

A Implementation Details

We select images containing only one labelled class as images from Google Open Images dataset are dense and we also filter some classes that are too general such as "person, people" or that often include other classes such as "hat, shoes" as they appear in dense images. Eventually, we select 2368 classes and 167.287 total images (ablations with more and fewer images shown in the Appendix E) and train the adapters for 1 epoch with a learning rate of $5e-3$. We report the average accuracy across 3 different random seeds and perform 10 random augmentations for each training sample. For the unsupervised training we use the same images but train for 10 epochs with learning rate of $5e-5$ and momentum teacher of 0.9998. Other hyperparameters are default ones from the official DINO implementation [4]. The backbone used in both settings is ViT-B/16, which is compatible with the bottleneck adapter. We used the adapter with the bottleneck of size 64 which achieved the best performance on classification tasks in the original paper.

B Performance Across Fine-grained Datasets

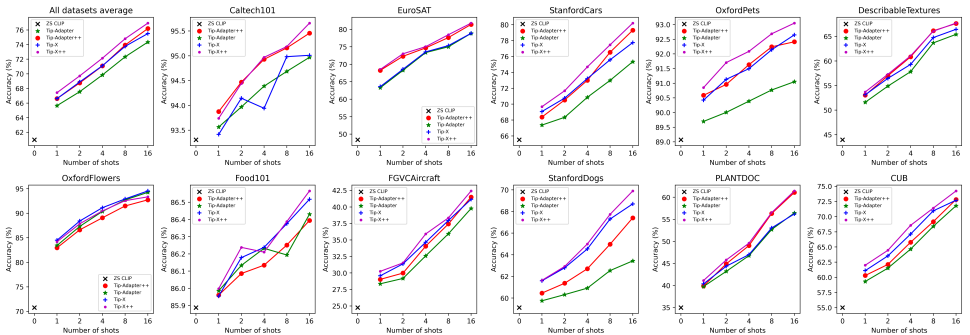


Figure 7: Performance comparison on 11 fine-grained datasets. Tip-Adapter++ consistently outperforms Tip-Adapter on 9 out of 11 fine-grained datasets with 1 dataset (Food101) achieving similar results and Tip-X++ consistently outperforms Tip-X on 10 out of 11 fine-grained datasets.

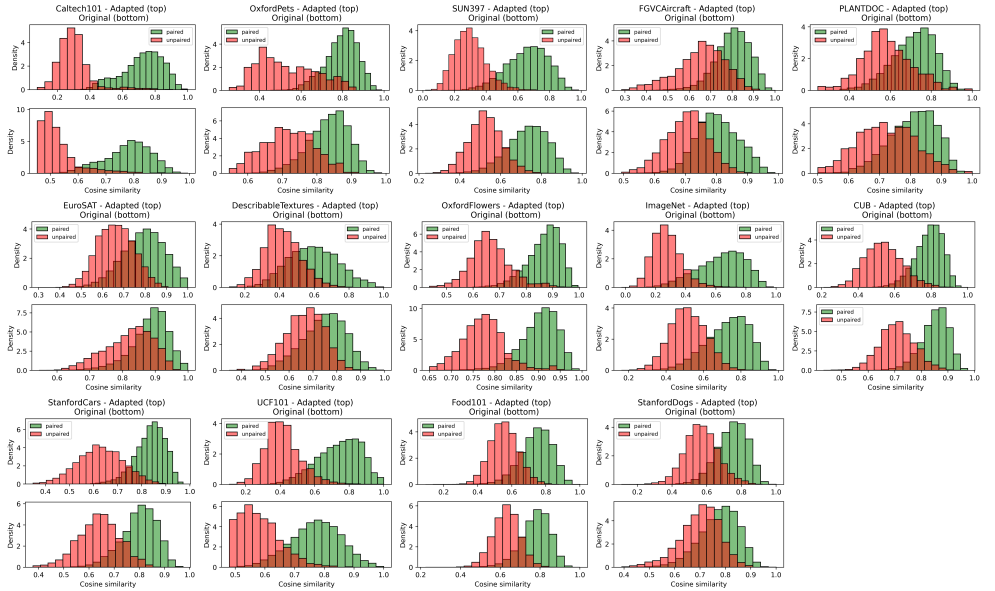
C Justification for few-shot CLIP learning

In [32] authors questioned the zero-shot generalization of multimodal models as classes and datasets used to test such capabilities could already be seen in the pretraining set. However, they did identify classes in the long tail of the distribution, where zero-shot performance was notably low, indicating that these classes were either rarely encountered or completely absent during pre-training. We argue that there is therefore still a case to improve the performance for such classes. We note that few-shot learning is valid especially where the difference between zero-shot and few-shot performance is significant, meaning that classes of those datasets are long tail. For instance, EuroSAT demonstrates low zero-shot performance, but

training-free few-shot learning leads to a substantial boost in accuracy of over 23%. Conversely, certain datasets such as Food101 already exhibit high zero-shot performance, with training-free few-shot learning resulting in only a marginal increase in accuracy of 0.5%. We improve upon existing training-free few-shot learning methods testing on a variety of datasets including both of these types.

D Intra-modal Overlap for All Datasets

In Fig. 3 we showed the intra-modal overlap (IMO) measured as an intersection area between cosine similarity distributions of paired and unpaired images for 4 datasets. In Fig. 8 we show the same for the remaining datasets, including the not fine-grained ones. The adaptation improves the IMO across 12 out of 14 datasets.



Dataset	Adapted Intersection Area (A)	Original Intersection Area (O)	$\Delta(O, A)$
Caltech101	0.127	0.108	-0.019
EuroSAT	0.482	0.6	0.119
StanfordCars	0.215	0.323	0.108
OxfordPets	0.358	0.386	0.028
DescribableTextures	0.47	0.633	0.163
UCF101	0.187	0.219	0.033
SUN397	0.146	0.26	0.114
OxfordFlowers	0.168	0.158	-0.01
Food101	0.282	0.295	0.013
FGVCaircraft	0.434	0.473	0.039
ImageNet	0.184	0.328	0.144
StanfordDogs	0.338	0.621	0.283
PLANTDOC	0.514	0.61	0.096
CUB	0.19	0.243	0.052

Figure 8: All datasets intra-modal overlap.

E Ablations

Other Datasets We conducted an ablation study across other standard datasets - Cifar100 and PascalVOC. Both of these datasets are of lower quality and less diverse compared to Google Open Images. Consequently, they were unable to decrease intra-modal overlap and improve accuracy to the same extent of Google Open Images when trained in a supervised way.

Training Dataset	Avg. IMO	Avg. $\Delta(TA++ , TA)$
Google Open Images	0.083	1.188
Cifar100	0.05	0.41
PascalVOC	0.01	0.12

Table 4: Aggregated performance and intra-modal overlap across all datasets and shots for Cifar100, PascalVoc and Google Open Images datasets trained in a supervised way.

Number of Samples Sensitivity In this analysis, we evaluate the impact of varying the number of samples from the Google Open Images dataset on performance and intra-modal overlap. We observed that an insufficient amount of data (80k samples) did not lead to significant performance improvement while increasing the dataset size to 200k samples did not yield much improvement compared to the 160k samples selected in our main experiments.

Number of samples	Avg. IMO	Avg. $\Delta(TA++ , TA)$
80k	0.059	0.5
160k	0.083	1.188
200k	0.076	0.82

Table 5: Aggregated performance and intra-modal overlap across all datasets and shots for different number of samples from Google Open Images trained in a supervised way.

F Granular Results & Performance with IMO Relation Across All Datasets

Intra-modal Overlap and Performance Relation When we include the not fine-grained datasets as observed in Fig. 9 the relation between intra-modal overlap reduction and performance improvement stays the same as for only the fine-grained ones reported in Fig. 4 in the main paper.

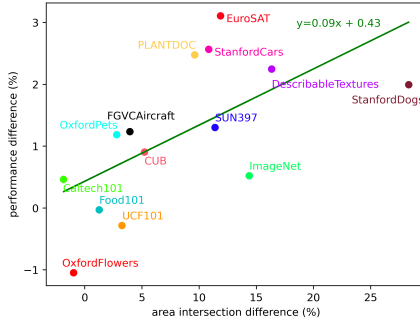


Figure 9: Relation between area intersection difference (intra-modal overlap reduction) between the original and adapted visual encoders vs average performance difference between Tip-Adapter++ and Tip-Adapter with supervised adaptation for all datasets.

Dataset	Shots	Zero-Shot	Tip-Adapter (TA)	Tip-Adapter++ (TA++)	Tip-X (TX)	Tip-X (TX++)	Δ (TA++, TA)	Δ (TX++, TX)	Δ (TA++,TX)
EuroSAT	1	48.383	63.288	68.259	63.597	68.527	4.971	4.93	4.663
EuroSAT	2	48.383	68.267	72.292	68.576	73.012	4.025	4.436	3.716
EuroSAT	4	48.383	73.354	74.683	73.547	75.041	1.329	1.494	1.136
EuroSAT	8	48.383	75.008	77.658	75.342	78.457	2.65	3.115	2.317
EuroSAT	16	48.383	78.852	81.407	78.864	81.782	2.556	2.918	2.543
StanfordCars	1	65.514	67.367	68.379	69.071	69.68	1.011	0.609	-0.692
StanfordCars	2	65.514	68.341	70.522	70.758	71.683	2.18	0.924	-0.236
StanfordCars	4	65.514	70.862	72.997	73.221	74.688	2.135	1.467	-0.224
StanfordCars	8	65.514	72.988	76.529	75.579	77.465	3.54	1.886	0.949
StanfordCars	16	65.514	75.347	79.306	77.752	80.201	3.959	2.45	1.555
PLANTDOC	1	34.994	39.78	39.888	40.384	41.138	0.108	0.755	-0.496
PLANTDOC	2	34.994	43.208	44.912	44.373	45.796	1.703	1.423	0.539
PLANTDOC	4	34.994	46.766	49.051	47.003	49.59	2.285	2.587	2.048
PLANTDOC	8	34.994	52.695	56.317	53.04	56.511	3.622	3.471	3.277
PLANTDOC	16	34.994	56.425	61.082	56.231	61.427	4.657	5.196	4.851
DescribableTextures	1	43.972	51.596	53.034	53.113	53.684	1.438	0.571	-0.079
DescribableTextures	2	43.972	54.886	56.994	56.462	57.289	2.108	0.827	0.532
DescribableTextures	4	43.972	57.821	60.835	59.299	61.032	3.014	1.734	1.537
DescribableTextures	8	43.972	63.672	66.135	64.756	66.056	2.463	1.3	1.379
DescribableTextures	16	43.972	65.406	67.612	66.43	67.691	2.206	1.261	1.182
StanfordDogs	1	59.117	59.749	60.461	61.596	61.636	0.712	0.04	-1.136
StanfordDogs	2	59.117	60.317	61.368	62.796	62.92	1.052	0.124	-1.428
StanfordDogs	4	59.117	60.917	62.708	64.539	64.999	1.791	0.46	-1.831
StanfordDogs	8	59.117	62.54	64.971	67.302	67.734	2.431	0.432	-2.331
StanfordDogs	16	59.117	63.436	67.414	68.706	69.902	3.979	1.196	-1.292
SUN397	1	62.579	65.529	66.713	66.584	67.058	1.184	0.474	0.128
SUN397	2	62.579	67.332	68.516	68.37	69.093	1.184	0.723	0.146
SUN397	4	62.579	68.791	70.35	70.025	70.929	1.559	0.904	0.325
SUN397	8	62.579	70.441	71.781	71.753	72.809	1.34	1.055	0.028
SUN397	16	62.579	71.635	72.874	72.955	73.776	1.239	0.821	-0.081
FGVC Aircraft	1	24.752	28.363	29.033	29.573	30.253	0.67	0.68	-0.54
FGVC Aircraft	2	24.752	29.173	29.983	31.383	31.523	0.81	0.14	-1.4
FGVC Aircraft	4	24.752	32.593	34.063	34.653	35.914	1.47	1.26	-0.59
FGVC Aircraft	8	24.752	35.934	37.424	37.954	38.344	1.49	0.39	-0.53
FGVC Aircraft	16	24.752	39.774	41.504	41.164	42.424	1.73	1.26	0.34
OxfordPets	1	89.071	89.697	90.588	90.424	90.851	0.89	0.427	0.164
OxfordPets	2	89.071	90.006	90.96	91.133	91.705	0.954	0.572	-0.173
OxfordPets	4	89.071	90.388	91.633	91.496	92.087	1.245	0.591	0.136
OxfordPets	8	89.071	90.77	92.241	92.141	92.686	1.472	0.545	0.1
OxfordPets	16	89.071	91.051	92.414	92.65	93.05	1.363	0.4	-0.236
CUB	1	55.009	59.318	60.301	61.103	61.995	0.983	0.892	-0.802
CUB	2	55.009	61.514	62.128	63.536	64.457	0.614	0.92	-1.408
CUB	4	55.009	64.652	65.781	67.127	68.57	1.129	1.443	-1.346
CUB	8	55.009	68.41	69.177	70.961	71.415	0.767	0.453	-1.785
CUB	16	55.009	71.798	72.823	72.711	74.238	1.025	1.527	0.112
ImageNet	1	68.804	69.28	69.536	69.389	69.568	0.256	0.179	0.147
ImageNet	2	68.804	69.477	69.805	69.509	69.812	0.328	0.303	0.297
ImageNet	4	68.804	69.791	70.359	69.864	70.359	0.569	0.495	0.495
ImageNet	8	68.804	70.249	70.949	70.459	71.012	0.699	0.553	0.489
ImageNet	16	68.804	70.753	71.505	70.973	71.587	0.753	0.613	0.532
Caltech101	1	93.306	93.563	93.874	93.414	93.739	0.311	0.325	0.46
Caltech101	2	93.306	93.969	94.469	94.145	94.442	0.5	0.297	0.325
Caltech101	4	93.306	94.388	94.929	93.942	94.97	0.541	1.028	0.987
Caltech101	8	93.306	94.686	95.159	94.983	95.186	0.473	0.203	0.176
Caltech101	16	93.306	94.97	95.456	95.01	95.659	0.487	0.649	0.446
Food101	1	85.888	85.986	85.96	85.955	85.998	-0.025	0.043	0.006
Food101	2	85.888	86.133	86.086	86.178	86.238	-0.047	0.059	-0.092
Food101	4	85.888	86.232	86.134	86.238	86.21	-0.098	-0.028	-0.103
Food101	8	85.888	86.194	86.251	86.375	86.387	0.057	0.012	-0.124
Food101	16	85.888	86.43	86.394	86.517	86.565	-0.036	0.048	-0.123
UCF101	1	67.46	71.716	72.024	72.553	72.667	0.308	0.115	-0.529
UCF101	2	67.46	73.777	73.857	75.17	75.24	0.079	0.07	-1.313
UCF101	4	67.46	74.007	73.795	75.399	75.17	-0.211	-0.229	-1.604
UCF101	8	67.46	77.284	76.509	78.298	78.377	-0.775	0.079	-1.789
UCF101	16	67.46	78.421	77.602	78.773	79.038	-0.819	0.264	-1.172
OxfordFlowers	1	70.767	83.435	82.961	84.504	84.193	-0.474	-0.311	-1.543
OxfordFlowers	2	70.767	87.319	86.615	88.145	87.86	-0.704	-0.555	-1.8
OxfordFlowers	4	70.767	90.378	89.078	91.135	90.472	-1.299	-0.663	-2.057
OxfordFlowers	8	70.767	92.719	91.487	92.922	92.57	-1.232	-0.352	-1.435
OxfordFlowers	16	70.767	94.262	92.732	94.546	93.341	-1.529	-1.204	-1.814
Average fine-grained	1	60.979	65.649	66.613	66.612	67.427	0.963	0.815	0.0
Average fine-grained	2	60.979	67.558	68.757	68.887	69.72	1.2	0.834	-0.13
Average fine-grained	4	60.979	69.85	71.081	71.109	72.143	1.231	1.034	-0.028
Average fine-grained	8	60.979	72.329	73.941	73.76	74.801	1.612	1.041	0.181
Average fine-grained	16	60.979	74.341	76.195	75.507	76.935	1.854	1.427	0.688
Average all	1	62.115	66.333	67.215	67.233	67.928	0.882	0.695	-0.018
Average all	2	62.115	68.123	69.179	69.343	70.076	1.056	0.733	-0.164
Average all	4	62.115	70.067	71.171	71.249	72.145	1.104	0.896	-0.078
Average all	8	62.115	72.399	73.756	73.705	74.644	1.357	0.939	0.052
Average all	16	62.115	74.183	75.723	75.235	76.477	1.541	1.243	0.489

Table 6: Average results by number of shots over 3 seeds.

G Unsupervised Training

Results In Fig. 10 and Table 7 we compare the performance of Tip-Adapter and Tip-Adapter++ (similar results for Tip-X vs Tip-X++ that we omit) observing that with unsupervised adaptation Tip-Adapter++ outperforms Tip-Adapter on 7 out of 14 datasets. These results are worse than the supervised counterpart, however, we believe that it is interesting to correct the intra-modal overlap through adaptation training adapters in an unsupervised way. As future work we will try to do it with a bigger and more diverse dataset.

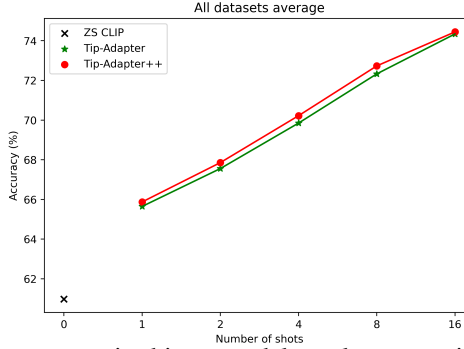


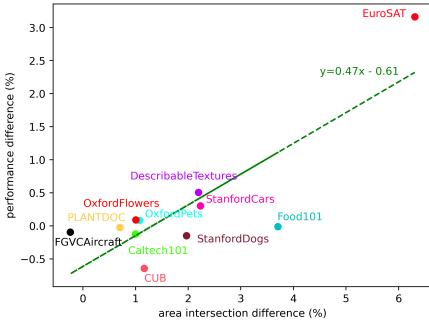
Figure 10: Performance unsupervised intra-modal overlap correction. Figure shows the average performance of Tip-Adapter and Tip-Adapter++ across different shots for fine-grained datasets.

Dataset	Zero-Shot	Tip-Adapter (TA)	Tip-Adapter++ (TA++)	Δ (TA++, TA)
EuroSAT	48.383	71.754	74.915	3.161
DescribableTextures	43.972	58.676	59.18	0.504
SUN397	62.579	68.783	69.115	0.332
StanfordCars	65.514	70.981	71.283	0.302
UCF101	67.46	75.041	75.286	0.245
OxfordFlowers	70.767	89.622	89.712	0.089
OxfordPets	89.071	90.382	90.464	0.082
Food101	85.888	86.195	86.182	-0.013
ImageNet	68.801	69.911	69.897	-0.014
PLANTDOC	34.994	47.775	47.749	-0.026
FGVCAircraft	24.752	33.167	33.071	-0.096
Caltech101	93.306	94.315	94.191	-0.124
StanfordDogs	59.117	61.392	61.242	-0.15
CUB	55.009	65.138	64.494	-0.644
Average fine-grained	60.979	69.945	70.226	0.281
Average all	62.115	70.224	70.484	0.261

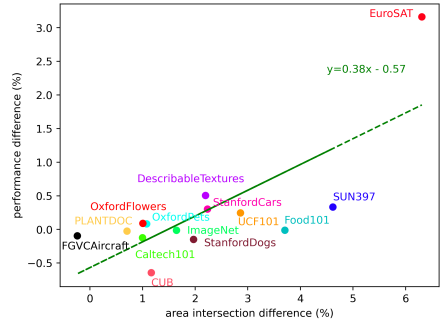
Table 7: Performance unsupervised intra-modal overlap correction. Table shows the comparison between average performance of Tip-Adapter and Tip-Adapter++ across different shots for all the datasets.

Performance and the Relation with Intra-modal Overlap of Unsupervised Adaptation

In Fig. 11 we observe a positive relation between the difference in intersection area and the average performance difference, mirroring the pattern seen in the supervised counterpart.



(a) Fine-grained datasets



(b) All datasets

Figure 11: Relation between area intersection difference (intra-modal overlap reduction) between the original and adapted visual encoders vs average performance difference between Tip-Adapter++ and Tip-Adapter with unsupervised adaptation. Fig. (a) shows this relation for fine-grained datasets while Fig. (b) for all the datasets.

H LoRA Adapter

We perform an ablation study implementing the LoRA [17] adapter rather than the bottleneck adapter [5]. LoRA adapter is applied to the self-attention at each layer of the visual encoder. The results presented in Table 8 indicate a significant degradation in performance compared to using the bottleneck adapter. We attribute the inferior performance of LoRA to the fact that the bottleneck adapter keeps the CLIP visual encoder weights frozen, maintaining extensive knowledge about different classes acquired during CLIP pretraining and only slightly adjusts the features with the effect of reducing the intra-modal overlap, while the application of LoRA adapters breaks that knowledge leading to inferior performance.

Dataset	Zero-Shot	Tip-Adapter (TA)	Tip-Adapter++ (TA++)	Δ (TA++, TA)
OxfordPets	89.071	90.382	89.97	-0.412
Food101	85.888	86.195	85.984	-0.211
Caltech101	93.306	94.315	93.915	-0.4
StanfordDogs	59.117	61.392	61.156	-0.235
ImageNet	68.804	69.911	69.374	-0.537
SUN397	62.579	68.783	66.516	-2.267
UCF101	67.46	75.041	71.478	-3.563
EuroSAT	48.383	71.754	69.165	-2.588
StanfordCars	65.514	70.981	67.798	-3.184
PLANTDOC	34.994	47.775	44.489	-3.286
CUB	55.009	65.138	58.444	-6.695
DescribableTextures	43.972	58.676	51.052	-7.624
FGVCAircraft	24.752	33.167	26.163	-7.005
OxfordFlowers	70.767	89.622	79.878	-9.744

Table 8: Performance comparison between average performance of Tip-Adapter and Tip-Adapter++ for each dataset across different shots using LoRA Adapter.

I APE Training-free Method

Method Description APE [35] is a training-free method where most discriminative features from the last vision and text CLIP layers are selected eliminating less discriminative feature channels based on a prior refinement module. They employ two criteria for this selection: inter-class similarity and variance. Inter-class similarity criterion focuses on extracting feature channels that minimize the inter-class similarity. On the other hand the inter-class variance criterion eliminates feature channels that exhibit minimal variation between categories as these channels have little impact on classification. These two criteria are then combined to extract the most discriminative features. With such refined features, indicated by ' symbol, the authors compute APE classification logits for a test image. These are given by the sum of CLIP zero-shot logits and Tip-Adapter affinity matrix but weighted by the uncertainty of CLIP logits based on few-shot training examples instead of training labels. To compute these weights, they calculate the Kullback-Leibler (KL) divergence between the zero-shot CLIP classification probabilities derived from training data features F_{train} as defined in Eq. 3 and classifier weight matrix W and the true labels L_{train} as defined in Eq. 4 in the main paper:

$$\text{APEweights} = \exp(\gamma D_{KL}(F'_{train} W'^T | L_{train})), \in \mathbb{R}^{1 \times NK} \quad (14)$$

Where ' indicates that the features were refined with the refinement module and γ is a smoothing factor.

These weights reflect the divergence between the true and zero-shot CLIP predicted labels. For classes where there is more uncertainty in zero-shot CLIP prediction, i.e., where the KL divergence is high, we need to rely more on the cache model and vice versa. Final prediction logits for APE are given by:

$$\text{APElogits} = \text{CLIPlogits} + \alpha A'(\text{diag}(\text{APEweights}) L_{train}) \quad (15)$$

Where A' is the affinity matrix as defined in Eq. 5 but with refined features, diag is the diagonalization operator and α is a weighting constant.

Replacing the affinity matrix A' with the intra-modal overlap corrected one, Y' , as in Eq. 10 we obtain APE++:

$$\text{APElogits++} = \text{CLIPlogits} + \alpha Y'(\text{diag}(\text{APEweights}) L_{train}) \quad (16)$$

Intra-modal Overlap After Features Pruning As discussed above authors of APE proposed a method to select more discriminative features by eliminating certain feature channels based on inter-class similarity criterion. This has the effect of shifting the unpaired distribution of cosine similarities to the left but, as we illustrate in Fig. 12 and in Tab. 9 it also moves the distribution of the paired images to the left thus either changing only slightly or making worse the intra-modal overlap in most cases.

Results In Tab. 10 we include the results with APE model for completeness. We can observe that in 10 out of 14 datasets APE++ outperforms APE although the margin of improvement is often smaller compared to the other training-free methods. This observed trend

Dataset	APE Intersection Area (APE)	Original Intersection Area (O)	Δ (O, APE)
Caltech101	0.36	0.108	-0.252
EuroSAT	0.61	0.6	-0.01
StanfordCars	0.484	0.323	-0.161
OxfordPets	0.464	0.386	-0.078
DescribableTextures	0.566	0.633	0.067
UCF101	0.311	0.219	-0.091
SUN397	0.232	0.26	0.027
OxfordFlowers	0.2	0.158	-0.042
Food101	0.26	0.295	0.035
FGVCAircraft	0.4731	0.473	-0.0001
ImageNet	0.292	0.328	0.036
StanfordDogs	0.571	0.621	0.05
PLANTDOC	0.644	0.61	-0.034
CUB	0.246	0.243	-0.003

Table 9: Intra-modal overlap after adaptive features refinement.

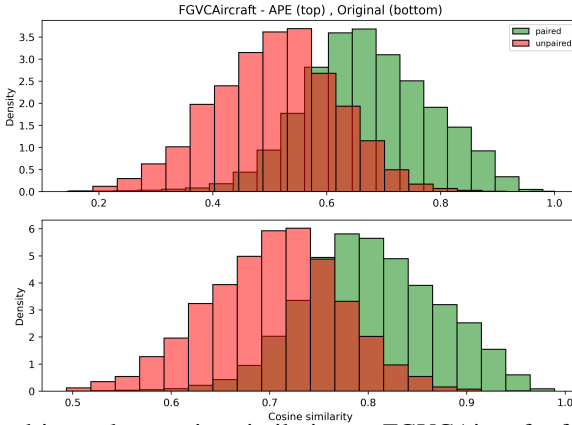


Figure 12: Intra and inter-class cosine similarity on FGVCaircraft after APE refinement. Both intra-class and inter-class similarity decreases almost not affecting the intra-modal overlap.

is attributed to the impact of features pruning. Indeed, as shown in Tab. 11 without feature pruning APE++ exhibits a more substantial performance improvement over APE, similar to the enhancements observed with Tip-Adapter and Tip-X. This is interesting as it indicates that by pruning features, while the intra-modal overlap is not reduced (implying the paired and unpaired samples are close), the features do lie on different sides of the decision boundary of the classifier. This would be a reduced sub-space of features that fits the features based on the decision boundary of the classifier. However, such an approach would not necessarily be robust or have the variance properties. We will investigate opportunities for residual subspace learning that are robust and with variance that explore the decision boundary of classifiers in the future.

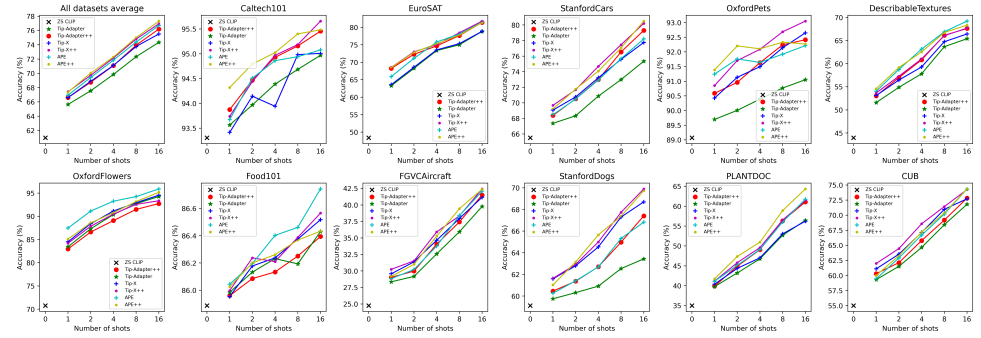


Figure 13: Performance comparison on fine-grained datasets including APE method

Dataset	Zero-Shot	Tip-Adapter (TA)	Tip-Adapter++ (TA++)	Tip-X (TX)	Tip-X++ (TX++)	APE	APE++	$\Delta(TA++, TA)$	$\Delta(TX++, TX)$	$\Delta(TA++, TX)$	$\Delta(APE++, APE)$
EuroSAT	48.383	71.754	74.86	71.985	75.364	74.486	75.165	3.106	3.379	2.875	0.679
StanfordCars	65.514	70.981	73.546	73.276	74.744	73.156	74.524	2.565	1.467	0.27	1.368
PLANTDOC	34.994	47.775	50.25	48.206	50.893	50.63	52.652	2.475	2.687	2.044	2.022
DescribableTextures	43.972	58.676	60.922	60.012	61.151	62.411	62.281	2.246	1.139	0.91	-0.13
StanfordDogs	59.117	61.392	63.385	64.988	65.438	63.304	65.39	1.993	0.45	-1.603	2.086
SUN397	62.579	68.746	70.047	69.938	70.733	70.447	71.016	1.301	0.795	0.109	0.569
FGVCAircraft	24.752	33.167	34.401	34.945	35.692	34.659	35.454	1.234	0.746	-0.544	0.795
OxfordPets	89.071	90.382	91.567	91.569	92.076	91.756	92.06	1.185	0.507	-0.002	0.304
CUB	55.009	65.138	66.042	67.088	68.135	66.709	67.033	0.904	1.047	-1.046	0.324
ImageNet	68.804	69.91	70.431	70.039	70.468	70.29	70.827	0.521	0.429	0.392	0.537
Caltech101	93.306	94.315	94.778	94.299	94.799	94.613	95.005	0.462	0.5	0.479	0.392
Food101	85.888	86.195	86.165	86.253	86.28	86.369	86.257	-0.03	0.027	-0.088	-0.112
UCF101	67.46	75.041	74.757	76.038	76.098	77.129	75.912	-0.284	0.06	-1.281	-1.217
OxfordFlowers	70.767	89.622	88.575	90.305	89.687	92.394	90.562	-1.048	-0.617	-1.73	-1.832
Average fine-grained	60.979	69.945	71.317	71.175	72.205	71.863	72.398	1.372	1.03	0.142	0.535
Average all	62.115	70.221	71.409	71.353	72.254	72.025	72.438	1.188	0.901	0.056	0.413

Table 10: Average performance datasets across all shots including APE.

Dataset	Zero-Shot	Tip-Adapter (TA)	Tip-Adapter++ (TA++)	Tip-X (TX)	Tip-X++ (TX++)	APE	APE++	$\Delta(TA++, TA)$	$\Delta(TX++, TX)$	$\Delta(TA++, TX)$	$\Delta(APE++, APE)$
EuroSAT	48.383	71.754	74.86	71.985	75.364	72.61	75.677	3.106	3.379	2.875	3.067
StanfordCars	65.514	70.981	73.546	73.276	74.744	71.596	73.935	2.565	1.467	0.27	2.339
PLANTDOC	34.994	47.775	50.25	48.206	50.893	48.491	51.397	2.475	2.687	2.044	2.906
DescribableTextures	43.972	58.676	60.922	60.012	61.151	59.421	61.446	2.246	1.139	0.91	2.025
StanfordDogs	59.117	61.392	63.385	64.988	65.438	61.815	64.314	1.993	0.45	-1.603	2.499
SUN397	62.579	68.746	70.047	69.938	70.733	69.52	70.855	1.301	0.795	0.109	1.335
FGVCAircraft	24.752	33.167	34.401	34.945	35.692	33.595	34.595	1.234	0.746	-0.544	1.0
OxfordPets	89.071	90.382	91.567	91.569	92.076	91.102	91.694	1.185	0.507	-0.002	0.592
CUB	55.009	65.138	66.042	67.088	68.135	65.466	66.46	0.904	1.047	-1.046	0.994
ImageNet	68.804	69.91	70.431	70.039	70.468	70.219	70.827	0.521	0.429	0.392	0.608
Caltech101	93.306	94.315	94.778	94.299	94.799	94.723	95.064	0.462	0.5	0.479	0.341
Food101	85.888	86.195	86.165	86.253	86.28	86.39	86.335	-0.03	0.027	-0.088	-0.055
UCF101	67.46	75.041	74.757	76.038	76.098	75.994	75.545	-0.284	0.06	-1.281	-0.449
OxfordFlowers	70.767	89.622	88.575	90.305	89.687	90.613	89.081	-1.048	-0.617	-1.73	-1.532
Average fine-grained	60.979	69.945	71.317	71.175	72.205	70.529	71.818	1.372	1.03	0.142	1.289
Average all	62.115	70.221	71.409	71.353	72.254	70.825	71.945	1.188	0.901	0.056	1.12

Table 11: Average performance datasets across all shots including APE without features pruning.

Dataset	Shots	Zero-Shot	Tip-Adapter (TA)	Tip-Adapter++ (TA++)	Tip-X (TX)	Tip-X (TX++)	APE	APE++	Δ (TA++, TA)	Δ (TX++, TX)	Δ (TA++, TX)	Δ (APE++, APE)
EuroSAT	1	48.383	63.288	68.259	63.597	68.527	65.901	68.465	4.971	4.93	4.663	2.564
EuroSAT	2	48.383	68.267	72.292	68.576	73.012	71.14	72.877	4.025	4.436	3.716	1.737
EuroSAT	4	48.383	73.354	74.683	73.547	75.041	75.802	75.292	1.329	1.494	1.136	-0.51
EuroSAT	8	48.383	75.008	77.658	75.342	78.457	78.095	77.802	2.65	3.115	2.317	-0.293
EuroSAT	16	48.383	78.852	81.407	78.864	81.782	81.494	81.387	2.556	2.918	2.543	-0.107
StanfordCars	1	65.514	67.367	68.379	69.071	69.68	68.478	69.22	1.011	0.609	-0.692	0.742
StanfordCars	2	65.514	68.341	70.522	70.758	71.683	70.489	71.77	2.18	0.924	-0.236	1.281
StanfordCars	4	65.514	70.862	72.997	73.221	74.688	72.935	74.07	2.135	1.467	-0.224	1.135
StanfordCars	8	65.514	72.988	76.529	75.579	77.465	75.671	77.063	3.54	1.886	0.994	1.392
StanfordCars	16	65.514	75.347	79.306	77.752	80.201	78.208	80.496	3.959	2.45	1.555	2.288
PLANTDOC	1	34.994	39.78	39.888	40.384	41.138	41.117	41.721	0.108	0.755	-0.496	0.604
PLANTDOC	2	34.994	43.208	44.912	44.373	45.796	45.127	47.348	1.703	1.423	0.539	2.221
PLANTDOC	4	34.994	46.766	49.051	47.003	49.59	49.116	50.949	2.285	2.587	2.048	1.833
PLANTDOC	8	34.994	52.695	56.317	53.04	56.511	56.037	58.905	3.622	3.471	3.277	2.868
PLANTDOC	16	34.994	56.425	61.082	56.231	61.427	61.751	64.338	4.657	5.196	4.851	2.587
DescribableTextures	1	43.972	51.596	53.034	53.113	53.684	54.039	54.59	1.438	0.571	-0.079	0.551
DescribableTextures	2	43.972	54.886	56.994	56.462	57.289	58.747	59.18	2.108	0.827	0.532	0.433
DescribableTextures	4	43.972	57.821	60.835	59.299	61.032	63.16	62.549	3.014	1.734	1.537	-0.611
DescribableTextures	8	43.972	63.672	66.135	64.756	66.056	66.903	66.745	2.463	1.3	1.379	-0.158
DescribableTextures	16	43.972	65.406	67.612	66.43	67.691	69.208	68.341	2.206	1.261	1.182	-0.867
StanfordDogs	1	59.117	59.749	60.461	61.596	61.636	60.261	61.028	0.712	0.04	-1.136	0.767
StanfordDogs	2	59.117	60.317	61.368	62.796	62.92	61.408	63.148	1.052	0.124	-1.428	1.74
StanfordDogs	4	59.117	60.917	62.708	64.539	64.999	62.696	65.659	1.791	0.46	-1.831	2.963
StanfordDogs	8	59.117	62.54	64.971	67.302	67.734	65.327	67.734	2.431	0.432	-2.331	2.047
StanfordDogs	16	59.117	63.436	67.414	68.706	69.902	66.827	69.742	3.979	1.196	-1.292	2.915
SUN397	1	62.579	65.529	66.713	66.584	67.058	66.687	67.453	1.184	0.474	0.128	0.766
SUN397	2	62.579	67.332	68.516	68.37	69.093	68.608	69.602	1.184	0.723	0.146	0.994
SUN397	4	62.579	68.791	70.35	70.025	70.929	70.94	71.8	1.559	0.904	0.325	0.886
SUN397	8	62.579	70.441	71.781	71.753	72.809	72.571	72.895	1.34	1.055	0.028	0.324
SUN397	16	62.579	71.635	72.874	72.955	73.776	73.429	73.332	1.239	0.821	-0.081	-0.097
FGVCAircraft	1	24.752	28.363	29.033	29.573	30.253	28.833	29.163	0.67	0.68	-0.54	0.33
FGVCAircraft	2	24.752	29.173	29.983	31.383	31.523	30.223	31.013	0.81	0.14	-1.4	0.79
FGVCAircraft	4	24.752	32.593	34.063	34.653	35.914	33.773	35.274	1.47	1.26	-0.59	1.501
FGVCAircraft	8	24.752	35.934	37.424	37.954	38.344	38.384	39.434	1.49	0.39	-0.53	1.05
FGVCAircraft	16	24.752	39.774	41.504	41.164	42.424	42.084	42.384	1.73	1.26	0.34	0.3
OxfordPets	1	89.071	89.697	90.588	90.424	90.851	91.242	91.387	0.89	0.427	0.164	0.145
OxfordPets	2	89.071	90.006	90.96	91.133	91.705	91.76	92.205	0.954	0.572	-0.173	0.445
OxfordPets	4	89.071	90.388	91.633	91.496	92.087	91.642	92.105	1.245	0.591	0.136	0.463
OxfordPets	8	89.071	90.77	92.241	92.141	92.686	91.923	92.332	1.472	0.545	0.1	0.409
OxfordPets	16	89.071	91.051	92.414	92.65	93.05	92.214	92.269	1.363	0.4	-0.236	0.055
CUB	1	55.009	59.318	60.301	61.103	61.995	59.437	59.993	0.983	0.892	-0.802	0.556
CUB	2	55.009	61.514	62.128	63.536	64.457	62.92	63.312	0.614	0.92	-1.408	0.392
CUB	4	55.009	64.652	65.781	67.127	68.57	66.681	67.172	1.129	1.443	-1.346	0.491
CUB	8	55.009	68.41	69.177	70.961	71.415	70.178	70.342	0.767	0.453	-1.785	0.164
CUB	16	55.009	71.798	72.823	72.711	74.238	74.33	74.345	1.025	1.527	0.112	0.015
ImageNet	1	68.804	69.28	69.536	69.389	69.568	69.493	69.822	0.256	0.179	0.147	0.329
ImageNet	2	68.804	69.477	69.805	69.509	69.812	69.804	70.289	0.328	0.303	0.297	0.485
ImageNet	4	68.804	69.791	70.359	69.864	70.359	70.247	70.845	0.569	0.495	0.495	0.598
ImageNet	8	68.804	70.249	70.949	70.459	71.012	70.81	71.367	0.699	0.553	0.489	0.557
ImageNet	16	68.804	70.753	71.505	70.973	71.587	71.094	71.811	0.753	0.613	0.532	0.717
Caltech101	1	93.306	93.563	93.874	93.414	93.739	93.671	94.32	0.311	0.325	0.46	0.649
Caltech101	2	93.306	93.969	94.469	94.145	94.442	94.51	94.794	0.5	0.297	0.325	0.284
Caltech101	4	93.306	94.388	94.929	93.942	94.97	94.861	95.024	0.541	1.028	0.987	0.163
Caltech101	8	93.306	94.686	95.159	94.983	95.186	94.943	95.402	0.473	0.203	0.176	0.459
Caltech101	16	93.306	94.97	95.456	95.01	95.659	95.078	95.483	0.487	0.649	0.446	0.405
Food101	1	85.888	85.986	85.96	85.955	85.998	86.044	86.025	-0.025	0.043	0.006	-0.019
Food101	2	85.888	86.133	86.086	86.178	86.238	86.196	86.2	-0.047	0.059	-0.092	0.004
Food101	4	85.888	86.232	86.134	86.238	86.21	86.403	86.261	-0.098	-0.028	-0.103	-0.142
Food101	8	85.888	86.194	86.251	86.375	86.387	86.461	86.369	0.057	0.012	-0.124	-0.092
Food101	16	85.888	86.43	86.394	86.517	86.565	86.743	86.432	-0.036	0.048	-0.123	-0.311
UCF101	1	67.46	71.716	72.024	72.553	72.667	73.187	73.055	0.308	0.115	-0.529	-0.132
UCF101	2	67.46	73.777	73.857	75.17	75.24	76.835	75.443	0.079	0.07	-1.313	-1.392
UCF101	4	67.46	74.007	73.795	75.399	75.17	76.853	75.178	-0.211	-0.229	-1.604	-1.675
UCF101	8	67.46	77.284	76.509	78.298	78.377	79.135	77.487	-0.775	0.079	-1.789	-1.648
UCF101	16	67.46	78.421	77.602	78.773	79.038	79.637	78.395	-0.819	0.264	-1.172	-1.242
OxfordFlowers	1	70.767	83.435	82.961	84.504	84.193	87.468	85.099	-0.474	-0.311	-1.543	-2.369
OxfordFlowers	2	70.767	87.319	86.615	88.415	87.86	91.122	88.578	-0.704	-0.555	-1.8	-2.544
OxfordFlowers	4	70.767	90.378	89.078	91.135	90.472	93.247	90.797	-1.299	-0.663	-2.057	-2.45
OxfordFlowers	8	70.767	92.719	91.487	92.922	92.57	94.248	93.166	-1.232	-0.352	-1.435	-1.082
OxfordFlowers	16	70.767	94.262	92.732	94.546	93.341	95.886	95.168	-1.529	-1.204	-1.814	-0.718
Average fine-grained	1	60.979	65.649	66.613	66.612	67.427	66.954	67.365	0.963	0.815	0.0	0.411
Average fine-grained	2	60.979	67.558	68.757	68.887	69.72	69.422	70.039	1.2	0.834	-0.13	0.617
Average fine-grained	4	60.979	69.85	71.081	71.109	72.143	71.847	72.287	1.231	1.034	-0.028	0.44
Average fine-grained	8	60.979	72.329	73.941	73.76	74.801	74.379	74.994	1.612	1.041	0.181	0.615
Average fine-grained	16	60.979	74.341	76.195	75.507	76.935	76.711	77.308	1.854	1.427	0.688	0.597
Average all	1	62.115	66.333	67.215	67.233	67.928	67.561	67.953	0.882	0.695	-0.018	0.392
Average all	2	62.115	68.123	69.179	69.343	70.076	69.921	70.411	1.056	0.733	-0.164	0.49
Average all	4	62.115	70.067	71.171	71.249	72.145	72.025	72.355	1.104	0.896	-0.078	0.33
Average all	8	62.115	72.399	73.756	73.705	74.644	74.335	74.763	1.357	0.939	0.052	0.428
Average all	16	62.115	74.183	75.723	75.235	76.477	76.284	76.709	1.541	1.243	0.489	0.425

Table 12: Average results by number of shots over 3 seeds including APE.