Bayesian Exploration of Pre-trained Models for Low-shot Image Classification

Supplementary Material

A. Datasets Preparation

The datasets employed in this work have been slightly modified to accommodate low-shot classification better. To ensure a fair comparison with previous works, in line with CaFo [?], we randomly sampled 1, 2, 4, 8, and 16 data points per class from ImageNet [?]. These sets are designated as 1, 2, 4, 8, and 16-shot training sets, with the ImageNet validation set serving as the test set. All samples from ImageNet-V2 [?] and ImageNet-Sketch [?] are exclusively used for testing purposes. For other datasets, we adhere to the same train/test/val splits as established by CaFo.

B. Additional Ablation Study

CLIP's Visual Encoders. For further performance enhancement on ImageNet [?], we attempt to change the backbone of the image encoder in CLIP from ResNet-50 to ViT-B/16. We provide the corresponding results in Tab. 1. It is easy to see that our method remains to surpass all the ensemble baselines consistently.

Shot	1	2	4	8	16
Ens-LP	41.60	51.75	59.82	65.42	69.86
Ens-LP†	69.81	71.11	71.45	73.05	74.20
Ens-CaFo	70.00	71.03	71.79	72.86	74.49
Ours	**70.70**	**71.48**	**72.62**	**73.96**	**75.22**

Table 1. Accuracy (%) on ImageNet when using the CLIP with a ViT-B/16 image encoder.

DALL-E Augmentation. Following CaFo [?], we also explore the impact of using synthetic images for data augmentation. According to [?], under the 1,2,4-shot setting, we use 8 synthetic images per class for augmentation. Under the 8, 16-shot setting, we use 2 synthetic images per class for augmentation. The results in Tab. 2 can serve as an ablation study on the DALL-E [?] augmentation. We can see that the use of synthetic images is intended to provide benefits when dealing with an extremely limited number of training samples, e.g., 1 or 2-shot setting. With data augmentation, our method also consistently outperforms other baselines. This demonstrates the effectiveness of our approach as well as its robustness against data augmentation.

Shot	1	2	4	8	16
Ens-LP	56.23	57.70	59.60	63.67	67.23
Ens-LP†	66.62	67.08	67.20	67.71	69.22
Ens-CaFo	65.19	66.02	66.65	67.45	68.85
Ours	**67.32**	**67.93**	**68.65**	**69.56**	**70.83**

Table 2. Accuracy (%) on ImageNet when using DALL-E augmentation.

C. Visualization of Uncertainty Estimates

We train our model on ImageNet [?] and then test on ImageNet-V2 [?], ImageNet-A [?], ImageNet-R [?], and Imagenet-Sketch [?] to get the uncertainty estimate distributions. The results in Fig. 1 align with the fact that ImageNet-V2 and ImageNet-A have similar distributions with ImageNet, while the distributions of ImageNet-R and Imagenet-Sketch are different from ImageNet.

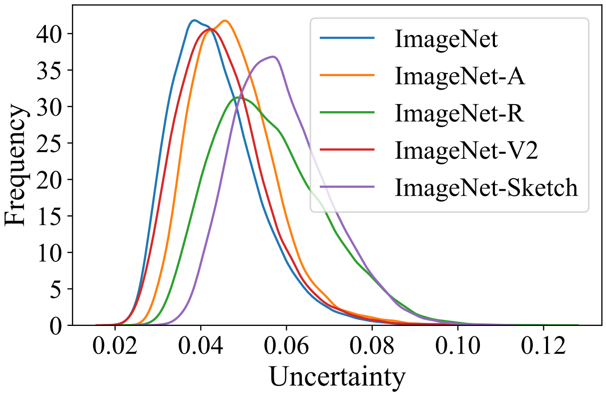

Figure 1. Histogram for uncertainty estimates. We evaluate our methods on ImageNet, ImageNet-V2, ImageNet-A, ImageNet-R, and Imagenet-Sketch.

To further evaluate the OOD detection capability of our method, we initially pre-train our model using the StanfordCars [?] dataset and subsequently evaluate its performance on various datasets to get histograms for uncertainty estimates. As depicted in Fig. 2, it is evident that our model distinguishes unique uncertainty distributions among the nine datasets and the StanfordCars dataset. The findings suggest that our model discerns dissimilarities, classifying the nine datasets as OOD data from the StanfordCars dataset.

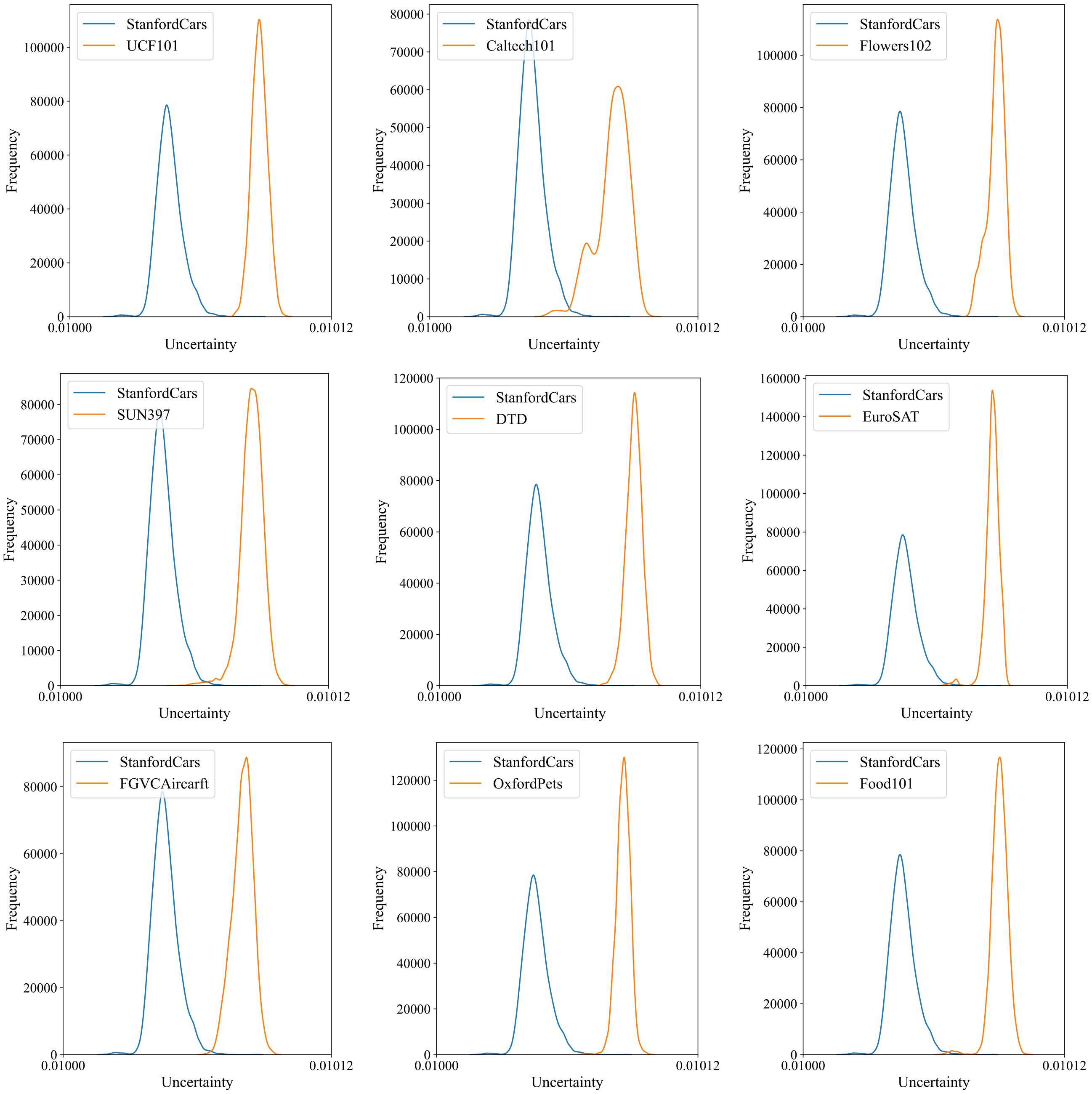

Figure 2. Histogram for uncertainty estimates. We evaluate our methods on StanfordCars and nine other datasets.