Appendix

A1. Training Procedure

In the few-shot image recognition setting, we first extract features from each image along with its naturalistic-views and geometric-views using the CLIP image encoder, capturing both natural variations and spatial transformations. Multi-variant semantic information of class names is simultaneously obtained via the CLIP text encoder. To handle outliers and enhance representation, uncertain classes are dynamically generated to absorb uncertain samples. A logistic regression classifier is then trained on the enriched support set features, and finally, the classifier is used to predict labels for the query images. The detailed description and algorithm of our MPA framework are provided in Algorithm 1.

Algorithm 1: The procedure of MPA to train the classifier.

Input: Target domain dataset D_t, CLIP image encoder f_θ, CLIP text encoder h_θ.

Output: Trained logistic regression parameters θ_{logic}.

while training **do**
  Randomly sample a few-shot task T from D_t.
  Enrich the support set representation via Hierarchical Multi-View Augmentation, as formulated in Eqs. (3) and (4).
  Extract multi-variant semantic representations by prompting the LLM with support set class names, as formulated in Eq. (2).
  Project the derived semantic embeddings into the visual space via h_θ; adapt dimensionality through vector duplication and incorporate into the support set.
  Synthesize uncertain-class samples leveraging the generative mechanism described in Eqs. (5) and (11), and integrate them into the support set.
  Perform supervised optimization over the enriched support set using a logistic regression classifier.
  Update the logistic regression parameters θ_{logic} accordingly.
end while
return θ_{logic} .

A2. AUCA Under Extreme Conditions

To evaluate AUCA's robustness, we introduce artificial uncertainty under the 5-way 5-shot setting. As shown in Table 9, AUCA consistently improves performance, achieving a 2.97% gain on CUB and a 2.39% average improvement on the target domain.

AUCA	CUB	Cars	CIFAR-FS	FC100
×	82.75	84.96	73.89	60.29
✓	**85.72**	**87.44**	**75.96**	**62.31**

Table 9: Average classification accuracy across target domain datasets, including those with artificially added uncertain samples, under the 5-way 5-shot setting.

Dataset	Recognition task	Mean	Variance
CUB	Fine-grained bird	0.3150	0.0011
Cars	Fine-grained car	0.3646	0.0014
Places	Scene	0.4727	0.0018
Plantae	Plantae	0.3236	0.0012
EuroSAT	Satellite images	0.2244	0.0013
CropDisease	Agricultural disease	0.2208	0.0014
miniImageNet	General few-shot	0.3947	0.0015
tieredImageNet	General few-shot	0.4257	0.0013
CIFAR-FS	General few-shot	0.3002	0.0009
FC100	General few-shot	0.2779	0.0011

Table 10: Mean values and variances of the dynamic factor λ across various datasets from different domains.

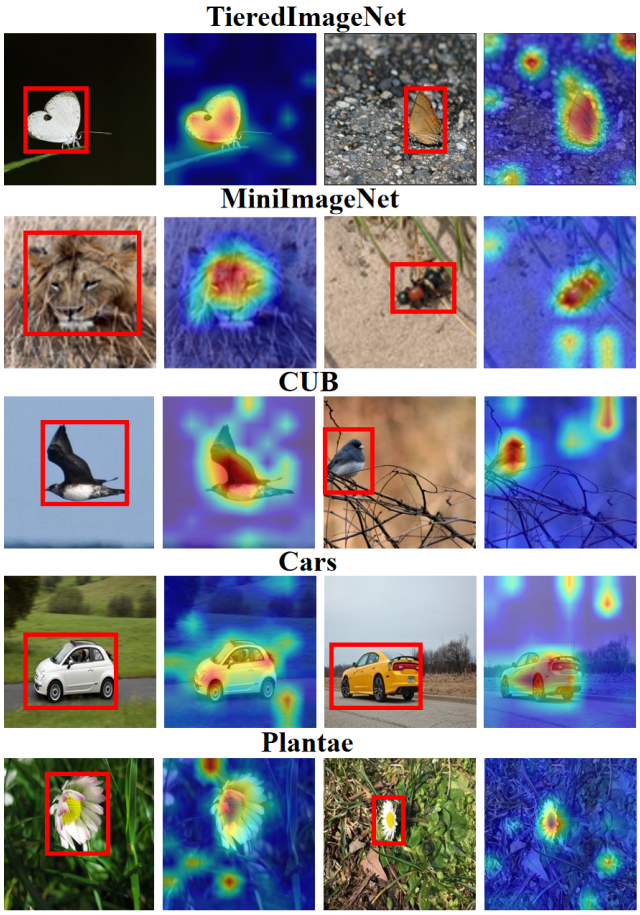

Figure 3: Feature visualization of MPA on public datasets.

A3. Statistical Analysis of the Dynamic Factor

In order to evaluate the adaptability of the AUCA module on different datasets, we calculate the mean and variance of λ. Specifically, we performed 1,000 tests per dataset to ensure reliable statistical results. As shown in Table 10, there are some differences in the mean values of λ on differ-

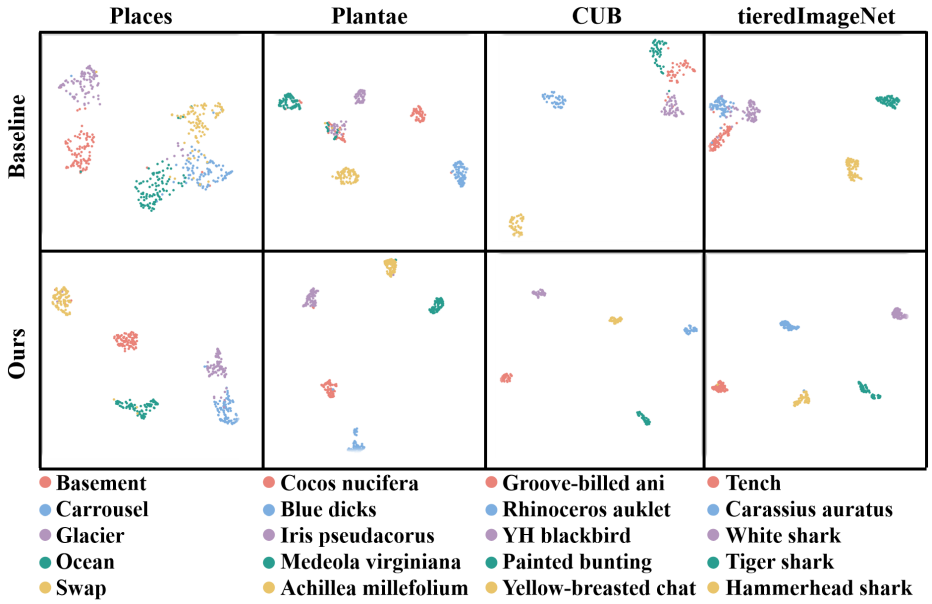

Figure 4: UMAP visualization.

ent domain datasets, which indicates that the AUCA module is able to adaptively generate uncertain classes according to the changes in the feature distributions, thus better adapting to different datasets. Moreover, the relatively small variances suggest that λ values remain stable within each dataset, demonstrating the robustness of the AUCA module in learning a consistent adaptation strategy.

A4. UMAP Visualization

Figure 4 presents 2D UMAP visualizations of MPA features on five target-domain datasets. For each dataset, compared with the baseline, MPA clearly separates the category boundaries, demonstrating the effectiveness of MPA.

A5. Feature Visualization

The results show that MPA effectively highlights target objects and captures more comprehensive semantic information, improving generalization. As shown in Fig. 3, MPA performs well on both single-domain and cross-domain datasets, highlighting its semantic understanding and robustness to data distribution shifts.

A6. Multimodal Information Richness

To highlight the comprehensive multimodal information captured by our method, we compare MPA with existing baseline methods in terms of the types of information utilized: raw images, multi-view images, class names, attributes, and multi-variant semantics.

As shown in Table 11, MPA leverages five types of information, yielding richer multimodal representations that improve prototype quality and few-shot performance.

A7. Efficiency Analysis

We evaluate computational efficiency on EuroSAT 5-way 1-shot. Table 12 shows memory (GB), runtime (s), and accuracy (%) for different module combinations. LMSE and HMA slightly increase cost while boosting performance.

Method	Img	MVI	Cls	Attr	MVS
SP-CLIP	✓	×	✓	×	×
CAML	✓	×	✓	×	×
SEVPro	✓	×	×	×	×
SemFew-Trans	✓	×	✓	✓	×
SPM	✓	×	✓	×	×
ECER-FSL	✓	×	×	✓	×
MLVLM	✓	×	✓	✓	×
MPA (Ours)	✓	✓	✓	✓	✓

Table 11: Comparison of the multimodal information richness across different methods. MVI represents multi-view images, and MVS represents multi-variant semantics.

LMSE	HMA	Mem. (G)	Time (s)	Acc. (%)
×	×	3.32	0.050	76.08
✓	×	3.36	0.058	83.03
×	✓	3.42	0.067	79.44

Table 12: Efficiency comparison of LMSE and HMA on EuroSAT 5-way 1-shot.

These results demonstrate that LMSE and HMA maintain high computational efficiency, with less than 3% increase in memory and under 0.02,s additional test time per image. The trade-off between accuracy and cost remains highly favorable, confirming that the proposed modules are lightweight and practical for few-shot learning tasks.