Supplementary Material for Fine-Grained Few-Shot Classification with Part Matching

We include additional results and information omitted due to the page length limitation.

Hyperparameter Sweep The SMPL part matching Transformer method relies on three main hyperparameters: number of layers, L, number of attention heads H, and temperature coefficient τ . Figure 1 shows the 5-shot classification accuracy on the CUB and Reptilia datasets as each parameter varies. In general, SMPL is not sensitive to hyperparameter settings. For the experiments in the paper, L = 4, H = 8, and $\tau = 20$.



Figure 1. 20-way, 5-shot accuracy of SMPL tuning the number of layers (left), number of attention heads (middle), and the temperature coefficient (right).

Semantic Prompts To identify parts in the images, we use GroundedSAM [34] with domain-specific part prompts. Table 1 shows the part names used for each of the four datasets in the experiments.

Additional Quantitative Results The table shows few-shot classification results for SMPL compared to the same baseline methods for 1-shot and 5-shot classification in 5-way and 20-way settings with DINOv2 ViT-B as the backbone.

Detect	Kauwords					
Dataset	Keywolus					
Hotel-8k	wall . bed . mattress . lamp . pillows . bed sheets . artwork . painting . blankets . nightstand . desk .					
	desk chair . floor . rug . closet . couch . coffee table . ottoman . hangers . dresser . mirror . television .					
	curtains . carpet . wall . wallpaper . ceiling . windows . door . sink . shower . bathtub . toilet . towel .					
	air conditioner . heater . headboard . fan . mini fridge . coffee maker . kettle . iron . ironing board .					
	luggage rack . lights . smoke detector . microwave . fridge . stove . cooktop . counter . light . cupboard .					
	sink . table . dishwasher . glasses . toaster					
CUB	bird. tail. head. wings. beak. talons. belly. eye					
Cars	car. truck. side mirror. wheel. hood. grill. headlight. windshield. seat. body. trunk. door. taillight.					
	bumper . grill . door handle . logo					
Reptilia	lizard . turtle . shell . head . neck . torso . tail . scales . claws . leg . feet . eyes . nostrils . ears . jaw .					
	spine . skin					
VGG-Flower	flower . stem . roots . petal . sepal . stamen . anther . filament . pistil . ovule . receptacle . nectar					

Table 1. GroundedSAM prompts used for each dataset to generate segmentation maps.

	Hotels-8k				Reptilia			
	5-way		20-way		5-way		20-way	
Method	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
R2D2	.259	.507	.120	.385	.829	.936	.621	.820
MetaOptNet	.259	.502	.118	.351	.829	.936	.623	.829
Linear Probing	.249	.436	.104	.276	.782	.897	.581	.799
FT	.268	.555	.123	.423	.820	.915	.623	.821
VPT	.258	.464	.099	.203	.731	.906	.612	.815
SSF	.257	.531	.116	.417	.778	.890	.598	.813
FORT-FT	.264	.536	.122	.416	.823	.916	.621	.819
FORT-SSF	.252	.445	.107	.326	.689	.846	.598	.780
ProtoNet	.447	.608	.233	.367	.658	.832	.405	.645
RelationNet	.487	.649	.258	.421	.796	.883	.510	.692
MATA	.616	.868	.410	.750	.832	.910	.595	.764
MetaBaseline	.754	.896	.589	.792	.812	.908	.586	.772
SMPL	.781	.922	.604	.819	.868	.938	.687	.835

Part Maching This figure shows examples of Hotels-8k few-shot classification in which the most similar matching parts between the query and support are highlighted in the same color to demonstrate SMPL matching across multiple support images.



Additional Qualitative Results This figure includes additional challenging 1-shot classification cases where SMPL accurately classified the query. Each row presents a query image, the correct match from the support class, and a false positive selected by multiple baseline methods.

