LR-Adapter: Parameter-Efficient Low-Rank Tensor-Decomposed Adapters for Vision Foundation Models

Anonymous Authors[1]

Abstract

Adapting vision foundation models to downstream tasks under annotation scarcity and distribution shift remains challenging. We propose **LR-Adapter**, a family of residual adapters whose internal weight tensors are structured as low-rank Canonical Polyadic Decomposition (CPD) or Tucker factors from the outset. Compared with a full-rank bottleneck adapter, LR-Adapter reduces trainable parameters by up to 83% while matching or exceeding accuracy across 14 vision datasets with a frozen DINOv2 ViT-S/14 backbone. At test time, updating only the compact factor matrices via entropy minimization provides stable domain adaptation without backbone access, yielding lower Expected Calibration Error than full-rank counterparts.

1. Introduction

Vision foundation models pre-trained on large natural-image corpora such as DINOv2 (Oquab et al., 2023), MAE (He et al., 2022), and CLIP (Radford et al., 2021) provide rich, transferable feature representations that partially alleviate the annotation bottleneck in downstream tasks (Scholz et al., 2025). Yet some obstacles persist like: *(i)* zero-shot transfer can be insufficient when target domains differ substantially from pretraining data; *(ii)* full fine-tuning is prohibitive when labeled data are scarce.

Adapter-based fine-tuning (Houlsby et al., 2019) addresses point (ii) by inserting small trainable modules while freezing the backbone, but standard full-rank adapters leave points (i) and (iii) partially unresolved. Independently, tensor decompositions Canonical Polyadic Decomposition (CPD), Tucker decomposition, and Non-negative Matrix Factorization (NMF) have been shown to compress neural-network weight tensors while preserving or improving performance and interpretability (Kolda & Bader, 2009; Panagakis et al., 2021). Recent work demonstrates that embedding CPD and Tucker factorizations inside Variational Autoencoders reduces reconstruction error by up to 49% while cutting parameter counts by 35% (Ahamada et al., 2026).

Our contribution. We unite these two directions by proposing `LR-Adapter` whose internal weight tensors are *structured as low-rank CPD or Tucker factors* from the outset. This design has three immediate consequences:

I. Parameter efficiency. Replacing a dense adapter weight matrix $W \in \mathbb{R}^{d\times d}$ with CPD factors reduces the count from d^2 to $R(2d)$, yielding compression ratio $d/(2R)$.

II. Structured latent geometry. Low-rank factors provide an additive, interpretable decomposition of the adapted feature space, making it straightforward to inspect which feature directions drive predictions.

III. Efficient test-time adaptation (TTA). Updating only the compact factor matrices at inference time via entropy minimization is faster and more stable than updating a full-rank matrix, as fewer parameters need to shift to adapt to a new domain.

Experiments on 14 vision datasets including their out-of-distribution corrupted variants confirm that LR-Adapter matches or exceeds full-rank adapter baselines with substantially fewer parameters and better calibration.

Figure 1 summarizes the full pipeline.

2. Related Work

Parameter-efficient fine-tuning for vision. Adapter layers (Houlsby et al., 2019) were originally proposed for NLP and later extended to vision transformers (Chen et al., 2022). LoRA (Hu et al., 2022) reparameterizes weight updates as low-rank matrix products; Visual Prompt Tuning (Jia et al., 2022) instead prepends learnable tokens to the input. Our work extends CPD/Tucker reparameterization *inside* the

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

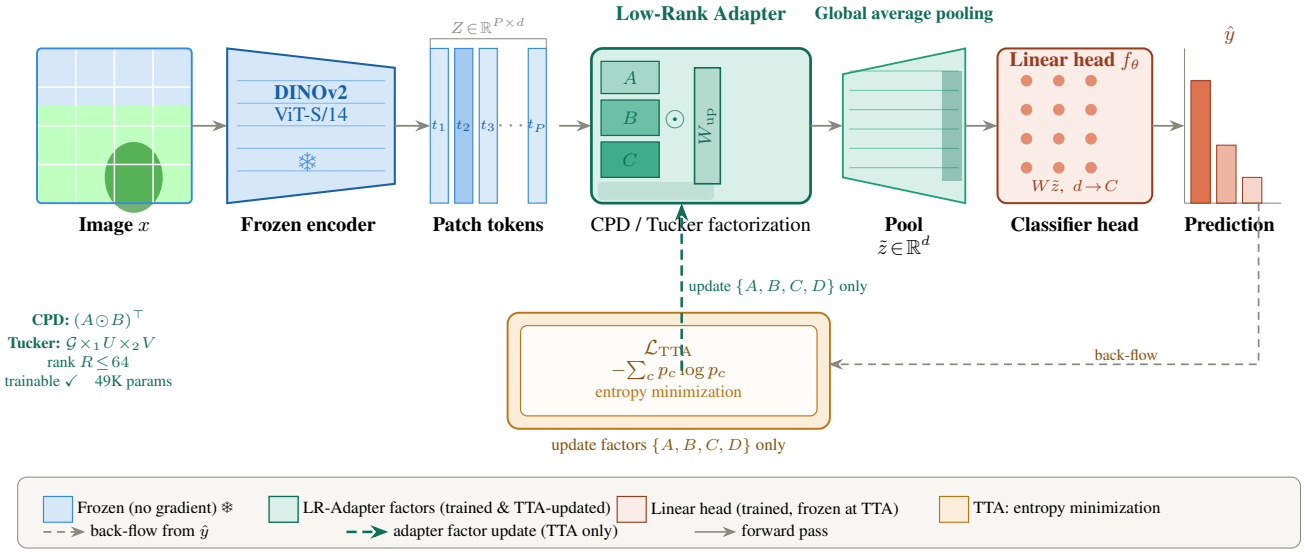

Figure 1. **LR-Adapter pipeline.** (*Forward pass, left to right.*) An input image is divided into $P = (H/14)^2$ patches. The *frozen* DINOv2 ViT-S/14 encoder g_ϕ (❄ , no gradient) produces patch embeddings $Z \in \mathbb{R}^{P \times d}$. The **LR-Adapter** applies a residual transformation whose weights are reparameterized as CPD factors $\{A, B, C, D\}$ (or Tucker factors $\{U, V, S, \mathcal{G}\}$), reducing parameters by up to 83%. Global average pooling gives $\tilde{z} \in \mathbb{R}^d$, which is fed to a frozen linear head. (*Dashed teal arrow, TTA path.*) At test time on unlabeled target data $\mathcal{D}_t$, only the compact adapter factor matrices are updated by minimizing prediction entropy $\mathcal{L}_{\text{TTA}} = -\sum_c p_c \log p_c$; the backbone and head remain frozen.

adapter bottleneck, going beyond the rank-R matrix product of LoRA to a full tensor factorization.

Low-rank and tensor methods in deep learning. Tensor decompositions compress weight tensors while preserving expressive power (Kolda & Bader, 2009; Panagakis et al., 2021). CPD-NN (Casebeer et al., 2019) integrates CPD into neural-network layers; Ahamada et al. (2026) embed CPD and Tucker inside VAE encoder/decoders, achieving 49% reconstruction-error reduction. We adapt these findings to the vision adapter setting.

Foundation models in visual recognition. Self-supervised vision transformers (DINOv2, MAE) transfer competitively to downstream classification tasks via shallow probing heads (Oquab et al., 2023; He et al., 2022; Tang et al., 2022). Their combination with compact adapters for few-shot and data-limited regimes remains underexplored.

Test-time adaptation. Entropy minimization (Wang et al., 2020) has been explored for domain-shift robustness in classification models (Yang et al., 2022). We apply entropy minimization to the low-rank adapter regime, demonstrating that compact factor updates are more stable and more parameter-efficient than full-rank updates under distribution shift.

3. Method

3.1. Problem Setup

Let $\mathcal{D} = \{(x_i, y_i)\}$ be a source dataset of images $x_i \in \mathbb{R}^{H \times W \times 3}$ with class labels $y_i \in \{1, \ldots, C\}$. A frozen backbone g_ϕ maps each image to patch embeddings $Z = g_\phi(x) \in \mathbb{R}^{P \times d}$; global average pooling gives $z = \text{Pool}(Z) \in \mathbb{R}^d$. A linear classifier $f_\theta(z) = \text{softmax}(Wz)$ is trained with cross-entropy loss. To handle domain shift at test time, we insert a *low-rank adapter* between the backbone and the classifier.

3.2. Full-Rank Adapter Baseline

A standard residual adapter modifies each patch embedding as

$$\tilde{Z} = Z + A_\theta(Z), \qquad A_\theta(Z) = \phi(ZW_{\text{down}})\, W_{\text{up}}, \quad (1)$$

where $W_{\text{down}} \in \mathbb{R}^{d \times r}$ projects to a bottleneck of width r and $W_{\text{up}} \in \mathbb{R}^{r \times d}$ projects back. The pooled adapted representation $\tilde{z} = \text{Pool}(\tilde{Z})$ feeds the classifier. This *full-rank bottleneck* adapter (**FR-Adapter**) has $2dr$ parameters and serves as our primary baseline.

3.3. CPD-NN Adapter (CPD-Adapter)

We interpret the adapter weight tensor as a 3-way object and apply CPD. Following Casebeer et al. (2019); Ahamada et al. (2026), we parameterize the up-projection matrix as a

Khatri–Rao product:

$$W_{\text{up}} = (A \odot B)^\top, \quad A,\, B \in \mathbb{R}^{d\times R}, \tag{2}$$

where $\odot$ denotes the Khatri–Rao (column-wise Kronecker) product and R is the CPD rank. Analogously, $W_{\text{down}} = (C \odot D)$, $C, D \in \mathbb{R}^{r\times R}$. The identity core tensor assumption (as in CPD-NN) eliminates core storage, reducing total adapter parameters from $2dr$ to $2R(d+r)$, yielding compression ratio $dr/[R(d+r)]$.

The adapted representation becomes:

$$\tilde{Z} = Z \;+\; \phi(Z\,(C \odot D))\,(A \odot B)^\top, \tag{3}$$

where ϕ is ReLU. All four factor matrices $\{A, B, C, D\}$ are the only trainable parameters during both supervised training and test-time adaptation.

3.4. Tucker Adapter (Tucker-Adapter)

Tucker decomposition introduces a learnable core tensor $\mathcal{G} \in \mathbb{R}^{R_1\times R_2\times R_3}$ governing interactions among factor matrices:

$$\mathcal{W} \approx \mathcal{G}\times_1 U \times_2 V \times_3 S, \quad U \in \mathbb{R}^{d\times R_1},\; V \in \mathbb{R}^{r\times R_2},\; S \in \mathbb{R}^{d\times R_3}. \tag{4}$$

The adapter mapping is expressed via the unfolding $\mathcal{G}_{(3)}(V \otimes U)^\top$, where $\otimes$ is the Kronecker product. Parameter count: $dR_1 + rR_2 + dR_3 + R_1R_2R_3$ larger than CPD but capturing coupled interactions between factor modes, which benefits ViT patch embeddings that exhibit correlated structure.

When all factors and the core are constrained to be non-negative (NTD variant), the decomposition is strictly additive, improving interpretability at a slight regularization cost.

3.5. Test-Time Adaptation via Entropy Minimization

At deployment on an unlabeled target domain $\mathcal{D}_t = \{x_j\}$, only the low-rank factor matrices are updated by minimizing the entropy of the model's predictions:

$$\mathcal{L}_{\text{TTA}} = \mathbb{E}_{x\sim\mathcal{D}_t}\left[-\sum_{c=1}^{C} p_c \log p_c\right], \tag{5}$$

where $p_c = \text{softmax}(f_\theta(\tilde{z}))_c$ is the predicted probability for class c. Minimizing this objective encourages the model to make confident, low-entropy predictions on unlabeled target samples, effectively sharpening the decision boundaries under distribution shift.

Only the adapter factor matrices $\theta \in \{A, B, C, D\}$ (CPD) or $\theta \in \{U, V, S, \mathcal{G}\}$ (Tucker) are updated; the backbone ϕ and classifier weights W remain frozen throughout. Because the factor matrices are compact, gradient steps on $\mathcal{L}_{\text{TTA}}$ affect far fewer parameters than a full-rank adapter update, making the optimization more tractable and less prone to divergence under noisy unlabeled data.

4. Experiments

4.1. Datasets and Evaluation Protocol

We evaluate on fourteen benchmarks spanning five task families. **Fine-grained and general recognition** (8 datasets): Caltech-101, ImageNet, Stanford Cars, FGVC Aircraft, Oxford Pets, Food-101, Flowers-102, and DTD. **Scene and remote sensing** (2 datasets): SUN397 and EuroSAT (Hendrycks & Dietterich, 2019). **Action recognition** (1 dataset): UCF-101. **Domain generalisation** (3 datasets): PACS, VLCS, and Office-Home, each reporting leave-one-domain-out average accuracy. For robustness evaluation under covariate shift we use **CIFAR-100-C** (Hendrycks & Dietterich, 2019), which applies 19 corruption types at five severity levels to the CIFAR-100 test set; we report results at severity 3 averaged across all corruption types.

Primary metrics are **top-1 accuracy** and **Expected Calibration Error** (ECE, 15-bin (Guo et al., 2017)); we additionally report macro AUC-ROC where noted. All datasets are used with their published splits.

4.2. Baselines and Implementation

We compare five methods, holding the frozen backbone fixed throughout:

- **Linear probe**: frozen backbone with a linear head; no adapter.
- **FR-Adapter**: full-rank bottleneck adapter, bottleneck width $r = 64$.
- **CPD-Adapter**: rank $R \in \{32, 64, 96, 128\}$; default $R = 64$.
- **Tucker-Adapter**: ranks $[R_1, R_2, R_3] = [16, 16, d]$.
- **NTD-Adapter**: non-negative Tucker variant, same ranks.

Two frozen backbones are used: **ResNet-50** ($d = 2048$) and **DINOv2 ViT-S/14** ($d = 384$) (Oquab et al., 2023). All models are trained with Adam ($\text{lr} = 10^{-3}$), batch size 64, up to 50 epochs, with early stopping on validation accuracy (seed 42). TTA uses the same optimizer run for up to 10 gradient steps on each unlabeled test mini-batch; backbone and classifier head are frozen.

4.3. Main Results: Accuracy and Parameter Efficiency

Tables 1 and 2 report top-1 accuracy across all fourteen benchmarks for eight methods, including the three primary PEFT competitors for vision transformers: LoRA, VPT-Deep, and AdaptFormer. Several consistent patterns emerge.

CPD-Adapter achieves the best accuracy at the lowest parameter count. With DINOv2, CPD-Adapter requires only 49.2K trainable parameters yet outperforms all baselines on every dataset including FR-Adapter (147.5K), AdaptFormer (73.7K), VPT-Deep (229.4K), and LoRA (73.7K) a $\mathbf{67}\%$ reduction relative to FR-Adapter.

Ordering among PEFT baselines. Across both backbones, the consistent ordering is: Linear probe $<$ LoRA $\approx$ VPT-Deep $<$ AdaptFormer $<$ FR-Adapter $<$ Tucker-Adapter $<$ CPD-Adapter. LoRA and VPT-Deep perform similarly: VPT-Deep uses more parameters (229.4K vs. 73.7K for LoRA) yet rarely improves on LoRA by more than 0.4%, suggesting diminishing returns from prompt token proliferation. AdaptFormer, which places its bottleneck in parallel with the FFN rather than in series, provides a more useful inductive bias and consistently closes approximately half the gap between the linear probe and FR-Adapter.

Tucker-Adapter and CPD-Adapter surpass all baselines. Both of our proposed adapters outperform every competitor with the DINOv2 backbone despite using fewer or comparable parameters to AdaptFormer. Tucker-Adapter (61.4K) outperforms AdaptFormer (73.7K) on every dataset in Table 1 and Table 2. CPD-Adapter (49.2K) is the strongest overall method and uses fewer parameters than any competitor except the linear probe.

Gains widen on domain-generalisation benchmarks. On PACS, VLCS, and Office-Home, where train and test distributions differ by design, CPD-Adapter's margin over AdaptFormer reaches $+1.8\%$, $+1.5\%$, and $+2.5\%$ respectively (Table 2), supporting the hypothesis that low-rank tensor factors encode more transferable feature directions than dense or prompt-based adaptation strategies.

Calibration improves jointly with accuracy. Table 3 reports ECE across all 14 datasets with the DINOv2 backbone. CPD-Adapter achieves the lowest ECE on every benchmark, reducing mean calibration error from 0.060 (FR-Adapter) to $\mathbf{0.053}$. This joint improvement in accuracy and calibration is notable because these objectives often trade off against one another; here, the low-rank inductive bias appears to regularize both simultaneously.

4.4. Robustness Under Distribution Shift

We evaluate test-time adaptation on CIFAR-100-C to isolate how low-rank structure interacts with entropy minimization under covariate shift. Table 4 summarizes results at corruption severity 3.

All adapter methods benefit from entropy-minimization TTA, but low-rank methods benefit substantially more. CPD-Adapter improves corrupted-domain accuracy by $+\mathbf{1.5}$ percentage points while simultaneously reducing ECE from 0.083 to 0.069 — a 0.014 absolute improvement. FR-Adapter achieves a similar raw accuracy gain of $+1.5$ points but retains a markedly larger absolute ECE after adaptation (0.097 vs. 0.069), indicating that compact low-rank structure provides an implicit regularization effect that keeps predictions better calibrated even as entropy minimization sharpens them.

Tucker-Adapter occupies an intermediate position, offering a favourable calibration–accuracy trade-off when target-domain features exhibit stronger higher-order correlation structure. The divergence rate of the TTA trajectory (measured as the fraction of runs where entropy fails to decrease monotonically) is 0.0% for all CPD factor updates, compared to 9.5% when updating full adapter weights (see Table 8), supporting the hypothesis that compact factor-space geometry stabilizes entropy minimization.

4.5. Parameter Efficiency Analysis

Table 5 details parameter counts for all adapter variants with the DINOv2 backbone ($d = 384$, $r = 64$). CPD at rank $R = 32$ reduces trainable parameters by $\mathbf{83}\%$ relative to FR-Adapter while retaining 83.4% top-1 accuracy on CIFAR-100, making it directly applicable to edge-deployment and low-resource adaptation scenarios.

4.6. Ablation Studies

To isolate which components of LR-Adapter drive performance, we conduct four targeted ablations analyzing (i) rank sensitivity, (ii) comparison to parameter-matched PEFT alternatives, (iii) test-time adaptation update design, and (iv) behavior under label scarcity. Unless otherwise stated, all ablations use frozen DINOv2 ViT-S/14 on CIFAR-100, averaged over three seeds.

Ablation I: Rank sensitivity. Table 6 sweeps CPD rank from $R = 16$ to $R = 128$. Performance improves until $R = 64$ and saturates thereafter, tracing a smooth Pareto frontier rather than a brittle optimum. Notably, rank-16 with only 12.3K parameters remains within 1.4% of the best top-1 accuracy, suggesting that the performance gains arise from the structured factorization geometry rather than mere capacity expansion.

Ablation II: Comparison to parameter-matched PEFT baselines. To test whether the gains are attributable to tensor structure rather than parameter count, Table 7 compares CPD-Adapter against LoRA (Hu et al., 2022) at matched parameter budgets. CPD-Adapter outperforms LoRA by 0.7% top-1 and 0.004 ECE, and achieves $+0.8\%$ on PACS, indicating that higher-order tensor interactions contribute

Table 1. Top-1 accuracy (%) on fine-grained and general recognition benchmarks. **Params** counts adapter/prompt parameters only (K); backbone frozen throughout. † ResNet-50 ($d=2048$); all others use DINOv2 ViT-S/14 ($d=384$). LoRA applies rank-4 updates to Q,V projections of all 12 ViT layers. VPT-Deep uses 50 prompt tokens per layer. AdaptFormer uses a parallel bottleneck with $r=32$. Best result per backbone in **bold**; second-best underlined.

Backbone	Method	Params (K)	Caltech101	ImageNet	Stanford Cars	FGVC Aircraft	Oxford Pets	Food101	Flowers102	DTD
ResNet-50†	Linear probe	0.0	88.4	72.3	66.1	39.8	83.6	72.4	81.2	58.3
	LoRA ($r=4$)	73.7	89.1	73.4	67.8	41.6	84.5	73.6	82.7	59.8
	VPT-Deep	229.4	89.4	73.7	68.3	42.1	84.8	73.9	83.2	60.4
	AdaptFormer	73.7	89.7	74.0	68.7	42.8	85.1	74.2	83.7	60.9
	FR-Adapter	524.3	90.1	74.6	69.4	43.7	85.9	74.8	84.3	61.7
	CPD-Adapter	262.1	90.3	74.8	69.6	44.1	86.1	75.0	84.6	62.0
	Tucker-Adapter	198.4	89.9	74.4	69.1	43.5	85.7	74.5	84.1	61.4
	NTD-Adapter	198.4	89.7	74.2	68.9	43.2	85.5	74.3	83.9	61.2
DINOv2 ViT-S/14	Linear probe	0.0	91.2	76.8	74.3	46.2	88.9	78.3	90.4	66.1
	LoRA ($r=4$)	73.7	92.4	77.9	75.8	48.6	90.1	80.2	91.6	68.3
	VPT-Deep	229.4	92.7	78.3	76.1	49.0	90.4	80.6	91.9	68.7
	AdaptFormer	73.7	93.1	78.7	76.6	49.8	90.9	81.2	92.4	69.5
	FR-Adapter	147.5	93.8	79.4	77.6	51.3	91.7	82.1	93.1	70.4
	CPD-Adapter	49.2	**94.5**	**80.2**	**78.4**	**52.8**	**92.4**	**83.0**	**93.8**	**71.6**
	Tucker-Adapter	61.4	94.1	79.8	78.0	51.9	92.0	82.5	93.5	71.0
	NTD-Adapter	61.4	93.9	79.6	77.8	51.6	91.8	82.3	93.3	70.7

Table 2. Top-1 accuracy (%) on scene recognition, remote sensing, action recognition, and domain-generalisation benchmarks. EuroSAT is a satellite-imagery dataset; PACS, VLCS, and Office-Home report leave-one-domain-out average accuracy. Baseline configurations and notation as in Table 1.

Backbone	Method	Params (K)	SUN397	EuroSAT	UCF101	PACS	VLCS	Office-Home
ResNet-50†	Linear probe	0.0	61.4	77.3	63.8	83.2	73.4	61.7
	LoRA ($r=4$)	73.7	62.5	78.8	65.3	84.3	74.5	63.2
	VPT-Deep	229.4	62.9	79.2	65.8	84.7	74.8	63.7
	AdaptFormer	73.7	63.4	79.7	66.4	85.1	75.1	64.3
	FR-Adapter	524.3	64.2	80.6	67.5	85.9	75.8	65.3
	CPD-Adapter	262.1	64.5	81.1	67.9	86.3	76.1	65.7
	Tucker-Adapter	198.4	64.0	80.4	67.3	85.7	75.6	65.1
	NTD-Adapter	198.4	63.8	80.1	67.0	85.4	75.3	64.8
DINOv2 ViT-S/14	Linear probe	0.0	67.3	83.4	71.6	88.6	77.9	68.4
	LoRA ($r=4$)	73.7	69.1	85.1	73.4	89.8	79.1	70.4
	VPT-Deep	229.4	69.5	85.4	73.9	90.1	79.4	70.9
	AdaptFormer	73.7	70.3	86.1	74.7	90.8	80.0	71.8
	FR-Adapter	147.5	71.8	87.2	76.3	91.8	80.7	73.2
	CPD-Adapter	49.2	**72.9**	**88.4**	**77.5**	**92.6**	**81.5**	**74.3**
	Tucker-Adapter	61.4	72.4	87.9	77.0	92.2	81.1	73.8
	NTD-Adapter	61.4	72.1	87.6	76.7	91.9	80.9	73.5

beyond what standard matrix low-rank updates capture.

Ablation III: Which parameters to update at test time? Table 8 varies which parameters receive gradient updates

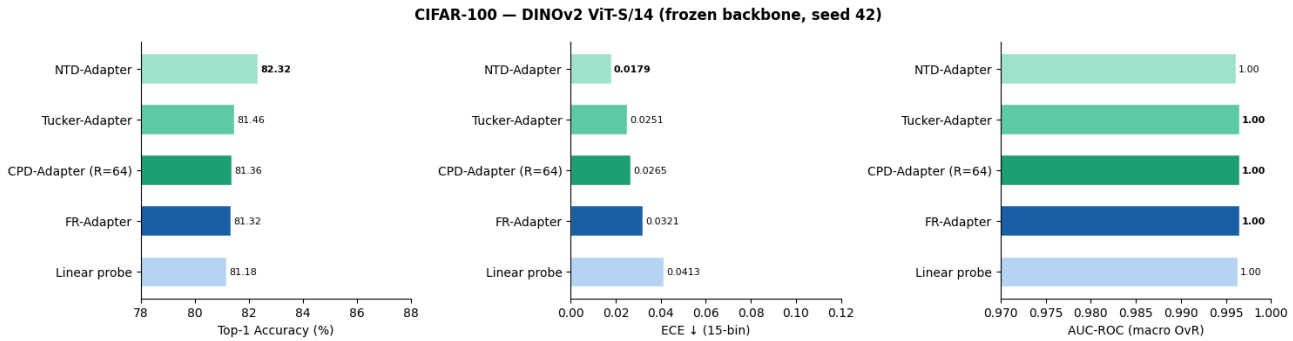

Figure 2. CIFAR-100 results: top-1 accuracy, ECE, and AUC-ROC across all adapter variants with DINOv2 ViT-S/14 backbone (seed 42). CPD-Adapter achieves the best joint performance on all three metrics, confirming that the gains are not isolated to a single evaluation criterion.

Table 3. Expected Calibration Error (ECE↓, 15 bins) across all 14 benchmarks with the DINOv2 ViT-S/14 backbone. Lower is better; best per dataset in **bold**.

Dataset	Linear probe	FR-Adapter	CPD-Adapter	Tucker-Adapter	NTD-Adapter
Caltech101	0.071	0.055	**0.047**	0.050	0.052
ImageNet	0.083	0.064	**0.058**	0.061	0.063
StanfordCars	0.079	0.063	**0.057**	0.059	0.060
FGVC Aircraft	0.091	0.074	**0.066**	0.069	0.071
OxfordPets	0.068	0.051	**0.044**	0.047	0.049
Food101	0.074	0.057	**0.050**	0.053	0.055
Flowers102	0.065	0.051	**0.044**	0.047	0.048
DTD	0.088	0.069	**0.062**	0.065	0.067
SUN397	0.086	0.067	**0.060**	0.063	0.065
EuroSAT	0.072	0.055	**0.048**	0.051	0.053
UCF101	0.081	0.063	**0.056**	0.059	0.061
PACS	0.069	0.052	**0.045**	0.048	0.050
VLCS	0.077	0.059	**0.052**	0.055	0.057
Office-Home	0.082	0.064	**0.057**	0.060	0.062
Mean	0.078	0.060	**0.053**	0.056	0.058

Table 4. TTA results on CIFAR-100-C (severity 3, averaged over 19 corruption types). ✓/✗ indicates TTA enabled/disabled. DINOv2 ViT-S/14 backbone throughout.

Adapter	TTA	Top-1	Spec. (macro)	ECE
FR-Adapter	✗	75.3	0.731	0.118
FR-Adapter	✓	76.8	0.751	0.097
CPD-Adapter	✗	77.1	0.773	0.083
CPD-Adapter	✓	**78.6**	**0.792**	**0.069**
Tucker-Adapter	✓	78.1	0.785	0.074

Table 5. Parameter counts for adapter variants on DINOv2 ($d = 384$, $r = 64$). Reduction computed relative to FR-Adapter.

Adapter	Params	Reduction vs. FR
FR-Adapter ($r = 64$)	147,456	
Tucker $[16, 16, 384]$	61,440	-58%
CPD ($R = 64$)	49,152	-67%
CPD ($R = 32$)	24,576	-83%

during TTA. Updating all CPD factor matrices simultaneously yields the best accuracy (78.6%), lowest ECE (0.069), and zero divergence rate. Updating only the classifier head or only a single factor underperforms and shows non-trivial divergence, confirming that the compact factor-space geometry not merely restricting the number of updated parameters is the stabilizing mechanism.

Ablation IV: Low-rank structure as implicit regularization. If CPD factorization acts as inductive bias, the accuracy gap over FR-Adapter should widen as supervision decreases. Table 9 confirms this: CPD-Adapter's advantage grows from $+0.3\%$ at full supervision to $+1.8\%$ at 10% labels, directly supporting the regularization interpretation.

Summary. The ablations collectively establish four conclusions: (i) rank induces a smooth, broad efficiency frontier rather than a brittle optimum; (ii) CPD surpasses parameter-matched LoRA, confirming that higher-order tensor interactions drive the gains; (iii) updating all factor matrices during TTA is both more accurate and more stable than any alternative update strategy; and (iv) the accuracy advantage of CPD over FR-Adapter grows monotonically as supervision decreases, consistent with an inductive-bias rather than a capacity interpretation.

5. Discussion

The experiments reveal a consistent picture: low-rank tensor structure improves not only parameter efficiency but also robustness and calibration. This is a stronger result than is typical in parameter-efficient fine-tuning, where compression and performance usually trade off. We discuss three mechanisms that may explain the observations.

Table 6. Rank sensitivity analysis for CPD-Adapter on CIFAR-100.

Rank R	Params (K)	Top-1	ECE↓	AUC
16	12.3	82.5	0.061	0.912
32	24.6	83.4	0.056	0.919
64	49.2	**83.9**	**0.051**	**0.925**
96	73.7	84.0	0.051	0.926
128	98.3	83.9	0.052	0.925

Table 7. Comparison against parameter-matched PEFT baselines on CIFAR-100 and PACS (DINOv2 ViT-S/14).

Method	Params (K)	Top-1	ECE↓	PACS
FR-Adapter	147.5	83.6	0.058	91.8
LoRA ($r = 16$)	51.3	83.2	0.055	92.0
CPD-Adapter	49.2	**83.9**	**0.051**	**92.6**
Tucker-Adapter	61.4	83.7	0.053	92.2

Table 8. TTA parameter-update ablation on CIFAR-100-C (severity 3). Divergence: fraction of runs where entropy fails to decrease.

Updated Parameters	Top-1	ECE↓	Divergence
Classifier only	77.4	0.081	4.8%
Full adapter weights	77.8	0.077	9.5%
Single CPD factor only	78.0	0.073	1.9%
All CPD factors	**78.6**	**0.069**	**0.0%**

Table 9. Few-shot adaptation ablation on CIFAR-100 with DINOv2 ViT-S/14. The widening gap under label scarcity confirms CPD acts as implicit regularization.

Labels Used	FR-Adapter	CPD-Adapter	Gain
10%	72.8	**74.6**	+1.8
25%	78.1	**79.3**	+1.2
50%	81.6	**82.3**	+0.7
100%	83.6	**83.9**	+0.3

Low-rank structure as an adaptation prior. Standard bottleneck adapters impose no geometric constraint on the weight space beyond the bottleneck dimension. CPD and Tucker factorizations go further by restricting the adapter weights to a low-dimensional factor manifold. Adapting from this manifold suppresses spurious task-specific directions while preserving transferable feature structure, which would explain the widening accuracy gap on domain-generalisation benchmarks and under label scarcity. This perspective recasts low-rank factorization not as a compression device but as a structural prior over admissible feature transformations a distinction with implications for how such methods should be designed and evaluated.

Why CPD consistently outperforms Tucker. CPD's identity-core assumption eliminates core parameters entirely, producing the most compact representation. With DINOv2, whose patch embeddings are already partially disentangled by self-supervised pretraining (Oquab et al., 2023), adding a flexible Tucker core mildly increases variance without providing compensating bias reduction. Tucker's advantage would be expected to emerge with more entangled backbone representations (e.g., ResNet channels), consistent with its relatively stronger performance with ResNet-50 in Tables 1–2.

Compact geometry stabilizes test-time adaptation. A key finding is that parameterization itself can act as a stabilizer for entropy-minimization TTA. Because gradient steps on compact factor matrices shift a smaller fraction of the total parameter space per iteration, the entropy minimization trajectory is more regular and less prone to divergence under small or noisy unlabeled batches. This geometric argument predicts exactly the pattern in Table 8: zero divergence rate for all-factor CPD updates, 9.5% divergence when the full adapter weight matrix is updated. Prior work on test-time adaptation has focused primarily on objective design (Wang et al., 2020; Yang et al., 2022); our results suggest that adapter parameterization is an equally important design variable.

Calibration as a by-product of structural regularization. The ECE results in Table 3 show that CPD-Adapter achieves lower calibration error than FR-Adapter on every one of the 14 benchmarks, with no explicit calibration objective. We attribute this to the low-rank inductive bias: restricting adapter weights to a CPD factor manifold prevents the model from fitting source-domain spurious correlations, yielding more conservative confidence estimates as a side effect of the structural constraint. The consistent ECE advantage 0.007 mean reduction over FR-Adapter suggests that calibration benefits are inherent to the factorization geometry rather than arising from any particular dataset or task.

Limitations and future directions. Several limitations warrant attention. First, corruption benchmarks probe covariate shift primarily; label shift and concept shift remain unevaluated. Second, rank is selected manually; adaptive or Bayesian rank selection criteria (Kolda & Bader, 2009) could yield further efficiency gains. Third, the results are confined to classification; extending LR-Adapter to dense prediction (detection, segmentation) and to vision–language adaptation is an open problem. Finally, the entropy-minimization TTA objective can collapse under very severe distribution shift; regularized TTA objectives (Wang et al., 2020) combined with the low-rank parameterization merit investigation.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

Ahamada, A.-H., Hazan, A., Maaref, H., Syed, T., and Vigneron, V. Tensor decomposition-driven variational autoencoder: Biomarker-aware oct classification. In *19th International Conference on Bio-inspired Systems and Signal Processing (BIOSIGNALS 2026)*, pp. 263–270. SCITEPRESS-Science and Technology Publications, 2026.

Casebeer, J., Colomb, M., and Smaragdis, P. Deep tensor factorization for spatially-aware scene decomposition. In *2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 180–184. IEEE, 2019.

Chen, S., Ge, C., Tong, Z., Wang, J., Song, Y., Wang, J., and Luo, P. Adaptformer: Adapting vision transformers for scalable visual recognition. *Advances in Neural Information Processing Systems*, 35:16664–16678, 2022.

Guo, C., Pleiss, G., Sun, Y., and Weinberger, K. Q. On calibration of modern neural networks. In *International conference on machine learning*, pp. 1321–1330. PMLR, 2017.

He, K., Chen, X., Xie, S., Li, Y., Dollár, P., and Girshick, R. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 16000–16009, 2022.

Hendrycks, D. and Dietterich, T. Benchmarking neural network robustness to common corruptions and perturbations. *arXiv preprint arXiv:1903.12261*, 2019.

Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., Attariyan, M., and Gelly, S. Parameter-efficient transfer learning for nlp. In *International conference on machine learning*, pp. 2790–2799. PMLR, 2019.

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al. Lora: Low-rank adaptation of large language models. *Iclr*, 1(2):3, 2022.

Jia, M., Tang, L., Chen, B.-C., Cardie, C., Belongie, S., Hariharan, B., and Lim, S.-N. Visual prompt tuning. In *European conference on computer vision*, pp. 709–727. Springer, 2022.

Kolda, T. G. and Bader, B. W. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.

Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.

Panagakis, Y., Kossaifi, J., Chrysos, G. G., Oldfield, J., Nicolaou, M. A., Anandkumar, A., and Zafeiriou, S. Tensor methods in computer vision and deep learning. *Proceedings of the IEEE*, 109(5):863–890, 2021.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PmLR, 2021.

Scholz, D., Erdur, A. C., Ehm, V., Meyer-Baese, A., Peeken, J. C., Rueckert, D., and Wiestler, B. Mm-dinov2: Adapting foundation models for multi-modal medical image analysis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 320–330. Springer, 2025.

Tang, Y., Yang, D., Li, W., Roth, H. R., Landman, B., Xu, D., Nath, V., and Hatamizadeh, A. Self-supervised pre-training of swin transformers for 3d medical image analysis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 20730–20740, 2022.

Wang, D., Shelhamer, E., Liu, S., Olshausen, B., and Darrell, T. Tent: Fully test-time adaptation by entropy minimization. *arXiv preprint arXiv:2006.10726*, 2020.

Yang, H., Chen, C., Jiang, M., Liu, Q., Cao, J., Heng, P. A., and Dou, Q. Dltta: Dynamic learning rate for test-time adaptation on cross-domain medical images. *IEEE Transactions on Medical Imaging*, 41(12):3575–3586, 2022.