ChimeraLoRA: Multi-Head LoRA-Guided Synthetic Datasets

Supplementary Material

Datasets	AIR	CAL	CAR	DTD	EUR	FLO	FOD	PET	**SUN**	AVG
IsSynth	34.5	95.7	65.9	67.6	71.3	90.0	85.4	92.1	72.2	75.0
DataDream	61.4	96.0	90.5	64.9	84.1	98.0	86.5	93.5	74.4	83.3
LoFT	66.1	96.7	89.3	70.5	86.8	98.0	86.0	93.2	75.5	84.7
Ours	**71.4**	**97.1**	**90.8**	**74.7**	**92.3**	**98.6**	**87.1**	**94.2**	**76.6**	**87.0**

Table 5. *Downstream performance with synthetic datasets under 16-shot scenarios.* Starting with 16-shot labels, we generate 500 additional images per class to train on 504 per class in total, improving accuracy by 2.3 percentage points on average over state-of-the-art baselines across nine datasets. Here, the baseline results are taken from the original papers.

A. Experiments on 16-Shot Setting

Table 5 shows that our method works well even in the 16-shot setting. Rather than using a single parameter α, we employ two concentration parameters, α and β, to define the Dirichlet distribution. Specifically, we sample weights $(w_1, \ldots, w_{16}) \sim \mathrm{Dir}(\mathbf{c})$, where $\mathbf{c} \in \mathbb{R}^{16}$ is defined as:

$$c_i = \begin{cases} \alpha & \text{if } i = k \\ \beta & \text{if } i \neq k \end{cases}, \quad k \sim \mathrm{Uniform}(\{1, \ldots, 16\}) . \tag{8}$$

We set $\alpha = 2$ and $\beta = 0.07$, yielding $\mathbb{E}[w_k] = \frac{\alpha}{\alpha+(16-1)\beta} \approx 66\%$ for the main adapter, with the remaining $\approx 34\%$ mass distributed over the other adapters. Choosing a small $\beta < 1$ promotes sparsity among the remaining adapters, i.e., the residual mass is typically concentrated on only a few adapters rather than being spread uniformly across all of them. In summary, as the number of shots increases, using a relatively larger α (relative to β) becomes beneficial for capturing fine-grained details by placing more emphasis on the main adapter, while still preventing too many adapters from being mixed in.

B. Computation Analyses

B.1. Resource Usage

As each image has its own LoRA head, resource usage naturally grows as the number of shots increases, leading to higher memory consumption and more trainable parameters. To keep this overhead manageable, we reduce the LoRA rank for multi-shot settings. In particular, the results reported in Table 1 for the 4-shot setting are obtained by setting the rank to one quarter of the baseline, which substantially lowers the per-head footprint while retaining competitive performance. Table 6 further summarizes the resulting parameter counts and training time, showing that our approach offers a practical middle ground between baselines in terms of both memory and speed.

Method	Rank	# parameters	Training time
DataDream	16	3.2M	**95s**
LoFT	4	**0.8M**	356s
Ours	4	1.9M	212s

Table 6. Trade-off between # parameters and training time.

B.2. Effect of Weight Sampling on Efficiency

In Section 4.2, when $\alpha = 1$ we draw $\mathbf{w} \sim \mathrm{Dirichlet}(\mathbf{1})$, i.e., $\mathbf{w}$ is uniformly distributed over the simplex Δ^{K-1}, and we compare three practical choices: *Uniform* w (fix $w_i = 1/K$), *Reuse* w (reuse a single $\mathbf{w} \sim \mathrm{Dirichlet}(\mathbf{1})$ for multiple images), and *New* w (sample a fresh $\mathbf{w}$ per image). Table 7 shows that moving from Uniform $\rightarrow$ Reuse $\rightarrow$ New increases the diversity of per-image mixtures and tends to improve accuracy, but also increases wall-time, where wall-time denotes the overall generation time, since per-image sampling reduces batching efficiency.

Method	Wall-Time	Accuracy
Uniform w	**4.6h**	59.37
Reuse w	5.8h	61.31
New w	6.1h	**61.51**

Table 7. Trade-off between wall-time and accuracy.

C. Additional Ablation Studies

C.1. Rank r Ablation

On the DTD dataset, we observe that increasing the LoRA rank r yields modest but consistent gains: with the number of synthetic images fixed at 100, performance improves from 55.07 at r=2 to 57.20 and 57.72 at r=8 and r=16, respectively. This trend aligns with the increased capacity of higher-rank adapters, but it also comes with higher computational cost, as larger r increases the number of trainable parameters and the associated memory and training-time overhead. Overall, we find that higher ranks provide incremental improvements, reflecting a typical trade-off between computation resources and performance.

C.2. α Ablation

In a symmetric Dirichlet distribution, α primarily controls the concentration of the mixture weights $\mathbf{w}$: smaller α produces spikier weights that are closer to selecting a single head, whereas larger α spreads mass more evenly and behaves like a more uniform ensemble. As summarized in Table 8, this induces a clear trade-off. With smaller α, synthesis tends to be more faithful and yields better FID, while larger α encourages broader mixing across heads and improves diversity, reflected by a higher distance score. Overall, α serves as a simple tunable parameter to balance fidelity and diversity depending on the desired operating point.

Metric	$\alpha = 0.2$	$\alpha = 1.0$	$\alpha = 5.0$
FID (Real vs. Synthetic, ↓)	**0.098**	<u>0.101</u>	0.146
Distance (Inter-class, ↑)	0.142	<u>0.153</u>	**0.199**

Table 8. *Comparison metrics under different α values.*

C.3. Semantic Boosting on Baselines

The table 9 summarizes the effect of applying semantic boosting to the baselines. Overall, semantic boosting consistently improves performance, and the gain is most pronounced for image-wise LoRA training, where optimization can otherwise overfit to spurious regions or drift away from the target instance.

Method	AIR	FLO
DataDream (+SB)	44.3 → 44.2	92.9 → <u>93.1</u>
LoFT (+SB)	41.7 → <u>44.4</u>	91.3 → 92.2
Ours	43.9 → **46.0**	93.1 → **93.4**

Table 9. *Effect of semantic boosting on baselines.*

D. Additional Qualitative Results

Figures 9 and 10 present additional qualitative results across diverse datasets.

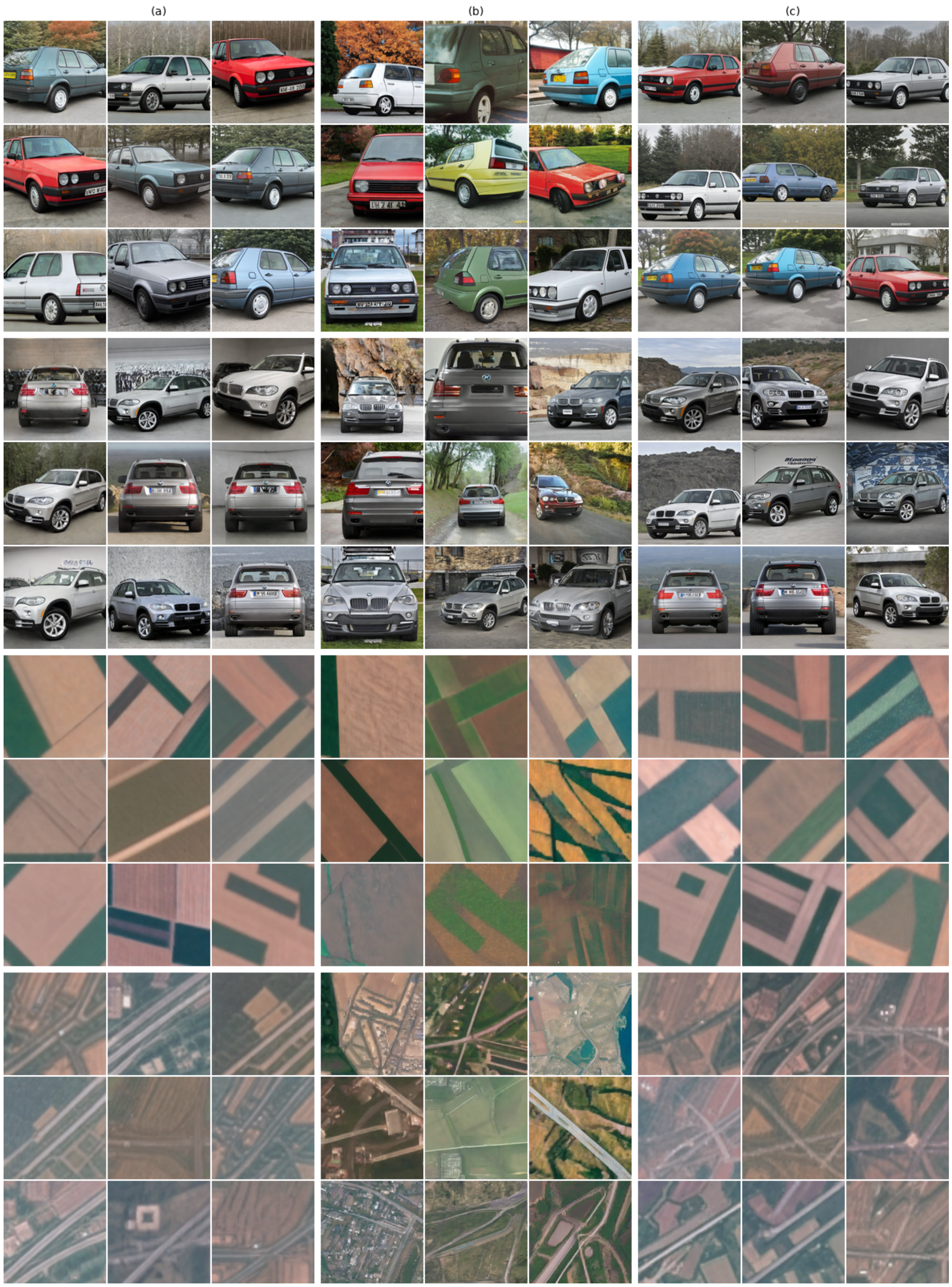

Figure 9. *Additional Qualitative Results.* (a) LoFT, (b) DataDream, (c) ChimeraLoRA. From top to bottom, the classes are 1991 Volkswagen Golf Hatchback and 2007 BMW X5 SUV from Cars, followed by annualcrop and highway from EuroSAT.

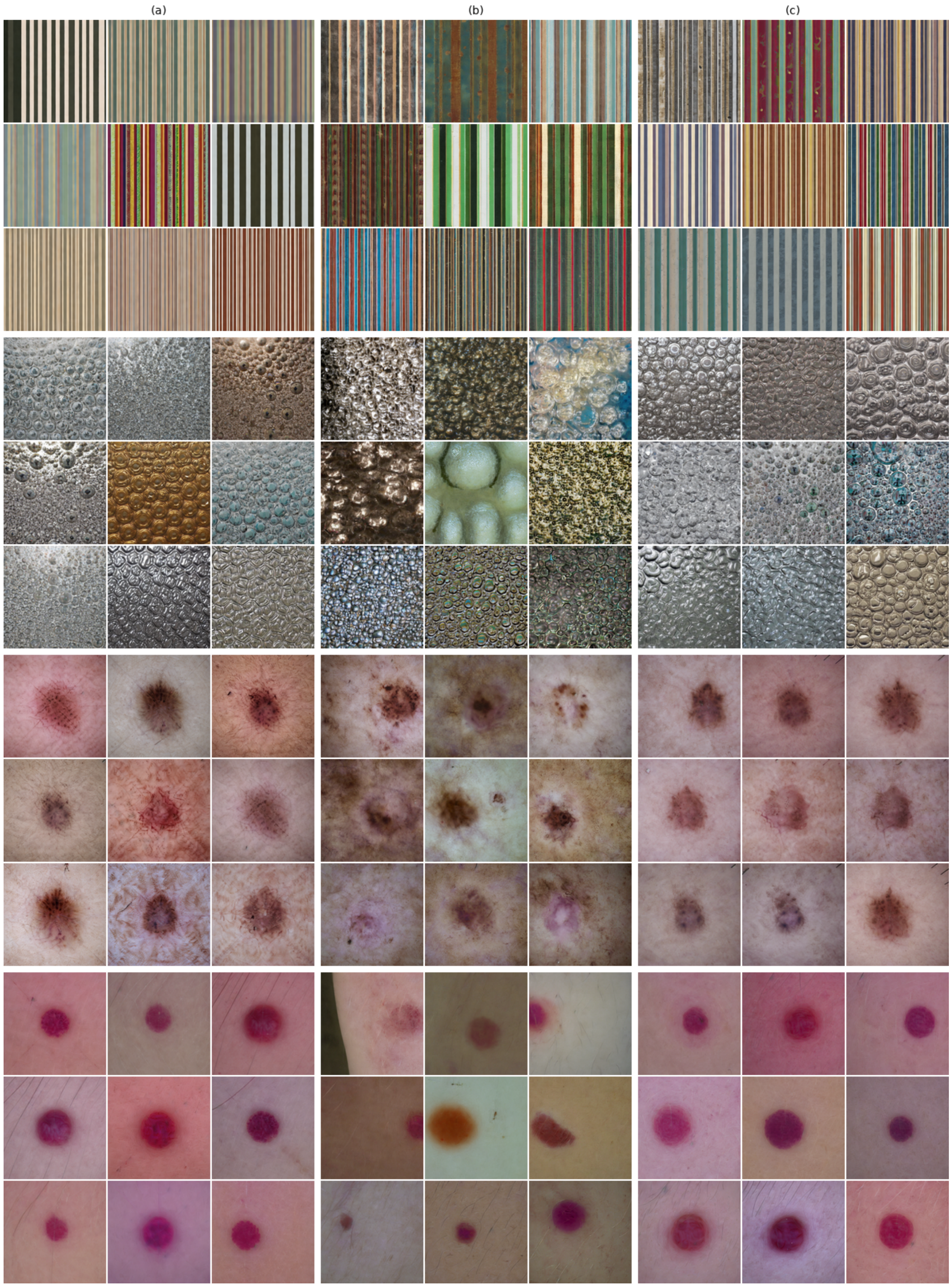

Figure 10. *Additional Qualitative Results.* (a) LoFT, (b) DataDream, (c) ChimeraLoRA. From top to bottom, the classes are banded and bubbly from DTD, followed by basal cell carcinoma and vascular lesion from ISIC.