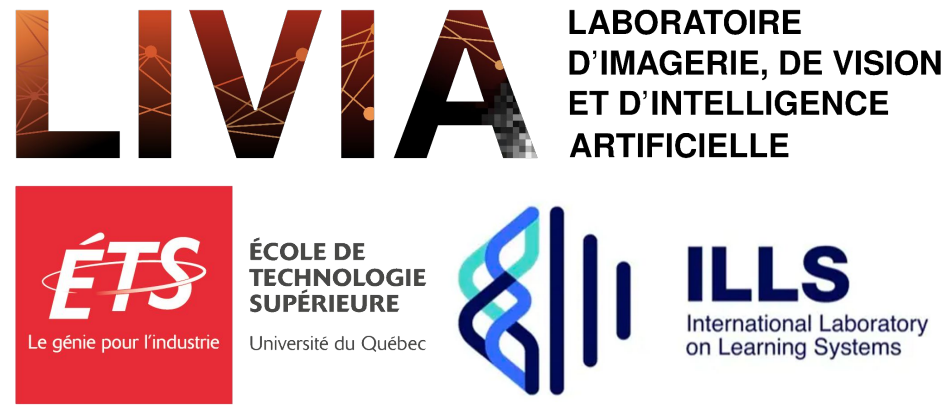


DogFit: Domain-guided Fine-tuning for Efficient Transfer Learning of Diffusion Models



Yara Bahram, Mohammadhadi Shateri, Eric Granger
yara.mohammadi-bahram@livia.etsmtl.ca



Main Contributions

- We propose domain-guided fine-tuning (**DogFit**), a controllable guidance mechanism for diffusion transfer learning that improves target domain alignment without needing double forward passes at test-time.
- We show that during fine-tuning, the source model offers stronger marginal estimates than the learning target model, making it better suited for generating guidance signals.
- For transfer learning across six target datasets on DiT and SiT models, **DogFit** outperforms state-of-the-art guidance methods in majority of cases while being up to 2x faster.

Motivation and Method

Diffusion Transfer Learning is Still Challenging

- **Fine-tuning** on small target datasets generalizes poorly and suffers from **low fidelity**.
- **Test-time guidances** like classifier-free guidance (CFG)^[1] and domain guidance (DoG)^[2] improve fidelity, but sampling needs **double forward passes**.
- **Training-time guidances** like model guidance (MG)^[3] do not require double forward passes, but exhibit **bad target alignment** and **no test-time guidance control**.

DogFit to the Rescue

- Improved fidelity
- Single forward pass sampling
- Target domain alignment
- Test-time guidance control

DogFit Tweaks the Denoising Training Loss

Learning Target Model

$$\mathcal{L}_{\text{DogFit}} = \mathbb{E}_{t, (x_0, c), \epsilon} \|\epsilon_{\theta}(x_t | c, \mathcal{D}^T) - \epsilon'\|^2,$$

$$\epsilon' = \epsilon + (w - 1) \cdot \text{sg}(\epsilon_{\theta}(x_t | c, \mathcal{D}^T) - \epsilon_{\theta_0}(x_t))$$

Guidance Weight

Stop Gradient

Source Model

Adding Test-time Guidance Control

Sample Guidance Weight

Input Guidance

$$\mathcal{L}_{\text{DogFit}} = \mathbb{E}_{t, (x_0, c, w), \epsilon} \|\epsilon_{\theta}(x_t | c, w, \mathcal{D}^T) - \epsilon'\|^2,$$

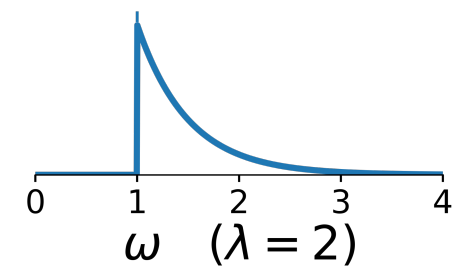
$$\epsilon' = \epsilon + (w - 1) \cdot \text{sg}(\epsilon_{\theta}(x_t | c, 1, \mathcal{D}^T) - \epsilon_{\theta_0}(x_t))$$

Input No Guidance

- Sample from shifted exponentially decaying distribution (SEDD):

$$w = 1 + z, \quad z \sim \mathcal{P}(z),$$

$$\mathcal{P}(z) = \lambda e^{-\lambda z}, \quad z \geq 0,$$



DogFit and Friends for Guiding Generation in Transfer Learning

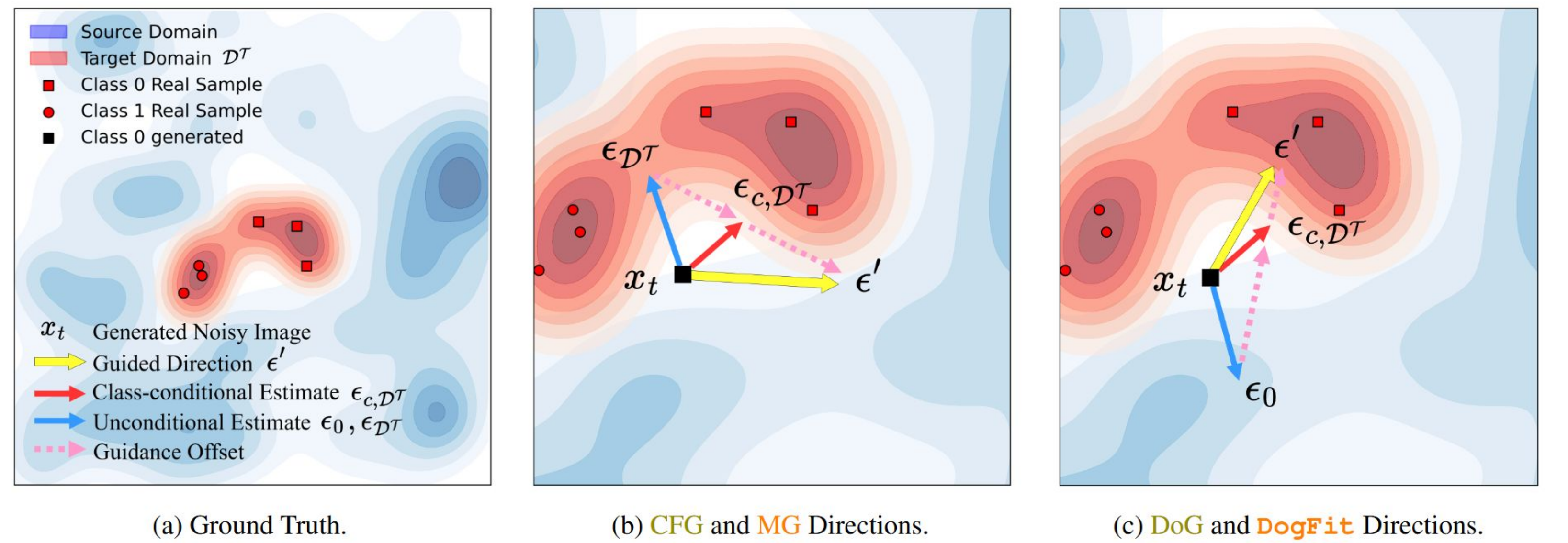


Figure 1: Transfer learning via guidance mechanisms for synthetic data created using a mixture of Gaussian distributions. (a) The red region defines the target domain, while the blue background represents the source distribution. (b) CFG and MG prioritize class separability without considering the source distribution, often pushing samples toward out-of-distribution areas in the target domain. (c) **DogFit** and DoG utilize source information to emphasize movement towards the core of the target manifold, improving domain alignment. (*) **Sampling-time** guidance methods (DoG and CFG) operate by computing the guidance offset, whereas **training-time** guidance methods (**DogFit** and MG) learn the **guided direction** directly.

Fast Sampling, Simple Training, and Guidance Control

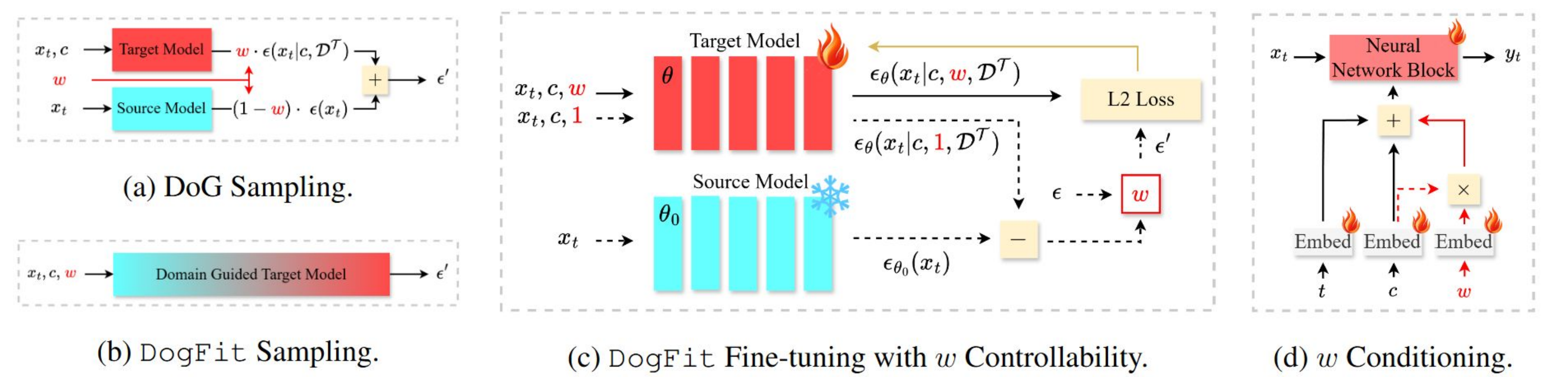


Figure 2: Illustration of DogFit during training and sampling. (a-b) DogFit internalizes guidance behavior, removing the need for inference-time double forward passes. (c) During training, the guidance value is treated as an input, allowing inference-time control. The simple case of our method, with fixed guidance, simply requires removing the w and 1 from the target model conditions. (d) w is embedded it as an extra input during fine-tuning, allowing inference-time control. (*) Dashed lines denote paths that do not propagate gradients.

Results

Higher Quality with Lower Sampling Cost

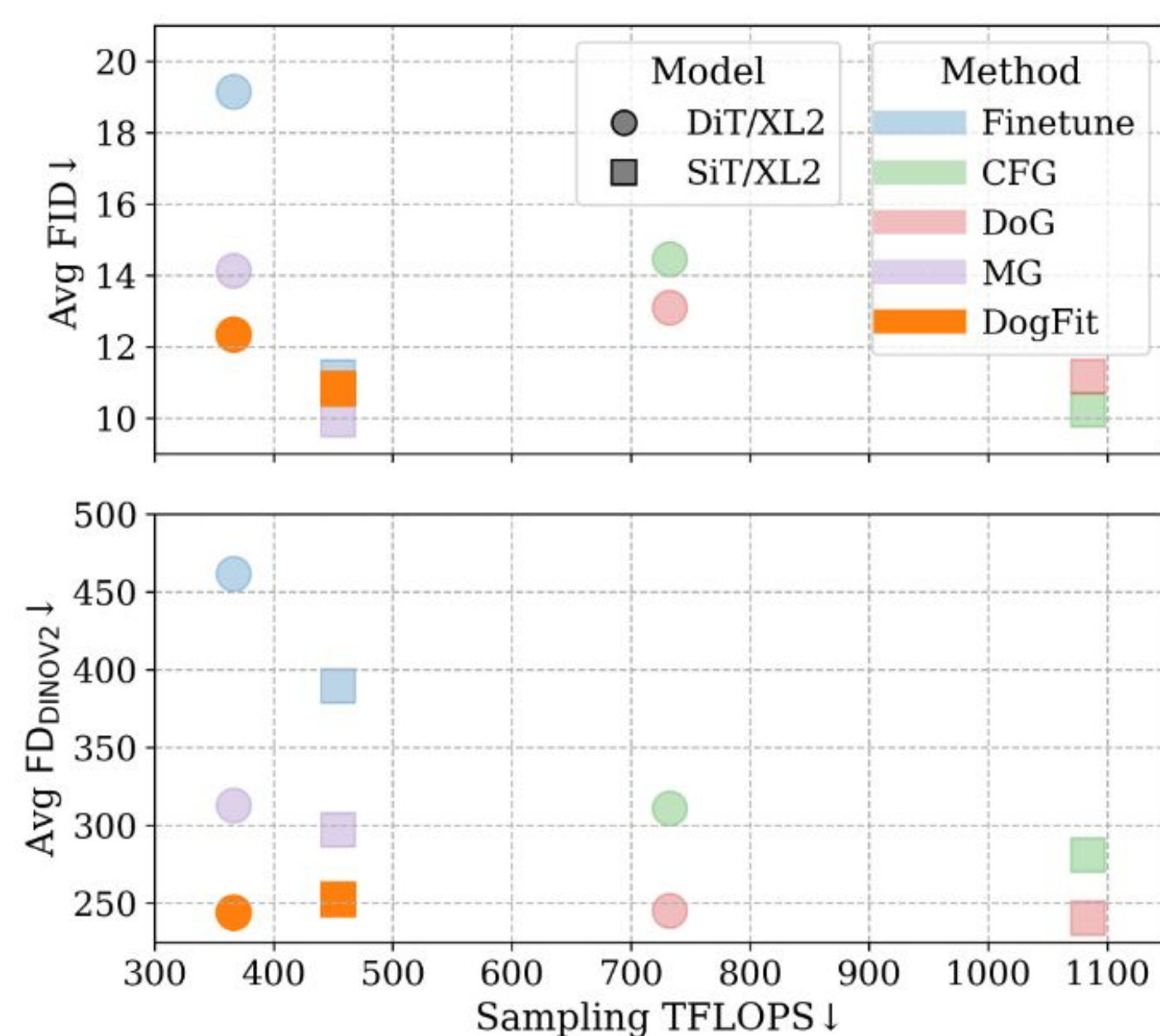


Figure 3: DogFit outperforms SOTA guidance methods in majority of cases for diffusion transfer learning, without increasing sampling overhead. Results are averaged across six target datasets.

Ablations on #Steps, PEFT^[4], SEDD

# Steps	Food FID↓		Art FID↓	
	MG	DogFit	MG	DogFit
10	60.60	57.91	108.26	88.84
25	21.13	18.97	39.80	29.11
50	11.13	10.64	19.91	16.32
100	7.60	8.09	13.38	12.70

Table 1: Varying the number of sampling steps for MG and DogFit.

DiT-XL2 Variant	FID↓		Train Params (M)
	Food	Art	
DogFit	10.64	16.32	675.42 (100%)
+ DiffFit	11.86	15.80	0.75 (0.11%)

Table 2: Using DogFit with Parameter-efficient Fine-tuning (PEFT).

Test-time ω	Uniform				SEDD ($\lambda = 2$)
	$\mathcal{U}[1, 1.1]$	$\mathcal{U}[1, 1.25]$	$\mathcal{U}[1, 1.5]$	$\mathcal{U}[1, 2]$	
1	19.02	16.68	20.08	22.08	12.98
1.5	19.62	13.84	18.27	18.05	10.94
2	29.65	15.97	26.29	24.23	13.05
3	61.56	28.28	53.28	48.00	19.81
4	85.62	53.24	70.60	65.12	24.84
5	114.97	98.75	88.39	77.53	27.95

Table 3: Ablating training-time guidance sampling methods, by comparing SEDD and uniform sampling. Results indicate FID↓ across a wide range of test-time ω values.

Controllable Guidance Knob at Test-time

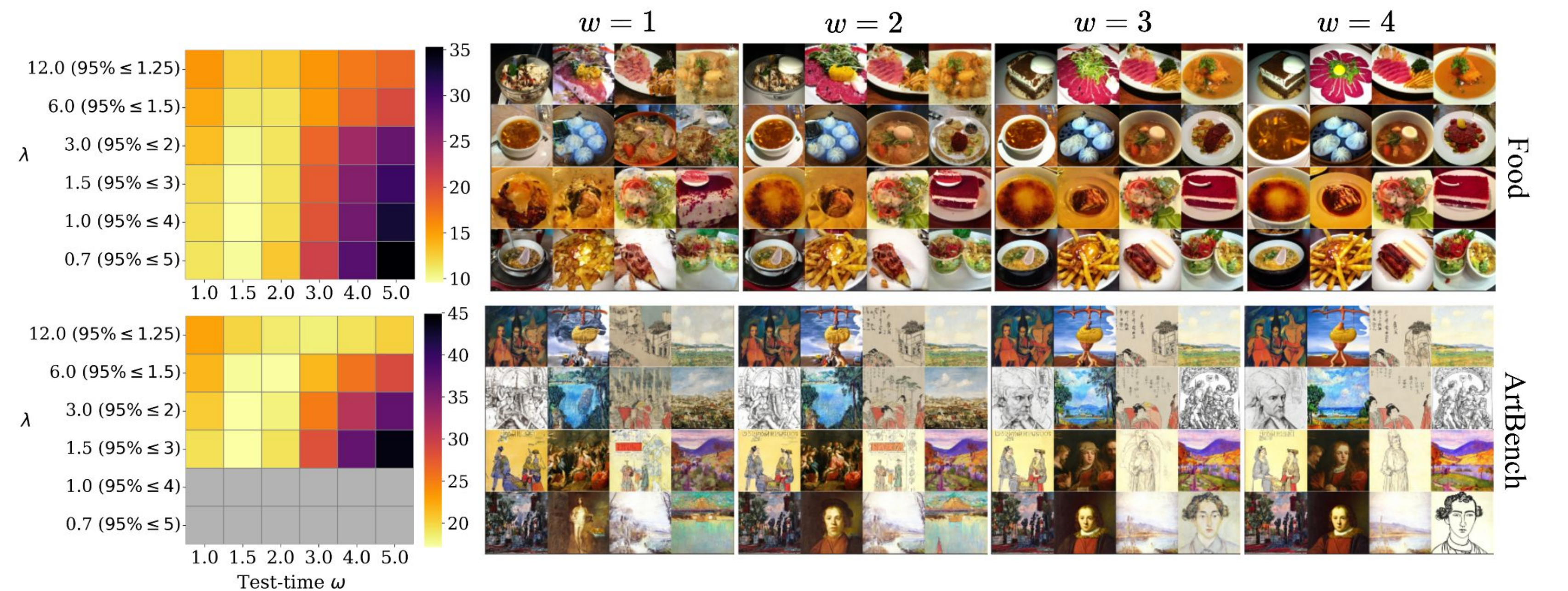


Figure 4: Effect of controllable guidance at test time for the Food and ArtBench domains. (Left) FID behavior across test-time w s with different training-time λ s. (Right) Corresponding generated samples for fixed $\lambda = 3$ (95% of sampled w values in $[1, 2]$), across varying test-time w . Gray cells indicate extreme FID values (≈ 350) resulting from collapsed generations.

When and Where to Guide Matters

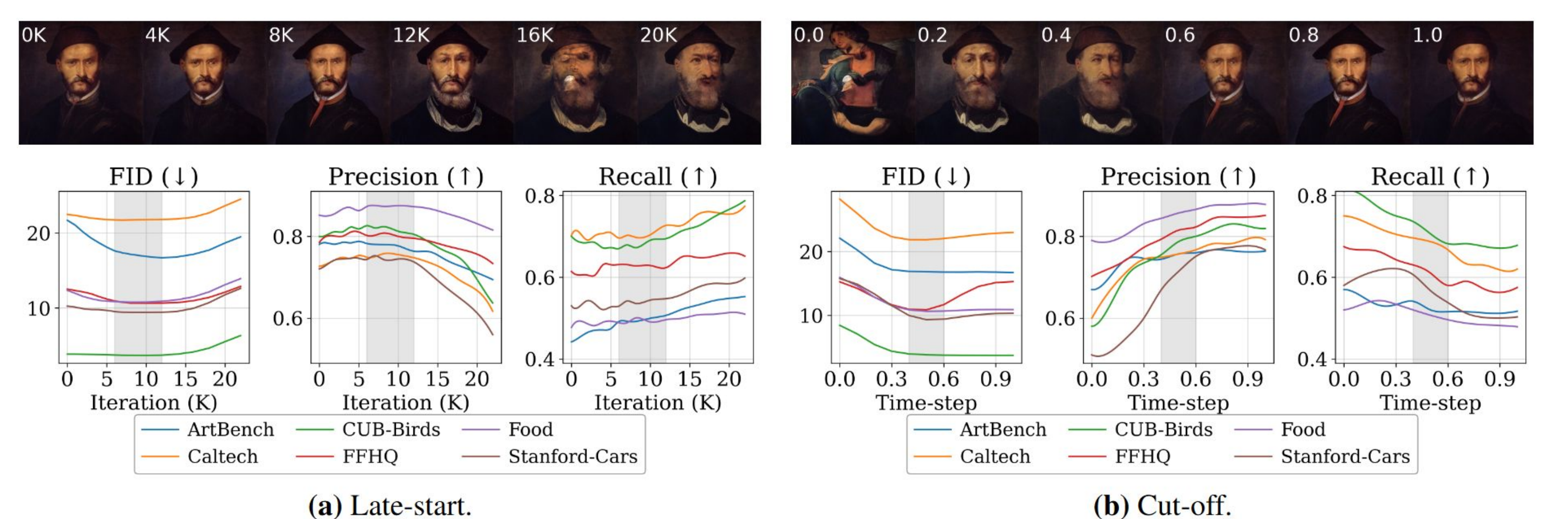


Figure 5: Ablation study on guidance schedules in DogFit. (a) Varying the late-start threshold τ_s to control when guidance begins. (b) Varying the cut-off threshold τ_c to restrict guidance to later denoising steps. Performed using FID on DiT-XL/2.

References

- Ho, Jonathan, and Tim Salimans. "Classifier-free diffusion guidance." *arXiv preprint* (2022).
- Zhong, Jincheng, et al. "Domain guidance: A simple transfer approach for a pre-trained diffusion model." *ICLR* 2025.
- Tang, Zhicong, et al. "Diffusion models without classifier-free guidance." *arXiv preprint* (2025).
- Xie, Enze, et al. "DiffFit: Unlocking transferability of large diffusion models via simple parameter-efficient fine-tuning." *ICCV* 2023.