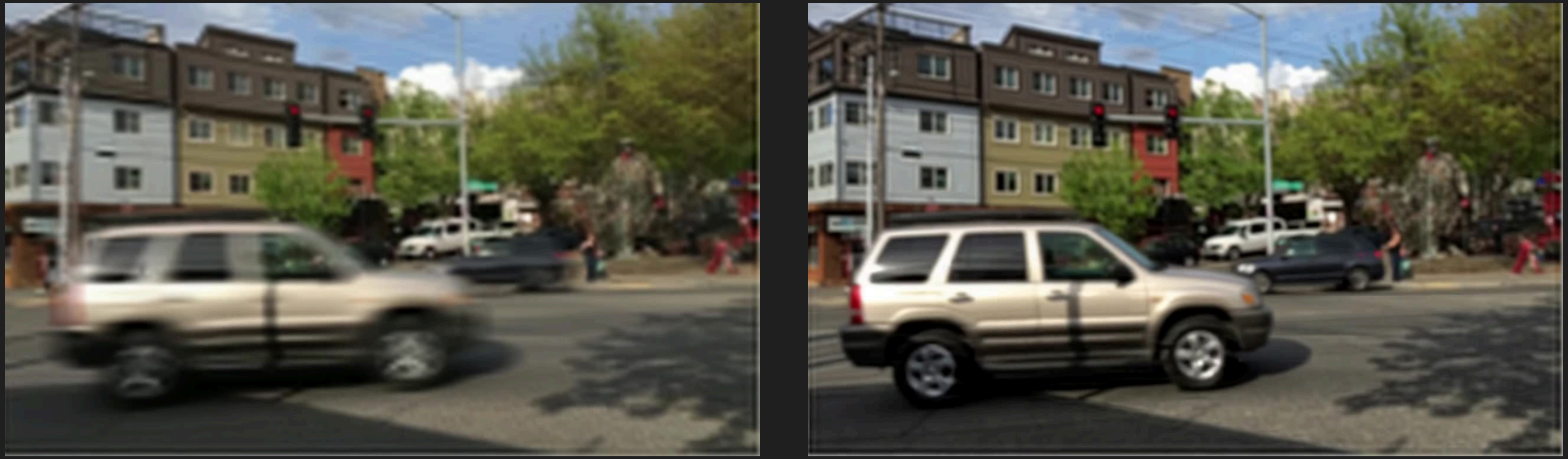


Spatial-temporal super-resolution x8 using video CM distilled from WAN 2.1 (see https://arxiv.org/pdf/2510.01339)

	Problem C: Temp. SR×8 + SR×8				
Method	NFE↓	FVMD↓	PSNR↑	SSIM↑	LPIPS↓
LVTINO	**7**	**602.5**	23.11	**0.697**	**0.411**
VISION-XL	8	1604	**23.38**	0.652	0.520
ADMM-TV	–	1645	18.15	0.663	0.439

EXTENSION TO VIDEO RESTORATION

Spatial-temporal super-resolution x8 as a challenging linear problem



OUTLINE

01 BACKGROUND

02 LATINO-PRO: LAtent consisTency INverse sOlver with PRompt Optimization

03 Learning few-step posterior samplers by unfolding and distillation

04 Conclusion

Learning few-step posterior samplers by unfolding and distillation of diffusion models

TMLR 2025.

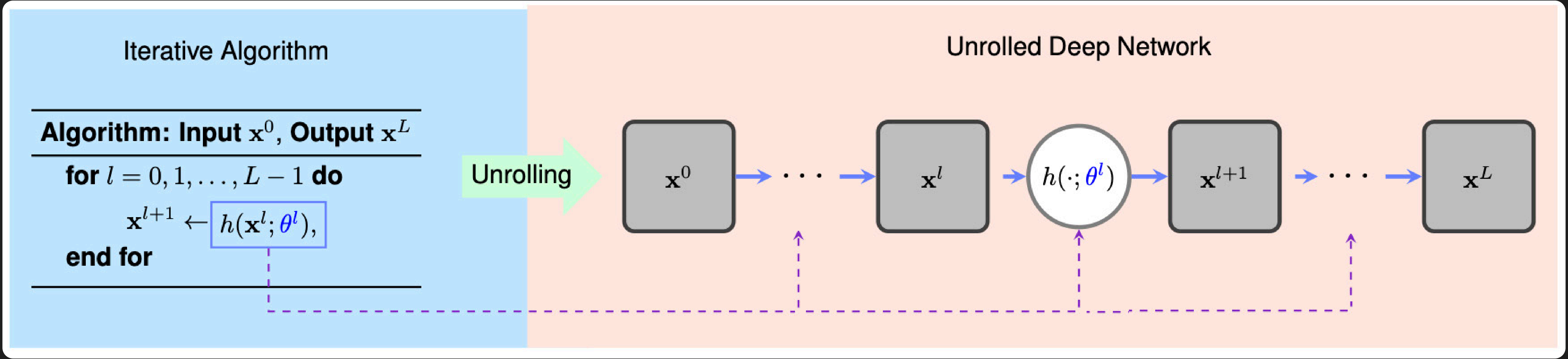
Joint work with Charlesquin Kemajou Mbakam and Jonny Spence.

BACKGROUND

Deep unfolding:

Transforms an iterative (optimization) algorithm with a fixed number of iterations into a deep neural network architecture, whereby the algorithm's steps become trainable layers that are trained end-to-end.

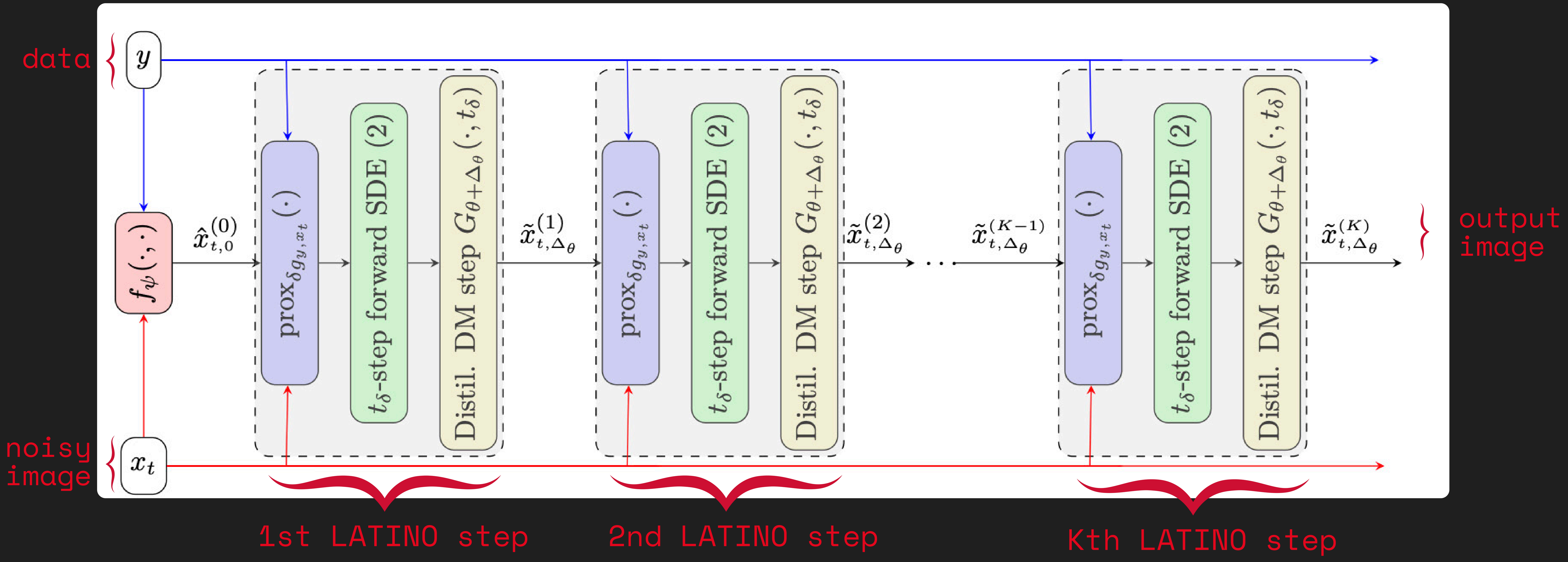
A powerful template for designing interpretable architectures that incorporate y and the degradation model explicitly during inference.



arXiv:1912.10557

CORE IDEA 1

We propose to unfold K LATINO modules. As prior, we use a DM $G_{\theta+\Delta_\theta}$ where Δ_θ is a *LoRA* on a frozen pre-trained DM G_θ . Warm-start with f.



CORE IDEA 2

We distil as conditional CM to sample $\mathbb{x}$ given y and a noisy copy of $\mathbb{x}$:

$$\underset{\Delta_\theta,\psi}{\arg\min} \quad \mathcal{L}^G(\Delta_\theta,\psi) \triangleq \mathcal{L}_{\text{Adv}}(\Delta_\theta,\psi,\phi) + \omega_{\ell_2}\mathcal{L}_{\ell_2}(\Delta_\theta,\psi) + \omega_{\text{PS}}\mathcal{L}_{\text{PS}}(\Delta_\theta,\psi)\,,$$

$$\underset{\phi}{\arg\max} \quad \mathcal{L}^D(\phi) \triangleq \mathcal{L}_{\text{Adv}}(\Delta_\theta,\vartheta,\phi) + \omega_{\text{GS}}\mathcal{L}_{\text{GS}}(\phi)\,,$$

Conditional consistency model objective

$$\mathcal{L}_{\ell_2}(\Delta_\theta,\psi) = \mathbb{E}_{t,\mathbf{x}_t,\mathbf{y},\mathbf{x}_0}\left[||\mathbf{x}_0 - L_{\Delta_\theta,\psi}(\mathbf{x}_t,\mathbf{y})||_2^2\right]$$

MSE quality loss

$$\mathcal{L}_{\text{Adv}}(\Delta_\theta,\vartheta,\phi) = \mathbb{E}_{\mathbf{x}_0,\mathbf{y}}\left[\log\left(\varsigma(D_\phi(\mathbf{x}_0;\mathbf{y}))\right)\right] + \mathbb{E}_{t,\mathbf{x}_t,\mathbf{y}}\left[\log\left(1-\varsigma(D_\phi(L_{\Delta_\theta,\psi}(\mathbf{x}_t,\mathbf{y}),\mathbf{y}))\right)\right]$$

adversarial (GAN) loss

$$\mathcal{L}_{\text{PS}}(\Delta_\theta,\psi) = \mathbb{E}_{t,\mathbf{x}_t,\mathbf{y},\mathbf{x}_0}\left[\text{LPIPS}(\mathbf{x}_0, L_{\Delta_\theta,\psi}(\mathbf{x}_t,\mathbf{y}))\right]$$

perceptual quality loss

$$\mathcal{L}_{\text{GS}}(\phi) = \mathbb{E}_{t,\mathbf{x}_t,\mathbf{y},\mathbf{x}_0,\mathbf{u}}\left[\left\|\nabla_x D_\phi(\mathbf{u}\mathbf{x} + (1-\mathbf{u})L_{\Delta_\theta,\psi}(\mathbf{x}_t,\mathbf{y}),\mathbf{y})\right\|^2\right]$$

Discriminator regularity loss

UNFOLDED DISTILLED DIFFUSION MODEL (UD2M)

Embed trained UD2M network within multi-step CM sampler. The data **y** and the **data fidelity term are specified during inference**.

Require: Observation y, Time-grid $0 = t_0 < t_1 < \cdots < t_N = T$,

Sample $x_{t_N} \sim \mathcal{N}(0, \boldsymbol{I})$ ▷ Initialize reversed diffusion

for $n = N, \ldots, 1$ **do**

  Set $\hat{x}_0 \leftarrow \tilde{x}^{\Delta_\theta}_{t_n,K}(x_{t_n}, y)$ using L_ϑ ▷ Unfolded sample targeting $p_0(x_0 \,|\, x_{t_n}, y)$

  Sample $x_{t_{n-1}} \sim p_{t_{n-1}}(x_{t_{n-1}} \,|\, \hat{x}_0, x_{t_n})$ ▷ Reverse DDIM step

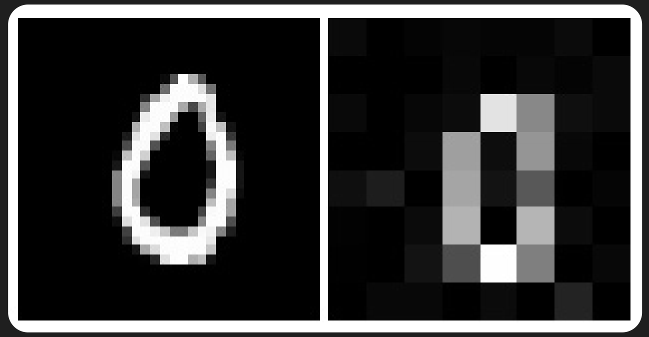
end for

return x_{t_0}

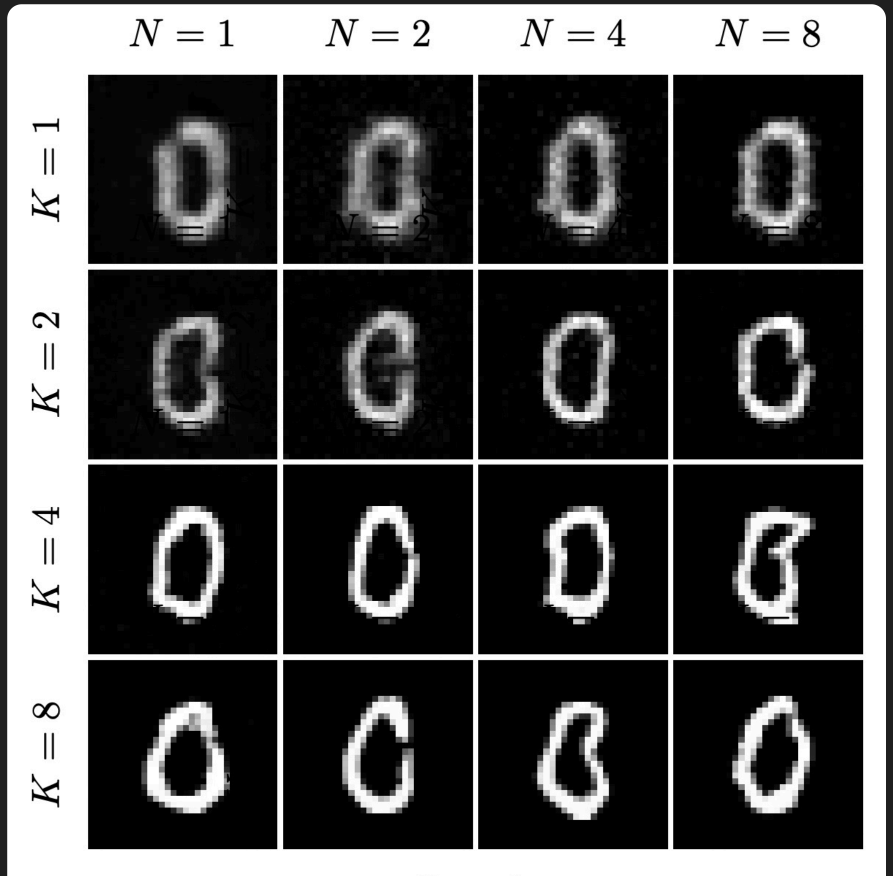
We typically unfold K = 3 LATINO modules with warm-starting and use N = 3 steps.

ILLUSTRATIVE EXAMPLES

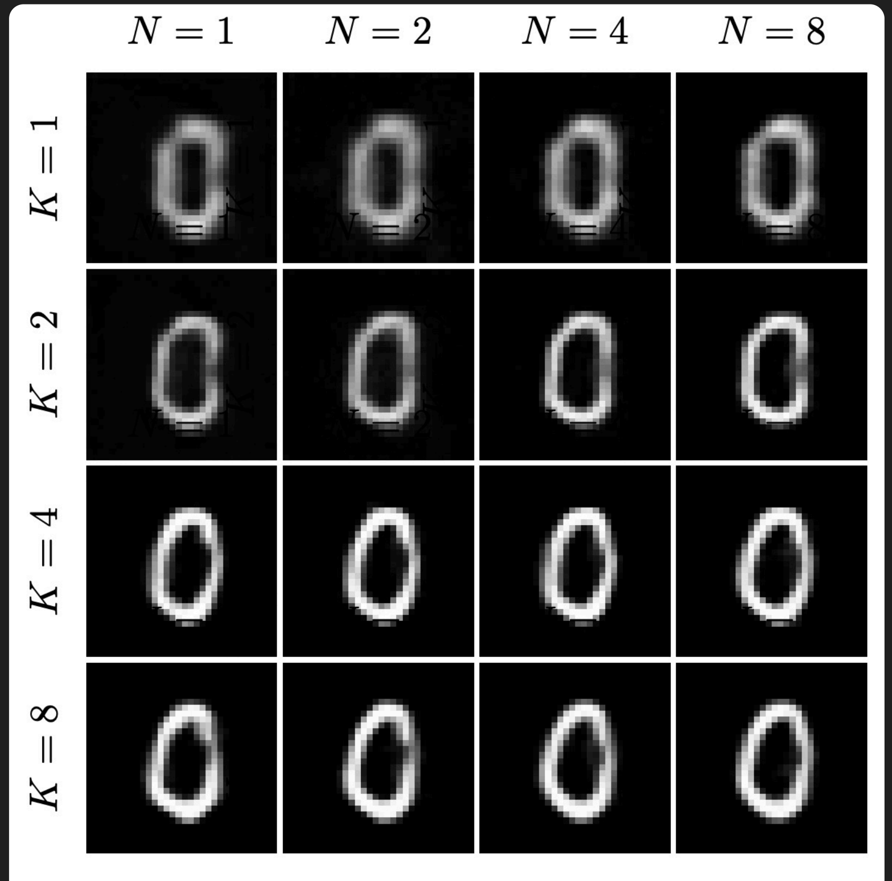
UD2M samples on MNIST SRx4.



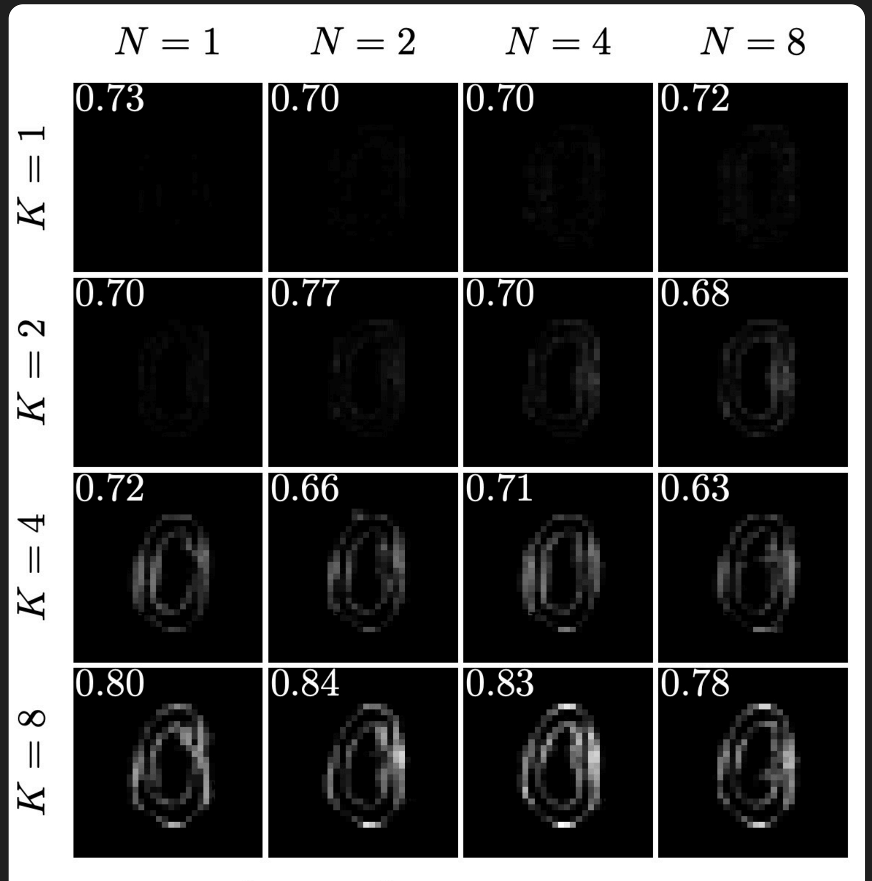
Reference Observed



Samples



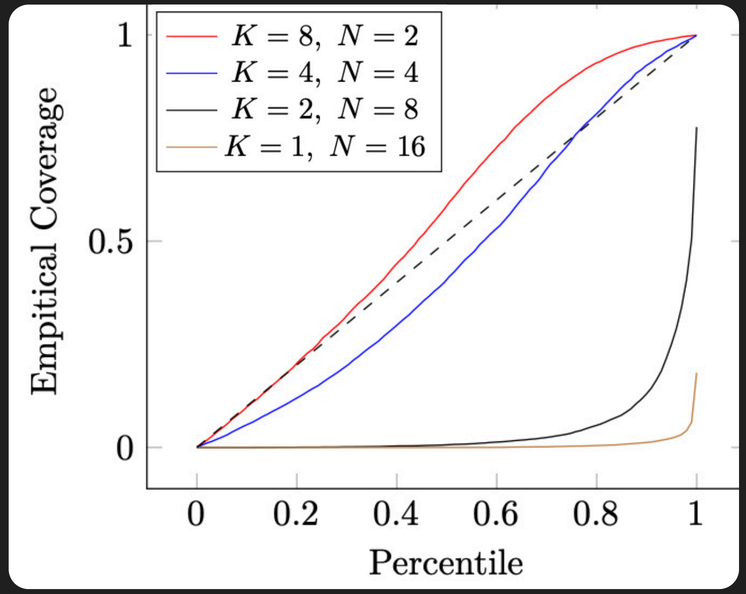
Sample Mean



Sample Standard Deviation

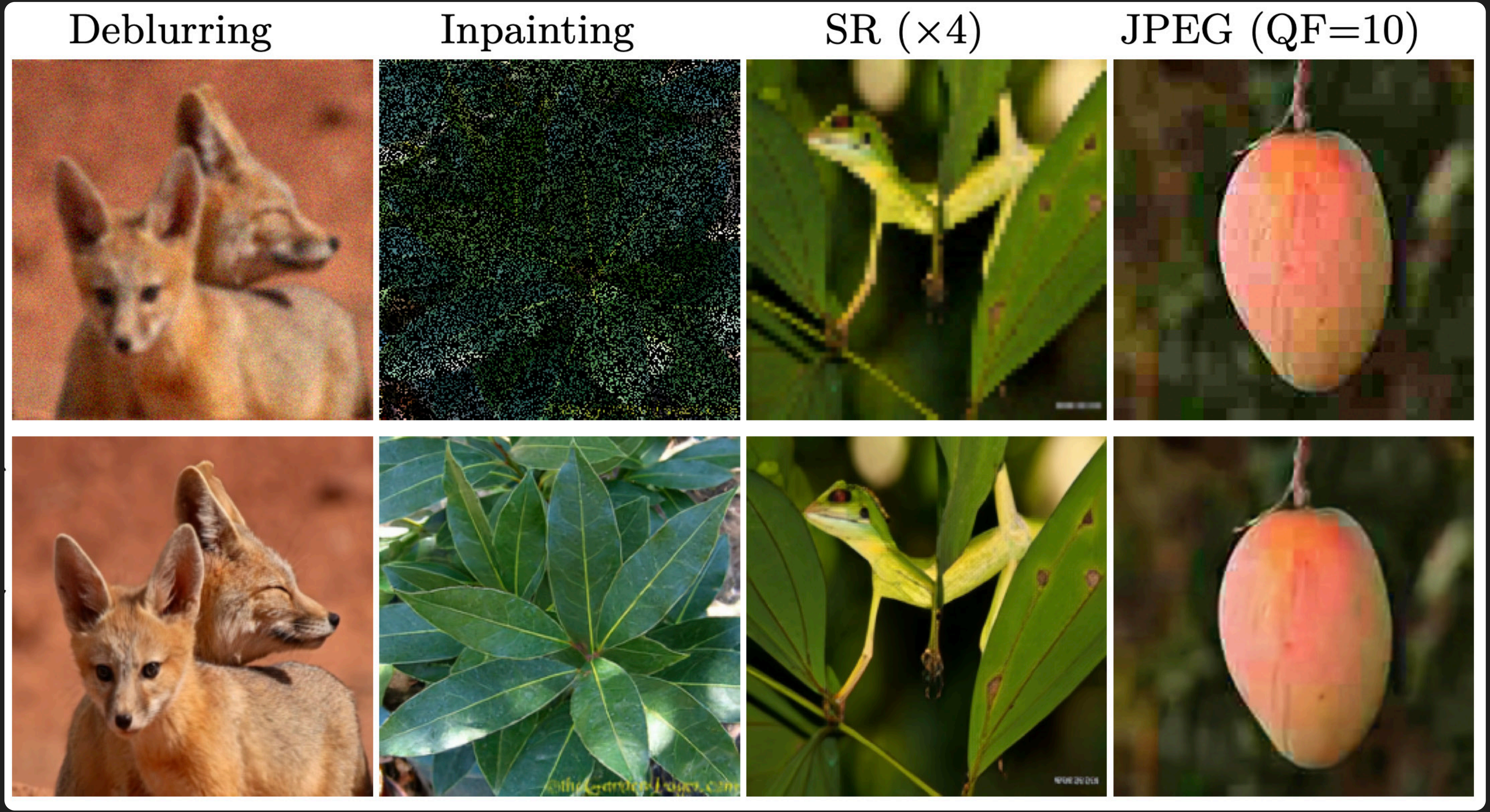
K \ N	1	2	4	8	16
1	6.91	5.16	4.83	4.08	3.53
2	3.65	2.25	1.47	1.01	0.94
4	0.47	0.38	0.37	0.36	0.35
8	0.22	0.22	0.22	0.22	0.22

Conditional W2 distance



ILLUSTRATIVE EXAMPLES

UD2M samples on the ImageNet 256 dataset for Gaussian Deblurring, random inpainting (70%), super-resolution (4×), and restoration of JPEG compression artifacts (QF=10). All methods are Iimplemented with an ImageNet DM in pixel domain (no text prompting).



		Deblurring						SR			JPEG		
		Gaussian			Uniform			×4			QF=10		
Methods	NFEs	PSNR	LPIPS	FID	PSNR	LPIPS	FID	PSNR	LPIPS	FID	PSNR	LPIPS	FID
Ours (wo RAM)	9	**38.77**	0.02	4.61	35.57	0.02	11.14	24.42	0.15	20.69	**27.52**	**0.18**	35.16
Ours (w/ RAM)	12	35.97	**0.01**	**3.30**	**36.96**	**0.01**	**2.69**	**26.70**	**0.08**	**11.9**	-	-	-
CDDB	1000	37.02	0.06	5.01	31.26	0.19	23.15	26.41	0.2	19.88	26.34	0.26	**19.48**
I2SB	1000	36.01	0.07	5.8	30.75	0.2	23.01	25.22	0.26	24.13	26.12	0.27	20.35
DiffPIR	100	28.10	0.13	21.53	31.44	0.10	20.20	20.39	0.36	70.45	-	-	-
DDRM	20	36.73	0.07	4.34	29.21	0.21	19.97	26.05	0.27	46.49	26.33	0.33	47.02

I2SB (Lui, ICML 2023) and CDDB (Chung, NeurIPS 2023) are Schrodinger bridges from **y** to **x**. DiffPIR (Zhu, CVPR 2023) and DDRM (Kawar, NeurIPS 2022) are zero-shot DM methods.

OUTLINE

01 BACKGROUND

02 LATINO-PRO: LAtent consisTency INverse sOlver with PRompt Optimization

03 Learning few-step posterior samplers by unfolding and distillation

04 Conclusion

CONCLUSION

Take-home 1: Deep generative image/video models provide a powerful framework for Bayesian inversion, via "prompting" with likelihood+data.

Take-home 2: Appropriate choice of SDE/ODEs and "distillation" are central to our approach and to achieving accurate results with a low cost.

NEWS

I am joining Imperial College London and the Imperial-CNRS Laboratory!

Please feel free to DM about opportunities in my group, to discuss an idea, or if you have any questions about our papers and codes.

m.pereyra@hw.ac.uk www.macs.hw.ac.uk/~mp71/

THANK YOU!