

Phased Consistency Models

Fu-Yun Wang¹ Zhaoyang Huang² Alexander William Bergman^{3,6} Dazhong Shen⁴ Peng Gao⁴ Michael Lingelbach^{3,6}
 Keqiang Sun Weikang Bian¹ Guanglu Song⁵ Yu Liu⁴ Xiaogang Wang¹ Hongsheng Li^{1,4,7}

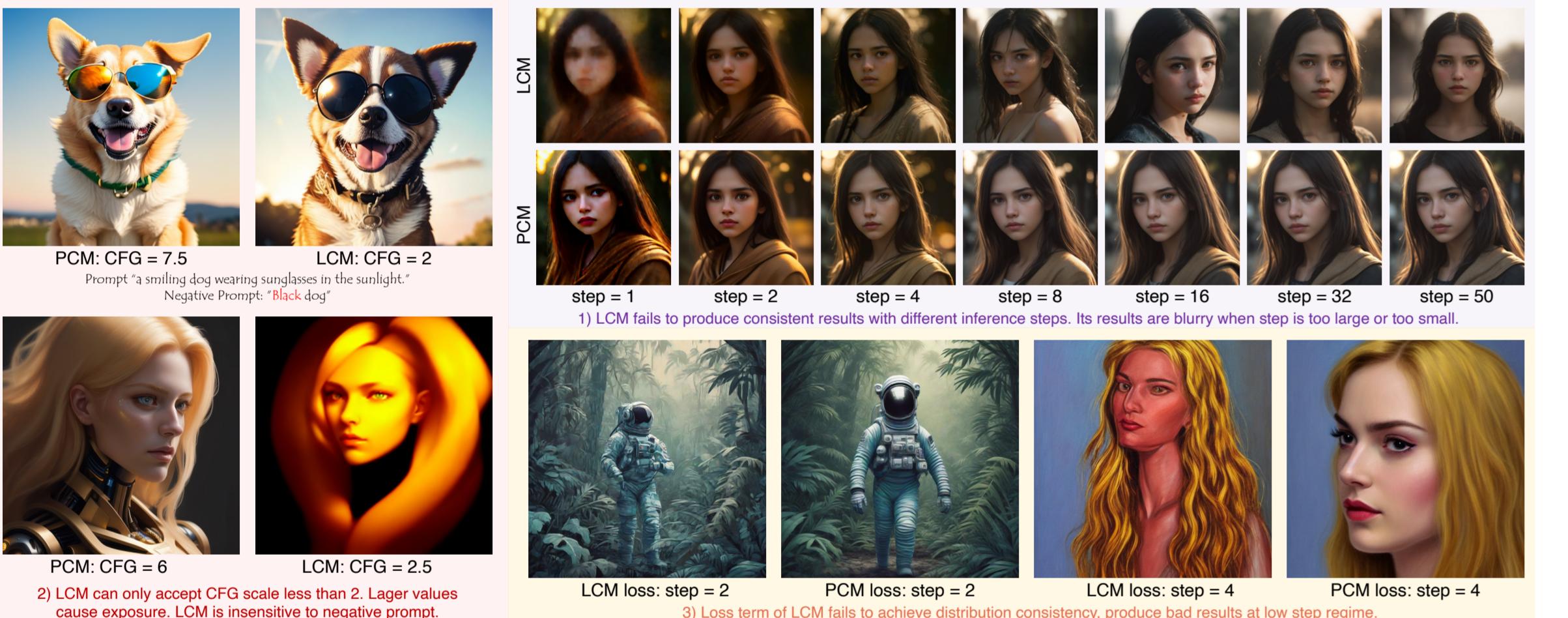
¹ CUHK MMLab ² Evolution AI ³ Hedra ⁴ Shanghai AI Lab ⁵ Sensemte Research ⁶Stanford University ⁷CPII under InnoHK



Background & Motivation

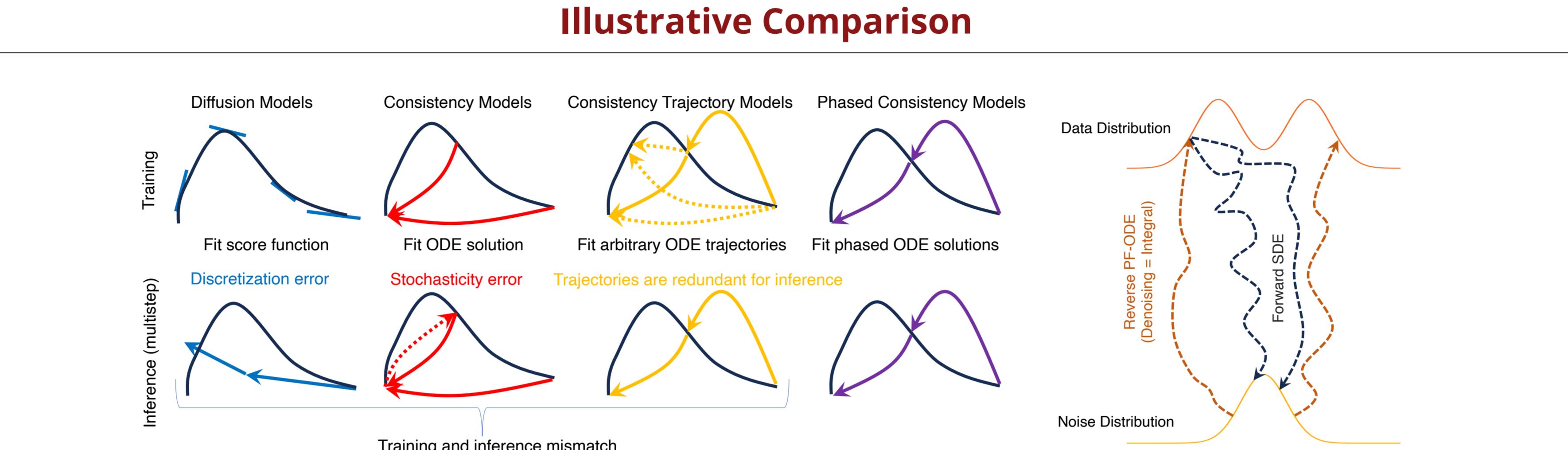
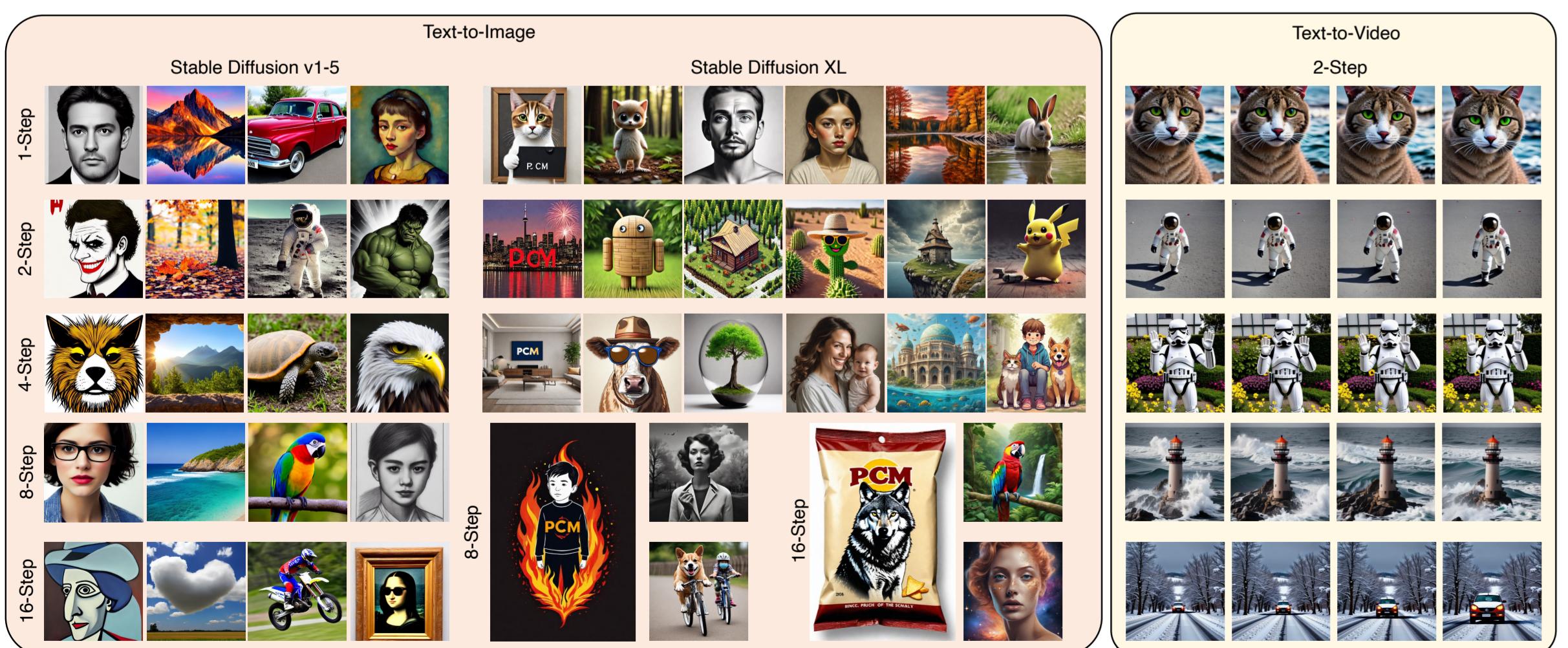
- Consistency Models (CMs) have made significant progress, capable of generating diverse high-fidelity samples in one step.
- Latent Consistency Models (LCMs) extend the scope of CMs to the high-resolution text-to-image generation. Yet the generation quality of LCMs is not satisfactory.

Limitations of Latent Consistency Models



LCMs face drawbacks in **controllability**, **consistency**, and **efficiency**. PCMs identify these limitations, generalize the design space, and tackle these limitations.

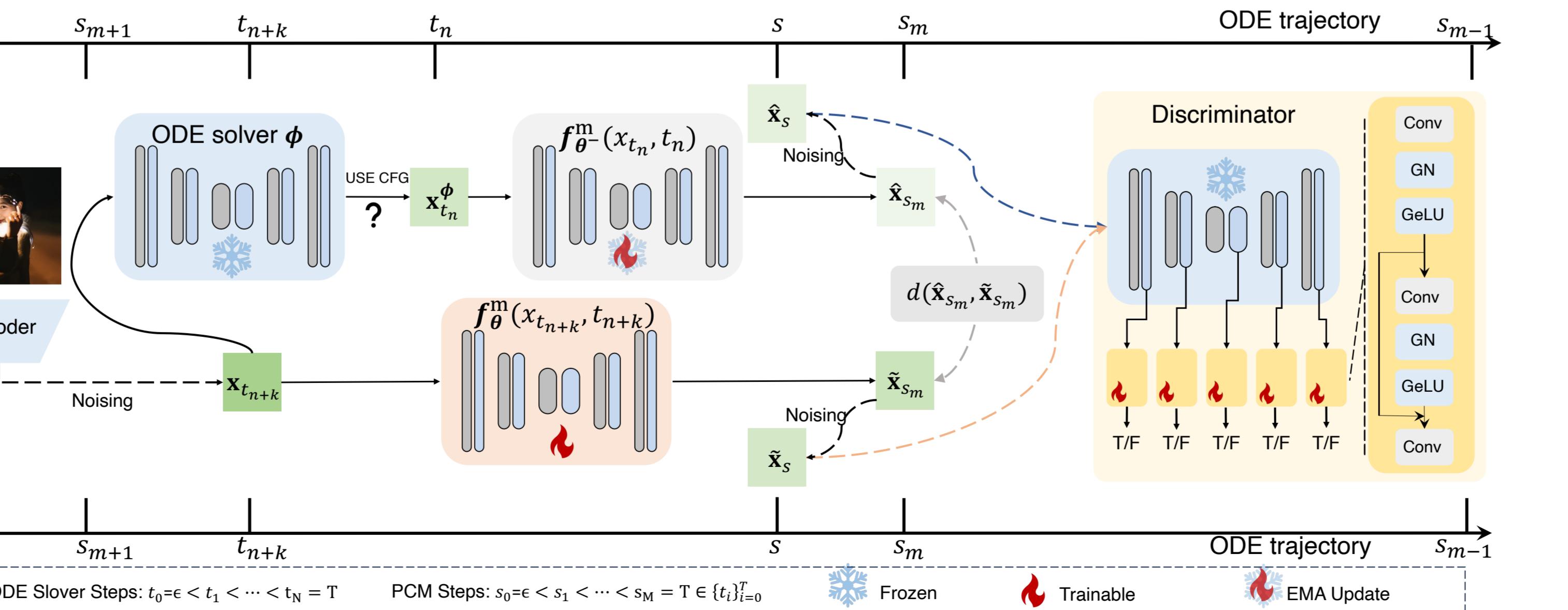
Text-to-Image and Text-to-Video in One Step



- (1) Diffusion models learn the gradient of PF-ODE, but face inevitable discretization errors in few-step settings.
- (2) Consistency models learn the solution point of PF-ODE but face stochasticity error in multistep sampling.
- (3) Consistency trajectory models learn arbitrary trajectories but is challenging to train.
- (4) Phased consistency models learn the deterministic multistep sampling and is easy to train.

Training Pipeline

- A VAE to encode the images into latents for efficient training.
- Denoising $\mathbf{x}_{t_{n+k}}$ with pretrained ODE solver ϕ to obtain $\mathbf{x}_{t_n}^\phi$.
- Penalizing the prediction distance between $\hat{\mathbf{x}}_{s_m} = \mathbf{f}_\theta^m(\hat{\mathbf{x}}_{t_n}, t_n)$ and $\tilde{\mathbf{x}}_{s_m} = \mathbf{f}_\theta^m(\hat{\mathbf{x}}_{t_{n+k}}, t_{n+k})$ to enforce self-consistency property.
- Adding noise to the latents to obtain $\mathbf{x}_{t_{n+k}}$.
- Latent adversarial consistency loss with a discriminator initialized with the pretrained diffusion models.



More Generation Results

