ON PROTOTYPE COMPOSITION FOR FEW-SHOT CLASS-INCREMENTAL LEARNING: BREAKING STATISTICAL BIAS VIA ATTENTIVE AND EXPANSIVE CALIBRATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Few-shot class-incremental learning (FSCIL) faces an inherent statistical imbalance: base classes are learned from abundant data, whereas each incremental class has only a few support examples. This imbalance produces two different vulnerabilities in prototype composition: base prototypes are trained without future classes, while centroids estimated from a few incremental supports may miss valid query directions. Their joint use can distort base–novel decisions in either direction. We refer to this decision-level asymmetry as *statistical bias in prototype composition*. To mitigate it, we propose BASIL *(Base-Prototype Alignment with Support-Informed Local Expansion)*, an attentive and expansive prototype-calibration framework. First, class-weighted alignment constrains the displacement of base prototypes from their empirical centroids without discarding base-class discrimination. Second, real supports and surrogate tangent references expand the incremental-class reference set while the mean-prototype score is retained. BASIL requires no training during incremental learning and achieves strong results on widely used FSCIL benchmarks.

1 INTRODUCTION

FSCIL requires a model to learn novel classes from only a few examples while preserving classification on previously acquired categories. Existing studies address this challenge through representation learning, forward compatibility, prototype refinement, and prompt adaptation. TOPIC (Tao et al., 2020), CEC (Zhang et al., 2021), FACT (Zhou et al., 2022), SAVC (Song et al., 2023), and CLOSER (Oh et al., 2024) improve the transferability of representations to future classes. SimpleCIL (Zhou et al., 2023a), RanPAC (McDonnell et al., 2023), and visual-prompt methods (Wang et al., 2022c;b; Smith et al., 2023; Jiang et al., 2025; Liu & Yang, 2025; He et al., 2025) further show that pretrained models can provide strong FSCIL representations.

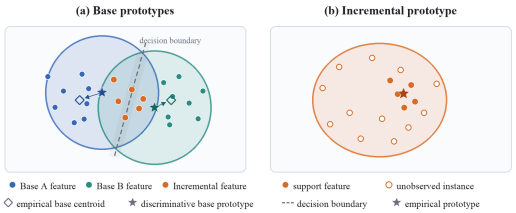


Figure 1: Conceptual illustration of the prototype-composition problem. Left: A novel class appears near a base-class decision boundary, motivating base-prototype calibration. Right: Sparse support sampling shifts the estimated novel-class center, motivating local reference expansion.

Because incremental classes are sampled sparsely, prototype-based classification is common in FSCIL. This makes the composition of base and incremental references central to prediction. Base

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.

classes contain abundant training data, leading to discriminative prototypes learned without seeing future incremental classes. In contrast, incremental-class mean prototypes depend on only K supports and can miss valid query directions. When these reference types compete in a shared classifier, either base-to-novel or novel-to-base errors may result. The objective is to improve their joint discrimination without assuming a loss of base-session accuracy.

Figure 1 illustrates the two vulnerabilities. The left panel places an incremental class near a boundary learned from base classes. The right panel shows how sparse supports can yield a prototype displaced from the center of the class distribution.

Prior work on prototype calibration (Wang et al., 2023), base–novel collision (Zhou et al., 2024), decision-boundary calibration (Wang et al., 2024; Zhou et al., 2023b), and prescribed classifier geometry (Yang et al., 2023) has examined unreliable competition between base and incremental classes. These studies address different aspects of the problem. Here, we examine what happens when a learned base prototype and a support-estimated incremental prototype contribute scores to the same classifier.

To address this problem, we propose BASIL (*Base-Prototype Alignment with Support-Informed Local Expansion*), combining attentive base alignment and expansive incremental scoring. First, attentive calibration fits the base cosine head with cross entropy and a class-weighted centroid-alignment penalty. The alignment is stronger when the base features are tightly clustered around their observed centroid, while dispersed classes retain more discriminative freedom. This step limits excessive prototype displacement while preserving base-class discrimination. Second, expansive calibration combines the mean prototype with scores from real supports and surrogate tangent references using a fixed interpolation coefficient. Together, the two steps address different errors that arise when these references are used in a shared classifier.

Our contributions are three-fold:

- We identify two sources of unreliable decisions in FSCIL prototype composition: densely supervised base prototypes are trained without future classes, while incremental prototypes are estimated from sparse supports. Their scores must nevertheless compete in one classifier.
- We propose BASIL, which uses concentration-weighted alignment to constrain base-prototype displacement and support-informed local scoring to supplement incremental mean prototypes.
- We demonstrate competitive performance across FSCIL benchmarks and validate the two calibration components through controlled ablations.

2 Related Work

FSCIL representations and pretrained backbones. TOPIC (Tao et al., 2020), CEC (Zhang et al., 2021), FACT (Zhou et al., 2022), SAVC (Song et al., 2023), ALICE (Peng et al., 2022), and CLOSER (Oh et al., 2024) improve adaptation to future classes through topology, graph propagation, forward-compatible training, open-set objectives, or geometric regularization. Prompt and parameter-efficient methods, including L2P (Wang et al., 2022c), DualPrompt (Wang et al., 2022b), CODA-Prompt (Smith et al., 2023), LGSP-Prompt (Jiang et al., 2025), SEC-Prompt (Liu & Yang, 2025), and DSS-Prompt (He et al., 2025), efficiently adapt pretrained models. SimpleCIL (Zhou et al., 2023a) and RanPAC (McDonnell et al., 2023) further establish strong frozen-feature baselines. BASIL also uses shared VPT (Jia et al., 2022), but focuses on how base and incremental references are combined for classification.

Prototype calibration and base–novel competition. Prototypical Networks (Snell et al., 2017), nearest-class-mean classification, and iCaRL (Rebuffi et al., 2017) establish class means as effective nonparametric classifiers. TEEN (Wang et al., 2023) calibrates few-shot prototypes, while LIMIT (Zhou et al., 2023b) maps base classifiers and incremental prototypes into a common score space. ConCM (Wang et al., 2026) uses memory-aware prototype calibration and dynamic structure matching to address prototype bias and cross-session structure inconsistency. NC-FSCIL (Yang et al., 2023) prescribes a unified classifier geometry, RDI (Zhou et al., 2024) studies base–novel

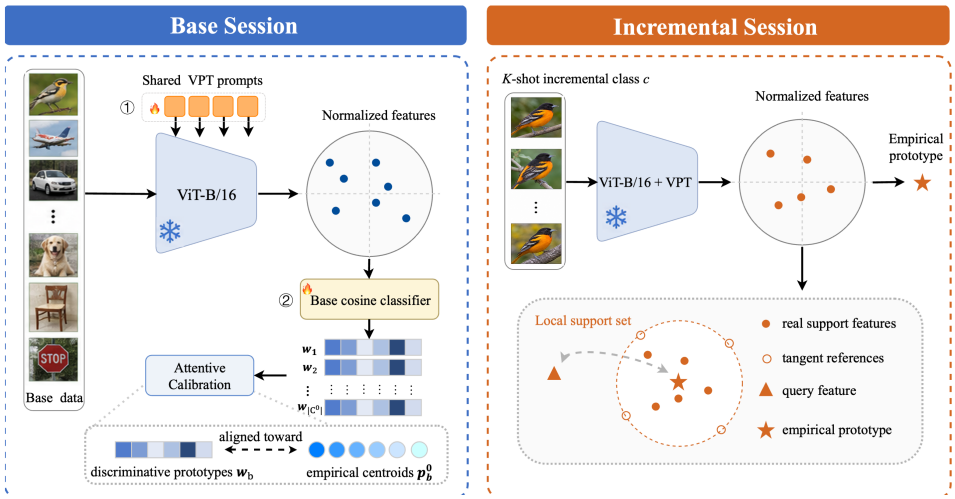


Figure 2: Attentive and expansive prototype calibration in BASIL. The base session constrains learned prototypes through concentration-weighted alignment. Incremental sessions combine a mean prototype with real support features and surrogate tangent references using a fixed interpolation coefficient.

collision, and Approximation Risk (Wang et al., 2024) analyzes old–new competition through class margins. BASIL specifically combines class-weighted alignment of learned base rows with mean-preserving local-reference scoring for incremental classes.

Few-shot prototype estimation and coverage expansion. Prototype calibration (Wang et al., 2023), generative augmentation (Bao et al., 2026), semantic knowledge transfer (Park et al., 2024), and virtual-class augmentation (Song et al., 2023) seek to recover within-class variation absent from sparse support sets. Instance-wise matching against individual supports or reference sets (Vinyals et al., 2016) can retrieve a query displaced from the support mean, but may also give an incorrect class an accidental near neighbor. BASIL synthesizes neither images nor class semantics. It keeps the support mean as the central estimate and allows real support features and surrogate tangent references to provide limited local compensation.

3 Method

3.1 Prototype Composition under Statistical Imbalance

We follow the standard FSCIL protocol (Jiang et al., 2025). The base session S_0 contains abundant labeled training examples from class set $\mathcal{C}_0$. Incremental session S_t introduces a disjoint class set $\mathcal{C}_t$, with only K support examples per class. At stage t, predictions are made over $\mathcal{C}_{\le t} = \bigcup_{j=0}^{t} \mathcal{C}_j$; we write $\mathcal{C}_{1:t} = \bigcup_{j=1}^{t} \mathcal{C}_j$ for the incremental classes seen so far. Incremental sessions have no access to historical training images, but their supports retain their labels. Prediction uses neither a test-time session identity nor a hidden task identifier.

Let $\mathbf{f}(x) \in \mathbb{R}^d$ be the unit-normalized feature extracted by the VPT (Jia et al., 2022)-adapted ViT (Dosovitskiy et al., 2021). The positive cosine-score scale τ is learned during the final base-head fit and shared by all class scores. For a base class $b \in \mathcal{C}_0$ with n_b training examples, define its feature mean $\bar{\mathbf{f}}_b = n_b^{-1} \sum_{i:y_i=b} \mathbf{f}(x_i)$ and unit empirical centroid $\mathbf{p}_b^0 = \bar{\mathbf{f}}_b / \|\bar{\mathbf{f}}_b\|_2$, assuming $\bar{\mathbf{f}}_b \neq \mathbf{0}$. Let $\mathbf{w}_b$ denote the unit base-class row fitted with cross entropy alone. It serves as the uncalibrated control. For an incremental class $c \in \mathcal{C}_{1:t}$, $\mathbf{p}_c$ is the unit-normalized mean of its K support features. Uncalibrated prototype composition scores a class j as

$$s_j^{\text{uncal}}(x) = \begin{cases} \tau \mathbf{f}(x)^\top \mathbf{w}_j, & j \in \mathcal{C}_0, \\ \tau \mathbf{f}(x)^\top \mathbf{p}_j, & j \in \mathcal{C}_{1:t}. \end{cases} \tag{1}$$

The base prototypes are learned without future classes, whereas each incremental prototype is estimated from only K supports. Their scores can therefore differ for two distinct reasons: displacement of a learned base prototype from its class centroid, and uncertainty in a few-shot novel centroid. Either can change a base–novel decision, depending on the query. The following two calibrations address these respective sources without assuming that one direction of error always dominates.

Figure 2 summarizes BASIL. In the base session, we first adapt the shared VPT prompts, then freeze the representation and fit the final base head with attentive calibration. In each incremental session, model parameters remain frozen. New classes retain their mean prototypes, while real supports and tangent references supplement their scores at inference.

3.2 Attentive Calibration for Base Prototypes

Base data provide two types of class-wise references. The empirical centroid $\mathbf{p}_b^0$ estimates the within-class direction, whereas a discriminative classifier row competes with other classes to form decision boundaries. The empirical centroid anchors the final head fit through a soft alignment penalty without replacing that row.

Base training proceeds in two steps. First, the VPT prompts are trained jointly with a temporary cosine classifier. Let $\mathbf{f}_\phi(x)$ and $\mathbf{f}_0(x)$ denote unit-normalized features of the backbone with and without VPT, respectively. We optimize

$$\mathcal{L}_{\mathrm{VPT}} = \mathcal{L}_{\mathrm{CE}}^{\mathrm{tmp}} + \lambda_{\mathrm{rc}} \frac{1}{N} \sum_{i=1}^{N} \left(1 - \mathbf{f}_\phi(x_i)^\top \mathbf{f}_0(x_i)\right), \tag{2}$$

where N is the number of base-session training samples, and λ_{rc} controls representation-consistency regularization. The cross-entropy term trains the temporary classifier and VPT prompts; the second term limits feature drift from the frozen pretrained backbone. We then discard the temporary head, freeze the backbone and VPT, and set $\mathbf{f}(x) = \mathbf{f}_\phi(x)$ for the remainder of the paper.

The reliability of empirical centroids varies across base classes. Suppose class b contains n_b training features. We measure its directional concentration using the sample mean resultant length

$$\kappa_b = \left\| \frac{1}{n_b} \sum_{i:y_i=b} \mathbf{f}(x_i) \right\|_2 = \frac{1}{n_b} \sum_{i:y_i=b} \mathbf{f}(x_i)^\top \mathbf{p}_b^0. \tag{3}$$

We normalize κ_b by its across-class mean and clip the result to obtain a class-wise alignment weight a_b, formulated as

$$\begin{aligned} \bar{\kappa} &= \frac{1}{|\mathcal{C}_0|} \sum_{j \in \mathcal{C}_0} \max(\kappa_j, \epsilon_{\mathrm{num}}), \\ a_b &= \mathrm{clip}\left(\frac{\max(\kappa_b, \epsilon_{\mathrm{num}})}{\bar{\kappa}}, a_{\min}, a_{\max} \right), \end{aligned} \tag{4}$$

where $\epsilon_{\mathrm{num}} > 0$ is a numerical floor, and $a_{\min}$ and $a_{\max}$ specify the clipping range. A larger κ_b indicates stronger directional concentration among the observed base-class features. We use it to weight alignment, not as a guarantee about unseen queries. In the second base-training step, the VPT features are frozen. We initialize the final cosine head from the empirical centroids and jointly optimize its rows under cross entropy and class-weighted alignment:

$$\mathcal{L}_{\mathrm{base}} = \mathcal{L}_{\mathrm{CE}} + \lambda_{\mathrm{att}} \frac{1}{|\mathcal{C}_0|} \sum_{b \in \mathcal{C}_0} a_b \left(1 - \widetilde{\mathbf{w}}_b^\top \mathbf{p}_b^0\right). \tag{5}$$

We denote the resulting unit row by $\widetilde{\mathbf{w}}_b$. The uncalibrated control $\mathbf{w}_b$ comes from a separate cross-entropy-only head fit on the same frozen features. Here, *attentive* refers to the class-dependent alignment weights a_b, determined by observed directional concentration, rather than query-level attention. Cross entropy preserves base-class separability, while alignment limits displacement from empirical centroids. Concentrated classes receive stronger centroid constraints, whereas dispersed classes retain more discriminative freedom.

For any unit query, the difference between the learned base score and the centroid score satisfies

$$\left| \tau \mathbf{f}(x)^\top \widetilde{\mathbf{w}}_b - \tau \mathbf{f}(x)^\top \mathbf{p}_b^0 \right| \le \tau \sqrt{2\left(1 - \widetilde{\mathbf{w}}_b^\top \mathbf{p}_b^0\right)}. \tag{6}$$

Eq. (3) supplies the observed class-specific weight, while Eq. (6) identifies the score discrepancy constrained by alignment.

3.3 Expansive Calibration for Incremental Prototypes

For an incremental class c, the empirical prototype $\mathbf{p}_c$ is the normalized mean of its K unit support features. It provides a central summary but may miss queries that deviate from the observed support mean.

BASIL therefore constructs a local support set $\mathcal{R}_c$ that contains all real support features and M tangent references around $\mathbf{p}_c$. Specifically, we sample $\boldsymbol{\eta}_{c,j} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$, project and normalize it in the tangent space, and map the resulting direction back to the unit sphere with tangent perturbation magnitude σ:

$$\begin{aligned} \mathbf{d}_{c,j} &= \frac{(\mathbf{I}_d - \mathbf{p}_c\mathbf{p}_c^\top)\boldsymbol{\eta}_{c,j}}{\|(\mathbf{I}_d - \mathbf{p}_c\mathbf{p}_c^\top)\boldsymbol{\eta}_{c,j}\|_2}, \\ \tilde{\mathbf{r}}_{c,j} &= \frac{\mathbf{p}_c + \sigma\mathbf{d}_{c,j}}{\|\mathbf{p}_c + \sigma\mathbf{d}_{c,j}\|_2}, \end{aligned} \tag{7}$$

$$\mathcal{R}_c = \{\mathbf{f}(x_{c,i})\}_{i=1}^{K} \cup \{\tilde{\mathbf{r}}_{c,j}\}_{j=1}^{M}, \tag{8}$$

where $\mathbf{I}_d$ is the $d \times d$ identity matrix, M is the number of tangent references, and $\sigma > 0$ determines their angular offset $\arctan\sigma$ from $\mathbf{p}_c$. By construction, $\mathbf{d}_{c,j} \perp \mathbf{p}_c$. These references probe nearby directions on the unit sphere around each incremental prototype. Each incremental class retains its labeled real-support features and a fixed set of tangent references sampled when that class enters the sequence. The references are reused during evaluation.

We then calibrate Eq. (1) by interpolating the mean-prototype and local-reference scores with a fixed coefficient α:

$$s_c^{\mathrm{exp}}(x) = (1-\alpha)\tau\mathbf{f}(x)^\top\mathbf{p}_c + \alpha\tau \max_{\mathbf{r}\in\mathcal{R}_c} \mathbf{f}(x)^\top\mathbf{r}. \tag{9}$$

The interpolation coefficient $\alpha \in [0,1]$ weights the local-reference score, while $1-\alpha$ retains the mean-prototype score. Its value is fixed before evaluation and shared by all queries, classes, and incremental sessions within each dataset.

Expansive calibration probes directions near the observed class prototype rather than adding independent observations. Real supports and tangent references are both derived from the same K-shot set, and their maximum can also increase a wrong class's score. We therefore retain the mean-prototype score through fixed interpolation instead of treating the expanded references as additional labeled data.

3.4 Attentive and Expansive Prototype Composition

The final classifier combines calibrated base rows with incremental scores. For any class $j \in \mathcal{C}_{\le t}$, base classes use attentively calibrated discriminative prototypes $\widetilde{\mathbf{w}}_j$, whereas incremental classes use the expanded-reference score s_j^{exp}:

$$\begin{aligned} s_j^{\mathrm{BASIL}}(x) &= \begin{cases} \tau\mathbf{f}(x)^\top\widetilde{\mathbf{w}}_j, & j \in \mathcal{C}_0, \\ s_j^{\mathrm{exp}}(x), & j \in \mathcal{C}_{1:t}, \end{cases} \\ \hat{y}(x) &= \arg\max_{j\in\mathcal{C}_{\le t}} s_j^{\mathrm{BASIL}}(x). \end{aligned} \tag{10}$$

The two stages address the same base–novel score competition from different sides. Attentive calibration constrains learned base prototypes toward their empirical centroids during base training. Expansive calibration then supplements each K-shot incremental mean without parameter updates. Neither step guarantees that every query is corrected.

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

Datasets and protocols. We evaluate BASIL on CUB-200 (Wah et al., 2011), CIFAR-100 (Krizhevsky, 2009), and mini-ImageNet (Vinyals et al., 2016), using their public FSCIL class orders and support indices. CUB-200 contains 100 base classes followed by ten 10-way 5-shot sessions. CIFAR-100 and mini-ImageNet contain 60 base classes followed by eight 5-way 5-shot sessions.

Implementation. We use an ImageNet-21K-pretrained ViT-B/16 (Dosovitskiy et al., 2021) and insert six shared VPT tokens (Jia et al., 2022) into Transformer blocks 1–8. Evaluation images and incremental supports use the 224-pixel evaluation transform of the corresponding pretrained `timm` model. Base training adapts VPT with a temporary cosine head for 10 epochs, freezes the representation, and then fits the final cosine head for 120 epochs. Incremental sessions update no model parameters and replay no images. They retain the class-indexed support features and fixed tangent references used for scoring. Inference uses neither test labels nor task/session identity. Experiments are conducted on NVIDIA RTX 3090 GPUs.

Metrics. We report top-1 accuracy for every session. *Avg.* is the arithmetic mean of the session accuracies listed from S_0 to the final session. For diagnostics and ablations, *Inc. Avg.* averages only $S_1, \ldots, S_T$, and *Final* is the accuracy at the final session.

4.2 BENCHMARK COMPARISONS

Tables 1–3 report session-wise accuracy and *Avg.* on CUB-200, CIFAR-100, and mini-ImageNet. The comparisons include conventional FSCIL methods, prompt-based continual-learning methods, and pretrained-ViT approaches. BASIL attains *Avg.* values of 86.62%, 90.60%, and 96.28%, respectively, with competitive base-session accuracy and strong retention across incremental sessions.

4.3 PROTOTYPE-COMPOSITION DIAGNOSTICS AND BIDIRECTIONAL CALIBRATION ABLATIONS

Table 4(a) fixes incremental classes to their mean prototypes and changes only the base-class reference. Replacing $\mathbf{w}_b$ with $\mathbf{p}_b^0$ raises CUB-200 *Inc. Avg.* from 79.02% to 84.82%, but lowers S_0 from 88.83% to 87.78%. This trade-off motivates constraining rather than replacing the discriminative base rows. In Table 4(b), attentive calibration improves mean-prototype *Inc. Avg.* on CUB-200, CIFAR-100, and mini-ImageNet by 6.41, 21.43, and 0.75 points, respectively. Adding expansive

Table 1: Session-wise top-1 accuracy (%) on CUB-200.

Method	S0	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Avg.
iCaRL (Rebuffi et al., 2017)	82.43	79.32	68.74	59.93	61.62	59.91	57.83	57.34	55.42	52.73	55.92	62.84
CEC (Zhang et al., 2021)	81.82	79.53	78.42	75.54	76.31	74.83	74.41	74.62	74.23	73.91	73.84	76.13
FOSTER (Wang et al., 2022a)	85.02	83.43	77.41	71.52	69.93	66.34	65.52	63.07	62.92	62.03	60.42	69.78
FACT (Zhou et al., 2022)	84.32	81.23	79.14	75.13	75.42	73.31	72.43	72.52	71.41	71.12	70.91	75.18
NC-FSCIL (Yang et al., 2023)	83.52	80.92	80.14	77.83	77.81	76.96	76.72	74.78	74.18	73.92	73.80	77.33
L2P (Wang et al., 2022c)	82.40	81.20	79.00	76.80	76.20	74.70	74.10	74.10	72.70	73.00	73.60	76.16
DualPrompt (Wang et al., 2022b)	83.50	82.20	80.90	79.50	78.60	77.00	76.30	77.00	75.70	76.10	76.50	78.48
CODA-Prompt (Smith et al., 2023)	79.60	78.10	76.40	75.60	75.00	73.10	72.50	72.80	72.00	72.40	72.90	74.58
TEEN (Wang et al., 2023)	88.80	86.20	85.50	83.00	83.00	81.70	81.50	79.70	79.90	79.50	80.00	82.62
PriViLege (Park et al., 2024)	82.30	81.30	80.50	77.80	77.80	76.00	75.70	76.00	75.20	75.20	75.10	77.54
LGSP-Prompt (Jiang et al., 2025)	85.72	84.31	83.21	81.33	81.80	80.33	79.89	80.09	79.18	79.74	79.72	81.39
ASP (Liu et al., 2024)	87.10	86.00	84.90	83.40	83.60	82.40	82.60	83.00	82.60	83.00	83.50	83.83
SEC-Prompt (Liu & Yang, 2025)	87.50	86.60	85.50	84.60	84.90	84.00	83.70	84.20	83.80	84.00	84.80	84.87
DSS-Prompt (He et al., 2025)	**90.95**	87.64	86.43	85.12	85.32	83.99	83.59	83.63	83.25	83.62	84.22	85.25
CASP (Huang et al., 2026)	89.40	87.90	87.40	86.80	86.80	85.00	85.00	85.40	**85.30**	**85.60**	**85.80**	86.40
BASIL	89.28	**88.67**	**88.51**	**86.83**	**87.09**	**85.37**	**85.41**	**85.56**	85.13	85.29	85.66	**86.62**

Table 2: Session-wise top-1 accuracy (%) on CIFAR-100.

Method	S0	S1	S2	S3	S4	S5	S6	S7	S8	Avg.
L2P (Wang et al., 2022c)	84.70	82.30	80.10	77.50	77.00	76.00	75.60	74.10	72.30	77.73
DualPrompt (Wang et al., 2022b)	86.00	83.60	82.90	80.20	80.60	80.20	80.50	79.00	77.40	81.16
CODA-Prompt (Smith et al., 2023)	86.00	83.60	81.60	79.20	79.10	78.50	78.30	77.00	75.40	79.86
TEEN (Wang et al., 2023)	92.90	90.20	88.40	86.80	86.40	86.00	85.80	85.10	84.00	87.29
PriViLege (Park et al., 2024)	90.90	89.40	89.00	87.60	87.90	87.40	87.60	87.20	86.10	88.12
ASP (Liu et al., 2024)	92.20	90.70	90.00	88.70	88.70	88.20	88.20	87.80	86.70	89.02
SEC-Prompt (Liu & Yang, 2025)	92.00	90.80	90.60	89.10	89.20	89.10	89.20	88.70	87.50	89.58
CASP (Huang et al., 2026)	**93.40**	**91.80**	91.00	89.70	89.80	89.50	89.50	89.30	88.10	90.23
BASIL	93.12	91.74	**91.33**	**90.17**	**90.29**	**90.13**	**90.13**	**89.82**	**88.71**	**90.60**

Table 3: Session-wise top-1 accuracy (%) on mini-ImageNet.

Method	S0	S1	S2	S3	S4	S5	S6	S7	S8	Avg.
iCaRL (Rebuffi et al., 2017)	61.31	46.32	42.94	37.63	30.49	24.00	20.89	18.80	17.20	33.29
TOPIC (Tao et al., 2020)	61.31	50.09	45.17	41.16	37.48	35.52	32.19	29.46	24.42	39.64
CEC (Zhang et al., 2021)	72.00	66.83	62.97	59.43	56.70	53.73	51.19	49.24	47.63	57.75
L2P (Wang et al., 2022c)	94.59	87.49	81.18	75.76	71.05	66.86	63.15	59.82	56.84	72.97
DualPrompt (Wang et al., 2022b)	95.05	87.81	81.51	76.07	71.38	67.19	63.45	60.15	57.14	73.31
WaRP (Kim et al., 2023)	83.30	80.53	77.22	74.99	73.64	71.52	69.16	68.79	67.97	74.12
TEEN (Wang et al., 2023)	73.53	70.55	66.37	63.23	60.53	57.95	55.24	53.44	52.08	61.44
GCMR (Bao et al., 2026)	87.38	77.44	74.45	71.11	68.80	66.59	64.45	62.23	59.85	70.26
PriViLege (Park et al., 2024)	96.68	96.49	95.65	95.54	95.54	94.91	94.33	94.19	94.10	95.27
BASIL	**97.28**	**97.18**	**96.46**	**96.41**	**96.41**	**95.99**	**95.66**	**95.62**	**95.49**	**96.28**

calibration after alignment changes *Inc. Avg.* by 0.92, 1.60, and 0.87 points on these datasets, respectively. On CUB-200, representation-consistency regularization further improves *Inc. Avg.* by 0.21 points and *Final* by 0.40 points. These results indicate that local reference relies on the consistent alignment of base classes first, and can further enhance the adaptability of novel classes in the subsequent stages.

Figure 3 separates score shifts from prediction changes at the final CUB-200 session. For each query, the left panel fixes the highest-scoring base class and novel class under uncalibrated base rows and novel mean prototypes, then tracks the change in their base-minus-novel score difference after each calibration step. A positive shift favors the selected base class, but this fixed-pair difference is not the true-class margin. The right panel instead measures top-1 accuracy changes over all seen classes, separately for base and novel queries. A corrective change turns an error into a correct prediction, while a harmful change does the reverse. Attentive calibration raises the median fixed-pair difference by $+0.329$ for base queries and $+0.049$ for novel queries. Relative to uncalibrated composition, it corrects 437 base-query predictions and harms 23 ($+14.46$ percentage points), but harms 33 novel-query predictions without correcting any (-1.13 points). Adding expansive calibration to the attentive model corrects 54 novel-query predictions and harms 13 ($+1.40$ points), while harming three base-query predictions (-0.10 points). Thus, the base-side gain from alignment and the novel-side gain from local scoring come with distinct, measurable trade-offs.

Table 5(a) measures the angle $\theta_b = \arccos(\widetilde{\mathbf{w}}_b^\top \mathbf{p}_b^0)$ between each calibrated base row and its empirical centroid. Attentive calibration reduces the mean angle from 52.45° to 4.70°, while mean cosine similarity rises from 0.6051 to 0.9948. Table 5(b) isolates local-reference scoring on CUB-200. Real support features raise *Inc. Avg.* from 85.43% to 85.50%, and adding tangent references raises it to 86.35%.

Figure 4 isolates the role of the empirical prototype in expansive calibration. We hold the attentive base head and local references fixed at CUB-200 S_{10}, and compare the local-reference maximum alone ($\alpha = 1$) with fixed interpolation ($\alpha = 0.75$). Both are evaluated against mean-prototype scoring ($\alpha = 0$). The maximum alone turns 329 errors into correct predictions but also reverses

Table 4: Prototype-composition diagnostics and component ablations. *Inc. Avg.* averages incremental sessions. *Final* is the last-session accuracy.

(a) Base-class reference on CUB-200 with mean-prototype novel classes

Base-class reference	S_0	Inc. Avg.	Final
Discriminative prototype $\mathbf{w}_b$	**88.83**	79.02	78.41
Empirical centroid $\mathbf{p}_b^0$	87.78	**84.82**	**84.14**

(b) Attentive and expansive calibration on three benchmarks

Dataset	Variant	Rep. Cons.	Att.	Exp.	S_0	Inc. Avg.	Final
CUB-200	No calibration	✓	–	–	88.83	79.02	78.41
	Expansive only	✓	–	✓	88.83	78.94	78.37
	Attentive only	✓	✓	–	89.28	85.43	84.47
	w/o Rep. Cons.	–	✓	✓	**89.35**	86.14	85.26
	BASIL	✓	✓	✓	89.28	**86.35**	**85.66**
CIFAR-100	No calibration	✓	–	–	92.17	67.26	61.82
	Expansive only	✓	–	✓	92.17	65.09	59.28
	Attentive only	✓	✓	–	93.12	88.69	87.31
	w/o Rep. Cons.	–	✓	✓	**93.18**	89.92	88.21
	BASIL	✓	✓	✓	93.12	**90.29**	**88.71**
mini-ImageNet	No calibration	✓	–	–	96.98	94.53	93.05
	Expansive only	✓	–	✓	96.98	93.85	92.03
	Attentive only	✓	✓	–	97.28	95.28	94.84
	w/o Rep. Cons.	–	✓	✓	**97.43**	95.96	95.17
	BASIL	✓	✓	✓	97.28	**96.15**	**95.49**

Table 5: Direct mechanism evidence for attentive and expansive calibration on CUB-200.

(a) Attentive base calibration

Base head	Cos.	Angle
No constraint	0.6051	52.45°
Attentive	0.9948	4.70°

(b) Expansive incremental calibration

	α	M	Inc. Avg.	Final	Avg.
Mean prototype	0	0	85.43	83.97	85.33
+ real support features	0.75	0	85.50	84.47	85.84
+ tangent references	0.75	96	**86.35**	**85.66**	**86.62**

172 correct predictions. Interpolation makes fewer corrective changes (277), yet reduces harmful changes to 72. Its net gain therefore rises from 157 to 205 correct predictions. This comparison shows that retaining the empirical-prototype score can limit harmful predictions from local-reference selection while preserving much of its corrective effect.

Table 6 examines the number of tangent references, the interpolation coefficient, and the tangent perturbation magnitude on CUB-200. The setting $M = 0$ retains real supports but removes tangent references. The endpoints $\alpha = 0$ and $\alpha = 1$ use only the mean-prototype score and only the local-reference maximum, respectively. Performance changes little over $M \in [32, 192]$ and $\sigma \in [0.01, 0.1]$, while an intermediate interpolation coefficient outperforms both endpoints.

5 Conclusion

BASIL addresses the statistical asymmetry between densely learned base prototypes and few-shot incremental mean prototypes in a shared FSCIL classifier. Attentive calibration constrains base-prototype displacement from empirical centroids, while expansive calibration supplements incremental mean-prototype scores with support-informed local scores through fixed interpolation. Diagnostics and ablations reveal the complementary gains and costs of these steps, and experiments support their combined use.

Table 6: Sensitivity of expansive calibration on CUB-200.

Parameter	Value	Inc. Avg.	Final
M	0	86.19	85.30
	32	86.33	85.62
	64	86.33	85.62
	96	**86.35**	**85.66**
	192	**86.35**	**85.66**
α	0.00	85.43	83.97
	0.25	86.01	85.19
	0.50	86.11	85.33
	0.75	**86.35**	**85.66**
	1.00	86.25	85.47
σ	0.010	**86.35**	85.64
	0.025	**86.35**	**85.66**
	0.050	86.32	85.62
	0.100	86.32	85.64

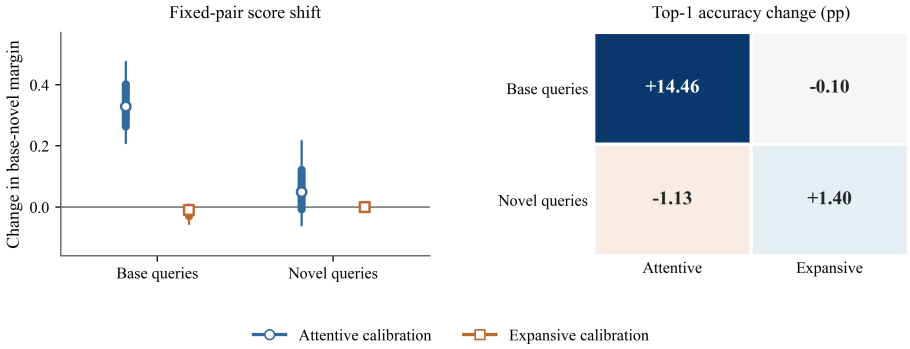


Figure 3: Fixed-pair score shifts and paired top-1 prediction changes on CUB-200 S_{10}.

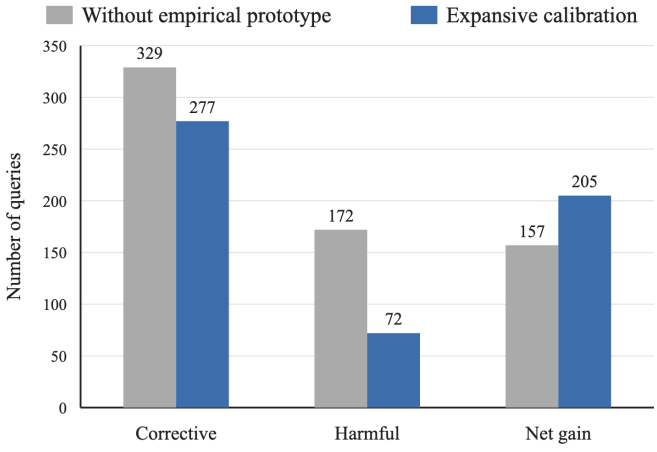


Figure 4: Corrective and harmful prediction changes with and without the empirical prototype on CUB-200 S_{10}.

AI USE STATEMENT

We used OpenAI generative AI tools to assist in the writing of experimental code and the polishing of paper manuscripts. The authors are responsible for the final methods, proofs, code, figures, reported results, and text.

REFERENCES

Kexin Bao, Yong Li, Dan Zeng, and Shiming Ge. Few-shot class-incremental learning via generative co-memory regularization. *arXiv preprint arXiv:2601.07117*, 2026.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021.

Linpu He, Yanan Li, Bingze Li, Elvis Han Cui, and Donghui Wang. Dss-prompt: Dynamic-static synergistic prompting for few-shot class-incremental learning. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 2703–2712, 2025.

Shuai Huang, Xuhan Lin, and Yuwu Lu. Casp: Few-shot class-incremental learning with cls token attention steering prompts. *Pattern Recognition*, pp. 113306, 2026.

Menglin Jia, Luming Tang, Bor-Chun Chen, Claire Cardie, Serge Belongie, Bharath Hariharan, and Ser-Nam Lim. Visual prompt tuning. In *European Conference on Computer Vision*, pp. 709–727, 2022.

Yongwei Jiang, Yixiong Zou, Yuhua Li, and Ruixuan Li. Revisiting pool-based prompt learning for few-shot class-incremental learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025.

Do-Yeon Kim, Dong-Jun Han, Jun Seo, and Jaekyun Moon. Warping the space: Weight space rotation for class-incremental few-shot learning. In *International Conference on Learning Representations*, 2023.

Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009.

Chenxi Liu, Zhenyi Wang, Tianyi Xiong, Ruibo Chen, Yihan Wu, Junfeng Guo, and Heng Huang. Few-shot class incremental learning with attention-aware self-adaptive prompt. In *European Conference on Computer Vision*, 2024.

Ye Liu and Meng Yang. Sec-prompt: Semantic complementary prompting for few-shot class-incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025.

Subhransu Maji, Esa Rahtu, Juho Kannala, Matthew Blaschko, and Andrea Vedaldi. Fine-grained visual classification of aircraft. *arXiv preprint arXiv:1306.5151*, 2013.

Mark D. McDonnell, Dong Gong, Amin Parvaneh, Ehsan Abbasnejad, and Anton van den Hengel. Ranpac: Random projections and pre-trained models for continual learning. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Junghun Oh, Sungyong Baik, and Kyoung Mu Lee. Closer: Towards better representation learning for few-shot class-incremental learning. In *European Conference on Computer Vision*, 2024.

Keon-Hee Park, Kyungwoo Song, and Gyeong-Moon Park. Pre-trained vision and language transformers are few-shot incremental learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 23881–23890, 2024.

Can Peng, Kun Zhao, Tianren Wang, Meng Li, and Brian C. Lovell. Few-shot class-incremental learning from an open-set perspective. In *European Conference on Computer Vision*, pp. 382–397, 2022.

Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H. Lampert. icarl: Incremental classifier and representation learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2001–2010, 2017.

James Seale Smith, Leonid Karlinsky, Vyshnavi Gutta, Paola Cascante-Bonilla, Donghyun Kim, Assaf Arbelle, Rameswar Panda, Rogerio Feris, and Zsolt Kira. Coda-prompt: Continual decomposed attention-based prompting for rehearsal-free continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.

Jake Snell, Kevin Swersky, and Richard S. Zemel. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Zeyin Song, Yifan Zhao, Yujun Shi, Peixi Peng, Li Yuan, and Yonghong Tian. Learning with fantasy: Semantic-aware virtual contrastive constraint for few-shot class-incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 24183–24192, 2023.

Yu-Ming Tang, Yi-Xing Peng, Jingke Meng, and Wei-Shi Zheng. Rethinking few-shot class-incremental learning: Learning from yourself. In *European Conference on Computer Vision*, 2024.

Xiaoyu Tao, Xiaopeng Hong, Xinyuan Chang, Songlin Dong, Xing Wei, and Yihong Gong. Few-shot class-incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12183–12192, 2020.

Grant Van Horn, Oisin Mac Aodha, Yang Song, Yin Cui, Chen Sun, Alex Shepard, Hartwig Adam, Pietro Perona, and Serge Belongie. The inaturalist species classification and detection dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8769–8778, 2018.

Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. In *Advances in Neural Information Processing Systems*, volume 29, 2016.

Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. Technical report, California Institute of Technology, 2011.

Fu-Yun Wang, Da-Wei Zhou, Han-Jia Ye, and De-Chuan Zhan. Foster: Feature boosting and compression for class-incremental learning. In *European Conference on Computer Vision*, 2022a.

Qi-Wei Wang, Da-Wei Zhou, Yi-Kai Zhang, De-Chuan Zhan, and Han-Jia Ye. Few-shot class-incremental learning via training-free prototype calibration. In *Advances in Neural Information Processing Systems*, volume 36, pp. 15060–15076, 2023.

Qinzhe Wang, Zixuan Chen, Keke Huang, Xiu Su, Chunhua Yang, and Chang Xu. Consistency-driven calibration and matching for few-shot class incremental learning. In *International Conference on Learning Representations*, pp. 1255–1279, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/030bfd0d78556d4cf3301cb61866600b-Paper-Conference.pdf.

Xuan Wang, Zhong Ji, Xiyao Liu, Yanwei Pang, and Jungong Han. On the approximation risk of few-shot class-incremental learning. In *European Conference on Computer Vision*, 2024.

Zifeng Wang, Zizhao Zhang, Sayna Ebrahimi, Ruoxi Sun, Han Zhang, Chen-Yu Lee, Xiaoqi Ren, Guolong Su, Vincent Perot, Jennifer Dy, and Tomas Pfister. Dualprompt: Complementary prompting for rehearsal-free continual learning. In *European Conference on Computer Vision*, 2022b.

Zifeng Wang, Zizhao Zhang, Chen-Yu Lee, Han Zhang, Ruoxi Sun, Xiaoqi Ren, Guolong Su, Vincent Perot, Jennifer Dy, and Tomas Pfister. Learning to prompt for continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 139–149, 2022c.

Yibo Yang, Haobo Yuan, Xiangtai Li, Zhouchen Lin, Philip H. S. Torr, and Dacheng Tao. Neural collapse inspired feature-classifier alignment for few-shot class-incremental learning. In *International Conference on Learning Representations*, 2023.

Chi Zhang, Nan Song, Guosheng Lin, Yun Zheng, Pan Pan, and Yinghui Xu. Few-shot incremental learning with continually evolved classifiers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12455–12464, 2021.

Da-Wei Zhou, Fu-Yun Wang, Han-Jia Ye, Liang Ma, Shiliang Pu, and De-Chuan Zhan. Forward compatible few-shot class-incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9046–9056, 2022.

Da-Wei Zhou, Zi-Wen Cai, Han-Jia Ye, De-Chuan Zhan, and Ziwei Liu. Revisiting class-incremental learning with pre-trained models: Generalizability and adaptivity are all you need. *arXiv preprint arXiv:2303.07338*, 2023a.

Da-Wei Zhou, Han-Jia Ye, Liang Ma, Di Xie, Shiliang Pu, and De-Chuan Zhan. Few-shot class-incremental learning by sampling multi-phase tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(11):12816–12831, 2023b.

Haichen Zhou, Yixiong Zou, Ruixuan Li, Yuhua Li, and Kui Xiao. Delve into base-novel confusion: Redundancy exploration for few-shot class-incremental learning. *arXiv preprint arXiv:2405.04918*, 2024.

APPENDIX

A IMPLEMENTATION DETAILS AND ADDITIONAL ANALYSES

A.1 DATASETS AND FSCIL PROTOCOLS

We evaluate on CUB-200 (Wah et al., 2011), CIFAR-100 (Krizhevsky, 2009), mini-ImageNet (Vinyals et al., 2016), FGVC-Aircraft (Maji et al., 2013), and iNF200, a 200-class subset of iNaturalist (Van Horn et al., 2018). Table 7 summarizes the dataset statistics and public FSCIL protocols. The "Base classes" column reports the number of classes in S_0, while "Incremental" reports the number of sessions and their way/shot configuration.

Table 7: Dataset statistics and FSCIL session configurations.

Dataset	Classes	Images	Train / test	Base classes	Incremental
CUB-200	200	11,788	5,994 / 5,794	100	$10\times$10-way 5-shot
CIFAR-100	100	60,000	50,000 / 10,000	60	$8\times$5-way 5-shot
mini-ImageNet	100	60,000	50,000 / 10,000	60	$8\times$5-way 5-shot
FGVC-Aircraft	100	10,000	3,334 / 3,333	50	$10\times$5-way 5-shot
iNF200	200	12,000	10,000 / 2,000	100	$10\times$10-way 5-shot

A.2 TRAINING DETAILS AND FINAL HYPERPARAMETERS

All experiments use an ImageNet-21K-pretrained ViT-B/16 with 6 shared deep VPT prompts inserted into Transformer blocks 1–8. CUB-200, CIFAR-100, and mini-ImageNet use the original pretrained ViT checkpoint and its corresponding evaluation transform. FGVC-Aircraft and iNF200 follow the benchmark protocol of LGSP-Prompt (Jiang et al., 2025). Base training has two steps: 10 epochs of VPT adaptation with representation-consistency regularization, followed by 120 epochs of cosine-head fitting on frozen features with attentive calibration. Incremental sessions update no parameters and retain class-indexed support features and fixed tangent references. For each base class b, the alignment weight a_b is its directional concentration κ_b divided by the across-class mean concentration, clipped to $[0.5, 1.5]$. A numerical floor of 10^{-6} is applied before taking the ratio.

Table 8 reports the settings used for each benchmark. The calibration mechanisms are shared across datasets, while the loss weights and interpolation coefficient are dataset-specific.

Table 8: Dataset-specific hyperparameters used by BASIL. Shared settings are listed in Table 9.

Dataset	λ_{rc}	λ_{att}	M	σ	α	Head LR	Head WD
CUB-200	0.05	16	96	0.025	0.75	0.03	10^{-4}
CIFAR-100	0.03	1	96	0.025	0.50	0.03	10^{-4}
mini-ImageNet	0.05	16	96	0.025	0.75	0.03	10^{-4}
FGVC-Aircraft	0.05	16	96	0.025	0.75	0.03	10^{-4}
iNF200	0.05	16	96	0.025	0.75	0.03	10^{-4}

Table 9: Shared optimization and scoring settings across datasets.

Setting	Value	Setting	Value
Feature dimension d	768	Support shots K	5
Score scale τ	learned	Head/evaluation seed	42
VPT learning rate	0.003	VPT batch size	48
VPT weight decay	10^{-4}	Optimizer	AdamW
Head batch size	512	VPT initialization std.	0.02
Head-scale init./cap	20 / 80	Image size	224
VPT epochs	10	Head epochs	120
VPT tokens	6	VPT blocks	1–8
Alignment clip $[a_{\min}, a_{\max}]$	$[0.5, 1.5]$	Numerical floor ϵ_{num}	10^{-6}

A.3 Additional Benchmark Results

Tables 10 and 11 present session-wise results on FGVC-Aircraft and iNF200. *Avg.* denotes the arithmetic mean over S_0–S_{10}.

Table 10: Session-wise top-1 accuracy (%) on FGVC-Aircraft.

Method	S0	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Avg.
Yourself (Tang et al., 2024)	26.29	20.37	17.05	14.82	13.16	11.87	10.83	9.97	9.24	8.62	8.08	13.66
Approximation-Risk (Wang et al., 2024)	21.67	19.48	18.90	17.57	16.50	16.68	15.82	15.21	14.20	13.85	12.97	16.62
PriViLege (Park et al., 2024)	22.75	20.79	21.85	20.96	20.27	22.32	21.53	21.67	20.43	20.18	20.04	21.16
LGSP-Prompt (Jiang et al., 2025)	27.67	24.82	26.30	27.38	25.25	26.76	26.37	25.49	25.50	24.83	23.87	25.84
BASIL	**45.56**	**42.12**	**42.10**	**40.86**	**39.31**	**40.36**	**38.52**	**38.02**	**36.00**	**35.31**	**34.92**	**39.37**

Table 11: Session-wise top-1 accuracy (%) on iNF200/iNaturalist.

Method	S0	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Avg.
Yourself (Tang et al., 2024)	61.90	55.73	50.50	45.31	40.93	35.07	30.75	26.18	22.72	19.47	16.10	36.79
Approximation-Risk (Wang et al., 2024)	66.30	60.18	57.25	53.62	50.43	47.67	44.88	42.41	40.11	38.32	36.75	48.90
PriViLege (Park et al., 2024)	63.10	60.73	59.50	58.31	57.93	56.07	53.75	52.18	51.72	49.47	48.10	55.53
LGSP-Prompt (Jiang et al., 2025)	67.00	63.27	62.83	60.92	59.57	57.67	56.19	53.88	53.83	50.05	48.60	57.62
BASIL	**75.80**	**71.91**	**71.00**	**69.62**	**68.86**	**67.67**	**65.75**	**63.88**	**62.50**	**59.79**	**58.70**	**66.86**

A.4 Incremental Reference Storage and Scoring Cost

For each newly introduced class, BASIL stores one mean prototype, $K = 5$ real-support features, and $M = 96$ tangent references in $d = 768$ dimensions. With FP32 storage, the local-reference set requires $(K + M)d \times 4 = 310{,}272$ bytes (303 KiB) per class, in addition to the 3 KiB mean prototype. Thus the total per-class feature payload is 306 KiB, compared with 3 KiB for a prototype-only classifier: the additional storage is 303 KiB, or 10,100% of the prototype-only payload. At the final CUB-200 session, 100 incremental classes require 29.59 MiB for their local references. The measured local-reference scoring time is 0.0544 ms per query at this session, excluding the shared ViT feature extraction.

B Supplementary Theoretical Analysis

B.1 Finite-Sample Prototype Error and Base-Side Displacement

This subsection distinguishes two quantities relevant to BASIL: the finite-sample uncertainty of an empirical prototype and the additional score displacement introduced by a learned base prototype.

Consider n independent unit features $\mathbf{f}_i$ from class c. Define the population mean vector, its length, and its direction by

$$\boldsymbol{\mu}_c = \mathbb{E}[\mathbf{f}_i], \qquad \rho_c = \|\boldsymbol{\mu}_c\|_2 > 0, \qquad \mathbf{p}_c^\star = \boldsymbol{\mu}_c/\rho_c. \tag{11}$$

Let $\bar{\mathbf{f}}_c = n^{-1}\sum_i \mathbf{f}_i$ and $\mathbf{p}_c^{(n)} = \bar{\mathbf{f}}_c/\|\bar{\mathbf{f}}_c\|_2$. For an incremental class, $n = K$ and $\mathbf{p}_c^{(n)} = \mathbf{p}_c$. For a base class, $n = n_c \gg K$, where n_c is its number of base-session training samples, and $\mathbf{p}_c^{(n)} = \mathbf{p}_c^0$. The unnormalized sample mean is unbiased, but normalization is nonlinear, so the relevant object is a high-probability directional error rather than an unbiasedness claim.

Because each coordinate of a unit feature lies in $[-1, 1]$, coordinate-wise Hoeffding bounds followed by a union bound imply that, for any $\delta \in (0, 1)$, with probability at least $1 - \delta$,

$$\|\bar{\mathbf{f}}_c - \boldsymbol{\mu}_c\|_2 \le \varepsilon_n(\delta) := \sqrt{\frac{2d\log(2d/\delta)}{n}}. \tag{12}$$

If $\varepsilon_n(\delta) \le \rho_c/2$, normalization is Lipschitz in this neighborhood, giving

$$\|\mathbf{p}_c^{(n)} - \mathbf{p}_c^\star\|_2 \le \frac{2\varepsilon_n(\delta)}{\rho_c}, \qquad \left|\tau\mathbf{f}(x)^\top(\mathbf{p}_c^{(n)} - \mathbf{p}_c^\star)\right| \le \frac{2\tau\varepsilon_n(\delta)}{\rho_c}. \tag{13}$$

The bound is intentionally conservative. Its role is to expose the $n^{-1/2}$ dependence. Under comparable dimension, confidence, and concentration ρ_c, a dense base centroid is expected to be more stable than a K-shot incremental prototype. This statement is probabilistic and does not require a strict class-wise ordering.

For a base class b and an incremental class c, the uncalibrated pairwise margin is

$$s_b^{\mathrm{uncal}}(x) - s_c^{\mathrm{uncal}}(x) = \tau\mathbf{f}(x)^\top(\mathbf{p}_b^0 - \mathbf{p}_c) + \tau\mathbf{f}(x)^\top(\mathbf{w}_b - \mathbf{p}_b^0). \tag{14}$$

The first term compares empirical class references. The second is a base-only score displacement caused by replacing the dense centroid $\mathbf{p}_b^0$ with the discriminative base prototype $\mathbf{w}_b$. Its sign depends on the query, so BASIL does not assume that discriminative base prototypes always increase base scores. Denote this pre-calibration displacement by $\delta_b^{\mathrm{pre}}(x) = \tau\mathbf{f}(x)^\top(\mathbf{w}_b - \mathbf{p}_b^0)$. After attentive calibration, the corresponding displacement is $\delta_b^{\mathrm{att}}(x) = \tau\mathbf{f}(x)^\top(\widetilde{\mathbf{w}}_b - \mathbf{p}_b^0)$, whose magnitude satisfies

$$\left|\tau\mathbf{f}(x)^\top\widetilde{\mathbf{w}}_b - \tau\mathbf{f}(x)^\top\mathbf{p}_b^0\right| \le \tau\|\widetilde{\mathbf{w}}_b - \mathbf{p}_b^0\|_2 = \tau\sqrt{2(1 - \widetilde{\mathbf{w}}_b^\top\mathbf{p}_b^0)}. \tag{15}$$

Thus, reducing the prototype–centroid angular error reduces the worst-case magnitude of the base-side displacement associated with Eq. (14). The result identifies what attentive calibration controls. It does not claim that every individual logit decreases. Expansive calibration addresses a different limitation: it does not create independent observations or improve the $K^{-1/2}$ sampling rate, but supplements local coverage around the observed incremental prototype, as analyzed next.

B.2 Geometry and Coverage of Tangent References

For an incremental prototype $\mathbf{p}$, BASIL samples an isotropic Gaussian vector, projects it with $\mathbf{I} - \mathbf{p}\mathbf{p}^\top$, and normalizes the projection to obtain a unit tangent direction $\mathbf{d} \perp \mathbf{p}$. Because $\|\mathbf{p}\|_2 = \|\mathbf{d}\|_2 = 1$, a reference with tangent perturbation magnitude σ has the exact form

$$\mathbf{r} = \frac{\mathbf{p} + \sigma\mathbf{d}}{\sqrt{1 + \sigma^2}}. \tag{16}$$

These references exist only in feature space. They are neither generated images nor additional semantic observations.

Let $\mathbf{u} = \mathbf{f}(x)$ be a unit query whose angle to $\mathbf{p}$ satisfies $\theta \in (0, \pi/2)$. There is a unit tangent direction $\mathbf{v} \perp \mathbf{p}$ such that

$$\mathbf{u} = \cos\theta\,\mathbf{p} + \sin\theta\,\mathbf{v}. \tag{17}$$

This is the orthogonal decomposition of a unit vector relative to $\mathbf{p}$, not an assumption about how the data are generated. Combining it with Eq. (16) yields

$$\mathbf{u}^\top\mathbf{r} - \mathbf{u}^\top\mathbf{p} = \frac{\cos\theta + \sigma\sin\theta\,\mathbf{v}^\top\mathbf{d}}{\sqrt{1 + \sigma^2}} - \cos\theta. \tag{18}$$

Consequently, a tangent reference scores above the empirical prototype exactly when

$$\mathbf{v}^\top \mathbf{d} > \frac{\cos\theta\left(\sqrt{1+\sigma^2}-1\right)}{\sigma \sin\theta}. \tag{19}$$

Whenever the right-hand side is below one, the successful directions form a nonempty spherical cap around the query's tangent direction $\mathbf{v}$. Define the probability that one isotropically sampled tangent direction hits this cap as

$$P_{\text{cap}}(\theta,\sigma) = \Pr_{\mathbf{d}}\left[\mathbf{v}^\top \mathbf{d} > \frac{\cos\theta\left(\sqrt{1+\sigma^2}-1\right)}{\sigma \sin\theta}\right]. \tag{20}$$

Then M independent references hit it at least once with probability

$$1-\left(1-P_{\text{cap}}(\theta,\sigma)\right)^M. \tag{21}$$

This gives a geometric condition under which a finite reference set can improve the similarity of a query with nonzero angular deviation from the empirical prototype. The hit probability can still be small in high dimensions, so the result does not guarantee coverage for every query. It also exposes the limitation of scoring solely by $\max_{\mathbf{r}\in\mathcal{R}_c} \mathbf{u}^\top \mathbf{r}$: an incorrect class may obtain a spuriously high similarity. BASIL therefore retains the empirical prototype and interpolates its score with the local score using the fixed interpolation coefficient α, rather than replacing the prototype score. This design balances local coverage against spurious competition.

C Additional Calibration Diagnostics

C.1 Final-Session Group Confusion

Figure 5 compares the final-session confusion matrices of uncalibrated prototype composition and BASIL on CUB-200. Base classes form group S_0, while each subsequent group contains the classes introduced in one incremental session. Row normalization makes errors across groups comparable. BASIL reduces base-to-incremental errors from 21.65% to 1.89%, although assignments from some incremental groups to base classes increase. The matrices characterize the combined calibration, rather than attributing the change to an individual component.

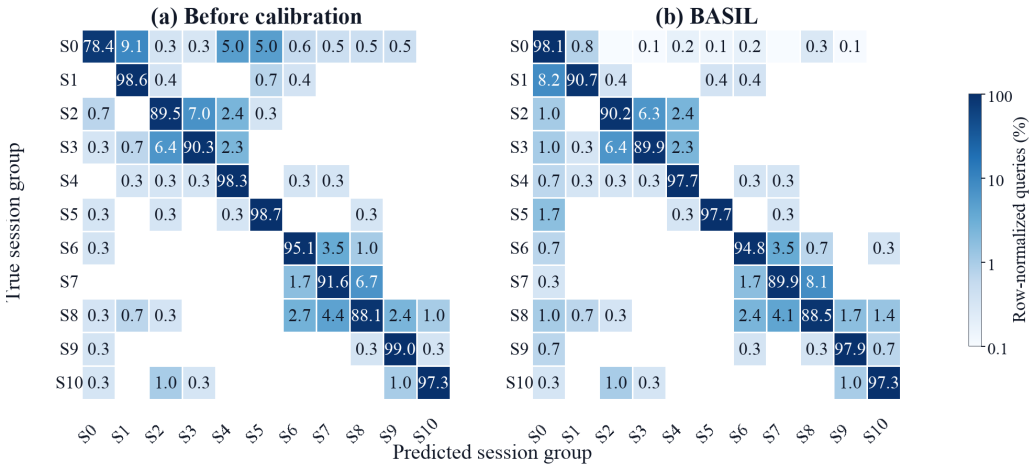


Figure 5: Row-normalized session-group confusion at the final CUB-200 session.

C.2 Attentive Calibration: Concentration-Aware Alignment

Figure 6 relates base-class concentration to the change in prototype–centroid angle relative to uniform alignment. A negative angle difference indicates stronger alignment under concentration-aware weighting. Across the 100 base classes, concentration and angle difference have Pearson correlation

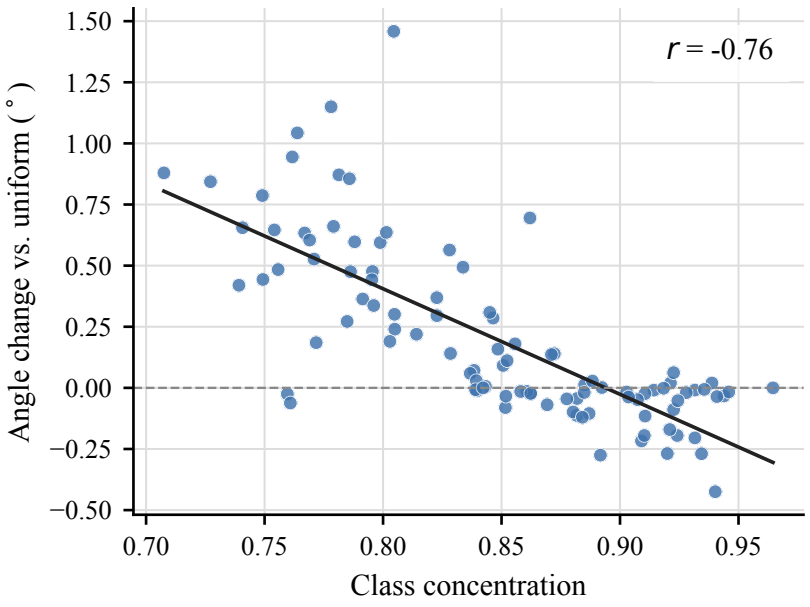


Figure 6: Base-class concentration versus the change in prototype–centroid angle relative to uniform alignment on CUB-200.

$r = -0.76$: concentrated classes are drawn closer to their empirical centroids, whereas dispersed classes retain more discriminative displacement.

We also compare concentration-aware alignment with a uniform weight $a_b = 1$ for every base class. Both variants use the same alignment strength and local-reference rule. Class-dependent weights yield a CUB-200 incremental-session average of 86.35%, compared with 85.26% for uniform alignment.

Table 12: Uniform versus concentration-aware base alignment on CUB-200.

Alignment weight	S_0	Inc. Avg.	Final
Uniform, $a_b = 1$	88.94	85.26	84.19
Concentration-aware, a_b	89.28	86.35	85.66

C.3 Expansive Calibration: Local Coverage of Incremental Queries

Figure 7 examines whether local references cover incremental-class queries that deviate from their empirical prototypes. The 2,930 queries at the final CUB-200 session are divided into five equal-sized groups by their angle to the true-class prototype. Ground-truth labels are used only for this analysis, not for prediction. Similarity gain is the best cosine similarity among a class's real supports, tangent references, or both, minus its mean-prototype similarity. For the combined reference set, mean gain increases from 0.0016 in the nearest-angle group to 0.0205 in the farthest, indicating greater local coverage away from the prototype.

For 28.67% of these queries, the best real-support score exceeds the mean-prototype score. For 72.46%, the best tangent-reference score exceeds the best real-support score. Taking the maximum over 96 tangent references also yields a score above the mean for every query. This alone does not establish better classification, because maximizing over references can also inflate incorrect-class scores. BASIL therefore interpolates the local-reference score with the empirical-prototype score using the fixed coefficient α, rather than replacing the latter with a maximum.

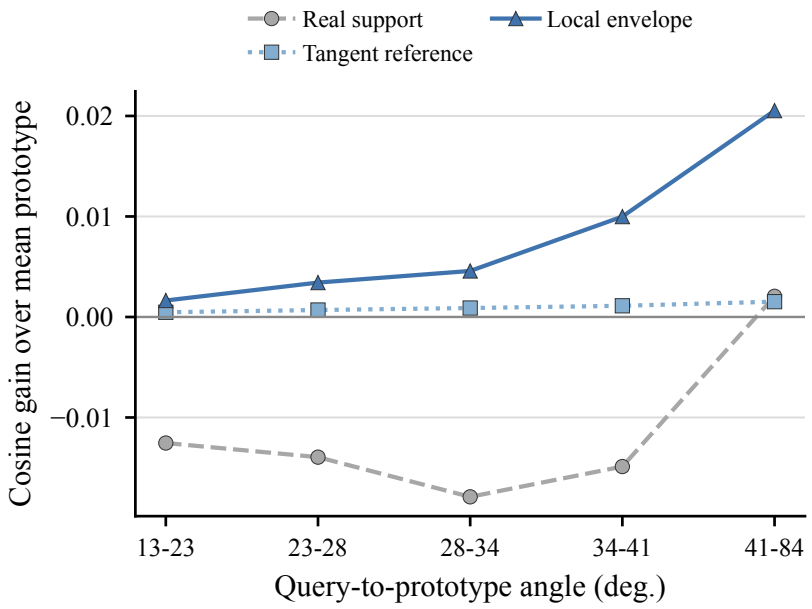


Figure 7: True-class similarity gain from real supports, tangent references, and their maximum relative to the mean prototype, grouped by query-to-prototype angle on CUB-200.

To examine reference-sampling variability, we keep the CUB-200 representation, classifier, and real supports fixed and evaluate five independently sampled tangent-reference sets. Table 13 reports the resulting mean and sample standard deviation. Tangent references improve the incremental-session average from 85.50% to 86.35% in this paired setting.

Table 13: CUB-200 tangent-reference sampling with fixed features, classifier, and supports.

Local references	Inc. Avg.	Final
Real supports only	85.50	84.47
Real supports and tangent references	86.35 ± 0.02	85.66 ± 0.02

C.4 Interaction Between Base Alignment and Local Scoring

Figure 8 reports the interaction between base-prototype alignment and local scoring across interpolation coefficients. A positive interaction means that alignment makes local scoring less harmful or more beneficial than it was without alignment; it does not establish an absolute accuracy gain. The coefficient sweep characterizes the trained classifiers and was not used to select the benchmark settings.

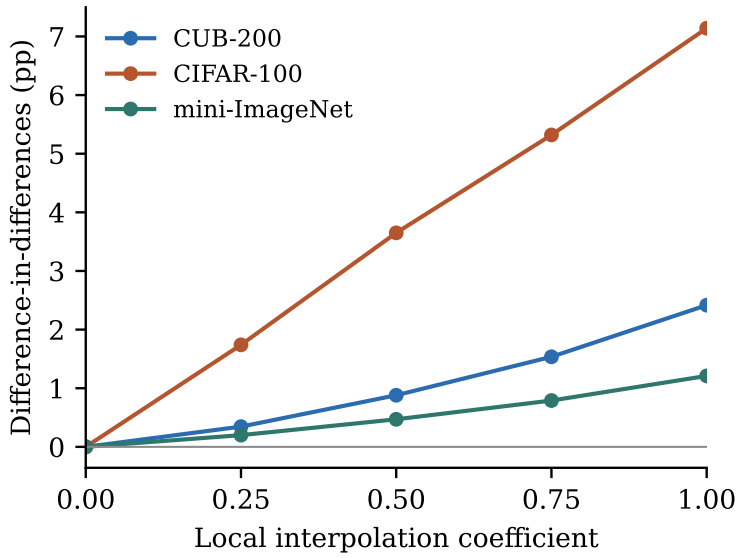


Figure 8: Accuracy interaction between attentive and expansive calibration across interpolation coefficients.