Charting the Right Manifold: Manifold Mixup for Few-shot Learning
Supplementary Material

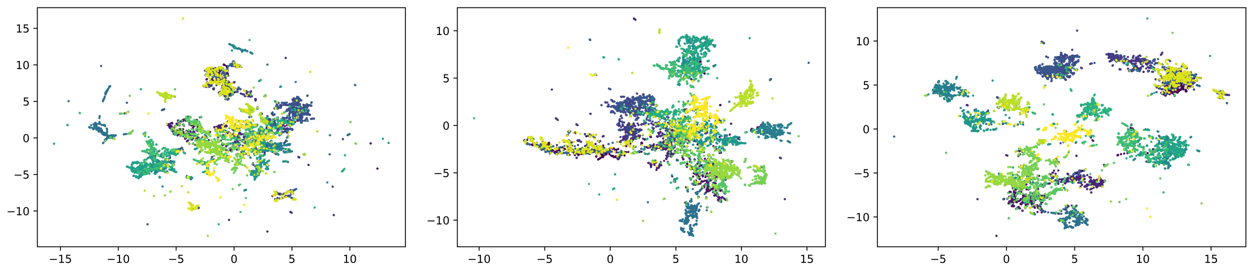

Figure 1: UMAP (2-dim) [2] plot for feature vectors of examples from novel classes of CIFAR-FS using Baseline++ [1], Rotation, *S2M2*$_R$ (top to bottom).

Method	5-way		10-way		15-way		20-way	
	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
Baseline++	67.5	80.08	53.39	68.89	44.73	60.59	38.22	54.68
Manifold Mixup	69.45	83.31	57.06	75.53	49.04	68.60	43.54	62.80
Rotation	70.5	84.03	57.37	73.60	48.49	66.25	42.28	61.10
S2M2$_R$	**74.45**	**87.50**	**62.28**	**78.47**	**53.49**	**71.88**	**47.59**	**66.37**

Table 1: Mean few-shot accuracy on CIFAR-FS as N increases in N-way K-shot classification.

S1. Ablation Studies

In this section, we perform additional experiments to study the efficacy of our approach *S2M2*$_R$.

S1.1. Effect of varying N in N-way classification on CIFAR-FS

We vary N in N-way K-shot evaluation criteria from 5 to 10, 15 and 20 for CIFAR-FS dataset. The corresponding results are reported in table 1. We observe that our approach *S2M2*$_R$ outperforms other techniques by a significant margin. The improvement becomes more pronounced for $N > 5$. Figure 1 shows the 2-dimensional UMAP [2] plot of feature vectors of novel classes obtained from different methods. We obtain similar results for CIFAR-FS as that in the case of *mini*-ImageNet. We show that our approach has more segregated clusters with less variance. This supports our hypothesis that using both self-supervision and Manifold Mixup regularization helps in learning feature representations with well separated margin between novel classes.

S1.2. Visualizing important regions in images responsible for classification

We visualize the relevant pixels responsible for classifying a particular image to the correct class. We define the relevance of the pixels as the top-1 percentile of the pixels sorted by the magnitude of the gradient with respect to the correct class of the image. For this experiment, we use models trained using the Baseline++ [1] and *S2M2*$_R$ methods to visualize the relevant pixels. In figure 2, we show the relevant pixels of the image *highlighted in white* for visualization. Qualitatively speaking, we observe that relevant pixels for model trained using *S2M2*$_R$ tends to focus more on the object belonging to the specified class and not in the background.

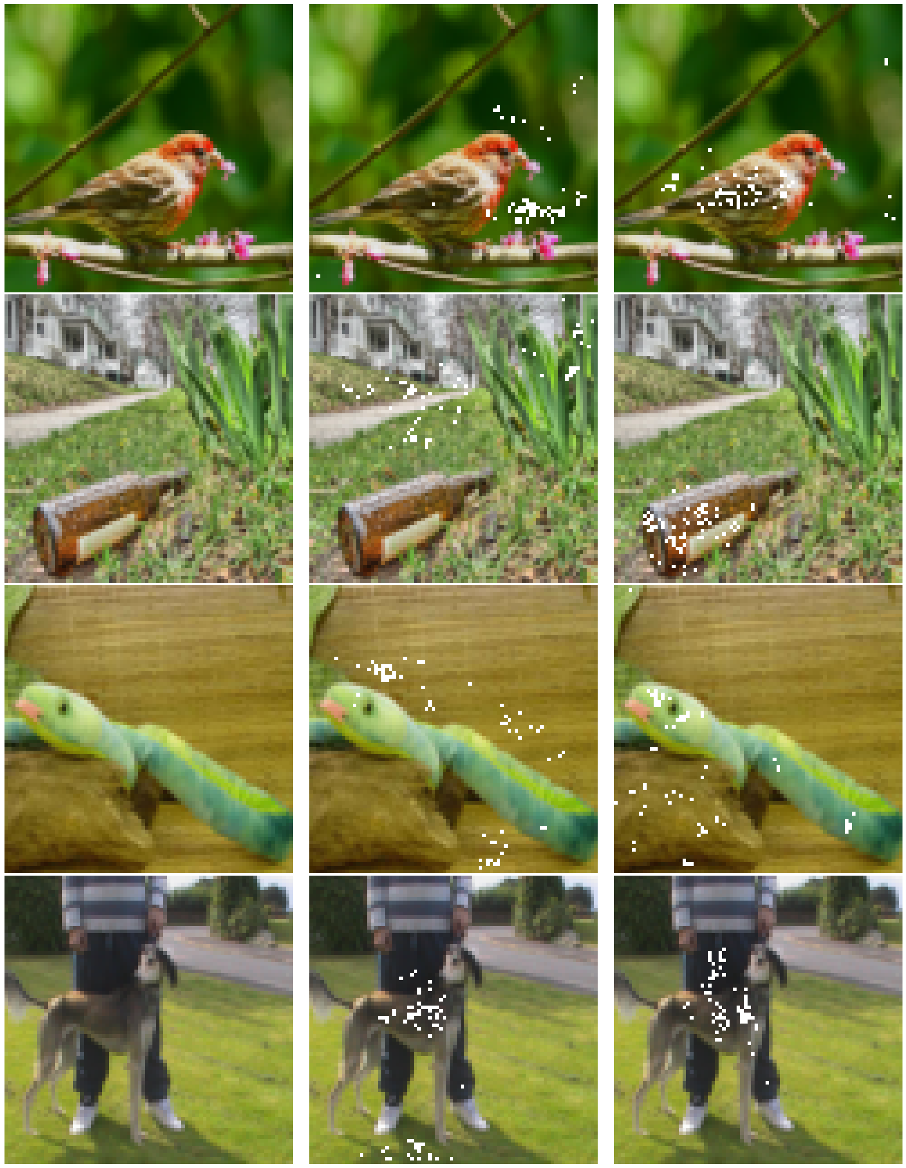

Figure 2: Each row visualizes the relevant pixels for classification with respect to a trained model for a image sampled from the base class of *mini*-ImageNet (house finch, beer bottle, green mamba, Saluki). Images in each row are arranged in the order with labels as original image, relevant pixels by Baseline++ model and relevant pixels by $S2M2_R$ model *(from left to right)* respectively. The relevant pixels is defined as the Top-1 percentile of pixels responsible for classification *(pixels marked in white color)*.

References

[1] W.-Y. Chen, Y.-C. Liu, Z. Kira, Y.-C. Wang, and J.-B. Huang. A closer look at few-shot classification. In *International Conference on Learning Representations*, 2019.

[2] L. McInnes, J. Healy, and J. Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.