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

Puneet Mangla* [1,2] **Abhishek Sinha** [2] **Mayank Singh** [2] **Nupur Kumari** [2] Vineeth N Balasubramanian [1] Balaji Krishnamurthy [2]

[1]IIT Hyderabad, India [2]Media and Data Science Research (MDSR), Adobe Noida, India

Github

cm cm

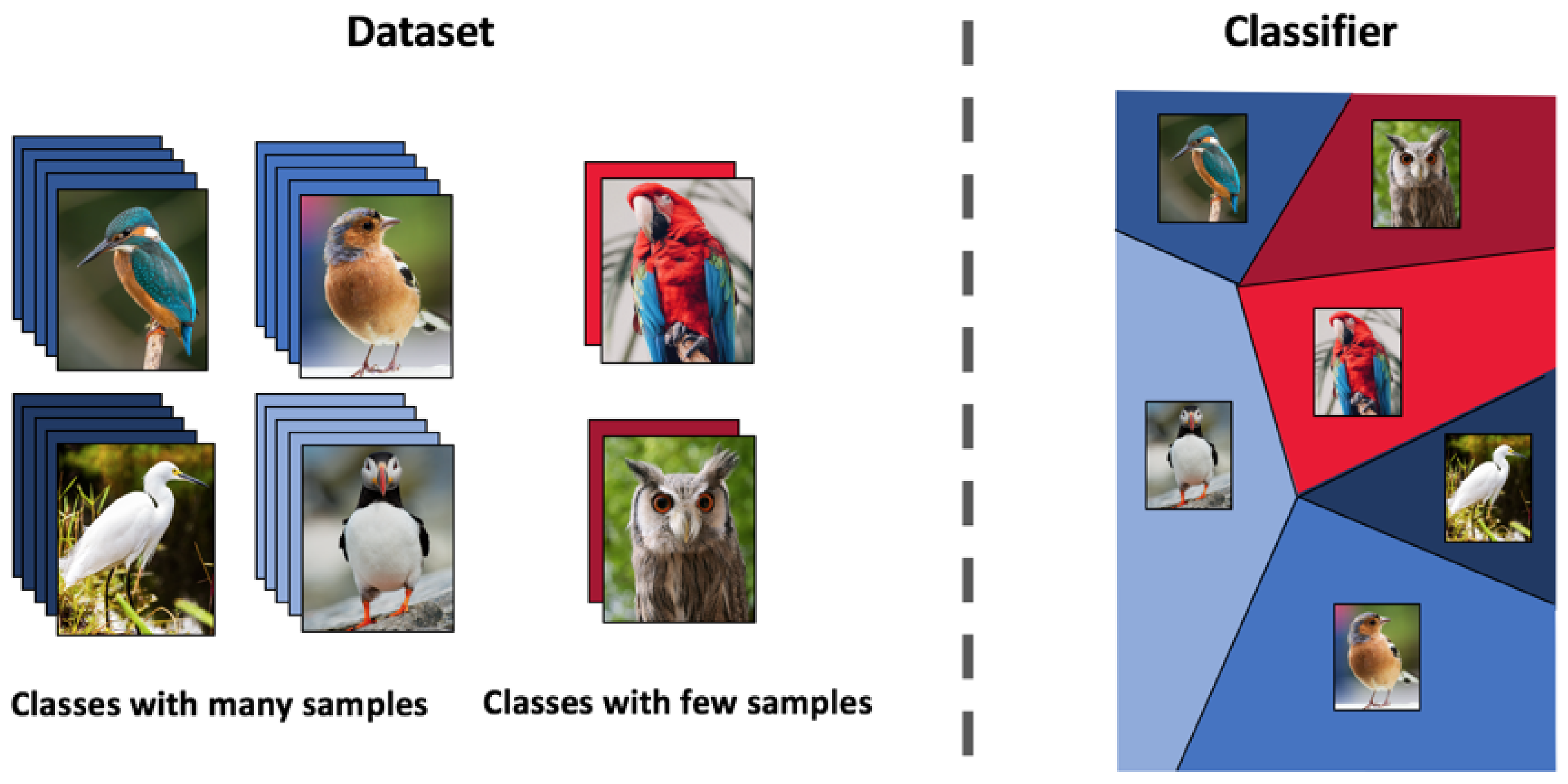
Puneet Mangla* [1,2] Abhishek Sinha [2] Mayank Singh [2] Nupur Kumari [2] Vineeth N Balasubramanian [1] Balaji Krishnamurthy [2]
[1]IIT Hyderabad, India [2]Media and Data Science Research (MDSR), Adobe Noida, India

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

Github

Problem Definition and Contribution

Few-Shot Learning: To learn model parameters capable of adapting to unseen classes with only a few labeled examples.



Key Contributions:

- Manifold Mixup over the feature manifold enriched via the self-supervision task improves few-shot classification.
- More pronounced effect on increasing N in the N-way K-shot evaluation
- Improves cross-domain few-shot task evaluation.

Notations and Preliminaries

Training:

- *First phase:* training a feature extractor f_θ over base class data $\mathcal{D}_b = \{(\mathbf{x}_i, y_i), i = 1, \cdots, m_b\}$.
- *Second phase:* Freezing feature extractor module and learning a new classifier for novel class data $\mathcal{D}_n = \{(\mathbf{x}_i, y_i), i = 1, \cdots, m_n\}$.

Manifold-Mixup: leverages linear interpolations in hidden layers of neural network to help the trained model generalize better.

$$L_{mm} = \mathbb{E}_{(x,y)\in\mathcal{D}_b}\Big[L\Big(Mix_\lambda(f^l_\theta(\mathbf{x}), f^l_\theta(\mathbf{x}')), Mix_\lambda(y, y')\Big)\Big]$$

where $Mix_\lambda(a, b) = \lambda \cdot a + (1-\lambda) \cdot b$ and L is cross-entropy loss.

Self Supervision: Towards the Right Manifold:

- Rotation: the input image is rotated, and the auxiliary task of the model is to predict the rotation.

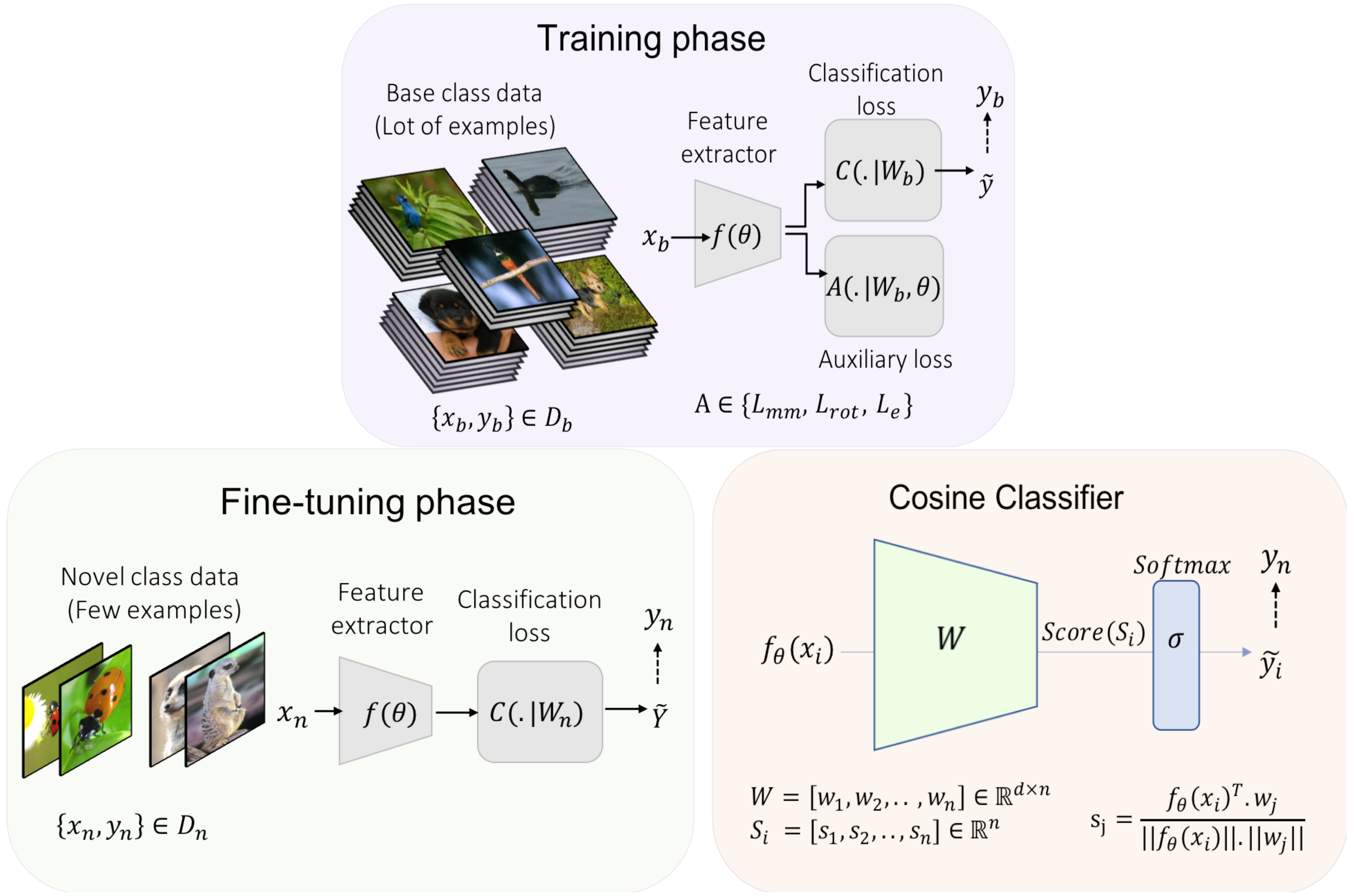
$$L_{rot} = \frac{1}{|C_R|} * \sum_{\mathbf{x}\in\mathcal{D}_b} \sum_{r\in C_R} L(c_{W_r}(f_\theta(\mathbf{x}^r)), r)$$

  where $\mathbf{x}^r$ is the image x rotated by r degrees and $r \in C_R = \{0°, 90°, 180°, 270°\}$
- Exemplar: feature representation of same image with random augmentations are promoted to be similar and different from other images through triplet loss L_e.

Cosine Classifier: Given an image x and $W_b = \{w_1, w_2 ... w_{N_b}\}$, final scores and training loss is given by:

$$s = SoftMax(\lambda \cdot \frac{f_\theta(x) \cdot w_j}{||f_\theta(x)|| \cdot ||w_j||})$$
$$L_{class} = \mathop{\mathbb{E}}_{x,y\in D_b} CrossEntropy(s, y)$$

Our Approach: Self-Supervised Manifold-Mixup (S2M2)



Step 1: Self-supervised training: Train feature extractor using self-supervision as an auxiliary loss along with classification loss.

Step 2: Fine-tuning with Manifold Mixup: Fine-tune the above model with Manifold-Mixup for a few more epochs.

After obtaining the backbone, a cosine classifier is learned over it to adapt to few-shot tasks.

Experiments

Comparison with prior/current state of the art methods on *mini*-ImageNet, *tiered*-ImageNet, and CUB dataset.

Method	*mini*-ImageNet		*tiered*-ImageNet		CUB	
	1-Shot	5-Shot	1-Shot	5-Shot	1-Shot	5-Shot
MAML	54.69 ± 0.89	66.62 ± 0.83	51.67 ± 1.81	70.30 ± 0.08	71.29 ± 0.95	80.33 ± 0.70
ProtoNet	54.16 ± 0.82	73.68±0.65	53.31 ± 0.89	72.69 ± 0.74	71.88±0.91	87.42 ± 0.48
RelationNet	52.19 ± 0.83	70.20 ± 0.66	54.48 ± 0.93	71.32 ± 0.78	68.65 ± 0.91	81.12 ± 0.63
LEO [3]	61.76 ± 0.08	77.59 ± 0.12	66.33 ± 0.05	81.44 ± 0.09	68.22 ± 0.22*	78.27 ± 0.16*
DCO [2]	62.64 ± 0.61	78.63 ± 0.46	65.99 ± 0.72	81.56 ± 0.53	-	-
Baseline++	57.53 ± 0.10	72.99 ± 0.43	60.98 ± 0.21	75.93 ± 0.17	70.4 ± 0.81	82.92 ± 0.78
Manifold Mixup	57.16 ± 0.17	75.89 ± 0.13	68.19 ± 0.23	84.61 ± 0.16	73.47 ± 0.89	85.42 ± 0.53
Rotation	63.9 ± 0.18	81.03 ± 0.11	73.04 ± 0.22	87.89 ± 0.14	77.61 ± 0.86	89.32 ± 0.46
S2M2$_R$	**64.93 ± 0.18**	**83.18 ± 0.11**	**73.71 ± 0.22**	**88.59 ± 0.14**	**80.68 ± 0.81**	**90.85 ± 0.44**

Ablation Studies

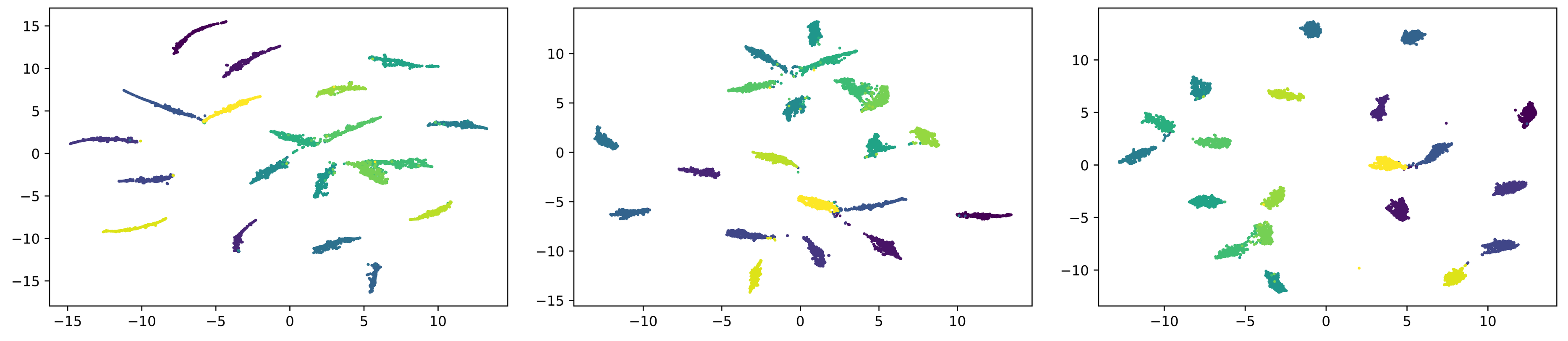
Effect of varying N in N-way Classification We test our proposed methodology in complex few-shot settings. We vary N in N-way K-shot evaluation criteria from 5 to 10, 15 and 20.

2*Method	10-way		15-way		20-way	
	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
Baseline++	40.43	56.89	31.96	48.2	26.92	42.8
LEO	45.26	64.36	36.74	56.26	31.42	50.48
DCO	44.83	64.49	36.88	57.04	31.5	51.25
Manifold Mixup	42.46	62.48	34.32	54.9	29.24	48.74
Rotation	47.77	67.2	38.4	59.59	33.21	54.16
S2M2$_R$	**50.4**	**70.93**	**41.65**	**63.32**	**36.5**	**58.36**

Cross-domain few-shot learning To further highlight the significance of selecting the right manifold for feature space, we evaluate the few-shot classification performance over cross-domain dataset : *mini*-ImageNet $\Longrightarrow$ CUB (coarse-grained to fine-grained distribution).

Method	*mini*-Imagenet $\Longrightarrow$ CUB	
	1-Shot	5-Shot
DCO	44.79 ± 0.75	64.98 ± 0.68
Baseline++	40.44 ± 0.75	56.64 ± 0.72
Manifold Mixup	46.21 ± 0.77	66.03 ± 0.71
Rotation	**48.42 ± 0.84**	68.40 ± 0.75
S2M2$_R$	48.24 ± 0.84	**70.44 ± 0.75**

Visualization of feature representations Our approach has more segregated clusters with less variance. Thus, using both self supervision and Manifold Mixup regularization helps in learning feature representations with well separated margin between novel classes.



UMAP (2-dim) plot for feature vectors of examples from novel classes of *mini*-Imagenet using Baseline++, Rotation, *S2M2$_R$* (left to right).

References

[1] Wei-Yu Chen, Yen-Cheng Liu, Zsolt Kira, Yu-Chiang Wang, and Jia-Bin Huang. A closer look at few-shot classification. In *International Conference on Learning Representations*, 2019.

[2] Kwonjoon Lee, Subhransu Maji, Avinash Ravichandran, and Stefano Soatto. Meta-learning with differentiable convex optimization. *CoRR*, abs/1904.03758, 2019.

[3] Andrei A. Rusu, Dushyant Rao, Jakub Sygnowski, Oriol Vinyals, Razvan Pascanu, Simon Osindero, and Raia Hadsell. Meta-learning with latent embedding optimization. In *International Conference on Learning Representations*, 2019.

[4] Vikas Verma, Alex Lamb, Christopher Beckham, Amir Najafi, Ioannis Mitliagkas, David Lopez-Paz, and Yoshua Bengio. Manifold mixup: Better representations by interpolating hidden states. In *International Conference on Machine Learning*, pages 6438–6447, 2019.

[5] Xiaohua Zhai, Avital Oliver, Alexander Kolesnikov, and Lucas Beyer. S ^{4} l: Self-supervised semi-supervised learning. *arXiv:1905.03670*, 2019.