Few-Shot Metric Learning: Online Adaptation of Embedding for Retrieval -*Supplementary Material*-

Deunsol Jung, Dahyun Kang, Suha Kwak, and Minsu Cho

Pohang University of Science and Technology (POSTECH), South Korea
{deunsol.jung,dahyun.kang,suha.kwak,mscho}@postech.ac.kr

1 Implementation details

Code base and GPUs. We implement FSML_{SFT}, $\text{FSML}_{\text{MAML}}$, and CRML on top of the code base from [2][1]. To implement FSML_{MTL}, we heavily borrow the core part from its public implementation from [14][2]. We use Nvidia TitanXP GPUs for experiments. Our code is available at https://github.com/hesedjds/FSML

Training details. To train models on *mini*ImageNet [10], CUB [15], the cross-domain setting [2], and *mini*DeepFashion datasets, we use Adam optimizer [5] with learning rate 10^{-3} from scratch unlike most conventional metric learning methods use ImageNet [6] pre-trained models. To train models on MPII dataset [1] for human pose retrieval, we train models using Adam with learning rate $5 \cdot 10^{-4}$ on top of ImageNet [11] pre-trained models. For $\text{FSML}_{\text{MAML}}$, we meta-update the initialization using Adam optimizer with the learning rate 10^{-3}. For FSML_{MTL}, we meta-update the initialization of the last layer and scaling and shifting (SS) parameters using Adam optimizer with the learning rate 10^{-3} for SS parameters and 10^{-4} for the initialization of the last layer from pre-trained model by the conventional deep metric learning procedure on meta-training set D^{mtr}. In the experiments on more than 5-way meta-testing settings, we meta-train $\text{FSML}_{\text{MAML}}$, FSML_{MTL}, and CRML with 5-way episodes to limit the extensive memory consumption.

Inference details. To construct meta-test episodes, we randomly sample N classes from $\mathcal{C}$. Among all instances in $\mathcal{C}$, we set a support set $\mathcal{S}$ as random NK instances, and a prediction set $\mathcal{P}$ as the rest to simulate a database in conventional metric learning. For *mini*DeepFashion, we randomly sample an attribute in the test attribute set, and then we compose an episode with $\mathcal{S}$ and $\mathcal{P}$ similarly as above. For (meta-)testing, we average model performances on 600 (meta-)test episodes for $N = 5$ and 50 for $N > 5$. For 20-way experiments on *mini*DeepFashion, we average model performances on 200 (meta-)test episodes. The class labels for each attribute are shown in Table S11. We use (meta-)test episode statistics for batch-normalization layers [3] during (meta-)testing.

[1] https://github.com/wyharveychen/CloserLookFewShot
[2] https://github.com/yaoyao-liu/meta-transfer-learning

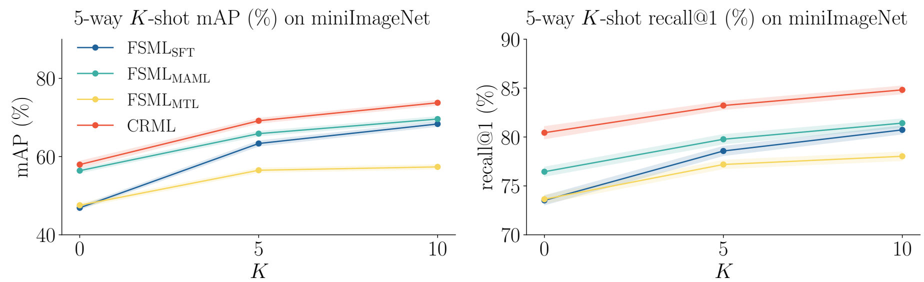

Fig. S1: 5-way K-shot mAP and recall@1 on *mini*ImageNet. Performances with $K = 0$ indicate performances before online adaptation.

Evaluation metrics for MPII Since the dataset is labeled with continuous real values, we employee two metrics defined on continuous labels following [7]: mean pose distance (mPD) and a modified version of normalized discounted cumulative gain (nDCG). The mPD_k evaluates the mean pose distance between a query and k nearest images. The modified nDCG_k evaluates the rank of the k nearest images and their relevance scores.

$$\text{nDCG}_k(q) = \frac{1}{Z_k} \sum_{i=1}^{k} \frac{2^{r_i}}{\log_2(i+1)}, \tag{1}$$

where Z_k is a normalization factor so that the maximum nDCG_k becomes 1 and $r_i = -\log_2(\|y_q - y_i\|_2 + 1)$ is the pose distance between the query q and the i^{th} nearest images. $\log_2(i + 1)$ is a discounting factor to give a higher weight to a higher rank.

2 Results of K-shot experiments

We conduct K-shot experiments with $K = \{0, 5, 10\}$ to see the effect of increasing number of target samples. Figure S1 shows the results of 5-way 5-shot and 10-shot on *mini*ImageNet when the multi-similarity loss is used. The performances with $K = 0$ indicate the performances from meta-learned initializations ($\text{FSML}_{\text{MAML}}$, FSML_{MTL}, and CRML) or the pre-trained model by conventional metric learning (FSML_{SFT}) before adaptation. All proposed four few-shot metric learning methods take advantage with increasing numbers of labels used. Note that CRML outperforms the three few-shot metric learning baselines on both image retrieval performance metrics including the performances before online adaptation. The comprehensive evaluation metrics with confidence intervals are provided in Table 1 and 2 In these tables, we refer to the cross-entropy loss and the multi-similarity loss [16] as CE and MS, respectively. “Pre-trained” refers to the pre-trained model trained by meta-training set of *mini*Imagenet using cross-entropy loss and evaluated the distance embedding right before the classifier.

Table 1: mAP (%) and Recall@K (%) on *mini*ImageNet.

Loss	Method	5-way 5-shot mAP	Recall@1	Recall@2	Recall@4	Recall@8
CE	Pre-trained	48.30 ± 0.51	72.85 ± 0.48	83.75 ± 0.32	91.10 ± 0.17	95.64 ± 0.08
	Baseline	53.14 ± 0.57	74.04 ± 0.49	84.19 ± 0.32	91.01 ± 0.17	95.27 ± 0.08
	Baseline++	65.22 ± 0.69	78.97 ± 0.49	86.88 ± 0.32	91.83 ± 0.18	95.03 ± 0.09
	MAML	57.03 ± 0.54	75.81 ± 0.47	86.16 ± 0.31	92.48 ± 0.17	96.10 ± 0.08
	MTL	47.93 ± 0.61	58.58 ± 0.62	74.01 ± 0.48	86.03 ± 0.29	93.45 ± 0.13
MS	DML	46.90 ± 0.45	73.52 ± 0.46	84.10 ± 0.30	91.24 ± 0.16	95.68 ± 0.08
	FSML_{SFT}	63.34 ± 0.66	78.57 ± 0.47	85.65 ± 0.30	91.77 ± 0.17	95.05 ± 0.09
	$\text{FSML}_{\text{MAML}}$	65.86 ± 0.60	79.77 ± 0.48	87.54 ± 0.27	92.24 ± 0.16	95.20 ± 0.09
	FSML_{MTL}	56.50 ± 0.55	77.19 ± 0.43	86.71 ± 0.27	92.61 ± 0.15	96.10 ± 0.07
	CRML	**69.15 ± 0.62**	**83.22 ± 0.41**	**89.91 ± 0.25**	**93.83 ± 0.14**	**96.24 ± 0.08**

Table 2: mAP (%) and Recall@K (%) on *mini*ImageNet.

Loss	Method	5-way 10-shot mAP	Recall@1	Recall@2	Recall@4	Recall@8
CE	Pre-trained	47.06 ± 0.47	72.32 ± 0.46	83.32 ± 0.30	90.88 ± 0.17	95.53 ± 0.08
	Baseline	56.57 ± 0.56	76.21 ± 0.45	85.50 ± 0.30	91.51 ± 0.16	95.34 ± 0.08
	Baseline++	71.51 ± 0.62	81.52 ± 0.43	88.26 ± 0.28	92.32 ± 0.17	94.97 ± 0.10
	MAML	57.15 ± 0.54	76.18 ± 0.47	86.31 ± 0.30	92.48 ± 0.17	96.08 ± 0.08
	MTL	52.96 ± 0.59	64.45 ± 0.60	78.34 ± 0.44	88.36 ± 0.25	94.42 ± 0.12
MS	DML	45.74 ± 0.40	73.08 ± 0.43	83.68 ± 0.29	90.95 ± 0.16	95.53 ± 0.08
	FSML_{SFT}	68.33 ± 0.60	80.73 ± 0.42	87.80 ± 0.27	92.17 ± 0.16	95.02 ± 0.09
	$\text{FSML}_{\text{MAML}}$	69.59 ± 0.61	81.42 ± 0.43	88.36 ± 0.27	92.51 ± 0.16	95.15 ± 0.09
	FSML_{MTL}	57.36 ± 0.56	78.03 ± 0.44	87.15 ± 0.27	92.72 ± 0.14	96.02 ± 0.07
	CRML	**73.75 ± 0.61**	**84.82 ± 0.40**	**90.67 ± 0.24**	**94.09 ± 0.13**	**98.25 ± 0.07**

3 Performance with confidence intervals

From Table S3 to S10, we provide all detailed quantitative results rounded up to the second digit after the decimal point as well as their 95% confidence intervals together which are not in the main paper. In these tables, we refer to the triplet loss [12], the lifted structured loss [13], and the multi-similarity loss as TR, LS, and MS, respectively. We refer to the conventional metric learning and few-shot metric learning as DML. We extensively verify that the few-shot metric learning improve mAP and Recall@K from deep metric learning models regardless of losses or datasets used as well as CRML outperforms the few-shot metric leaning baselines. Note that we empirically find that proxy-based losses [9, 4] are hardly optimized from scratch and overfit to a support set during online adaptation, resulting in poor performances on a prediction set. Thus, we cannot report the result with the state-of-the-art proxy-based losses.

4 Qualitative results

From Fig. S2 to S9, we compare the conventional deep metric learning (DML), $\text{FSML}_{\text{MAML}}$ as a representative few-shot metric learning baseline, and CRML by demonstrating each of their 8-nearest-neighbor retrieval results and t-SNE [8] visualization in a grid.

The results suggest that retrieval of DML relies on superficial aspects of images, *e.g.*, background and color only, since it does not adapt on target classes online. For example, given a query of a scoreboard in Fig. S2, DML fails to retrieve other scoreboards and instead get superficially similar images having similar grid visuals or the sunset background, whereas CRML successfully captures the context from lighting and play scores in scoreboards. Also, DML confuses image similarity with the majority of pixels in background colors or bird shapes, and $\mathrm{FSML}_{\mathrm{MAML}}$ retrieves visually similar birds of inaccurate species in Fig. S3. In contrast, CRML captures class-specific information by a few-shot adaptation and retrieves correct images. Especially, in *mini*DeepFashion examples (Fig. S4 and S5), DML is misled by similar colors or shapes without adapting to target attributes, only retrieving images of common patterns for category and texture. In contrast, CRML adapts to attribute-specific data, thus retrieving correct images.

Figures S8 and S9 show t-SNE visualization of the adapted embedding space on category and texture attributes, respectively. They show that FSML successfully learns attribute-specific embedding spaces.

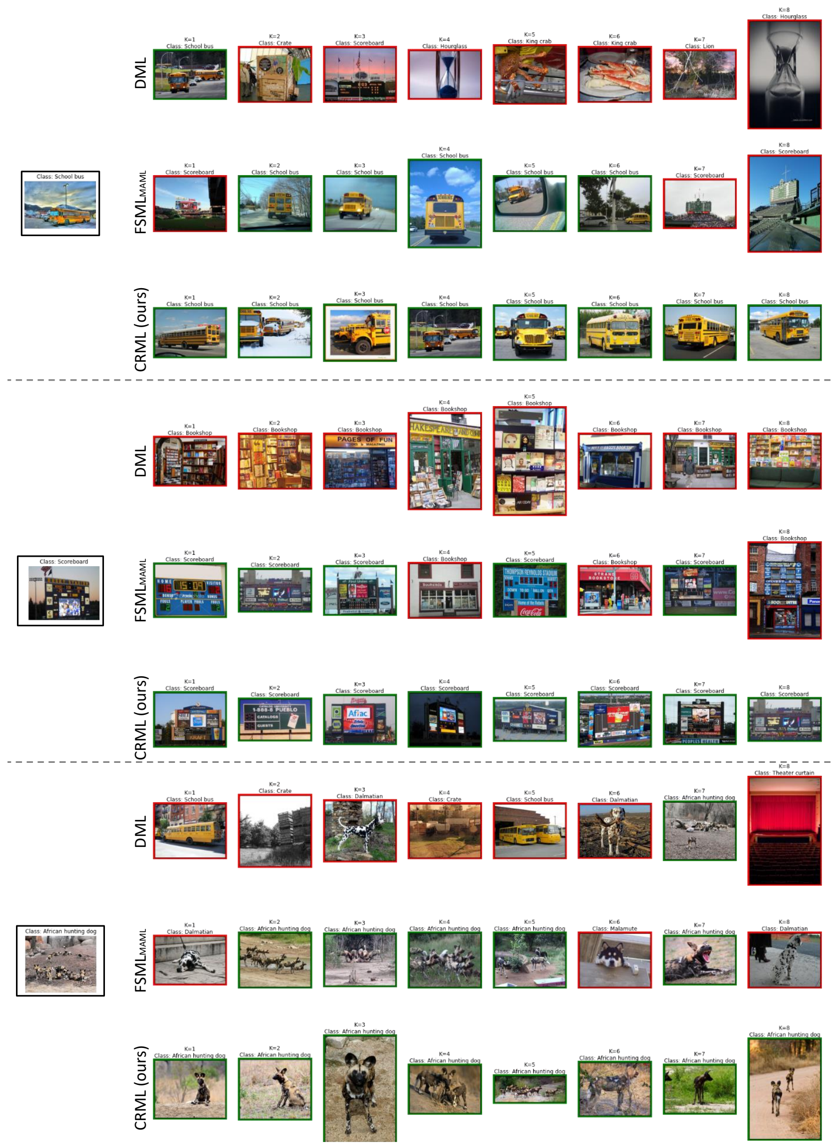

Fig. S2: Retrieval results on *mini*ImageNet [10]. The leftmost images are queries, and the right eight images are top-eight nearest neighbors. Red bounding boxes are negative images, and green boxes are positive. Best viewed in pdf.

Fig. S3: Retrieval results on CUB [15]. The leftmost images are queries, and the right eight images are top-eight nearest neighbors. Red bounding boxes are negative images, and green boxes are positive. Best viewed in pdf.

Fig. S4: Retrieval results on the texture attribute in *mini*DeepFashion. The leftmost images are queries, and the right eight images are top-eight nearest neighbors. Red bounding boxes are negative images, and green boxes are positive. Best viewed in pdf.

Fig. S5: Retrieval results on the texture attribute in *mini*DeepFashion. The leftmost images are queries, and the right eight images are top-eight nearest neighbors. Red bounding boxes are negative images, and green boxes are positive. Best viewed in pdf.

Fig. S6: t-SNE [8] visualization of the adapted embedding space on *mini*ImageNet [10].

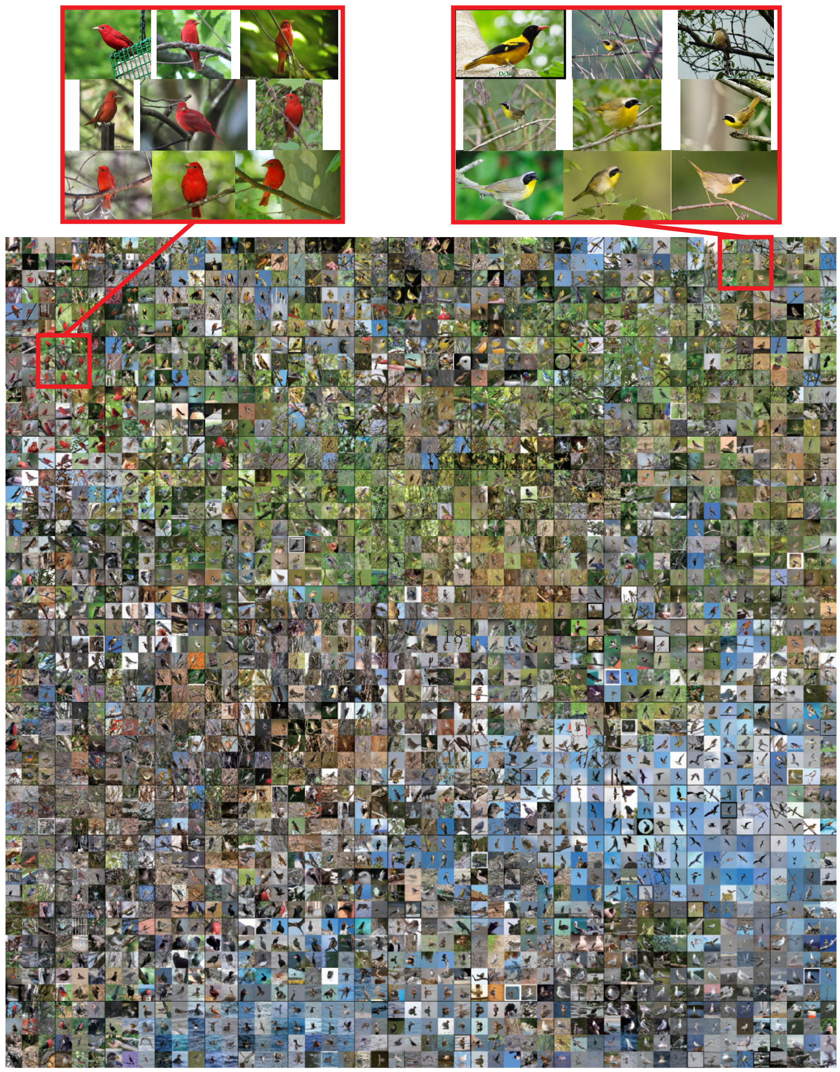

Fig. S7: t-SNE [8] visualization of the adapted embedding space on the cross-domain setting [2].

Fig. S8: t-SNE [8] visualization of the adapted embedding space on the category attribute in *mini*DeepFashion.

Fig. S9: t-SNE [8] visualization of the adapted embedding space on the texture attribute in *mini*DeepFashion.

Table S3: nDCG_1 (%) and mPD_1 on MPII [1]. For nDCG_1, the higher the better, and for mPD_1, the lower the better.

	#(pair label used)					
Metric	0	25	50	100	200	300
nDCG_1	40.45 ± 0.81	41.21 ± 1.20	41.41 ± 1.21	41.89 ± 1.22	42.44 ± 1.23	42.95 ± 1.21
mPD_1	31.54 ± 1.03	30.89 ± 1.16	30.80 ± 1.17	30.69 ± 1.24	30.27 ± 1.29	29.89 ± 1.35

Table S4: mAP (%) and Recall@K (%) on *mini*ImageNet [10].

		20-way 5-shot				
Loss	Method	mAP	Recall@1	Recall@2	Recall@4	Recall@8
TR	DML	21.48 ± 0.01	44.99 ± 0.09	59.38 ± 0.09	72.53 ± 0.08	82.89 ± 0.06
	$\mathrm{FSML}_{\mathrm{SFT}}$	25.66 ± 0.22	48.22 ± 0.18	62.10 ± 0.17	74.01 ± 0.15	83.36 ± 0.11
	$\mathrm{FSML}_{\mathrm{MAML}}$	26.07 ± 0.27	50.62 ± 0.30	64.71 ± 0.24	76.53 ± 0.18	85.43 ± 0.14
	$\mathrm{FSML}_{\mathrm{MTL}}$	24.17 ± 0.00	49.84 ± 0.01	64.03 ± 0.01	76.36 ± 0.02	85.41 ± 0.01
	CRML	29.86 ± 0.29	53.55 ± 0.25	66.98 ± 0.21	77.92 ± 0.14	**86.09 ± 0.10**
LS	DML	14.72 ± 0.01	45.19 ± 0.09	57.87 ± 0.09	69.50 ± 0.08	79.21 ± 0.07
	$\mathrm{FSML}_{\mathrm{SFT}}$	23.87 ± 0.27	50.30 ± 0.21	62.48 ± 0.20	72.82 ± 0.16	81.31 ± 0.12
	$\mathrm{FSML}_{\mathrm{MAML}}$	26.97 ± 0.28	52.79 ± 0.22	66.03 ± 0.19	76.81 ± 0.14	84.91 ± 0.12
	$\mathrm{FSML}_{\mathrm{MTL}}$	22.43 ± 0.01	52.87 ± 0.02	65.62 ± 0.02	76.10 ± 0.02	84.32 ± 0.02
	CRML	25.32 ± 0.34	50.69 ± 0.28	62.40 ± 0.26	72.56 ± 0.21	80.95 ± 0.17
MS	DML	21.01 ± 0.01	49.22 ± 0.09	62.57 ± 0.08	74.15 ± 0.07	83.42 ± 0.05
	$\mathrm{FSML}_{\mathrm{SFT}}$	29.88 ± 0.44	52.58 ± 0.31	65.07 ± 0.26	75.25 ± 0.19	83.15 ± 0.14
	$\mathrm{FSML}_{\mathrm{MAML}}$	28.56 ± 0.35	52.71 ± 0.31	65.61 ± 0.26	76.31 ± 0.19	84.45 ± 0.13
	$\mathrm{FSML}_{\mathrm{MTL}}$	24.03 ± 0.01	49.99 ± 0.03	64.64 ± 0.02	76.19 ± 0.02	85.10 ± 0.01
	CRML	**30.69 ± 0.35**	**56.27 ± 0.24**	**68.61 ± 0.20**	**78.48 ± 0.15**	85.93 ± 0.12

Table S5: mAP (%) and Recall@K (%) on CUB-200-2011 [15].

		50-way 5-shot				
Loss	Method	mAP	Recall@1	Recall@2	Recall@4	Recall@8
TR	DML	25.10 ± 0.05	45.20 ± 0.18	58.87 ± 0.14	70.66 ± 0.12	79.45 ± 0.15
	FSML_{SFT}	27.19 ± 0.17	47.89 ± 0.33	61.20 ± 0.28	72.26 ± 0.22	80.77 ± 0.20
	$\text{FSML}_{\text{MAML}}$	29.66 ± 0.14	50.99 ± 0.25	64.21 ± 0.23	75.08 ± 0.18	83.39 ± 0.19
	FSML_{MTL}	28.47 ± 0.05	49.65 ± 0.12	62.65 ± 0.13	73.45 ± 0.10	81.43 ± 0.08
	CRML	29.25 ± 0.17	49.74 ± 0.29	62.31 ± 0.26	72.84 ± 0.19	81.33 ± 0.18
LS	DML	22.28 ± 0.05	49.79 ± 0.16	63.81 ± 0.16	75.76 ± 0.13	84.83 ± 0.12
	FSML_{SFT}	28.08 ± 0.21	53.29 ± 0.33	66.20 ± 0.28	**76.66 ± 0.26**	**84.58 ± 0.21**
	$\text{FSML}_{\text{MAML}}$	27.06 ± 0.21	48.38 ± 0.32	60.72 ± 0.28	71.55 ± 0.26	80.36 ± 0.24
	FSML_{MTL}	28.20 ± 0.05	52.07 ± 0.13	64.83 ± 0.11	75.98 ± 0.09	83.06 ± 0.09
	CRML	30.85 ± 0.25	53.34 ± 0.38	65.26 ± 0.28	75.09 ± 0.23	82.72 ± 0.18
MS	DML	26.26 ± 0.06	51.76 ± 0.16	62.48 ± 0.15	72.23 ± 0.13	80.25 ± 0.12
	FSML_{SFT}	31.14 ± 0.20	55.34 ± 0.22	66.18 ± 0.23	74.84 ± 0.19	81.95 ± 0.18
	$\text{FSML}_{\text{MAML}}$	33.17 ± 0.21	54.56 ± 0.35	66.27 ± 0.32	75.72 ± 0.23	82.94 ± 0.19
	FSML_{MTL}	30.18 ± 0.05	54.29 ± 0.10	65.16 ± 0.12	73.59 ± 0.10	81.55 ± 0.08
	CRML	**33.87 ± 0.28**	**58.08 ± 0.31**	**68.42 ± 0.27**	76.50 ± 0.23	83.09 ± 0.17

Table S6: mAP (%) and Recall@K (%) on the cross-domain setting [2].

		5-way 5-shot				
Loss	Method	mAP	Recall@1	Recall@2	Recall@4	Recall@8
TR	DML	31.07 ± 0.27	48.46 ± 0.58	65.91 ± 0.51	81.80 ± 0.34	93.01 ± 0.17
	FSML_{SFT}	37.75 ± 0.44	54.56 ± 0.66	70.42 ± 0.53	83.87 ± 0.34	93.21 ± 0.16
	$\text{FSML}_{\text{MAML}}$	46.80 ± 0.56	62.68 ± 0.67	76.79 ± 0.49	87.31 ± 0.30	94.03 ± 0.15
	FSML_{MTL}	39.38 ± 0.48	60.52 ± 0.60	75.32 ± 0.47	86.80 ± 0.29	94.18 ± 0.14
	CRML	50.35 ± 0.72	65.44 ± 0.76	78.01 ± 0.54	87.31 ± 0.32	93.56 ± 0.17
LS	DML	31.80 ± 0.25	55.16 ± 0.59	71.57 ± 0.50	85.52 ± 0.32	94.77 ± 0.15
	FSML_{SFT}	43.20 ± 0.56	62.28 ± 0.68	76.03 ± 0.52	86.83 ± 0.30	94.01 ± 0.15
	$\text{FSML}_{\text{MAML}}$	45.35 ± 0.63	64.42 ± 0.69	78.01 ± 0.50	88.18 ± 0.29	94.73 ± 0.15
	FSML_{MTL}	39.77 ± 0.49	61.41 ± 0.59	75.19 ± 0.45	86.05 ± 0.28	93.29 ± 0.15
	CRML	**58.91 ± 0.83**	**72.19 ± 0.74**	81.51 ± 0.50	88.06 ± 0.30	92.68 ± 0.17
MS	DML	36.15 ± 0.32	57.48 ± 0.61	73.23 ± 0.50	86.13 ± 0.32	94.87 ± 0.15
	FSML_{SFT}	49.44 ± 0.61	65.21 ± 0.67	77.10 ± 0.49	85.97 ± 0.30	92.23 ± 0.16
	$\text{FSML}_{\text{MAML}}$	51.48 ± 0.67	66.95 ± 0.70	78.77 ± 0.50	87.22 ± 0.30	93.12 ± 0.17
	FSML_{MTL}	40.29 ± 0.45	62.72 ± 0.61	76.98 ± 0.45	87.75 ± 0.27	94.67 ± 0.13
	CRML	56.37 ± 0.71	71.03 ± 0.69	**81.53 ± 0.46**	**88.92 ± 0.27**	**93.75 ± 0.16**

Table S7: mAP (%) and Recall@K (%) on the cross-domain setting [2].

		50-way 5-shot				
Loss	Method	mAP	Recall@1	Recall@2	Recall@4	Recall@8
TR	DML	4.42 ± 0.01	12.53 ± 0.12	20.43 ± 0.16	30.88 ± 0.18	44.58 ± 0.20
	FSML$_{\text{SFT}}$	6.53 ± 0.06	15.75 ± 0.24	24.94 ± 0.25	36.90 ± 0.30	50.83 ± 0.28
	FSML$_{\text{MAML}}$	9.12 ± 0.09	21.35 ± 0.23	32.56 ± 0.28	46.11 ± 0.30	60.76 ± 0.28
	FSML$_{\text{MTL}}$	6.37 ± 0.01	17.89 ± 0.12	28.67 ± 0.11	41.14 ± 0.12	55.03 ± 0.01
	CRML	**10.96 ± 0.22**	24.05 ± 0.41	35.49 ± 0.47	48.76 ± 0.47	62.45 ± 0.44
LS	DML	5.19 ± 0.01	18.49 ± 0.15	27.34 ± 0.18	38.90 ± 0.19	52.86 ± 0.17
	FSML$_{\text{SFT}}$	6.91 ± 0.09	20.76 ± 0.25	30.99 ± 0.28	43.53 ± 0.28	57.77 ± 0.30
	FSML$_{\text{MAML}}$	7.14 ± 0.06	21.23 ± 0.20	31.90 ± 0.29	44.66 ± 0.31	59.00 ± 0.32
	FSML$_{\text{MTL}}$	6.66 ± 0.05	19.81 ± 0.10	30.27 ± 0.14	42.45 ± 0.11	56.85 ± 0.12
	CRML	8.13 ± 0.18	21.24 ± 0.39	31.34 ± 0.46	43.65 ± 0.45	57.27 ± 0.44
MS	DML	6.05 ± 0.02	19.58 ± 0.15	29.04 ± 0.15	41.08 ± 0.17	54.68 ± 0.17
	FSML$_{\text{SFT}}$	9.76 ± 0.12	24.06 ± 0.28	35.15 ± 0.28	47.79 ± 0.31	61.00 ± 0.28
	FSML$_{\text{MAML}}$	9.99 ± 0.10	23.66 ± 0.23	35.40 ± 0.26	48.84 ± 0.27	62.83 ± 0.27
	FSML$_{\text{MTL}}$	6.77 ± 0.01	20.00 ± 0.09	29.91 ± 0.10	42.76 ± 0.09	57.60 ± 0.11
	CRML	10.87 ± 0.18	**27.01 ± 0.41**	**38.30 ± 0.46**	**51.08 ± 0.49**	**64.30 ± 0.47**

Table S8: mAP (%) and Recall@K (%) on *mini*DeepFashion.

		5-way 5-shot				
Loss	Method	mAP	Recall@1	Recall@2	Recall@4	Recall@8
TR	DML	34.66 ± 0.66	48.19 ± 0.90	64.43 ± 0.77	79.19 ± 0.48	90.13 ± 0.20
	FSML$_{\text{SFT}}$	36.53 ± 0.77	48.67 ± 0.94	64.68 ± 0.80	79.19 ± 0.50	90.00 ± 0.20
	FSML$_{\text{MAML}}$	38.09 ± 0.87	50.22 ± 1.00	65.87 ± 0.84	79.80 ± 0.52	90.17 ± 0.21
	FSML$_{\text{MTL}}$	36.16 ± 0.78	50.38 ± 0.98	65.95 ± 0.83	79.86 ± 0.51	90.27 ± 0.21
	CRML	38.11 ± 0.84	50.48 ± 0.99	66.19 ± 0.82	80.12 ± 0.50	90.40 ± 0.21
LS	DML	34.93 ± 0.66	53.81 ± 0.84	68.76 ± 0.71	82.16 ± 0.45	92.17 ± 0.21
	FSML$_{\text{SFT}}$	38.77 ± 0.87	54.40 ± 0.90	68.86 ± 0.73	81.53 ± 0.45	91.11 ± 0.19
	FSML$_{\text{MAML}}$	**40.17 ± 0.97**	54.76 ± 1.01	69.08 ± 0.80	81.46 ± 0.48	90.75 ± 0.19
	FSML$_{\text{MTL}}$	37.76 ± 0.84	**55.69 ± 0.92**	**69.69 ± 0.73**	**81.96 ± 0.46**	91.20 ± 0.19
	CRML	39.91 ± 0.92	55.11 ± 0.95	69.17 ± 0.76	81.45 ± 0.47	90.70 ± 0.20
MS	DML	31.77 ± 0.38	50.26 ± 0.56	65.79 ± 0.48	80.20 ± 0.32	**91.32 ± 0.14**
	FSML$_{\text{SFT}}$	35.20 ± 0.52	51.27 ± 0.60	66.08 ± 0.50	79.78 ± 0.31	90.40 ± 0.13
	FSML$_{\text{MAML}}$	38.22 ± 0.89	53.45 ± 0.95	67.56 ± 0.77	80.19 ± 0.46	89.57 ± 0.19
	FSML$_{\text{MTL}}$	35.21 ± 0.76	52.23 ± 0.83	66.78 ± 0.68	79.92 ± 0.44	90.16 ± 0.18
	CRML	38.25 ± 0.84	50.70 ± 0.99	66.29 ± 0.82	80.18 ± 0.50	90.43 ± 0.21

Table S9: mAP (%) and Recall@K (%) on *mini*DeepFashion.

Loss	Method	mAP	Recall@1	Recall@2	Recall@4	Recall@8
				20-way 5-shot		
TR	DML	12.09 ± 0.31	22.85 ± 0.54	34.69 ± 0.72	48.76 ± 0.79	63.70 ± 0.69
	FSML_{SFT}	12.78 ± 0.36	23.05 ± 0.57	34.82 ± 0.75	48.86 ± 0.83	63.85 ± 0.73
	$\text{FSML}_{\text{MAML}}$	12.69 ± 0.38	22.65 ± 0.58	34.10 ± 0.77	48.02 ± 0.85	63.21 ± 0.76
	FSML_{MTL}	13.38 ± 0.64	24.90 ± 1.09	36.93 ± 1.38	50.78 ± 1.46	65.31 ± 1.26
	CRML	12.93 ± 0.63	23.72 ± 1.00	35.63 ± 1.30	49.75 ± 1.42	64.65 ± 1.24
LS	DML	12.72 ± 0.38	28.00 ± 0.55	39.51 ± 0.69	52.69 ± 0.72	66.60 ± 0.62
	FSML_{SFT}	13.74 ± 0.43	28.27 ± 0.58	39.69 ± 0.71	52.63 ± 0.73	66.16 ± 0.61
	$\text{FSML}_{\text{MAML}}$	**14.75 ± 0.49**	28.69 ± 0.69	40.42 ± 0.82	53.53 ± 0.83	66.99 ± 0.68
	FSML_{MTL}	14.13 ± 0.79	29.39 ± 1.05	40.73 ± 1.26	53.55 ± 1.30	67.04 ± 1.11
	CRML	14.58 ± 0.79	**30.03 ± 1.07**	**41.75 ± 1.30**	**54.75 ± 1.33**	**68.07 ± 1.13**
MS	DML	11.34 ± 0.31	26.10 ± 0.46	37.26 ± 0.59	50.19 ± 0.63	64.16 ± 0.56
	FSML_{SFT}	12.46 ± 0.38	26.42 ± 0.49	37.77 ± 0.62	50.41 ± 0.66	64.31 ± 0.58
	$\text{FSML}_{\text{MAML}}$	13.30 ± 0.42	27.72 ± 0.54	39.12 ± 0.69	52.09 ± 0.74	65.59 ± 0.65
	FSML_{MTL}	12.27 ± 0.66	26.83 ± 0.81	38.01 ± 1.02	50.62 ± 1.08	64.20 ± 0.94
	CRML	12.99 ± 0.66	27.75 ± 0.87	39.13 ± 1.16	52.03 ± 1.16	65.61 ± 1.02

Table S10: mAP (%) and Recall@K (%) on CUB-200-2011 [15].

Loss	Method	mAP	Recall@1	Recall@2	Recall@4	Recall@8
				5-way 5-shot		
CE	DML	64.72 ± 0.69	81.41 ± 0.55	88.63 ± 0.36	93.45 ± 0.21	96.58 ± 0.12
	FSML_{SFT}	68.14 ± 0.70	81.60 ± 0.56	88.47 ± 0.36	92.83 ± 0.21	95.74 ± 0.13
	$\text{FSML}_{\text{SFT++}}$	79.23 ± 0.62	86.75 ± 0.46	91.34 ± 0.29	93.94 ± 0.20	95.67 ± 0.14
	$\text{FSML}_{\text{MAML}}$	69.73 ± 0.68	82.37 ± 0.54	89.63 ± 0.34	94.02 ± 0.20	96.76 ± 0.11
	FSML_{MTL}	71.85 ± 0.72	80.15 ± 0.63	88.09 ± 0.38	92.42 ± 0.22	94.97 ± 0.15
MS	DML	57.75 ± 0.55	81.80 ± 0.47	89.59 ± 0.30	94.63 ± 0.17	**97.69 ± 0.09**
	FSML_{SFT}	79.94 ± 0.58	87.68 ± 0.40	91.58 ± 0.26	93.84 ± 0.19	95.41 ± 0.14
	$\text{FSML}_{\text{MAML}}$	81.98 ± 0.57	89.50 ± 0.40	93.22 ± 0.25	95.37 ± 0.16	96.78 ± 0.11
	FSML_{MTL}	71.93 ± 0.61	86.24 ± 0.40	91.40 ± 0.25	94.54 ± 0.15	96.69 ± 0.10
	CRML	**82.65 ± 0.53**	**89.99 ± 0.36**	**93.48 ± 0.23**	**95.52 ± 0.15**	96.87 ± 0.11

Table S11: Class configurations for each attribute in *mini*DeepFashion.

split	attribute (# classes)	classes
$\mathcal{C}^{\mathrm{mtr}}$	fabric (99)	beaded, brocade, burnout, cable, cable-knit, canvas, chambray, chiffon, chino, chunky, clean, corduroy, cotton, crepe, crochet, crocheted, damask, denim, distressed, dye, embellished, embroidered, embroidery, eyelet, faded, faux, feather, french, fur, fuzzy, gauze, georgette, gingham, glitter, heathered, jacquard, knit, lace, lace-paneled, layered, leather, loose, mesh, mesh-paneled, metallic, neon, neoprene, nets, nylon, organza, panel, paneled, patchwork, pintuck, pintucked, plaid, pleat, pleated, pointelle, ponte, quilted, rib, ribbed, ripped, ruched, ruffle, ruffled, sateen, satin, scuba, seamless, seersucker, semi-sheer, sequin, sequined, sheer, shirred, sleek, slub, slub-knit, stone, stones, stretch, stretch-knit, studded, suede, tartan, terry, textured, tie-dye, tiered, tile, tulle, tweed, twill, velvet, wash, washed, woven
$\mathcal{C}^{\mathrm{mtr}}$	part (95)	arrow collar, back cutout, backless, batwing, bell, belted, boat neck, bow, braided, button, button-front, buttoned, cap-sleeve, collar, collared, collarless, contrast, crew, crisscross, crochet-trimmed, crossback, cuffed, cutout-back, dolman, dolphin, double-breasted, draped, drawstring, drop waist, elephant, flat, flat-front, flounce, flounced, fluted, flutter, fringe, fringed, gathered waistline, hem, high-neck, high-slit, high-waist, high-waisted, hood, hooded, illusion, keyhole, knit open, knotted, lace-trim, lace-trimmed, lace-up, lapel, long-sleeve, long-sleeved, neckline, off-the-shoulder, open-back, open-front, open-knit, peplum, pin, pocket, racer-back, raglan, scallop, scalloped, scoop, self-tie, shawl, shoulder, side-slit, sleeve, sleeveless, slit, split, strap, strapless, strappy, surplice, tassel, tie-front, trim, trimmed, turtle-neck, twisted, two-button, v-back, v-neck, zip, zip-front, zip-up, zipper, zippered
$\mathcal{C}^{\mathrm{mtr}}$	style (127)	angeles, art, athletic, babydoll, barbie, baseball, basic, beach, beatles, bella, biker, boho, bold, boyfriend, brooklyn, california, candy, cat, chic, city, classic, coffee, cute, dainty, dark, darling, desert, destroyed, doll, dream, eagle, edge, elegant, everyday, fancy, field, fisherman, flirty, fox, fresh, galaxy, garden, girl, girls, grunge, inset, island, isle, joie, kiss, la, lady, laser, life, light, logo, lounge, love, luxe, mandarin, mob, mod, moon, night, ny, nyc, oxford, pan, paris, party, performance, pineapple, pink, pj, please, posh, power, raga, rainbow, red, relaxed, retro, reverse, reversible, roll, rose, rugby, run, running, sea, shopping, sky, smart, snap, snoopy, soft, solid, sporty, standout, star, stars, studio, summer, sun, sunflower, sweet, sweetheart, swim, swiss, thermal, tokyo, track, tree, trench, triangle, tropical, tupac, utility, varsity, wave, weekend, west, wild, workout, yoga, yoke, youth
$\mathcal{C}^{\mathrm{mvl}}$	shape (76)	a-line, ankle, asymmetric, asymmetrical, baja, bandage, bermuda, bib, big, bodycon, box, boxy, bustier, caged, cami, capri, cargo, chiffon-paneled, combo, cover-up, cozy, crochet-paneled, crop, cropped, cut, cutoff, cutout, drapey, fit, fitted, flare, flared, flowy, foldover, harem, high-low, longline, maxi, medium, midi, mini, moto, muscle, oversized, peasant, pencil, polo, popover, puffer, pullover, raw, rise, round, sheath, shift, shirt, skater, skinny, skort, slim, slip, slouchy, smock, smocked, square, swing, trapeze, trouser, tube, tulip, tunic, vertical, wide-leg, windbreaker, windowpane, wrap
$\mathcal{C}$	category (37)	Blazer, Blouse, Bomber, Button-Down, Cardigan, Flannel, Halter, Hoodie, Jacket, Jersey, Parka, Peacoat, Sweater, Tank, Tee, Top, Turtleneck, Capris, Culottes, Cutoffs, Gauchos, Jeggings, Jodhpurs, Leggings, Sarong, Skirt, Sweatpants, Sweatshorts, Trunks, Caftan, Coverup, Jumpsuit, Kaftan, Kimono, Nightdress, Romper, Shirtdress
$\mathcal{C}$	texture (57)	abstract, animal, bandana, baroque, bird, butterfly, camo, camouflage, checked, checkered, chevron, circle, colorblock, colorblocked, daisy, diamond, dot, dots, dotted, floral, floral-embroidered, flower, geo, graphic, grid, heart, houndstooth, ikat, leaf, leopard, linen, linen-blend, marled, mixed, ombre, ornate, paint, paisley, pattern, patterned, pinstripe, print, printed, ringer, sophisticated, southwestern, speckled, spotted, stripe, striped, stripes, structured, tribal, two-tone, varsity-striped, watercolor, zigzag

References

1. Andriluka, M., Pishchulin, L., Gehler, P., Schiele, B.: 2d human pose estimation: New benchmark and state of the art analysis. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2014)
2. Chen, W.Y., Liu, Y.C., Kira, Z., Wang, Y.C., Huang, J.B.: A closer look at few-shot classification. In: International Conference on Learning Representations (ICLR) (2019)
3. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning (ICML) (2015)
4. Kim, S., Kim, D., Cho, M., Kwak, S.: Proxy anchor loss for deep metric learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3238–3247 (2020)
5. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. International Conference on Learning Representation (ICLR) (2015)
6. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems (NeurIPS) (2012)
7. Kwak, S., Cho, M., Laptev, I.: Thin-slicing for pose: Learning to understand pose without explicit pose estimation. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
8. Maaten, L.v.d., Hinton, G.: Visualizing data using t-sne. Journal of Machine Learning Research (2008)
9. Movshovitz-Attias, Y., Toshev, A., Leung, T.K., Ioffe, S., Singh, S.: No fuss distance metric learning using proxies. In: The IEEE International Conference on Computer Vision (ICCV) (2017)
10. Ravi, S., Larochelle, H.: Optimization as a model for few-shot learning. In: International Conference on Learning Representations (ICLR) (2017)
11. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al.: Imagenet large scale visual recognition challenge. International Journal of Computer Vision (IJCV) **115**(3), 211–252 (2015)
12. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015)
13. Song, H.O., Xiang, Y., Jegelka, S., Savarese, S.: Deep metric learning via lifted structured feature embedding. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
14. Sun, Q., Liu, Y., Chua, T.S., Schiele, B.: Meta-transfer learning for few-shot learning. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
15. Wah, C., Branson, S., Welinder, P., Perona, P., Belongie, S.: The caltech-ucsd birds-200-2011 dataset (2011)
16. Wang, X., Han, X., Huang, W., Dong, D., Scott, M.R.: Multi-similarity loss with general pair weighting for deep metric learning. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019)