Appendix

A Experiment Settings

A.1 Training Settings

General setups for DP (dataset pruning). Regardless of the choice of the downstream tasks and DP methods, each DP process will go through a three-stage paradigm, namely ❶ source dataset pruning, ❷ pretraining based on the pruned source dataset, and ❸ finetuning on the target dataset using either LP (linear probing) or FF (full finetuning). Regarding baseline methods for source dataset pruning, we strictly follow the baselines' settings provided in the literature. As stated before, we utilize a small surrogate model (ResNet-18) pretrained on the full ImageNet to conduct DP. The proportion of pruned class numbers to the total source class numbers determines pruning ratios for our methods: LM (label mapping) and FM (feature mapping). To ensure a fair comparison with other non-class-wise baselines (RANDOM, GRAND, MODERATE), the same pruning ratio is applied to the total number of training data points.

Pretraining setups. We keep our pretraining procedure consistent for all models and strictly follow the settings released by the existing work [39]. All the models are trained from scratch using Stochastic Gradient Descent (SGD) to minimize the standard cross-entropy loss in a multi-class classification setting. We use a batch size of 1024, momentum of 0.9, and a weight decay of 5×10^{-4}. The training utilizes a cyclic learning rate schedule, which begins with an initial learning rate of 0.5 and peaks at the second epoch. During training, we incorporate data augmentations such as random resized cropping and random horizontal flipping.

Downstream task finetuning settings. Our approach to finetuning the pretrained model on downstream tasks involves using LP and FF. Details of the downstream datasets and the training configurations are presented in Tab. A1, following [96]. For LP, we employ the Adam optimizer, a multi-step decaying scheduler, and an initial learning rate of 0.1 across 50 total training epochs. As for FF, we utilize the Adam optimizer over 200 epochs with a cosine-annealing scheduler, an initial learning rate of 0.01, and a weight decay of 5×10^{-4}. All finetuning experiments employ a batch size of 256 and standard data augmentations, such as random resized cropping and horizontal flipping.

Training settings for SSL (self-supervised learning) and ViT. Our SSL training settings follow the configurations provided by MOCOV2. Details of the pretraining and finetuning stages can be accessed at https://github.com/facebookresearch/moco. For the training of ViTs, we rely on the setting released in the original ViT paper [108] (see ViT/B in Table 3).

Table A1: Dataset attributes and training configurations of 8 downstream image classification datasets considered in this work.

Dataset	Train Size	Test Size	Class Number	Batch Size	Rescaled Resolution
Flowers102	4093	2463	102	256	224×224
DTD	2820	1692	47	256	224×224
UCF101	7639	3783	101	256	224×224
Food101	50500	30300	101	256	224×224
OxfordPets	2944	3669	37	256	224×224
StanfordCars	6509	8041	196	256	224×224
SUN397	15888	19850	397	256	224×224
CIFAR10	50000	10000	10	256	160×160

B Additional Results

Expanded performance evaluation of in-domain DP methods across all downstream datasets. **Fig. A1** expands on the performance comparisons in **Fig. 2**, providing a more thorough evaluation of various in-domain DP methods on all eight downstream datasets. The trends observed are consistent with those in **Fig. 2**: Random pruning shows a strong baseline method in DP for transfer learning compared to other state-of-the-art DP methods designed for non-transfer learning. This observation prompts us to explore more effective dataset pruning strategies for transfer learning. MODERATE and

GRAND are also demonstrating strong baselines, motivating us to choose them as the default DP baselines in Section 5.

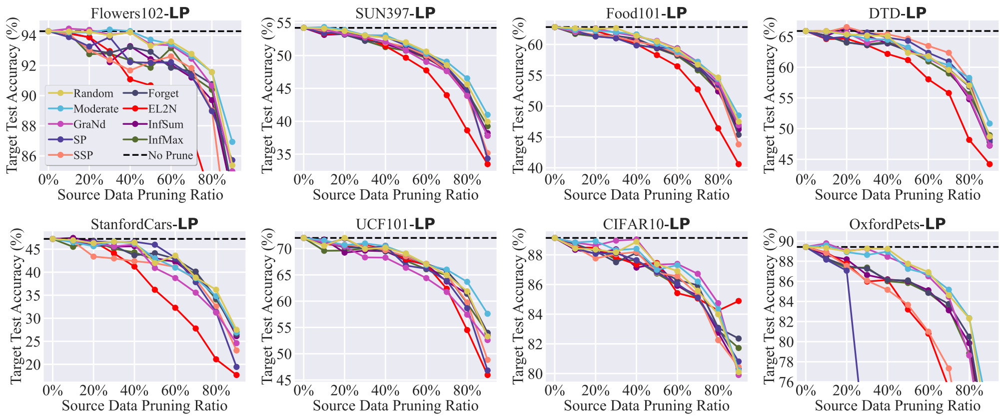


Figure A1: Extended performance comparison of different in-domain DP methods for transfer learning across all downstream tasks. Experiment settings are consistent with Fig. 4.

The main results reported in Fig. 4 are stable: analysis with detailed numerical results and standard deviations. In **Tab. A2**, we provide the exact numerical results used to generate **Fig. 4**. These numbers give a more granular view of the performance comparisons. The results, calculated over three independent trials, show that the magnitude of the standard deviations is quite small relative to the mean values. This indicates that the trends and conclusions drawn from **Fig. 4** are generally valid and not significantly affected by trial variations. To maintain readability and clarity, we choose not to include these minor standard deviations in **Fig. 4**.

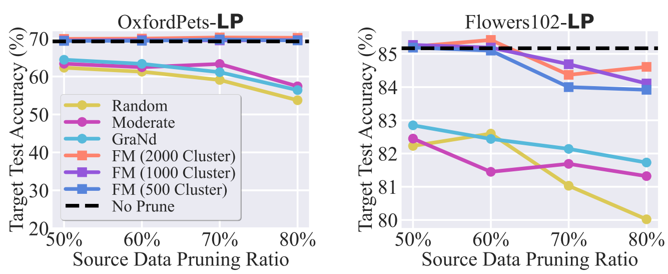


Figure A2: Ablation study on the sensitivity of FM against the choice of cluster numbers. For FM, the original dataset is pre-clustered into K clusters, with $K = 2000$ as the default setting in this study. Here, various K values (500, 1000, and 2000) are studied in this experiment. Other settings are aligned with Tab. 2.

FM performance is robust to the choice of cluster number. The results presented in **Fig. A2** provide insights into the sensitivity of FM performance to variations of the cluster number (K). This extended study from Tab. 2 shows that FM performance remains robust as the cluster number varies from 500 to 2000, preserving the benefits of DP for transfer learning without much performance degradation. This implies that FM can provide reliable and stable results even when the hyperparameter settings are not optimal.

LM identifies the most influential source classes. To validate that the classes with the highest LM scores are the most influential, we present the pruning trajectory in **Fig. A3**, where pruning is executed in the reverse order of the class scores, different from the proposed DP implementation. That is, classes with the smallest scores are retained, while the ones with the highest scores are pruned. Remarkably, even a slight pruning ratio (such as 10%) leads to a significant degradation in downstream performance. This evidence underscores the benefits of source classes with high LM scores in promoting the downstream performance.

Subsets obtained by LM improve the flatness of loss landscape on downstream tasks. **Fig. A4** evaluates the flatness of the loss landscape of the model pretrained on an 80%-pruned source dataset

Table A2: Exact numbers and standard deviations for Fig. 4.

Method	Pruning Ratio 5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%
OxfordPets - LP																	
RANDOM	89.81±0.22	89.32±0.18	89.92±0.12	88.99±0.34	89.26±0.13	89.26±0.11	88.66±0.38	89.23±0.28	88.23±0.07	87.76±0.05	87.08±0.27	86.92±0.23	86.10±0.03	84.63±0.31	82.56±0.09	82.31±0.2	76.59±0.35
MODERATE	89.67±0.36	89.59±0.29	89.86±0.15	88.91±0.14	88.96±0.10	88.63±0.25	88.93±0.32	88.96±0.05	87.79±0.37	87.24±0.17	88.01±0.28	86.73±0.02	86.59±0.04	85.17±0.08	83.84±0.04	82.34±0.33	78.71±0.27
GRAND	89.70±0.20	89.78±0.16	89.78±0.24	89.04±0.38	89.40±0.11	89.23±0.07	88.72±0.04	88.44±0.39	88.93±0.28	87.63±0.20	86.86±0.29	86.51±0.17	84.85±0.02	84.46±0.14	81.36±0.33	78.63±0.13	74.19±0.08
LM	89.72±0.1	89.34±0.04	90.27±0.37	90.11±0.27	89.97±0.38	90.02±0.19	89.72±0.15	89.62±0.24	89.46±0.16	89.45±0.06	89.92±0.11	90.02±0.27	89.89±0.18	90.19±0.14	90.16±0.38	90.11±0.03	88.85±0.37
FM	90.35±0.32	89.78±0.07	89.70±0.24	89.45±0.29	90.02±0.22	89.83±0.37	89.89±0.03	89.72±0.27	89.56±0.06	89.72±0.19	89.94±0.11	90.08±0.36	90.62±0.03	89.78±0.10	90.32±0.38	89.64±0.15	90.16±0.07
SUN397 - LP																	
RANDOM	51.25±0.14	51.23±0.37	51.23±0.27	51.16±0.32	51.06±0.21	50.88±0.08	51.11±0.34	51.32±0.28	50.80±0.19	50.31±0.27	50.90±0.36	50.33±0.17	50.41±0.07	49.88±0.24	49.11±0.10	49.45±0.14	48.63±0.31
MODERATE	51.34±0.10	51.17±0.35	51.36±0.38	51.73±0.07	51.45±0.20	51.48±0.36	51.04±0.12	51.16±0.15	50.68±0.14	50.70±0.21	50.87±0.27	50.06±0.36	50.34±0.38	49.83±0.12	50.16±0.34	49.51±0.25	49.00±0.20
GRAND	51.38±0.30	51.25±0.24	51.15±0.33	51.42±0.14	51.44±0.38	51.20±0.24	51.17±0.13	50.63±0.32	50.68±0.29	50.28±0.36	49.81±0.38	50.28±0.10	50.02±0.28	49.84±0.12	48.87±0.21	48.92±0.11	47.95±0.27
LM	51.70±0.18	51.78±0.26	51.76±0.20	51.67±0.35	51.52±0.32	51.55±0.28	51.87±0.30	51.40±0.16	51.46±0.32	51.62±0.14	51.48±0.23	51.39±0.22	51.65±0.18	50.95±0.33	49.62±0.16	49.53±0.21	48.93±0.27
FM	51.88±0.26	51.72±0.31	51.68±0.18	51.50±0.14	51.70±0.37	51.79±0.30	51.74±0.24	51.63±0.38	51.54±0.33	51.61±0.28	51.72±0.27	51.42±0.32	51.57±0.23	51.08±0.22	49.72±0.18	49.73±0.21	48.97±0.35
Food101 - LP																	
RANDOM	62.22±0.30	62.64±0.21	62.32±0.18	62.53±0.33	62.23±0.17	62.65±0.27	61.33±0.31	61.42±0.39	60.91±0.26	60.50±0.21	60.19±0.34	59.20±0.20	58.15±0.32	56.74±0.27	54.70±0.19	54.64±0.24	51.12±0.22
MODERATE	62.47±0.19	62.34±0.28	62.49±0.27	62.22±0.16	62.45±0.36	62.00±0.24	61.76±0.30	61.63±0.29	60.75±0.16	60.13±0.26	59.42±0.24	58.78±0.35	58.80±0.30	57.19±0.25	55.73±0.37	54.43±0.24	51.79±0.16
GRAND	62.70±0.18	62.10±0.24	62.30±0.33	62.61±0.22	62.25±0.27	61.90±0.24	62.27±0.31	61.32±0.34	61.60±0.22	60.58±0.24	59.72±0.25	59.32±0.33	58.19±0.27	57.17±0.21	55.54±0.19	53.84±0.30	50.79±0.28
LM	63.17±0.16	63.39±0.27	63.60±0.18	63.55±0.32	63.30±0.16	63.50±0.33	63.97±0.23	63.08±0.29	63.34±0.25	63.47±0.33	63.09±0.24	63.05±0.36	62.78±0.26	62.14±0.30	61.51±0.33	61.15±0.22	58.78±0.30
FM	62.89±0.28	63.29±0.24	63.55±0.16	63.53±0.37	63.26±0.30	63.34±0.27	63.20±0.35	63.24±0.22	62.73±0.32	62.84±0.24	63.06±0.23	62.55±0.16	61.82±0.35	61.71±0.32	61.38±0.24	60.50±0.36	58.24±0.32
DTD - LP																	
RANDOM	66.13±0.20	65.90±0.17	66.84±0.21	65.37±0.23	65.07±0.30	65.19±0.19	64.42±0.22	65.08±0.18	63.30±0.25	62.35±0.32	61.95±0.29	61.82±0.24	62.00±0.28	59.57±0.31	58.69±0.34	57.03±0.33	52.54±0.26
MODERATE	66.73±0.18	65.84±0.30	65.72±0.16	65.31±0.33	65.60±0.22	64.66±0.27	65.25±0.16	64.48±0.35	63.00±0.22	63.24±0.36	63.00±0.33	61.35±0.24	60.87±0.25	60.34±0.16	58.92±0.34	58.27±0.28	54.26±0.24
GRAND	66.67±0.24	65.90±0.22	65.96±0.17	65.66±0.26	65.13±0.27	64.01±0.35	64.42±0.16	64.66±0.22	63.30±0.28	62.65±0.16	61.47±0.23	61.41±0.29	60.40±0.19	59.75±0.35	57.27±0.16	55.08±0.28	51.95±0.26
LM	66.73±0.17	66.21±0.21	66.43±0.22	66.54±0.25	67.02±0.33	66.25±0.28	67.43±0.22	66.38±0.24	66.33±0.16	66.25±0.21	65.80±0.22	65.48±0.16	64.24±0.32	62.88±0.35	62.17±0.26	60.64±0.28	58.87±0.32
FM	66.19±0.23	67.38±0.32	66.67±0.30	66.22±0.16	66.31±0.22	66.13±0.19	66.55±0.23	66.72±0.17	66.61±0.27	66.77±0.33	66.78±0.22	64.95±0.29	63.48±0.30	63.30±0.35	62.07±0.26	60.34±0.18	60.05±0.23
StanfordCars - LP																	
RANDOM	46.47±0.23	46.95±0.20	46.32±0.21	46.33±0.19	46.49±0.32	46.63±0.28	45.39±0.30	46.46±0.18	43.76±0.25	42.00±0.28	43.74±0.33	43.61±0.22	41.80±0.35	38.84±0.29	37.81±0.30	36.20±0.31	32.76±0.28
MODERATE	46.66±0.16	46.51±0.24	46.85±0.30	45.64±0.16	46.03±0.21	46.67±0.22	45.21±0.17	46.23±0.32	45.47±0.29	43.03±0.33	42.35±0.30	41.00±0.28	41.09±0.24	38.69±0.26	37.89±0.33	34.81±0.22	29.97±0.32
GRAND	46.70±0.27	46.77±0.19	46.62±0.20	46.33±0.30	45.55±0.16	45.59±0.33	45.94±0.22	46.02±0.28	44.40±0.24	40.92±0.30	39.93±0.31	38.73±0.19	37.71±0.28	35.56±0.34	32.51±0.16	31.23±0.28	28.06±0.26
LM	47.60±0.28	48.60±0.24	47.52±0.30	48.12±0.33	46.81±0.20	47.49±0.22	48.78±0.27	47.36±0.23	47.63±0.26	47.27±0.29	47.42±0.30	46.88±0.24	47.56±0.26	46.45±0.33	46.80±0.18	45.07±0.25	42.87±0.32
FM	47.69±0.22	47.98±0.16	48.69±0.33	47.21±0.19	47.23±0.32	48.17±0.18	47.80±0.30	47.41±0.28	47.63±0.26	48.46±0.21	47.08±0.30	47.37±0.16	46.93±0.31	46.64±0.33	46.47±0.22	44.45±0.24	42.10±0.28
UCF101 - LP																	
RANDOM	71.96±0.31	70.53±0.28	71.53±0.33	72.03±0.30	70.37±0.24	69.94±0.31	70.84±0.32	70.16±0.27	69.07±0.29	69.07±0.28	67.27±0.25	67.06±0.24	65.98±0.26	64.97±0.30	63.84±0.22	61.88±0.23	58.60±0.32
MODERATE	72.72±0.24	71.56±0.29	71.58±0.28	70.68±0.33	70.13±0.25	70.98±0.27	70.18±0.29	70.47±0.26	70.31±0.32	69.05±0.24	67.78±0.31	66.88±0.28	66.03±0.29	66.01±0.28	64.63±0.27	63.68±0.33	60.38±0.24
GRAND	71.93±0.22	71.19±0.25	71.00±0.33	70.39±0.29	69.52±0.24	68.33±0.30	68.81±0.32	68.28±0.22	67.14±0.26	66.35±0.31	64.42±0.33	64.39±0.30	63.28±0.31	61.75±0.33	58.95±0.26	57.44±0.33	55.70±0.28
LM	73.32±0.28	73.54±0.29	72.77±0.25	72.82±0.30	72.69±0.29	72.58±0.24	72.35±0.32	72.83±0.29	72.51±0.27	71.82±0.32	71.76±0.31	71.48±0.33	70.98±0.24	69.44±0.28	68.75±0.26	66.64±0.29	64.95±0.30
FM	72.72±0.32	72.83±0.27	73.14±0.30	72.53±0.32	71.82±0.29	72.11±0.26	72.00±0.31	71.97±0.30	72.16±0.28	71.35±0.33	71.00±0.24	69.76±0.31	69.94±0.33	69.84±0.30	68.33±0.28	66.67±0.29	65.03±0.33
CIFAR10 - LP																	
RANDOM	88.9±0.24	88.43±0.28	88.46±0.27	88.36±0.29	88.27±0.31	88.17±0.33	88.64±0.26	88.88±0.27	88.01±0.30	87.38±0.24	86.77±0.28	86.91±0.31	86.38±0.24	85.58±0.32	85.18±0.27	83.99±0.26	82.85±0.33
MODERATE	89.17±0.30	88.83±0.32	88.74±0.26	88.92±0.33	88.82±0.27	88.32±0.28	88.09±0.29	88.41±0.24	87.63±0.26	86.99±0.27	86.85±0.29	87.31±0.32	86.36±0.33	86.20±0.26	84.75±0.30	84.35±0.26	82.14±0.31
GRAND	89.35±0.29	88.83±0.24	88.69±0.27	88.63±0.28	88.91±0.30	88.96±0.26	89.04±0.33	89.04±0.29	88.27±0.32	87.34±0.27	87.75±0.24	87.39±0.31	86.86±0.27	86.72±0.33	85.61±0.26	84.74±0.31	82.72±0.32
LM	89.51±0.31	89.97±0.33	89.42±0.24	89.79±0.27	89.42±0.26	89.31±0.32	89.14±0.30	89.15±0.31	88.81±0.33	88.74±0.30	88.81±0.27	88.54±0.26	88.33±0.32	87.43±0.29	87.03±0.30	86.62±0.27	85.87±0.31
FM	89.53±0.28	89.32±0.30	89.32±0.31	89.47±0.29	89.49±0.32	89.25±0.31	89.23±0.30	89.05±0.27	89.41±0.33	88.89±0.26	88.48±0.29	88.13±0.30	87.79±0.27	87.39±0.33	86.50±0.30	85.95±0.32	84.84±0.27
Flowers102 - LP																	
RANDOM	94.14±0.24	94.25±0.28	94.32±0.27	94.23±0.29	94.03±0.31	93.99±0.33	93.96±0.26	94.20±0.27	93.91±0.30	93.30±0.24	93.26±0.28	93.59±0.31	93.26±0.24	92.77±0.32	91.51±0.27	91.56±0.26	88.96±0.33
MODERATE	94.24±0.30	94.19±0.32	94.6±0.26	94.19±0.33	94.19±0.27	94.38±0.28	94.23±0.29	94.28±0.24	93.3±0.26	93.71±0.27	93.67±0.29	93.42±0.32	92.73±0.33	92.61±0.26	92±0.30	91.56±0.26	88.63±0.31
GRAND	94.03±0.29	94.44±0.24	94.11±0.27	94.36±0.28	94.68±0.30	94.03±0.26	94.4±0.33	94.2±0.29	94.28±0.32	93.4±0.27	93.59±0.24	93.34±0.31	93.02±0.27	92.45±0.33	91.03±0.26	90.66±0.31	89.28±0.32
LM	94.11±0.31	94.60±0.33	94.31±0.26	94.32±0.30	94.68±0.26	94.23±0.32	94.29±0.27	94.32±0.33	94.32±0.26	94.64±0.31	94.67±0.30	94.15±0.24	93.99±0.32	93.54±0.31	93.5±0.26	91.96±0.29	90.91±0.30
FM	94.36±0.26	94.59±0.33	94.40±0.24	94.71±0.32	94.60±0.31	94.28±0.33	94.34±0.26	94.21±0.31	94.60±0.30	94.44±0.33	94.32±0.29	93.91±0.26	93.42±0.31	93.34±0.24	93.50±0.30	92.05±0.27	90.22±0.26

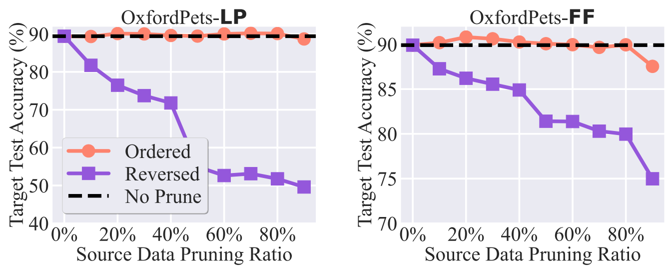


Figure A3: Performance comparison on the downstream performance of LM with the “Ordered” and the “Reversed” pruning order. Here, the "Ordered" strategy retains source classes with the highest scores, while the "Reversed" order prunes these classes first. Other settings are aligned with Fig. 4.

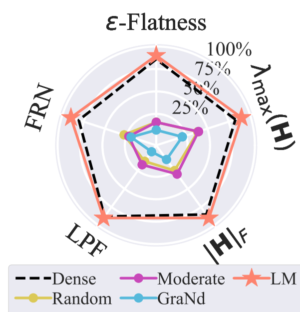


Figure A4: Flatness evaluations on the models pretrained with the pruned source data using different methods. The models are first pretrained and then finetuned using LP. The flatness are evaluated with respect to the downstream training loss and are quantified by the reversed sharpness evaluated through five widely acknowledged sharpness metrics. The results are normalized to $0\% \sim 100\%$, with 100% denoting the highest (best) flatness across all the methods given one specific flatness metric.

and subsequently finetuned on the downstream OxfordPets dataset [54]. Better flatness in the loss landscape is typically associated with better transferability. We quantify this flatness using the measure of *reversed sharpness* calculated via five widely accepted metrics: ϵ-sharpness [113], low pass filter-based measure (LPF) [114], the max eigenvalue of the Hessian ($\lambda_{\max}(\mathbf{H})$) [115], the Frobenius norm of the Hessian ($|\mathbf{H}|_F$) [115], and Fisher Rao Norm (FRN) [116]. The results suggest that pretraining on the subset selected by LM leads the model towards a flatter region in the downstream loss landscape, potentially contributing to the superior transferability of LM when compared to the baseline methods.

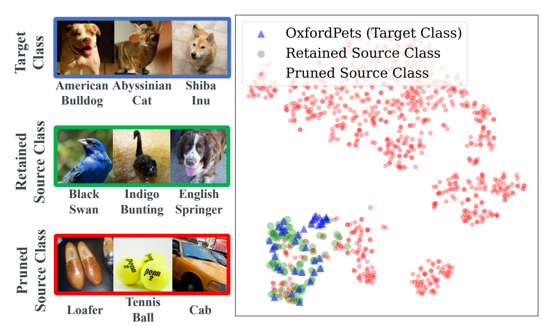
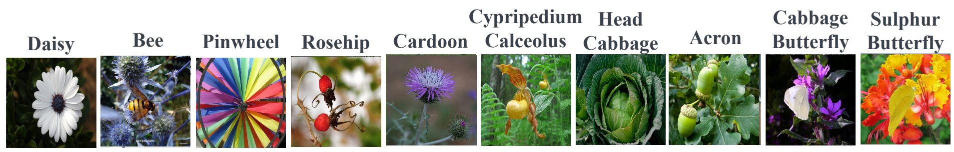


Figure A5: (Left) Interpretation merit of the data pruning strategy by LM. (Right) Feature distribution visualization using t-SNE for the source class selection by LM for OxfordPets with 90% pruning ratio.

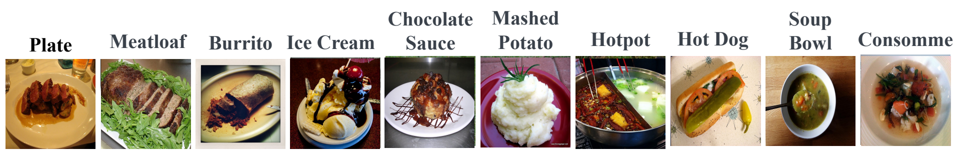
Feature distribution analysis. **Fig. A5** provides visual explanations of DP at both deep representation (feature) and data levels. Here we focus on the LM method with a pruning ratio of 90% given the downstream dataset OxfordPets. The other settings are consistent with Fig. 4. In **Fig. A5 (right)**, we visualize the source ImageNet classes (including pruned ones and retrained ones) and target OxfordPets classes in terms of their class-wise feature centroids in a 2D space achieved by t-SNE[117]. The class feature centroid is obtained by averaging the features of data points within a class, extracted by the pretrained source model on the full ImageNet dataset. As we can see, all retained source classes are grouped together and are close to the target classes. This indicates that the source classes that the pruning process share the most resemblance with the target data. In contrast, the pruned source classes are more dispersed and located further away from the target data classes. Furthermore, **Fig. A5 (left)** exhibits image examples of target classes as well as pruned and retrained source classes. We observe that image examples in the retained source classes (*e.g.*, relating to animals) are semantically closer to the target data points (relating to pets) than the pruned ones. This highlights the ability of LM to effectively identify and retain the most relevant classes for the downstream tasks. We provide more examples for FM in Fig. A6.

Examining the top selected classes by FM and their image examples. In **Fig. A6**, we showcase the top-10 source classes chosen by FM, as determined by the endowed scores. These selected classes closely correspond to the downstream datasets' subjects, demonstrating FM's effectiveness in identifying relevant classes for transfer learning. This finding also aligns with our observations in Fig. A5, showing that FM identifies source classes resembling downstream data.

(a) DTD.



(b) Flowers102.



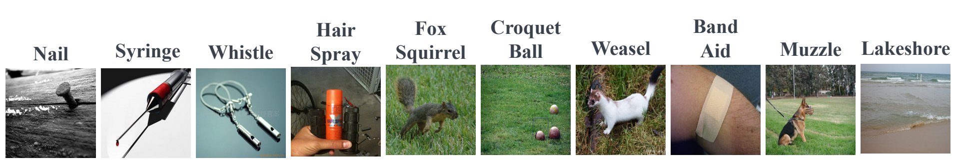
(c) Food101.

(d) OxfordPets.

(e) Stanfordcars.

(f) Sun397.

(g) UCF101.



(h) CIFAR-10.

Figure A6: The source classes with top-10 scores selected by FM for the 8 downstream tasks studied in this work. For each class, the class label (name) as well as an representative image example is presented.