CENWRITE: SUPPORT-CENTERED LORA FOR CROSS-DOMAIN FEW-SHOT LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Cross-domain few-shot learning adapts a pretrained vision model to a new domain with scarce labels. Such adaptation must use the same small support set to accommodate domain differences and distinguish target classes. It must also produce reliable predictive probabilities despite having limited labeled evidence about the target domain. Low-rank adaptation (LoRA) offers a parameter-efficient approach, but uncentered support activations make the support-set mean response and sample-specific deviations jointly shape its low-rank factors. Direct centering changes this learning path but also removes the shared response from the forward pass. We propose CenWrite to center factor learning while preserving the shared response and enabling boundary adjustment. CenWrite learns low-rank factors from centered support activations, restores the shared response through stop-gradient compensation, and learns a separate boundary residual. Across four domains, CenWrite has lower average probability errors than six adapters on DINOv3 and higher average accuracy than these adapters on DINOv2. Component controls support probability gains from centering and compensation; the residual improves mean macro-F1 in the full method, with a probability cost.

1 INTRODUCTION

Cross-domain few-shot learning adapts a pretrained visual representation to target classes using only a handful of labeled images. The same support set must reveal how the target domain differs from pretraining and which features distinguish its classes (Tseng et al., 2020; Guo et al., 2020). An update must accommodate both demands with little independent evidence against which to check it. The resulting predictions should also carry useful probabilities: two models can predict the same class while assigning quite different probabilities to that class, a distinction that accuracy does not capture (Guo et al., 2017; Chidambaram & Ge, 2025). Low-rank adaptation (LoRA) limits the number of trained parameters (Hu et al., 2022), making it attractive when support labels are scarce. Yet limiting update capacity does not determine how that support evidence shapes the factors, or how the resulting update affects classification and probability quality.

To examine this learning signal, consider an activation h, its support-set mean μ, and a low-rank weight update ΔW. The adapter response is

$$\Delta W h = \underbrace{\Delta W(h-\mu)}_{\text{sample-dependent response}} + \underbrace{\Delta W \mu}_{\text{shared mean response}} . \tag{1}$$

The first term varies across inputs; the second is shared at that layer and can affect later predictions through subsequent transformations. Ordinary LoRA sends gradients from both terms to the same factors, coupling the learning of sample differences with changes in the shared response. Directly replacing h by $h-\mu$ removes the mean term from factor learning, but also removes it from the forward pass. That response is not known to be expendable: a mean estimated from a few images can reflect both useful common structure and sampling variation, while deviations need not contain only class information. Figure 1 illustrates the resulting design question: how can centered factor learning preserve the shared forward response while allowing an independent boundary update?

We propose **CenWrite**: learn low-rank factors from $h-\mu$, restore $\Delta W\mu$ without its direct factor gradient, and train a separate boundary residual. The method preserves the shared forward response while changing the factor-learning signal. Complete comparisons show lower mean probability

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

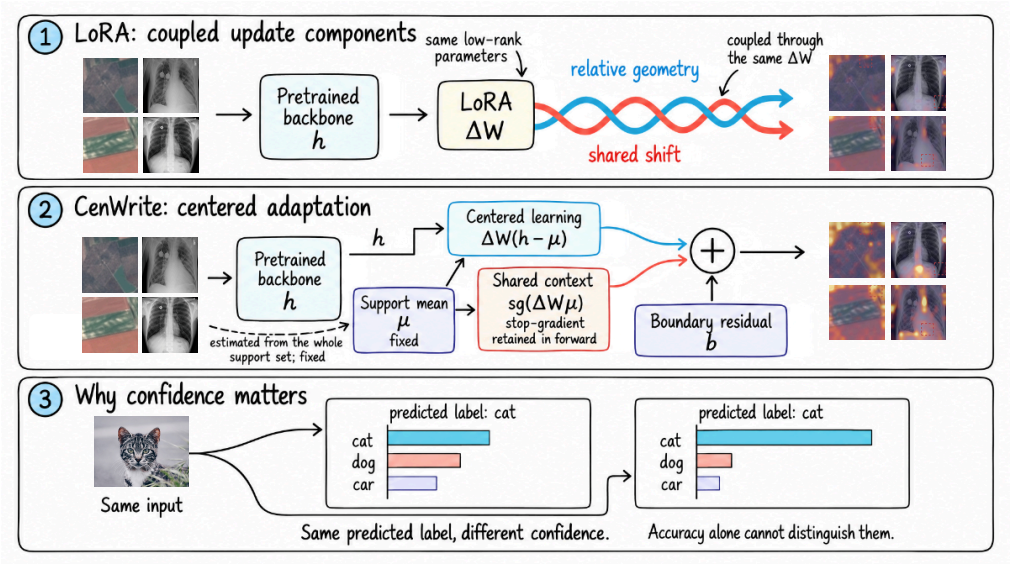


Figure 1: Motivation and design of CenWrite. (1) LoRA couples sample-dependent variation ("relative geometry") and the shared mean response ("shared shift"). (2) The full-support mean is fixed during adaptation. CenWrite combines centered factor learning, stop-gradient compensation ("shared context"), and a separate boundary residual; the frozen projection is omitted. (3) Identical class predictions can carry different probabilities, motivating evaluation beyond accuracy. Heatmaps and probability bars are schematic, not measured model outputs.

errors than six adapters on DINOv3 and higher mean accuracy on DINOv2. Matched component controls connect the probability gains to centering and compensation, and a macro-F1 benefit to the residual, with measured tradeoffs. Two material-image evaluations further test the centered learning route on microstructure and morphology recognition. These results concern the stated support-adaptation protocols; they do not establish a separation of domain and class information or superiority for every metric and dataset.

2 Related Work

Few-shot visual adaptation. Cross-domain few-shot learning tests whether features learned from source data can support new classes under a target-domain shift (Tseng et al., 2020; Guo et al., 2020). SimpleShot centers features for nearest-neighbor classification (Wang et al., 2019); CenWrite instead uses a support-set activation mean inside backbone adaptation. Neural fine-tuning search selects adapter placement and trainable layers (Eustratiadis et al., 2024), while visual prompting adapts the input or token space (Han et al., 2024; Jin et al., 2025). These choices alter where the small support set can change a pretrained representation. MPS-Tuning preserves batch- and token-level Gram matrices and sculpts image–text class similarities (Chen et al., 2026). It addresses the tension between retaining useful relations and learning target distinctions through explicit manifold losses and a text encoder. Flow Matching Alignment learns cross-modal transport (Jiang et al., 2026). CenWrite instead changes the factor gradients of an image-only adapter while keeping its shared mean response in the forward pass. These approaches use different information and objectives; their reported results do not constitute matched comparisons with CenWrite. Appendix K details these distinctions.

Low-rank capacity and learning dynamics. LoRA factorizes weight increments (Hu et al., 2022), but a low parameter count alone does not specify the update's capacity or learning trajectory (Zeng & Lee, 2024). FacT shares tensor factors (Jie & Deng, 2023), BOFT learns structured orthogonal transformations (Liu et al., 2024), and FourierFT represents updates with sparse frequencies (Gao et al., 2024). RandLoRA combines random bases (Albert et al., 2025), while ReFTA

avoids full tensor reconstruction (Zheng et al., 2026). These methods change the representation of the update. A separate line changes how factorized updates are learned: LoRA-Pro adjusts factor gradients (Wang et al., 2025), LoRA-RITE uses invariant preconditioning (Yen et al., 2025), Stable-LoRA shrinks the nonzero initial factor (Wu et al., 2026), and LoFT projects optimizer moments (Tastan et al., 2026). CenWrite retains LoRA factors but removes the shared mean response's direct factor gradient, restores that response in the forward pass, and adds a trainable boundary residual. Its effect is therefore distinct from reducing rank or changing the factor optimizer. Whether the intervention complements these optimizers remains untested.

Probability quality under adaptation. Accuracy does not establish whether predicted probabilities are reliable (Guo et al., 2017). Expected calibration error (ECE) summarizes confidence–frequency agreement under a specified binning scheme, whereas negative log-likelihood (NLL) and the Brier score evaluate the full predicted distribution. Their behavior can differ, so a single calibration number does not settle a comparison (Chidambaram & Ge, 2025; Tao et al., 2024). C-TPT uses text-feature dispersion to improve calibration (Yoon et al., 2024); A-TPT studies angular diversity in unlabeled test-time prompt tuning (Ahamed et al., 2026). Both use vision–language information and adaptation conditions that differ from CenWrite's labeled-support, image-encoder setting. CenWrite uses support cross-entropy rather than a calibration-specific loss or post-hoc temperature fit. We evaluate accuracy and macro-F1 alongside ECE, NLL, and Brier to expose tradeoffs between class decisions and probability estimates.

3 METHOD

3.1 TASK AND SUPPORT-SET CENTERS

An episode provides a support set $S = \{(x_i, y_i)\}_{i=1}^{CK}$ with C classes and K labeled images per class, and a disjoint query set Q. We adapt a pretrained image encoder and a linear classifier using support cross-entropy; queries enter only after adaptation. The classifier, adapter, and optimizer reset for every episode. Before optimization, we estimate a center at each adapted projection using the full support set, then hold it fixed while learning the factors and residual. Figure 2 connects this reference to the forward and backward paths. Support labels affect the loss and classifier, whereas center estimation uses all support examples irrespective of class.

For a projection with frozen weight $W_0 \in \mathbb{R}^{d_{\text{out}} \times d_{\text{in}}}$, we suppress its frozen bias and write the LoRA increment as

$$\Delta W = sBA, \qquad s = \alpha/r, \quad A \in \mathbb{R}^{r \times d_{\text{in}}}, \quad B \in \mathbb{R}^{d_{\text{out}} \times r}. \tag{2}$$

The layer center averages projection inputs over all support images, sampled augmented views, and token positions,

$$\mu = \frac{1}{M} \sum_{i=1}^{CK} \sum_{v \in \mathcal{V}} \sum_{p \in \mathcal{P}(x_i, v)} h_p(v(x_i)), \tag{3}$$

where $\mathcal{V}$ indexes views, $\mathcal{P}$ indexes tokens, and M counts the collected activation vectors. Each projection has its own center. The reference uses all support classes and is not a class prototype or a guaranteed separation of domain and class information. Because earlier layers subsequently adapt, the fixed-reference deviation $h - \mu$ need not retain zero empirical mean throughout optimization.

3.2 CENTERED LEARNING WITH FORWARD COMPENSATION

CenWrite separates the signal for factor learning from preservation of the shared forward response using the compensated layer

$$f(h) = W_0 h + \Delta W(h - \mu) + \text{sg}(\Delta W \mu) + b, \tag{4}$$

where sg preserves its argument's value while blocking its gradient, and $b \in \mathbb{R}^{d_{\text{out}}}$ is a separate trainable residual. At matched inputs and parameters, the forward value equals $W_0 h + \Delta W h + b$. Compensation restores the response removed by centering without directly updating the factors through that response. It is recomputed from the current factors and the fixed center, introduces no trainable parameters, and requires no additional support pass.

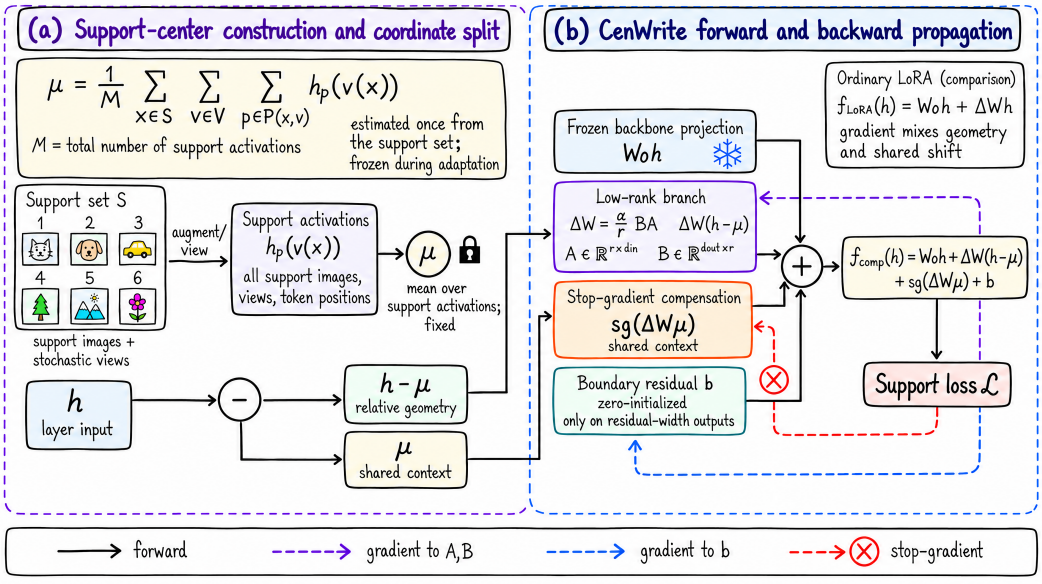


Figure 2: Support-center construction and CenWrite propagation. (a) Support images, augmented views, and tokens define the fixed mean μ; $h - \mu$ is the centered activation, labeled "relative geometry." (b) The frozen projection, centered low-rank response, detached compensation, and boundary residual sum to Equation 4. The upper-right ordinary-LoRA box is a comparison, not an added computational branch. The fixed μ is estimated before subtraction and supplies its second input; it is not the subtraction output. Dashed arrows indicate gradients, and the red cross blocks compensation gradients. "Shared context" denotes the mean or its response. Support-image icons illustrate samples without specifying the episode's class count or shot number.

For a fixed local input h, let g be the incoming output gradient of the support loss. Automatic differentiation of Equation 4 gives the factor gradients

$$\nabla_A \mathcal{L} = sB^\top g(h-\mu)^\top, \qquad \nabla_B \mathcal{L} = s\, g\, [A(h-\mu)]^\top. \tag{5}$$

Ordinary LoRA replaces $h - \mu$ by h in both expressions. The detached mean response changes this local learning signal while preserving the forward value at a fixed state. Gradients still accumulate over support activations, and upstream adapted layers interact through the network. The local input derivative remains $W_0 + \Delta W$. These identities specify the intervention; the matched controls in Section 5.2 test its predictive benefit.

3.3 Independent Boundary Adjustment

The residual learns a shared correction with local gradient $\nabla_b \mathcal{L} = g$. It is inserted only at attention-output and second-MLP projections, which return to the residual width. Subsequent nonlinear layers can turn that shared correction into sample-dependent logit changes. Both paths can therefore affect predictions; their independence concerns parameters and direct gradients, not exclusive control of decision boundaries. We initialize A with Kaiming uniform and set $B = b = 0$, initially preserving the pretrained projection. The classifier, factors, and residual are then optimized together using the settings in Appendix A.

4 Experimental Setup

Domains and adaptation. We evaluate EuroSAT and PlantVillage (Helber et al., 2019; Mohanty et al., 2016), ISIC 2018 Task 3 and single-label NIH ChestX (Codella et al., 2019; Wang et al., 2017). Each five-way task has one or five labeled support images and 15 queries per class, with 200 matched episodes per setting. Classes, order, and support/query identities are shared across methods. Available leaf, lesion, or patient identifiers separate support and query samples; missing

mappings and EuroSAT use image identities. DINOv3 ViT-B/16 is the primary encoder (Siméoni et al., 2025); backbone weights remain frozen during support-only adaptation. Queries enter after 50 AdamW steps, and no post-hoc calibration is applied.

Comparators and outcomes. Ordinary LoRA matches CenWrite's rank, targets, and optimizer settings. FacT, BOFT, FourierFT, RandLoRA, and ReFTA share its classifier, episodes, and training schedule, with different projection targets and parameter budgets. Adapter learning rates are selected on held-out images within reused support tasks, excluding formal queries, then fixed across settings. The seven-method comparison is repeated on DINOv2 ViT-S/14 (Oquab et al., 2024). Accuracy and macro-F1 measure classification; NLL, unscaled multiclass Brier, and 15-bin top-label ECE measure probability quality. Aggregate means weight all settings equally. Table boldface and underlining rank displayed means, not statistical significance. Paired bootstrap intervals are simultaneous within specified comparison families; Appendix A gives configurations and statistical scope.

Material-morphology evaluation. UHCS steel microstructures and NFFA SEM morphology (Hecht et al., 2017; Aversa et al., 2018) contain 601 and 11,540 prepared images. Each five-way setting uses 200 episodes, with 10 queries per class for UHCS and 15 for NFFA. One-shot support is nested in five-shot support with shared queries. UHCS separates recorded sample groups; NFFA separates detected image families. Two additional UHCS splits test sensitivity. Both use DINOv3 and transferred 50-step settings for LoRA, BOFT, ReFTA, and component variants. Each shot is one image; it need not represent an independent specimen. Appendices E and F give the complete protocols.

5 Results

5.1 Classification and Probability Quality

Table 1: DINOv3 results (200 matched episodes per setting). Top: equal-weight means; bottom: domain–shot results. Bold/underline: best/second-best displayed means, including ties, not significance. Complete metrics and intervals are in Appendix B.

Dataset, shot	Metric	LoRA	FacT	BOFT	FourierFT	RandLoRA	ReFTA	CenWrite
Eight-setting equal-weight means								
NLL ↓		1.2955	1.8333	1.5175	1.5400	3.2041	1.2797	**1.2032**
Brier ↓		0.5583	0.6347	0.6006	0.6104	0.8232	0.5666	**0.5380**
ECE (%) ↓		22.79	29.07	26.22	26.79	39.53	22.43	**20.59**
Accuracy (%) ↑		**61.35**	60.95	61.09	60.53	53.83	60.28	61.27
Macro-F1 (%) ↑		**60.08**	59.56	59.77	59.16	51.11	58.92	59.91
Per-setting means								
EuroSAT, 1	Acc. (%) ↑	73.41	**73.45**	73.11	72.79	58.49	70.65	72.81
	NLL ↓	**0.7414**	0.9416	0.7931	0.7865	2.7299	0.8335	0.7587
EuroSAT, 5	Acc. (%) ↑	**91.73**	91.45	91.68	90.31	70.86	90.15	91.28
	NLL ↓	**0.2451**	0.3631	0.2796	0.3099	1.2935	0.2886	0.2562
PlantVillage, 1	Acc. (%) ↑	90.85	87.93	88.58	87.69	78.85	89.20	**91.11**
	NLL ↓	0.2659	0.4057	0.3276	0.3626	1.3108	0.3134	**0.2603**
PlantVillage, 5	Acc. (%) ↑	**97.59**	96.90	97.49	97.16	94.09	97.14	**97.59**
	NLL ↓	0.0738	0.1156	0.0797	0.0885	0.3307	0.0876	**0.0726**
ISIC, 1	Acc. (%) ↑	**36.18**	35.78	36.10	35.38	32.49	35.58	35.97
	NLL ↓	1.9810	2.9349	2.2024	2.2353	5.6700	1.9722	**1.9362**
ISIC, 5	Acc. (%) ↑	54.28	54.73	54.61	54.13	49.47	52.23	**54.77**
	NLL ↓	1.5065	2.2838	1.8052	1.8187	3.3959	1.5357	**1.3787**
ChestX, 1	Acc. (%) ↑	21.83	**22.27**	21.76	21.62	21.91	21.95	21.36
	NLL ↓	2.7546	3.7548	3.1159	3.2897	6.1366	2.7729	**2.5977**
ChestX, 5	Acc. (%) ↑	24.90	25.09	**25.40**	25.17	24.47	25.31	25.25
	NLL ↓	2.7960	3.8672	3.5367	3.4289	4.7654	2.4336	**2.3651**

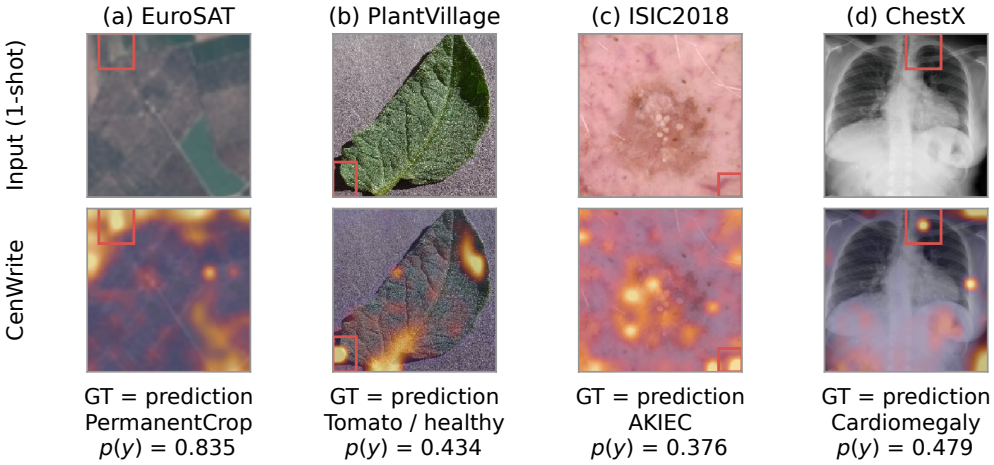


Figure 3: Successful four-domain predictions after 50 steps of 1-shot adaptation. Each input is paired with its true-class logit gradient response. Red boxes mark a common local window, not a component-specific attribution. Predictions, probabilities, and display construction are documented in Appendix H.

CenWrite achieves the lowest mean NLL, ECE, and Brier among the seven DINOv3 adaptation methods (Table 1). Relative to ordinary LoRA, eight-setting mean NLL decreases from 1.2955 to 1.2032 and Brier from 0.5583 to 0.5380, reductions of 7.13% and 3.63%; ECE falls by 2.20 percentage points. All 18 mean probability-score improvements have positive simultaneous intervals in the supplementary 30-endpoint analysis. NLL and Brier gains are largest on ISIC and ChestX. Accuracy and macro-F1 also improve over FourierFT, RandLoRA, and ReFTA with interval support. The advantage is strongest in mean probability quality. Complete metrics and intervals are in Appendix B.

Successful predictions connect the aggregate results to actual target images (Figure 3). Each domain pairs a query with CenWrite's prediction, true-class probability, and input-gradient response after 1-shot adaptation. Red boxes locate a shared observation window in the input and response. The maps show sensitivity of the true-class logit to the input. These selected successful predictions come from the existing qualitative evaluation. Appendix H documents their calculation, selection, and paired LoRA responses. Averaged over four domains and both support sizes, CenWrite improves all three probability scores against each of the six adapters.

5.2 Component Contributions

Matched component controls distinguish the learning route from extra residual parameters (Table 2). Full CenWrite has the lowest 1-shot mean NLL and Brier, 1.3882 and 0.6232. With the residual present, compensated centering reduces equal-shot mean NLL by 0.0963 and Brier by 0.0212 relative to uncentered learning. Without a residual, it also improves all three mean probability scores over LoRA. Forward compensation further reduces the three probability errors relative to direct centering in both residual conditions; full CenWrite's NLL gain over the no-compensation control is 0.0940. These results support separately controlling factor learning and the shared response. Complete effects and optimization trajectories are in Appendix C.

The independent residual adds a distinct classification adjustment within the complete path. Relative to the no-residual variant, full CenWrite increases mean macro-F1 by 0.152 percentage points, with a simultaneous 95% interval of $[0.044, 0.260]$. Both macro-F1 interaction intervals are also positive: its residual benefit is larger with centering and compensation together than under uncentered or uncompensated updates. This supports the combination of operations more directly than ranking their mean scores alone. Centering and compensation account for the clearest probability gains, while the residual improves macro-F1 in the full method. All settings and interactions remain in Appendix C.

Table 2: Four-domain component means by shot. C/P/R: centering, detached compensation, residual; ✓/×: enabled/disabled. Variant definitions are in Appendix A.4. Bold/underline: best/second-best means, including ties. Gray identifies full CenWrite (F).

Shot	Variant	C	P	R	NLL ↓	Brier ↓	ECE ↓ (%)	Acc. ↑ (%)	Macro-F1 ↑ (%)
1-shot	L	×	×	×	1.4357	0.6348	25.39	55.57	53.61
	NC	×	×	✓	1.4389	0.6356	25.49	55.58	53.61
	P	✓	×	×	1.4410	0.6318	25.34	**55.83**	**53.96**
	NP	✓	×	✓	1.4453	0.6329	25.37	55.82	53.94
	NB	✓	✓	×	1.3902	0.6237	**24.00**	55.23	52.94
	F	✓	✓	✓	**1.3882**	**0.6232**	24.09	55.31	53.11
5-shot	L	×	×	×	1.1553	0.4818	20.18	67.13	66.54
	NC	×	×	✓	1.1600	0.4828	20.21	67.13	66.55
	P	✓	×	×	1.1395	0.4745	19.53	67.47	66.96
	NP	✓	×	✓	1.1489	0.4763	19.70	**67.48**	**66.97**
	NB	✓	✓	×	**1.0108**	**0.4509**	**16.89**	67.12	66.57
	F	✓	✓	✓	1.0182	0.4528	17.08	67.22	66.71

Table 3: Backbone comparisons. (a) Four-domain mean accuracy by shot; bold/underline: best/second-best means, including ties. (b) Equal-shot accuracy gains; †: simultaneous 95% interval excludes zero within the backbone's 30-endpoint family. DINOv3: ViT-B/16; DINOv2: ViT-S/14.

(a) Four-domain mean accuracy (%)

Backbone	Shot	LoRA	FacT	BOFT	FourierFT	RandLoRA	ReFTA	CenWrite
DINOv3	1-shot	**55.57**	54.86	54.89	54.37	47.93	54.34	55.31
DINOv3	5-shot	67.13	67.04	**67.30**	66.70	59.72	66.21	67.22
DINOv2	1-shot	51.76	50.92	50.89	34.70	29.29	51.64	**52.14**
DINOv2	5-shot	65.19	64.37	65.10	46.20	42.15	64.85	**65.37**

(b) Equal-shot accuracy gains over each baseline (pp)

Backbone	LoRA	FacT	BOFT	FourierFT	RandLoRA	ReFTA
DINOv3	−0.08	+0.32	+0.18	+0.74†	+7.44†	+0.99†
DINOv2	+0.28†	+1.11†	+0.76†	+18.30†	+23.03†	+0.51†

5.3 Generalization across Backbone Configurations

The complete seven-method comparison tests applicability to a second encoder, DINOv2 ViT-S/14, under the same domains, support sizes, and matched tasks (Table 3). Each method is adapted separately on that encoder. CenWrite reaches 52.138% mean accuracy in 1-shot and 65.368% in 5-shot, the highest displayed means in both conditions, exceeding LoRA by 0.380 and 0.175 percentage points. These matched comparisons support the accuracy gain under DINOv2.

The equal-shot aggregate further supports this classification advantage. CenWrite reaches 58.753% mean accuracy on DINOv2, with positive simultaneous intervals against all six baselines. Its gains over LoRA, BOFT, FacT, and ReFTA are 0.28, 0.76, 1.11, and 0.51 percentage points, respectively. Together with the DINOv3 probability results, this supports the method's applicability to two encoder configurations. The favorable metric is probability quality on DINOv3 and accuracy on DINOv2. We report per-shot mean rankings separately from the aggregate intervals, and each backbone retains its own 30-endpoint family. Appendix D includes all five metrics and domain-level contrasts; the encoders differ in architecture, capacity, and pretraining.

Table 4: Material adapter comparisons (200 matched tasks per shot; UHCS fold 5). Bold/underline: best/second-best means, including ties. Complete material results are in Appendices E and F.

Dataset	Shot	Metric	LoRA	BOFT	ReFTA	CenWrite
UHCS	1-shot	NLL ↓	<u>1.3225</u>	1.4956	1.3558	**1.2755**
		Brier ↓	<u>0.6545</u>	0.7003	0.6693	**0.6397**
		ECE (%) ↓	<u>23.50</u>	27.16	23.58	**22.35**
		Acc. (%) ↑	<u>52.31</u>	50.58	50.17	**52.71**
		Macro-F1 (%) ↑	<u>49.67</u>	47.47	47.05	**50.05**
UHCS	5-shot	NLL ↓	<u>0.8942</u>	1.1873	0.9239	**0.8331**
		Brier ↓	<u>0.4534</u>	0.5199	0.4693	**0.4434**
		ECE (%) ↓	<u>18.20</u>	23.24	18.64	**17.24**
		Acc. (%) ↑	**68.58**	66.94	67.09	<u>68.13</u>
		Macro-F1 (%) ↑	**67.53**	64.97	65.72	<u>67.31</u>
NFFA	1-shot	NLL ↓	<u>1.1173</u>	1.2214	1.1890	**1.1134**
		Brier ↓	<u>0.5498</u>	0.5931	0.5873	**0.5485**
		ECE (%) ↓	**15.04**	16.80	15.52	<u>15.08</u>
		Acc. (%) ↑	<u>58.18</u>	55.42	54.97	**58.32**
		Macro-F1 (%) ↑	<u>57.00</u>	53.80	53.45	**57.19**
NFFA	5-shot	NLL ↓	<u>0.6495</u>	0.7555	0.7078	**0.6342**
		Brier ↓	<u>0.3155</u>	0.3538	0.3491	**0.3115**
		ECE (%) ↓	<u>11.40</u>	13.36	12.16	**11.32**
		Acc. (%) ↑	**78.01**	76.08	75.31	<u>77.87</u>
		Macro-F1 (%) ↑	**77.98**	75.99	75.25	<u>77.84</u>

5.4 Discussion

Material microstructure and morphology recognition test the same adaptation design in an application with scarce labeled images. UHCS steel micrographs and NFFA SEM images use matched 1-shot and 5-shot tasks, DINOv3, and 50 support-only optimization steps. The comparisons in Table 4 include ordinary LoRA, butterfly orthogonal adaptation (BOFT), and tensor adaptation (ReFTA), with each support size reported separately. Complete protocols and all reported variants are retained in Appendices E and F.

Probability-quality gains extend to both material evaluations. Relative to ordinary LoRA, CenWrite has lower mean NLL and Brier at both support sizes. Equal-shot mean NLL decreases by 0.0541 on UHCS and 0.0096 on NFFA, with positive simultaneous intervals for these improvements. Against BOFT and ReFTA, all five aggregate metrics improve on both datasets. Material component controls link these gains to the proposed learning route: compensated centering lowers NLL and Brier relative to uncentered learning in both residual conditions. Two additional UHCS group splits retain the three mean probability gains over LoRA, providing split-sensitivity evidence within the same collection. Appendices E, F, and G report complete split, component, and trajectory results.

Paired-shot examples illustrate how the probability changes appear on the same real query image (Figure 4). For the UHCS example, CenWrite assigns the true class probabilities of 80.4% and 62.3% in 1-shot and 5-shot, both above LoRA's corresponding values. For NFFA, CenWrite recognizes fibres in both conditions and corrects LoRA's porous-sponge prediction in 5-shot. The full class probabilities and paired NLL/Brier changes expose the predictions behind these examples. They were selected post hoc for success in both support conditions, so the complete material comparisons, rather than the examples' frequency, support the aggregate conclusions. Their sources, selection rule, and complete records appear in Appendix H.

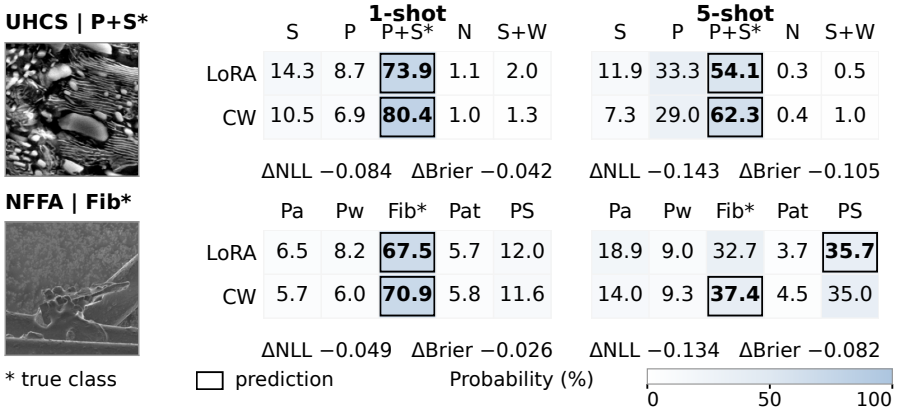


Figure 4: Paired-shot material predictions. Boxes mark predictions; stars mark true classes: pearlite+spheroidite (P+S) and fibres (Fib). NLL/Brier changes are CenWrite minus LoRA (lower is better). These are selected successes; class keys, sources, and selection are in Appendix H.

6 Conclusion

CenWrite separates centered factor learning from the shared forward response and an independent boundary adjustment. The complete DINOv3 comparison substantiates lower average probability errors against six baselines; DINOv2 supports higher average accuracy against the same methods. UHCS and NFFA further support lower NLL and Brier than ordinary LoRA. Their component controls corroborate the centered learning route in both residual conditions. The original-domain ablations support a probability benefit from compensation and a macro-F1 benefit from the residual. These findings substantiate the proposed separation under scarce labels.

7 Limitations

Benefits depend on the metric, encoder, and comparator. DINOv3 classification differences from LoRA, BOFT, and FacT are unresolved, and EuroSAT probability gains are not uniform. On DINOv2, NLL/ECE exceed LoRA's, NLL exceeds BOFT's, and all three errors exceed ReFTA's. The full residual improves original-domain macro-F1 at a probability cost. On UHCS, classification gains over LoRA are supported only in fold 31; compensation costs accuracy/F1 and the residual costs NLL/Brier. NFFA compensation/residual effects remain unresolved. The complete comparisons support metric-specific benefits under the tested settings.

Fixed pools, reused support tasks, unequal budgets, and transferred settings limit generalization. Development/evaluation disjointness is unverified; episode intervals exclude seed and pool uncertainty. Encoder architecture, capacity, and pretraining change together. UHCS has small group pools and magnification/detector confounding; NFFA lacks specimen identities. Historical probability exports are incomplete, and some rank runs differ from planned episode counts. A support mean can mix domain and class information. Selected successful cases do not estimate success frequency; input gradients and local algebra do not causally establish the measured gains. Neither acquisition-independent transfer, materials-property utility, nor clinical utility is established.

8 Future Work

Future work should test independent data pools, repeated seeds, and new materials. Matched-capacity encoders can separate pretraining from architecture. Comparing fixed, updated, and regularized centers with factor optimizers (Yen et al., 2025; Tastan et al., 2026) under support-only selection would assess complementary classification and probability gains.

AI USE STATEMENT

We used generative AI tools to assist with parts of the experimental workflow and manuscript preparation. Applications included experiment-related assistance, sentence revision, section compression to meet conference page limits, and formalization of diagnostic explanations regarding baseline behavior. All empirical findings, numerical results, tables, figures, model checkpoints, and reported analyses are derived from actual experimental runs and execution logs. No empirical results, data, checkpoints, figures, or conclusions were generated, fabricated, or synthetically constructed by AI. The authors conducted and verified the experiments, reviewed all AI-assisted outputs, understand the role of the AI tools in this work, and determined and verified the final interpretations and conclusions. The authors take full responsibility for the accuracy, originality, and integrity of the final submission.

REFERENCES

Soroush Abbasi Koohpayegani, K L Navaneet, Parsa Nooralinejad, Soheil Kolouri, and Hamed Pirsiavash. NOLA: Compressing LoRA using Linear Combination of Random Basis. In *International Conference on Learning Representations*, 2024.

Shihab Ahamed, Udaya Sampath K Perera Miriya Thanthrige, Ranga Rodrigo, and Muhammad Haris Khan. A-TPT: Angular Diversity Calibration Properties for Test-Time Prompt Tuning of Vision-Language Models. In *International Conference on Learning Representations*, 2026.

Paul Albert, Frederic Zhang, Hemanth Saratchandran, Cristian Rodriguez-Opazo, Anton van den Hengel, and Ehsan Abbasnejad. RandLoRA: Full rank parameter-efficient fine-tuning of large models. In *International Conference on Learning Representations*, 2025.

Rossella Aversa, Mohammad Hadi Modarres, Stefano Cozzini, and Regina Ciancio. NFFA-EUROPE – 100% SEM Dataset. NFFA-EUROPE Project, EUDAT B2SHARE, 2018. URL https://b2share.eudat.eu/records/zja8y-53j14. Version 2.0.

Massimo Bini, Leander Girrbach, and Zeynep Akata. Decoupling Angles and Strength in Low-rank Adaptation. In *International Conference on Learning Representations*, 2025.

Quentin Bouniot, Pavlo Mozharovskyi, and Florence d'Alché Buc. Tailoring Mixup to Data for Calibration. In *International Conference on Learning Representations*, 2025.

Dexia Chen, Qianjie Zhu, Weibing Li, Yue Yu, Tong Zhang, and Ruixuan Wang. Preserve and Sculpt: Manifold-Aligned Fine-tuning of Vision-Language Models for Few-Shot Learning. In *International Conference on Learning Representations*, 2026.

Zhixiang Chi, Li Gu, Tao Zhong, Huan Liu, Yuanhao Yu, Konstantinos Plataniotis, and Yang Wang. Adapting to Distribution Shift by Visual Domain Prompt Generation. In *International Conference on Learning Representations*, 2024.

Zhixiang Chi, Li Gu, Huan Liu, Ziqiang Wang, Yanan Wu, Yang Wang, and Konstantinos Plataniotis. Learning to Adapt Frozen CLIP for Few-Shot Test-Time Domain Adaptation. In *International Conference on Learning Representations*, 2025.

Muthu Chidambaram and Rong Ge. Reassessing How to Compare and Improve the Calibration of Machine Learning Models. In *International Conference on Learning Representations*, 2025.

Kamil Ciosek, Nicolò Felicioni, Sina Ghiassian, Juan Elenter, Francesco Tonolini, David Gustafsson, Eva Garcia-Martin, Carmen Gonzalez, and Raphaëlle Bertrand-Lalo. Measuring Uncertainty Calibration. In *International Conference on Learning Representations*, 2026.

Noel C. F. Codella, Veronica Rotemberg, Philipp Tschandl, M. Emre Celebi, Stephen Dusza, David Gutman, Brian Helba, Aadi Kalloo, Konstantinos Liopyris, Michael Marchetti, et al. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC). *CoRR*, abs/1902.03368, 2019.

Zhekai Du, Yinjie Min, Jingjing Li, Ke Lu, Changliang Zou, Liuhua Peng, Tingjin Chu, and Mingming Gong. LoCA: Location-Aware Cosine Adaptation for Parameter-Efficient Fine-Tuning. In *International Conference on Learning Representations*, 2025.

Panagiotis Eustratiadis, Ł ukasz Dudziak, Da Li, and Timothy Hospedales. Neural Fine-Tuning Search for Few-Shot Learning. In *International Conference on Learning Representations*, 2024.

Ziqi Gao, Qichao Wang, Aochuan Chen, Zijing Liu, Bingzhe Wu, Liang Chen, and Jia Li. Parameter-efficient fine-tuning with discrete fourier transform. In *International Conference on Machine Learning*, pp. 14884–14901, 2024.

Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On calibration of modern neural networks. In *International Conference on Machine Learning*, 2017.

Yunhui Guo, Noel Codella, Leonid Karlinsky, James V. Codella, John R. Smith, Kate Saenko, Tajana Rosing, and Rogério Feris. A broader study of cross-domain few-shot learning. In *European Conference on Computer Vision*, pp. 124–141, 2020.

Cheng Han, Qifan Wang, Yiming Cui, Wenguan Wang, Lifu Huang, Siyuan Qi, and Dongfang Liu. Facing the Elephant in the Room: Visual Prompt Tuning or Full finetuning? In *International Conference on Learning Representations*, 2024.

Matthew D. Hecht, Brian L. DeCost, Toby Francis, Elizabeth A. Holm, Yoosuf N. Picard, and Bryan A. Webler. Ultrahigh carbon steel micrographs. https://hdl.handle.net/11256/940, 2017. Micrograph dataset.

Patrick Helber, Benjamin Bischke, Andreas Dengel, and Damian Borth. EuroSAT: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(7):2217–2226, 2019.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022.

Jincheng Huang, Jie Xu, Xiaoshuang Shi, Ping Hu, Lei Feng, and Xiaofeng Zhu. Revisiting Confidence Calibration for Misclassification Detection in VLMs. In *International Conference on Learning Representations*, 2026.

Junseo Hwang, Wonguk Cho, and Taesup Kim. PiCa: Parameter-Efficient Fine-Tuning with Column Space Projection. In *International Conference on Learning Representations*, 2026.

Ziqi Jiang, Yanghao Wang, and Long Chen. Exploring Cross-Modal Flows for Few-Shot Learning. In *International Conference on Learning Representations*, 2026.

Shibo Jie and Zhi-Hong Deng. FacT: Factor-tuning for lightweight adaptation on vision transformer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 1060–1068, 2023.

Can Jin, Ying Li, Mingyu Zhao, Shiyu Zhao, Zhenting Wang, Xiaoxiao He, Ligong Han, Tong Che, and Dimitris Metaxas. LoR-VP: Low-Rank Visual Prompting for Efficient Vision Model Adaptation. In *International Conference on Learning Representations*, 2025.

Shiwei Li, Xiandi Luo, Haozhao Wang, Xing Tang, Ziqiang Cui, Dugang Liu, Yuhua Li, Yichen Li, Xiuqiang He, and Ruixuan Li. BoRA: Towards More Expressive Low-Rank Adaptation with Block Diversity. In *International Conference on Learning Representations*, 2026a.

Wenhao Li, Xianjing Meng, Qiangchang Wang, Zhongyi Han, Zhibin Wu, and Yilong Yin. DVLA-RL: Dual-Level Vision–Language Alignment with Reinforcement Learning Gating for Few-Shot Learning. In *International Conference on Learning Representations*, 2026b.

Yushu Li, Yongyi Su, Adam Goodge, Kui Jia, and Xun Xu. Efficient and Context-Aware Label Propagation for Zero-/Few-Shot Training-Free Adaptation of Vision-Language Model. In *International Conference on Learning Representations*, 2025.

Vladislav Lialin, Sherin Muckatira, Namrata Shivagunde, and Anna Rumshisky. ReLoRA: High-Rank Training Through Low-Rank Updates. In *International Conference on Learning Representations*, 2024.

Muqing Liu, Chongjie Si, and Yuheng Jia. FlexLoRA: Entropy-Guided Flexible Low-Rank Adaptation. In *International Conference on Learning Representations*, 2026.

Weiyang Liu, Zeju Qiu, Yao Feng, Yuliang Xiu, Yuxuan Xue, Longhui Yu, Haiwen Feng, Zhen Liu, Juyeon Heo, Songyou Peng, Yandong Wen, Michael J Black, Adrian Weller, and David Ha. Parameter-Efficient Orthogonal Finetuning via Butterfly Factorization. In *International Conference on Learning Representations*, 2024.

Xu-Hui Liu, Yali Du, Jun Wang, and Yang Yu. On the Optimization Landscape of Low Rank Adaptation Methods for Large Language Models. In *International Conference on Learning Representations*, 2025.

Sharada P. Mohanty, David P. Hughes, and Marcel Salathé. Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, 7:1419, 2016. doi: 10.3389/fpls.2016.01419.

Lukas Muttenthaler, Robert A Vandermeulen, Qiuyi (Richard) Zhang, Thomas Unterthiner, and Klaus R Muller. Set Learning for Accurate and Calibrated Models. In *International Conference on Learning Representations*, 2024.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy V. Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. DINOv2: Learning robust visual features without supervision. *Transactions on Machine Learning Research*, 2024.

Stylianos Poulakakis-Daktylidis and Hadi J. Rad. BECLR: Batch Enhanced Contrastive Few-Shot Learning. In *International Conference on Learning Representations*, 2024.

Steffen Schotthöfer, Emanuele Zangrando, Gianluca Ceruti, Francesco Tudisco, and Jonas Kusch. GeoLoRA: Geometric integration for parameter efficient fine-tuning. In *International Conference on Learning Representations*, 2025.

Chongjie Si, Zhiyi Shi, Shifan Zhang, Xiaokang Yang, Hanspeter Pfister, and Wei Shen. Unleashing the Power of Task-Specific Directions in Parameter Efficient Fine-tuning. In *International Conference on Learning Representations*, 2025.

Oriane Siméoni, Huy V. Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seungeun Yi, Michaël Ramamonjisoa, Francisco Massa, et al. DINOv3. *arXiv preprint arXiv:2508.10104*, 2025.

Haobo Song, Hao Zhao, Soumajit Majumder, and Tao Lin. Increasing Model Capacity for Free: A Simple Strategy for Parameter Efficient Fine-tuning. In *International Conference on Learning Representations*, 2024.

Chengli Tan, Yubo Zhou, Haishan Ye, Guang Dai, Junmin Liu, Zengjie Song, Jiangshe Zhang, Zixiang Zhao, Yunda Hao, and Yong Xu. Towards Understanding The Calibration Benefits of Sharpness-Aware Minimization. In *International Conference on Learning Representations*, 2026.

Linwei Tao, Younan Zhu, Haolan Guo, Minjing Dong, and Chang Xu. A Benchmark Study on Calibration. In *International Conference on Learning Representations*, 2024.

Nurbek Tastan, Stefanos Laskaridis, Martin Takáč, Karthik Nandakumar, and Samuel Horváth. LoFT: Low-Rank Adaptation That Behaves Like Full Fine-Tuning. In *International Conference on Learning Representations*, 2026.

Hung-Yu Tseng, Hsin-Ying Lee, Jia-Bin Huang, and Ming-Hsuan Yang. Cross-domain few-shot classification via learned feature-wise transformation. In *International Conference on Learning Representations*, 2020.

Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M. Summers. ChestX-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2097–2106, 2017.

Yan Wang, Wei-Lun Chao, Kilian Q. Weinberger, and Laurens van der Maaten. SimpleShot: Revisiting nearest-neighbor classification for few-shot learning. *arXiv preprint arXiv:1911.04623*, 2019. URL https://arxiv.org/abs/1911.04623v2.

Zhengbo Wang, Jian Liang, Ran He, Zilei Wang, and Tieniu Tan. LoRA-Pro: Are Low-Rank Adapters Properly Optimized? In *International Conference on Learning Representations*, 2025.

Yize Wu, Ke Gao, Ling Li, and Yanjun Wu. Stable-LoRA: Stabilizing Feature Learning of Low-Rank Adaptation. In *International Conference on Learning Representations*, 2026.

Zhuoyan Xu, Zhenmei Shi, Junyi Wei, Fangzhou Mu, Yin Li, and Yingyu Liang. Towards Few-Shot Adaptation of Foundation Models via Multitask Finetuning. In *International Conference on Learning Representations*, 2024.

Bangguo Ye, Yuanwei Zhang, and Xiaoqun Zhang. TRAC: Tensor-Train based Across-layer Compression for Parameter-Efficient Fine-Tuning. In *International Conference on Learning Representations*, 2026a.

Jingqi Ye, Haonan He, Minglei Li, Fujun Han, Tao Chen, and Peng Ye. Gradient Intrinsic Dimensionality Alignment: Narrowing The Gap Between Low-Rank Adaptation and Full Fine-Tuning. In *International Conference on Learning Representations*, 2026b.

Jui-Nan Yen, Si Si, Zhao Meng, Felix Yu, Venkata Sai Surya Subramanyam Duvvuri, Inderjit Dhillon, Cho-Jui Hsieh, and Sanjiv Kumar. LoRA Done RITE: Robust Invariant Transformation Equilibration for LoRA Optimization. In *International Conference on Learning Representations*, 2025.

Hee Suk Yoon, Eunseop Yoon, Joshua Tian Jin Tee, Mark Hasegawa-Johnson, Yingzhen Li, and Chang Yoo. C-TPT: Calibrated Test-Time Prompt Tuning for Vision-Language Models via Text Feature Dispersion. In *International Conference on Learning Representations*, 2024.

Yuchen Zeng and Kangwook Lee. The Expressive Power of Low-Rank Adaptation. In *International Conference on Learning Representations*, 2024.

Min Zhang, Haoxuan Li, Fei Wu, and Kun Kuang. MetaCoCo: A New Few-Shot Classification Benchmark with Spurious Correlation. In *International Conference on Learning Representations*, 2024.

Jingjing Zheng, Anda Tang, Qiangqiang Mao, Zhouchen Lin, and Yankai Cao. ReFTA: Breaking the weight reconstruction bottleneck in tensorized parameter-efficient fine-tuning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 26369–26378, 2026.

Jinyang Zheng and Tong Wu. LoRA-S: An Efficient Low Rank Adaptation scheme via Sylvester equation. In *International Conference on Learning Representations*, 2026.

A Implementation and Evaluation Details

Table 5: Evidence guide. Each entry links a main-text question to the complete results and protocol details used to assess it. Appendix numbering follows the reading order rather than the order in which experiments finished.

Question	Main-text evidence	Complete evidence
Overall performance	Section 5.1, Table 1	Appendix B
Component roles	Section 5.2, Table 2	Appendix C
Backbone dependence	Section 5.3	Appendix D
Material morphology	Section 5.4, Table 4	Appendices E, F
Prediction examples	Figure 4	Appendix H
Process, cost and rank	Supplementary diagnostics	Appendices G, I, J

A.1 SUPPORT ADAPTATION

Table 6 summarizes the common CenWrite configuration. Inputs use a bicubic random resized crop to 224×224, with crop scale $[0.8, 1.0]$ and ImageNet normalization. Horizontal flips are disabled for ChestX. Query preprocessing resizes to 256 and takes a 224×224 center crop. The classifier, adapters, and optimizer reset for each episode. The support batch contains all 5 or 25 support images; query batch size is 64. AdamW uses weight decay 0.01 and gradient-norm clipping at 1.0, with bfloat16 autocasting and deterministic CUDA settings. Adaptation runs for 50 steps with classifier rate 10^{-3} and adapter and boundary rate 1.5×10^{-4}. Rank and scale are both 16, with zero adapter dropout. These settings apply to all eight primary tasks. Backbone weights remain frozen; queries are evaluated after support fitting.

Table 6: CenWrite adaptation settings used across the eight primary settings. The centering entry specifies the executed view count: three paired draws produce six augmented views. "Carrier" refers to the branch parameterized by the low-rank factors.

Setting	Value
Backbone	DINOv3 ViT-B/16, LVD-1689M; 12 blocks, width 768, 12 heads
Input	224×224 RGB; patch size 16; 4 storage tokens
Datasets	EuroSAT, PlantVillage, ISIC2018, ChestX
Episode protocol	5-way, 1/5-shot; 15 query images/class; 200 episodes/setting
Optimization	AdamW; 50 support-only steps; gradient norm cap 1.0
Batch sizes	Support: 5/25 for 1/5-shot; query: 64
Adapted modules	Fused QKV, attention output, MLP FC1/FC2; learned boundaries
Carrier	Rank 16; alpha 16; dropout 0
Learning rates	Head: 1e-3; carrier and boundary: 1.5e-4
Weight decay	0.01 for head, carrier, and boundary
Support centering	3 paired draws (6 augmented views); fixed full-support mean
Seed / precision	Base seed 20260906; strict determinism; BF16 autocast
Augmentation / calibration	Horizontal flip except ChestX; no post-hoc calibration

The center-collection loop takes three paired draws, forwarding both augmented views in each draw. It therefore averages six views per support image, including all token positions at each adapted layer. Centers are collected before optimization and kept fixed. The random-number-generator state is restored after collection so this step does not advance the adaptation augmentation sequence. The mean is not rescaled. CenWrite has 2,359,296 low-rank parameters, 18,432 boundary parameters across 24 residual-output projections, and a 3,845-parameter five-way classifier, for 2,381,573 trainable parameters in total. The 48 adapted projections comprise one fused QKV projection, the attention output, and the two MLP projections in each of 12 transformer blocks. The main model uses the official DINOv3 ViT-B/16 LVD-1689M weights. The secondary backbone is DINOv2 ViT-S/14; its capacity and pretraining differ.

A.2 TRANSFERRED BASELINES AND SUPPORT-BASED SELECTION

All baseline classifiers use rate 10^{-3}. FacT-TT uses rank 16, scale 1, dropout 0.1, shared input/output factors across Q/K/V/O and both MLP projections, adapter rate 0.003, and weight decay 0.0001. BOFT uses the same targets, block size 4, two butterfly factors, dropout 0.1, target-only trainable biases, rate 0.001, and decay 0.02. Its exact portable block-diagonal kernel replaces the custom CUDA extension. FourierFT uses 3,000 frequencies on Q/V, scale 300, location seed 2024, Gaussian spectrum initialization, rate 0.06, and decay 0.0003. RandLoRA uses $r = 8$, $\alpha = 10$, zero dropout, dense shared random bases with seed 0, fused QKV and both MLP projections, rate 0.003, and decay 0.1; its r is not LoRA rank. ReFTA uses global rank 15 per Q/V stack, fixed cosine transforms with corrected inverse orientation, rate 0.001, and weight decay 0.01. These transferred configurations use different trainable parameter budgets and projection targets.

The search compares adapter learning rates under the common 50-step schedule while holding the classifier rate and each baseline's other settings fixed. For FacT, BOFT, and ReFTA, the three candidate adapter learning rates are 0.0003, 0.001, and 0.003. FourierFT uses 0.06, 0.2, and 0.6; RandLoRA uses 0.003, 0.01, and 0.03. Ten reused support tasks, five each from EuroSAT and PlantVillage, provide four training and one held-out image per class. Five methods, three candidates, and ten tasks give 150 fits of 50 steps. Equal-domain held-out NLL selects one rate per method for all settings, with ties resolved toward the smaller rate. Other settings are fixed. Formal query labels, query results, and comparisons with CenWrite are excluded. This is validation within reused support tasks, not evidence of an independent development pool. The selected rates are carried unchanged into both shot counts; no separate query-based retuning is performed for the medical-image domains.

A.3 Metrics, pairing, and statistical scope

For a query set with N examples and C classes, let p_{ic} be the predicted class probability and y_i the true label. With $\hat{y}_i = \arg\max_c p_{ic}$, the probability scores are

$$\mathrm{NLL} = -\frac{1}{N}\sum_{i=1}^{N} \log p_{iy_i}, \tag{6}$$

$$\mathrm{Brier} = \frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} \left(p_{ic} - \mathbf{1}\{y_i = c\}\right)^2, \tag{7}$$

$$\mathrm{ECE} = \sum_{m=1}^{15} \frac{|\mathcal{B}_m|}{N} \left|\mathrm{acc}(\mathcal{B}_m) - \mathrm{conf}(\mathcal{B}_m)\right|. \tag{8}$$

Here $\mathcal{B}_m$ groups top-label probabilities into 15 equal-width bins on $[0, 1]$. The bin accuracy is the fraction with $\hat{y}_i = y_i$; bin confidence is the mean of $\max_c p_{ic}$. Empty bins contribute zero. Brier is not divided by the class count. Accuracy is the fraction of correct top-label predictions, and macro-F1 is the arithmetic mean of the five classwise F1 scores. ECE depends on binning and the finite query sample; NLL and Brier provide complementary proper-score evidence. NLL penalizes assigning little probability to the true class, whereas Brier scores the full class-probability vector. Both therefore capture changes that can leave the top-label prediction and its accuracy unchanged. ECE summarizes binned top-label calibration, whereas NLL and Brier are proper scores for probability forecasts.

Every metric is computed per episode before averaging. Pairing uses identical episode classes, labels, support/query identities, and order. The bootstrap draws episode indices jointly across the methods and metrics within each setting, then averages the eight setting means with equal weights. Updated baseline and component analyses use 20,000 draws with seed 2026092301. A fixed-standard-error maximum absolute standardized deviation gives approximate simultaneous 95% intervals within each stated family. Pointwise percentile intervals are also retained for descriptive setting-level comparisons. Resampling separately across the non-nested shot settings is an approximation; shared seeds do not establish their independence. Reused-image dependence and independent training-seed variation are not modeled, so the intervals remain conditional on the fixed data pools and task protocol, not variation across independently collected datasets.

The DINOv3 family includes six comparators and five metrics, giving 30 aggregate endpoints. Added after viewing the four-domain results, it is supplementary rather than preregistered. DINOv2 uses its separately specified 30-endpoint family for the same methods and metrics. Their intervals do not give joint coverage for 60 endpoints. The component family was fixed before the supplementary runs: eight direct contrasts and two interactions on five metrics give 50 endpoints. This local prospective definition also includes comparisons that reuse known historical results. Positive plotted effects mean higher accuracy/F1 or lower NLL/ECE/Brier. Accuracy, F1, and ECE differences use percentage points; NLL and Brier use native units unless explicitly scaled by 10^{-3}. Historical rank and resource summaries retain their stated protocols. Single-method intervals are descriptive, not paired-difference intervals; their overlap is not used to test a contrast between methods.

A.4 COMPONENT-REMOVAL DEFINITIONS

For the complete adapter in Equation 4, five controls separate factor learning, forward compensation, and the independent residual:

$$f_{\text{LoRA}}(h) = W_0 h + \Delta W h, \tag{9}$$
$$f_{\text{pure center}}(h) = W_0 h + \Delta W(h - \mu), \tag{10}$$
$$f_{\text{no center}}(h) = W_0 h + \Delta W h + b, \tag{11}$$
$$f_{\text{no compensation}}(h) = W_0 h + \Delta W(h - \mu) + b, \tag{12}$$
$$f_{\text{no boundary}}(h) = W_0 h + \Delta W(h - \mu) + \text{sg}(\Delta W \mu). \tag{13}$$

Ordinary LoRA and pure centering do not contain the independent residual. The no-centering variant differentiates through the full uncentered response; it does not add a second copy of $\Delta W \mu$. All other training settings and episodes are matched. We abbreviate the six variants as L (LoRA), NC (no centering), P (pure centering), NP (no compensation), NB (no residual), and F (full CenWrite). The direct contrasts are F–L, F–NC, NB–L, F–NP, NB–P, F–NB, NC–L, and NP–P. Together they test the complete method and the centered learning, compensation, and residual changes under their matched counterpart rules. They do not assume that a component has the same effect in every setting or combination of update rules and support conditions.

Let U_M be the mean for variant M, expressed in the reported units and oriented so larger values indicate higher accuracy/F1 or lower probability errors. The two interaction contrasts are

$$I_1 = (U_F - U_{NB}) - (U_{NC} - U_L),$$
$$I_2 = (U_F - U_{NB}) - (U_{NP} - U_P). \tag{14}$$

They compare the residual effect under the complete rule with its effect under uncentered and uncompensated updates, respectively. The resulting ten contrasts across five metrics form the prespecified 50-endpoint family. Table 20 gives every aggregate endpoint with its pointwise and simultaneous intervals. Appendix C also retains the absolute means and all setting-level contrasts, preventing aggregate improvements from concealing domain-specific reversals. A positive interaction describes a larger residual effect relative to another update rule, not superiority over every variant.

B COMPLETE DINOV3 BASELINE COMPARISONS

The complete DINOv3 comparison retains seven methods, five metrics, and all eight domain–shot settings, with 200 matched episodes per setting. Setting-level intervals are pointwise; the equal-setting comparisons use the separate supplementary 30-endpoint family defined in Appendix A.3. The 18 aggregate probability comparisons favor CenWrite. Classification intervals favor it over FourierFT, RandLoRA, and ReFTA. The tables report both aggregate comparisons and the complete results for every domain and support size.

Table 7: DINOv3 Accuracy (%) ↑ across all domains and support sizes. Bold/underline: best/second-best displayed means, including ties.

Dataset	Shot	LoRA	FacT	BOFT	FourierFT	RandLoRA	ReFTA	CenWrite
EuroSAT	1	<u>73.41</u>	**73.45**	73.11	72.79	58.49	70.65	72.81
EuroSAT	5	**91.73**	91.45	<u>91.68</u>	90.31	70.86	90.15	91.28
PlantVillage	1	<u>90.85</u>	87.93	88.58	87.69	78.85	89.20	**91.11**
PlantVillage	5	**97.59**	96.90	<u>97.49</u>	97.16	94.09	97.14	**97.59**
ISIC	1	**36.18**	35.78	<u>36.10</u>	35.38	32.49	35.58	35.97
ISIC	5	54.28	<u>54.73</u>	54.61	54.13	49.47	52.23	**54.77**
ChestX	1	21.83	**22.27**	21.76	21.62	21.91	<u>21.95</u>	21.36
ChestX	5	24.90	25.09	**25.40**	25.17	24.47	<u>25.31</u>	25.25
Equal-setting mean		**61.35**	60.95	61.09	60.53	53.83	60.28	<u>61.27</u>

Table 8: DINOv3 Macro-F1 (%) ↑ across all domains and support sizes. Bold/underline: best/second-best displayed means, including ties.

Dataset	Shot	LoRA	FacT	BOFT	FourierFT	RandLoRA	ReFTA	CenWrite
EuroSAT	1	71.91	**71.97**	71.48	71.22	54.70	68.93	71.30
EuroSAT	5	**91.64**	91.37	91.58	90.20	68.07	90.02	91.19
PlantVillage	1	90.38	87.19	87.93	87.08	77.88	88.65	**90.63**
PlantVillage	5	**97.55**	96.83	97.43	97.09	93.94	97.08	**97.55**
ISIC	1	**33.22**	32.70	33.21	32.46	28.22	32.67	32.78
ISIC	5	53.51	53.65	53.68	53.23	47.42	51.36	**54.04**
ChestX	1	18.93	**19.41**	19.01	18.82	17.85	18.88	17.72
ChestX	5	23.47	23.34	23.86	23.19	20.80	23.77	**24.06**
Equal-setting mean		**60.08**	59.56	59.77	59.16	51.11	58.92	59.91

Table 9: DINOv3 NLL ↓ across all domains and support sizes. Bold/underline: best/second-best displayed means, including ties.

Dataset	Shot	LoRA	FacT	BOFT	FourierFT	RandLoRA	ReFTA	CenWrite
EuroSAT	1	**0.7414**	0.9416	0.7931	0.7865	2.7299	0.8335	0.7587
EuroSAT	5	**0.2451**	0.3631	0.2796	0.3099	1.2935	0.2886	0.2562
PlantVillage	1	0.2659	0.4057	0.3276	0.3626	1.3108	0.3134	**0.2603**
PlantVillage	5	0.0738	0.1156	0.0797	0.0885	0.3307	0.0876	**0.0726**
ISIC	1	1.9810	2.9349	2.2024	2.2353	5.6700	1.9722	**1.9362**
ISIC	5	1.5065	2.2838	1.8052	1.8187	3.3959	1.5357	**1.3787**
ChestX	1	2.7546	3.7548	3.1159	3.2897	6.1366	2.7729	**2.5977**
ChestX	5	2.7960	3.8672	3.5367	3.4289	4.7654	2.4336	**2.3651**
Equal-setting mean		1.2955	1.8333	1.5175	1.5400	3.2041	1.2797	**1.2032**

Table 10: DINOv3 ECE (%) ↓ across all domains and support sizes. Bold/underline: best/second-best displayed means, including ties.

Dataset	Shot	LoRA	FacT	BOFT	FourierFT	RandLoRA	ReFTA	CenWrite
EuroSAT	1	13.88	18.42	14.99	15.25	34.41	14.94	**13.80**
EuroSAT	5	**6.70**	7.43	6.79	7.42	19.21	7.22	7.27
PlantVillage	1	**7.79**	9.43	8.44	8.93	19.28	8.73	7.82
PlantVillage	5	2.76	2.87	**2.41**	2.83	5.31	3.31	2.78
ISIC	1	33.32	46.75	38.20	39.49	62.63	33.33	**31.78**
ISIC	5	25.90	35.11	31.36	31.68	43.53	25.83	**22.12**
ChestX	1	46.58	56.87	52.53	54.62	69.20	47.06	**42.95**
ChestX	5	45.37	55.67	55.05	54.11	62.64	38.97	**36.17**
Equal-setting mean		22.79	29.07	26.22	26.79	39.53	22.43	**20.59**

Table 11: DINOv3 Brier ↓ across all domains and support sizes. Bold/underline: best/second-best displayed means, including ties.

Dataset	Shot	LoRA	FacT	BOFT	FourierFT	RandLoRA	ReFTA	CenWrite
EuroSAT	1	**0.3800**	0.4178	0.3926	0.3951	0.7269	0.4201	<u>0.3878</u>
EuroSAT	5	**0.1259**	0.1429	0.1316	0.1498	0.4534	0.1473	<u>0.1314</u>
PlantVillage	1	<u>0.1349</u>	0.1891	0.1658	0.1816	0.3820	0.1581	**0.1327**
PlantVillage	5	**0.0366**	0.0501	0.0385	0.0429	0.1037	0.0435	<u>0.0367</u>
ISIC	1	<u>0.8950</u>	1.0417	0.9430	0.9573	1.2698	0.8982	**0.8839**
ISIC	5	<u>0.6712</u>	0.7600	0.7219	0.7271	0.9030	0.6974	**0.6423**
ChestX	1	<u>1.1292</u>	1.2586	1.2006	1.2278	1.4274	1.1364	**1.0886**
ChestX	5	1.0933	1.2171	1.2110	1.2018	1.3191	<u>1.0317</u>	**1.0007**
Equal-setting mean		<u>0.5583</u>	0.6347	0.6006	0.6104	0.8232	0.5666	**0.5380**

Table 12: DINOv3 equal-setting effects with approximate 95% simultaneous max-$|T|$ intervals for the supplementary, post-hoc 30-endpoint family. Positive values favor CenWrite. Accuracy/F1/ECE use percentage points; NLL/Brier use native units.

Baseline	Accuracy (pp)	Macro-F1 (pp)	NLL	ECE (pp)	Brier
LoRA	−0.08 [−0.33, 0.17]	−0.17 [−0.43, 0.09]	+0.0923 [0.0846, 0.1001]	+2.20 [1.92, 2.48]	+0.0203 [0.0177, 0.0228]
FacT	+0.32 [−0.06, 0.69]	+0.35 [−0.05, 0.76]	+0.6302 [0.6063, 0.6540]	+8.48 [8.07, 8.89]	+0.0967 [0.0912, 0.1022]
BOFT	+0.18 [−0.18, 0.53]	+0.14 [−0.25, 0.52]	+0.3143 [0.2984, 0.3303]	+5.63 [5.26, 6.00]	+0.0626 [0.0577, 0.0676]
FourierFT	+0.74 [0.33, 1.14]	+0.75 [0.31, 1.19]	+0.3368 [0.3179, 0.3558]	+6.20 [5.78, 6.62]	+0.0724 [0.0665, 0.0783]
RandLoRA	+7.44 [6.61, 8.27]	+8.80 [7.84, 9.75]	+2.0009 [1.9367, 2.0651]	+18.94 [18.22, 19.66]	+0.2852 [0.2726, 0.2977]
ReFTA	+0.99 [0.63, 1.36]	+0.99 [0.60, 1.38]	+0.0765 [0.0618, 0.0912]	+1.84 [1.45, 2.23]	+0.0286 [0.0237, 0.0335]

Table 13: DINOv3 paired effects relative to LoRA, with unadjusted 95% pointwise intervals over 200 episodes per setting. Positive values favor CenWrite. Accuracy/F1/ECE use percentage points; NLL/Brier use native units.

Setting	NLL	Brier	ECE (pp)	Accuracy (pp)	Macro-F1 (pp)
EuroSAT 1-shot	−0.0174 [−0.0254, −0.0092]	−0.0078 [−0.0120, −0.0035]	+0.08 [−0.41, 0.56]	−0.61 [−1.10, −0.11]	−0.61 [−1.15, −0.06]
EuroSAT 5-shot	−0.0111 [−0.0172, −0.0050]	−0.0055 [−0.0087, −0.0022]	−0.57 [−0.87, −0.26]	−0.45 [−0.79, −0.11]	−0.45 [−0.79, −0.09]
PlantVillage 1-shot	+0.0056 [0.0019, 0.0093]	+0.0022 [0.0001, 0.0043]	−0.03 [−0.31, 0.25]	+0.25 [0.01, 0.49]	+0.25 [−0.03, 0.52]
PlantVillage 5-shot	+0.0013 [−0.0013, 0.0039]	−0.0001 [−0.0017, 0.0014]	−0.02 [−0.13, 0.10]	0.00 [−0.17, 0.17]	−0.01 [−0.19, 0.17]
ISIC 1-shot	+0.0448 [0.0308, 0.0587]	+0.0112 [0.0062, 0.0161]	+1.53 [0.93, 2.14]	−0.21 [−0.71, 0.29]	−0.44 [−0.99, 0.10]
ISIC 5-shot	+0.1277 [0.1131, 0.1424]	+0.0288 [0.0233, 0.0344]	+3.78 [3.20, 4.36]	+0.49 [−0.07, 1.06]	+0.53 [−0.05, 1.12]

Continued on the following page.

Table 13 continued

Setting	NLL	Brier	ECE (pp)	Accuracy (pp)	Macro-F1 (pp)
ChestX 1-shot	+0.1569 [0.1338, 0.1794]	+0.0406 [0.0342, 0.0469]	+3.62 [2.95, 4.30]	−0.47 [−0.96, 0.01]	−1.21 [−1.69, −0.72]
ChestX 5-shot	+0.4309 [0.4074, 0.4545]	+0.0926 [0.0859, 0.0992]	+9.19 [8.46, 9.91]	+0.35 [−0.27, 0.98]	+0.59 [−0.06, 1.23]

Table 14: DINOv3 paired effects relative to FacT, with unadjusted 95% pointwise intervals over 200 episodes per setting. Positive values favor CenWrite. Accuracy/F1/ECE use percentage points; NLL/Brier use native units.

Setting	NLL	Brier	ECE (pp)	Accuracy (pp)	Macro-F1 (pp)
EuroSAT 1-shot	+0.1828 [0.1428, 0.2241]	+0.0301 [0.0159, 0.0442]	+4.62 [3.69, 5.55]	−0.65 [−1.64, 0.34]	−0.66 [−1.80, 0.46]
EuroSAT 5-shot	+0.1069 [0.0850, 0.1300]	+0.0115 [0.0043, 0.0189]	+0.17 [−0.28, 0.63]	−0.17 [−0.70, 0.37]	−0.18 [−0.73, 0.38]
PlantVillage 1-shot	+0.1454 [0.1204, 0.1711]	+0.0564 [0.0468, 0.0662]	+1.61 [0.95, 2.28]	+3.18 [2.53, 3.85]	+3.44 [2.69, 4.21]
PlantVillage 5-shot	+0.0430 [0.0307, 0.0563]	+0.0134 [0.0092, 0.0176]	+0.09 [−0.15, 0.34]	+0.69 [0.39, 0.97]	+0.72 [0.41, 1.03]
ISIC 1-shot	+0.9987 [0.9424, 1.0558]	+0.1578 [0.1457, 0.1698]	+14.97 [14.02, 15.90]	+0.19 [−0.61, 1.00]	+0.08 [−0.76, 0.93]
ISIC 5-shot	+0.9051 [0.8523, 0.9587]	+0.1177 [0.1068, 0.1287]	+12.99 [12.18, 13.79]	+0.04 [−0.70, 0.80]	+0.39 [−0.37, 1.18]
ChestX 1-shot	+1.1571 [1.0974, 1.2174]	+0.1701 [0.1594, 0.1805]	+13.92 [13.07, 14.77]	−0.91 [−1.51, −0.31]	−1.69 [−2.35, −1.02]
ChestX 5-shot	+1.5021 [1.4491, 1.5554]	+0.2163 [0.2072, 0.2253]	+19.49 [18.67, 20.29]	+0.16 [−0.51, 0.82]	+0.72 [0.02, 1.41]

Table 15: DINOv3 paired effects relative to BOFT, with unadjusted 95% pointwise intervals over 200 episodes per setting. Positive values favor CenWrite. Accuracy/F1/ECE use percentage points; NLL/Brier use native units.

Setting	NLL	Brier	ECE (pp)	Accuracy (pp)	Macro-F1 (pp)
EuroSAT 1-shot	+0.0344 [0.0070, 0.0629]	+0.0049 [−0.0073, 0.0169]	+1.19 [0.44, 1.94]	−0.31 [−1.29, 0.67]	−0.17 [−1.25, 0.89]
EuroSAT 5-shot	+0.0234 [0.0084, 0.0391]	+0.0002 [−0.0068, 0.0074]	−0.48 [−0.87, −0.08]	−0.40 [−0.96, 0.17]	−0.39 [−0.98, 0.21]
PlantVillage 1-shot	+0.0673 [0.0539, 0.0810]	+0.0331 [0.0261, 0.0401]	+0.62 [0.22, 1.03]	+2.53 [1.94, 3.11]	+2.70 [2.05, 3.35]
PlantVillage 5-shot	+0.0071 [0.0000, 0.0145]	+0.0018 [−0.0015, 0.0050]	−0.37 [−0.56, −0.16]	+0.09 [−0.17, 0.35]	+0.11 [−0.16, 0.38]
ISIC 1-shot	+0.2663 [0.2361, 0.2970]	+0.0592 [0.0491, 0.0692]	+6.41 [5.56, 7.25]	−0.13 [−0.85, 0.59]	−0.43 [−1.21, 0.34]
ISIC 5-shot	+0.4264 [0.3963, 0.4569]	+0.0796 [0.0701, 0.0892]	+9.24 [8.52, 9.96]	+0.17 [−0.47, 0.79]	+0.36 [−0.30, 1.00]
ChestX 1-shot	+0.5182 [0.4741, 0.5622]	+0.1120 [0.1009, 0.1228]	+9.58 [8.59, 10.53]	−0.40 [−1.01, 0.22]	−1.28 [−1.91, −0.67]
ChestX 5-shot	+1.1716 [1.1288, 1.2145]	+0.2103 [0.2009, 0.2199]	+18.88 [18.10, 19.67]	−0.15 [−0.81, 0.53]	+0.20 [−0.51, 0.91]

Table 16: DINOv3 paired effects relative to FourierFT, with unadjusted 95% pointwise intervals over 200 episodes per setting. Positive values favor CenWrite. Accuracy/F1/ECE use percentage points; NLL/Brier use native units.

Setting	NLL	Brier	ECE (pp)	Accuracy (pp)	Macro-F1 (pp)
EuroSAT 1-shot	+0.0278 [−0.0054, 0.0615]	+0.0074 [−0.0075, 0.0221]	+1.45 [0.63, 2.28]	+0.02 [−1.07, 1.11]	+0.08 [−1.17, 1.31]
EuroSAT 5-shot	+0.0537 [0.0349, 0.0731]	+0.0184 [0.0095, 0.0274]	+0.16 [−0.30, 0.63]	+0.97 [0.33, 1.60]	+0.99 [0.33, 1.65]
PlantVillage 1-shot	+0.1023 [0.0853, 0.1202]	+0.0489 [0.0400, 0.0580]	+1.10 [0.68, 1.52]	+3.41 [2.73, 4.11]	+3.54 [2.82, 4.29]
PlantVillage 5-shot	+0.0160 [0.0082, 0.0243]	+0.0061 [0.0026, 0.0098]	+0.05 [−0.16, 0.27]	+0.43 [0.13, 0.72]	+0.46 [0.15, 0.76]
ISIC 1-shot	+0.2991 [0.2627, 0.3357]	+0.0734 [0.0613, 0.0857]	+7.70 [6.72, 8.68]	+0.59 [−0.27, 1.45]	+0.32 [−0.63, 1.26]
ISIC 5-shot	+0.4400 [0.4047, 0.4765]	+0.0848 [0.0738, 0.0959]	+9.56 [8.73, 10.39]	+0.64 [−0.15, 1.43]	+0.81 [−0.01, 1.61]
ChestX 1-shot	+0.6921 [0.6391, 0.7443]	+0.1392 [0.1264, 0.1517]	+11.66 [10.60, 12.70]	−0.26 [−0.94, 0.41]	−1.09 [−1.78, −0.43]
ChestX 5-shot	+1.0638 [1.0154, 1.1132]	+0.2011 [0.1901, 0.2122]	+17.94 [17.01, 18.85]	+0.08 [−0.65, 0.81]	+0.86 [0.09, 1.63]

Table 17: DINOv3 paired effects relative to RandLoRA, with unadjusted 95% pointwise intervals over 200 episodes per setting. Positive values favor CenWrite. Accuracy/F1/ECE use percentage points; NLL/Brier use native units.

Setting	NLL	Brier	ECE (pp)	Accuracy (pp)	Macro-F1 (pp)
EuroSAT 1-shot	+1.9712 [1.8303, 2.1151]	+0.3391 [0.3086, 0.3697]	+20.61 [18.98, 22.31]	+14.32 [12.17, 16.50]	+16.61 [14.08, 19.20]
EuroSAT 5-shot	+1.0373 [0.9341, 1.1482]	+0.3220 [0.2876, 0.3578]	+11.94 [10.45, 13.49]	+20.42 [17.63, 23.35]	+23.12 [19.92, 26.47]
PlantVillage 1-shot	+1.0505 [0.9508, 1.1513]	+0.2492 [0.2267, 0.2721]	+11.46 [10.23, 12.67]	+12.26 [11.04, 13.50]	+12.75 [11.44, 14.08]
PlantVillage 5-shot	+0.2581 [0.2133, 0.3073]	+0.0670 [0.0535, 0.0818]	+2.53 [1.92, 3.20]	+3.49 [2.67, 4.43]	+3.61 [2.75, 4.58]
ISIC 1-shot	+3.7338 [3.6123, 3.8534]	+0.3859 [0.3689, 0.4031]	+30.84 [29.76, 31.95]	+3.49 [2.55, 4.45]	+4.55 [3.45, 5.66]
ISIC 5-shot	+2.0172 [1.9045, 2.1286]	+0.2607 [0.2408, 0.2808]	+21.41 [20.11, 22.70]	+5.31 [4.02, 6.65]	+6.62 [5.17, 8.16]
ChestX 1-shot	+3.5389 [3.3861, 3.6837]	+0.3388 [0.3188, 0.3575]	+26.25 [24.60, 27.78]	−0.55 [−1.18, 0.06]	−0.13 [−0.86, 0.60]
ChestX 5-shot	+2.4003 [2.2689, 2.5300]	+0.3184 [0.3022, 0.3344]	+26.46 [25.25, 27.68]	+0.78 [0.00, 1.57]	+3.25 [2.40, 4.10]

Table 18: DINOv3 paired effects relative to ReFTA, with unadjusted 95% pointwise intervals over 200 episodes per setting. Positive values favor CenWrite. Accuracy/F1/ECE use percentage points; NLL/Brier use native units.

Setting	NLL	Brier	ECE (pp)	Accuracy (pp)	Macro-F1 (pp)
EuroSAT 1-shot	+0.0748 [0.0476, 0.1028]	+0.0324 [0.0201, 0.0451]	+1.15 [0.43, 1.89]	+2.16 [1.17, 3.16]	+2.37 [1.29, 3.46]
EuroSAT 5-shot	+0.0324 [0.0188, 0.0462]	+0.0159 [0.0090, 0.0229]	−0.05 [−0.40, 0.30]	+1.13 [0.57, 1.71]	+1.18 [0.59, 1.77]

Continued on the following page.

Table 18 continued

Setting	NLL	Brier	ECE (pp)	Accuracy (pp)	Macro-F1 (pp)
PlantVillage 1-shot	+0.0531 [0.0434, 0.0630]	+0.0254 [0.0202, 0.0306]	+0.91 [0.58, 1.24]	+1.91 [1.38, 2.45]	+1.98 [1.40, 2.57]
PlantVillage 5-shot	+0.0151 [0.0099, 0.0209]	+0.0068 [0.0041, 0.0095]	+0.54 [0.36, 0.72]	+0.45 [0.24, 0.67]	+0.47 [0.25, 0.70]
ISIC 1-shot	+0.0360 [0.0109, 0.0616]	+0.0144 [0.0055, 0.0233]	+1.55 [0.72, 2.38]	+0.39 [−0.37, 1.15]	+0.11 [−0.70, 0.91]
ISIC 5-shot	+0.1569 [0.1336, 0.1806]	+0.0551 [0.0463, 0.0640]	+3.72 [2.98, 4.45]	+2.55 [1.81, 3.27]	+2.68 [1.92, 3.43]
ChestX 1-shot	+0.1752 [0.1311, 0.2200]	+0.0479 [0.0356, 0.0600]	+4.11 [3.04, 5.17]	−0.59 [−1.23, 0.05]	−1.16 [−1.83, −0.48]
ChestX 5-shot	+0.0685 [0.0282, 0.1083]	+0.0310 [0.0206, 0.0412]	+2.79 [1.77, 3.78]	−0.06 [−0.70, 0.59]	+0.29 [−0.39, 0.98]

C Component Ablation Results

The six implementations share the same support/query tasks and 50-step adaptation schedule. Complete results include the five aggregate means, all 50 prespecified effects and interactions, 48 setting–variant means, and 400 setting-specific effects. Aggregate effects use simultaneous intervals; setting-level intervals are pointwise. Both macro-F1 interactions have positive simultaneous intervals, supporting a larger residual benefit under the complete update rule. The accompanying tables provide every component contrast under the same paired protocol.

a Variants and mean performance

Variant	Center	Comp.	b	NLL ↓	Brier ↓	Acc. ↑ (%)
L LoRA	✗	✗	✗	1.296	0.558	61.35
NC No centering	✗	✗	✓	1.299	0.559	61.36
P Pure centering	✓	✗	✗	1.290	0.553	**61.65**
NP No compensation	✓	✗	✓	1.297	0.555	**61.65**
NB No residual	✓	✓	✗	**1.201**	**0.537**	61.17
F CenWrite	✓	✓	✓	1.203	0.538	61.27

b Paired component effects

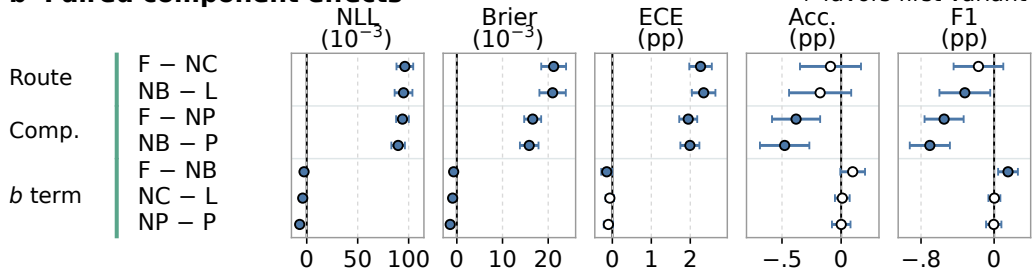


Figure 5: Component definitions, eight-setting mean scores, and paired effects. In (a), checkmarks and crosses indicate enabled and disabled operations; gray identifies full CenWrite. Best means are bold and second-best means underlined. In (b), positive effects favor the first variant. Bars are simultaneous 95% intervals from the 50-endpoint family; filled points exclude zero. NLL/Brier effects use 10^{-3} units and the other effects use percentage points. Every setting contains 200 matched episodes.

Table 19: Component means over eight equally weighted settings. Accuracy/F1/ECE are percentages; NLL/Brier use native units. Variants follow Appendix A.4. Bold/underline: best/second-best displayed means, including ties.

Variant	Acc.	Macro-F1	NLL	ECE	Brier
L	61.35	<u>60.08</u>	1.296	22.79	0.558
NC	<u>61.36</u>	<u>60.08</u>	1.299	22.85	0.559
P	**61.65**	**60.46**	1.290	22.43	0.553
NP	**61.65**	**60.46**	1.297	22.53	0.555
NB	61.17	59.76	**1.201**	**20.44**	**0.537**
F	61.27	59.91	<u>1.203</u>	<u>20.59</u>	<u>0.538</u>

Table 20: Aggregate component effects: estimate, pointwise 95% interval, then simultaneous 95% interval for the prespecified 50-endpoint family. Positive direct effects favor the first variant; interactions follow Equation 14.

Contrast	Acc. (pp)	F1 (pp)	NLL (10^{-3})	ECE (pp)	Brier (10^{-3})
$F-L$	−0.080 [−0.242,0.078] [−0.335,0.175]	−0.168 [−0.338,−0.001] [−0.437,0.100]	+92.33 [87.34,97.30] [84.33,100.34]	+2.199 [2.016,2.377] [1.910,2.488]	+20.27 [18.60,21.91] [17.62,22.92]
$F-NC$	−0.090 [−0.253,0.072] [−0.348,0.168]	−0.172 [−0.343,−0.001] [−0.444,0.100]	+96.28 [91.15,101.38] [88.07,104.49]	+2.264 [2.081,2.443] [1.975,2.553]	+21.19 [19.48,22.87] [18.47,23.91]
$NB-L$	−0.178 [−0.343,−0.013] [−0.440,0.085]	−0.320 [−0.497,−0.147] [−0.598,−0.043]	+94.98 [89.48,100.43] [86.24,103.73]	+2.343 [2.153,2.532] [2.041,2.645]	+20.96 [19.13,22.76] [18.07,23.86]
$F-NP$	−0.381 [−0.508,−0.253] [−0.584,−0.177]	−0.547 [−0.681,−0.413] [−0.761,−0.333]	+93.95 [90.06,97.78] [87.71,100.19]	+1.947 [1.801,2.093] [1.717,2.178]	+16.60 [15.44,17.77] [14.74,18.46]
$NB-P$	−0.478 [−0.608,−0.347] [−0.686,−0.269]	−0.704 [−0.842,−0.567] [−0.923,−0.484]	+89.69 [85.58,93.78] [83.05,96.33]	+1.990 [1.836,2.142] [1.746,2.234]	+15.84 [14.57,17.12] [13.81,17.88]
$F-NB$	+0.098 [0.032,0.164] [−0.007,0.202]	+0.152 [0.083,0.220] [0.044,0.260]	−2.65 [−3.61,−1.68] [−4.20,−1.10]	−0.144 [−0.234,−0.054] [−0.285,−0.003]	−0.70 [−1.02,−0.37] [−1.21,−0.18]
$NC-L$	+0.010 [−0.030,0.049] [−0.052,0.072]	+0.003 [−0.039,0.044] [−0.062,0.068]	−3.95 [−4.37,−3.53] [−4.61,−3.28]	−0.065 [−0.132,0.002] [−0.171,0.042]	−0.92 [−1.06,−0.78] [−1.14,−0.70]
$NP-P$	+0.001 [−0.048,0.050] [−0.078,0.079]	−0.004 [−0.057,0.049] [−0.088,0.080]	−6.91 [−7.55,−6.28] [−7.93,−5.89]	−0.102 [−0.176,−0.028] [−0.220,0.017]	−1.45 [−1.66,−1.25] [−1.77,−1.13]
I_1	+0.087 [0.011,0.165] [−0.034,0.209]	+0.149 [0.071,0.228] [0.023,0.275]	+1.30 [0.28,2.32] [−0.34,2.93]	−0.079 [−0.190,0.034] [−0.257,0.098]	+0.23 [−0.11,0.57] [−0.31,0.76]
I_2	+0.097 [0.017,0.178] [−0.032,0.225]	+0.156 [0.072,0.240] [0.023,0.290]	+4.26 [3.40,5.11] [2.88,5.63]	−0.042 [−0.160,0.076] [−0.228,0.144]	+0.76 [0.48,1.04] [0.31,1.20]

Table 21: Component means by domain and shot, with 200 matched episodes per setting. Accuracy/F1/ECE are percentages; NLL/Brier use native units. Bold/underline: best/second-best displayed means, including ties.

Setting	Variant	Acc.	Macro-F1	NLL	ECE	Brier
EuroSAT 1-shot	L	73.41	71.91	0.741	13.88	0.380
	NC	73.43	71.93	0.740	13.85	0.380
	P	**73.85**	**72.47**	**0.729**	13.85	**0.374**
	NP	73.84	72.46	**0.729**	**13.65**	**0.374**
	NB	72.69	71.18	0.765	13.91	0.390
	F	72.81	71.30	0.759	13.80	0.388
EuroSAT 5-shot	L	**91.73**	**91.64**	0.245	6.70	0.126
	NC	91.71	91.61	0.246	**6.57**	0.126
	P	91.58	91.47	0.244	6.73	**0.125**
	NP	91.62	91.51	**0.243**	6.70	**0.125**
	NB	91.13	91.04	0.259	7.47	0.133
	F	91.28	91.19	0.256	7.27	0.131
PlantVillage 1-shot	L	90.85	90.38	0.266	7.79	0.135
	NC	90.88	90.40	0.267	7.73	0.135
	P	**91.17**	**90.73**	**0.255**	**7.63**	**0.130**
	NP	91.15	90.69	**0.255**	7.67	**0.130**
	NB	91.03	90.53	0.261	7.82	0.133
	F	91.11	90.63	0.260	7.82	0.133
PlantVillage 5-shot	L	97.59	97.55	0.074	2.76	0.037
	NC	97.57	97.54	0.074	2.75	0.037
	P	**97.67**	97.63	**0.071**	2.74	**0.036**
	NP	**97.67**	**97.64**	**0.071**	**2.73**	**0.036**
	NB	97.57	97.53	0.073	2.86	0.037
	F	97.59	97.55	0.073	2.78	0.037
ISIC 2018 1-shot	L	36.18	33.22	1.981	33.32	0.895
	NC	36.17	33.20	1.987	33.50	0.897
	P	**36.45**	33.56	1.981	33.05	0.892
	NP	**36.45**	**33.57**	1.988	33.28	0.894
	NB	35.86	32.61	**1.933**	**31.43**	**0.883**
	F	35.97	32.78	1.936	31.78	0.884
ISIC 2018 5-shot	L	54.28	53.51	1.506	25.90	0.671
	NC	54.27	53.48	1.512	26.08	0.672
	P	55.15	54.45	1.471	24.92	0.659
	NP	**55.23**	**54.54**	1.481	25.22	0.661
	NB	54.90	54.16	**1.365**	**21.75**	**0.639**
	F	54.77	54.04	1.379	22.12	0.642
ChestX 1-shot	L	21.83	18.93	2.755	46.58	1.129
	NC	**21.85**	18.93	2.762	46.88	1.131
	P	21.83	**19.07**	2.800	46.81	1.132
	NP	21.84	19.04	2.810	46.86	1.134
	NB	21.33	17.42	2.602	**42.82**	**1.088**
	F	21.36	17.72	**2.598**	42.95	1.089
ChestX 5-shot	L	24.90	23.47	2.796	45.37	1.093
	NC	24.98	23.55	2.808	45.44	1.096
	P	**25.47**	**24.29**	2.772	43.72	1.079
	NP	25.39	24.20	2.800	44.17	1.084
	NB	24.87	23.57	**2.347**	**35.47**	**0.995**
	F	25.25	24.06	2.365	36.17	1.001

Table 22: EuroSAT, 1-shot component effects with unadjusted 95% pointwise intervals (200 paired episodes). Positive direct effects favor the first variant; interaction signs follow Equation 14.

Contrast	Acc. (pp)	F1 (pp)	NLL (10^{-3})	ECE (pp)	Brier (10^{-3})
$F - L$	−0.607 [−1.100,−0.107]	−0.610 [−1.148,−0.061]	−17.35 [−25.37,−9.18]	+0.078 [−0.411,0.560]	−7.76 [−12.04,−3.47]
$F - NC$	−0.627 [−1.147,−0.107]	−0.626 [−1.204,−0.044]	−18.32 [−26.86,−9.62]	+0.050 [−0.443,0.544]	−8.25 [−12.83,−3.68]
$NB - L$	−0.727 [−1.260,−0.180]	−0.734 [−1.317,−0.138]	−23.25 [−32.67,−13.76]	−0.033 [−0.547,0.480]	−10.43 [−15.45,−5.42]
$F - NP$	−1.033 [−1.393,−0.667]	−1.154 [−1.574,−0.735]	−29.90 [−35.16,−24.66]	−0.145 [−0.585,0.299]	−14.10 [−16.92,−11.26]
$NB - P$	−1.167 [−1.560,−0.780]	−1.293 [−1.719,−0.865]	−35.99 [−42.06,−29.88]	−0.059 [−0.522,0.396]	−16.68 [−19.88,−13.47]
$F - NB$	+0.120 [−0.060,0.300]	+0.124 [−0.066,0.316]	+5.90 [3.76,8.06]	+0.111 [−0.211,0.426]	+2.66 [1.60,3.75]
$NC - L$	+0.020 [−0.100,0.140]	+0.017 [−0.118,0.152]	+0.96 [−0.04,1.98]	+0.028 [−0.242,0.298]	+0.48 [0.00,0.97]
$NP - P$	−0.013 [−0.153,0.127]	−0.015 [−0.171,0.138]	−0.19 [−1.66,1.23]	+0.197 [−0.089,0.486]	+0.08 [−0.63,0.79]
I_1	+0.100 [−0.113,0.313]	+0.108 [−0.123,0.337]	+4.94 [2.77,7.13]	+0.083 [−0.327,0.494]	+2.18 [1.12,3.27]
I_2	+0.133 [−0.087,0.353]	+0.139 [−0.092,0.373]	+6.09 [4.61,7.62]	−0.086 [−0.548,0.366]	+2.58 [1.86,3.32]

Table 23: EuroSAT, 5-shot component effects with unadjusted 95% pointwise intervals (200 paired episodes). Positive direct effects favor the first variant; interaction signs follow Equation 14.

Contrast	Acc. (pp)	F1 (pp)	NLL (10^{-3})	ECE (pp)	Brier (10^{-3})
$F - L$	−0.453 [−0.787,−0.113]	−0.452 [−0.793,−0.095]	−11.14 [−17.17,−4.99]	−0.566 [−0.867,−0.263]	−5.46 [−8.74,−2.15]
$F - NC$	−0.427 [−0.767,−0.080]	−0.423 [−0.772,−0.062]	−10.13 [−16.32,−3.80]	−0.693 [−1.001,−0.387]	−5.12 [−8.43,−1.72]
$NB - L$	−0.607 [−0.947,−0.253]	−0.607 [−0.962,−0.247]	−13.93 [−20.58,−7.12]	−0.773 [−1.059,−0.483]	−6.79 [−10.34,−3.25]
$F - NP$	−0.340 [−0.613,−0.060]	−0.321 [−0.607,−0.030]	−12.91 [−17.31,−8.52]	−0.568 [−0.861,−0.269]	−6.78 [−8.97,−4.63]
$NB - P$	−0.453 [−0.747,−0.160]	−0.437 [−0.734,−0.136]	−14.92 [−19.58,−10.25]	−0.744 [−1.004,−0.480]	−7.65 [−9.96,−5.32]
$F - NB$	+0.153 [0.007,0.307]	+0.155 [0.004,0.310]	+2.79 [1.62,3.97]	+0.207 [−0.014,0.425]	+1.33 [0.70,1.94]
$NC - L$	−0.027 [−0.113,0.060]	−0.029 [−0.115,0.056]	−1.01 [−1.59,−0.47]	+0.128 [−0.017,0.276]	−0.34 [−0.62,−0.07]
$NP - P$	+0.040 [−0.067,0.140]	+0.038 [−0.068,0.144]	+0.79 [0.04,1.53]	+0.031 [−0.144,0.206]	+0.47 [0.07,0.86]
I_1	+0.180 [0.013,0.353]	+0.184 [0.016,0.359]	+3.81 [2.71,4.90]	+0.079 [−0.175,0.330]	+1.67 [1.06,2.27]
I_2	+0.113 [−0.073,0.300]	+0.117 [−0.070,0.309]	+2.00 [1.11,2.88]	+0.176 [−0.111,0.457]	+0.86 [0.34,1.38]

Table 24: PlantVillage, 1-shot component effects with unadjusted 95% pointwise intervals (200 paired episodes). Positive direct effects favor the first variant; interaction signs follow Equation 14.

Contrast	Acc. (pp)	F1 (pp)	NLL (10^{-3})	ECE (pp)	Brier (10^{-3})
$F - L$	+0.253 [0.007,0.493]	+0.251 [−0.031,0.521]	+5.57 [1.85,9.30]	−0.032 [−0.310,0.251]	+2.19 [0.06,4.28]
$F - NC$	+0.227 [−0.020,0.467]	+0.234 [−0.047,0.502]	+6.36 [2.45,10.29]	−0.093 [−0.380,0.191]	+2.58 [0.34,4.79]
$NB - L$	+0.173 [−0.087,0.427]	+0.155 [−0.138,0.438]	+4.56 [0.50,8.63]	−0.025 [−0.289,0.249]	+1.73 [−0.56,4.00]
$F - NP$	−0.040 [−0.193,0.107]	−0.063 [−0.240,0.107]	−5.54 [−7.18,−4.00]	−0.155 [−0.368,0.056]	−2.51 [−3.48,−1.61]
$NB - P$	−0.147 [−0.307,0.013]	−0.198 [−0.391,−0.017]	−6.61 [−8.38,−4.94]	−0.183 [−0.390,0.026]	−3.00 [−4.04,−2.05]
$F - NB$	+0.080 [0.000,0.167]	+0.096 [0.010,0.187]	+1.00 [0.49,1.53]	−0.008 [−0.150,0.133]	+0.45 [0.16,0.74]
$NC - L$	+0.027 [−0.040,0.093]	+0.017 [−0.053,0.089]	−0.79 [−1.17,−0.42]	+0.061 [−0.078,0.197]	−0.39 [−0.61,−0.18]
$NP - P$	−0.027 [−0.087,0.033]	−0.040 [−0.106,0.023]	−0.06 [−0.50,0.37]	−0.035 [−0.179,0.107]	−0.04 [−0.29,0.21]
I_1	+0.053 [−0.053,0.160]	+0.079 [−0.032,0.191]	+1.80 [1.28,2.30]	−0.069 [−0.272,0.130]	+0.85 [0.57,1.13]
I_2	+0.107 [0.007,0.213]	+0.135 [0.029,0.249]	+1.07 [0.79,1.34]	+0.028 [−0.182,0.240]	+0.49 [0.35,0.64]

Table 25: PlantVillage, 5-shot component effects with unadjusted 95% pointwise intervals (200 paired episodes). Positive direct effects favor the first variant; interaction signs follow Equation 14.

Contrast	Acc. (pp)	F1 (pp)	NLL (10^{-3})	ECE (pp)	Brier (10^{-3})
$F - L$	+0.000 [−0.173,0.167]	−0.005 [−0.188,0.169]	+1.28 [−1.34,3.95]	−0.017 [−0.128,0.096]	−0.09 [−1.67,1.42]
$F - NC$	+0.013 [−0.160,0.187]	+0.011 [−0.174,0.190]	+1.28 [−1.43,4.04]	−0.025 [−0.142,0.094]	−0.07 [−1.69,1.48]
$NB - L$	−0.020 [−0.207,0.153]	−0.024 [−0.215,0.156]	+1.15 [−1.64,3.98]	−0.101 [−0.218,0.015]	−0.13 [−1.80,1.47]
$F - NP$	−0.087 [−0.180,0.000]	−0.089 [−0.191,0.000]	−1.68 [−2.80,−0.57]	−0.048 [−0.149,0.053]	−0.84 [−1.50,−0.24]
$NB - P$	−0.100 [−0.193,−0.020]	−0.102 [−0.203,−0.014]	−2.03 [−3.19,−0.88]	−0.117 [−0.230,−0.003]	−0.95 [−1.63,−0.32]
$F - NB$	+0.020 [−0.027,0.073]	+0.019 [−0.029,0.073]	+0.13 [−0.15,0.41]	+0.084 [0.006,0.166]	+0.04 [−0.12,0.20]
$NC - L$	−0.013 [−0.040,0.013]	−0.016 [−0.045,0.010]	+0.00 [−0.17,0.17]	+0.008 [−0.054,0.066]	−0.02 [−0.11,0.08]
$NP - P$	+0.007 [−0.033,0.053]	+0.007 [−0.038,0.053]	−0.23 [−0.49,0.03]	+0.015 [−0.048,0.077]	−0.07 [−0.21,0.08]
I_1	+0.033 [−0.020,0.093]	+0.035 [−0.022,0.097]	+0.13 [−0.13,0.38]	+0.076 [−0.023,0.181]	+0.06 [−0.11,0.22]
I_2	+0.013 [−0.047,0.073]	+0.012 [−0.048,0.073]	+0.35 [0.21,0.50]	+0.069 [−0.026,0.169]	+0.11 [0.02,0.20]

Table 26: ISIC 2018, 1-shot component effects with unadjusted 95% pointwise intervals (200 paired episodes). Positive direct effects favor the first variant; interaction signs follow Equation 14.

Contrast	Acc. (pp)	F1 (pp)	NLL (10^{-3})	ECE (pp)	Brier (10^{-3})
$F - L$	−0.207 [−0.707,0.287]	−0.442 [−0.988,0.096]	+44.82 [30.79,58.67]	+1.534 [0.925,2.139]	+11.17 [6.22,16.13]
$F - NC$	−0.193 [−0.693,0.287]	−0.421 [−0.961,0.107]	+50.88 [36.36,65.20]	+1.719 [1.133,2.304]	+12.69 [7.60,17.80]
$NB - L$	−0.320 [−0.833,0.180]	−0.604 [−1.168,−0.046]	+48.27 [32.87,63.35]	+1.882 [1.258,2.495]	+11.66 [6.19,17.15]
$F - NP$	−0.473 [−0.847,−0.107]	−0.791 [−1.196,−0.385]	+51.86 [40.88,62.84]	+1.498 [1.034,1.956]	+9.75 [5.98,13.58]
$NB - P$	−0.593 [−0.987,−0.207]	−0.949 [−1.382,−0.532]	+48.08 [36.25,59.89]	+1.617 [1.135,2.096]	+8.43 [4.30,12.64]
$F - NB$	+0.113 [−0.080,0.307]	+0.162 [−0.043,0.366]	−3.45 [−5.63,−1.28]	−0.349 [−0.614,−0.082]	−0.49 [−1.34,0.35]
$NC - L$	−0.013 [−0.127,0.100]	−0.021 [−0.137,0.093]	−6.06 [−7.02,−5.09]	−0.186 [−0.400,0.029]	−1.52 [−1.88,−1.16]
$NP - P$	−0.007 [−0.160,0.153]	+0.004 [−0.170,0.184]	−7.23 [−8.64,−5.84]	−0.230 [−0.461,−0.006]	−1.80 [−2.32,−1.30]
I_1	+0.127 [−0.080,0.333]	+0.183 [−0.038,0.404]	+2.61 [0.52,4.69]	−0.163 [−0.502,0.181]	+1.03 [0.20,1.84]
I_2	+0.120 [−0.120,0.360]	+0.159 [−0.099,0.420]	+3.78 [2.06,5.49]	−0.119 [−0.457,0.222]	+1.31 [0.65,1.98]

Table 27: ISIC 2018, 5-shot component effects with unadjusted 95% pointwise intervals (200 paired episodes). Positive direct effects favor the first variant; interaction signs follow Equation 14.

Contrast	Acc. (pp)	F1 (pp)	NLL (10^{-3})	ECE (pp)	Brier (10^{-3})
$F - L$	+0.493 [−0.067,1.060]	+0.529 [−0.054,1.118]	+127.71 [113.09,142.43]	+3.779 [3.203,4.364]	+28.85 [23.31,34.42]
$F - NC$	+0.507 [−0.060,1.080]	+0.557 [−0.034,1.153]	+133.11 [118.07,148.25]	+3.959 [3.390,4.539]	+30.03 [24.34,35.73]
$NB - L$	+0.620 [0.027,1.220]	+0.650 [0.038,1.264]	+141.49 [125.66,157.51]	+4.144 [3.551,4.747]	+32.21 [26.38,38.07]
$F - NP$	−0.460 [−0.913,−0.007]	−0.501 [−0.975,−0.023]	+102.39 [92.44,112.48]	+3.100 [2.695,3.513]	+18.51 [15.01,22.05]
$NB - P$	−0.247 [−0.713,0.227]	−0.287 [−0.774,0.198]	+105.80 [95.20,116.47]	+3.169 [2.723,3.626]	+19.60 [15.84,23.40]
$F - NB$	−0.127 [−0.327,0.073]	−0.121 [−0.332,0.094]	−13.78 [−15.89,−11.66]	−0.365 [−0.670,−0.059]	−3.36 [−4.19,−2.54]
$NC - L$	−0.013 [−0.147,0.113]	−0.028 [−0.160,0.101]	−5.40 [−6.33,−4.46]	−0.180 [−0.397,0.034]	−1.18 [−1.58,−0.77]
$NP - P$	+0.087 [−0.087,0.253]	+0.093 [−0.084,0.266]	−10.37 [−11.84,−8.92]	−0.296 [−0.536,−0.052]	−2.27 [−2.84,−1.71]
I_1	−0.113 [−0.353,0.127]	−0.093 [−0.341,0.157]	−8.38 [−10.54,−6.17]	−0.185 [−0.564,0.203]	−2.18 [−3.09,−1.28]
I_2	−0.213 [−0.460,0.040]	−0.214 [−0.476,0.050]	−3.41 [−5.17,−1.63]	−0.069 [−0.442,0.311]	−1.09 [−1.80,−0.36]

Table 28: ChestX, 1-shot component effects with unadjusted 95% pointwise intervals (200 paired episodes). Positive direct effects favor the first variant; interaction signs follow Equation 14.

Contrast	Acc. (pp)	F1 (pp)	NLL (10^{-3})	ECE (pp)	Brier (10^{-3})
$F-L$	−0.473 [−0.960,0.013]	−1.206 [−1.692,−0.716]	+156.92 [133.77,179.38]	+3.625 [2.948,4.298]	+40.63 [34.16,46.86]
$F-NC$	−0.493 [−0.980,−0.007]	−1.210 [−1.698,−0.727]	+163.91 [139.95,186.99]	+3.924 [3.250,4.582]	+42.51 [35.92,48.85]
$NB-L$	−0.507 [−1.034,0.020]	−1.505 [−2.049,−0.966]	+152.27 [126.32,177.29]	+3.756 [2.991,4.508]	+41.27 [34.13,48.26]
$F-NP$	−0.480 [−0.953,−0.013]	−1.313 [−1.792,−0.845]	+212.07 [191.68,232.10]	+3.903 [3.292,4.511]	+45.74 [40.28,51.14]
$NB-P$	−0.507 [−1.000,−0.020]	−1.644 [−2.154,−1.142]	+197.42 [175.44,219.09]	+3.988 [3.303,4.650]	+43.63 [37.57,49.61]
$F-NB$	+0.033 [−0.207,0.273]	+0.299 [0.058,0.542]	+4.65 [−0.62,10.03]	−0.131 [−0.391,0.139]	−0.64 [−2.15,0.88]
$NC-L$	+0.020 [−0.113,0.160]	+0.004 [−0.141,0.149]	−6.99 [−8.81,−5.21]	−0.299 [−0.480,−0.118]	−1.88 [−2.35,−1.42]
$NP-P$	+0.007 [−0.160,0.173]	−0.032 [−0.218,0.153]	−10.00 [−13.27,−6.66]	−0.046 [−0.241,0.154]	−2.75 [−3.49,−2.01]
I_1	+0.013 [−0.260,0.294]	+0.295 [0.019,0.572]	+11.64 [6.20,17.24]	+0.168 [−0.151,0.492]	+1.24 [−0.27,2.80]
I_2	+0.027 [−0.240,0.293]	+0.331 [0.047,0.614]	+14.65 [9.86,19.46]	−0.085 [−0.428,0.262]	+2.11 [0.79,3.45]

Table 29: ChestX, 5-shot component effects with unadjusted 95% pointwise intervals (200 paired episodes). Positive direct effects favor the first variant; interaction signs follow Equation 14.

Contrast	Acc. (pp)	F1 (pp)	NLL (10^{-3})	ECE (pp)	Brier (10^{-3})
$F-L$	+0.353 [−0.273,0.980]	+0.587 [−0.063,1.231]	+430.87 [407.41,454.52]	+9.192 [8.461,9.913]	+92.63 [85.89,99.25]
$F-NC$	+0.273 [−0.360,0.907]	+0.507 [−0.149,1.165]	+443.15 [419.42,467.20]	+9.270 [8.528,10.002]	+95.17 [88.37,101.88]
$NB-L$	−0.033 [−0.627,0.547]	+0.105 [−0.508,0.718]	+449.33 [424.39,474.33]	+9.895 [9.144,10.634]	+98.20 [91.08,105.17]
$F-NP$	−0.133 [−0.620,0.360]	−0.149 [−0.648,0.349]	+435.32 [418.04,452.41]	+7.994 [7.486,8.497]	+83.05 [78.93,87.13]
$NB-P$	−0.607 [−1.067,−0.140]	−0.720 [−1.186,−0.242]	+425.79 [408.29,443.21]	+8.248 [7.713,8.779]	+83.37 [78.98,87.78]
$F-NB$	+0.387 [0.113,0.667]	+0.482 [0.200,0.771]	−18.46 [−22.58,−14.31]	−0.703 [−1.006,−0.406]	−5.57 [−6.76,−4.38]
$NC-L$	+0.080 [−0.067,0.227]	+0.081 [−0.078,0.241]	−12.29 [−14.38,−10.12]	−0.078 [−0.286,0.130]	−2.54 [−3.11,−1.95]
$NP-P$	−0.087 [−0.287,0.113]	−0.089 [−0.292,0.117]	−28.00 [−30.94,−25.11]	−0.450 [−0.708,−0.199]	−5.24 [−6.09,−4.39]
I_1	+0.307 [−0.020,0.633]	+0.401 [0.062,0.743]	−6.17 [−10.94,−1.46]	−0.626 [−1.009,−0.249]	−3.03 [−4.40,−1.69]
I_2	+0.473 [0.113,0.833]	+0.571 [0.208,0.938]	+9.54 [5.62,13.32]	−0.253 [−0.641,0.138]	−0.33 [−1.51,0.84]

C.1 Support-only process diagnostics

Support-only diagnostics trace the recorded adaptation states (Figures 6–8). At step 50, CenWrite's B-factor mean/centered gradient norm ratio is lower than LoRA's in all eight settings: 0.62–1.15 versus 1.37–2.36. These local raw-gradient decompositions exclude the mean term from centered variants' actual factor gradients. Mean-response probes use all 48 adapted projections; residual

probes use the 24 containing b. The support reference stays fixed while later activation means can change, so learning does not recenter every state. Each curve summarizes three fixed tasks with repeated layer measurements; these update-path illustrations do not replace the 200-episode paired comparisons or establish significance.

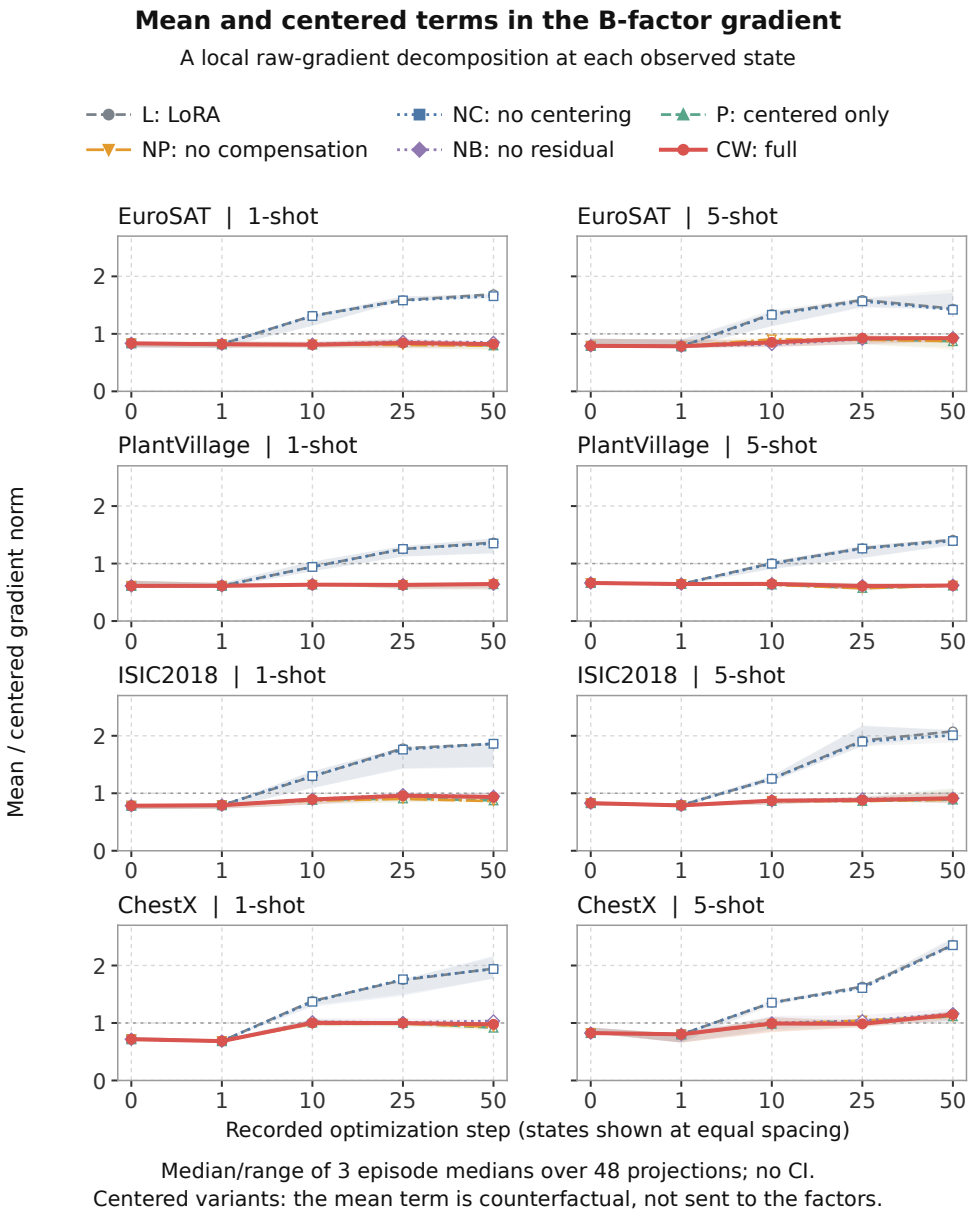


Figure 6: Observed factor-gradient decomposition. $\|G_B^{\text{mean}}\|_F / \|G_B^{\text{centered}}\|_F$ is evaluated at each variant's current state. For centered variants, the mean term is a counterfactual raw-update component, not an applied factor gradient. Lines are medians of three fixed-episode medians over 48 projections; bands are min–max ranges, not confidence intervals. All eight domain/shot settings and six variants are shown. Five recorded states are equally spaced; lines do not imply intermediate observations.

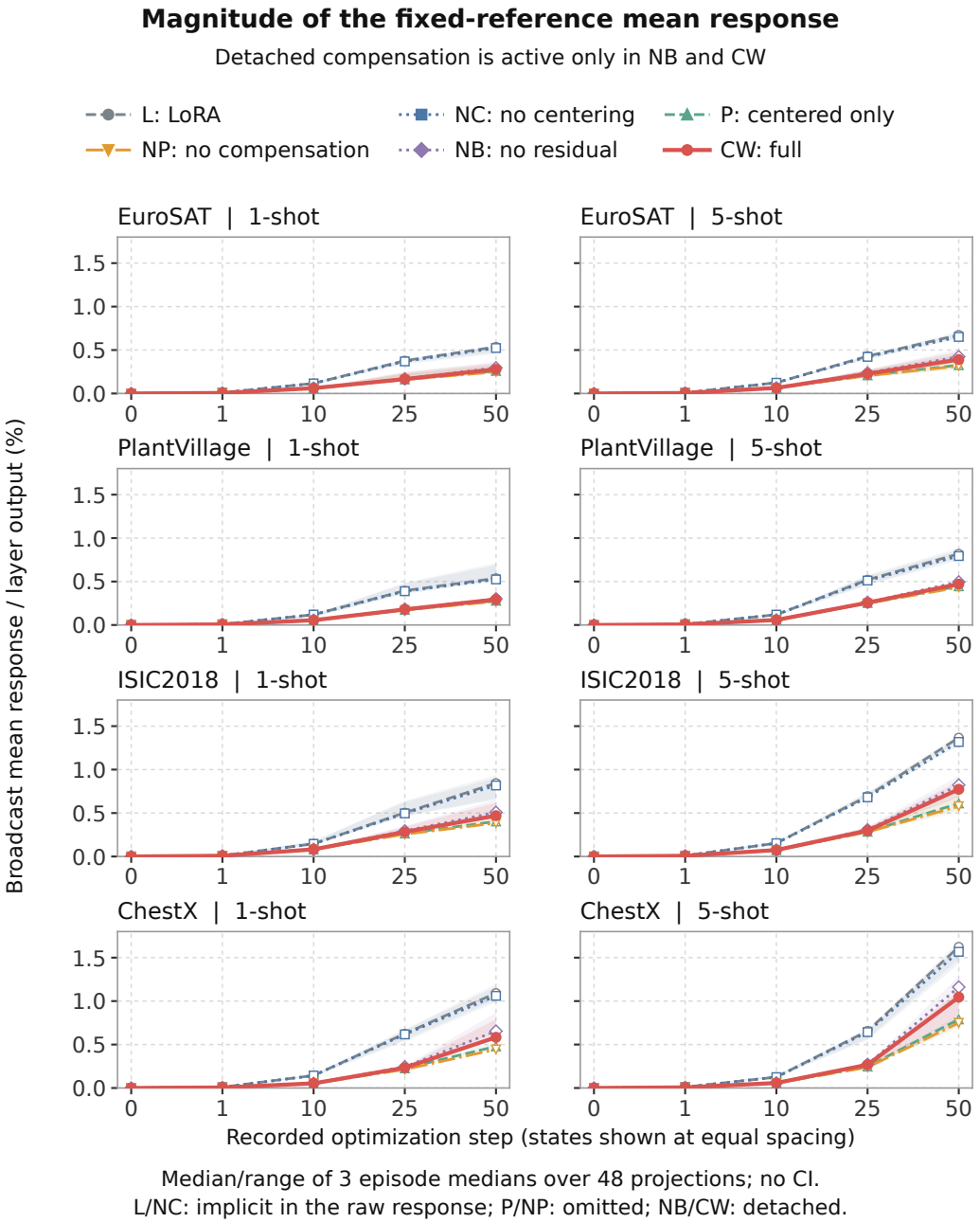


Figure 7: Observed fixed-reference mean response. Broadcast $\Delta W\mu$ Frobenius norm as a percentage of full adapted-layer output norm. The term is implicit in L/NC, omitted in P/NP and restored with stopped gradients in NB/CW; a nonzero diagnostic does not imply active compensation. Fixed-task medians, ranges and equal state spacing follow Figure 6. Magnitudes describe computation, not predictive benefit.

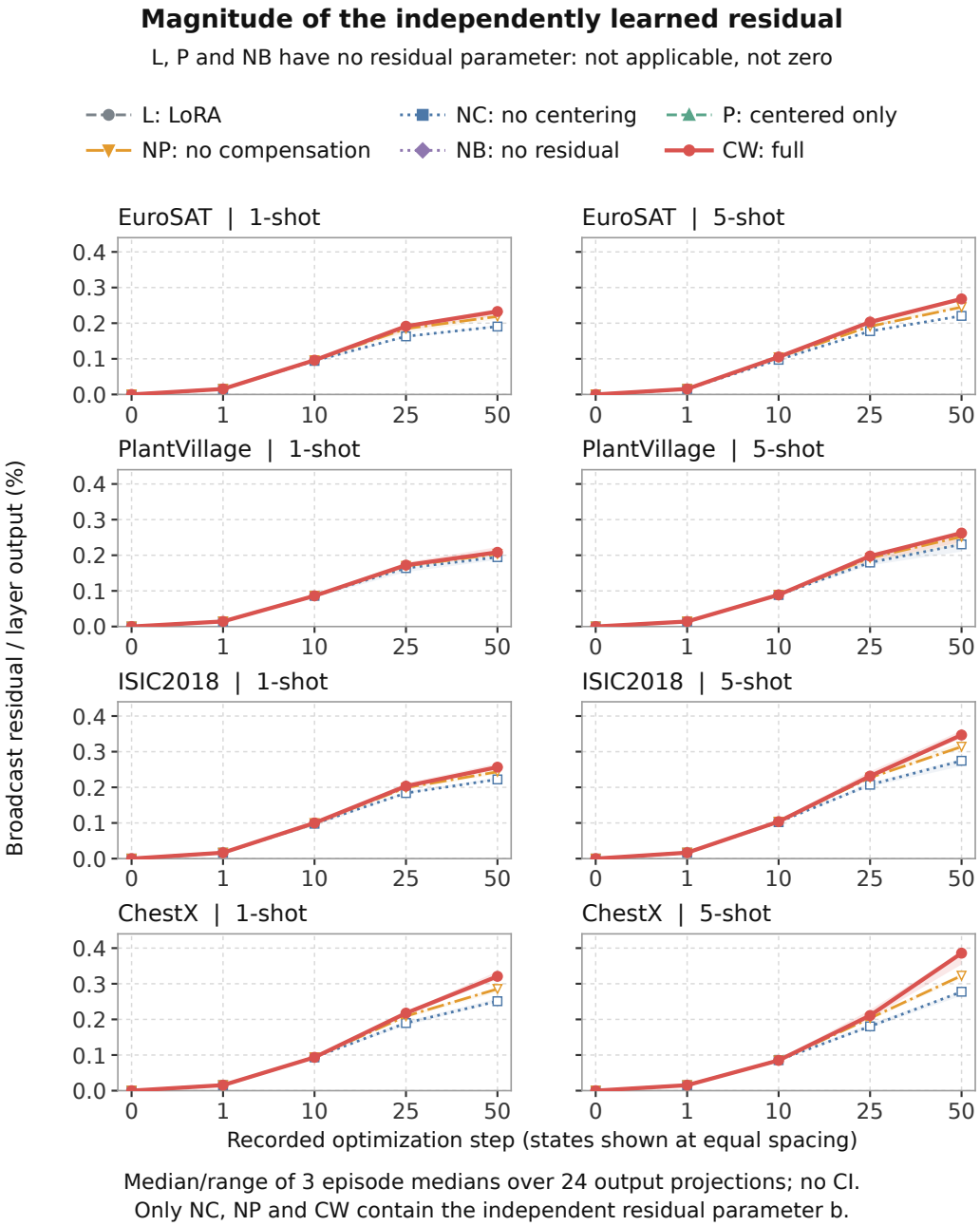


Figure 8: Observed independent residual. Broadcast b Frobenius norm as a percentage of full adapted-layer output norm. NC, NP and CW learn b; L, P and NB lack it and are not plotted as zero curves. Each task's median uses 24 output projections containing b, excluding the other 24 missing values. Lines/bands summarize the median/range across three fixed tasks at five recorded states, not confidence intervals; predictive effects use the separate 200-episode paired comparisons.

D COMPLETE DINOV2 BASELINE COMPARISONS

The secondary-encoder comparison covers seven methods, four domains, two support sizes and 200 episodes per setting. All five metrics and equal eight-setting means are reported in Tables 30–34. Differences use percentage points for accuracy, F1 and ECE and native units for NLL and Brier; shading identifies CenWrite, not significance. Aggregate intervals use a separate DINOv2 30-endpoint family with 20,000 paired bootstrap draws (seed 2026092301), preserving matched methods and metrics within each resampled setting (Table 35). The 240 setting-level differences have unadjusted pointwise intervals (Tables 36–41). Neither these intervals nor the two separate backbone families provide joint coverage over all settings or the 60 endpoints in Figure 9.

CenWrite has the highest mean accuracy, supported against all six adapters by simultaneous intervals. Macro-F1 improves over five adapters. The accuracy gain over LoRA is 0.28 percentage points $[0.004, 0.55]$. All three probability errors improve over FacT and FourierFT; NLL and Brier also improve over RandLoRA. These results establish classification gains under a second pretrained encoder using the same support-adaptation protocol. The following domain/shot tables give all five metrics and paired differences for the transferred configurations.

Table 30: DINOv2 Accuracy (%) ↑ across all domains and support sizes. Bold/underline: best/second-best displayed means, including ties.

Dataset	Shot	LoRA	FacT	BOFT	FourierFT	RandLoRA	ReFTA	CenWrite
EuroSAT	1	67.63	<u>67.94</u>	67.15	45.67	39.24	67.11	**68.13**
EuroSAT	5	**88.20**	87.07	87.79	59.55	53.95	87.75	<u>87.85</u>
PlantVillage	1	84.85	80.97	82.43	45.53	33.38	<u>84.89</u>	**85.25**
PlantVillage	5	**97.03**	95.33	96.69	65.75	60.51	<u>96.87</u>	96.68
ISIC	1	33.12	<u>33.15</u>	32.24	27.49	24.91	32.81	**33.48**
ISIC	5	50.48	50.00	<u>51.15</u>	37.76	33.22	49.76	**51.67**
ChestX	1	21.43	21.61	**21.74**	20.11	19.61	<u>21.73</u>	21.69
ChestX	5	25.06	<u>25.08</u>	24.76	21.74	20.93	25.01	**25.27**
Equal-setting mean		<u>58.48</u>	57.64	58.00	40.45	35.72	58.24	**58.75**

Table 31: DINOv2 Macro-F1 (%) ↑ across all domains and support sizes. Bold/underline: best/second-best displayed means, including ties.

Dataset	Shot	LoRA	FacT	BOFT	FourierFT	RandLoRA	ReFTA	CenWrite
EuroSAT	1	65.49	<u>65.83</u>	65.03	42.34	33.50	65.17	**66.02**
EuroSAT	5	**88.01**	86.85	87.61	57.95	49.91	87.58	<u>87.65</u>
PlantVillage	1	83.81	79.92	81.36	42.80	26.23	<u>83.83</u>	**84.23**
PlantVillage	5	**96.97**	95.19	96.61	64.59	57.36	<u>96.82</u>	96.61
ISIC	1	29.53	29.69	29.04	24.39	18.64	<u>29.77</u>	**29.85**
ISIC	5	49.38	48.66	<u>50.05</u>	35.59	29.00	48.77	**50.62**
ChestX	1	18.19	17.99	<u>18.34</u>	16.53	13.29	**19.02**	18.28
ChestX	5	23.23	22.69	21.90	18.58	15.26	<u>23.58</u>	**23.88**
Equal-setting mean		<u>56.83</u>	55.85	56.24	37.85	30.40	56.82	**57.14**

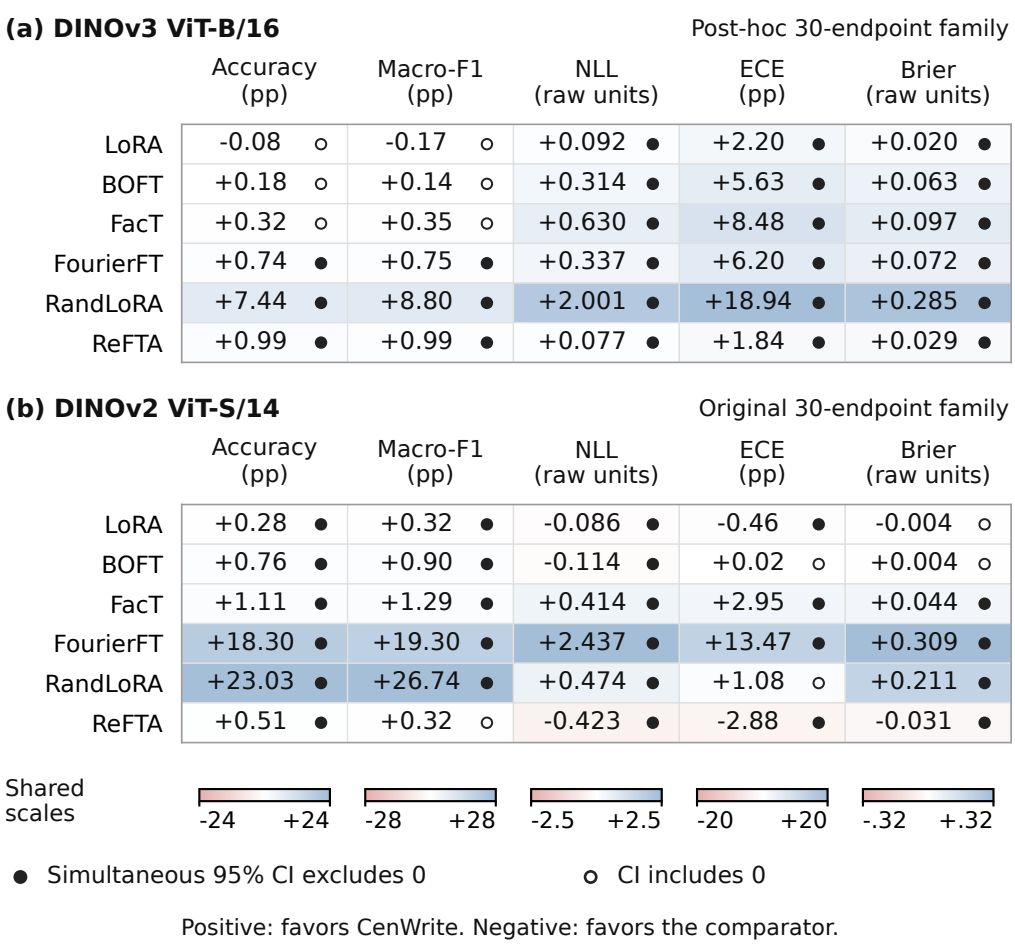


Figure 9: Relative effects across two backbones. Each cell is CenWrite's equal-setting average advantage over the named comparator across four domains and two shot counts. Positive values favor CenWrite: Accuracy/F1 use CenWrite minus comparator, while NLL/ECE/Brier reverse the subtraction. Accuracy, F1 and ECE use percentage points; NLL and Brier retain their original units. Each metric shares a symmetric color scale across panels. Filled/open circles indicate whether its simultaneous 95% interval excludes/includes zero. DINOv3 uses a supplementary post-hoc 30-endpoint family; DINOv2 has its own 30-endpoint family with separate coverage.

Table 32: DINOv2 NLL ↓ across all domains and support sizes. Bold/underline: best/second-best displayed means, including ties.

Dataset	Shot	LoRA	FacT	BOFT	FourierFT	RandLoRA	ReFTA	CenWrite
EuroSAT	1	1.2736	1.6104	<u>1.2714</u>	5.7259	3.7622	**1.0944**	1.2957
EuroSAT	5	<u>0.4386</u>	0.6485	0.4970	2.2982	1.6947	**0.3894**	0.4706
PlantVillage	1	<u>0.5367</u>	0.8574	0.5873	5.4034	2.8841	**0.4791**	0.5515
PlantVillage	5	<u>0.1033</u>	0.2183	0.1185	2.0920	1.6037	**0.1024**	0.1137
ISIC	1	3.2397	4.3271	<u>3.1381</u>	8.1084	3.7927	**2.8143**	3.4912
ISIC	5	2.3389	3.1510	2.4012	3.4002	<u>2.3208</u>	**2.0766**	2.4555
ChestX	1	4.3900	5.3911	4.3194	7.1970	**3.2809**	<u>3.6688</u>	4.6942
ChestX	5	4.6205	4.7426	4.3885	<u>2.9047</u>	**2.0869**	3.6201	4.5601
Equal-setting mean		2.1177	2.6183	<u>2.0902</u>	4.6412	2.6783	**1.7806**	2.2041

Table 33: DINOv2 ECE (%) ↓ across all domains and support sizes. Bold/underline: best/second-best displayed means, including ties.

Dataset	Shot	LoRA	FacT	BOFT	FourierFT	RandLoRA	ReFTA	CenWrite
EuroSAT	1	22.11	24.87	22.95	48.18	41.88	**19.58**	22.46
EuroSAT	5	9.36	11.05	9.97	30.35	26.41	**8.65**	9.77
PlantVillage	1	10.79	15.23	11.69	47.90	36.25	**10.19**	10.86
PlantVillage	5	**2.88**	4.33	3.07	27.46	23.24	3.04	3.04
ISIC	1	48.56	54.66	48.97	64.07	**42.73**	45.27	50.15
ISIC	5	36.40	41.01	36.64	41.02	**29.13**	34.12	36.33
ChestX	1	59.95	64.56	60.13	63.83	**38.03**	55.23	61.06
ChestX	5	58.36	59.97	58.88	37.08	**23.08**	52.95	58.44
Equal-setting mean		31.05	34.46	31.54	44.99	32.60	**28.63**	31.51

Table 34: DINOv2 Brier ↓ across all domains and support sizes. Bold/underline: best/second-best displayed means, including ties.

Dataset	Shot	LoRA	FacT	BOFT	FourierFT	RandLoRA	ReFTA	CenWrite
EuroSAT	1	0.5107	0.5339	0.5178	0.9910	0.9612	**0.4918**	0.5096
EuroSAT	5	0.1901	0.2186	0.2003	0.6685	0.6706	**0.1860**	0.1957
PlantVillage	1	0.2348	0.3144	0.2662	0.9901	0.9542	**0.2261**	0.2317
PlantVillage	5	**0.0475**	0.0794	0.0521	0.5866	0.5886	0.0492	0.0512
ISIC	1	1.0843	1.1623	1.0889	1.3224	1.1008	**1.0464**	1.1028
ISIC	5	0.8057	0.8647	0.8068	0.9692	0.8901	**0.7915**	0.8022
ChestX	1	1.3028	1.3632	1.3015	1.3702	**1.0761**	1.2379	1.3163
ChestX	5	1.2601	1.2800	1.2679	1.0388	**0.9153**	1.1906	1.2572
Equal-setting mean		0.6795	0.7271	0.6877	0.9921	0.8946	**0.6525**	0.6833

Table 35: DINOv2 equal-setting effects with approximate 95% simultaneous intervals for its 30-endpoint family. Positive values favor CenWrite. Accuracy/F1/ECE use percentage points; NLL/Brier use native units.

Baseline	Accuracy (pp)	Macro-F1 (pp)	NLL	ECE (pp)	Brier
LoRA	+0.28 [0.004, 0.55]	+0.32 [0.02, 0.62]	−0.0864 [−0.1105, −0.0623]	−0.46 [−0.77, −0.15]	−0.0038 [−0.0081, 0.0005]
FacT	+1.11 [0.67, 1.55]	+1.29 [0.81, 1.77]	+0.4142 [0.3702, 0.4583]	+2.95 [2.49, 3.41]	+0.0437 [0.0360, 0.0515]
BOFT	+0.76 [0.32, 1.19]	+0.90 [0.43, 1.38]	−0.1139 [−0.1561, −0.0718]	+0.02 [−0.44, 0.49]	+0.0044 [−0.0032, 0.0120]
FourierFT	+18.30 [17.58, 19.03]	+19.30 [18.52, 20.07]	+2.4372 [2.3284, 2.5459]	+13.47 [12.66, 14.28]	+0.3088 [0.2954, 0.3221]
RandLoRA	+23.03 [22.17, 23.90]	+26.74 [25.74, 27.75]	+0.4742 [0.3748, 0.5736]	+1.08 [−0.05, 2.22]	+0.2113 [0.1965, 0.2261]
ReFTA	+0.51 [0.17, 0.86]	+0.32 [−0.05, 0.70]	−0.4234 [−0.4510, −0.3959]	−2.88 [−3.25, −2.51]	−0.0309 [−0.0363, −0.0255]

Table 36: DINOv2 paired effects relative to LoRA, with unadjusted 95% pointwise intervals over 200 episodes per setting. Positive values favor CenWrite. Accuracy/F1/ECE use percentage points; NLL/Brier use native units.

Setting	NLL	Brier	ECE (pp)	Accuracy (pp)	Macro-F1 (pp)
EuroSAT 1-shot	−0.0221 [−0.0583, 0.0142]	+0.0011 [−0.0088, 0.0112]	−0.35 [−1.02, 0.33]	+0.50 [−0.14, 1.13]	+0.53 [−0.23, 1.27]
EuroSAT 5-shot	−0.0321 [−0.0487, −0.0156]	−0.0056 [−0.0119, 0.0007]	−0.41 [−0.77, −0.06]	−0.35 [−0.78, 0.09]	−0.36 [−0.80, 0.08]
PlantVillage 1-shot	−0.0148 [−0.0328, 0.0032]	+0.0031 [−0.0034, 0.0096]	−0.08 [−0.49, 0.34]	+0.40 [−0.09, 0.88]	+0.42 [−0.12, 0.96]
PlantVillage 5-shot	−0.0103 [−0.0187, −0.0025]	−0.0037 [−0.0064, −0.0010]	−0.16 [−0.31, −0.01]	−0.35 [−0.57, −0.14]	−0.36 [−0.60, −0.14]
ISIC 1-shot	−0.2515 [−0.3039, −0.1994]	−0.0185 [−0.0274, −0.0098]	−1.59 [−2.30, −0.90]	+0.36 [−0.21, 0.93]	+0.32 [−0.28, 0.92]
ISIC 5-shot	−0.1166 [−0.1498, −0.0836]	+0.0035 [−0.0042, 0.0112]	+0.07 [−0.52, 0.66]	+1.19 [0.68, 1.70]	+1.25 [0.69, 1.80]
ChestX 1-shot	−0.3042 [−0.3892, −0.2199]	−0.0135 [−0.0234, −0.0036]	−1.11 [−1.88, −0.35]	+0.26 [−0.27, 0.79]	+0.09 [−0.50, 0.68]
ChestX 5-shot	+0.0605 [0.0064, 0.1157]	+0.0029 [−0.0055, 0.0115]	−0.08 [−0.74, 0.58]	+0.21 [−0.37, 0.77]	+0.66 [0.08, 1.23]

Table 37: DINOv2 paired effects relative to FacT, with unadjusted 95% pointwise intervals over 200 episodes per setting. Positive values favor CenWrite. Accuracy/F1/ECE use percentage points; NLL/Brier use native units.

Setting	NLL	Brier	ECE (pp)	Accuracy (pp)	Macro-F1 (pp)
EuroSAT 1-shot	+0.3147 [0.2323, 0.3990]	+0.0243 [0.0051, 0.0442]	+2.42 [1.31, 3.54]	+0.19 [−0.93, 1.33]	+0.19 [−1.04, 1.43]
EuroSAT 5-shot	+0.1779 [0.1370, 0.2197]	+0.0229 [0.0116, 0.0342]	+1.28 [0.69, 1.87]	+0.78 [0.08, 1.48]	+0.81 [0.08, 1.53]
PlantVillage 1-shot	+0.3059 [0.2513, 0.3613]	+0.0828 [0.0664, 0.0991]	+4.37 [3.53, 5.21]	+4.29 [3.25, 5.35]	+4.31 [3.15, 5.49]
PlantVillage 5-shot	+0.1046 [0.0814, 0.1299]	+0.0282 [0.0214, 0.0353]	+1.29 [0.95, 1.66]	+1.35 [0.93, 1.79]	+1.41 [0.97, 1.88]
ISIC 1-shot	+0.8359 [0.7402, 0.9314]	+0.0595 [0.0439, 0.0751]	+4.51 [3.45, 5.57]	+0.33 [−0.49, 1.13]	+0.16 [−0.76, 1.06]
ISIC 5-shot	+0.6955 [0.6167, 0.7750]	+0.0625 [0.0481, 0.0770]	+4.69 [3.80, 5.58]	+1.67 [0.83, 2.52]	+1.96 [1.03, 2.91]
ChestX 1-shot	+0.6969 [0.5649, 0.8318]	+0.0469 [0.0333, 0.0611]	+3.49 [2.58, 4.45]	+0.08 [−0.54, 0.70]	+0.29 [−0.43, 1.02]
ChestX 5-shot	+0.1825 [0.0904, 0.2752]	+0.0228 [0.0102, 0.0357]	+1.53 [0.69, 2.39]	+0.19 [−0.55, 0.93]	+1.19 [0.43, 1.96]

Table 38: DINOv2 paired effects relative to BOFT, with unadjusted 95% pointwise intervals over 200 episodes per setting. Positive values favor CenWrite. Accuracy/F1/ECE use percentage points; NLL/Brier use native units.

Setting	NLL	Brier	ECE (pp)	Accuracy (pp)	Macro-F1 (pp)
EuroSAT 1-shot	−0.0244 [−0.1009, 0.0524]	+0.0083 [−0.0130, 0.0296]	+0.49 [−0.68, 1.68]	+0.97 [−0.29, 2.23]	+0.99 [−0.41, 2.40]
EuroSAT 5-shot	+0.0263 [−0.0058, 0.0584]	+0.0046 [−0.0057, 0.0149]	+0.20 [−0.36, 0.76]	+0.07 [−0.60, 0.73]	+0.05 [−0.65, 0.74]

Continued on the following page.

Table 38 continued

Setting	NLL	Brier	ECE (pp)	Accuracy (pp)	Macro-F1 (pp)
PlantVillage 1-shot	+0.0358 [−0.0070, 0.0776]	+0.0345 [0.0212, 0.0479]	+0.83 [0.14, 1.53]	+2.82 [1.95, 3.71]	+2.87 [1.87, 3.86]
PlantVillage 5-shot	+0.0048 [−0.0094, 0.0196]	+0.0009 [−0.0039, 0.0056]	+0.03 [−0.24, 0.30]	−0.01 [−0.34, 0.31]	−0.005 [−0.35, 0.34]
ISIC 1-shot	−0.3532 [−0.4444, −0.2634]	−0.0138 [−0.0289, 0.0012]	−1.17 [−2.24, −0.11]	+1.24 [0.45, 2.04]	+0.81 [−0.09, 1.71]
ISIC 5-shot	−0.0543 [−0.1122, 0.0050]	+0.0046 [−0.0087, 0.0185]	+0.32 [−0.52, 1.18]	+0.52 [−0.26, 1.32]	+0.58 [−0.25, 1.43]
ChestX 1-shot	−0.3748 [−0.5127, −0.2402]	−0.0148 [−0.0294, 0.00004]	−0.93 [−1.94, 0.07]	−0.05 [−0.71, 0.62]	−0.06 [−0.78, 0.66]
ChestX 5-shot	−0.1716 [−0.2704, −0.0719]	+0.0107 [−0.0034, 0.0248]	+0.44 [−0.50, 1.37]	+0.51 [−0.25, 1.27]	+1.99 [1.22, 2.76]

Table 39: DINOv2 paired effects relative to FourierFT, with unadjusted 95% pointwise intervals over 200 episodes per setting. Positive values favor CenWrite. Accuracy/F1/ECE use percentage points; NLL/Brier use native units.

Setting	NLL	Brier	ECE (pp)	Accuracy (pp)	Macro-F1 (pp)
EuroSAT 1-shot	+4.4302 [4.1619, 4.7058]	+0.4814 [0.4507, 0.5125]	+25.73 [24.03, 27.43]	+22.45 [20.72, 24.19]	+23.68 [21.85, 25.51]
EuroSAT 5-shot	+1.8276 [1.7190, 1.9360]	+0.4729 [0.4532, 0.4921]	+20.58 [19.52, 21.64]	+28.30 [27.05, 29.53]	+29.70 [28.38, 30.99]
PlantVillage 1-shot	+4.8519 [4.6291, 5.0791]	+0.7584 [0.7274, 0.7902]	+37.04 [35.45, 38.64]	+39.72 [37.95, 41.53]	+41.43 [39.46, 43.43]
PlantVillage 5-shot	+1.9783 [1.8690, 2.0913]	+0.5354 [0.5107, 0.5601]	+24.42 [23.23, 25.60]	+30.93 [29.44, 32.41]	+32.02 [30.46, 33.58]
ISIC 1-shot	+4.6171 [4.3630, 4.8681]	+0.2196 [0.1947, 0.2446]	+13.92 [12.41, 15.41]	+5.99 [4.72, 7.27]	+5.46 [4.09, 6.83]
ISIC 5-shot	+0.9448 [0.8030, 1.0902]	+0.1670 [0.1450, 0.1892]	+4.69 [3.17, 6.22]	+13.91 [12.57, 15.24]	+15.03 [13.59, 16.47]
ChestX 1-shot	+2.5028 [2.2299, 2.7806]	+0.0539 [0.0312, 0.0765]	+2.77 [1.03, 4.48]	+1.59 [0.85, 2.33]	+1.75 [0.97, 2.53]
ChestX 5-shot	−1.6553 [−1.7816, −1.5261]	−0.2184 [−0.2366, −0.2000]	−21.37 [−23.02, −19.69]	+3.53 [2.70, 4.35]	+5.30 [4.47, 6.15]

Table 40: DINOv2 paired effects relative to RandLoRA, with unadjusted 95% pointwise intervals over 200 episodes per setting. Positive values favor CenWrite. Accuracy/F1/ECE use percentage points; NLL/Brier use native units.

Setting	NLL	Brier	ECE (pp)	Accuracy (pp)	Macro-F1 (pp)
EuroSAT 1-shot	+2.4664 [2.2203, 2.7258]	+0.4516 [0.4205, 0.4832]	+19.43 [17.30, 21.59]	+28.89 [27.01, 30.76]	+32.52 [30.39, 34.66]
EuroSAT 5-shot	+1.2241 [1.1533, 1.2950]	+0.4749 [0.4571, 0.4922]	+16.64 [15.52, 17.72]	+33.91 [32.44, 35.41]	+37.75 [35.99, 39.53]
PlantVillage 1-shot	+2.3326 [2.1525, 2.5207]	+0.7225 [0.6905, 0.7558]	+25.38 [22.85, 27.90]	+51.87 [49.90, 53.86]	+58.00 [55.65, 60.38]
PlantVillage 5-shot	+1.4900 [1.3638, 1.6303]	+0.5375 [0.5039, 0.5710]	+20.20 [18.58, 21.88]	+36.17 [33.64, 38.71]	+39.24 [36.42, 42.11]

Continued on the following page.

Table 40 continued

Setting	NLL	Brier	ECE (pp)	Accuracy (pp)	Macro-F1 (pp)
ISIC 1-shot	+0.3015 [0.0314, 0.5803]	−0.0019 [−0.0327, 0.0296]	−7.42 [−10.12, −4.69]	+8.57 [7.41, 9.73]	+11.21 [9.88, 12.53]
ISIC 5-shot	−0.1347 [−0.2556, −0.0103]	+0.0879 [0.0655, 0.1104]	−7.19 [−9.18, −5.19]	+18.45 [17.08, 19.83]	+21.63 [20.02, 23.24]
ChestX 1-shot	−1.4133 [−1.6331, −1.1865]	−0.2402 [−0.2685, −0.2115]	−23.03 [−25.58, −20.45]	+2.08 [1.31, 2.85]	+4.99 [4.11, 5.87]
ChestX 5-shot	−2.4732 [−2.5726, −2.3734]	−0.3419 [−0.3589, −0.3246]	−35.36 [−36.95, −33.74]	+4.33 [3.48, 5.20]	+8.62 [7.63, 9.62]

Table 41: DINOv2 paired effects relative to ReFTA, with unadjusted 95% pointwise intervals over 200 episodes per setting. Positive values favor CenWrite. Accuracy/F1/ECE use percentage points; NLL/Brier use native units.

Setting	NLL	Brier	ECE (pp)	Accuracy (pp)	Macro-F1 (pp)
EuroSAT 1-shot	−0.2014 [−0.2470, −0.1577]	−0.0177 [−0.0306, −0.0050]	−2.87 [−3.69, −2.06]	+1.01 [0.20, 1.83]	+0.85 [−0.10, 1.79]
EuroSAT 5-shot	−0.0812 [−0.1038, −0.0592]	−0.0097 [−0.0172, −0.0022]	−1.13 [−1.57, −0.68]	+0.10 [−0.43, 0.65]	+0.07 [−0.48, 0.64]
PlantVillage 1-shot	−0.0725 [−0.1011, −0.0442]	−0.0056 [−0.0148, 0.0038]	−0.67 [−1.13, −0.21]	+0.37 [−0.31, 1.03]	+0.40 [−0.35, 1.15]
PlantVillage 5-shot	−0.0112 [−0.0223, −0.0011]	−0.0020 [−0.0059, 0.0019]	+0.002 [−0.22, 0.22]	−0.19 [−0.49, 0.10]	−0.21 [−0.52, 0.09]
ISIC 1-shot	−0.6769 [−0.7407, −0.6138]	−0.0563 [−0.0676, −0.0451]	−4.88 [−5.70, −4.08]	+0.67 [0.01, 1.33]	+0.07 [−0.67, 0.82]
ISIC 5-shot	−0.3789 [−0.4187, −0.3396]	−0.0107 [−0.0202, −0.0012]	−2.20 [−2.91, −1.50]	+1.91 [1.27, 2.55]	+1.85 [1.19, 2.52]
ChestX 1-shot	−1.0254 [−1.1149, −0.9359]	−0.0783 [−0.0904, −0.0665]	−5.84 [−6.74, −4.94]	−0.03 [−0.67, 0.61]	−0.74 [−1.45, −0.04]
ChestX 5-shot	−0.9400 [−1.0022, −0.8785]	−0.0666 [−0.0776, −0.0553]	−5.49 [−6.33, −4.66]	+0.25 [−0.47, 0.97]	+0.30 [−0.41, 1.01]

E UHCS Steel-Microstructure Evaluation

E.1 Data, groups, and statistical protocol

We use a fixed five-class subset of Ultrahigh Carbon Steel Micrographs (Hecht et al., 2017) from the author-recommended Citrination mirror. Cropping 613 selected images from 645×522 to 645×484 removes the SEM text/scale-bar strip; whole-view deduplication leaves 601 images. Labels are network, pearlite, pearlite+spheroidite, spheroidite and spheroidite+widmanstatten: dominant field-of-view morphologies rather than composition, properties or five pure phases. Sample identifiers define 38 source groups, with samples 24/25 conservatively merged for matching processing strings. These are provenance groups, not verified manufacturing batches. Original images are retained; CC BY 3.0 US applies, with crop and resize changes documented.

Every task uses five classes, one or five support images and ten queries per class. Source-group-disjoint support/query pools are fixed globally within each fold, across classes and shots. Methods share 200 tasks per shot, nested one-/five-shot supports and identical queries. A shot is an image, not an independent specimen. Primary fold 5 reports eight methods or variants; folds 31/35 rotate the pearlite+spheroidite support group and compare four methods. Reused images/groups make these split-sensitivity analyses, not independent datasets. Pearlite's five support images per fold require reuse and motivate ten queries per class. All runs use the fixed DINOv3 ViT-B/16 encoder, 224-pixel inputs and 50-step adaptation.

Each fold reports 15 performance endpoints (three baselines, five metrics). Their intervals retain the original, larger 20-endpoint adjustment rather than recalculating a less conservative threshold for the reported subset. Fold 5 adds a separate 50-endpoint component family. The 20,000 paired bootstrap draws share episode indices across methods and nested shots. Equal-shot effects use approximate 95% simultaneous max-$|T|$ intervals; shot-specific intervals are unadjusted pointwise descriptions. No adjustment pools folds/datasets or assumes cross-fold independence. Intervals describe episode resampling from fixed pools, not a population of independent specimens. The parameter-free frozen cosine-prototype reference uses $T = 1$; its uncalibrated scores lie outside adjusted families.

Recorded magnifications span $49\times$–$246\times$ for network and $1964\times$–$34372\times$ for pearlite, with no common five-class range. A common-magnification five-class/five-shot/ten-query evaluation therefore cannot preserve source-group separation. A magnification-only nearest-support diagnostic, using no pixels or learned parameters, reaches 43.72–46.75% accuracy over the six fold/shot settings (20% chance). The acquisition metadata and source-group counts below specify the evaluated image pools.

E.2 Complete results and scope of the component evidence

CenWrite lowers NLL, ECE and Brier relative to LoRA in each fold's separate simultaneous family. Fold 5 reductions are 0.05407 $[0.04193, 0.06622]$ for NLL and 0.01240 $[0.00686, 0.01794]$ for Brier. Fold 31 additionally supports accuracy and macro-F1 gains over LoRA. Primary-fold results favor CenWrite over BOFT and ReFTA on all five metrics. Both centered-route contrasts, $F - NC$ and $NB - L$, improve NLL, ECE and Brier in the primary component family. Compensation improves NLL with or without the residual. These comparisons connect the observed probability gains to the centered update route and mean-response compensation. The tables report complete primary/rotation means, family and shot-level intervals for every baseline, direct component contrast and interaction, alongside the $T = 1$ prototype reference.

Table 42: UHCS deduplicated pools (images/source groups). N: network; Pe: pearlite; S: spheroidite; W: widmanstatten. Each fold has 18 support/20 query groups, disjoint across pools. Shared groups within a pool cannot be summed across classes.

	Fold 5		Fold 31		Fold 35	
Class	Support	Query	Support	Query	Support	Query
N	38/11	74/15	46/11	66/15	42/11	70/15
Pe	5/3	11/3	5/3	11/3	5/3	11/3
Pe+S	15/1	13/2	6/1	22/2	7/1	21/2
S	212/15	156/15	213/16	155/14	216/16	152/14
S+W	20/10	57/12	20/10	57/12	22/11	55/11

Table 43: UHCS acquisition metadata: 601 images, including 560 with known magnification. SE/BSE/? denote secondary-electron/backscattered-electron/missing detector counts. Classes follow Table 42.

Class	Images	Groups	Mag. min/median/max (×)	Missing	SE/BSE/?
N	112	26	49/98/246	4	96/11/5
Pe	16	6	1964/7856/34372	5	16/0/0
Pe+S	28	3	491/1964/4910	4	28/0/0
S	368	30	246/1964/19641	26	328/19/21
S+W	77	22	246/982/11785	2	76/0/1

Table 44: UHCS fold 5: means on 200 matched five-way tasks per support size, with 10 queries/class. Accuracy/F1/ECE are percentages; NLL/Brier use native units. Variants follow Table 45. Bold/underline: best/second-best displayed means, including ties.

1-shot Metric	L	BOFT	ReFTA	NC	P	NP	NB	CenWrite
Accuracy ↑	52.31	50.58	50.17	52.36	**53.39**	<u>53.24</u>	52.65	52.71
Macro-F1 ↑	49.67	47.47	47.05	49.75	**50.87**	<u>50.66</u>	49.97	50.05
NLL ↓	1.3225	1.4956	1.3558	1.3264	1.2894	1.2960	**1.2674**	<u>1.2755</u>
ECE ↓	23.50	27.16	23.58	23.81	23.11	23.16	**22.17**	<u>22.35</u>
Brier ↓	0.6545	0.7003	0.6693	0.6556	0.6411	0.6432	**0.6368**	<u>0.6397</u>
5-shot Metric	L	BOFT	ReFTA	NC	P	NP	NB	CenWrite
Accuracy ↑	68.58	66.94	67.09	68.42	<u>69.90</u>	**69.93**	67.98	68.13
Macro-F1 ↑	67.53	64.97	65.72	67.34	<u>68.68</u>	**68.72**	67.23	67.31
NLL ↓	0.8942	1.1873	0.9239	0.8976	0.8393	0.8459	**0.8264**	<u>0.8331</u>
ECE ↓	18.20	23.24	18.64	18.20	**16.87**	<u>17.09</u>	17.15	17.24
Brier ↓	0.4534	0.5199	0.4693	0.4539	**0.4349**	<u>0.4363</u>	0.4427	0.4434

Table 45: UHCS variant and contrast definitions. Positive direct effects favor the first variant (higher Accuracy/F1, lower NLL/ECE/Brier); I1/I2 compare residual effects across update rules.

Symbol	Variant
F	Full CenWrite
L	Ordinary LoRA, without boundary residual
NC	Uncentered low-rank update with boundary residual
P	Centering only, without compensation or boundary residual
NP	Centered update with boundary residual, without compensation
NB	Centered update with compensation, without boundary residual
Contrast	**Linear combination of metric utilities**
C1	$F - L$
C2	$F - NC$
C3	$NB - L$
C4	$F - NP$
C5	$NB - P$
C6	$F - NB$
C7	$NC - L$
C8	$NP - P$
I1	$(F - NC) - (NB - L)$
I2	$(F - NP) - (NB - P)$

Table 46: UHCS sensitivity folds 31 and 35: means on 200 matched tasks per support size, with 10 queries/class. Accuracy/F1/ECE are percentages; NLL/Brier use native units. Bold/underline: best/second-best displayed means, including ties.

Fold	Shot	Metric	L	BOFT	ReFTA	CenWrite
31	1-shot	Accuracy ↑	**53.23**	51.21	51.69	53.04
		Macro-F1 ↑	**50.97**	48.46	48.55	50.88
		NLL ↓	1.3505	1.5594	1.3699	**1.2995**
		ECE ↓	24.16	28.25	23.69	**22.84**
		Brier ↓	0.6554	0.7087	0.6619	**0.6421**
31	5-shot	Accuracy ↑	69.46	67.42	69.86	**71.02**
		Macro-F1 ↑	68.66	66.19	68.91	**70.42**
		NLL ↓	1.0150	1.3515	0.9832	**0.8753**
		ECE ↓	19.26	24.78	18.66	**17.20**
		Brier ↓	0.4650	0.5318	0.4513	**0.4273**
35	1-shot	Accuracy ↑	**53.82**	52.00	53.23	53.58
		Macro-F1 ↑	**51.37**	49.19	50.52	51.32
		NLL ↓	1.3084	1.5135	1.3202	**1.2733**
		ECE ↓	23.19	27.36	22.30	**22.23**
		Brier ↓	0.6383	0.6919	0.6392	**0.6304**
35	5-shot	Accuracy ↑	69.10	68.91	**70.16**	69.91
		Macro-F1 ↑	68.32	67.99	69.25	**69.46**
		NLL ↓	0.9634	1.2586	0.9659	**0.8688**
		ECE ↓	18.44	22.73	17.87	**16.63**
		Brier ↓	0.4545	0.4990	0.4421	**0.4294**

Table 47: Frozen cosine-prototype reference on 200 matched UHCS tasks per fold/shot, with fixed encoder and $T = 1$. Accuracy/F1/ECE are percentages; NLL/Brier use native units. Values are descriptive and outside the adjusted comparison families.

Fold	Shot	Accuracy	Macro-F1	NLL	ECE	Brier
5	1	51.82	48.54	1.5134	28.58	0.7600
5	5	66.66	66.10	1.4907	43.69	0.7507
31	1	50.52	47.98	1.5112	27.17	0.7588
31	5	68.05	67.80	1.4888	44.92	0.7496
35	1	51.32	48.46	1.5100	27.87	0.7582
35	5	65.43	65.24	1.4877	42.09	0.7491

Table 48: UHCS equal-shot means for each fold. Accuracy/F1/ECE are percentages; NLL/Brier use native units. Bold/underline: best/second-best displayed means, including ties.

Primary fold 5, equal shot weights

Metric	L	BOFT	ReFTA	NC	P	NP	NB	CenWrite
Accuracy ↑	60.45	58.76	58.63	60.39	**61.65**	<u>61.59</u>	60.32	60.42
Macro-F1 ↑	58.60	56.22	56.38	58.54	**59.77**	<u>59.69</u>	58.60	58.68
NLL ↓	1.1084	1.3415	1.1398	1.1120	1.0643	1.0709	**1.0469**	<u>1.0543</u>
ECE ↓	20.85	25.20	21.11	21.01	19.99	20.13	**19.66**	<u>19.79</u>
Brier ↓	0.5540	0.6101	0.5693	0.5548	**0.5380**	<u>0.5397</u>	0.5398	0.5416

Sensitivity fold 31, equal shot weights

Metric	L	BOFT	ReFTA	CenWrite
Accuracy ↑	<u>61.35</u>	59.32	60.78	**62.03**
Macro-F1 ↑	<u>59.81</u>	57.33	58.73	**60.65**
NLL ↓	1.1827	1.4555	<u>1.1766</u>	**1.0874**
ECE ↓	21.71	26.51	<u>21.18</u>	**20.02**
Brier ↓	0.5602	0.6203	<u>0.5566</u>	**0.5347**

Sensitivity fold 35, equal shot weights

Metric	L	BOFT	ReFTA	CenWrite
Accuracy ↑	61.46	60.46	<u>61.70</u>	**61.75**
Macro-F1 ↑	59.84	58.59	<u>59.88</u>	**60.39**
NLL ↓	<u>1.1359</u>	1.3860	1.1431	**1.0711**
ECE ↓	20.81	25.04	<u>20.09</u>	**19.43**
Brier ↓	0.5464	0.5955	<u>0.5406</u>	**0.5299**

Table 49: UHCS fold 5: equal-shot component effects with approximate 95% simultaneous max-$|T|$ intervals over 50 endpoints. Accuracy/F1/ECE use percentage points; NLL/Brier use native units. Signs and interactions follow Table 45.

Contrast	Accuracy (pp)	Macro-F1 (pp)	NLL	ECE (pp)	Brier
C1	−0.03 [−0.67, 0.62]	+0.08 [−0.60, 0.76]	+0.0541 [0.0410, 0.0671]	+1.06 [0.38, 1.74]	+0.0124 [0.0065, 0.0183]
C2	+0.03 [−0.62, 0.68]	+0.13 [−0.56, 0.82]	+0.0577 [0.0441, 0.0713]	+1.22 [0.50, 1.93]	+0.0132 [0.0070, 0.0194]
C3	−0.13 [−0.84, 0.58]	−0.0008 [−0.75, 0.74]	+0.0615 [0.0471, 0.0758]	+1.19 [0.45, 1.94]	+0.0142 [0.0077, 0.0208]
C4	−1.17 [−1.67, −0.66]	−1.01 [−1.54, −0.49]	+0.0167 [0.0092, 0.0241]	+0.33 [−0.28, 0.95]	−0.0019 [−0.0052, 0.0015]
C5	−1.33 [−1.82, −0.84]	−1.18 [−1.71, −0.65]	+0.0174 [0.0094, 0.0255]	+0.33 [−0.25, 0.91]	−0.0017 [−0.0054, 0.0019]
C6	+0.11 [−0.16, 0.37]	+0.08 [−0.20, 0.36]	−0.0074 [−0.0095, −0.0052]	−0.13 [−0.59, 0.32]	−0.0018 [−0.0028, −0.0008]
C7	−0.06 [−0.22, 0.11]	−0.06 [−0.24, 0.13]	−0.0036 [−0.0046, −0.0027]	−0.16 [−0.48, 0.16]	−0.0008 [−0.0012, −0.0003]
C8	−0.06 [−0.26, 0.14]	−0.08 [−0.30, 0.13]	−0.0066 [−0.0082, −0.0050]	−0.14 [−0.54, 0.27]	−0.0017 [−0.0024, −0.0010]
I1	+0.16 [−0.15, 0.47]	+0.13 [−0.20, 0.47]	−0.0038 [−0.0057, −0.0018]	+0.03 [−0.56, 0.62]	−0.0010 [−0.0020, −0.0001]
I2	+0.17 [−0.15, 0.48]	+0.16 [−0.17, 0.50]	−0.0008 [−0.0020, 0.0004]	+0.005 [−0.58, 0.59]	−0.0001 [−0.0007, 0.0005]

Table 50: UHCS fold 5: CenWrite effects relative to three baselines. Upper: equal-shot effects with approximate 95% simultaneous max-$|T|$ intervals (15 reported endpoints; original 20-endpoint adjustment retained within fold). Lower: shot-wise effects with unadjusted 95% pointwise intervals. Positive values favor CenWrite. Accuracy/F1/ECE use percentage points; NLL/Brier use native units.

Baseline shot	Accuracy (pp)	Macro-F1 (pp)	NLL	ECE (pp)	Brier
Equal-shot effects: simultaneous intervals, 15 reported endpoints, original adjustment retained					
LoRA Mean	−0.03 [−0.63, 0.58]	+0.08 [−0.55, 0.71]	+0.0541 [0.0419, 0.0662]	+1.06 [0.42, 1.69]	+0.0124 [0.0069, 0.0179]
BOFT Mean	+1.66 [0.67, 2.65]	+2.45 [1.39, 3.52]	+0.2872 [0.2544, 0.3199]	+5.40 [4.52, 6.28]	+0.0685 [0.0566, 0.0804]
ReFTA Mean	+1.79 [0.78, 2.80]	+2.29 [1.21, 3.38]	+0.0856 [0.0662, 0.1049]	+1.31 [0.55, 2.08]	+0.0277 [0.0179, 0.0376]
Individual shots: pointwise intervals, no multiplicity adjustment					
LoRA 1-shot	+0.40 [−0.20, 1.01]	+0.38 [−0.26, 1.02]	+0.0470 [0.0372, 0.0568]	+1.15 [0.52, 1.79]	+0.0148 [0.0100, 0.0197]
LoRA 5-shot	−0.45 [−1.07, 0.17]	−0.22 [−0.88, 0.43]	+0.0612 [0.0491, 0.0734]	+0.96 [0.37, 1.55]	+0.0100 [0.0046, 0.0154]
BOFT 1-shot	+2.13 [1.10, 3.17]	+2.58 [1.47, 3.67]	+0.2201 [0.1950, 0.2455]	+4.81 [3.93, 5.68]	+0.0606 [0.0496, 0.0716]
BOFT 5-shot	+1.19 [0.34, 2.04]	+2.33 [1.40, 3.28]	+0.3543 [0.3208, 0.3894]	+6.00 [5.20, 6.81]	+0.0765 [0.0652, 0.0881]
ReFTA 1-shot	+2.54 [1.64, 3.47]	+3.00 [2.00, 4.01]	+0.0803 [0.0624, 0.0984]	+1.23 [0.41, 2.04]	+0.0296 [0.0207, 0.0386]
ReFTA 5-shot	+1.04 [0.17, 1.92]	+1.59 [0.67, 2.52]	+0.0908 [0.0728, 0.1087]	+1.40 [0.73, 2.06]	+0.0259 [0.0173, 0.0346]

Table 51: UHCS fold 31: CenWrite effects relative to three baselines. Upper: equal-shot effects with approximate 95% simultaneous max-$|T|$ intervals (15 reported endpoints; original 20-endpoint adjustment retained within fold). Lower: shot-wise effects with unadjusted 95% pointwise intervals. Positive values favor CenWrite. Accuracy/F1/ECE use percentage points; NLL/Brier use native units.

Baseline shot	Accuracy (pp)	Macro-F1 (pp)	NLL	ECE (pp)	Brier
Equal-shot effects: simultaneous intervals, 15 reported endpoints, original adjustment retained					
LoRA Mean	+0.68 [0.16, 1.21]	+0.83 [0.27, 1.39]	+0.0954 [0.0840, 0.1067]	+1.69 [1.14, 2.24]	+0.0255 [0.0210, 0.0300]
BOFT Mean	+2.71 [1.84, 3.59]	+3.32 [2.35, 4.29]	+0.3681 [0.3351, 0.4011]	+6.49 [5.60, 7.38]	+0.0856 [0.0741, 0.0971]
ReFTA Mean	+1.25 [0.40, 2.11]	+1.91 [0.99, 2.84]	+0.0892 [0.0676, 0.1108]	+1.15 [0.41, 1.90]	+0.0219 [0.0126, 0.0311]
Individual shots: pointwise intervals, no multiplicity adjustment					
LoRA 1-shot	−0.19 [−0.69, 0.31]	−0.09 [−0.61, 0.45]	+0.0510 [0.0431, 0.0590]	+1.32 [0.75, 1.88]	+0.0133 [0.0099, 0.0170]
LoRA 5-shot	+1.56 [1.03, 2.09]	+1.76 [1.20, 2.30]	+0.1398 [0.1273, 0.1523]	+2.06 [1.54, 2.59]	+0.0377 [0.0328, 0.0425]
BOFT 1-shot	+1.83 [0.94, 2.73]	+2.42 [1.44, 3.41]	+0.2600 [0.2329, 0.2879]	+5.41 [4.56, 6.28]	+0.0666 [0.0559, 0.0775]
BOFT 5-shot	+3.60 [2.90, 4.29]	+4.22 [3.47, 4.96]	+0.4762 [0.4466, 0.5058]	+7.58 [6.87, 8.27]	+0.1045 [0.0957, 0.1134]
ReFTA 1-shot	+1.35 [0.46, 2.24]	+2.33 [1.37, 3.28]	+0.0705 [0.0515, 0.0896]	+0.85 [0.08, 1.64]	+0.0198 [0.0109, 0.0287]

Continued on the following page.

UHCS fold 31: continuation of Table 51

Baseline shot	Accuracy (pp)	Macro-F1 (pp)	NLL	ECE (pp)	Brier
ReFTA 5-shot	+1.16 [0.45, 1.88]	+1.50 [0.73, 2.27]	+0.1079 [0.0900, 0.1262]	+1.46 [0.87, 2.05]	+0.0240 [0.0169, 0.0311]

Table 52: UHCS fold 35: CenWrite effects relative to three baselines. Upper: equal-shot effects with approximate 95% simultaneous max-$|T|$ intervals (15 reported endpoints; original 20-endpoint adjustment retained within fold). Lower: shot-wise effects with unadjusted 95% pointwise intervals. Positive values favor CenWrite. Accuracy/F1/ECE use percentage points; NLL/Brier use native units.

Baseline shot	Accuracy (pp)	Macro-F1 (pp)	NLL	ECE (pp)	Brier
Equal-shot effects: simultaneous intervals, 15 reported endpoints, original adjustment retained					
LoRA Mean	+0.29 [−0.26, 0.83]	+0.55 [−0.01, 1.11]	+0.0649 [0.0526, 0.0772]	+1.38 [0.79, 1.97]	+0.0165 [0.0112, 0.0217]
BOFT Mean	+1.29 [0.38, 2.20]	+1.80 [0.81, 2.79]	+0.3150 [0.2844, 0.3455]	+5.61 [4.87, 6.35]	+0.0656 [0.0544, 0.0768]
ReFTA Mean	+0.05 [−0.79, 0.89]	+0.51 [−0.41, 1.43]	+0.0720 [0.0509, 0.0931]	+0.66 [−0.07, 1.39]	+0.0108 [0.0019, 0.0196]
Individual shots: pointwise intervals, no multiplicity adjustment					
LoRA 1-shot	−0.24 [−0.72, 0.24]	−0.05 [−0.56, 0.48]	+0.0352 [0.0267, 0.0435]	+0.96 [0.37, 1.55]	+0.0078 [0.0041, 0.0115]
LoRA 5-shot	+0.81 [0.25, 1.36]	+1.15 [0.57, 1.71]	+0.0945 [0.0817, 0.1075]	+1.81 [1.24, 2.36]	+0.0252 [0.0197, 0.0307]
BOFT 1-shot	+1.58 [0.60, 2.55]	+2.13 [1.06, 3.20]	+0.2402 [0.2146, 0.2660]	+5.13 [4.32, 5.97]	+0.0615 [0.0506, 0.0728]
BOFT 5-shot	+1.00 [0.27, 1.75]	+1.47 [0.71, 2.25]	+0.3898 [0.3614, 0.4184]	+6.10 [5.46, 6.74]	+0.0696 [0.0604, 0.0788]
ReFTA 1-shot	+0.35 [−0.50, 1.19]	+0.81 [−0.16, 1.78]	+0.0470 [0.0275, 0.0657]	+0.07 [−0.70, 0.85]	+0.0087 [0.0001, 0.0171]
ReFTA 5-shot	−0.25 [−0.89, 0.40]	+0.21 [−0.45, 0.90]	+0.0970 [0.0789, 0.1165]	+1.25 [0.67, 1.83]	+0.0128 [0.0057, 0.0201]

Table 53: UHCS fold 5: component effects by support size, with unadjusted 95% pointwise intervals. Accuracy/F1/ECE use percentage points; NLL/Brier use native units. Signs and interactions follow Table 45.

Contrast shot	Accuracy (pp)	Macro-F1 (pp)	NLL	ECE (pp)	Brier
C1 1-shot	+0.40 [−0.20, 1.01]	+0.38 [−0.26, 1.02]	+0.0470 [0.0372, 0.0568]	+1.15 [0.52, 1.79]	+0.0148 [0.0100, 0.0197]
C1 5-shot	−0.45 [−1.07, 0.17]	−0.22 [−0.88, 0.43]	+0.0612 [0.0491, 0.0734]	+0.96 [0.37, 1.55]	+0.0100 [0.0046, 0.0154]
C2 1-shot	+0.35 [−0.25, 0.96]	+0.29 [−0.35, 0.94]	+0.0509 [0.0406, 0.0613]	+1.47 [0.79, 2.15]	+0.0159 [0.0109, 0.0210]
C2 5-shot	−0.29 [−0.92, 0.34]	−0.03 [−0.69, 0.63]	+0.0645 [0.0520, 0.0772]	+0.97 [0.39, 1.54]	+0.0105 [0.0050, 0.0161]
C3 1-shot	+0.34 [−0.29, 0.99]	+0.30 [−0.36, 0.97]	+0.0551 [0.0441, 0.0663]	+1.33 [0.63, 2.03]	+0.0177 [0.0123, 0.0232]
C3 5-shot	−0.60 [−1.26, 0.05]	−0.30 [−1.01, 0.38]	+0.0678 [0.0547, 0.0811]	+1.05 [0.44, 1.67]	+0.0107 [0.0049, 0.0166]
C4 1-shot	−0.53 [−0.97, −0.10]	−0.62 [−1.08, −0.16]	+0.0205 [0.0141, 0.0267]	+0.82 [0.28, 1.36]	+0.0035 [0.0005, 0.0064]
C4 5-shot	−1.80 [−2.28, −1.31]	−1.41 [−1.91, −0.90]	+0.0128 [0.0065, 0.0194]	−0.15 [−0.64, 0.35]	−0.0072 [−0.0101, −0.0043]
C5 1-shot	−0.74 [−1.18, −0.30]	−0.90 [−1.36, −0.43]	+0.0220 [0.0152, 0.0288]	+0.94 [0.39, 1.49]	+0.0043 [0.0012, 0.0075]
C5 5-shot	−1.92 [−2.40, −1.44]	−1.45 [−1.97, −0.94]	+0.0129 [0.0060, 0.0200]	−0.28 [−0.80, 0.25]	−0.0078 [−0.0110, −0.0046]
C6 1-shot	+0.06 [−0.17, 0.29]	+0.07 [−0.17, 0.33]	−0.0081 [−0.0100, −0.0063]	−0.17 [−0.59, 0.25]	−0.0029 [−0.0038, −0.0021]
C6 5-shot	+0.15 [−0.08, 0.38]	+0.08 [−0.15, 0.32]	−0.0066 [−0.0084, −0.0049]	−0.09 [−0.49, 0.32]	−0.0007 [−0.0016, 0.0001]
C7 1-shot	+0.05 [−0.10, 0.19]	+0.08 [−0.09, 0.25]	−0.0039 [−0.0047, −0.0031]	−0.32 [−0.61, −0.02]	−0.0011 [−0.0015, −0.0007]
C7 5-shot	−0.16 [−0.31, −0.01]	−0.19 [−0.35, −0.04]	−0.0033 [−0.0042, −0.0025]	−0.003 [−0.32, 0.31]	−0.0005 [−0.0009, −0.0001]
C8 1-shot	−0.15 $[-0.31, 1.0\times10^{-7}]$	−0.20 [−0.37, −0.04]	−0.0066 [−0.0080, −0.0052]	−0.05 [−0.41, 0.32]	−0.0021 [−0.0027, −0.0014]
C8 5-shot	+0.03 [−0.17, 0.22]	+0.04 [−0.17, 0.24]	−0.0066 [−0.0079, −0.0053]	−0.22 [−0.55, 0.12]	−0.0014 [−0.0019, −0.0008]
I1 1-shot	+0.01 [−0.24, 0.26]	−0.01 [−0.29, 0.27]	−0.0042 [−0.0058, −0.0027]	+0.14 [−0.36, 0.66]	−0.0018 [−0.0026, −0.0010]
I1 5-shot	+0.31 [0.03, 0.59]	+0.27 [−0.02, 0.57]	−0.0033 [−0.0051, −0.0016]	−0.09 [−0.61, 0.43]	−0.0002 [−0.0011, 0.0007]
I2 1-shot	+0.21 [−0.06, 0.48]	+0.28 [−0.01, 0.57]	−0.0015 [−0.0025, −0.0005]	−0.12 [−0.67, 0.44]	−0.0008 [−0.0013, −0.0004]
I2 5-shot	+0.12 [−0.16, 0.41]	+0.05 [−0.25, 0.35]	−0.00005 [−0.0012, 0.0011]	+0.13 [−0.39, 0.65]	+0.0006 [0.00003, 0.0012]

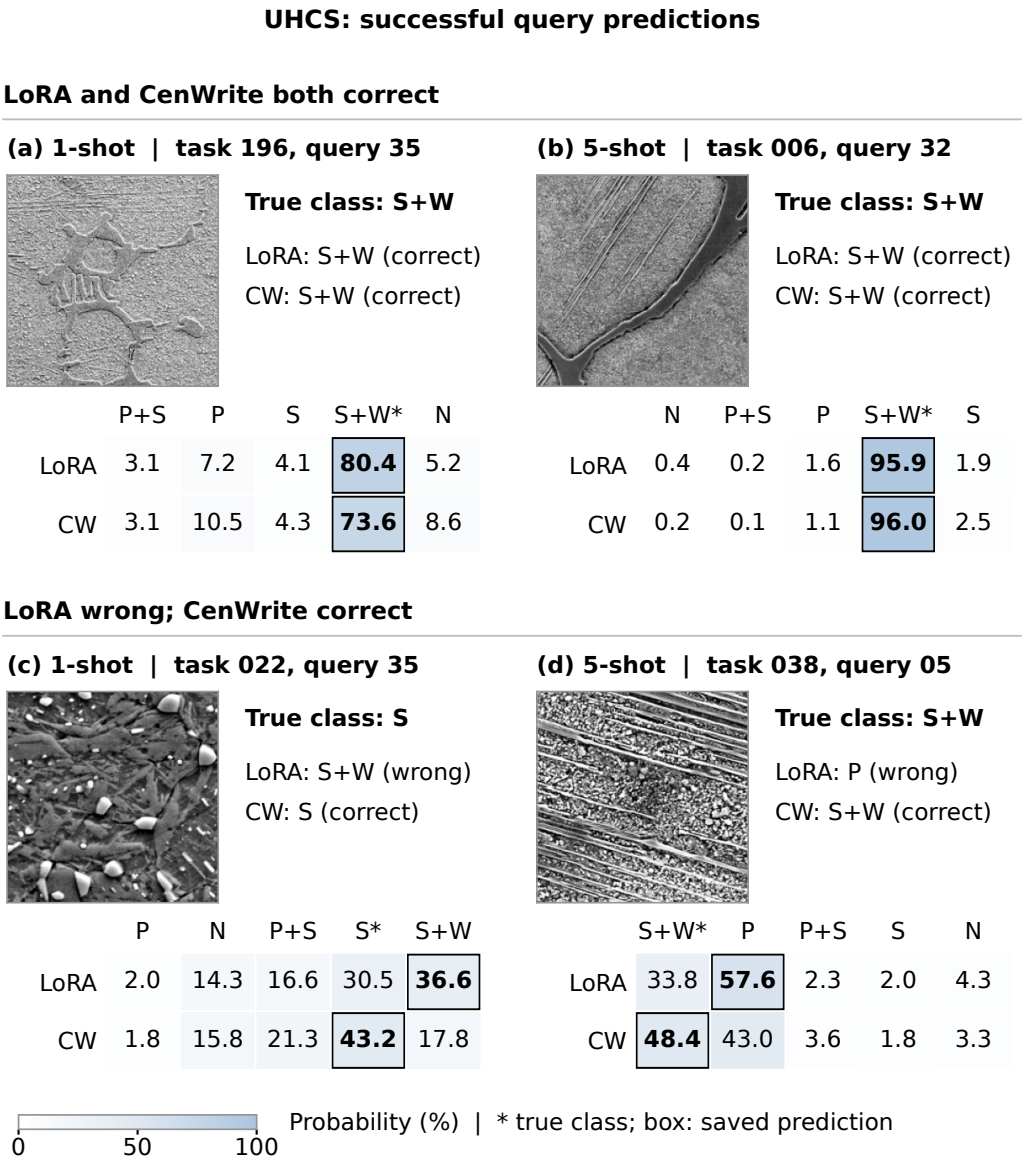


Figure 10: UHCS successful query predictions. Panels (a)–(d) show shared correct predictions and LoRA errors corrected by CenWrite in 1-/5-shot conditions. Exported model inputs accompany saved five-class probabilities for LoRA and CenWrite on a shared 0–100% scale; boxes mark predictions and asterisks mark truth. Four selected conditions are displayed. Appendix H gives the selection rules and complete records.

F NFFA MATERIAL-MORPHOLOGY EVALUATION

F.1 DATA AND EPISODE PROTOCOL

We use NFFA-EUROPE 100% SEM version 2.0 (Aversa et al., 2018), retaining fibres, films/coated surfaces, nanowires, particles, patterned surfaces, porous sponges and powder. These are morphology labels, not compositions or property levels. Fixed preparation removes the lower 30% of each image, excludes reviewed residual overlays/unsuitable content and groups detected duplicates or overlapping views. The resulting 11,540 images form 9,753 detected families, assigned wholly to globally disjoint support/query pools. Families group images rather than documented specimens; undetected related views and acquisition effects can remain.

Each episode chooses five of seven classes, one or five support images and 15 queries per class. All eight reported methods/variants share 200 episodes per shot, nested supports, class order and query identities. Each run resets the classifier/adaptation parameters of the same pretrained DINOv3 ViT-B/16 and uses support-only adaptation for 50 AdamW steps. Method settings transfer from fixed EuroSAT configurations without NFFA query-based selection. CenWrite/LoRA use rank 16 and $\alpha = 16$ (unit scaling). After the source crop, query preprocessing resizes the short edge to 256 and center-crops to 224 pixels.

Means weight shots equally. The 20,000 paired bootstrap draws (seed 2026092301) share sampled episode indices across methods and nested shots. Fixed-standard-error max-$|T|$ intervals provide approximate 95% simultaneous coverage. The 15 reported baseline endpoints (LoRA, BOFT and ReFTA; five metrics) retain the original, larger 20-endpoint adjustment without recalculating its threshold. A separate family covers 50 component endpoints (eight direct/two interaction contrasts). Families are not pooled across datasets; per-shot percentile intervals are unadjusted. Intervals condition on fixed image pools and episodes, excluding uncertainty about specimens, acquisition sessions and pool selection.

F.2 COMPLETE RESULTS AND COMPONENT COMPARISONS

CenWrite reduces NLL by 0.00962 $[0.00485, 0.01439]$ and Brier by 0.00267 $[0.00048, 0.00486]$ relative to LoRA under simultaneous intervals. Both centered-route contrasts ($F - NC$, $NB - L$) reduce NLL/Brier, supporting the role of centered factor learning in these probability improvements. Positive NLL/Brier interactions show a more favorable residual effect under the full rule than under the corresponding uncentered or uncompensated rule. All methods, metrics, shots and component contrasts are reported in the tables below.

Table 54: NFFA means on 200 matched five-way tasks per support size. Accuracy/F1/ECE are percentages; NLL/Brier use native units. Bold/underline: best/second-best displayed means, including ties.

1-shot Metric	L	BOFT	ReFTA	NC	P	NP	NB	CenWrite
Accuracy ↑	58.18	55.42	54.97	58.13	58.28	58.25	**58.33**	<u>58.32</u>
Macro-F1 ↑	57.00	53.80	53.45	56.94	57.13	57.09	<u>57.18</u>	**57.19**
NLL ↓	1.1173	1.2214	1.1890	1.1190	1.1151	1.1163	**1.1129**	<u>1.1134</u>
ECE ↓	**15.04**	16.80	15.52	15.26	15.18	15.19	15.12	<u>15.08</u>
Brier ↓	0.5498	0.5931	0.5873	0.5504	<u>0.5486</u>	0.5489	**0.5485**	**0.5485**
5-shot Metric	L	BOFT	ReFTA	NC	P	NP	NB	CenWrite
Accuracy ↑	<u>78.01</u>	76.08	75.31	77.95	**78.23**	**78.23**	77.93	77.87
Macro-F1 ↑	<u>77.98</u>	75.99	75.25	77.91	**78.19**	**78.19**	77.91	77.84
NLL ↓	0.6495	0.7555	0.7078	0.6523	<u>0.6340</u>	0.6365	**0.6330**	0.6342
ECE ↓	11.40	13.36	12.16	11.56	11.47	11.54	**11.25**	<u>11.32</u>
Brier ↓	0.3155	0.3538	0.3491	0.3165	**0.3095**	<u>0.3103</u>	0.3114	0.3115

Table 55: NFFA variant and contrast definitions. Direct effects use metric utility; positive values favor the first variant. Interaction signs measure changes in a component effect.

Symbol	Variant
F	Full CenWrite
L	Ordinary LoRA, without boundary residual
NC	Uncentered low-rank update with boundary residual
P	Centering only, without compensation or boundary residual
NP	Centered update with boundary residual, without compensation
NB	Centered update with compensation, without boundary residual
Contrast	**Linear combination of metric utilities**
C1	$F - L$
C2	$F - NC$
C3	$NB - L$
C4	$F - NP$
C5	$NB - P$
C6	$F - NB$
C7	$NC - L$
C8	$NP - P$
I1	$(F - NC) - (NB - L)$
I2	$(F - NP) - (NB - P)$

Table 56: NFFA baseline effects averaged equally over both support sizes, with approximate 95% simultaneous intervals retaining the original 20-endpoint adjustment for the 15 reported performance endpoints. Positive effects favor CenWrite. NLL and Brier use native units; other effects use percentage points.

Baseline	Accuracy (pp)	Macro-F1 (pp)	NLL	ECE (pp)	Brier
LoRA	0.00 [−0.33, 0.33]	+0.03 [−0.32, 0.37]	+0.0096 [0.0049, 0.0144]	+0.02 [−0.43, 0.47]	+0.0027 [0.0005, 0.0049]
BOFT	+2.35 [1.68, 3.02]	+2.62 [1.91, 3.32]	+0.1147 [0.0990, 0.1304]	+1.88 [1.19, 2.57]	+0.0435 [0.0371, 0.0498]
ReFTA	+2.96 [2.25, 3.67]	+3.16 [2.42, 3.91]	+0.0746 [0.0610, 0.0882]	+0.64 [0.06, 1.22]	+0.0382 [0.0314, 0.0450]

Table 57: All 30 NFFA baseline effects by support size, with unadjusted 95% pointwise intervals. Positive effects favor CenWrite. NLL and Brier use native units; other effects use percentage points.

Baseline	Shot	Accuracy (pp)	Macro-F1 (pp)	NLL	ECE (pp)	Brier
LoRA	1	+0.14 [−0.16, 0.44]	+0.19 [−0.13, 0.50]	+0.0039 [0.0007, 0.0072]	−0.04 [−0.48, 0.41]	+0.0013 [−0.0003, 0.0029]
LoRA	5	−0.14 [−0.45, 0.17]	−0.14 [−0.45, 0.18]	+0.0153 [0.0102, 0.0206]	+0.07 [−0.34, 0.48]	+0.0040 [0.0016, 0.0064]
BOFT	1	+2.90 [2.24, 3.55]	+3.39 [2.67, 4.10]	+0.1080 [0.0954, 0.1208]	+1.72 [1.07, 2.38]	+0.0447 [0.0393, 0.0501]
BOFT	5	+1.79 [1.25, 2.34]	+1.85 [1.30, 2.40]	+0.1213 [0.1073, 0.1355]	+2.04 [1.49, 2.59]	+0.0423 [0.0366, 0.0479]
ReFTA	1	+3.35 [2.67, 4.04]	+3.73 [3.00, 4.46]	+0.0757 [0.0632, 0.0883]	+0.44 [−0.14, 1.02]	+0.0388 [0.0326, 0.0452]
ReFTA	5	+2.57 [1.98, 3.16]	+2.59 [1.99, 3.21]	+0.0736 [0.0622, 0.0857]	+0.83 [0.33, 1.34]	+0.0376 [0.0319, 0.0435]

Table 58: NFFA equal-shot component effects with approximate 95% simultaneous intervals over 50 endpoints. Accuracy/F1/ECE use percentage points; NLL/Brier use native units. Signs follow Table 55.

Contrast	Accuracy (pp)	Macro-F1 (pp)	NLL	ECE (pp)	Brier
C1	0.00 [−0.36, 0.36]	+0.03 [−0.34, 0.39]	+0.0096 [0.0045, 0.0147]	+0.02 [−0.47, 0.50]	+0.0027 [0.0003, 0.0050]
C2	+0.06 [−0.33, 0.44]	+0.09 [−0.31, 0.48]	+0.0119 [0.0063, 0.0174]	+0.21 [−0.30, 0.71]	+0.0035 [0.0009, 0.0060]
C3	+0.03 [−0.36, 0.42]	+0.05 [−0.35, 0.46]	+0.0104 [0.0046, 0.0163]	+0.04 [−0.48, 0.55]	+0.0027 [0.00002, 0.0054]
C4	−0.14 [−0.41, 0.13]	−0.12 [−0.41, 0.17]	+0.0027 [−0.0005, 0.0058]	+0.16 [−0.28, 0.61]	−0.0004 [−0.0020, 0.0012]
C5	−0.12 [−0.41, 0.17]	−0.11 [−0.42, 0.19]	+0.0016 [−0.0018, 0.0050]	+0.14 [−0.31, 0.60]	−0.0009 [−0.0027, 0.0008]
C6	−0.03 [−0.18, 0.12]	−0.03 [−0.18, 0.13]	−0.0008 [−0.0018, 0.0001]	−0.02 [−0.36, 0.32]	−0.00003 [−0.0005, 0.0004]
C7	−0.06 [−0.17, 0.05]	−0.06 [−0.17, 0.05]	−0.0022 [−0.0029, −0.0016]	−0.19 [−0.44, 0.06]	−0.0008 [−0.0011, −0.0005]
C8	−0.02 [−0.15, 0.12]	−0.02 [−0.17, 0.12]	−0.0019 [−0.0027, −0.0011]	−0.04 [−0.32, 0.24]	−0.0006 [−0.0009, −0.0002]
I1	+0.02 [−0.16, 0.20]	+0.03 [−0.15, 0.22]	+0.0014 [0.0007, 0.0021]	+0.17 [−0.25, 0.59]	+0.0008 [0.0004, 0.0011]
I2	−0.02 [−0.21, 0.17]	−0.01 [−0.21, 0.19]	+0.0010 [0.0006, 0.0015]	+0.02 [−0.40, 0.44]	+0.0005 [0.0003, 0.0008]

F.3 Real query predictions

The four successful display conditions comprise one-/five-shot examples of shared correct predictions and LoRA errors corrected by CenWrite (Figure 11). Displays reconstruct the recorded query transform and use saved predictions, showing LoRA and CenWrite's five-class probabilities on a common 0–100% scale. These four conditions use three distinct source images: panels (a)/(b) show the same complete query image under 1-/5-shot support, respectively. The source/crop/input mapping identifies each displayed image, and Appendix H records the case-selection rules. Complete quantitative comparisons appear in the preceding tables.

Table 59: NFFA component effects by support size, with unadjusted 95% pointwise intervals. Accuracy/F1/ECE use percentage points; NLL/Brier use native units. Signs follow Table 55.

Contrast	Shot	Accuracy (pp)	Macro-F1 (pp)	NLL	ECE (pp)	Brier
C1	1	+0.14 [−0.16, 0.44]	+0.19 [−0.13, 0.50]	+0.0039 [0.0007, 0.0072]	−0.04 [−0.48, 0.41]	+0.0013 [−0.0003, 0.0029]
C1	5	−0.14 [−0.45, 0.17]	−0.14 [−0.45, 0.18]	+0.0153 [0.0102, 0.0206]	+0.07 [−0.34, 0.48]	+0.0040 [0.0016, 0.0064]
C2	1	+0.19 [−0.13, 0.51]	+0.24 [−0.09, 0.58]	+0.0056 [0.0021, 0.0092]	+0.18 [−0.31, 0.68]	+0.0020 [0.0002, 0.0037]
C2	5	−0.07 [−0.39, 0.25]	−0.07 [−0.40, 0.26]	+0.0181 [0.0126, 0.0237]	+0.23 [−0.19, 0.64]	+0.0050 [0.0024, 0.0075]
C3	1	+0.15 [−0.17, 0.47]	+0.18 [−0.16, 0.52]	+0.0044 [0.0005, 0.0082]	−0.08 [−0.55, 0.39]	+0.0013 [−0.0006, 0.0032]
C3	5	−0.08 [−0.41, 0.25]	−0.07 [−0.41, 0.27]	+0.0165 [0.0108, 0.0224]	+0.15 [−0.24, 0.54]	+0.0041 [0.0015, 0.0068]
C4	1	+0.07 [−0.19, 0.33]	+0.10 [−0.18, 0.38]	+0.0030 [0.0007, 0.0053]	+0.11 [−0.33, 0.54]	+0.0005 [−0.0007, 0.0016]
C4	5	−0.35 [−0.57, −0.13]	−0.35 [−0.56, −0.12]	+0.0023 [−0.0005, 0.0051]	+0.21 [−0.14, 0.57]	−0.0013 [−0.0027, 0.0001]
C5	1	+0.05 [−0.21, 0.31]	+0.05 [−0.23, 0.33]	+0.0022 [−0.0002, 0.0046]	+0.06 [−0.35, 0.47]	+0.0001 [−0.0011, 0.0013]
C5	5	−0.29 [−0.54, −0.05]	−0.28 [−0.53, −0.03]	+0.0010 [−0.0020, 0.0041]	+0.22 [−0.15, 0.60]	−0.0019 [−0.0035, −0.0004]
C6	1	−0.01 [−0.15, 0.13]	+0.01 [−0.14, 0.16]	−0.0005 [−0.0013, 0.0003]	+0.04 [−0.29, 0.37]	+0.00004 [−0.0004, 0.0005]
C6	5	−0.06 [−0.19, 0.07]	−0.07 [−0.20, 0.07]	−0.0012 [−0.0019, −0.0005]	−0.08 [−0.35, 0.19]	−0.0001 [−0.0005, 0.0003]
C7	1	−0.05 [−0.16, 0.07]	−0.06 [−0.17, 0.05]	−0.0017 [−0.0022, −0.0012]	−0.22 [−0.47, 0.03]	−0.0006 [−0.0009, −0.0004]
C7	5	−0.07 [−0.15, 0.02]	−0.07 [−0.16, 0.02]	−0.0028 [−0.0033, −0.0022]	−0.16 [−0.33, 0.03]	−0.0009 [−0.0012, −0.0007]
C8	1	−0.03 [−0.17, 0.11]	−0.05 [−0.19, 0.10]	−0.0013 [−0.0019, −0.0006]	−0.01 [−0.26, 0.24]	−0.0003 [−0.0007, 0.00001]
C8	5	0.00 [−0.09, 0.09]	0.00 [−0.10, 0.10]	−0.0025 [−0.0031, −0.0019]	−0.07 [−0.31, 0.17]	−0.0008 [−0.0011, −0.0005]
I1	1	+0.04 [−0.13, 0.21]	+0.07 [−0.11, 0.25]	+0.0013 [0.0007, 0.0018]	+0.26 [−0.18, 0.70]	+0.0007 [0.0004, 0.0010]
I1	5	+0.01 [−0.15, 0.16]	0.00 [−0.15, 0.16]	+0.0016 [0.0010, 0.0022]	+0.08 [−0.24, 0.39]	+0.0008 [0.0005, 0.0012]
I2	1	+0.03 [−0.16, 0.22]	+0.05 [−0.15, 0.26]	+0.0008 [0.0004, 0.0011]	+0.05 [−0.37, 0.47]	+0.0004 [0.0002, 0.0006]
I2	5	−0.06 [−0.21, 0.09]	−0.07 [−0.22, 0.08]	+0.0013 [0.0009, 0.0017]	−0.01 [−0.35, 0.32]	+0.0007 [0.0004, 0.0009]

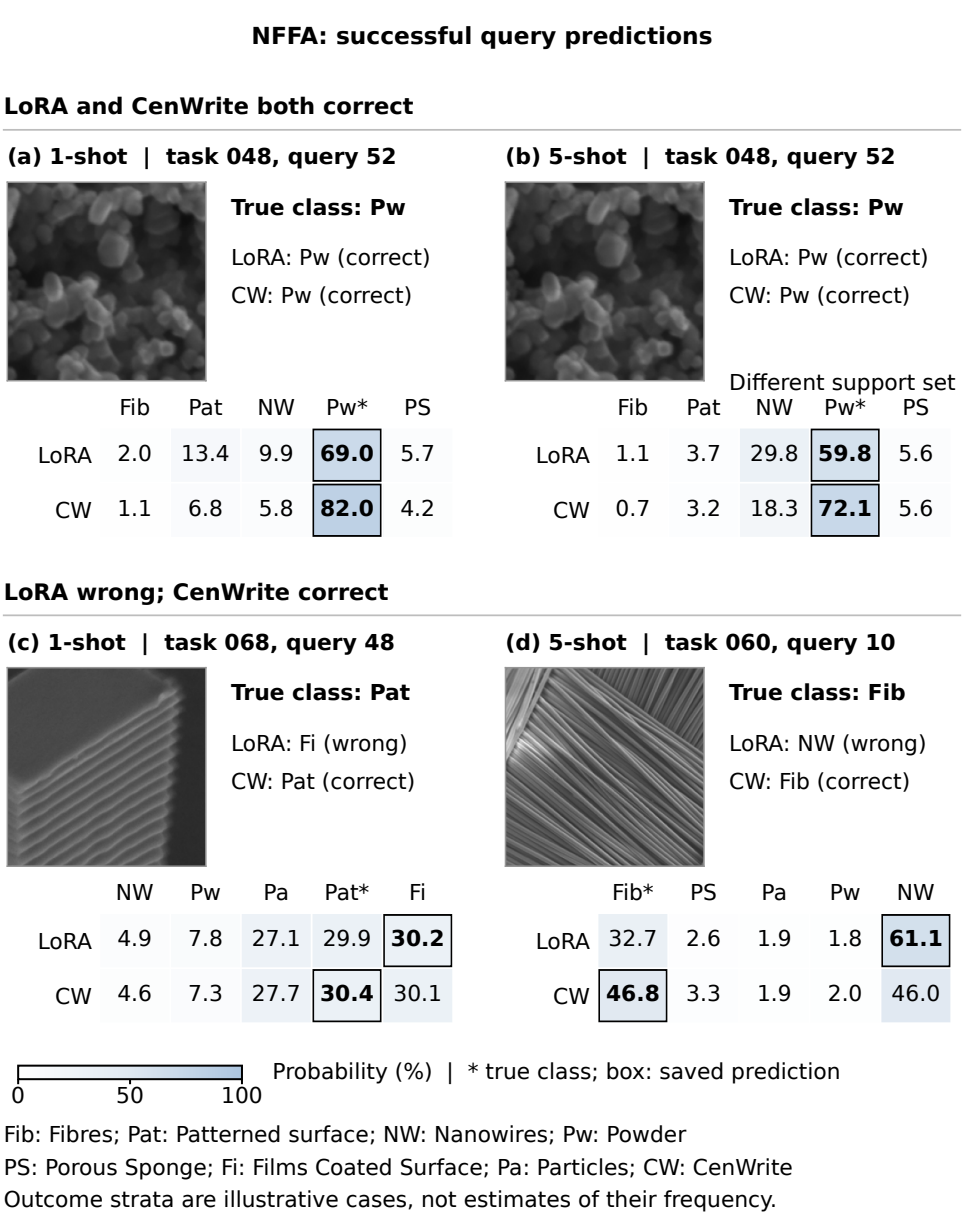


Figure 11: NFFA successful query predictions. Panels (a)–(d) show shared correct predictions and LoRA errors corrected by CenWrite in 1-/5-shot conditions. Exported model inputs accompany saved five-class probabilities for LoRA and CenWrite (0–100%); boxes mark predictions and asterisks mark truth. The same real query image is shown in both panels (a)/(b), under 1-/5-shot support, respectively. The four conditions use three distinct images. Appendix H gives the selection rules; Appendix F reports the complete quantitative comparisons.

G Material Adaptation Diagnostics

Material probes retain all six component variants at both support sizes (Figures 12 and 13). The UHCS probes use the designated primary split, fold 5; they do not cover the other splits. Each setting uses fixed episodes 0, 1, and 2, observed at steps 0, 1, 10, 25, and 50 on the same first training support view. Each episode is first summarized by the median across its eligible projections; lines then show the median of the three episode values, and bands show their minimum and maximum. The same projection is observed repeatedly across states, so neither its states nor the individual projections are

independent tasks. Gradient and mean-response diagnostics use 48 adapted projections per episode, while residual diagnostics use only the 24 output projections containing b. Variants without b have missing residual values, not zeros. The displayed ranges describe three trajectories; they are not confidence intervals. Predictive benefits require the complete paired evaluations.

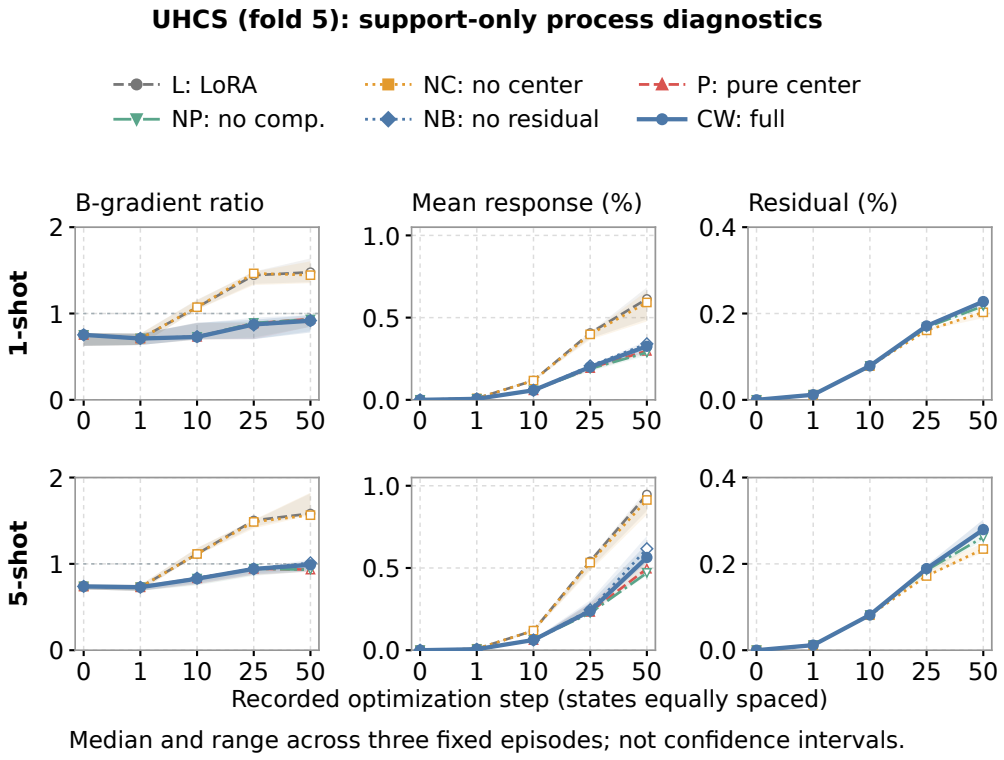


Figure 12: UHCS process diagnostics on the primary split. The first column shows the mean-to-centered B-gradient Frobenius-norm ratio. The other columns show broadcast $\Delta W\mu$ and b norms relative to the full adapted-layer output, in percent. The mean response is implicit in L/NC, omitted in P/NP, and restored with stopped gradients in NB/CW; its diagnostic magnitude does not identify whether compensation is active. Residual curves include only NC, NP, and CW. All curves use the three fixed-episode summaries described above; bands are ranges, not confidence intervals. Recorded states have equal spacing, so connecting lines imply no intermediate observations.

At step 50, CenWrite's mean-to-centered B-gradient norm ratios are 0.916 and 0.989 on UHCS, and 0.653 and 0.679 on NFFA, for one- and five-shot tasks respectively. Ordinary LoRA's corresponding ratios are 1.478, 1.583, 1.185, and 1.137. Table 60 retains every variant and its descriptive range. For centered variants, the mean term is a counterfactual raw-gradient component, not an applied factor gradient. The reference μ stays fixed even as the probe's activation mean changes; Table 61 distinguishes the distance from this reference from the change relative to the initial probe mean. Their initial values need not match because the fixed probe need not reproduce the center-collection views. LoRA's reference is diagnostic only and never enters its training update. These smaller gradient ratios and response norms describe the observed trajectories, but alone they establish neither better predictions nor the causal contribution of an individual component.

H Paired-Shot Material Case Selection

The material examples in Figure 4 were selected after completing the experiments. Matching episode, query index, ground truth and class order across methods and nested support sizes gives 10,000 paired episode–query records for UHCS (200 episodes, 50 queries each) and 15,000 for

Table 60: Step-50 material diagnostics: median [minimum, maximum] over three fixed-episode projection medians. Mean-response/residual norms are percentages of the layer-output norm; the gradient ratio is unitless. Ranges are descriptive; dashes mean no residual parameter.

Setting	Variant	B mean / centered	Mean response (%)	Residual (%)
UHCS (fold 5)		1.478	0.614	
1-shot	L	[1.363, 1.638]	[0.489, 0.684]	—
		1.444	0.592	0.203
	NC	[1.351, 1.605]	[0.477, 0.662]	[0.189, 0.207]
		0.936	0.298	
	P	[0.846, 0.965]	[0.281, 0.307]	—
		0.931	0.288	0.217
	NP	[0.846, 0.955]	[0.269, 0.298]	[0.209, 0.218]
		0.919	0.341	
	NB	[0.795, 0.984]	[0.306, 0.360]	—
		0.916	0.326	0.228
	CW	[0.780, 0.983]	[0.292, 0.340]	[0.216, 0.232]
UHCS (fold 5)		1.583	0.947	
5-shot	L	[1.580, 1.817]	[0.847, 0.978]	—
		1.564	0.914	0.234
	NC	[1.563, 1.825]	[0.824, 0.953]	[0.226, 0.248]
		0.934	0.496	
	P	[0.908, 1.001]	[0.472, 0.570]	—
		0.932	0.470	0.262
	NP	[0.915, 0.969]	[0.452, 0.543]	[0.254, 0.278]
		1.012	0.617	
	NB	[0.936, 1.053]	[0.568, 0.695]	—
		0.989	0.564	0.279
	CW	[0.914, 1.031]	[0.533, 0.647]	[0.279, 0.303]
		1.185	0.497	
NFFA 1-shot	L	[1.121, 1.199]	[0.460, 0.525]	—
		1.171	0.485	0.238
	NC	[1.124, 1.194]	[0.450, 0.511]	[0.211, 0.248]
		0.643	0.262	
	P	[0.630, 0.663]	[0.211, 0.307]	—
		0.638	0.254	0.248
	NP	[0.632, 0.672]	[0.203, 0.301]	[0.216, 0.258]
		0.665	0.284	
	NB	[0.656, 0.668]	[0.227, 0.331]	—
		0.653	0.273	0.257
	CW	[0.647, 0.674]	[0.217, 0.320]	[0.219, 0.266]
		1.137	0.684	
NFFA 5-shot	L	[1.119, 1.160]	[0.598, 0.799]	—
		1.133	0.667	0.277
	NC	[1.097, 1.150]	[0.585, 0.779]	[0.272, 0.304]
		0.667	0.350	
	P	[0.567, 0.674]	[0.331, 0.363]	—
		0.657	0.338	0.300
	NP	[0.554, 0.676]	[0.320, 0.351]	[0.286, 0.328]
		0.697	0.383	
	NB	[0.570, 0.774]	[0.381, 0.407]	—
		0.679	0.366	0.322
	CW	[0.567, 0.741]	[0.362, 0.382]	[0.295, 0.341]

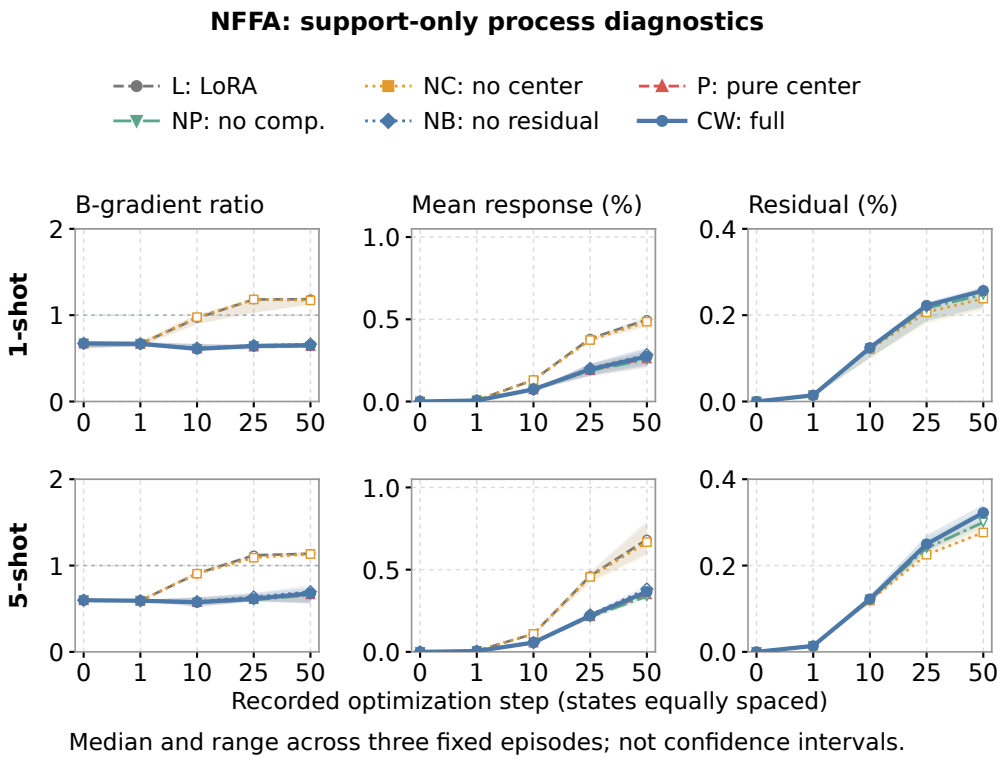


Figure 13: NFFA process diagnostics. Definitions, six variants, and within-episode aggregation match Figure 12. The same metric uses the same vertical scale across the two materials. Lines and bands show the median and range of episodes 0, 1, and 2, without statistical inference. Gradient and mean-response summaries use 48 projections; residual summaries use the 24 eligible output projections, excluding absent parameters. The stored support reference is fixed, not repeatedly estimated from the evolving probe activations. The observed magnitudes describe the adaptation paths and are interpreted alongside the complete component results.

Table 61: Medians of projection norms for the fixed reference μ and recorded support-probe mean $\bar{h}_t$. The reference is fixed throughout adaptation; the initial probe may differ from center-collection views.

Setting	Variant	$\|\mu\|_2$	$\|\bar{h}_0 - \mu\|_2$	$\|\bar{h}_{50} - \mu\|_2$	$\|\bar{h}_{50} - \bar{h}_0\|_2$
UHCS (fold 5)					
1-shot	L	6.460	0.457	1.496	1.416
	CW	6.460	0.457	0.729	0.547
UHCS (fold 5)					
5-shot	L	6.255	0.190	1.863	1.857
	CW	6.255	0.190	0.894	0.859
NFFA 1-shot	L	5.935	$< 10^{-6}$	1.343	1.343
	CW	5.935	$< 10^{-6}$	0.603	0.603
NFFA 5-shot	L	5.487	$< 10^{-6}$	1.536	1.536
	CW	5.487	$< 10^{-6}$	0.592	0.592

NFFA (200 episodes, 75 queries each). These cover 311 and 5,171 distinct query-image indices, respectively: an image can recur across tasks, and each record includes both shots rather than counting them twice. Selection uses saved predictions without additional training, inference or significance tests; aggregate comparisons provide the performance evidence.

The previously selected records remain unchanged: UHCS episode 41/query 23 and NFFA episode 52/query 37. In both support conditions, CenWrite correctly classifies each selected query and has lower per-query NLL and Brier than the retained LoRA, BOFT and ReFTA comparisons. These examples were selected post hoc and were not reselected when the reported comparator set changed. They illustrate successful paired-shot predictions rather than estimating an overall success rate. The figure contains two images in four support conditions; aggregate results provide the performance evidence.

The main display gives complete five-class LoRA/CenWrite vectors; case records retain the eight reported methods. UHCS abbreviations are spheroidite (S), pearlite (P), their mixture (P+S), network (N), and spheroidite+widmanstatten (S+W); NFFA abbreviations are particles (Pa), powder (Pw), fibres (Fib), patterned surfaces (Pat), and porous sponges (PS). NLL is $-\log p_y$, and Brier is $\sum_c (p_c - \mathbf{1}[c = y])^2$, without division by the five classes. Displayed changes are CenWrite minus LoRA. Both methods correctly classify UHCS in both shots, with lower CenWrite NLL and Brier. Both correctly classify NFFA in 1-shot; in 5-shot, LoRA predicts porous sponge and CenWrite predicts the true fibres class, with true-class probabilities 0.3267 and 0.3735. Aggregate ECE and macro-F1 are reported in the complete quantitative evaluation.

H.1 Image provenance and display preprocessing

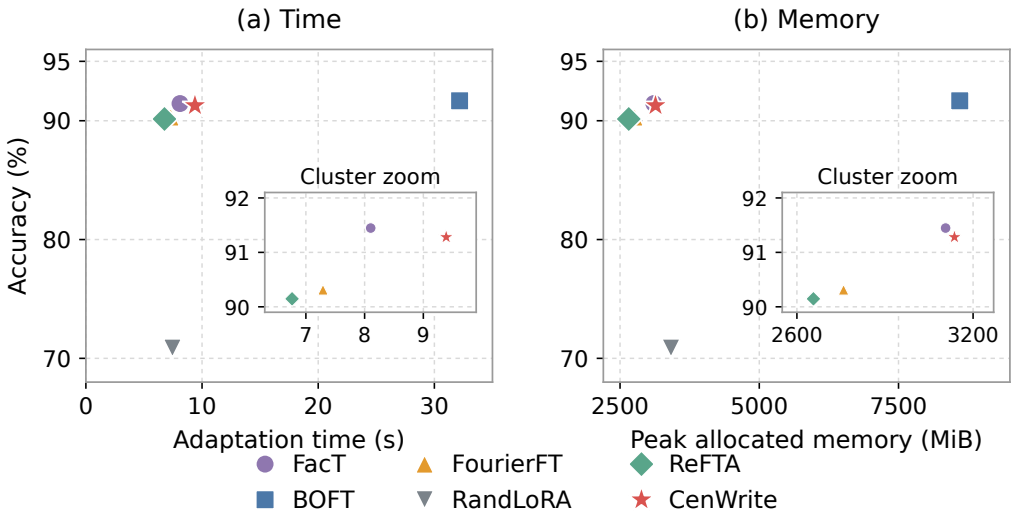
The UHCS example, `micrograph1503.png`, is labelled pearlite+spheroidite and belongs to sample 35. The source is *Ultrahigh Carbon Steel Micrographs* (Hecht et al., 2017), distributed by the NIST Materials Data Repository under CC BY 3.0 US, as recorded in the Materials Data Facility metadata. Preparation retains the upper 645×484 pixels of the 645×522 source, removing bottom metadata. The processing label is AC1 970C 90M FC and the detector is SE; magnification is unrecorded.

The NFFA fibres example comes from *NFFA-EUROPE—100% SEM Dataset*, version 2.0 (Aversa et al., 2018), available through B2SHARE under CC BY 4.0. Preparation retains the upper 1024×537 pixels of the 1024×768 source. Its final query index is 11462; preparation index 74 belongs to an earlier indexing stage. Original filename and image-family identifier remain in the case manifest. The specimen is unknown, and the label describes morphology, not composition or material properties. Both datasets are credited for these modified displays without implying their creators' endorsement.

Display reconstruction resizes the short edge to 256 pixels with bicubic interpolation and center-crops to 224×224, using Pillow 12.3.0. The UHCS/NFFA resized dimensions are 341×256/488×256 and crop boxes are $(58, 16, 282, 240)$/$(132, 16, 356, 240)$, in left/top/right/bottom coordinates. These RGB displays reproduce recorded query geometry rather than newly captured forward tensors. No enhancement, generated content or additional region selection is applied. Each image is shared across shots; probabilities come from the completed runs, not a rerun on the displays.

H.2 Four-domain response displays

The four successful examples in Figure 3 retain the existing DINOv3 ViT-B/16, 1-shot, 50-step predictions. Responses are the input-channel L_2 norm of the true-class logit's input gradient. The original patch aggregation, color map and transparency are retained; each display uses the larger of the paired LoRA/CenWrite response 99th percentiles for normalization, although only CenWrite is shown. The red box marks the highest-mean 16×16 response patch and its radius-one patch neighborhood, clipped at image boundaries; identical coordinates are shown on the input. This provides a deterministic response-location reference. Saved probabilities and response arrays are reused without new inference or gradient computation.



Method	Acc. (%)	F1 (%)	Adapt (s)	Query (s)	Peak (MiB)	Parameters
FacT	91.45	91.37	8.104	0.189	3106.3	65,285
BOFT	**91.68**	**91.58**	32.172	0.870	8599.2	805,637
FourierFT	90.31	90.20	7.291	0.185	2758.9	75,845
RandLoRA	70.86	68.07	7.452	0.199	3414.7	2,685,701
ReFTA	90.15	90.02	**6.765**	**0.174**	**2656.2**	**49,955**
CenWrite	91.28	91.19	9.389	0.193	3136.5	2,381,573

Figure 14: Accuracy and resource costs on EuroSAT 5-shot. Panels (a) and (b) compare adaptation time and peak memory; insets enlarge the closely grouped high-accuracy configurations. The lower panel gives the measured resource summaries. Bold/underline mark the best/second-best displayed values (ties included), not significance.

H.3 Successful material-query displays

The supplementary UHCS and NFFA displays each show four successful conditions across 1-shot and 5-shot support (Figures 10 and 11). They include shared correct predictions and queries correctly classified by CenWrite after LoRA assigns another label. Both displays retain LoRA and CenWrite's complete five-class probability vectors. NFFA panels (a) and (b) repeat the same real image under the two support sizes, with the image fully displayed in each panel. Selection follows the recorded success strata. Dataset-level conclusions use the complete paired quantitative comparisons in Appendices E and F.

I Resource Measurements

Resource measurements use the same first ten EuroSAT 5-shot episodes on an RTX 5090, including the first run. Median adaptation and query times for CenWrite are 9.389 and 0.193 seconds, respectively, with 3,136.5 MiB peak allocated GPU memory and 2,381,573 trainable parameters. Query time covers all 75 queries and memory includes resident model tensors. Accuracy and F1 use all 200 episodes. The measurements are specific to these implementations and this hardware; BOFT uses the portable kernel described in Appendix A.2.

Table 62: CenWrite backbone sensitivity. Panel (a) reports DINOv2 ViT-S/14 means; panel (b) reports DINOv3 ViT-B/16 minus DINOv2 with pointwise 95% intervals. Accuracy/F1/ECE means are percentages and their differences are percentage points. There are 200 pairs per setting.

(a) DINOv2 ViT-S/14 means

Dataset / shot	Acc. ↑	F1 ↑	NLL ↓	ECE ↓	Brier ↓
EuroSAT, 1-shot	68.13	66.02	1.296	22.46	0.510
EuroSAT, 5-shot	87.85	87.65	0.471	9.77	0.196
PlantVillage, 1-shot	85.25	84.23	0.552	10.86	0.232
PlantVillage, 5-shot	96.68	96.61	0.114	3.04	0.051
ISIC2018, 1-shot	33.48	29.85	3.491	50.15	1.103
ISIC2018, 5-shot	51.67	50.62	2.455	36.33	0.802
ChestX, 1-shot	21.69	18.28	4.694	61.06	1.316
ChestX, 5-shot	25.27	23.88	4.560	58.44	1.257

(b) DINOv3 minus DINOv2: differences [95% CI]

Dataset / shot	Acc. ↑	F1 ↑	NLL ↓	ECE ↓	Brier ↓
EuroSAT 1-shot	4.680 [3.373, 6.000]	5.285 [3.798, 6.804]	−0.537 [−0.604, −0.473]	−8.658 [−9.769, −7.542]	−0.122 [−0.142, −0.102]
EuroSAT 5-shot	3.427 [2.700, 4.134]	3.538 [2.786, 4.278]	−0.214 [−0.244, −0.185]	−2.508 [−3.078, −1.941]	−0.064 [−0.075, −0.053]
PlantVillage 1-shot	5.853 [5.026, 6.700]	6.402 [5.448, 7.387]	−0.291 [−0.335, −0.248]	−3.038 [−3.778, −2.321]	−0.099 [−0.111, −0.087]
PlantVillage 5-shot	0.907 [0.613, 1.220]	0.938 [0.627, 1.275]	−0.041 [−0.056, −0.028]	−0.264 [−0.530, −0.013]	−0.014 [−0.019, −0.010]
ISIC2018 1-shot	2.493 [1.560, 3.413]	2.928 [1.942, 3.886]	−1.555 [−1.629, −1.481]	−18.363 [−19.414, −17.299]	−0.219 [−0.233, −0.205]
ISIC2018 5-shot	3.100 [2.260, 3.913]	3.414 [2.573, 4.219]	−1.077 [−1.131, −1.022]	−14.207 [−15.001, −13.396]	−0.160 [−0.172, −0.148]
ChestX 1-shot	−0.333 [−1.027, 0.367]	−0.557 [−1.326, 0.225]	−2.097 [−2.216, −1.979]	−18.107 [−19.307, −16.896]	−0.228 [−0.243, −0.212]
ChestX 5-shot	−0.013 [−0.760, 0.733]	0.172 [−0.541, 0.892]	−2.195 [−2.271, −2.119]	−22.268 [−23.132, −21.413]	−0.256 [−0.267, −0.246]

J Backbone and Rank Sensitivity

Within CenWrite, DINOv3 has higher accuracy than DINOv2 in six settings with positive paired intervals; the two ChestX accuracy differences remain unresolved. NLL and Brier are lower on DINOv3 in all eight settings. This sensitivity comparison changes encoder architecture, capacity, and pretraining together. The seven-method comparison in Appendix D instead evaluates relative method performance within DINOv2 and supports Section 5.3.

Rank sensitivity uses 5-shot tasks with ranks 4, 8, 16, and 32 and $\alpha = r$. Changes are measured relative to rank 16. ISIC and ChestX contrasts each contain 100 matched episodes. On ChestX, rank 16 raises ECE by 11.290 percentage points relative to rank 4 with an interval excluding zero, while the accuracy difference is unresolved. EuroSAT and PlantVillage have 200 saved records for ranks 4, 8, and 32 despite a plan for 100; their full curves are retained as descriptive results. Setting $\alpha = r$ fixes the explicit scale, not the parameter budget.

K Further Connections to Recent ICLR Work

K.1 A close recent comparison: MPS-Tuning

MPS-Tuning provides a close comparison in few-shot visual adaptation (Chen et al., 2026). Its manifold regularizer matches original and adapted CLIP Gram matrices across images and among tokens within each image. A second objective increases same-class image–image and image–text

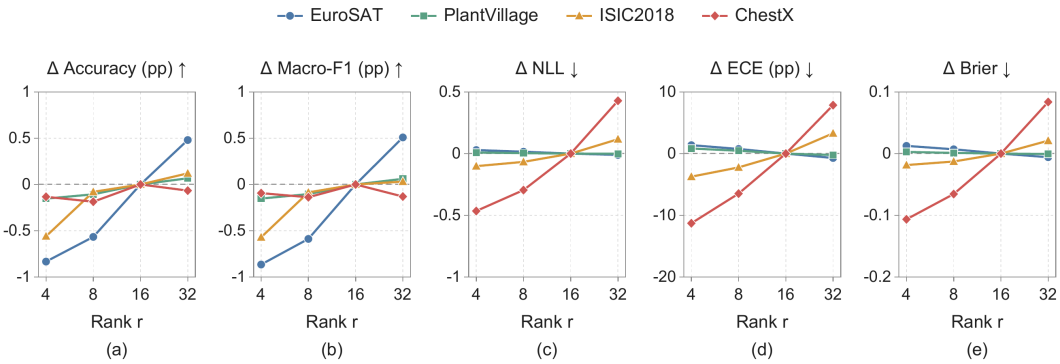


Figure 15: Rank sensitivity relative to rank 16 on the four 5-shot settings. Panels show accuracy, macro-F1, NLL, ECE, and Brier. Accuracy/F1/ECE changes are percentage points. Positive probability changes indicate worse scores. EuroSAT and PlantVillage curves are descriptive because their saved record counts differ from the planned episode counts.

Table 63: Paired rank comparisons on ISIC 2018 and ChestX. Entries are rank 16 minus ranks 4, 8, or 32 with pointwise 95% intervals from 100 matched episodes. Accuracy/F1/ECE differences are percentage points. Signs are opposite to the changes from rank 16 in Figure 15.

Dataset / rank	Acc. ↑	F1 ↑	NLL ↓	ECE ↓	Brier ↓
ISIC2018	0.560	0.571	0.102	3.738	0.019
16 vs. 4	[−0.027, 1.173]	[−0.049, 1.211]	[0.087, 0.117]	[3.034, 4.442]	[0.013, 0.024]
ISIC2018	0.080	0.086	0.067	2.238	0.013
16 vs. 8	[−0.360, 0.507]	[−0.370, 0.532]	[0.058, 0.077]	[1.611, 2.854]	[0.009, 0.016]
ISIC2018	−0.120	−0.031	−0.117	−3.284	−0.021
16 vs. 32	[−0.640, 0.387]	[−0.586, 0.497]	[−0.130, −0.103]	[−3.791, −2.774]	[−0.025, −0.017]
ChestX	0.133	0.094	0.465	11.290	0.106
16 vs. 4	[−0.507, 0.773]	[−0.573, 0.756]	[0.438, 0.494]	[10.564, 12.019]	[0.100, 0.113]
ChestX	0.187	0.138	0.296	6.483	0.066
16 vs. 8	[−0.347, 0.707]	[−0.410, 0.673]	[0.279, 0.313]	[6.006, 6.972]	[0.062, 0.069]
ChestX	0.067	0.131	−0.428	−7.852	−0.084
16 vs. 32	[−0.507, 0.640]	[−0.470, 0.729]	[−0.453, −0.404]	[−8.464, −7.226]	[−0.089, −0.078]

similarities, with a pseudo-forward projection extending it to intermediate layers; prediction mixes original and adapted logits. CenWrite instead preserves the current adapter's shared mean response at each adapted projection. It changes the factor-gradient route without matching pretrained Gram matrices or constraining manifold distance. The methods thus separate retained information from learned discrimination through different quantities and objectives.

MPS-Tuning uses CLIP ViT-B/16, 1/2/4/8/16 labeled examples per class, full test sets, and three seeds across eleven datasets (Chen et al., 2026). Its Table 6 combines the regularizers with LoRA, increasing average accuracy from 71.96 to 72.20 at 1-shot and from 84.17 to 85.13 at 16-shot. These published results use a different protocol from CenWrite's image-only DINOv3 encoder, 5-way episodes, and 50-step support adaptation, and do not establish NLL, ECE, or Brier gains in our setting. Evaluating a combination would require a common backbone, information budget, and selection protocol, with matched adaptation data and held-out queries.

K.2 Information available during few-shot adaptation

Permitted information distinguishes adaptation protocols even at the same label count. Multi-task fine-tuning selects relevant auxiliary tasks (Xu et al., 2024); BECLR combines contrastive learning with optimal-transport alignment at few-shot inference (Poulakakis-Daktylidis & J. Rad, 2024). MetaCoCo distinguishes domain changes from shifts in spurious concept–context correlations (Zhang et al., 2024). A support mean can therefore mix task structure, context, and sampling

effects rather than represent a pure domain signal. CenWrite changes direct factor gradients but retains shared forward responses, without identifying causes or recovering causal representations.

Visual domain prompt generation conditions adaptation on a source knowledge bank and few unlabeled target observations (Chi et al., 2024). L2C adds a side branch and refined text features to frozen CLIP (Chi et al., 2025), while ECALP propagates labels through a graph of text prompts, few-shot examples, and test samples (Li et al., 2025). CenWrite computes centers from labeled support images without using labels in the mean, then optimizes support cross-entropy before reading queries. Our transferred baselines share the image encoder, support/query partition, and fixed schedule without extra source-domain evidence; comparisons with these alternatives would first need aligned information access.

DVLA-RL builds attributes and descriptions from class names and support samples and learns cross-modal fusion through reinforcement learning (Li et al., 2026b). Flow Matching Alignment learns a gradual cross-modal transformation (Jiang et al., 2026). Both use semantic assistance beyond a mean of image activations. Update location also matters: LoR-VP factorizes image-space prompts (Jin et al., 2025), whereas CenWrite factorizes weight increments inside the encoder. Shared few-shot objectives or low-rank matrices therefore do not imply equivalent adaptation rules.

K.3 Capacity, parameterization, and factor optimization

Low parameter count does not uniquely determine capacity. LoRA's expressive-power analysis relates approximation ability to rank and architecture (Zeng & Lee, 2024). NOLA combines fixed random bases (Abbasi Koohpayegani et al., 2024), while CAPABOOST uses static masks on shared weights to increase update diversity (Song et al., 2024). These methods alter reachable updates or their parameter budget. CenWrite retains the LoRA factorization, changes the input component driving its factors, and adds a separately counted residual vector. Equal nominal rank does not equalize capacity; our resource analysis reports actual trainable counts.

BoRA increases factorized-update diversity through block-specific diagonal matrices (Li et al., 2026a). TRAC shares tensor-train cores across layers with lightweight layer-specific control (Ye et al., 2026a), and LoCA learns cosine coefficients and their locations (Du et al., 2025). These constructions redistribute update capacity, unlike CenWrite's centered input and detached compensation within the same factor form. This distinction motivates including factorized and frequency-based comparators while retaining their actual configurations: rank and frequency counts are not a common budget.

Optimization can change within a compact update form. DeLoRA separates angular learning from adaptation strength using normalized factors (Bini et al., 2025). GeoLoRA adapts rank through dynamical low-rank approximation (Schotthöfer et al., 2025); LoRA-S derives optimizers from quotient geometry and a Sylvester equation (Zheng & Wu, 2026). Their geometries concern factorized weight updates. CenWrite instead uses an empirical support activation reference, rather than invariance to alternative factorizations of the same matrix. Its local derivatives specify why identical forward values at a fixed state can produce different learning trajectories.

PiCa projects gradients onto principal column spaces of pretrained weights (Hwang et al., 2026), and LoRA-Dash emphasizes task-specific directions (Si et al., 2025). Landscape analyses compare factorized adaptation and gradient projection under stated assumptions (Liu et al., 2025). CenWrite operates in activation space: it neither projects a full fine-tuning gradient onto a learned principal subspace nor estimates task directions from weight spectra. Equation 5 gives the direct local derivative; matched component comparisons measure probability effects.

RaLoRA aligns layer ranks with estimated gradient intrinsic dimensions (Ye et al., 2026b); FlexLoRA prunes and expands ranks using entropy (Liu et al., 2026); ReLoRA accumulates low-rank updates for high-rank training (Lialin et al., 2024). Their capacity controls and training budgets differ from a fixed short support-adaptation episode. Our rank analysis shows that a larger rank can worsen probability scores without a resolved classification gain, motivating joint evaluation of both objectives. Comparing static and adaptive ranks would also need matched total budgets and support selection, including their interaction with centered learning.

K.4 PROBABILITY QUALITY UNDER ADAPTATION

Accuracy alone does not measure probability quality (Guo et al., 2017). Recent analyses motivate proper scores alongside binned calibration error (Chidambaram & Ge, 2025; Tao et al., 2024). C-TPT changes text-feature dispersion (Yoon et al., 2024), while A-TPT studies angular diversity during unlabeled test-time prompt tuning (Ahamed et al., 2026). Their text-based mechanisms and data access differ from CenWrite's labeled support adaptation in an image encoder. We use support cross-entropy without a calibration loss or post-hoc temperature fit, and report NLL, Brier, and ECE alongside accuracy and macro-F1 to distinguish probability gains from changes in class decisions.

Training objectives provide other routes to calibration. Odd-k-out learning applies cross-entropy to sets of examples (Muttenthaler et al., 2024); tailored mixup adjusts interpolation to data similarity (Bouniot et al., 2025). Analyses of sharpness-aware minimization connect calibration benefits to predictive entropy (Tan et al., 2026). CenWrite uses none of these ingredients. Its ablations compare gradient and forward operations under a shared loss, identifying effects of those changes rather than a unique cause of miscalibration.

Calibration estimation and confidence ranking answer different questions. Finite-sample bounds for binary calibration under explicit conditions (Ciosek et al., 2026) do not directly certify five-way, 75-query ECE estimates. VLM misclassification-detection work also distinguishes calibration from separating correct and incorrect predictions (Huang et al., 2026). Our binned ECE, NLL and Brier scores and paired episode intervals describe benchmark probability quality. They do not assess selective risk, out-of-distribution detection or certified calibration bounds.