Supplementary Material: Learning to Hallucinate Examples from Extrinsic and Intrinsic Supervision

Liangke Gui[1*] Adrien Bardes[2*†] Ruslan Salakhutdinov[1]
Alexander Hauptmann[1] Martial Hebert[1] Yu-Xiong Wang[3]
[1] CMU [2] Facebook AI Research, Inria [3] UIUC
{liangkeg, rsalakhu, alex, hebert}@cs.cmu.edu adrien.bardes@inria.fr yxw@illinois.edu

A. Additional Implementation Details

Training Time. As the DMAS hallucinator is a multi-layer perceptron (MLP), the training is very fast. We use Nvidia Quadro RTX 8000 with 48GB memory to train our models. The training time for each iteration on a single GPU is 0.7 seconds. When training our model for $12,000$ iterations, the total training time on the *mini*ImageNet dataset is around 2.5 hours.

Training of Mentor Classifiers. The mentor classifier is a large-sample classifier and is thus trained in a standard batch mode. We use a mini-batch with 256 image features which is randomly sampled from $\mathcal{S}_{\text{large}}$.

B. Inductive/Transductive Methods

For a fair comparison, we only use the training set as the meta-training set (*i.e.*, the inductive learning scenario). Note that to further boost the few-shot learning accuracy, recent methods consider leveraging the transducive learning scenario, where they have access to the test data. For example, transducive fine-tuning [4] fine-tunes the network on the meta-testing set and uses information from the testing data. It performs gradient updates during the fine-tuning phase, which makes it slow (*e.g.*, 50x slower for a single query shot) at inference time [4]. SIB + E^3BM [10] meta-learns an ensemble of epoch-wise empirical Bayes models (E^3BM) to achieve robust predictions. The comparison with these methods is shown in Table A. Under the inductive learning setting without having access to the meta-testing set, we outperform the state-of-the-art methods by a large margin. *Notably, our inductive approach with a shallow network achieves comparable performance with and in some cases even outperforms state-of-the-art transductive learning methods.*

Note that, there are other transductive learning approaches by having access to the meta-testing set [6, 7, 11, 5, 9], learning with external data [1, 12], and model ensemble [8]. In our work, we only consider inductive learning without accessing any meta-testing set or external data. On the other hand, *our data hallucination approach is agnostic to the choice of learning settings, and can be further improved by leveraging transductive learning*. We leave this as future work.

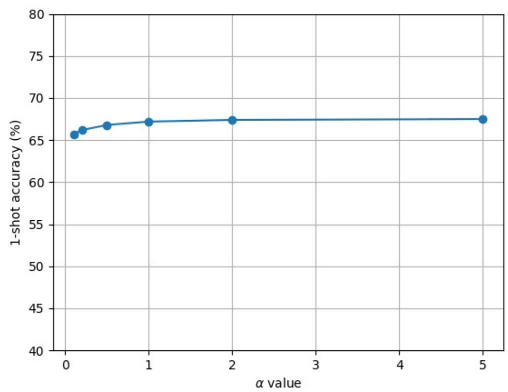
C. Additional Comparisons and Ablation Studies

Additional Comparisons with State of the Art. In the main paper, we provided extensive comparisons with state-of-the-art methods. Here in Table A, we present additional comparisons on *mini*ImageNet and *tiered*ImageNet. Our model consistently outperforms these methods as well by large margins. In addition, we further combine DMAS with DeepEMDv2-Samping in [14] – a more advanced variant of the best-performing baseline DeepEMD [15]. Table B shows that our DMAS can work with DeepEMDv2-Samping and improves its performance, which is consistent with the observations in Table 4 in the main paper.

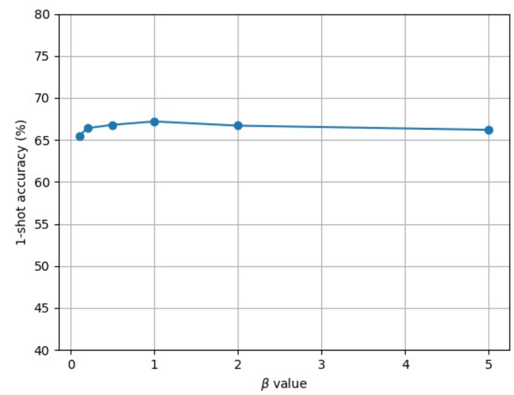
Additional Comparisons with Data Augmentation. We also compare our DMAS with other data augmentation strategies such as RandAugment [3]. With the same ResNet12 backbone, Table C shows that DMAS consistently outperforms RandAugment by large margins.

Analysis of Hyper-Parameter Sensitivity. We conduct sensitivity experiments on the *mini*ImageNet dataset for the hyper-parameters α and β, which trade off different loss components in the overall objective of our DMAS hallucinator. We vary one of the hyper-parameters while fixing the remaining one to its cross-validated value. As shown in Figure A, the performance is *stable* over different hyper-parameter values. Across the board with different hyper-parameter values, our DMAS consistently and significantly outperforms the baselines shown in the main paper. In addition, we use *the same set of hyper-parameter values for all the datasets*, further showing the generalizability of our approach.

*equal contribution
†work done while visiting CMU



(a) We fix β as 1 and evaluate the 1-shot classification accuracy with different α



(b) We fix α as 5 and evaluate the 1-shot classification accuracy with different β.

Figure A: Sensitivity analysis of the trade-off hyper-parameters α and β in the overall objective on *mini*ImageNet. The performance of our DMAS is *stable* over different hyper-parameter values. In addition, we use *the same set of hyper-parameter values for all the datasets*, further showing the generalizability of our approach.

Learning setting	Method	Backbone	*mini*ImageNet		*tiered*ImageNet	
			k=1	5	k=1	5
Transductive learning	Trans-FT [4]	WRN-28-10	65.73 ± 0.68	78.40 ± 0.52	73.34 ± 0.71	85.50 ± 0.50
	SIB+E^{3}BM [10]	ResNet25	**71.4** ± **–**	81.2 ± –	**75.6** ± **–**	84.3 ± –
Inductive learning	Trans-FT [4]	WRN-28-10	57.73 ± 0.62	78.17 ± 0.49	66.58 ± 0.70	85.55 ± 0.48
	MTL+E^{3}BM [10]	ResNet25	64.3 ± –	81.0 ± –	70.0 ± –	85.0 ± –
	Meta-baseline [2]	ResNet12	63.17 ± 0.23	79.26 ± 0.17	68.62 ± 0.27	83.29 ± 0.18
Inductive learning	**DMAS (Ours)**	ResNet12	67.42 ± 0.28	**83.74** ± **0.20**	73.54 ± 0.73	**86.27** ± **0.47**

Table A: Comparison of inductive/transductive learning methods. Our model outperforms other baseline methods under the same inductive learning setting, while achieving comparable performance with and in some cases even outperforming the transductive learning methods. In addition, our data hallucination approach is agnostic to the choice of learning settings, and can be further improved by leveraging transductive learning.

Method	*mini*ImageNet		CUB	
	k=1	5	k=1	5
DeepEMDv2-Sampling [14]	68.77 ± 0.29	84.13 ± 0.53	79.27 ± 0.29	89.80 ± 0.51
DeepEMDv2-Sampling + DMAS (Ours)	**69.45** ± **0.15**	**84.50** ± **0.20**	**80.05** ± **0.62**	**90.75** ± **0.35**

Table B: Additional ablation study on the generalizability of our approach and comparison with DeepEMDv2-Sampling – a more advanced variant of the best-performing baseline DeepEMD [15]. Our DMAS hallucinator can combine with DeepEMDv2-Sampling and improve its performance as well.

Method	*mini*ImageNet		*tiered*ImageNet	
	k=1	5	k=1	5
RandAugment [3]	62.72 ± 0.57	79.60 ± 0.25	70.34 ± 0.71	84.92 ± 0.59
DMAS (Ours)	**67.42** ± **0.28**	**83.74** ± **0.20**	**73.54** ± **0.73**	**86.27** ± **0.47**

Table C: Ours outperforms RandAugment by large margins.

D. Visualization of Hallucinated Examples

As shown in Figure B, we visualize the hallucinated examples for novel classes using t-SNE [13]. The hallucinated examples introduce variations to the few real examples. By jointly leveraging the complementary extrinsic and intrinsic supervision, the hallucinated examples are able to help the classification algorithm learn better classifier decision boundaries.

References

[1] Da Chen, Yuefeng Chen, Yuhong Li, Feng Mao, Yuan He, and Hui Xue. Self-supervised learning for few-shot image classification. In *ICASSP*, 2021. 1

[2] Yinbo Chen, Zhuang Liu, Huijuan Xu, Trevor Darrell, and Xiaolong Wang. Meta-baseline: Rethinking the effectiveness of simple meta-learning for few-shot learning. In *ICCV*, 2021. 2

[3] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V

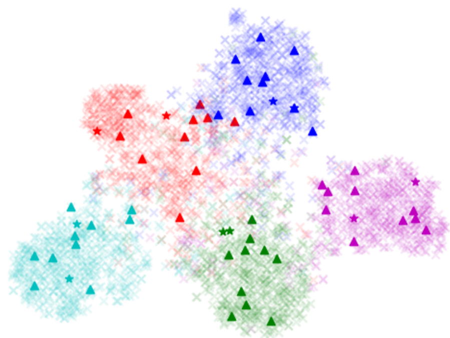


Figure B: t-SNE visualizations of hallucinated examples for five novel classes on *mini*ImageNet. Seeds as stars, real examples as crosses, hallucinated examples as triangles.

Le. RandAugment: Practical automated data augmentation with a reduced search space. In *CVPR Workshops*, 2020. 1, 2

[4] Guneet S Dhillon, Pratik Chaudhari, Avinash Ravichandran, and Stefano Soatto. A baseline for few-shot image classification. In *ICLR*, 2020. 1, 2

[5] Yuqing Hu, Vincent Gripon, and Stéphane Pateux. Graph-based interpolation of feature vectors for accurate few-shot classification. In *ICPR*, 2020. 1

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

[7] Seong Min Kye, Hae Beom Lee, Hoirin Kim, and Sung Ju Hwang. Meta-learned confidence for few-shot learning. *arXiv preprint arXiv:2002.12017*, 2020. 1

[8] Jialin Liu, Fei Chao, and Chih-Min Lin. Task augmentation by rotating for meta-learning. *arXiv preprint arXiv:2003.00804*, 2020. 1

[9] Jinlu Liu, Liang Song, and Yongqiang Qin. Prototype rectification for few-shot learning. In *ECCV*, 2020. 1

[10] Yaoyao Liu, Bernt Schiele, and Qianru Sun. An ensemble of epoch-wise empirical Bayes for few-shot learning. In *ECCV*, 2020. 1, 2

[11] Pau Rodríguez, Issam Laradji, Alexandre Drouin, and Alexandre Lacoste. Embedding propagation: Smoother manifold for few-shot classification. In *ECCV*, 2020. 1

[12] Eli Schwartz, Leonid Karlinsky, Rogerio Feris, Raja Giryes, and Alex M Bronstein. Baby steps towards few-shot learning with multiple semantics. *arXiv preprint arXiv:1906.01905*, 2019. 1

[13] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *JMLR*, 9(11):2579–2605, 2008. 2

[14] Chi Zhang, Yujun Cai, Guosheng Lin, and Chunhua Shen. DeepEMD: Differentiable earth mover's distance for few-shot learning. *arXiv preprint arXiv:2003.06777*, 2020. 1, 2

[15] Chi Zhang, Yujun Cai, Guosheng Lin, and Chunhua Shen. DeepEMD: Few-shot image classification with differentiable earth mover's distance and structured classifiers. In *CVPR*, 2020. 1, 2