ICML
International Conference
On Machine Learning

SHENZHEN UNIVERSITY
深圳大学

Bayes-inspired Integration of Pretrained Priors and Few-Shot Evidence for Few-Shot Classification

Authors: Mingyang Zhou, Xiaoxuan Zhang, Gang Liu, Yuhong Feng, Xiaoqun Wu, Hao Liao, Rui Mao

Shenzhen University

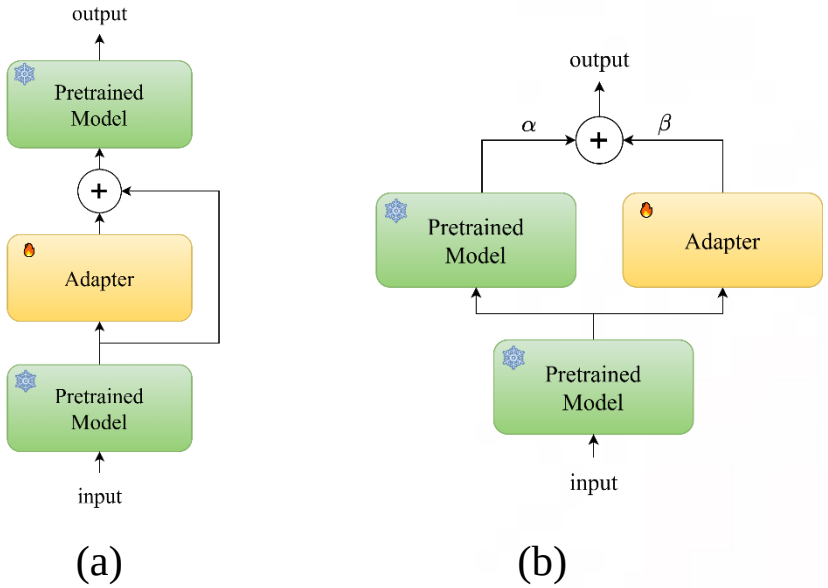
zmy@szu.edu.cn (Mingyang Zhou), 2410104019@mails.szu.edu.cn (Xiaoxuan Zhang)

July 6–11, 2026

Motivation

Existing few-shot methods combine pretrained models and support sets through heuristics such as residual connections, as shown in Fig. 1 (a), or weighted averaging, as shown in Fig. 1 (b).

Question: What is the relationship between pre-trained knowledge and few-shot evidence, and how to *optimally* integrate few-shot evidence into pre-trained models?



(a) (b)

Fig. 1: Heuristic strategies for combining pretrained models and few-shot adapter evidence

Problem Definition: A pretrained model is learned from a source dataset $D_0 = \{(\boldsymbol{x}_i^0, y_i^0)\}_{i=1}^{N_0}$ with labels $y_i^0 \in Y_0$. During few-shot adaptation, the model receives a support set $D_{train} = \{(\boldsymbol{x}_i^s, y_i^s)\}_{i=1}^{CK}$ with C novel classes and K labeled examples per class, where $y_i^s \in Y_{novel}$ and $Y_0 \cap Y_{novel} = \emptyset$.

The goal is to estimate the posterior over novel classes for a test input $\boldsymbol{x}$, conditioned on both information sources (Fig. 2):

$$P(\mathrm{y}|\boldsymbol{x}, D_0, D_{train}), \qquad y \in Y_{novel} \qquad (1)$$

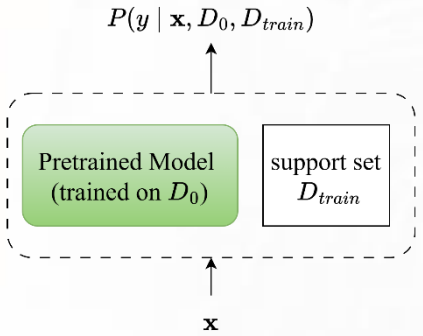


Fig. 2: Bayesian posterior estimation using pretrained model and support set.

Method

Assumption 1 (Conditional Independence of Information Sources).

$$P(D_0, D_{train}|y, \boldsymbol{x}) = P(D_0|y, \boldsymbol{x}) \cdot P(D_{train}|y, \boldsymbol{x}) \qquad (2)$$

This is justified via a latent variable model:

- In latent variable models, observed data are generated from shared latent variables z, which represent the underlying real-world structure.
- Different datasets capture the same world from different viewpoints; they depend only on z, not directly on each other.
- Therefore, conditional independence holds given z:

$$P(D_0, D_{train}|z) = P(D_0|z) \cdot P(D_{train}|z) \qquad (3)$$

We can derive

$$P(D_0, D_{train}|(\boldsymbol{x}, y)) \propto P(D_0|(\boldsymbol{x}, y)) \cdot P(D_{train}|(\boldsymbol{x}, y)) \qquad (4)$$

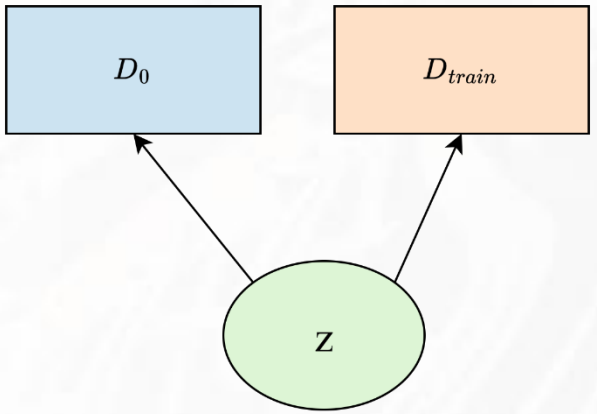


Fig. 3: Conditional independence of D_0 and D_{train} under the latent visual world variable z

Method

Assumption 2 (Weakly Informative Prior).

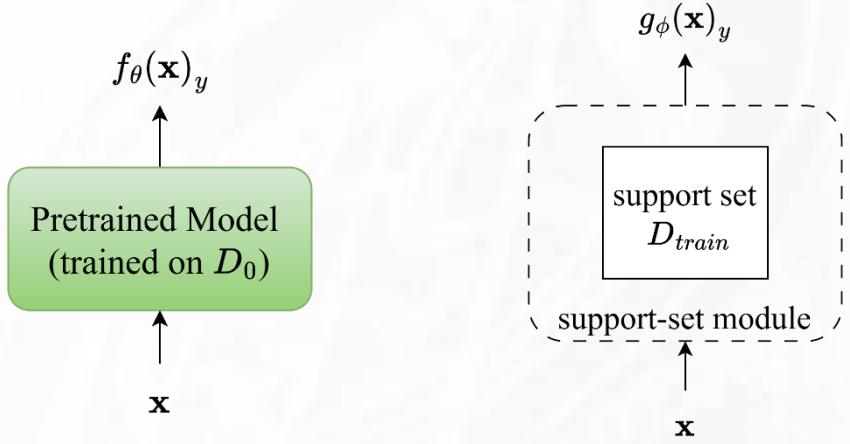
$$P(y|\boldsymbol{x}) \approx \frac{1}{|Y_{novel}|} \quad (5)$$

Under Assumptions 1 and 2, the joint posterior is approximately proportional to the product of the individual posteriors:

$$P(y|\boldsymbol{x}, D_0, D_{train}) \approx \frac{1}{\tilde{Z}(\boldsymbol{x})} \cdot P(y|\boldsymbol{x}, D_0) \cdot P(y|\boldsymbol{x}, D_{train}) \quad (6)$$

Then we use pretrained model $f_\theta(\cdot)$ as Approximate Posterior, support-set module $g_\phi(\cdot)$ as Likelihood Estimator:

$$f_\theta(\boldsymbol{x})_{\mathrm{y}} \approx P(\mathrm{y}|\boldsymbol{x}, D_0), \quad g_\phi(\boldsymbol{x})_y \approx P(\mathrm{y}|\boldsymbol{x}, D_{train}) \quad (7)$$



(a) pretrained model (b) support-set module

Under Eq. (6) and Eq. (7), our **Central Design Principle (Additive Logit Integration)** takes the following form:

$$\log P(\mathrm{y}|\boldsymbol{x}, D_0, D_{train}) \approx \log f_\theta(\boldsymbol{x})_{\mathrm{y}} + \log g_\phi(\boldsymbol{x})_y + constant. \quad (8)$$

Method

Our Central Design Principle yields two actionable guidelines:

- **Independent design:** The pretrained model f_θ and the support-set module g_ϕ can be developed separately.
- **Additive combination:** Optimal prediction emerges from simply summing their logits.

Although conditional independence is an approximation, it provides a principled rationale for decoupling the two information sources.

In practice, we permit limited interaction through joint optimization to capture residual dependencies.

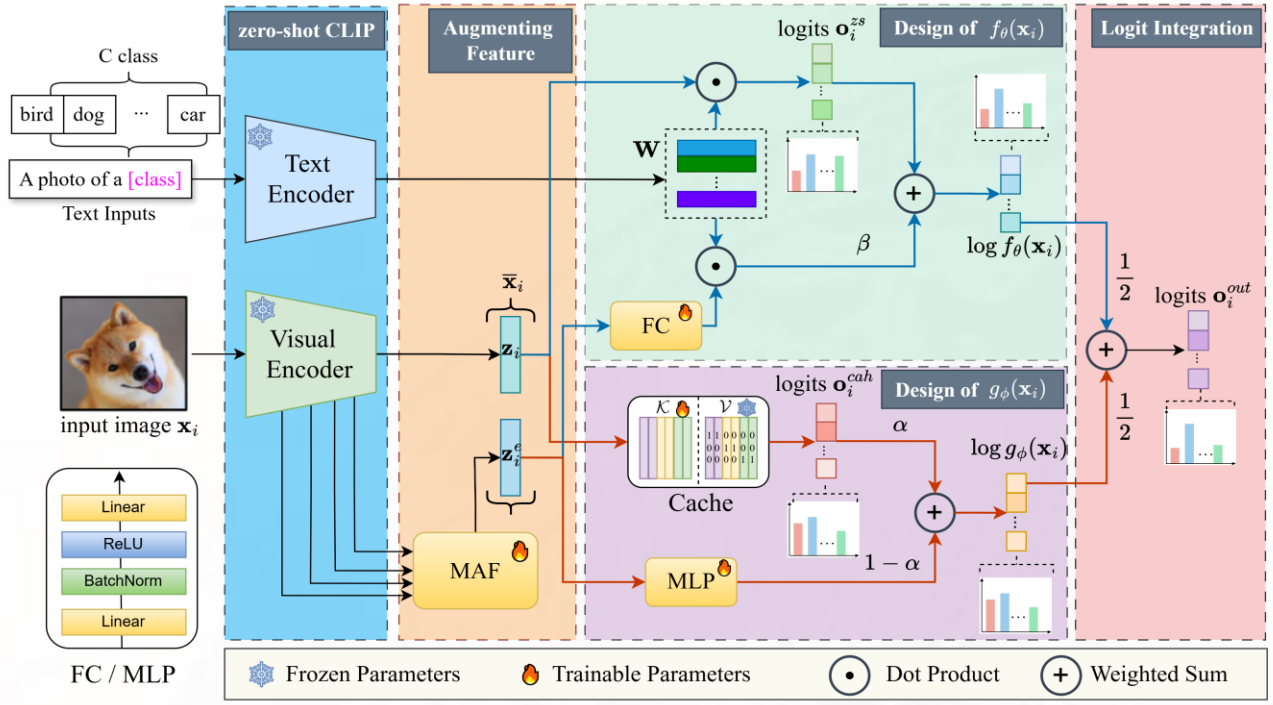


Fig. 4: **The Bayesian-inspired Optimal Integration Framework (BOIF)**. The architecture utilizes a frozen CLIP backbone to extract features. The framework consists of two independent paths: $f_\theta(\boldsymbol{x}_i)$ generates the refined prior logits; $g_\theta(\boldsymbol{x}_i)$ derives the support-set posterior logits.

深圳大学 SHENZHEN UNIVERSITY

Results

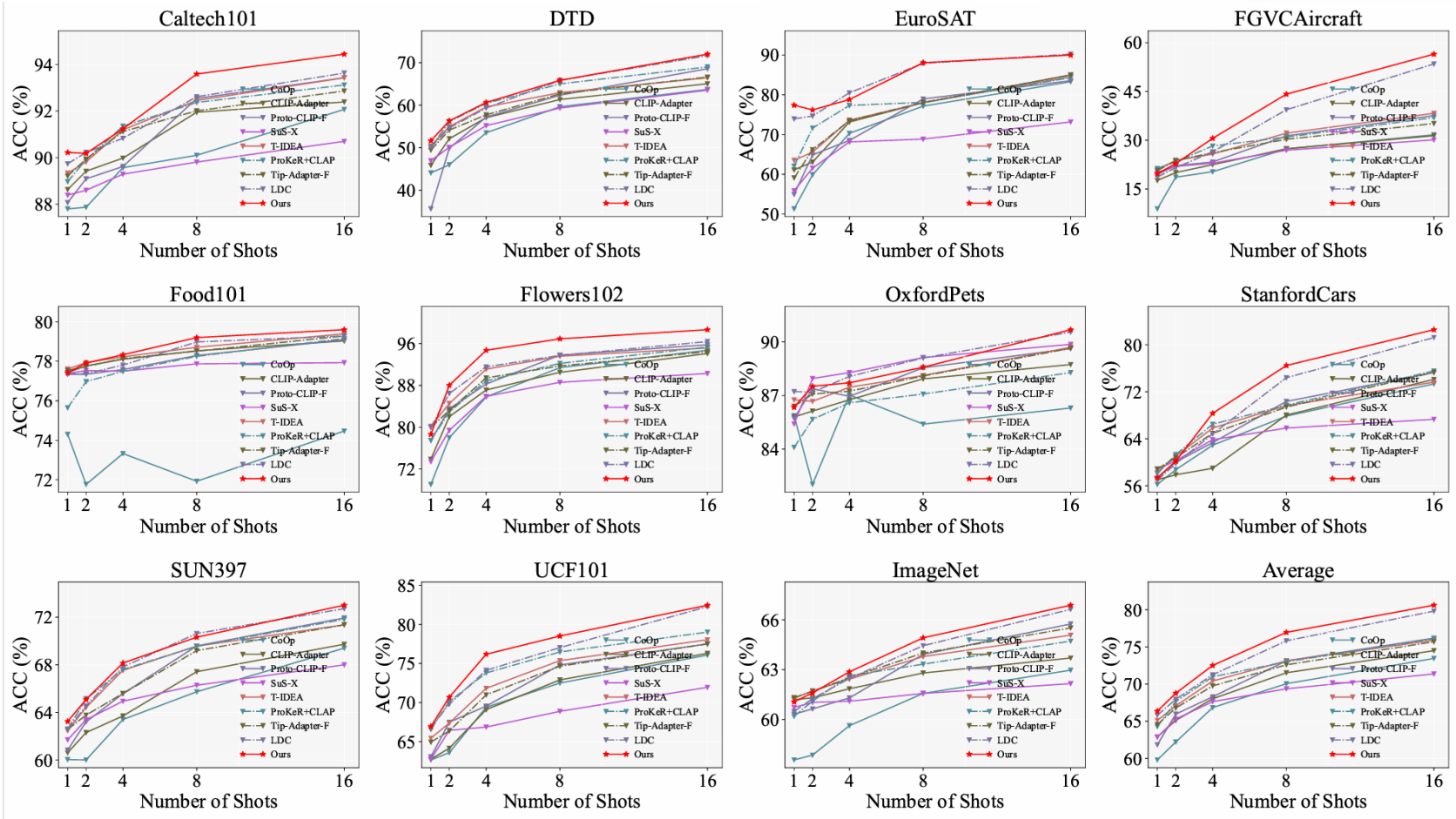


Fig. 5: Classification accuracy performance comparison on 11 datasets, and the last one is the average performance on these 11 datasets.

深圳大学 SHENZHEN UNIVERSITY

Results

Table 1. Accuracy comparison of different methods over 11 datasets. '-' denotes unavailable results. A higher value is better.

Method	K-Shot Accuracy (%)					
	0	1	2	4	8	16
ZS-CLIP[ICML2021]	58.87	-	-	-	-	-
LP-CLIP[ICML2021]	-	36.67	47.61	57.19	64.98	71.10
CoOp[IJCV2022]	-	59.80	62.21	66.84	70.05	73.45
Tip-Adapter-F[ECCV2022]	-	64.55	66.79	69.76	72.59	75.69
SuS-X[ICCV2023]	-	62.87	65.29	67.64	69.37	71.36
CLIP-Adapter[IJCV2024]	-	62.90	65.11	68.02	71.52	74.50
Proto-CLIP-F[IROS2024]	-	61.84	65.96	68.29	73.13	76.18
ProKeR+CLAP[CVPR2025]	-	64.28	67.66	71.07	73.09	76.11
LDC[CVPR2025]	-	65.71	67.92	71.17	75.79	79.78
T-IDEA[PR2026]	-	65.11	67.07	70.41	73.12	75.91
Ours	-	**66.35↑**	**68.77↑**	**72.50↑**	**76.96↑**	**80.61↑**

Table 2. Comparison of different methods under OOD setting. A higher value is better.

	Method	Source	Target	
		ImageNet	V2	Sketch
ResNet-50	ZS-CLIP	60.33	53.27	35.44
	LP-CLIP	56.13	45.61	19.13
	CoOp	62.95	54.58	31.04
	CoCoOp	62.81	55.72	34.48
	CALIP-FS	65.81	55.98	35.37
	Tip-Adapter	62.03	54.60	35.90
	Tip-Adapter-F	65.51	57.11	36.00
	CLIP-Adapter	63.59	55.69	35.68
	ProKeR	64.47	56.08	**36.01**
	ProKeR+CLAP	64.72	56.12	35.32
	LDC	66.63	58.03	35.52
	Ours	**66.84**	**58.45**	36.00
ViT-B/16	ZS-CLIP	66.73	60.83	46.15
	CoOp	71.51	64.20	47.99
	CoCoOp	71.02	64.07	48.75
	MaPLe	70.72	64.07	49.15
	MMA	71.00	64.33	49.13
	RPO	71.67	65.13	49.27
	LDC	73.88	66.10	48.85
	Ours	**74.34**	**66.25**	**49.41**

Thank you!

zmy@szu.edu.cn (Mingyang Zhou), 2410104019@mails.szu.edu.cn (Xiaoxuan Zhang)