Phased Consistency Models

Fu-Yun Wang[1] Zhaoyang Huang[2] Alexander William Bergman[3,6] Dazhong Shen[4] Peng Gao[4] Michael Lingelbach[3,6]
Keqiang Sun Weikang Bian[1] Guanglu Song[5] Yu Liu[4] Xiaogang Wang[1] Hongsheng Li[1,4,7]

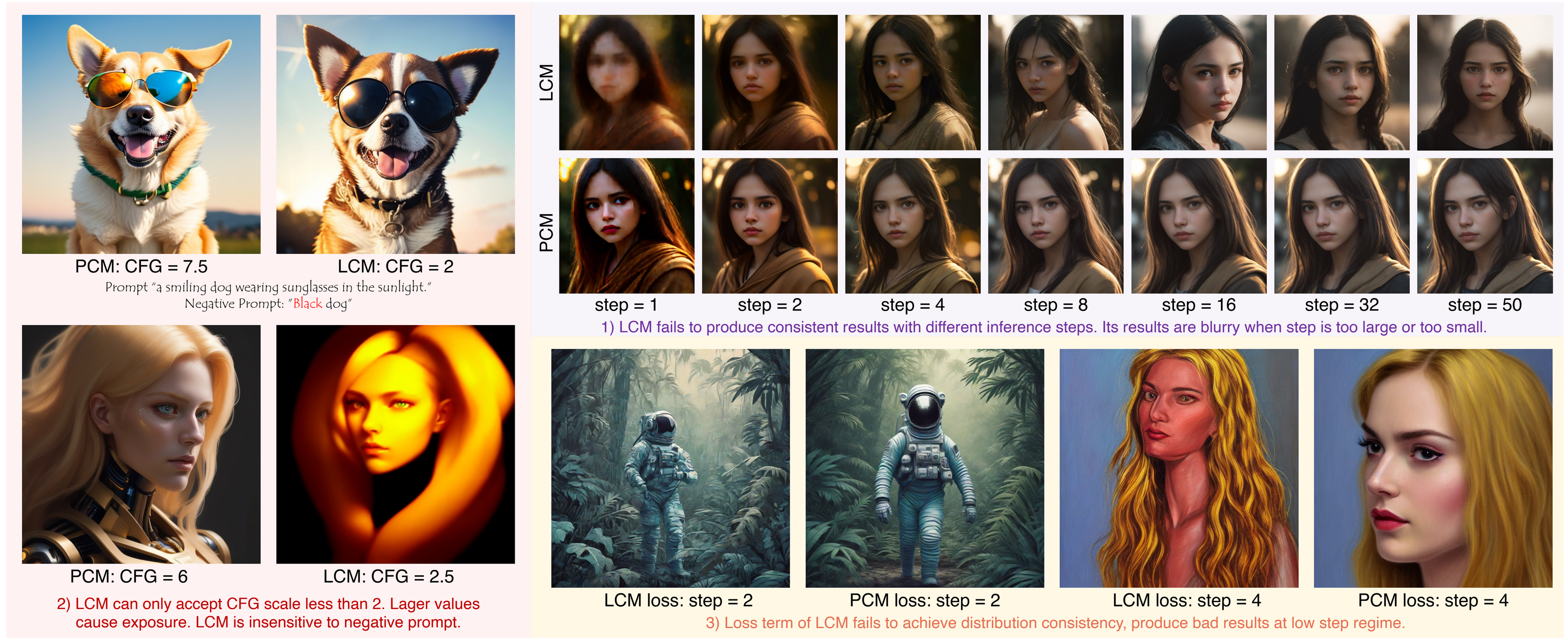
[1] CUHK MMLab [2] Avolution AI [3] Hedra [4] Shanghai AI Lab [5] Sensetime Research [6]Stanford University [7]CPII under InnoHK

SCAN ME

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.

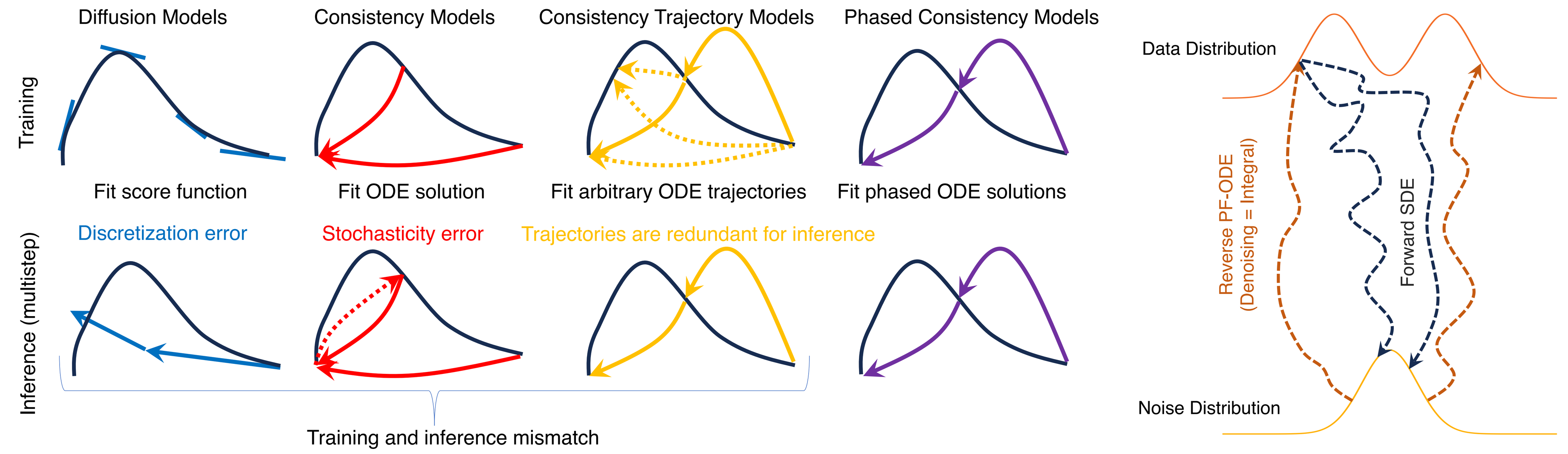
Limitations of Latent Consistency Models



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

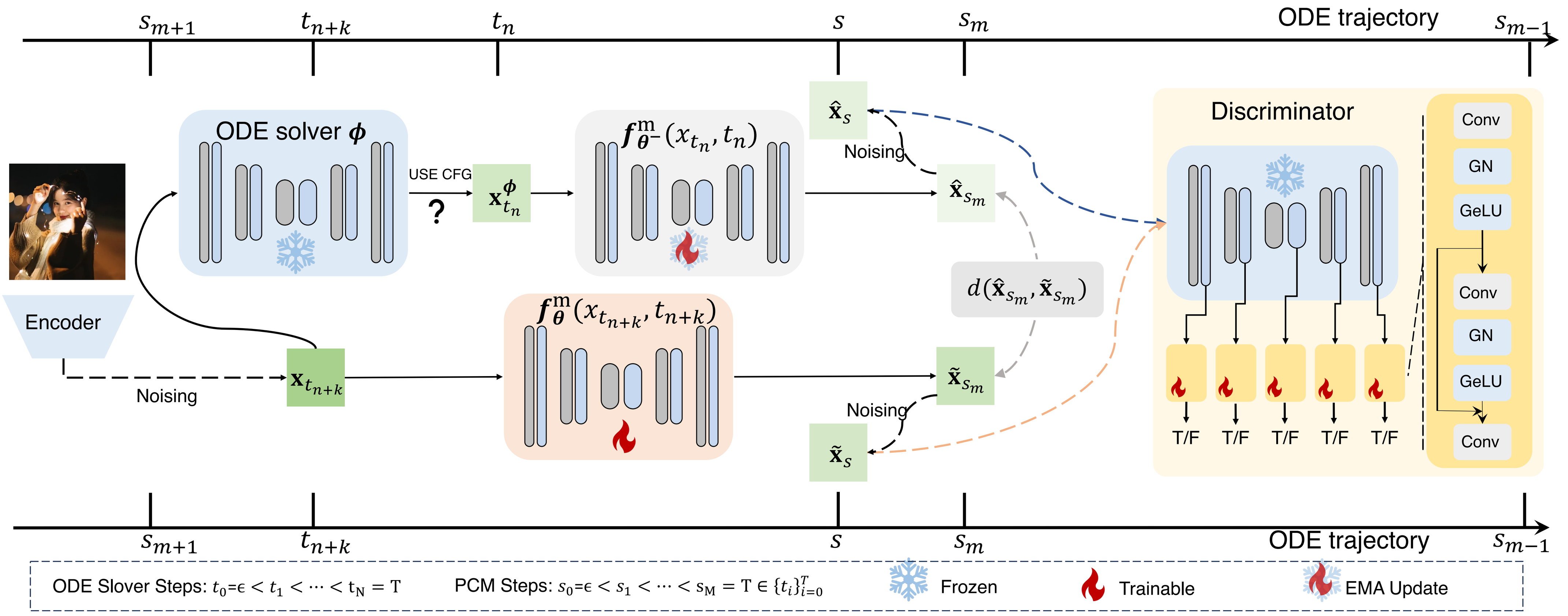
Illustrative Comparison



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



More Generation Results