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

Yara Bahram, Mohammadhadi Shateri, Eric Granger

yara.mohammadi-bahram@livia.etsmtl.ca

LIVIA LABORATOIRE D'IMAGERIE, DE VISION ET D'INTELLIGENCE ARTIFICIELLE
ÉTS Le génie pour l'industrie | ÉCOLE DE TECHNOLOGIE SUPÉRIEURE Université du Québec | ILLS International Laboratory on Learning Systems

Main Contributions

- ➔ We propose a controllable training-time guidance 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 marginals for guidance than the learning target. We analyze the timing and placement of this guidance.
- ➔ For transfer learning across six target datasets on DiT and SiT models, `DogFit` outperforms SOTA guidance methods in majority of cases while being up to 2x faster.

Motivation and Method

Diffusion transfer learning is still annoying

- ➔ **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**.

Domain-guided Fine-tuning (`DogFit`) to the rescue

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

`DogFit` tweaks the diffusion training loss

$$\mathcal{L}_{\text{DogFit}} = \mathbb{E}_{t,(x_0,c),\epsilon}\left\|\boldsymbol{\epsilon}_\theta(x_t|c,\mathcal{D}^{\mathcal{T}}) - \boldsymbol{\epsilon}'\right\|^2,$$
$$\boldsymbol{\epsilon}' = \boldsymbol{\epsilon} + (w-1)\cdot \text{sg}\left(\boldsymbol{\epsilon}_\theta(x_t|c,\mathcal{D}^{\mathcal{T}}) - \boldsymbol{\epsilon}_{\theta_0}(x_t)\right).$$

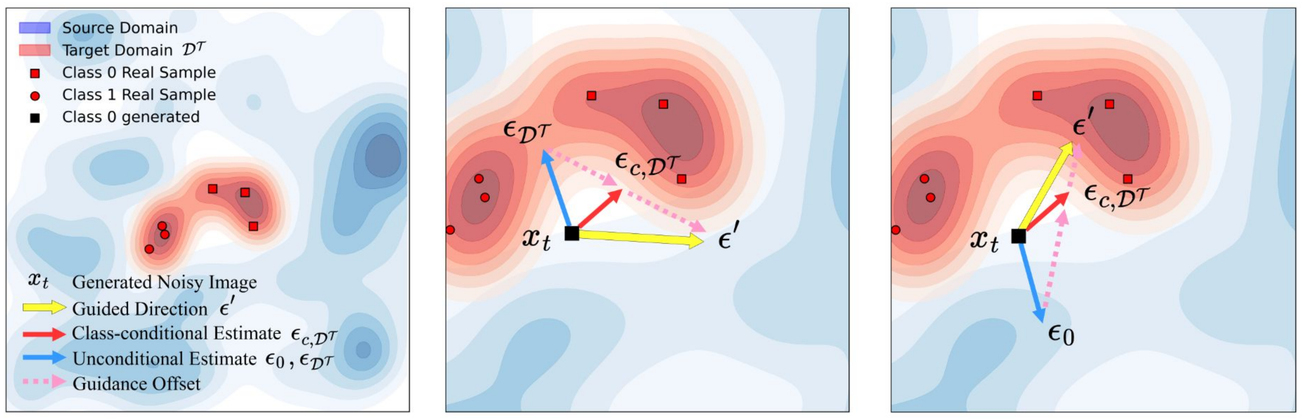
Adding test-time guidance control

$$\mathcal{L}_{\text{DogFit}} = \mathbb{E}_{t,(x_0,c,\boldsymbol{w}),\epsilon}\left\|\boldsymbol{\epsilon}_\theta(x_t|c,\boldsymbol{w},\mathcal{D}^{\mathcal{T}}) - \boldsymbol{\epsilon}'\right\|^2,$$
$$\boldsymbol{\epsilon}' = \boldsymbol{\epsilon} + (\boldsymbol{w}-1)\cdot \text{sg}\left(\boldsymbol{\epsilon}_\theta(x_t|c,\mathbf{1},\mathcal{D}^{\mathcal{T}}) - \boldsymbol{\epsilon}_{\theta_0}(x_t)\right),$$

➔ Sample from Shifted exponentially decaying distribution:

$$\boldsymbol{w} = 1 + z, \quad z \sim \mathcal{P}(z),$$
$$\mathcal{P}(z) = \lambda e^{-\lambda z}, \quad z \geq 0,$$

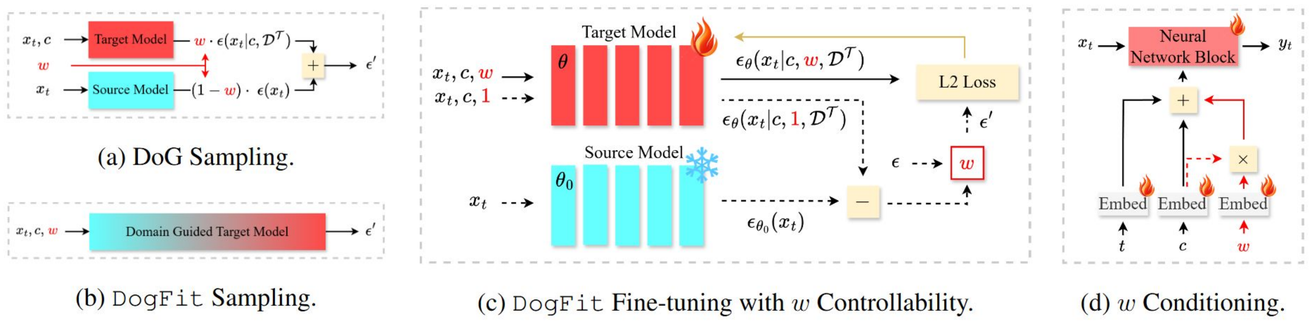
`DogFit` and friends for guiding generation in transfer learning



(a) Ground Truth. (b) CFG and MG Directions. (c) DoG and `DogFit` Directions.

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



(a) DoG Sampling. (b) `DogFit` Sampling. (c) `DogFit` Fine-tuning with w Controllability. (d) w Conditioning.

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 less sampling cost

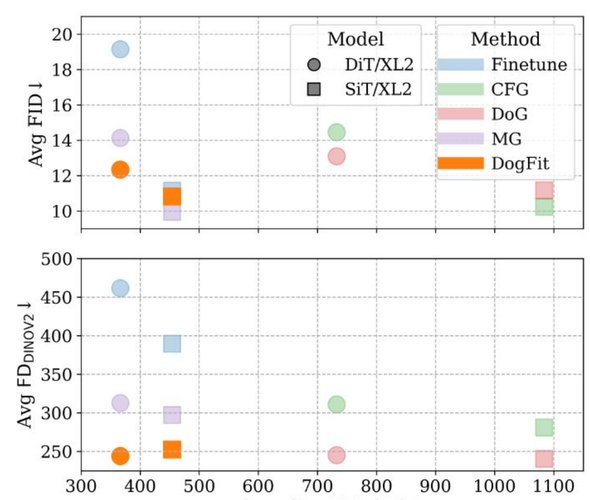


Figure 3: `DogFit` achieves competitive average FID and FD_{DINOV2} when compared to SOTA guidance mechanisms for diffusion transfer learning, all without increasing the computational complexity (sampling TFLOPS). Averages are computed over six target datasets.

Can be combined with PEFT[4] for training efficiency

DiT-XL/2 Variant	FID↓ Food	FID↓ Art	Train Params (M)
`DogFit`	**10.64**	16.32	675.42 (100%)
+ DiffFit	11.86	**15.80**	**0.75 (0.11%)**

Table 3: Combining `DogFit` with DiffFit.

Controllable guidance knob at test-time

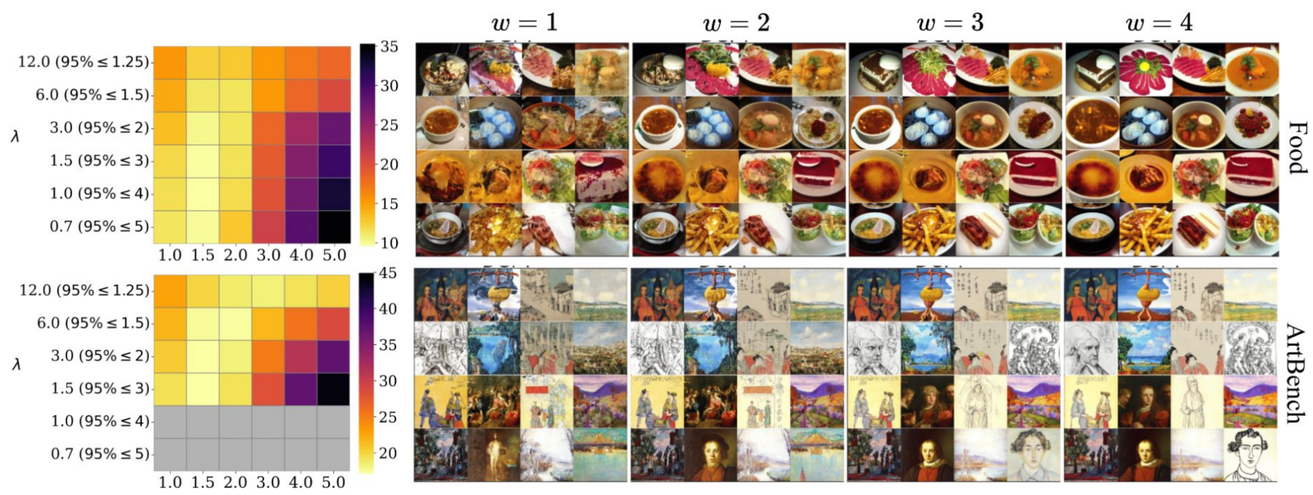
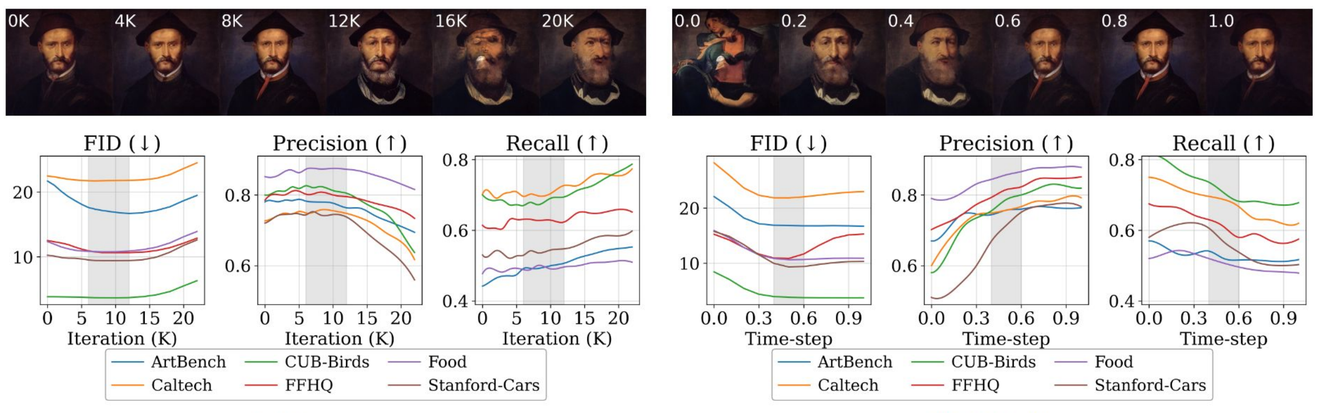


Figure 4: Effect of controllable guidance at test time for the Food and ArtBench domains. (Left) FID behavior across test-time ws with different training-time λs. (Right) Corresponding generated samples for fixed $\lambda = 3$ (95% of sampled w values in $[1, 2]$), across varying test-time w. Gray cells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

When and where to guide matters



(a) Late-start. **(b)** Cut-off.

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

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