Data-Centric Meta-Learning for Robust Few-Shot Generalization

Supplementary Material

A. Additional Dataset Descriptions

miniImageNet The miniImageNet [19] dataset consists of 100 classes randomly sampled from ImageNet [16], each containing 600 images resized to 84×84 pixels. We follow the standard 64/16/20 split of classes into meta-training, meta-validation, and meta-test sets, respectively, with no overlap between the splits.

tieredImageNet Similar to miniImageNet, tieredImageNet [15] is also constructed from ImageNet [16] but is larger in scale. It contains 779,165 images from 608 classes, grouped into 34 high-level categories. These categories are split into 20 for meta-training, 6 for meta-validation, and 8 for meta-testing.

CIFAR-FS The CIFAR-FS [4] dataset is a few-shot benchmark derived from CIFAR-100 [10]. It contains 100 classes with 600 images per class, and all images are resized from 32×32 to 84×84 pixels. We use 64 classes for meta-training, 16 for meta-validation, and 20 for meta-testing.

FC100 The FC100 dataset [13] is another CIFAR-100-based benchmark designed to increase task difficulty by grouping classes into superclasses. It includes 100 classes with 600 images per class, resized to 84×84 pixels. We follow the 60/20/20 split into meta-training, meta-validation, and meta-test classes.

CUB The Caltech-UCSD Birds 200-2011 (CUB) dataset [20] contains 200 bird species (classes) and a total of 11,788 images. Following prior work [11], we split the dataset into 100 meta-training classes, 50 meta-validation classes, and 50 meta-test classes, with all images resized to 84×84.

B. Additional Implementation Details

Backbone Configurations We use Conv4 and ResNet12 as the backbone network f_θ. The 4-layer convolutional network (4-CONV) consists of four convolutional blocks, each composed of a 3×3 convolution with 128 filters, followed by batch normalization [7], a ReLU activation, and a 2×2 max pooling operation. This architecture is widely adopted in few-shot learning [8]. A fully connected layer and a softmax classifier are appended after the final convolutional block for classification.

The ResNet-12 [6] backbone is composed of four residual blocks. Each residual block contains three convolutional layers with 3×3 filters, where each convolution is followed by batch normalization [7] and a ReLU activation. A skip connection is applied in each block, consisting of a 1×1 convolution, batch normalization, and a ReLU. Each residual block concludes with a 2×2 max pooling layer. The number of filters in the residual blocks is set to 64, 128, 256, and 512, respectively. The final representation is aggregated by global average pooling and passed through a fully connected layer with softmax for classification.

Training Details We adopt SGD for inner-level updates and Adam [9] for outer-level updates with a weight decay of 0.0005. Meta-training is performed for 150 epochs on miniImageNet, CIFAR-FS, and FC100, and for 250 epochs on tieredImageNet, each with 500 iterations per epoch. The meta-batch size is set to 2 for 1-shot and 4 for 5-shot tasks. As data augmentation, we apply random horizontal and vertical flipping during meta-training.

Reproducibility and Hardware Following standard few-shot evaluation protocols [5, 8], we evaluate all methods on 600 randomly sampled 5-way 1-shot and 5-shot tasks with 15 query images per class. Accuracy is computed on the query set after adaptation using the support set. We report the mean and 95% confidence interval across the 600 tasks. For each run, we select the top 5 models based on validation accuracy and report the result of the best-performing model.

C. Additional Results

In-domain Results. We present additional experiments using the ResNet-12 backbone to further validate the effectiveness and generality of DCML. As shown in Tables 1 and 2, DCML consistently achieves strong performance on miniImageNet, tieredImageNet, CIFAR-FS, and FC100 under both 1-shot and 5-shot settings. Across all benchmarks, DCML either matches or surpasses representative optimization-based meta-learning methods, demonstrating that the proposed data-centric alignment mechanism remains effective even when applied to deeper feature extractors. The improvement is consistent across datasets and shot settings, indicating that DCML adapts reliably to diverse task distributions. Although several recent approaches benefit from external pre-training, DCML attains competitive performance without such reliance, highlighting its ability to acquire generalizable prior knowledge purely through meta-training.

Table 1. Few-shot in-domain classification accuracy on miniImageNet and tieredImageNet using ResNet-12 backbone.

Method	miniImageNet		tieredImageNet	
	1-shot	5-shot	1-shot	5-shot
MAML [5]	58.37±0.49	69.76±0.46	58.58±0.49	71.24±0.43
ANIL [14]	51.58±0.50	72.22±0.45	58.97±0.00	71.80±0.50
BOIL [12]	55.86±0.50	69.94±0.46	59.63±0.49	73.86±0.44
L2F [2]	59.71±0.49	77.04±0.42	64.04±0.48	81.13±0.39
ALFA [1]	59.74±0.49	77.96±0.42	64.62±0.49	82.48±0.38
MeTAL [3]	59.64±0.38	76.20±0.19	63.89±0.43	80.14±0.40
CxGrad [11]	60.19±0.45	75.17±0.40	65.47±0.44	82.52±0.35
Meta-AdaM [17]	59.89±0.49	77.92±0.43	65.31±0.48	85.24±0.35
MetaDiff [21]†	*64.99±0.77*	*81.21±0.56*	*72.33±0.92*	*86.31±0.62*
DCML (Ours)	60.38±0.49	77.54±0.43	66.24±0.47	84.06±0.37

† Baseline MetaDiff [21] utilizes a pre-trained backbone, whereas all other methods, including ours, are trained from scratch without external pre-training.

Table 2. Few-shot in-domain classification accuracy on CIFAR-FS and FC100 using ResNet-12 backbone.

Method	CIFAR-FS		FC100	
	1-shot	5-shot	1-shot	5-shot
MAML [5]	64.33±0.48	76.38±0.42	37.92±0.48	52.62±0.50
ANIL [14]	62.38±0.48	74.87±0.11	37.98±0.98	52.67±0.50
BOIL [12]	64.21±0.49	77.38±0.10	39.88±0.49	52.00±0.50
L2F [2]	67.48±0.46	82.79±0.38	41.89±0.47	54.68±0.50
ALFA [1]	64.14±0.48	78.11±0.41	40.57±0.49	53.19±0.50
MeTAL [3]	67.97±0.47	82.17±0.38	39.98±0.39	53.85±0.36
MetaDiff [21]†	*73.60±0.85*	*87.43±0.60*	*44.42±0.77*	*61.02±0.75*
DCML (Ours)	69.87±0.45	83.56±0.36	42.40±0.49	56.68±0.50

† Baseline MetaDiff [21] utilizes a pre-trained backbone, whereas all other methods, including ours, are trained from scratch without external pre-training.

Table 3. Few-shot cross-domain classification accuracy under the 5-way 5-shot setting using ResNet-12 backbone.

Method	miniImagenet	
	→ CUB	→ CIFAR-FS
MAML [5]	53.83±0.32	61.66±0.49
ANIL [14]	54.69±0.50	64.09±0.48
BOIL [12]	55.25±0.50	59.61±0.49
L2F [2]	62.12±0.21	65.90±0.46
ALFA [1]	61.22±0.22	64.74±0.46
MeTAL [3]	61.29±0.21	64.98±0.46
DCML (Ours)	64.39±0.48	66.50±0.47

Cross-domain Results. We further evaluate cross-domain generalization using the ResNet-12 backbone by meta-training on miniImageNet and testing on both CUB and CIFAR-FS. As reported in Table 3, DCML achieves the highest accuracy on the miniImageNet→CUB transfer and also delivers strong performance on the miniImageNet→CIFAR-FS transfer, demonstrating improved robustness under distribution shift. These results indicate that DCML maintains its effectiveness even when the target domain differs significantly from the meta-training distribution, owing to the generalizable prior knowledge acquired during meta-training.

Large-scale Dataset Results. We evaluate DCML on Meta-Dataset [18], a large-scale dataset benchmark that samples few-shot tasks from multiple visual domains. This benchmark introduces substantial variation in visual appearance and task difficulty, providing a more realistic evaluation beyond standard benchmarks. All models are meta-trained on ILSVRC-2012 following the standard protocol, using the same backbone as fo-Proto-MAML [18] for fair comparison. As reported in Table 4, DCML maintains consistent performance across domains, highlighting that the robust generalization induced by data-centric meta-learning remains important even in large-scale evaluation settings with highly diverse task distributions.

Method	ILSVRC	Omniglot	Aircraft	Birds	Texture	Quick Draw	Fungi	VGG Flower	Traffic Signs	MSCOCO
fo-Proto-MAML [18]	49.53±1.05	63.37±1.33	55.95±0.99	68.66±0.96	66.49±0.83	51.52±1.00	39.96±1.14	87.15±0.69	48.83±1.09	43.74±1.12
ALFA [1]	52.80±1.11	61.87±1.51	63.43±1.10	69.75±1.05	70.78±0.88	59.17±1.16	41.49±1.17	85.96±0.77	60.78±1.29	48.11±1.14
GAP [8]†	52.75±1.13	63.74±1.42	64.51±1.10	69.15±1.05	71.32±0.88	60.05±1.15	40.53±1.06	87.52±0.78	59.66±1.22	47.10±1.16
DCML (Ours)	54.00±1.23	70.95±1.51	66.70±1.09	70.63±1.15	72.05±0.87	61.13±1.12	41.82±1.07	88.03±0.78	62.36±1.23	49.12±1.15

† Reproduced results.

Table 4. Few-shot classification accuracy on Meta-Dataset.

References

[1] Sungyong Baik, Myungsub Choi, Janghoon Choi, Heewon Kim, and Kyoung Mu Lee. Meta-learning with adaptive hyperparameters. *Advances in neural information processing systems*, 33:20755–20765, 2020. 2, 3

[2] Sungyong Baik, Seokil Hong, and Kyoung Mu Lee. Learning to forget for meta-learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2379–2387, 2020. 2

[3] Sungyong Baik, Janghoon Choi, Heewon Kim, Dohee Cho, Jaesik Min, and Kyoung Mu Lee. Meta-learning with task-adaptive loss function for few-shot learning. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9465–9474, 2021. 2

[4] Luca Bertinetto, Joao F Henriques, Philip HS Torr, and Andrea Vedaldi. Meta-learning with differentiable closed-form solvers. *arXiv preprint arXiv:1805.08136*, 2018. 1

[5] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning*, pages 1126–1135. PMLR, 2017. 1, 2

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 1

[7] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. pmlr, 2015. 1

[8] Suhyun Kang, Duhun Hwang, Moonjung Eo, Taesup Kim, and Wonjong Rhee. Meta-learning with a geometry-adaptive preconditioner. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16080–16090, 2023. 1, 3

[9] Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 1

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

[11] Sanghyuk Lee, Seunghyun Lee, and Byung Cheol Song. Contextual gradient scaling for few-shot learning. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 834–843, 2022. 1, 2

[12] Jaehoon Oh, Hyungjun Yoo, ChangHwan Kim, and Se-Young Yun. Boil: Towards representation change for few-shot learning. *arXiv preprint arXiv:2008.08882*, 2020. 2

[13] Boris Oreshkin, Pau Rodríguez López, and Alexandre Lacoste. Tadam: Task dependent adaptive metric for improved few-shot learning. *Advances in neural information processing systems*, 31, 2018. 1

[14] Aniruddh Raghu, Maithra Raghu, Samy Bengio, and Oriol Vinyals. Rapid learning or feature reuse? towards understanding the effectiveness of maml. *arXiv preprint arXiv:1909.09157*, 2019. 2

[15] 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

[16] 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*, 115:211–252, 2015. 1

[17] Siyuan Sun and Hongyang Gao. Meta-adam: An meta-learned adaptive optimizer with momentum for few-shot learning. *Advances in Neural Information Processing Systems*, 36:65441–65455, 2023. 2

[18] Eleni Triantafillou, Tyler Zhu, Vincent Dumoulin, Pascal Lamblin, Utku Evci, Kelvin Xu, Ross Goroshin, Carles Gelada, Kevin Swersky, Pierre-Antoine Manzagol, et al. Meta-dataset: A dataset of datasets for learning to learn from few examples. *arXiv preprint arXiv:1903.03096*, 2019. 2, 3

[19] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. *Advances in neural information processing systems*, 29, 2016. 1

[20] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. 2011. 1

[21] Baoquan Zhang, Chuyao Luo, Demin Yu, Xutao Li, Huiwei Lin, Yunming Ye, and Bowen Zhang. Metadiff: Meta-learning with conditional diffusion for few-shot learning. In *Proceedings of the AAAI conference on artificial intelligence*, pages 16687–16695, 2024. 2