Supplementary Materials for EASE: Unsupervised Discriminant Subspace Learning for Transductive Few-Shot Learning

Hao Zhu[†,§], Piotr Koniusz[*,§,†]
[†]The Australian National University [§]Data61/CSIRO
allenhaozhu@gmail.com, firstname.lastname@anu.edu.au

Below are additional details regarding our model and experiments.

A. Datasets

mini-ImageNet. This dataset [10, 14] is widely used in few-shot classification. It contains 100 randomly chosen classes from ImageNet [12] . There are 64 training (base) classes, 16 validation (novel) classes, and 20 test (novel) classes among the 100 classes. There are 600 images in each class. We adopt the split provided in [10].

tiered-ImageNet. An ImageNet with a hierarchical structure was used to create the tiered-ImageNet. Categories of classes are divided into 34 categories, each of which contains 351, 97, and 160 classes for training, validation, and test, respectively. Please note the training and test classes are semantically disjoint. We follow the common split in [1] and 84 by 84px resolution.

CIFAR-FS. This dataset has 100 classes, each with 600 examples in CIFAR-100 [5], on which this dataset is based on. We use the 64 training, 16 validation, and 20 test classes provided by [2].

CUB. There are 200 classes, each representing a bird species, in this fine-grained dataset. Following the setting in [2] , we divide our classes into three groups: 100 training, 50 validation, and 50 testing classes.

B. Feature Extraction and Pre-processing

ResNet-12 ResNet-12A is the pre-trained backbone network used in [16]. For all of our transductive and semi-supervised experiments using this network, we adopt exactly the same pre-processing as [16], which includes normalizing feature vectors by their ℓ_2 norms.

[*]The corresponding author. Code: `https://github.com/allenhaozhu/EASE`

WRN-28-10 WRN-28-10 is the pre-trained network used in [8] and [3]. To provide fair comparisons with PT+MAP [3], we adopt exactly the same pre-processing as [3]. In the all experiments, we apply the power transform and normalize feature vectors by their ℓ_2 norms.

C. Hyperparameters

In Eq. (8), we have a parameter for the rank-K. In fact, K is the same as K-way. In Eq. (6), there is a parameter α that controls the trade-off between the similarity matrix and the dissimilarity matrix, and we set it as 40 for all experiments. In Algorithm 1, we set $\lambda = 10$, $\epsilon = 0.001$ and $\alpha = 0.2$ (please note here the α is the parameter to control updating the new estimate center).

D. Other Experiments

We evaluate our approach on two few-shot classification benchmarks, mini-ImageNet [14] and tiered-ImageNet [11], all used in transductive and semi-supervised FSL works [4,7,9,11,13]. On these benchmarks, we use the standard evaluation protocols [6]. The results of the transductive and semi-supervised FSL evaluation, together with comparisons to previous methods, are summarized in Tables 1, and are detailed and discussed in the following sections. The performance numbers are given as accuracy %, and the 0.95 confidence intervals are reported. The tests are performed on 10,000 random 5-way episodes, with 1 or 5 shots (number of support examples per class), and with 15 queries per episode (unless otherwise specified). For each dataset, the standard train, validation and test splits are used. Training subset was used to pre-train the DenseNet backbone (from scratch) with a regular multi-class classifier on all training classes, same as in [15]. The validation data was used to select the best model along the training epochs and to choose the hyper-parameters. Episodes generated from the test data (with test categories unseen during training and validation) were used for meta-testing to obtain the final performance.

Comparison with other clustering Methods. To demonstrate the effectiveness of conStraIned wAsserstein MEan Shift clustEring (SIAMESE), we integrated EASE with clustering methods other than SIAMESE. Specifically, we compared SIAMESE with k-means and Mean Shift Propagation (MSP) mentioned in [6]. As can be seen from Table 1, on both transductive FSL benchmarks, the top performing variant of EASE consistently outperforms all the previous (transductive and inductive) SOTA and SIAMESE significantly outperforms other clustering methods.

Unbalanced (long-tail) test classes distribution in unlabeled data. In previous transductive FSL works, balanced test tasks (in terms of the number of queries for each class) were used. While this is fine for the experimental evaluation, in practical applications, it is unrealistic that the bulk of queries sent for an off-line evaluation will be class-balanced. In fact, class-skew (lack of balance) is a very likely case. To test the effect of query set skew, we have evaluated the proposed method under varying levels of class-skew controlled through the so-called 'unbalanced factor' parameter R: in each test episode, for each class 15 + uni([0, R]) query samples were randomly chosen (here uni refers to the uniform distribution). Figure 1 shows the effect of varying R from 10 to 50, while at the extreme setting (50) above factor 4 skew is possible between the classes in terms of the number of associated queries. Nevertheless, as can be seen from the figure, the effect of lack of balance on the performance of EASE is minimal, leading to at most 2% performance loss at $R = 50$. Interestingly, when the imbalance factor is 10, EASE appears performing better than in the balanced setting (mini-ImageNet). This is likely because although the queries are unbalanced, the larger number of queries still provides an advantage to our method.

Out-of-distribution noise (distraction classes) in unlabeled data. Unlabeled data may become contaminated with samples 'unrelated' to the few-shot task T target classes distribution in many applications. This is most likely to happen in a semi-supervised FSL context. As in transductive FSL, the unlabeled samples are the queries which are typically assumed to solely belong to the target classes distribution. This form of the noise is examined in the semi-supervised FSL literature using additional random samples from random 'distracting' classes introduced into the unlabeled set. We set the number of distracting classes between 0 and 7, and show the comparison between our method and other STOA methods of semi-supervised few-shot learning. Fig. 2 shows that in this setting the performance of EASE drops very quickly as distracting classes increasing because distracting classes ruins the hypothesis of block diagonal prior. At the same time, we can see that other methods are also very sensitive to this setting and also suffer the similar performance drops.

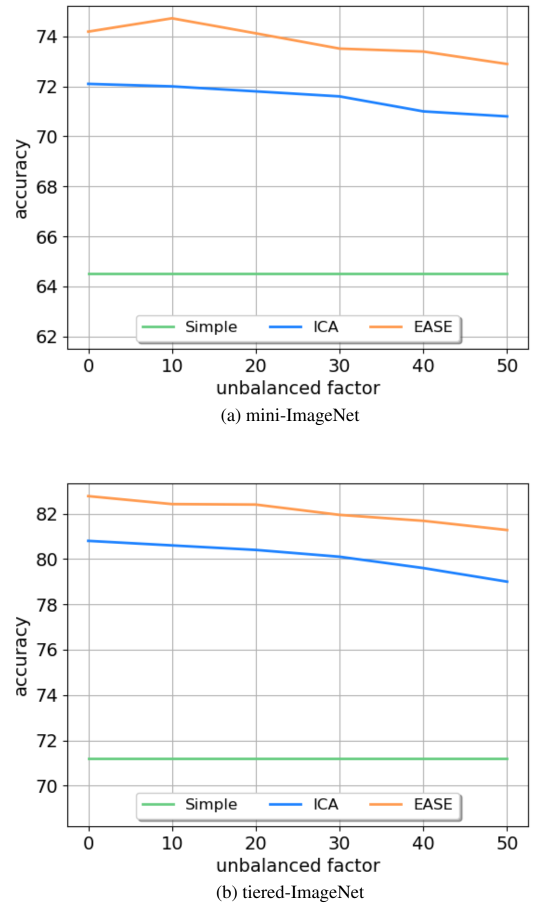

Figure 1. Unbalanced setting performance on (a) mini-ImageNet and (b) tiered-ImageNet.

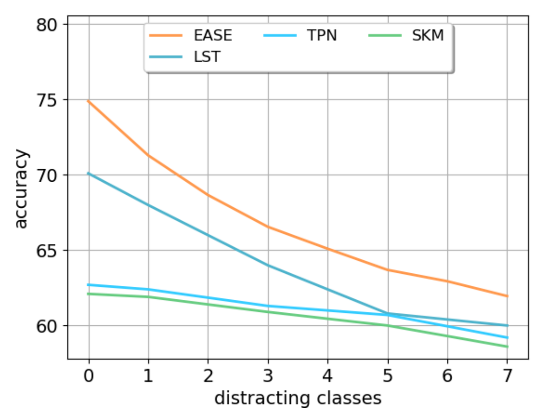

Figure 2. Performance w.r.t. the unlabeled data noise.

References

[1] Chaofan Chen, Xiaoshan Yang, Changsheng Xu, Xuhui Huang, and Zhe Ma. Eckpn: Explicit class knowledge prop-

Table 1. Comparison of test accuracy against state-of-the-art methods for 1-shot and 5-shot classification given EASE (ours)+*clustering*, where SIAMESE is our clustering variant. In fact, all variants of EASE are our variants but k-means and MSP were used before in FSL.

		mini-ImageNet		tiered-ImageNet	
Methods	Network	1-shot	5-shot	1-shot	5-shot
SimpleShot(Baseline) [15]	DenseNet	65.77 ± 0.19	82.23 ± 0.13	71.20 ± 0.22	86.33 ± 0.15
LaplacianShot [17]	DenseNet	75.57 ± 0.19	84.72 ± 0.13	80.30 ± 0.20	87.93 ± 0.15
TAFSSL(PCA) [6]	DenseNet	70.53 ± 0.25	80.71 ± 0.16	80.07 ± 0.25	86.42 ± 0.17
TAFSSL(ICA) [6]	DenseNet	72.10 ± 0.25	81.85 ± 0.16	80.82 ± 0.25	86.97 ± 0.17
EASE	DenseNet	74.30 ± 0.26	82.08 ± 0.17	82.67 ± 0.25	87.60 ± 0.17
EASE+k-means	DenseNet	76.41 ± 0.27	84.09 ± 0.15	83.57 ± 0.25	87.94 ± 0.15
EASE+MSP	DenseNet	77.54 ± 0.26	85.35 ± 0.14	84.63 ± 0.25	88.25 ± 0.15
EASE+SIAMESE (ours)	DenseNet	79.42 ± 0.27	86.76 ± 0.14	86.17 ± 0.25	90.54 ± 0.15

agation network for transductive few-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6596–6605, 2021. 1

[2] Wei-Yu Chen, Yen-Cheng Liu, Zsolt Kira, Yu-Chiang Frank Wang, and Jia-Bin Huang. A closer look at few-shot classification. *arXiv preprint arXiv:1904.04232*, 2019. 1

[3] Yuqing Hu, Vincent Gripon, and Stéphane Pateux. Leveraging the feature distribution in transfer-based few-shot learning. *arXiv preprint arXiv:2006.03806*, 2020. 1

[4] Jongmin Kim, Taesup Kim, Sungwoong Kim, and Chang D Yoo. Edge-labeling graph neural network for few-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11–20, 2019. 1

[5] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009. 1

[6] Moshe Lichtenstein, Prasanna Sattigeri, Rogerio Feris, Raja Giryes, and Leonid Karlinsky. Tafssl: Task-adaptive feature sub-space learning for few-shot classification. In *European Conference on Computer Vision*, pages 522–539. Springer, 2020. 1, 2, 3

[7] Yanbin Liu, Juho Lee, Minseop Park, Saehoon Kim, Eunho Yang, Sung Ju Hwang, and Yi Yang. Learning to propagate labels: Transductive propagation network for few-shot learning. *arXiv preprint arXiv:1805.10002*, 2018. 1

[8] Puneet Mangla, Nupur Kumari, Abhishek Sinha, Mayank Singh, Balaji Krishnamurthy, and Vineeth N Balasubramanian. Charting the right manifold: Manifold mixup for few-shot learning. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2218–2227, 2020. 1

[9] Limeng Qiao, Yemin Shi, Jia Li, Yaowei Wang, Tiejun Huang, and Yonghong Tian. Transductive episodic-wise adaptive metric for few-shot learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3603–3612, 2019. 1

[10] Sachin Ravi and Hugo Larochelle. Optimization as a model for few-shot learning. 2017. 1

[11] Mengye Ren, Eleni Triantafillou, Sachin Ravi, Jake Snell, Kevin Swersky, Joshua B Tenenbaum, Hugo Larochelle, and Richard S Zemel. Meta-learning for semi-supervised few-shot classification. *arXiv preprint arXiv:1803.00676*, 2018. 1

[12] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015. 1

[13] Christian Simon, Piotr Koniusz, Richard Nock, and Mehrtash Harandi. Adaptive subspaces for few-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4136–4145, 2020. 1

[14] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Koray Kavukcuoglu, and Daan Wierstra. Matching networks for one shot learning. *arXiv preprint arXiv:1606.04080*, 2016. 1

[15] Yan Wang, Wei-Lun Chao, Kilian Q Weinberger, and Laurens van der Maaten. Simpleshot: Revisiting nearest-neighbor classification for few-shot learning. *arXiv preprint arXiv:1911.04623*, 2019. 1, 3

[16] Yikai Wang, Chengming Xu, Chen Liu, Li Zhang, and Yanwei Fu. Instance credibility inference for few-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12836–12845, 2020. 1

[17] Imtiaz Ziko, Jose Dolz, Eric Granger, and Ismail Ben Ayed. Laplacian regularized few-shot learning. In *International Conference on Machine Learning*, pages 11660–11670. PMLR, 2020. 3