LoFT: LoRA-Fused Training Dataset Generation with Few-shot Guidance

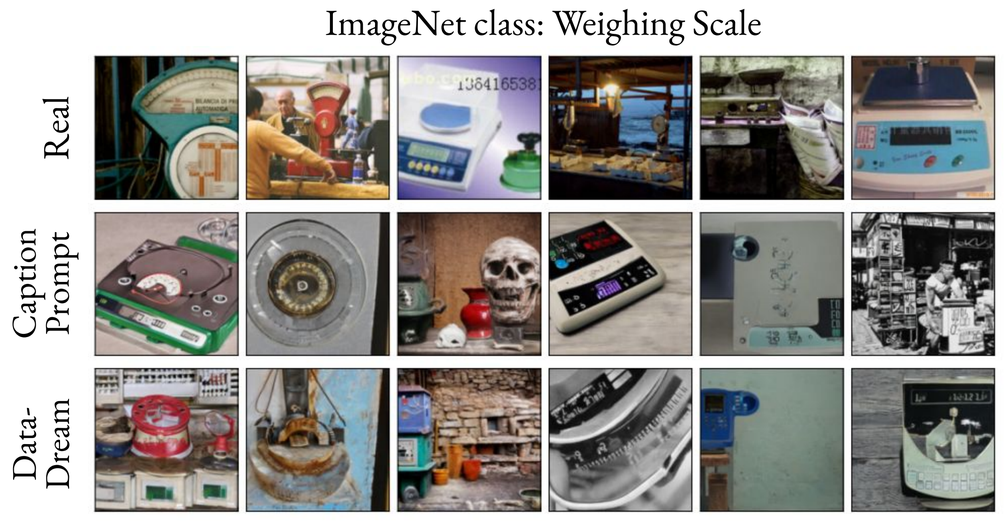
Jae Myung Kim[1,2] Stephan Alaniz[5] Cordelia Schmid[4] Zeynep Akata[2,3]

[1] Eberhard Karls Universität Tübingen [2] Helmholtz Munich [3] Technical University of Munich [4] Inria [5] Institut Polytechnique de Paris

Motivation

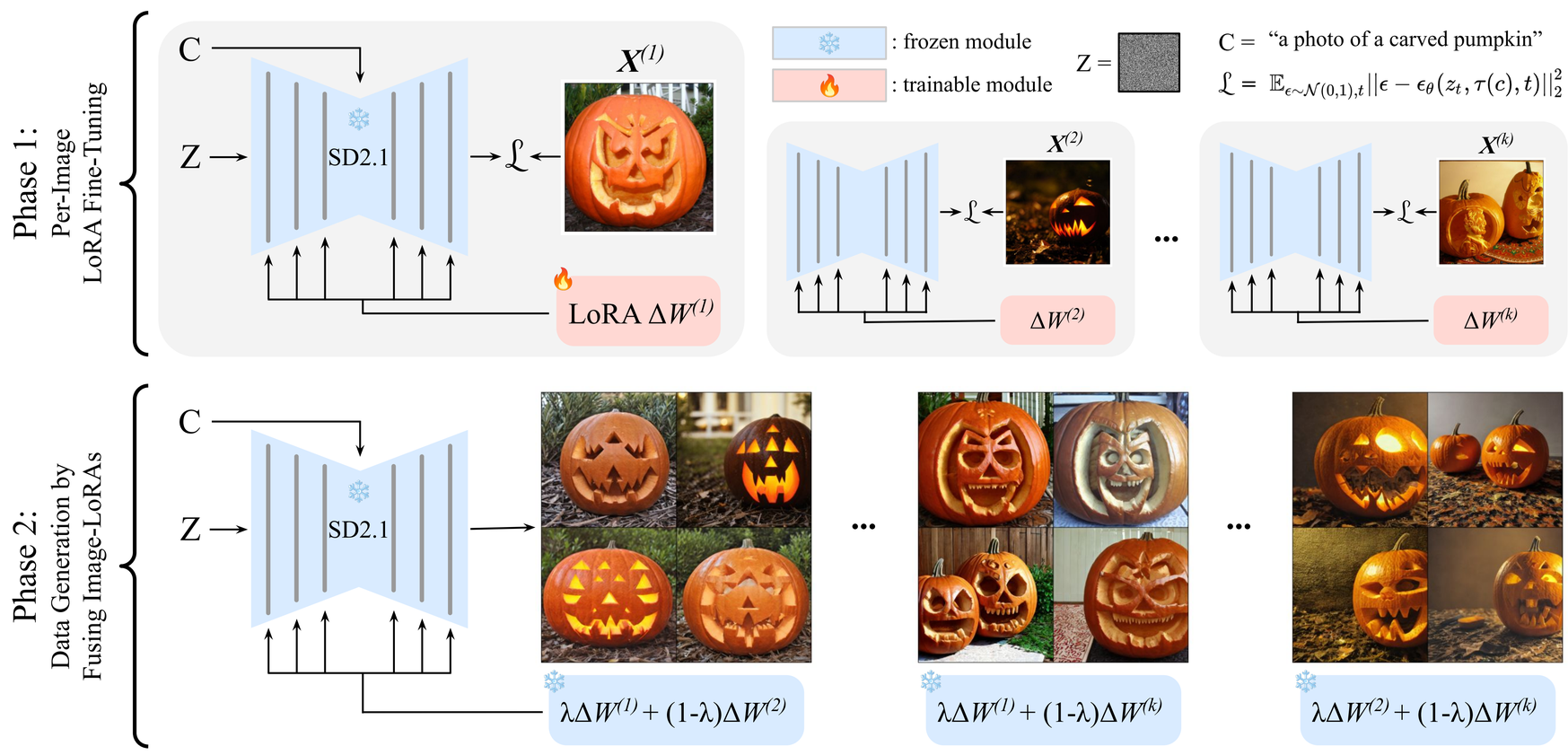
- Synthetic data is now a common choice for training models.
- Yet, it rarely provides substantial gain in supervised learning.
- Even SOTA DataDream [1] cannot reliably generate in-distribution images due to underfitting.

[1] DataDream: Few-shot Guided Dataset Generation, ECCV, 2024.
* CaptionPrompt uses LLM-generated textual prompt to generate images.
* DataDream finetunes generative model with few-shot real data.

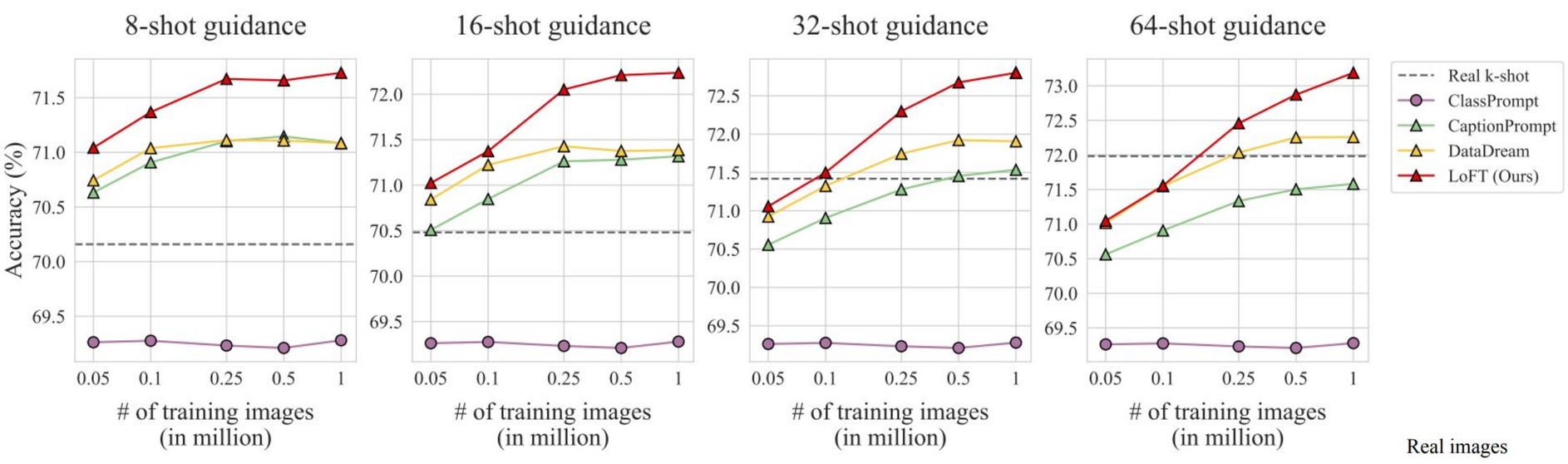


LoFT: LoRA-Fused Training Dataset Generation

- Goal: Generate synthetic data for classification task.
- Leverage few-shot real data.
- Phase1: Overfit generative model to each single image.
- Phase2: Fuse LoRA weights in the image generation phase.
- Fidelity is achieved in phase 1.
- Diversity is achieved in phase 2, by LoRA weight combination and noise random seed.



Evaluation Results

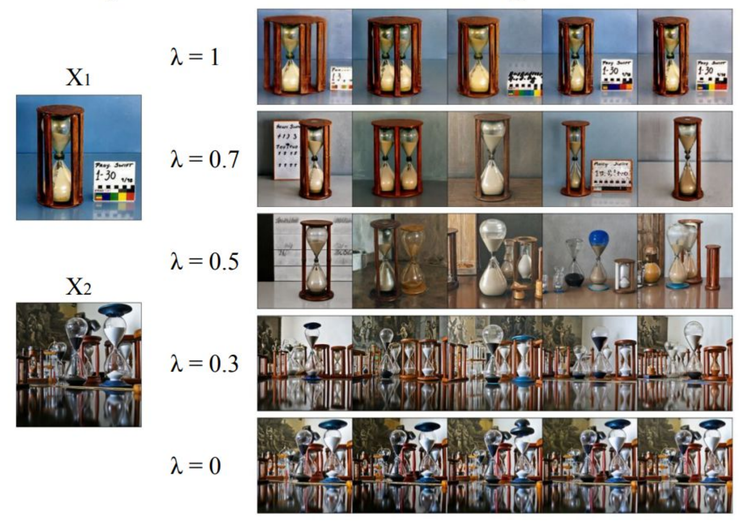


More details ↓

On ImageNet, LoFT consistently outperforms other methods in varied number of real few-shots and different number of synthetic images.

Method	Cal	DTD	Eur	Air	Pet	Car	SUN	Food	Flo	Avg
CLIP zero-shot	93.0	44.4	47.6	24.7	89.2	65.2	62.6	86.1	71.4	64.9
ClassPrompt	$93.9_{\pm 0.2}$	$53.7_{\pm 0.5}$	$46.2_{\pm 0.7}$	$26.5_{\pm 0.1}$	$92.5_{\pm 0.1}$	$74.1_{\pm 0.5}$	$67.0_{\pm 0.0}$	$85.1_{\pm 0.0}$	$71.9_{\pm 0.5}$	$67.9_{\pm 0.1}$
CaptionPrompt	$95.4_{\pm 0.3}$	$63.9_{\pm 1.0}$	$46.6_{\pm 1.7}$	$26.6_{\pm 0.1}$	$92.9_{\pm 0.1}$	$74.4_{\pm 0.1}$	$73.9_{\pm 0.2}$	$85.4_{\pm 0.0}$	$72.3_{\pm 0.5}$	$70.2_{\pm 0.1}$
DataDream	$96.0_{\pm 0.4}$	$64.9_{\pm 0.2}$	$84.1_{\pm 3.4}$	$61.4_{\pm 0.9}$	$\mathbf{93.5}_{\pm 0.2}$	$\mathbf{90.5}_{\pm 0.4}$	$74.4_{\pm 0.2}$	$\mathbf{86.5}_{\pm 0.0}$	$\mathbf{98.0}_{\pm 0.2}$	$83.2_{\pm 0.4}$
LoFT (Ours)	$\mathbf{96.7}_{\pm 0.2}$	$\mathbf{70.5}_{\pm 0.1}$	$\mathbf{86.8}_{\pm 2.2}$	$\mathbf{66.1}_{\pm 1.5}$	$93.2_{\pm 0.1}$	$89.3_{\pm 0.5}$	$\mathbf{75.5}_{\pm 0.0}$	$86.0_{\pm 0.0}$	$\mathbf{98.0}_{\pm 0.2}$	$\mathbf{84.7}_{\pm 0.2}$

LoFT outperforms other methods on 9 different benchmarks. 16-shot with 500 synthetic images.



Qualitative examples of hourglass.