Supplementary Material

Embedding Propagation: Smoother Manifold for Few-Shot Classification

Pau Rodríguez[1], Issam Laradji[1,2], Alexandre Drouin[1], and Alexandre Lacoste[1]

[1] Element AI, Montreal, Canada
[2] University of British Columbia, Canada
{pau.rodriguez,issam.laradji,adrouin,allac}@elementai.com

In the following sections, we provide results in the 10-shot scenario as well as more challenging settings, such as 10-way, 15-way, and 20-way classification. We also illustrate the effect of ablating different parts of the propagator matrix. Finally, we report the CO_2 emissions to produce this research.

10-shot test accuracy.

We report the results of EPNet and EPNet$_{SSL}$ in Table 1. Our methods improve the accuracy over *TADAM* [6] and *Discriminative* [1] by 5% accuracy. Since not many methods have been evaluated on this benchmark, it is challenging to illustrate the impact of embedding propagation. Further, we see that EPNet$_{SSL}$ does not improve much over EPNet, suggesting that embedding propagation has larger impact with fewer labeled data.

Table 1: 10-shot test accuracy on *mini*Imagenet, *tiered*Imagenet, and CUB.

	mini	**tiered**	**CUB**
RESNET-12			
Discriminative [1]	78.50	-	-
TADAM [6]	80.80	-	-
EPNet (ours)	**85.39**	**88.39**	**92.68**
EPNet$_{SSL}$ (ours)	**87.34**	**89.24**	**92.88**
WRN-28-10			
EPNet (ours)	**87.03**	**89.46**	**93.99**
EPNet$_{SSL}$ (ours)	**89.02**	**89.56**	**94.11**

Higher-way results

For higher way setups, we compare against previous state-of-the-art methods on *mini*Imagenet (Table 2). For a fair comparison, we use a WRN-28-10 [8] as our feature extractor. Our results show that EPNet attains higher test accuracies than previous state-of-the-art in all settings. For instance, in the 1-shot 20-way scenario, EPNet improves results from 36.5% to 38.6% accuracy. This suggests that embedding propagation generalizes effectively to higher number of classes.

Table 2: *mini*Imagenet 1-shot and 5-shot test accuracies for the 10-way, 15-way and 20-way scenarios. We report the accuracy with 95% confidence intervals over 600 episodes.

	10-way		**15-way**		**20-way**	
	1-shot	**5-shot**	**1-shot**	**5-shot**	**1-shot**	**5-shot**
Baseline++ [2]	40.43	56.89	31.96	48.20	26.92	42.80
LEO [7]	45.26	64.36	36.74	56.26	31.42	50.48
DCO [4]	44.83	64.49	36.88	57.04	31.50	51.25
Manifold mixup [5]	50.40	70.93	41.65	63.32	36.50	58.36
EPNet (ours)	**53.70** ± 0.59	**72.17** ± 0.44	**44.55** ± 0.28	**64.44** ± 0.34	**38.55** ± 0.19	**59.01** ± 0.27

Table 3: Propagator ablation with resnet-12 on 5-shot *mini*Imagenet. Pretraining accuracy with (1) the full propagator matrix, (2) removing the diagonal of the propagator matrix, (3) removing the off-diagonal of the propagator matrix. As shown, our method leverages information from the neighborhood to attain optimal performance

VER	OFF-DIAG	DIAG	**ACC**
1	✓	✓	75.95 ±0.56
2	✓	-	74.66 ±0.38
3	-	✓	73.80 ±0.29

Additional ablation experiments

While in Table 6 of the main text we ablate all the components of the proposed model, here (in Table 3) we focus only on EP and whether it leverages information from neighboring embeddings (off-diagonal) or it just rescales them (diagonal). We show that neighbor information is important for the performance of EP. We train three versions of EPNet, (i) one with the full propagator matrix, (ii) one with only the off-diagonal of the matrix, and (iii) one with the diagonal matrix only. Hence, the second version only relies on information from neighboring embeddings to make predictions. The third version is equivalent to multiplying the original embeddings by a scalar. As seen in Table 3, the best performance is obtained with the first version, confirming that information from neighboring nodes is used.

CO2 Emission Related to Experiments

Experiments were conducted using a private infrastructure (located in Quebec, Canada) with a carbon emission factor of 0.02 kg/kWh. A cumulative of 24480 hours of computation was performed on hardware of type Tesla V100 (with a TDP of 250 W). Total emissions are estimated to be 146.88 kgCO_2eq, and 1000 kgCO_2eq (685%) were offsetted through Gold Standard. Estimations were obtained using the MachineLearning Impact calculator [3].

References

1. Bauer, M., Rojas-Carulla, M., Światkowski, J.B., Schölkopf, B., Turner, R.E.: Discriminative k-shot learning using probabilistic models. arXiv preprint arXiv:1706.00326 (2017)
2. Chen, W.Y., Liu, Y.C., Kira, Z., Wang, Y.C.F., Huang, J.B.: A closer look at few-shot classification. In: ICLR (2019)
3. Lacoste, A., Luccioni, A., Schmidt, V., Dandres, T.: Quantifying the carbon emissions of machine learning. arXiv preprint arXiv:1910.09700 (2019)
4. Lee, K., Maji, S., Ravichandran, A., Soatto, S.: Meta-learning with differentiable convex optimization. In: CVPR. pp. 10657–10665 (2019)
5. Mangla, P., Singh, M., Sinha, A., Kumari, N., Balasubramanian, V.N., Krishnamurthy, B.: Charting the right manifold: Manifold mixup for few-shot learning. arXiv preprint arXiv:1907.12087 (2019)
6. Oreshkin, B., López, P.R., Lacoste, A.: Tadam: Task dependent adaptive metric for improved few-shot learning. In: NeurIPS. pp. 721–731 (2018)
7. Rusu, A.A., Rao, D., Sygnowski, J., Vinyals, O., Pascanu, R., Osindero, S., Hadsell, R.: Meta-learning with latent embedding optimization. In: ICLR (2018)
8. Zagoruyko, S., Komodakis, N.: Wide residual networks. In: BMVC. pp. 87.1–87.12. BMVA Press (September 2016). https://doi.org/10.5244/C.30.87