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

Yara Bahram
PhD student

Mohammadhadi Shateri
Professor

Eric Granger
Professor

LIVIA, ILLS, ÉTS Montréal, Canada

yara.mohammadi-bahram@livia.etsmtl.ca

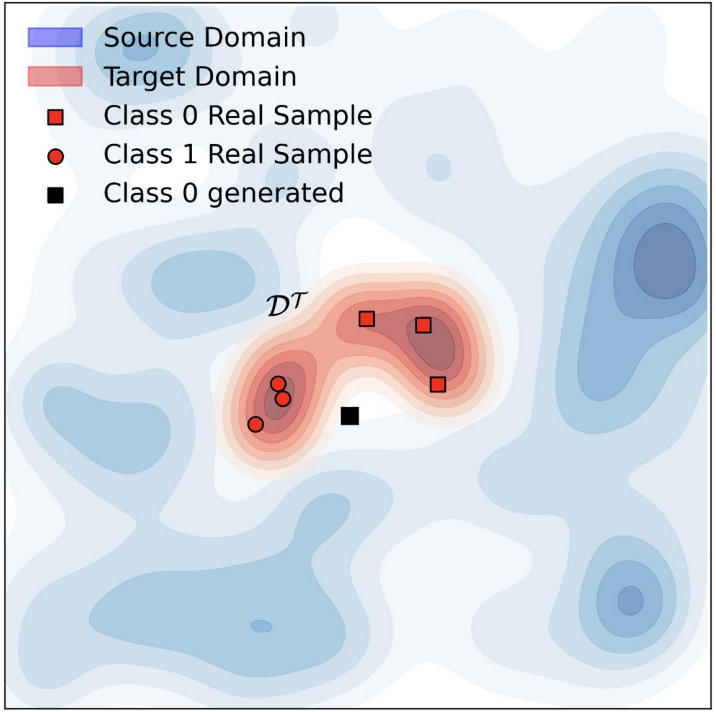
LIVIA LABORATOIRE D'IMAGERIE, DE VISION ET D'INTELLIGENCE ARTIFICIELLE

ÉTS Le génie pour l'industrie — ÉCOLE DE TECHNOLOGIE SUPÉRIEURE Université du Québec

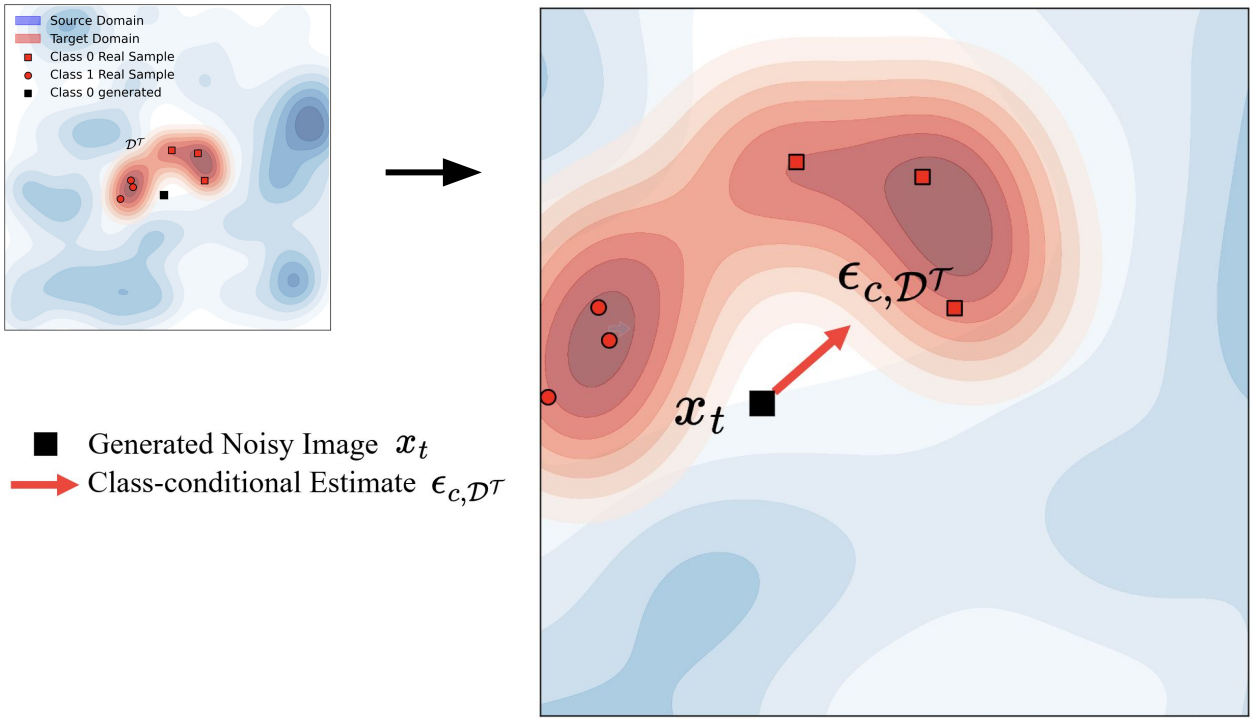
ILLS International Laboratory on Learning Systems

The 40th Annual AAAI Conference on Artificial Intelligence
JANUARY 20 – JANUARY 27, 2026 | SINGAPORE

Transfer Learning for Generation

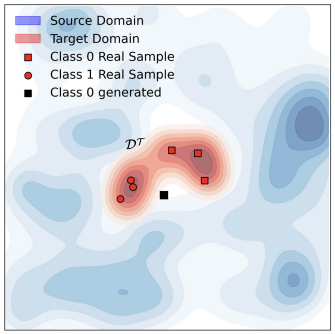


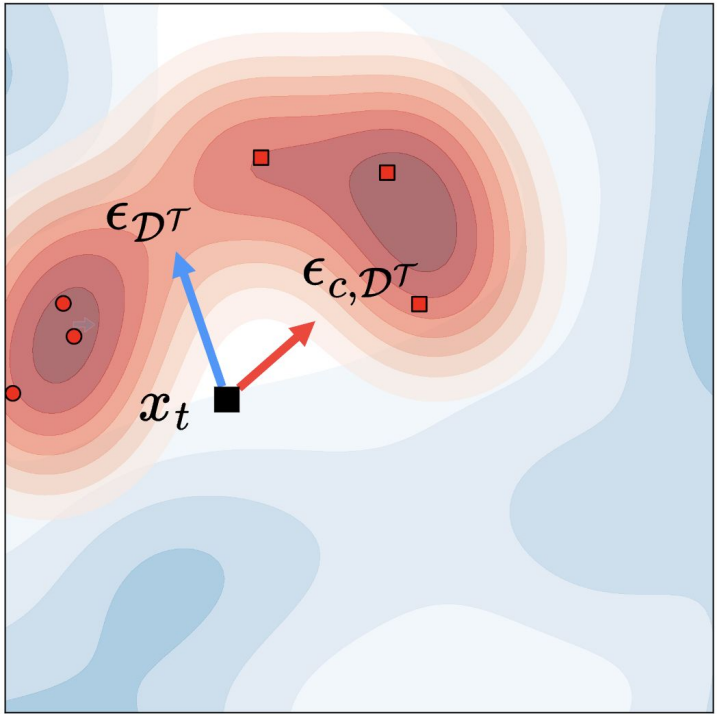
Sampling via Classifier-free Guidance [1]



[1] Ho, Jonathan, and Tim Salimans. "Classifier-free diffusion guidance." *arXiv* preprint (2022).

Sampling via Classifier-free Guidance [1]

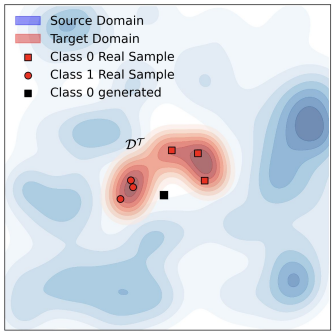


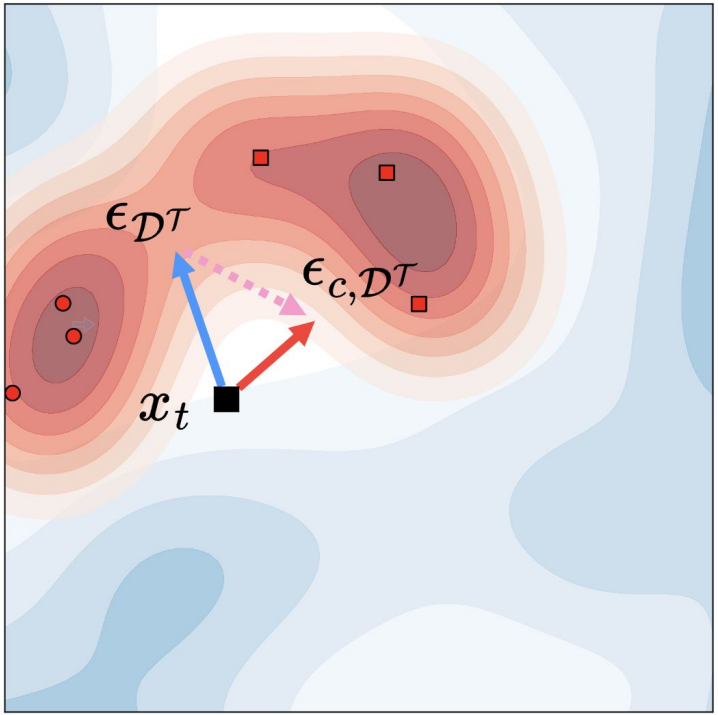


Generated Noisy Image x_t
Class-conditional Estimate $\epsilon_{c,\mathcal{D}^{\mathcal{T}}}$
Unconditional Estimate ϵ_0 , $\epsilon_{\mathcal{D}^{\mathcal{T}}}$

[1] Ho, Jonathan, and Tim Salimans. "Classifier-free diffusion guidance." *arXiv* preprint (2022).

Sampling via Classifier-free Guidance [1]

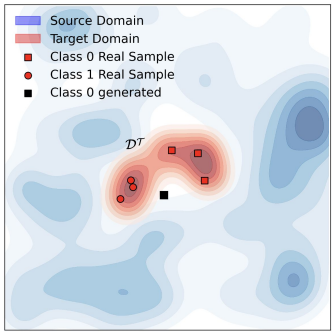


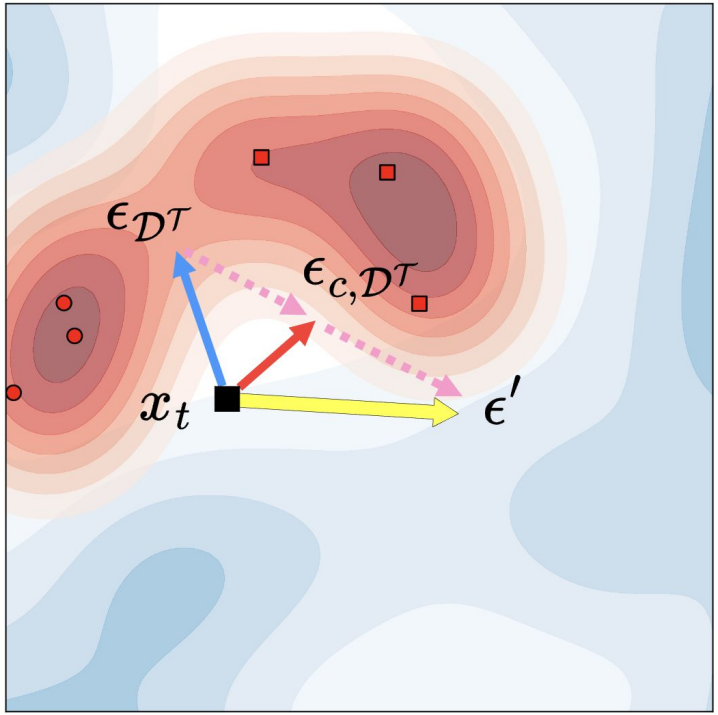


Generated Noisy Image x_t
Class-conditional Estimate $\epsilon_{c,\mathcal{D}^{\mathcal{T}}}$
Unconditional Estimate ϵ_0 , $\epsilon_{\mathcal{D}^{\mathcal{T}}}$
Guidance Offset

[1] Ho, Jonathan, and Tim Salimans. "Classifier-free diffusion guidance." *arXiv* preprint (2022).

Sampling via Classifier-free Guidance [1]

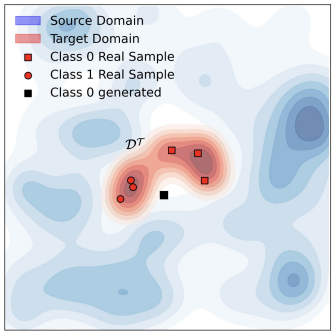


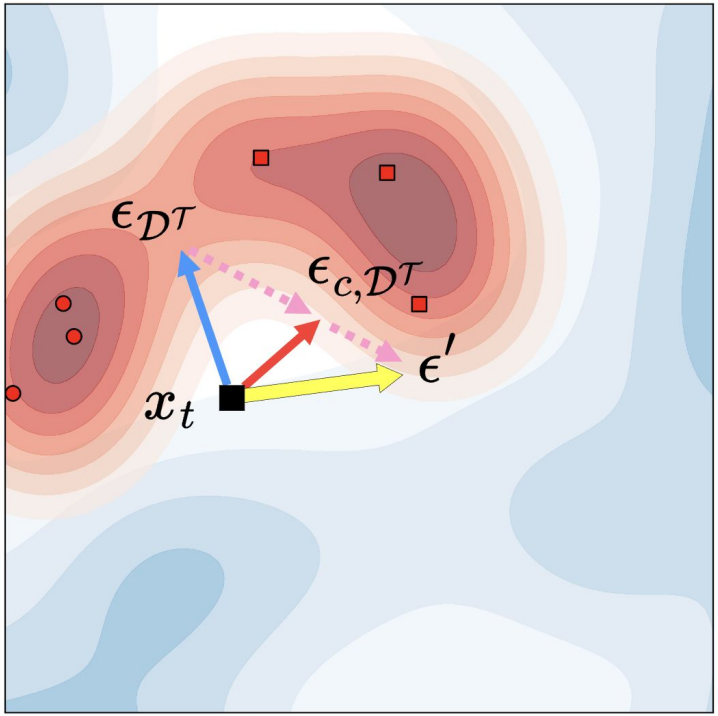


- Generated Noisy Image x_t
- Class-conditional Estimate $\epsilon_{c,\mathcal{D}^{\mathcal{T}}}$
- Unconditional Estimate ϵ_0 , $\epsilon_{\mathcal{D}^{\mathcal{T}}}$
- Guidance Offset
- Guided Direction ϵ'

[1] Ho, Jonathan, and Tim Salimans. "Classifier-free diffusion guidance." *arXiv* preprint (2022).

Sampling via Classifier-free Guidance [1]

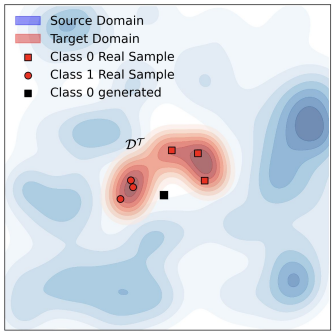


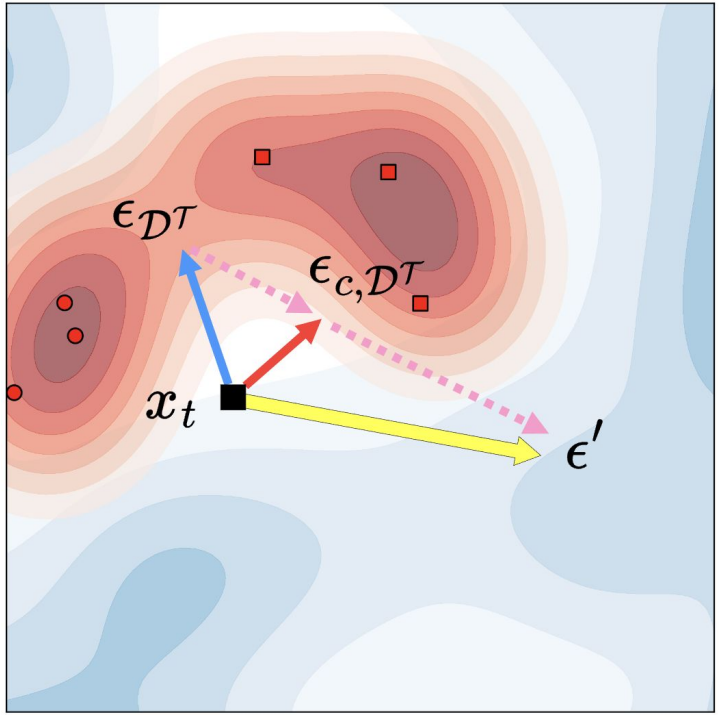


- Generated Noisy Image x_t
- Class-conditional Estimate $\epsilon_{c,\mathcal{D}^{\mathcal{T}}}$
- Unconditional Estimate ϵ_0 , $\epsilon_{\mathcal{D}^{\mathcal{T}}}$
- Guidance Offset
- Guided Direction ϵ'

[1] Ho, Jonathan, and Tim Salimans. "Classifier-free diffusion guidance." *arXiv* preprint (2022).

Sampling via Classifier-free Guidance [1]



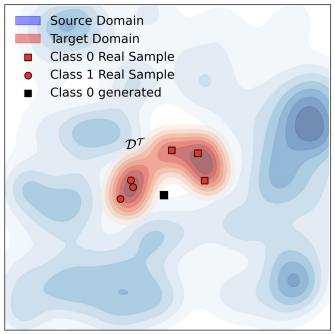


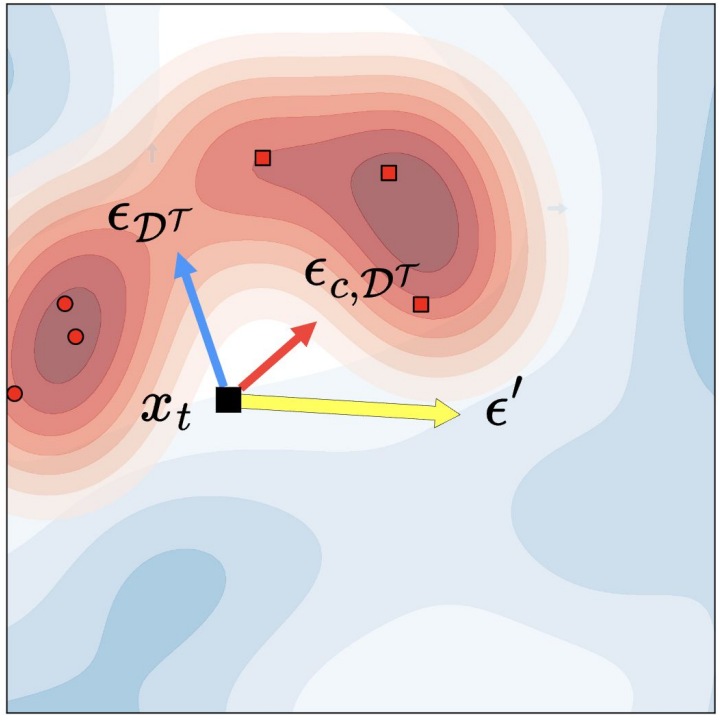
- Generated Noisy Image x_t
- Class-conditional Estimate $\epsilon_{c,\mathcal{D}^{\mathcal{T}}}$
- Unconditional Estimate ϵ_0 , $\epsilon_{\mathcal{D}^{\mathcal{T}}}$
- Guidance Offset
- Guided Direction ϵ'

- ➔ Costly sampling
- ➔ Weak unconditional estimate:
  - ◆ OOD generation
  - ◆ Not enough target data

[1] Ho, Jonathan, and Tim Salimans. "Classifier-free diffusion guidance." *arXiv* preprint (2022).

Sampling via Model Guidance [1]



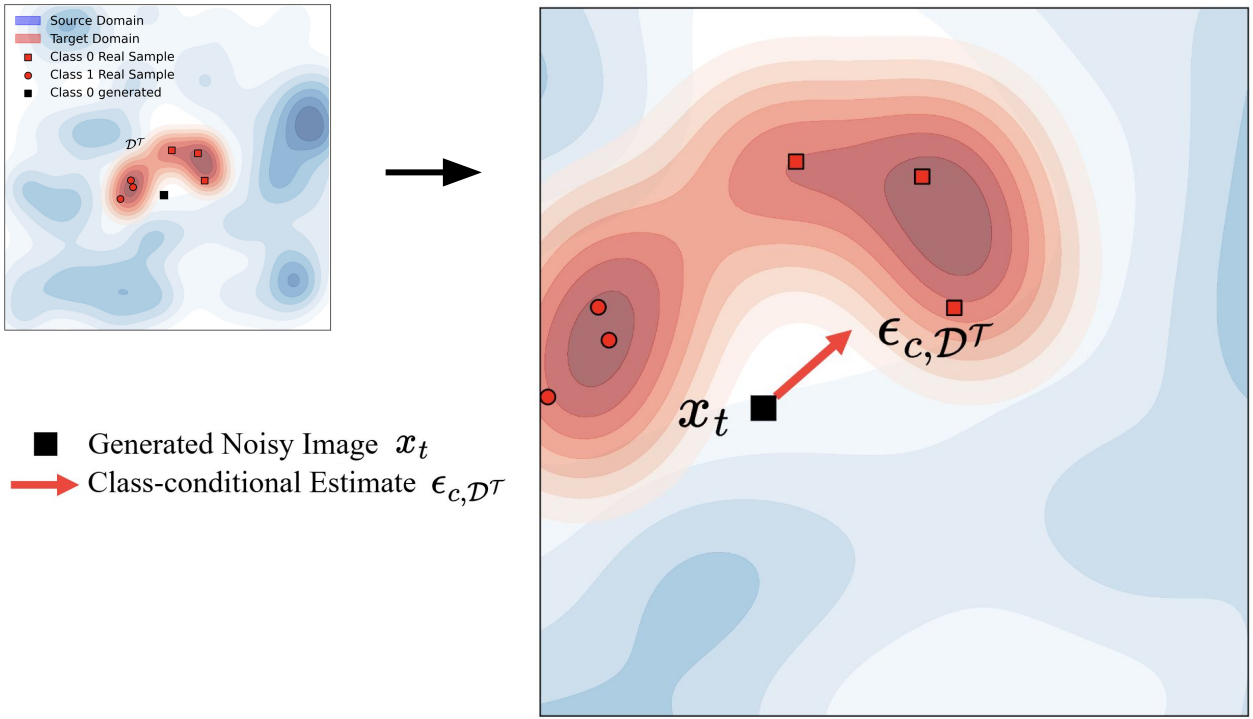


- Generated Noisy Image x_t
- Class-conditional Estimate $\epsilon_{c,\mathcal{D}^{\mathcal{T}}}$
- Unconditional Estimate ϵ_0 , $\epsilon_{\mathcal{D}^{\mathcal{T}}}$
- Guidance Offset
- Guided Direction ϵ'

➔ No test-time control

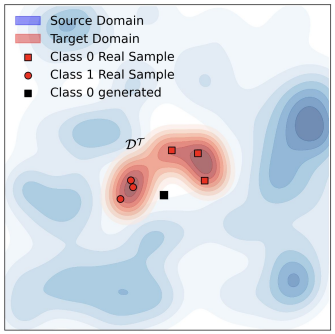
[1] Tang, Zhicong, et al. "Diffusion models without classifier-free guidance." *arXiv* preprint arXiv:2502.12154 (2025).

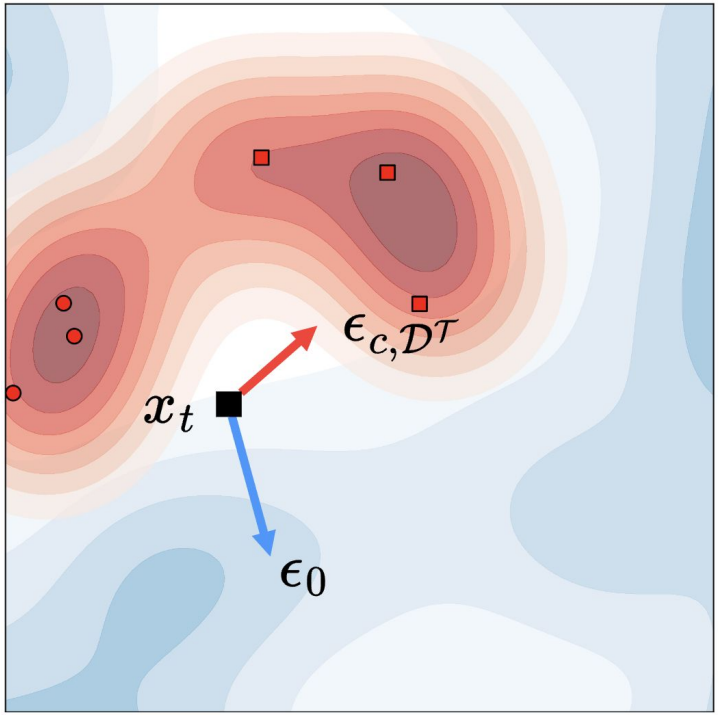
Sampling via Domain Guidance [1]



[1] Zhong, Jincheng, et al. "Domain Guidance: A Simple Transfer Approach for a Pre-trained Diffusion Model." ICLR. 2025

Sampling via Domain Guidance [1]

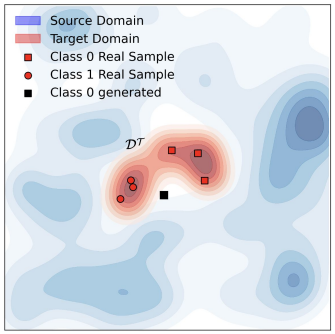


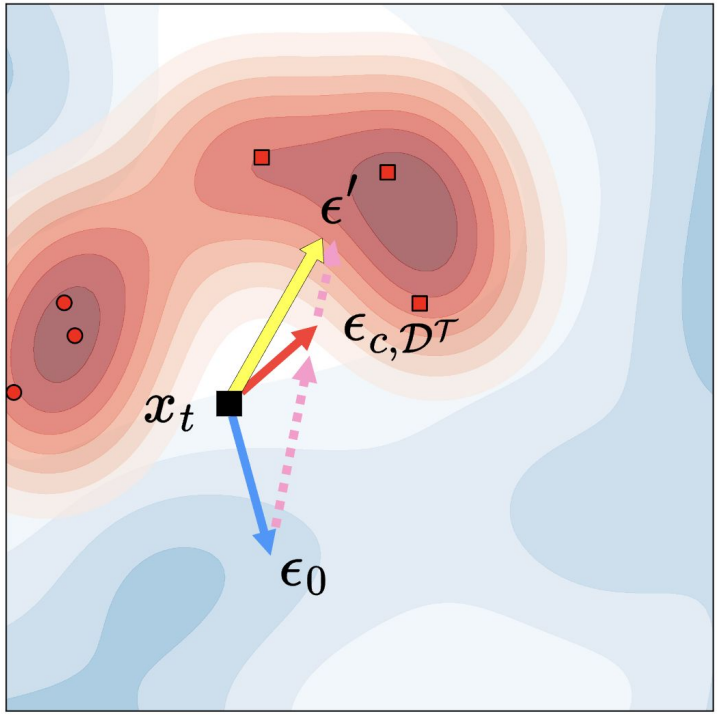


Generated Noisy Image x_t
Class-conditional Estimate $\epsilon_{c,\mathcal{D}^{\mathcal{T}}}$
Unconditional Estimate ϵ_0 , $\epsilon_{\mathcal{D}^{\mathcal{T}}}$

[1] Zhong, Jincheng, et al. "Domain Guidance: A Simple Transfer Approach for a Pre-trained Diffusion Model." ICLR. 2025

Sampling via Domain Guidance [1]

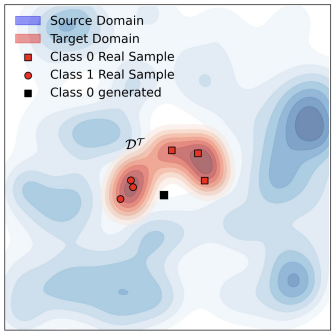


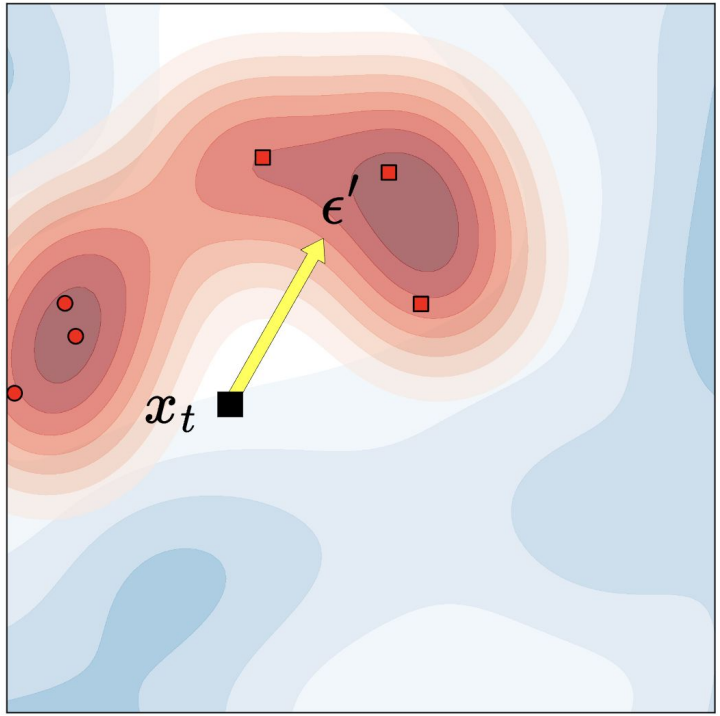


- Generated Noisy Image x_t
- Class-conditional Estimate $\epsilon_{c,\mathcal{D}^{\mathcal{T}}}$
- Unconditional Estimate ϵ_0 , $\epsilon_{\mathcal{D}^{\mathcal{T}}}$
- Guidance Offset
- Guided Direction ϵ'

[1] Zhong, Jincheng, et al. "Domain Guidance: A Simple Transfer Approach for a Pre-trained Diffusion Model." ICLR. 2025

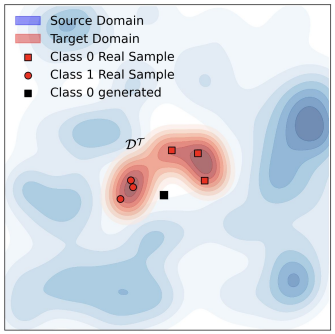
Sampling via Domain-guided Fine-tuning (`DogFit`)

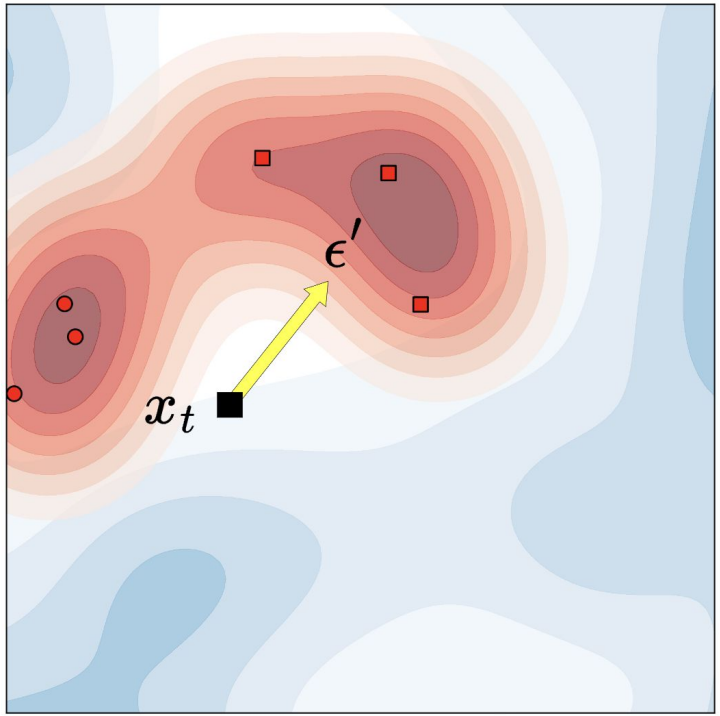




- Generated Noisy Image x_t
- Guided Direction ϵ'
- Unconditional Estimate ϵ_0 , $\epsilon_{\mathcal{D}^{\mathcal{T}}}$
- Guidance Offset
- Guided Direction ϵ'

Sampling via Domain-guided Fine-tuning (`DogFit`)

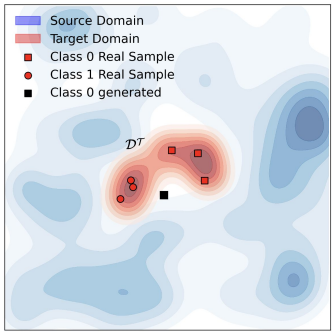


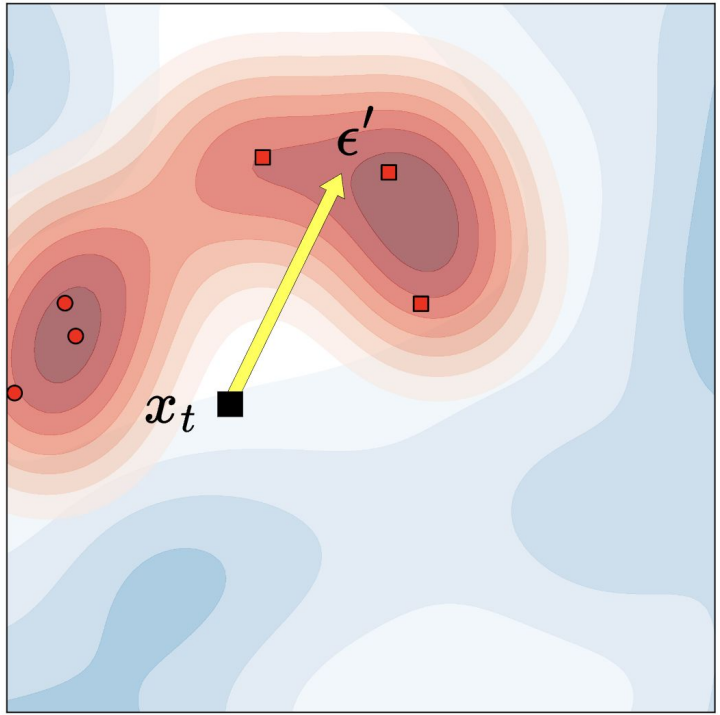


Generated Noisy Image x_t

Guided Direction ϵ'

Sampling via Domain-guided Fine-tuning (`DogFit`)

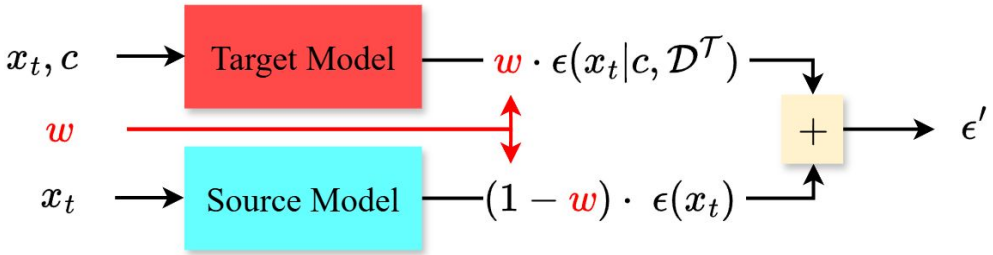




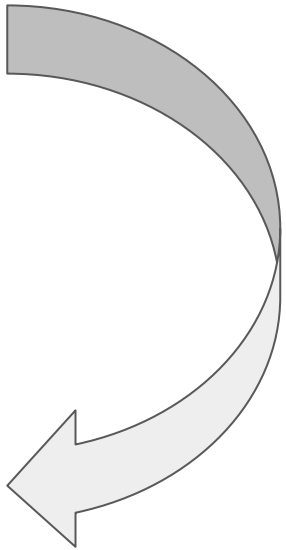
Generated Noisy Image x_t

Guided Direction ϵ'

Domain Guidance Sampling



`DogFit` sampling



DogFit Training

Learning Target Model

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

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

Guidance Weight · Stop Gradient · Source Model

Adding Test-time Guidance Control

Sample Guidance Weight · Input Guidance

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

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

Input No Guidance

Equivalence to Domain Guidance

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

Assume:

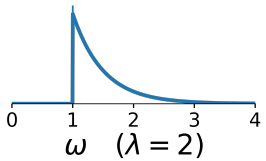
1. When $w = 1$, the model accurately recovers ϵ in training.
2. The model's prediction varies linearly with the guidance strength.

➔ High exposure to $w = 1$.
➔ Smaller w values [1].

Then: For any sampled w at test-time, **`DogFit`** ≈ Domain Guidance

➔ Sample w from shifted exponentially decaying distribution (SEDD):

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



[1] Zheng, Candi, and Yuan Lan. "Characteristic guidance: non-linear correction for diffusion model at large guidance scale." ICML. 2024

Relation to Model Guidance (MG)

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

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

$$\epsilon'_{\texttt{DogFit}} = \epsilon'_{\text{MG}} - \sigma(w-1)\cdot\nabla_{x_t}\log p_\theta(\mathcal{D}^{\mathcal{T}}|x_t).$$

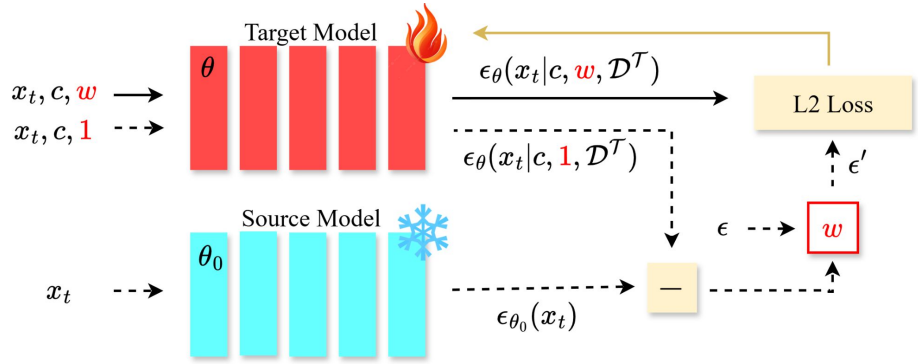
Domain Alignment Score

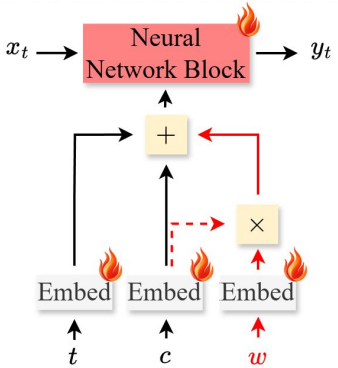
Domain Classifier Gradient

DogFit:

- ➔ Maintains MG's class-conditional guidance.
- ➔ Avoids OOD generation by pushing toward the core of the target domain manifold.

`DogFit` Training on Diffusion Transformers





Experimental Setting

Backbones:

- ➔ Diffusion Transformers (DiT-XL/2)
- ➔ Scalable Interpolant Transformers (SiT-XL/2)

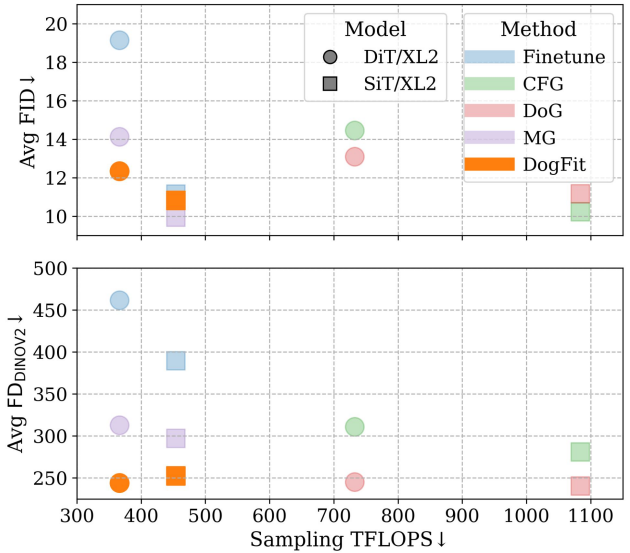
Metrics: FID, FD_{DINOV2}

Source domain: ImageNet256

Target domains:

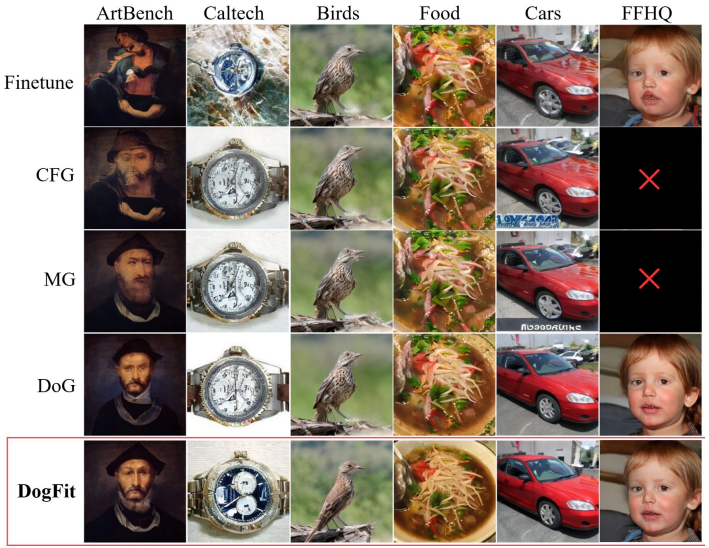
- ➔ Birds, ArtBench, Caltech, CUB-birds, Food,Stanford-Cars, FFHQ

Higher Quality with Lower Sampling Cost

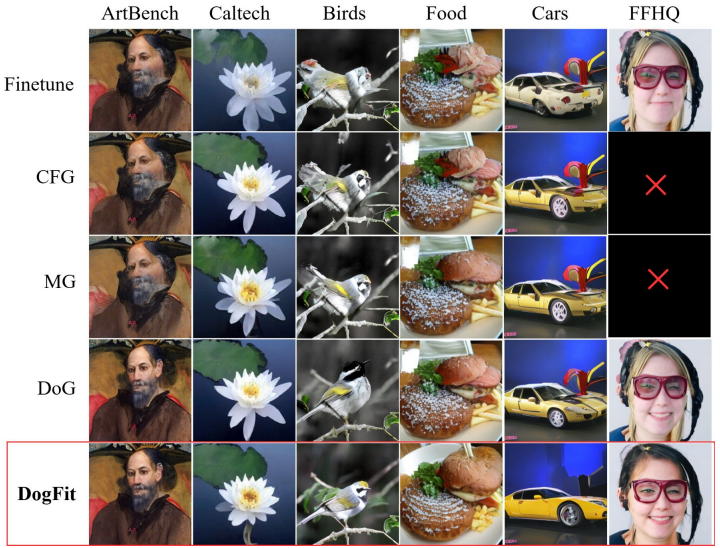


	Metric	Method	Unlabeled	Labeled						Sampling Cost	
			FFHQ	ArtBench	Caltech	CUB-Birds	Food	Stanford-Cars	Avg.	Passes	TFLOPS
DiT/XL-2	FID ↓	Fine-tuning	15.94	23.36	30.02	9.35	16.75	16.24	19.14	**x1**	**366.14**
		+ CFG (Ho and Salimans 2022)	–	20.83	24.07	5.03	11.77	10.60	14.46	x2	732.28
		+ DoG (Zhong et al. 2025)	13.87	17.30	23.76	**3.65**	10.97	9.77	13.09	x2	732.28
		MG (Tang et al. 2025)	–	19.91	23.71	4.85	11.13	11.03	14.13	**x1**	**366.14**
		DogFit	**10.48**	**16.32**	**21.68**	3.69	10.64	**9.35**	**12.34**	**x1**	**366.14**
		DogFit + Control	13.03	16.98	21.98	3.69	**10.44**	10.05	12.63	**x1**	**366.14**
	FD$_{DINOV2}$ ↓	Fine-tuning	461.45	360.77	529.85	428.21	610.31	378.11	461.45	**x1**	**366.14**
		+ CFG (Ho and Salimans 2022)	–	314.32	401.96	200.92	410.48	227.46	311.03	x2	732.28
		+ DoG (Zhong et al. 2025)	**273.93**	266.25	377.49	**135.74**	314.16	**132.91**	**245.31**	x2	732.28
		MG (Tang et al. 2025)	–	299.06	409.20	220.40	379.39	255.88	312.78	**x1**	**366.14**
		DogFit	282.32	269.06	**377.15**	143.42	**293.24**	147.20	246.01	**x1**	**366.14**
		DogFit + Control	274.67	**261.72**	378.86	144.39	314.12	141.08	248.03	**x1**	**366.14**
SiT/XL-2	FID ↓	Fine-tuning	**7.63**	9.66	26.10	4.92	7.76	10.53	11.79	**x1**	**454.01**
		+ CFG (Ho and Salimans 2022)	-	9.28	23.21	3.49	7.95	9.71	10.73	x2	1083.77
		+ DoG (Zhong et al. 2025)	**7.63**	11.22	23.53	**3.39**	7.95	9.71	11.16	x2	1083.77
		MG (Tang et al. 2025)	-	9.64	**23.10**	3.60	6.84	**8.62**	**10.36**	**x1**	**454.01**
		DogFit	12.44	9.62	23.45	3.52	7.85	10.05	10.91	**x1**	**454.01**
		DogFit + Control	10.8	**9.03**	23.15	3.59	**6.52**	10.10	10.48	**x1**	**454.01**
	FD$_{DINOV2}$ ↓	Fine-tuning	335.15	271.92	519.54	405.07	522.47	283.29	400.46	**x1**	**454.01**
		+ CFG (Ho and Salimans 2022)	-	236.13	406.63	203.47	368.26	190.24	280.94	x2	1083.77
		+ DoG (Zhong et al. 2025)	335.15	**215.12**	**387.89**	160.85	298.65	**139.09**	240.32	x2	1083.77
		MG (Tang et al. 2025)	-	232.17	418.86	229.32	359.90	203.23	288.70	**x1**	**454.01**
		DogFit	**278.10**	222.20	403.44	181.48	296.49	159.83	252.29	**x1**	**454.01**
		DogFit + Control	319.16	230.23	399.20	**152.50**	**234.98**	183.41	**240.06**	**x1**	**454.01**

Qualitative Comparison



(**a**) DiT-XL/2



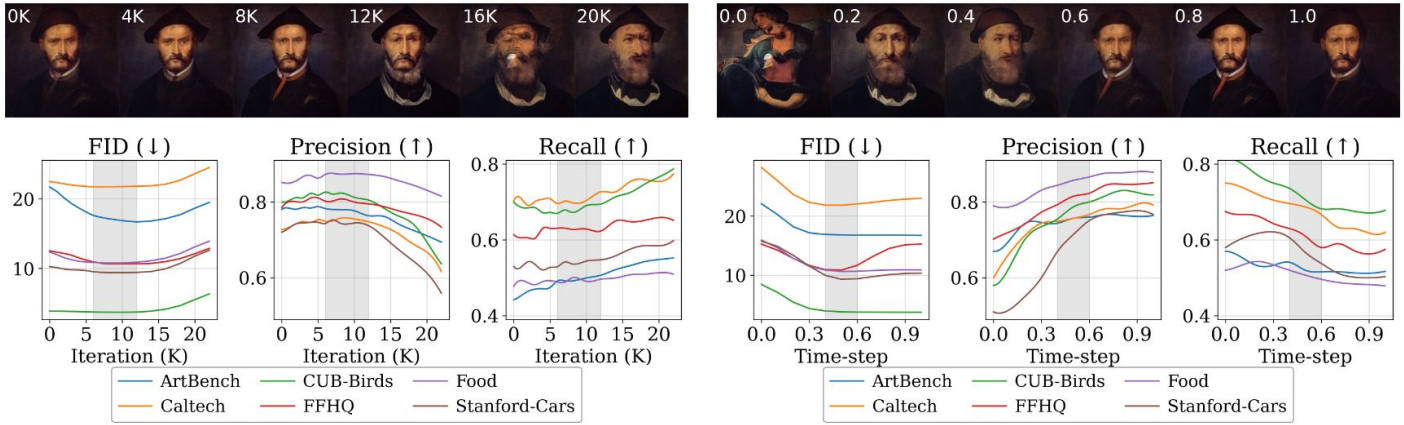
(**b**) SiT-XL/2

Controllable Guidance Knob at Test-time

(a) Training Guidance Values

(b) Sampling Guidance Values

When and Where to Guide Matters



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

Ablations on #Steps, PEFT [1], 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-XL/2 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.

[1] Xie, Enze, et al. "Difffit: Unlocking transferability of large diffusion models via simple parameter-efficient fine-tuning." *ICCV* 2023.

Summary

- ➔ We propose **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.

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

Yara Bahram
PhD student

Mohammadhadi Shateri
Professor

Eric Granger
Professor

LIVIA, ILLS, ÉTS Montréal, Canada

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

The 40th Annual AAAI Conference on Artificial Intelligence
JANUARY 20 – JANUARY 27, 2026 | SINGAPORE

Training Algorithm

Supplementary

Algorithm 1: `DogFit` training with Controllable Guidance and Scheduling

Require: Target dataset $\{\mathcal{D}^{\mathcal{T}}\}$, noise schedule $\bar{\alpha}$, fixed source model ϵ_{θ_0}, training model ϵ_θ initialized by ϵ_{θ_0}, guidance cut-off τ_{c}, late-start step τ_{s}.

for training step $s = 1$ to S **do**
  Sample target data $(x_0, c) \sim \{\mathcal{D}^{\mathcal{T}}\}$
  Sample noise $\epsilon \sim \mathcal{N}(0, 1)$ and time-step $t \sim \mathcal{U}(0, 1)$
  Sample guidance strength $\boldsymbol{w} = 1 + z, \quad z \sim \mathcal{P}(z)$
  Add noise: $x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} \cdot \epsilon$
  if $s > \tau_{\text{s}}$ **and** $t < \tau_{\text{c}}$ **then**
    Compute guided target:
    $\epsilon' = \epsilon + (\boldsymbol{w} - 1) \cdot \text{sg}\left(\epsilon_\theta(x_t \mid c, \mathbf{1}, \mathcal{D}^{\mathcal{T}}) - \epsilon_{\theta_0}(x_t)\right)$
  else
    $\epsilon' = \epsilon$
  end if
  Compute loss: $\mathcal{L}_{\texttt{DogFit}} = \left\|\epsilon_\theta(x_t \mid c, \boldsymbol{w}, \mathcal{D}^{\mathcal{T}}) - \epsilon'\right\|^2$
  Backpropagate and update: $\theta = \theta - \eta \nabla_\theta \mathcal{L}_{\texttt{DogFit}}$
end for

More Ablations

Supplementary

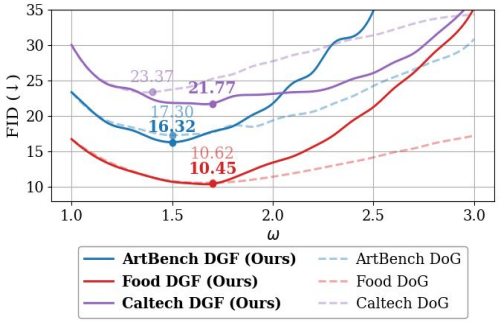


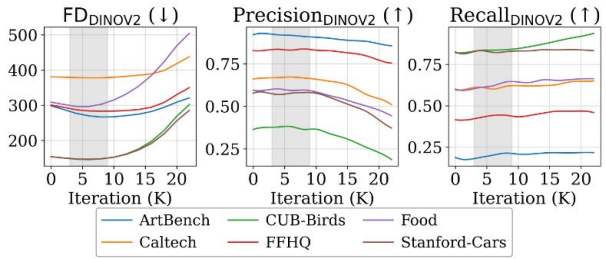
Figure 10: Comparing `DogFit` (without guidance control) and DoG in sensitivity of FID to w.

Method	FID↓	KID↓ ($\times 10^{-4}$)	FD_{DINOv2}↓	#Pass
Fine-tuning	15.73	90.41	510.60	**x1**
+ CFG	11.93 (+24.16%)	57.69 (+36.19%)	**366.00** (+28.32%)	x2
`DogFit`	**11.06** (+29.69%)	**43.15** (+52.27%)	380.44 (+25.49%)	**x1**

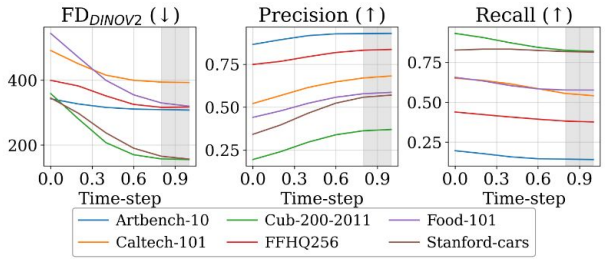
Table 6: Adapting DiT-XL/2 to Food101 at 512×512 resolution. Percent improvements relative to Fine-tuning are in parentheses.

Method	FID↓	KID↓ ($\times 10^{-4}$)	FD_{DINOv2}↓	#Pass
Fine-tuning	36.08	24.23	258.62	**x1**
+ CFG	33.38 (+7.48%)	21.25 (+12.30%)	257.73 (+0.34%)	x2
`DogFit`	**29.30** (+18.79%)	**16.51** (+31.86%)	**221.03** (+14.53%)	**x1**

Table 7: Adapting DiT-XL/2 to the distant EuroSat domain at 256×256.



(a) Late-start ablation



(b) Cut-off ablation

Figure 9: Ablation 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 FD_{DINOV2} on DiT-XL/2.

Precision and Recall Analysis on DiT/XL-2 Supplementary

Table 4: Precision and Recall results on DiT/XL-2 backbone using Inception and DINOV2 metrics.

Metric	Method	Unlabelled	Labelled					
		FFHQ	ArtBench	Caltech	CUB-Birds	Food	Stanford-Cars	Avg.
Precision (Inception)	Fine-tuning	0.70	0.67	0.60	0.58	0.79	0.51	0.63
	+ CFG (Ho and Salimans 2022)	-	0.68	0.76	0.72	0.82	0.63	0.72
	+ DoG (Zhong et al. 2025)	**0.83**	**0.77**	**0.81**	0.82	**0.88**	0.77	**0.81**
	MG (Tang et al. 2025)	-	0.68	0.76	0.71	0.81	0.62	0.72
	DogFit	0.81	0.76	0.75	0.82	0.87	0.74	0.79
	DogFit + Control	**0.83**	0.76	0.78	**0.83**	0.87	**0.80**	**0.81**
Precision (DINOV2)	Fine-tuning	0.74	0.84	0.48	0.16	0.41	0.30	0.44
	+ CFG (Ho and Salimans 2022)	-	0.83	0.67	0.30	0.48	0.47	0.52
	+ DoG (Zhong et al. 2025)	0.82	0.91	**0.73**	**0.39**	**0.59**	**0.61**	**0.65**
	MG (Tang et al. 2025)	-	0.84	0.63	0.28	0.51	0.44	0.54
	DogFit	0.83	**0.92**	0.67	0.38	0.58	0.58	0.62
	DogFit + Control	**0.86**	0.90	0.67	0.38	0.58	0.59	0.62
Recall (Inception)	Fine-tuning	**0.67**	**0.57**	**0.75**	**0.81**	0.52	0.58	**0.65**
	+ CFG (Ho and Salimans 2022)	-	0.56	0.66	0.76	0.53	**0.59**	0.62
	+ DoG (Zhong et al. 2025)	0.61	0.50	0.65	0.68	0.48	0.52	0.57
	MG (Tang et al. 2025)	-	0.56	0.67	0.75	**0.55**	**0.59**	0.624
	DogFit	0.63	0.51	0.70	0.68	0.50	0.55	0.59
	DogFit + Control	0.57	0.51	0.68	0.67	0.50	0.48	0.57
Recall (DINOV2)	Fine-tuning	**0.46**	0.21	**0.66**	**0.95**	**0.67**	0.82	**0.66**
	+ CFG (Ho and Salimans 2022)	-	**0.22**	0.59	0.87	0.61	**0.84**	0.63
	+ DoG (Zhong et al. 2025)	0.44	0.21	0.57	0.83	0.64	0.83	0.62
	MG (Tang et al. 2025)	-	**0.22**	0.60	**0.89**	0.60	**0.84**	0.63
	DogFit	0.43	0.21	0.60	0.83	0.64	0.83	0.62
	DogFit + Control	0.43	0.21	0.61	0.85	0.64	0.83	0.63

Text-to-Image Initial Results

Supplementary

Method	DINO ↑	CLIP-I ↑	CLIP-T ↑	#Pass
Fine-tuning	0.383	0.697	0.275	**x1**
+ CFG	**0.529** (+38.1%)	**0.799** (+14.6%)	**0.304** (+10.5%)	x2
DogFit	0.516 (+34.7%)	0.758 (+8.8%)	0.269 (−2.2%)	**x1**

Table 8: Comparison on the text-to-image DreamBooth benchmark for fine-tuning Stable Diffusion V1.5. Results are based on the best guidance strength per method ($\omega_{\text{CFG}} = 7.5,\ \omega_{\texttt{DogFit}} = 1.65$).

Guidance	CLIP-T ↑	#Pass
Real data	**0.347**	–
Fine-tuning	0.276	**x1**
+ CFG	0.343 (+24.3%)	x2
DogFit	0.319 (+15.6%)	**x1**

Table 9: Comparison on text-to-image LoRA fine-tuning of Stable Diffusion v1.5 to the Pokémon image–caption dataset. Results are based on the best guidance strength per method ($\omega_{\text{CFG}} = 7.5,\ \omega_{\texttt{DogFit}} = 1.75$).

Figure 12: Qualitative DreamBooth text-to-image generations (SD v1.5) comparing CFG and DogFit. Black images correspond to Not-Safe-For-Work (NSFW) content detection using hugging face generators.

End.