A. Full Dataset Details

We describe the full details of our multi-task meta-dataset in Table 6 and provide further high-level details in this section.

- Classification datasets: We reuse datasets selected in the initial release of Meta-Album [65]. We split BCT (microscopy – bacteria) [83], BRD (large-animals – birds) [51] and CRS (vehicles – cars) [28] datasets into meta-training, in-domain meta-validation and in-domain meta-testing splits. We perform the splits randomly and in terms of classes – 70% for training, 15% validation and testing each. FLW (plants – flowers) [46], MD-Mix (OCR) [60] and PLK (small animals – plankton) [24] datasets are used for out-domain meta-validation. PLT-VIL (plant diseases) [18], RESISC (remote sensing) [12], SPT (human actions – sports) [50] and TEX (manufacturing – textures) [17, 29, 31, 40] for out-domain meta-testing. We use the middle version ("Mini") of these datasets as processed by the authors of Meta-Album [65], which allows us to keep the overall size of Meta Omniumsufficiently small.

- Segmentation datasets: We first split FSS1000 [37] dataset into in-domain train, validation, and test sets, i.e. FSS1000-Trn, FSS1000-Val, FSS1000-Test. We use Vizwiz [64] dataset for out-of-domain validation, and a modified version of Pascal 5i [58] and PH2 [41] datasets for out-of-domain testing. We exclude the object classes from the out-of-domain datasets that overlap with FSS1000 to ensure the classes during validation and testing are never seen during training.

- Keypoint estimation datasets: We use three keypoint datasets in the paper, including animal pose [10], synthetic animal pose [44] and human pose [4]. A single animal/human image is cropped from the original picture according to absolute maximum and minimum keypoint coordinates. The boundary is extended with 5 more pixels to avoid losing important information at object edges. Different keypoint datasets would have various target keypoints, so we cannot have a trivial solution like classification with a N-way K-shot setting, which stands for sampling K samples from N categories. Instead, we sample each keypoint task from one object category with only a fixed number of keypoints. In detail, we randomly select 5 keypoints per task, and train and fit the model to predict only 5 keypoints. This method leads to a general meta-learning keypoint prediction model that learns to predict corresponding keypoints from the limited support labels, which makes it possible for an arbitrary number of keypoint prediction tasks when conducted on more complex keypoint datasets.

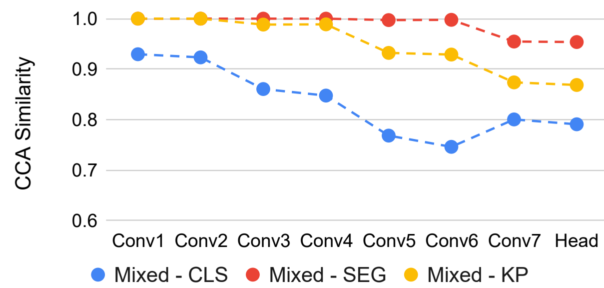

Figure 4. Analysis of the layer adaptation by MAML in Meta Omnium.

- Regression datasets: We use regression datasets only for out-of-task (OOT) meta-test evaluation, so they are not used during meta-training. More specifically we use ShapeNet1D [19], ShapeNet2D [19], Distractor [19] and Pascal1D datasets [76]. Because regression problems typically require larger number of examples for adaptation, we use 5-times as many support examples compared to the other cases (e.g. instead of 5-shot we have 25-shot case). For our analysis experiments we consider the equivalent of variable 1-to-5-shot setting: variable 5-to-25-shot setting.

B. Additional Analysis

How do gradient-based meta-learners adapt their layers? A recent debate in few-shot meta-learning has been around whether gradient-based meta-learners really learn to adapt, or simply reuse features without adaptation. [53] claimed that feature reuse was the dominant effect after measuring the representational change pre- and post-adaptation and finding that representational change was primarily in the output layer. We analyze this using Canonical Correlation Analysis (CCA) [43, 54] for Meta Omnium, reporting the representational change of multi-task MAML by layer for each task family during meta-testing. From the results in Figure 4, we observe that: (1) The degree of representational change varies substantially with tasks, (2) Similar to [53], there is greater representational change at the later layers, especially the final output layer. However, significant amount of adaptation is done also in the earlier layers, which we attribute to the greater diversity of tasks and visual domains in Meta Omnium compared to the simple recognition episodes in miniImageNet studied by [53].

Task Family	Dataset Name	Domain	# Classes	# Images	Cardinality	Role	Size (MB)
Classification	BCT-Trn	Microscopy	23	920	(5)	Meta-train	8
	BRD-Trn	Bird	220	8800	(5)	Meta-train	72
	CRS-Trn	Car	137	5480	(5)	Meta-train	44
	BCT-Val	Microscopy	5	200	(5)	ID Meta-val	1.7
	BRD-Val	Bird	47	1880	(5)	ID Meta-val	15
	CRS-Val	Car	29	1160	(5)	ID Meta-val	9
	FLW	Flowers	102	4080	(5)	OD Meta-val	39
	MD-MIX	OCR	706	28240	(5)	OD Meta-val	479
	PLK	Plankton	86	3440	(5)	OD Meta-val	36
	BCT-Test	Microscopy	5	200	(5)	ID Meta-test	1.7
	BRD-Test	Bird	48	1920	(5)	ID Meta-test	16
	CRS-Test	Car	30	1200	(5)	ID Meta-test	10
	PLT-VIL	Plant Disease	38	1520	(5)	OD Meta-test	14
	RESISC	Remote Sensing	45	1800	(5)	OD Meta-test	17
	SPT	Sports	73	2920	(5)	OD Meta-test	27
	TEX	Textures	64	2560	(5)	OD Meta-test	26
Segmentation	FSS1000-Trn	Natural Image	520	5200	(2)	Meta-train	331
	FSS1000-Val	Natural Image	240	2400	(2)	ID Meta-val	150
	FSS1000-Test	Natural Image	240	2400	(2)	ID Meta-test	53
	Pascal 5i	Natural Image	6	7247	(2)	OD Meta-test	563
	Vizwiz	Natural Image	22	862	(2)	OD Meta-val	24
	PH2 (Skin)	Medical Image	3	200	(2)	OD Meta-test	114
Keypoint Regression	Animal pose - Trn	Animal	2	3237	(20, 2)	Meta-train	112
	Animal pose - Val	Animal	2	2038	(20, 2)	ID Meta-val	54
	Animal pose - Test	Animal	1	842	(20, 2)	ID Meta-test	18
	Synthetic Animal Pose	Synthetic Animal	2	20000	(18, 2)	OD Meta-val	627
	MPII	Human	1	28882	(16, 2)	OD Meta-test	265
Regression	ShapeNet1D-Test	Synthetic Image	60	3000	(2)	OOT Meta-test	8
	ShapeNet2D-Test	Synthetic Image	300	9000	(4)	OOT Meta-test	29
	Distractor-Test	Synthetic Image	200	7200	(2)	OOT Meta-test	93
	Pascal1D-Test	Synthetic Image	15	1500	(1)	OOT Meta-test	4

Table 6. Details of all task families included in Meta Omnium.

C. Additional Experimental Details

C.1. Hyperparameter Optimization (HPO)

The details of how we perform HPO are described in Section 3.6, and in this section we provide additional details. The search space for HPO is as follows (note that momentum is only used if SGD optimizer is selected):

- MAML and Meta-Curvature: meta-learning rate $\in (10^{-4}, 10^{-1})$ (log scale), meta optimizer $\in \{\text{Adam}, \text{SGD}\}$, momentum $\in \{0.0, 0.9, 0.99\}$, inner-loop learning rate $\in (10^{-3}, 0.5)$ (log scale)

- Proto-MAML: same as MAML and also parameter $\lambda \in (0.01, 100)$ (log scale) that influences the prototype calculation in the case of keypoint estimation

- ProtoNet: meta-learning rate $\in (10^{-4}, 10^{-1})$ (log scale), meta optimizer $\in \{\text{Adam}, \text{SGD}\}$, momentum $\in \{0.0, 0.9, 0.99\}$ and distance temperature $\in (0.1, 10.0)$ (log scale) that is used for keypoint estimation

- DDRR: meta-learning rate $\in (10^{-4}, 10^{-1})$ (log scale), meta optimizer $\in \{\text{Adam}, \text{SGD}\}$, momentum $\in \{0.0, 0.9, 0.99\}$ and $\lambda \in (0.01, 100)$ (log scale)

- Proto-FineTuning: learning rate $\in (10^{-4}, 10^{-1})$ (log scale), optimizer $\in \{\text{Adam}, \text{SGD}\}$, momentum $\in \{0.0, 0.9, 0.99\}$ and $\lambda \in (0.01, 100)$ (log scale)

- FineTuning: learning rate $\in (10^{-4}, 10^{-1})$ (log scale), optimizer $\in \{\text{Adam}, \text{SGD}\}$ and momentum $\in \{0.0, 0.9, 0.99\}$

- Linear-Readout and TFS: same as FineTuning

After training a model with the candidate configuration for 5,000 iterations, we evaluate its validation performance. We use 100 tasks for evaluating the in-domain validation performance, and additional 100 tasks for evaluation of out-domain performance. As part of our multi-objective HPO, we minimize the validation error rates (or appropriate equivalent) and use each dataset as a separate objective. We perform HPO on the primary variable 1-to-5 shot setting. We use the same hyperparameters also for the 1-shot and 5-shot settings.

Our HPO is reasonably fast, and it generally takes between a few hours up to two days in the slowest cases (using a single NVIDIA 1080 Ti GPU with 12GB memory and using 4 CPUs). As a result, it is feasible to run the HPO even with modest resources when designing new approaches for our multi-task scenario. We provide the found hyperparameters within the released code.

C.2. Experimental Settings

Many of our experimental settings follow Meta-Album [65]), whose code-base we have also used as the starting point. All approaches use one task in a meta-batch. We use 5 inner-loop steps during meta-training and 10 inner-loop steps during evaluation for MAML, Proto-MAML and Meta-Curvature. We use gradient-clipping of 5. DDRR uses an adjustment layer, the scale of which is initialized to 5.0 (with the adjust base set to 1.0). Proto-FineTuning, FineTuning, Linear-Readout and TFS use 20 fine-tuning steps during evaluation. The training minibatch size for these approaches is 16, while the testing minibatch size is 4. We use standard ImageNet normalization for segmentation tasks, but we do not use normalization in the other cases, following earlier work [65].

We train each model for 30,000 iterations and evaluate the model on validation data after every 2,500 tasks, including at the beginning and the end (used for early stopping – model selection). We use 5-way tasks during both training and evaluation. The number of shots is between 1 and 5 during meta-training, and we consider 3 setups for evaluation: variable 1-to-5-shot (primary), 1-shot and 5-shot (presented in the appendix). The query size is 5 examples per category and this has been selected to be consistent across the different datasets. Validation uses 600 tasks for each of in-domain and out-domain evaluation. Testing uses 600 tasks per dataset to provide more rigorous evaluation.

During evaluation, we randomly initialize the top layer weights (classifier) to enable any-way predictions, in line with previous literature [65]. We do this for the approaches that perform fine-tuning (e.g. MAML or Fine-Tuning baseline). Note that in approaches such as Proto-MAML the top layer is initialized using weights derived from the prototypes or ridge regression solution.

D. Detailed Per-Dataset Results

We include detailed per-dataset results (various-shot evaluation), first showing the single-task learning results for classification, segmentation and keypoint estimation, followed by multi-task learning results. In each case, we separately report the results for in-domain and out-of-domain evaluation. We also include detailed results for our out-of-task evaluation using regression datasets. Summary 1-shot and 5-shot results are included for the single and multi-task settings.

E. Full Acknowledgements

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) Grant number EP/S000631/1; the MOD University Defence Research Collaboration (UDRC) in Signal Processing; EPSRC Centre for Doctoral Training in Data Science, funded by the UK Engineering and Physical Sciences Research Council (grant EP/L016427/1) and the University of Edinburgh; and the United Kingdom Research and Innovation (grant EP/S02431X/1), UKRI Centre for Doctoral Training in Biomedical AI at the University of Edinburgh, School of Informatics, and Samsung AI Center, Cambridge. This project was supported by the Royal Academy of Engineering under the Research Fellowship programme. This work was also supported by Key Project Plan of Blockchain in Ministry of Education of the People's Republic of China (Grant No. 2020KJ010802), Innovation and Transformation Fund of Peking University Third Hospital (Grant No. BYSYZHKC2021115) and China Scholarship Council.

Method	BCT-Test	BRD-Test	CRS-Test
MAML	78.09 ± 0.75	64.14 ± 1.17	33.87 ± 0.94
Proto-MAML	74.29 ± 0.83	51.99 ± 1.16	25.25 ± 0.61
Meta-Curvature	85.31 ± 0.66	71.99 ± 1.09	37.19 ± 0.95
ProtoNet	81.53 ± 0.66	75.39 ± 1.07	54.35 ± 1.12
DDRR	76.49 ± 0.80	69.25 ± 1.15	43.52 ± 1.03
Proto-FineTuning	42.92 ± 0.89	68.69 ± 1.17	40.81 ± 1.03
FineTuning	41.48 ± 0.85	52.51 ± 1.15	32.97 ± 0.78
Linear-Readout	45.44 ± 0.87	64.33 ± 1.12	36.11 ± 0.80
TFS	34.19 ± 0.86	35.14 ± 0.93	25.25 ± 0.71

Table 7. In-domain single-task classification results. Mean test accuracy (%) and 95% confidence interval across test tasks.

Method	PLT_VIL	RESISC	SPT	TEX
MAML	62.69 ± 1.14	51.83 ± 1.06	46.24 ± 1.05	85.49 ± 1.02
Proto-MAML	46.59 ± 1.00	39.79 ± 0.94	35.24 ± 0.91	77.11 ± 1.14
Meta-Curvature	61.88 ± 1.07	52.00 ± 1.13	45.11 ± 1.06	86.55 ± 0.98
ProtoNet	59.68 ± 1.15	51.17 ± 1.04	43.65 ± 1.06	83.02 ± 1.03
DDRR	60.28 ± 1.19	48.70 ± 1.04	42.83 ± 1.04	83.17 ± 1.10
Proto-FineTuning	51.50 ± 1.27	41.92 ± 1.06	39.54 ± 1.03	69.77 ± 1.39
FineTuning	46.83 ± 1.13	41.37 ± 0.96	36.03 ± 0.90	68.39 ± 1.23
Linear-Readout	52.68 ± 1.02	46.24 ± 1.00	41.39 ± 0.94	73.45 ± 1.13
TFS	43.87 ± 1.03	36.36 ± 0.90	35.07 ± 0.91	52.54 ± 1.33

Table 8. Out-of-domain single-task classification results. Mean test accuracy (%) and 95% confidence interval across test tasks.

Method	FSS1000-Test
MAML	54.70 ± 1.68
Proto-MAML	46.40 ± 1.62
Meta-Curvature	65.57 ± 1.21
ProtoNet	75.84 ± 0.98
DDRR	66.71 ± 1.20
Proto-FineTuning	59.96 ± 1.55
FineTuning	50.52 ± 1.59
Linear-Readout	34.00 ± 1.85
TFS	42.80 ± 1.52

Table 9. In-domain single-task segmentation results. Mean test mIoU (%) and 95% confidence interval across test tasks. Larger mIoU is better.

Method	Pascal 5i	PH2
MAML	15.27 ± 1.29	68.88 ± 1.25
Proto-MAML	22.80 ± 1.19	65.46 ± 1.19
Meta-Curvature	27.66 ± 1.22	71.92 ± 0.80
ProtoNet	36.49 ± 1.39	77.82 ± 0.79
DDRR	29.07 ± 1.12	66.95 ± 0.77
Proto-FineTuning	21.03 ± 1.24	65.79 ± 1.21
FineTuning	16.23 ± 1.24	63.68 ± 1.11
Linear-Readout	5.67 ± 0.80	39.67 ± 1.91
TFS	15.45 ± 1.03	59.77 ± 1.18

Table 10. Out-of-domain single-task segmentation results. Mean test mIoU (%) and 95% confidence interval across test tasks. Larger mIoU is better.

Method	Animal pose - Test
MAML	25.36 ± 0.93
Proto-MAML	23.63 ± 0.84
Meta-Curvature	43.47 ± 0.99
ProtoNet	27.79 ± 0.89
DDRR	20.53 ± 0.72
Proto-FineTuning	21.27 ± 0.74
FineTuning	25.69 ± 0.90
Linear-Readout	22.09 ± 0.74
TFS	20.98 ± 0.63

Table 11. In-domain single-task keypoint estimation results. Mean test PCK (%) and 95% confidence interval across test tasks. Larger PCK is better.

Method	MPII
MAML	33.04 ± 0.64
Proto-MAML	22.48 ± 0.64
Meta-Curvature	16.00 ± 0.39
ProtoNet	33.33 ± 0.71
DDRR	31.88 ± 0.63
Proto-FineTuning	33.10 ± 0.71
FineTuning	30.03 ± 0.53
Linear-Readout	26.86 ± 0.46
TFS	25.95 ± 0.52

Table 12. Out-of-domain single-task keypoint estimation results. Mean test PCK (%) and 95% confidence interval across test tasks. Larger PCK is better.

Method	FSS1000-Test	BCT-Test	BRD-Test	CRS-Test	Animal pose - Test
MAML	43.31 ± 1.60	89.05 ± 0.61	59.94 ± 0.99	28.19 ± 0.75	24.25 ± 0.79
Proto-MAML	53.03 ± 1.51	84.71 ± 0.70	59.79 ± 1.12	30.87 ± 0.87	21.63 ± 0.76
Meta-Curvature	42.60 ± 1.74	85.43 ± 0.66	76.85 ± 1.06	48.97 ± 1.09	18.21 ± 0.47
ProtoNet	63.32 ± 1.09	81.95 ± 0.68	72.31 ± 1.05	43.58 ± 1.04	20.10 ± 0.75
DDRR	40.39 ± 1.11	77.19 ± 0.73	51.47 ± 1.11	29.59 ± 0.83	22.77 ± 0.73
Proto-FineTuning	44.80 ± 1.62	41.11 ± 0.91	71.21 ± 1.21	44.85 ± 1.06	21.16 ± 0.74
FineTuning	41.31 ± 1.74	42.84 ± 0.89	55.41 ± 1.22	34.05 ± 0.81	18.05 ± 0.49
Linear-Readout	41.53 ± 1.66	39.07 ± 0.78	63.60 ± 1.04	35.18 ± 0.81	19.89 ± 0.52
TFS	38.66 ± 1.56	22.45 ± 0.54	22.74 ± 0.49	20.39 ± 0.39	14.09 ± 0.75

Table 13. In-domain multi-task learning results. Mean test score (%) and 95% confidence interval across test tasks. Larger score is better in all cases.

Method	PLT_VIL	RESISC	SPT	TEX	Pascal 5i	PH2	MPII
MAML	60.81 ± 1.11	48.19 ± 1.04	39.23 ± 0.91	85.59 ± 1.02	15.57 ± 1.11	59.28 ± 1.32	23.85 ± 0.47
Proto-MAML	65.18 ± 1.18	54.04 ± 1.10	49.84 ± 1.09	85.85 ± 0.95	21.90 ± 1.16	64.51 ± 1.04	33.34 ± 0.68
Meta-Curvature	70.98 ± 1.09	56.05 ± 1.15	51.09 ± 1.18	89.63 ± 0.80	13.29 ± 1.12	55.78 ± 1.52	25.29 ± 0.39
ProtoNet	60.55 ± 1.10	50.13 ± 1.04	41.92 ± 1.05	82.71 ± 1.00	30.46 ± 1.11	68.95 ± 0.87	33.00 ± 0.69
DDRR	53.19 ± 1.13	41.49 ± 0.98	35.73 ± 0.98	77.05 ± 1.19	20.19 ± 0.68	54.35 ± 0.78	30.08 ± 0.59
Proto-FineTuning	55.05 ± 1.21	44.17 ± 1.05	41.79 ± 1.04	71.64 ± 1.33	15.19 ± 1.06	60.47 ± 1.30	30.04 ± 0.60
FineTuning	50.88 ± 1.16	43.59 ± 1.00	38.07 ± 0.94	72.41 ± 1.18	11.68 ± 1.02	60.58 ± 1.39	20.46 ± 0.33
Linear-Readout	49.78 ± 1.00	44.57 ± 0.98	40.83 ± 0.97	68.58 ± 1.12	14.37 ± 1.07	50.73 ± 1.96	23.47 ± 0.35
TFS	23.15 ± 0.51	23.01 ± 0.47	22.45 ± 0.49	26.63 ± 0.71	13.12 ± 0.96	58.41 ± 1.25	11.04 ± 0.29

Table 14. Out-of-domain multi-task learning results. Mean test score (%) and 95% confidence interval across test tasks. Larger score is better in all cases.

Method	ShapeNet2D-Test	Distractor-Test	ShapeNet1D-Test	Pascal1D-Test
MAML	95.44 ± 2.82	38.68 ± 0.60	54.20 ± 2.10	2.88 ± 0.10
Proto-MAML	69.94 ± 1.50	39.58 ± 0.69	63.17 ± 2.33	51.74 ± 1.15
Meta-Curvature	70.55 ± 2.32	38.73 ± 0.66	43.17 ± 2.03	12.04 ± 0.61
ProtoNet	63.50 ± 1.15	38.53 ± 0.58	84.53 ± 1.92	2.53 ± 0.06
DDRR	64.41 ± 1.64	41.67 ± 0.68	46.08 ± 1.90	2.11 ± 0.07
Proto-FineTuning	61.94 ± 2.25	39.44 ± 0.69	58.33 ± 2.54	4.17 ± 0.18
FineTuning	63.36 ± 1.54	51.52 ± 1.44	81.36 ± 2.12	6.54 ± 0.31
Linear-Readout	68.66 ± 1.94	40.53 ± 0.71	46.93 ± 2.05	2.62 ± 0.09
TFS	133.67 ± 2.67	95.36 ± 0.91	88.25 ± 1.88	6.63 ± 0.31

Table 15. Evaluation of multi-task models on out-of-task regression datasets, using variable 5-to-25-shot episodes. Lower value is better.

		Classification		**Segmentation**		**Keypoints**		**Average Rank**		
		ID	OOD	ID	OOD	ID	OOD	ID	OOD	AVG
Single-Task	MAML	50.8	50.8	44.1	33.6	34.7	33.4	4.3	4.0	4.2
	Proto-MAML	53.5	52.4	46.0	39.2	23.5	14.8	4.7	4.7	4.7
	Meta-Curvature	58.0	51.6	60.5	40.1	38.1	16.1	**1.7**	4.3	**3.0**
	ProtoNet	61.7	50.2	73.7	52.5	22.5	31.9	2.7	**3.7**	3.2
	DDRR	54.7	48.9	60.1	42.2	22.1	32.2	4.7	4.0	4.3
	Proto-FineTuning	47.0	50.4	50.5	36.6	22.4	32.8	5.7	4.3	5.0
	FineTuning	35.5	43.2	42.4	36.6	34.6	33.8	6.0	4.7	5.3
	Linear-Readout	46.2	48.0	30.3	18.3	26.5	26.7	6.7	7.3	7.0
	TFS	27.2	36.4	31.5	30.3	19.5	20.0	8.7	8.0	8.3
Multi-Task	MAML	56.4	54.0	35.0	28.8	29.1	29.0	3.0	4.3	3.7
	Proto-MAML	50.5	51.7	43.6	34.2	22.5	32.7	3.0	**2.3**	2.7
	Meta-Curvature	62.4	56.1	29.2	25.6	16.0	22.3	5.7	5.7	5.7
	ProtoNet	60.8	50.8	59.2	42.2	22.5	31.9	**2.3**	2.7	**2.5**
	DDRR	47.3	46.2	36.4	33.2	19.6	29.3	5.0	4.7	4.8
	Proto-FineTuning	47.2	52.1	33.6	33.9	19.2	28.1	6.3	3.7	5.0
	FineTuning	37.2	43.9	38.2	33.4	23.5	23.8	4.3	5.7	5.0
	Linear-Readout	40.3	43.7	24.5	22.3	21.3	22.8	7.0	8.0	7.5
	TFS	22.0	24.2	30.2	30.2	9.4	9.3	8.3	8.0	8.2

Table 16. 5-way 1-shot results, reporting the same metrics as in our primary table with variable-shot results.

		Classification		Segmentation		Keypoints		Average Rank		
		ID	OOD	ID	OOD	ID	OOD	ID	OOD	AVG
Single-Task	MAML	63.2	67.7	57.6	45.0	22.2	33.6	4.7	3.3	4.0
	Proto-MAML	57.0	52.3	49.9	46.7	22.3	29.6	5.0	5.7	5.3
	Meta-Curvature	69.0	67.2	73.5	53.5	43.7	16.3	1.7	4.3	3.0
	ProtoNet	74.3	64.0	75.9	55.9	29.4	33.9	**1.3**	**2.0**	**1.7**
	DDRR	68.0	65.7	69.4	50.5	22.0	32.0	4.3	3.7	4.0
	Proto-FineTuning	52.9	52.0	65.9	48.7	22.2	33.9	5.0	4.3	4.7
	FineTuning	43.8	50.0	55.3	42.7	22.1	33.1	6.7	6.3	6.5
	Linear-Readout	53.6	55.0	32.3	32.8	20.0	27.1	8.0	7.3	7.7
	TFS	33.8	44.4	47.4	40.5	20.9	28.2	8.3	8.0	8.2
Multi-Task	MAML	68.3	72.1	52.0	42.1	20.7	31.4	3.3	3.0	3.2
	Proto-MAML	67.0	71.2	63.0	48.5	23.2	34.0	**2.7**	2.3	**2.5**
	Meta-Curvature	76.7	73.8	49.7	38.1	19.6	27.7	4.3	5.3	4.8
	ProtoNet	71.0	63.4	64.7	52.4	19.7	34.5	3.0	**2.0**	**2.5**
	DDRR	58.0	59.2	42.5	38.3	23.5	30.0	4.7	6.0	5.3
	Proto-FineTuning	52.7	51.3	50.4	40.6	21.3	32.5	4.3	5.0	4.7
	FineTuning	47.8	54.2	46.6	41.5	18.1	22.3	7.3	6.0	6.7
	Linear-Readout	48.2	50.9	45.8	38.4	19.7	25.5	6.3	7.0	6.7
	TFS	22.4	23.9	40.7	38.3	15.8	12.1	9.0	8.3	8.7

Table 17. 5-way 5-shot results, reporting the same metrics as in our primary table with variable-shot results.