Task-Specific Knowledge Improves Generalization: A Logits-Based Framework for Continual Learning of Vision-Language Models

Supplementary Material

7. More Analysis

Fusion Ratio Generation Methods. We investigate strategies for generating fusion coefficients λ_{min} and λ_{ori}, which correspond to the selected task-specific logits l_{min} and the vanilla CLIP logits l_{ori} respectively. Besides using constant coefficients in our main approach, we evaluate two adaptive methods. These methods use the minimum entropy H_{min} from the fine-tuned model and the entropy H_{ori} from vanilla CLIP, the goal is to assign higher weights to logits with lower entropy, since lower entropy reflects higher confidence. The first adaptive strategy is a reciprocal-based normalization, where the coefficients are defined as:

$$\lambda_{\text{min}} = \frac{1/H_{\text{min}}}{1/H_{\text{min}} + 1/H_{\text{ori}}} \tag{17}$$

$$\lambda_{\text{ori}} = \frac{1/H_{\text{ori}}}{1/H_{\text{min}} + 1/H_{\text{ori}}} \tag{18}$$

The second is a softmax-based normalization, defined as:

$$\lambda_{\text{min}}, \lambda_{\text{ori}} = \text{softmax}\big([-H_{\text{min}}, -H_{\text{ori}}]\big). \tag{19}$$

As reported in Figure 7, both adaptive strategies yield suboptimal performance compared to our primary method. The use of constant coefficients consistently achieves superior results, highlighting its robustness and effectiveness.

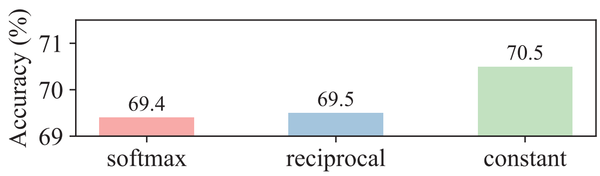

Figure 7. Effect of different coefficient-generation methods on “Transfer” metric.

Analysis of Generalization improvement. Figure 8 compares the generalization performance of different methods with vanilla CLIP, evaluated using the “Transfer” metric on all three settings in Order-I. The figure shows the difference between the “Transfer” accuracy of each method and that of vanilla CLIP, only positive differences are displayed for clarify.

We observe that most prior methods are primarily designed to preserve the original zero-shot capabilities of the model rather than enhance them. Although some recent approaches surpass the CLIP baseline on certain datasets, these improvements lack consistency across the full benchmark. In contrast, the proposed entropy-guided logits ensemble method consistently exceeds zero-shot performance on a majority of the datasets, especially on MTIL-FS and X-TAIL settings. This highlights the strong capability of the method to leverage prior task knowledge and enhance classification performance on new unseen tasks.

Routing Mechanisms. We analyze the routing results based on Mahalanobis distance. As previously mentioned, beyond the original distributions for each downstream task, we construct an ImageNet-based distribution to serve as a proxy for OOD tasks. However, due to complex latent correlations among tasks, not all samples from unseen tasks are processed by the logits ensemble strategy.

Table 6 presents the routing results on test samples after training on each task. Each entry in row i, column j represents: after the model has been trained on task i, the fraction of test samples from task j that are routed to the logits ensemble strategy. The diagonal elements are highlighted in YellowGreen. Ideally, all test samples from seen tasks should be recognized as ID, and all test samples from unseen tasks should be recognized as OOD. This is reflected in the content, where the diagonal and elements below the diagonal ideally equal 0%, and the elements above the diagonal ideally equal 100%.

We observe that all test samples from seen tasks are correctly routed to their respective fine-tuned PEFT modules, for samples ideally belonging to OOD tasks, some or all are routed to PEFT modules of other seen tasks, with corresponding values highlighted in an asterisk (*) in *italics*, primarily involving EuroSAT and MNIST datasets. Specifically, after training on CIFAR-100, approximately 40% of EuroSAT test samples are processed by the CIFAR-100-trained PEFT module, while the remaining 60% utilize the logits ensemble strategy. For MNIST, all samples are routed to the CIFAR-100-trained PEFT module. This occurs because CIFAR-100's broad category coverage makes it highly relevant to EuroSAT and MNIST, providing more specific knowledge than the general ImageNet distribution.

Adaptive Nature of Routing Mechanisms. As shown in Table 7, based on the routing mechanisms analyzed above, we compared the transfer accuracy under two configurations: the theoretical logits ensemble and the practically adopted routing-based logits ensemble. Under the theoretical condition, all test samples from unseen tasks are processed through the logits ensemble strategy, which corresponds to the ”Routing-based” row in the table. In practice, however, the routing mechanism exhibits different behavior. As shown in the analysis above, when an unseen task demonstrates high similarity to certain seen tasks, the cor-

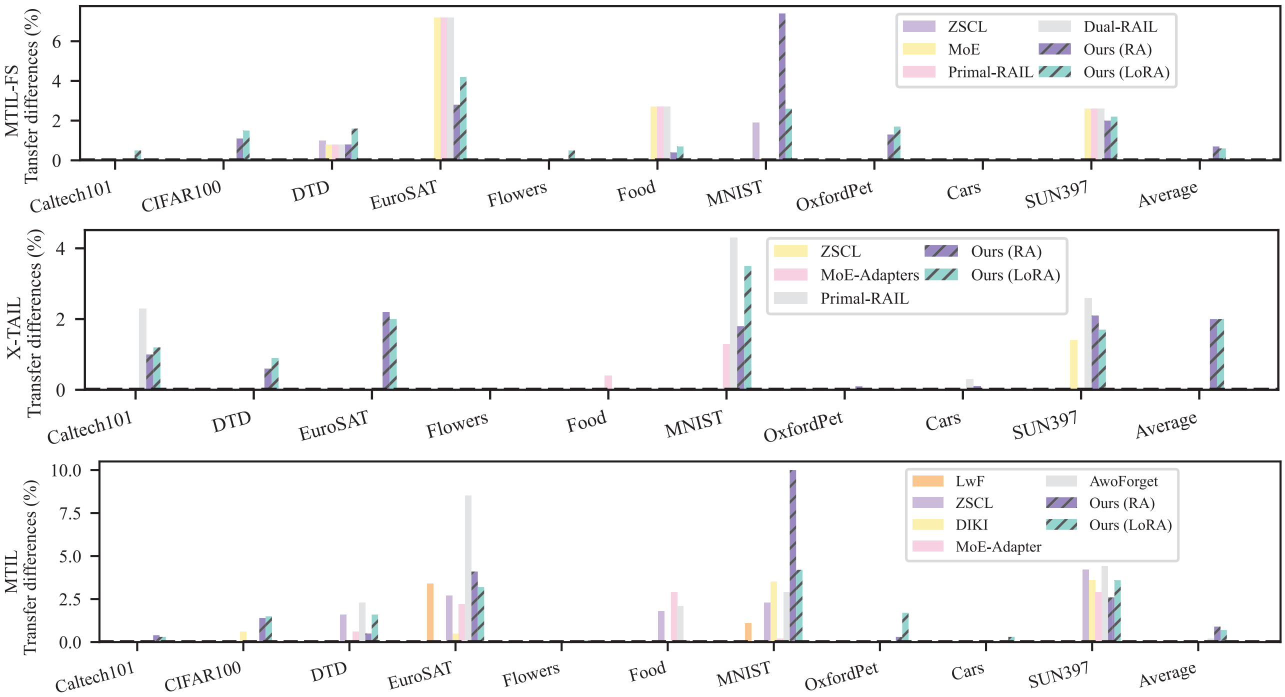

Figure 8. Comparison of the generalization performance between different methods and vanilla CLIP, evaluated using the "Transfer" metric on all three settings in Order-I. Each bar shows the difference between the "Transfer" accuracy of each method and that of vanilla CLIP, only positive differences are displayed for clarify. It can be observed that, unlike previous methods, our method surpasses vanilla CLIP on the "Transfer" metric over most datasets, especially on the MTIL-FS and X-TAIL settings.

Table 6. Routing results of the improved Mahalanobis distance-based method. Each row corresponds to the dataset used in training, and each column corresponds to the dataset used in inference. The value in each cell denotes the percentage of test samples from the column dataset that are routed as OOD after the model is trained on the row dataset. Outliers are marked with an asterisk (*) in *italics*. Diagonal elements are highlighted with YellowGreen background.

Train \ Test	Aircraft	Caltech101	CIFAR100	DTD	EuroSAT	Flowers	Food	MNIST	OxfordPet	Cars	SUN397
Aircraft	0	100	100	100	100	100	100	100	100	100	100
Caltech101	0	0	100	100	100	100	100	100	100	100	100
CIFAR100	0	0	0	100	*60.05**	100	100	*0**	100	100	100
DTD	0	0	0	0	*60.05**	100	100	*0**	100	100	100
EuroSAT	0	0	0	0	0	100	100	*0**	100	100	100
Flowers	0	0	0	0	0	0	100	*0**	100	100	100
Food	0	0	0	0	0	0	0	*0**	100	100	100
MNIST	0	0	0	0	0	0	0	0	100	100	100
OxfordPet	0	0	0	0	0	0	0	0	0	100	100
Cars	0	0	0	0	0	0	0	0	0	0	100
SUN397	0	0	0	0	0	0	0	0	0	0	0

Table 7. Comparison of "Transfer" metric on unseen datasets. "Routing-based" refers to applying the proposed routing and logits ensemble strategy during inference, whereas "Ensemble for all" indicates disabling routing and applying the logits ensemble strategy to all test samples. "Δ" denotes the change in accuracy, with positive values highlighted in **bold** and negative values underlined.

Method	Aircraft	Caltech101	CIFAR100	DTD	EuroSAT	Flowers	Food	MNIST	OxfordPet	Cars	SUN397	Average
Routing-based	—	93.3	69.8	44.3	51.8	69.9	85.9	69.5	89.4	65.7	65.2	70.5
Ensemble for all	—	93.2	69.7	44.5	49.2	70.0	86.1	68.6	89.5	65.5	65.1	70.1
Δ	—	-0.1	-0.1	**+0.2**	-2.6	**+0.1**	**+0.2**	-0.9	**+0.1**	-0.2	-0.1	-0.4

responding test samples are partially or fully handled by the PEFT module associated with those similar seen tasks instead of undergoing logits ensemble, this practical scenario corresponds to the ”Ensemble for all” row in the table.

The results align with our routing analysis, for most datasets, the performance change is negligible ($< 0.2\%$), except for EuroSAT and MNIST, which show performance drops of 2.6% and 0.9% respectively under the theoretical condition. This behavior reveals the adaptive nature of our routing strategy: the weighting coefficient λ for $l_{\min}$ dynamically adjusts based on task similarity. When samples closely resemble a seen task and are not recognized as OOD, logits are generated entirely by the PEFT module of the corresponding task, resulting in $\lambda = 1.0$ to maximize knowledge transfer. Conversely, when test samples are recognized as OOD due to lower inter-task similarity, λ takes intermediate values to balance the utilization of learned knowledge and the overconfidence of fine-tuned model.

8. More Results

8.1. Comparison with State-of-the-Arts

MTIL-FS. We further evaluate our method under the more challenging few-shot learning setting MTIL-FS, where models are trained with limited few data. As reported in Table 8 for Order-I, our method demonstrates substantial performance improvements over the leading baseline, Dual-RAIL. Specifically, the “Transfer” metric improves by 0.9% (RA) and 0.8% (LoRA), while the “Average” metric sees increases of 0.8% (RA) and 1.1% (LoRA). Our model also demonstrates a notable advantage in knowledge retention, with the “Last” metric improving by 0.6% (RA) and 1.5% (LoRA).

X-TAIL. As shown in Table 9, excluding Dual-RAIL due to its prohibitive memory requirement for storing all feature representations, our method consistently surpasses the strongest feasible baseline, Primal-RAIL, by +2.0% / +2.0% on “Transfer”, +3.2% / +3.6% on “Average”, and +1.9% / +2.6% on “Last”. Notably, gains are simultaneously achieved on both adaptation-oriented (“Last”) and transfer-oriented (“Transfer”) metrics, suggesting that our framework effectively mitigates the plasticity–stability trade-off rather than shifting the balance toward one side.

Order-II Comparison Result. Table 10-12 presents the performance comparison under the Order-II setting across the three configurations: MTIL, MTIL-FS, and X-TAIL. The results indicate that our method consistently outperforms other approaches, demonstrating stable and uniform performance improvements across all settings.

8.2. Detailed Results.

Table 13-24 present our detailed experimental results across the MTIL, MTIL-FS, and X-TAIL settings, including two task orders for each setting and the two variants of our proposed method, ours (RA) and ours (LoRA), resulting in a total of twelve detailed result tables.

9. Limitation and Future Work

Our analysis shows that our method can fully adapt to downstream ID tasks, while also effectively leveraging previously learned task-specific knowledge to improve generalization on OOD tasks. The entropy-guided logits ensemble strategy, while effective in improving generalization, suffers from certain disadvantages in inference time. In contrast, the Mahalanobis distance-based ensemble strategy surpasses current SOTA baselines while maintaining higher time efficiency. Analysis of Table 6 and 7 further shows that for some test samples, processing entirely with the PEFT module of highly similar seen tasks yields better generalization than using the logits ensemble strategy. These findings highlight the need for more precise identification of similarity between test samples and seen tasks to enable more appropriate processing. This motivates enhancements to our routing strategy, aiming for constant-time complexity in task-specific PEFT module selection and adaptive weight allocation based on similarity rather than fixed values. Implementing these strategies could further improve overall model performance and serves as a direction for future work.

Table 8. Comparison with state-of-the-art methods on MTIL-FS setting in Order-I, in terms of “Transfer”, “Average”, and “Last” scores (%). The **best** and the **second best** results are highlighted in **red** and **blue**, respectively.

Method	Aircraft	Caltech101	CIFAR100	DTD	EuroSAT	Flowers	Food	MNIST	OxfordPet	Cars	SUN397	*Average*
Zero-shot	24.8	92.9	68.4	43.8	47.7	71.4	85.8	59.5	89.1	65.8	62.6	64.7
Transfer												
Lwf [19]	–	72.1	49.2	35.9	44.5	41.1	66.6	50.5	69.0	19.0	51.7	50.0
WiSE-FT [37]	–	77.6	60.0	41.3	39.4	53.0	76.6	58.1	75.5	37.3	58.2	57.7
ZSCL [46]	–	84.0	68.1	**44.8**	46.8	63.6	84.9	61.4	81.4	55.5	62.2	65.3
MoE-Adapters [40]	–	87.9	68.2	44.1	48.1	64.7	**88.8**	**69.0**	89.1	64.5	**65.1**	68.9
Primal-RAIL [39]	–	88.4	68.2	44.6	**54.9**	71.0	**88.5**	59.6	89.0	64.7	**65.2**	69.4
Dual-RAIL [39]	–	88.4	68.2	44.6	**54.9**	71.0	**88.5**	59.6	89.0	64.7	**65.2**	69.4
Ours (RA)	–	**92.9**	**69.5**	44.6	50.5	**71.2**	86.2	**66.9**	**90.4**	**65.8**	64.6	**70.3**
Ours (LoRA)	–	**93.4**	**69.9**	**45.4**	**51.9**	**71.9**	86.5	62.1	**90.8**	**65.4**	64.8	**70.2**
Average												
LwF [19]	23.5	77.4	43.5	41.7	43.5	52.2	54.6	63.4	68.0	21.3	52.6	49.2
WiSE-FT [37]	32.0	87.7	61.0	55.8	68.1	69.3	76.8	71.5	77.6	42.0	59.3	63.7
ZSCL [46] DIKI [30]	28.2	86.6	66.5	53.5	56.3	73.4	**83.1**	56.4	82.4	57.5	62.9	64.4
MoE-Adapters [40]	30.0	89.6	73.9	58.7	69.3	79.3	88.1	**76.5**	89.1	65.3	**65.8**	71.4
Primal-RAIL [39]	32.9	**94.5**	69.9	58.1	**71.8**	**84.4**	**88.5**	70.4	89.0	66.1	**65.7**	71.9
Dual-RAIL [39]	36.0	94.2	70.9	**58.8**	70.6	**84.3**	**88.5**	70.3	89.7	66.5	**65.8**	72.3
Ours (RA)	**36.4**	94.4	**75.1**	**59.5**	**72.1**	83.3	85.7	**74.1**	**90.6**	**68.0**	65.3	**73.1**
Ours (LoRA)	**38.0**	**95.1**	**76.3**	58.3	70.9	**84.3**	86.4	73.5	**90.9**	**67.7**	65.6	**73.4**
Last												
LwF [19]	22.1	58.2	17.9	32.1	28.1	66.7	46.0	84.3	64.1	31.5	60.1	46.5
WiSE-FT [37]	30.8	88.9	59.6	60.3	80.9	81.7	77.1	**94.9**	83.2	62.8	70.0	71.9
ZSCL [46]	26.8	88.5	63.7	55.7	60.2	82.1	82.6	58.6	85.9	66.7	70.4	67.4
MoE-Adapters [40]	30.1	89.3	74.9	64.0	**82.3**	89.4	**87.1**	89.0	89.1	69.5	72.5	76.1
Primal-RAIL [39]	32.9	**95.1**	70.3	63.2	81.5	**95.6**	**88.5**	89.7	89.0	72.5	71.0	77.2
Dual-RAIL [39]	36.0	94.8	71.5	**64.1**	79.5	**95.3**	**88.5**	89.4	**91.5**	71.1	71.1	77.9
Ours (RA)	**36.5**	94.4	**76.3**	**65.0**	**84.5**	93.4	85.2	86.7	**91.2**	**77.9**	**72.7**	**78.5**
Ours (LoRA)	**38.1**	**95.2**	**77.8**	63.2	81.7	94.6	86.3	**93.4**	**91.2**	**78.0**	**73.4**	**79.4**

Table 9. Comparison with state-of-the-art methods on X-TAIL setting in Order-I, in terms of “Transfer”, “Average”, and “Last” scores (%). The **best** and the **second best** results are highlighted in **red** and **blue**, respectively.

Method	Aircraft	Caltech101	DTD	EuroSAT	Flowers	Food	MNIST	OxfordPet	Cars	SUN397	*Average*
CLIP											
Zero-shot	24.7	74.5	37.8	45.4	71.3	85.2	42.4	88.8	65.8	61.1	59.7
Transfer											
LwF [19]	–	62.4	27.8	10.7	52.0	76.0	25.4	68.3	30.4	54.5	45.3
WiSE-FT [37]	–	59.9	25.4	10.2	43.9	67.9	29.4	57.4	24.1	50.3	40.9
ZSCL [46]	–	71.3	33.8	33.0	66.2	**85.2**	40.2	81.9	57.3	62.5	59.0
MoE-Adapters [40]	–	69.5	30.8	19.0	60.3	**85.6**	43.7	85.6	55.5	57.3	56.4
Primal-RAIL [39]	–	**76.8**	37.3	36.7	63.6	84.0	**46.7**	86.7	**66.1**	**63.7**	**62.4**
Ours (RA)	–	75.5	**38.4**	**47.6**	**70.7**	**85.2**	44.2	**88.9**	**65.9**	**63.2**	**64.4**
Ours (LoRA)	–	**75.7**	**38.7**	**47.4**	**70.4**	**85.2**	**45.9**	**88.8**	65.1	62.8	**64.4**
Average											
LwF [19]	25.8	81.2	48.1	33.1	57.0	75.1	54.9	74.5	35.5	57.1	54.2
WiSE-FT [37]	14.9	79.8	45.3	17.1	55.7	70.1	52.0	68.0	35.6	53.3	49.2
ZSCL [46]	39.7	80.8	52.9	40.8	79.3	**88.0**	51.4	85.5	62.9	64.1	64.5
MoE-Adapters [40]	52.4	79.4	57.7	42.7	81.1	86.6	64.8	86.7	61.3	59.0	67.2
Primal-RAIL [39]	48.8	89.6	59.0	74.4	84.0	86.1	65.6	89.5	68.9	62.3	72.8
Ours (RA)	**53.4**	**93.3**	**67.2**	**82.5**	**86.3**	**87.3**	**66.4**	**90.5**	**69.1**	**64.5**	**76.0**
Ours (LoRA)	**55.1**	**93.0**	**68.2**	**82.4**	**86.7**	87.2	**67.7**	**90.6**	**68.9**	**64.2**	**76.4**
Last											
LwF [19]	9.6	77.1	55.3	38.7	60.5	83.1	**99.5**	85.9	49.6	**80.0**	63.9
WiSE-FT [37]	18.1	84.9	53.4	27.0	69.6	88.0	88.4	91.5	76.7	**80.2**	67.8
ZSCL [46]	33.8	80.4	60.2	31.1	85.8	**91.3**	80.4	93.7	**84.9**	79.0	72.1
MoE-Adapters [40]	51.9	79.0	64.2	51.5	95.1	87.6	96.4	89.1	**84.4**	74.0	77.3
Primal-RAIL [39]	45.8	94.1	70.7	**94.2**	96.5	89.0	98.1	93.5	82.0	76.5	84.0
Ours (RA)	**53.4**	**95.3**	**74.4**	**97.9**	**96.7**	**89.3**	**99.3**	**94.3**	82.0	**76.3**	**85.9**
Ours (LoRA)	**55.2**	**94.9**	**75.5**	**97.9**	**97.7**	**89.3**	**99.3**	**94.8**	84.1	**77.4**	**86.6**

Table 10. Comparison with state-of-the-art methods on MTIL benchmark in Order-II, in terms of “Transfer”, “Average”, and “Last” scores (%).

Method	Cars	Food	MNIST	OxfordPet	Flowers	SUN397	Aircraft	Caltech101	DTD	EuroSAT	CIFAR100	*Average*
CLIP												
Zero-shot	65.8	85.8	59.5	89.1	71.4	62.6	24.8	92.9	43.8	47.7	68.4	64.7
Transfer												
LwF [19]	–	87.8	58.5	71.9	46.6	57.3	12.8	81.4	34.5	34.5	46.8	53.2
iCaRL [28]	–	86.1	51.8	67.6	50.4	57.9	11.0	72.3	31.2	32.7	48.1	50.9
WiSE-FT [37]	–	87.2	57.6	67.0	45.0	54.0	12.9	78.6	35.5	28.4	44.3	51.1
ZSCL [46]	–	88.3	57.5	84.7	68.1	**64.8**	21.1	88.2	**45.3**	**55.2**	68.2	64.1
DIKI [30]	–	85.8	59.8	89.1	**71.8**	62.6	24.3	**93.3**	42.7	46.8	67.8	64.4
MoE-Adapters [40]	–	**88.8**	59.5	89.1	69.9	64.4	18.1	86.9	43.7	**54.6**	68.2	64.3
AwoForget [45]	–	**88.6**	61.7	87.5	70.7	**66.5**	22.3	89.3	**45.5**	53.6	68.7	65.4
Ours (RA)	–	86.0	**64.9**	**89.4**	**70.9**	64.5	**24.9**	**93.7**	44.4	49.8	**70.1**	**65.9**
Ours (LoRA)	–	85.7	**64.6**	**90.9**	**71.8**	64.7	**24.3**	93.2	44.8	50.7	**70.6**	**66.1**
Average												
LwF [19]	49.0	77.0	92.1	85.9	66.5	67.2	20.9	84.7	44.6	45.5	50.5	62.2
iCaRL [28]	52.0	75.9	77.4	74.6	58.4	59.3	11.7	79.6	42.1	43.2	51.7	56.9
WiSE-FT [37]	52.6	79.3	91.9	83.9	63.4	65.2	23.3	83.7	45.4	40.0	48.2	61.5
ZSCL [46]	81.7	**91.3**	91.1	91.0	82.9	**72.5**	33.6	89.7	53.3	**62.8**	69.9	74.5
DIKI [30]	81.9	88.9	92.1	**92.8**	87.7	70.3	34.3	**94.2**	51.5	56.1	69.5	74.5
MoE-Adapters [40]	84.9	89.9	89.3	91.4	86.2	72.2	33.4	89.4	**53.3**	61.4	69.9	74.7
AwoForget [45]	**86.1**	**91.8**	92.0	91.8	85.2	**74.7**	36.6	90.9	**54.1**	**61.7**	70.5	76.0
Ours (RA)	**86.1**	88.9	**93.2**	**92.8**	**88.1**	71.6	**38.8**	**94.7**	53.0	58.7	**71.7**	**76.2**
Ours (LoRA)	**86.7**	89.5	**93.1**	**93.8**	**88.3**	72.1	**39.7**	**94.2**	**53.3**	59.2	**72.1**	**76.5**
Last												
LwF [19]	34.6	69.6	99.3	88.7	61.1	72.5	32.5	88.1	65.6	90.9	**87.9**	71.9
iCaRL [28]	46.0	81.5	91.3	82.8	66.5	72.2	16.3	91.6	68.1	83.2	**87.8**	71.6
WiSE-FT [37]	35.6	76.9	**99.5**	89.1	62.1	71.8	27.8	90.8	67.0	85.6	87.6	72.2
ZSCL [46]	78.2	**91.1**	97.6	92.5	87.4	**78.2**	45.0	92.3	72.7	96.2	86.3	83.4
DIKI [30]	81.9	89.2	**99.4**	**94.3**	96.8	76.7	46.3	95.9	74.8	**98.3**	86.6	85.5
MoE-Adapters [40]	84.1	88.5	94.0	91.8	94.1	77.8	50.4	93.3	**77.1**	87.7	86.6	84.1
AwoForget [45]	83.9	**91.5**	97.6	92.7	89.2	80.7	50.6	92.9	75.3	97.0	**87.8**	85.4
Ours (RA)	**86.1**	89.2	**99.5**	94.1	**97.9**	77.5	**55.4**	**96.6**	**76.0**	**98.8**	87.3	**87.1**
Ours (LoRA)	**86.7**	89.9	**99.4**	**94.9**	**97.8**	**78.3**	**58.2**	**96.0**	75.7	97.8	87.4	**87.4**

Table 11. Comparison with state-of-the-art methods on MTIL-FS setting in Order-II, in terms of “Transfer”, “Average”, and “Last” scores (%).

Method	Cars	Food	MNIST	OxfordPet	Flowers	SUN397	Aircraft	Caltech101	DTD	EuroSAT	CIFAR100	*Average*
CLIP												
Zero-shot	65.8	85.8	59.5	89.1	71.4	62.6	24.8	92.9	43.8	47.7	68.4	64.7
Transfer												
LwF [19]	–	64.2	59.1	68.1	38.4	54.9	6.7	78.0	35.5	33.5	47.4	48.6
WiSE-FT [37]	–	77.3	60.0	76.9	54.2	58.0	11.1	81.8	37.6	31.7	48.1	53.7
ZSCL [46]	–	**87.3**	**64.8**	85.3	67.9	**64.5**	18.9	86.6	43.6	43.2	65.7	62.8
MoE-Adapters [40]	–	**88.8**	59.5	89.1	**71.2**	**65.3**	18.2	87.9	44.2	**54.6**	68.2	64.7
Ours (RA)	–	86.0	**64.9**	**89.8**	**71.5**	**65.3**	**24.7**	93.1	**44.5**	48.0	**69.6**	**65.8**
Ours (LoRA)	–	85.5	60.5	**91.2**	**71.5**	**65.3**	**24.0**	**93.5**	**45.4**	**50.0**	**70.4**	**65.7**
Average												
LwF [19]	64.1	55.0	79.5	69.2	55.7	58.3	10.8	81.7	41.3	39.2	47.4	54.7
WiSE-FT [37]	59.3	64.7	77.4	70.3	51.3	58.6	10.8	84.2	42.0	38.6	49.1	55.1
ZSCL [46]	70.0	85.0	79.8	86.1	79.4	68.3	21.8	88.8	48.8	49.3	66.5	67.6
MoE-Adapters [40]	61.2	**87.0**	**87.3**	89.1	79.3	68.5	23.4	89.4	49.9	60.8	68.8	69.5
Ours (RA)	**77.8**	85.1	82.4	**91.3**	**86.0**	**69.5**	**30.5**	**93.5**	**49.7**	**54.6**	**70.1**	**71.9**
Ours (LoRA)	**77.6**	**86.1**	**88.2**	**91.8**	**85.9**	**69.7**	**31.2**	**94.0**	**50.2**	**56.5**	**71.0**	**72.9**
Last												
LwF [19]	57.1	40.1	84.1	58.1	50.5	57.6	14.3	87.9	54.7	64.0	47.0	56.8
WiSE-FT [37]	48.1	47.7	66.9	59.8	25.0	56.1	7.4	88.5	52.2	66.8	59.4	51.8
ZSCL [46]	67.4	82.7	78.7	85.7	81.3	71.2	25.0	92.5	62.0	72.2	74.4	71.8
MoE-Adapters [40]	59.4	**87.0**	**91.8**	89.0	84.1	71.9	29.4	91.4	**64.2**	**88.8**	**75.0**	75.7
Ours (RA)	**77.8**	85.0	86.3	**91.9**	**94.3**	**72.9**	**37.6**	**94.2**	**63.4**	83.9	74.5	**78.3**
Ours (LoRA)	**77.6**	**86.1**	**94.3**	**92.0**	**94.0**	**73.3**	**39.9**	**95.0**	62.9	**86.0**	**76.7**	**79.8**

Table 12. Comparison with state-of-the-art methods on full-shot X-TAIL setting in Order-II, in terms of “Transfer”, “Average”, and “Last” scores (%).

Method	Cars	Aircraft	OxfordPet	Food	SUN397	MNIST	Flowers	DTD	Caltech101	EuroSAT	*Average*
CLIP											
Zero-shot	65.8	24.7	88.8	85.2	61.1	42.4	71.3	37.8	74.5	45.4	59.7
Transfer											
LwF [19]	–	16.7	78.5	76.0	59.7	41.3	46.6	27.3	63.3	10.4	46.6
WiSE-FT [37]	–	0.3	77.9	75.7	55.6	39.6	45.0	25.4	58.9	8.3	45.2
ZSCL [46]	–	21.7	83.2	**85.6**	**63.0**	39.3	61.8	34.3	72.2	26.4	54.2
MoE-Adapters [40]	–	17.5	87.1	**86.8**	58.2	**44.2**	63.4	33.9	67.9	15.3	52.7
Primal-RAIL [39]	–	23.8	**87.8**	83.4	60.8	43.9	64.1	36.4	74.3	37.4	57.7
Ours (RA)	–	**24.7**	**89.3**	85.3	**62.8**	**46.3**	**71.2**	**39.2**	**76.2**	**44.9**	**60.0**
Ours (LoRA)	–	**24.8**	**89.3**	**85.6**	62.1	42.8	**71.4**	**38.7**	**75.5**	**46.0**	**59.6**
Average											
LwF [19]	38.6	21.6	73.2	75.9	67.7	69.4	62.2	36.8	68.0	15.1	52.9
WiSE-FT [37]	47.1	30.9	77.9	76.6	65.0	59.0	58.7	36.1	65.4	9.8	52.7
ZSCL [46]	77.8	43.1	90.3	**89.5**	**71.8**	61.3	73.6	42.9	74.2	26.8	65.1
MoE-Adapters [40]	84.2	**47.4**	89.0	88.0	65.2	70.7	76.3	43.4	71.5	14.8	65.1
Primal-RAIL [39]	83.0	45.7	92.1	87.1	70.0	61.6	76.8	46.4	78.3	43.1	68.4
Ours (RA)	**85.9**	45.8	**93.1**	87.9	71.1	**73.0**	**81.6**	**49.6**	**79.6**	**50.2**	**71.8**
Ours (LoRA)	**86.7**	**47.6**	**93.5**	**88.1**	**71.5**	**71.6**	**82.0**	**49.9**	**79.4**	**51.2**	**72.1**
Last											
LwF [19]	13.4	6.9	58.2	74.5	71.2	**99.4**	76.3	53.2	85.4	57.0	59.6
WiSE-FT [37]	19.3	2.0	70.5	77.6	67.9	72.1	66.2	52.7	90.6	22.7	54.2
ZSCL [46]	75.6	31.7	90.4	**90.7**	**77.0**	75.4	88.3	60.5	82.1	29.6	70.1
MoE-Adapters [40]	84.1	**50.6**	88.9	88.1	68.7	97.2	95.5	65.6	86.2	10.3	73.5
Primal-RAIL [39]	81.9	46.1	93.3	89.0	76.6	98.2	96.6	70.6	**94.1**	94.1	84.1
Ours (RA)	**86.0**	48.2	**94.1**	89.0	76.6	99.3	**97.2**	**73.8**	93.1	**98.1**	**85.5**
Ours (LoRA)	**86.7**	**50.0**	**94.6**	**89.2**	**77.6**	**99.5**	**97.9**	**75.8**	**94.8**	**98.2**	**86.4**

Table 13. Accuracy (%) of ours (RA) on the MTIL benchmark with order-I. Each row represents the performance on every dataset of the model trained after the corresponding task. Transfer, Average, and Last metrics are shown in color.

	Aircraft	Caltech101	CIFAR100	DTD	EuroSAT	Flowers	Food	MNIST	OxfordPet	Cars	SUN397	
Transfer		93.3	69.8	44.3	51.8	69.9	85.9	69.5	89.4	65.7	65.2	70.5
Aircraft	56.4	93.3	69.6	44.3	50.4	69.7	85.9	68.8	88.9	66.1	64.1	
Caltech101	56.1	96.8	70.0	44.7	50.1	70.2	86.1	68.9	89.2	66.1	64.8	
CIFAR100	56.2	96.8	87.2	43.9	52.6	70.1	86.0	69.8	89.3	65.8	65.3	
DTD	56.3	96.8	87.2	75.9	53.9	70.0	85.9	69.9	89.1	65.8	65.3	
EuroSAT	56.3	96.8	87.2	75.9	98.1	69.8	85.8	69.8	89.1	65.7	65.3	
Flowers	56.3	96.8	87.2	75.9	98.1	98.3	85.9	69.8	89.9	65.7	65.2	
Food	56.2	96.8	87.2	75.9	98.1	98.3	89.3	69.8	89.8	65.6	65.4	
MNIST	56.1	96.8	87.2	76.0	98.1	98.3	89.3	99.4	89.9	65.6	65.4	
OxfordPet	56.2	96.8	87.2	75.9	98.1	98.3	89.3	99.4	94.4	65.3	65.4	
Cars	56.2	96.8	87.2	76.0	98.1	98.3	89.3	99.4	94.4	86.3	65.4	
SUN397	56.3	96.8	87.2	75.8	98.1	98.3	89.3	99.4	94.4	86.3	77.2	87.2
Average	56.2	96.5	84.0	67.3	81.2	85.4	87.4	80.4	90.8	69.5	66.2	78.6

Table 14. Accuracy (%) of ours (LoRA) on the MTIL benchmark with order-I. Each row represents the performance on every dataset of the model trained after the corresponding task. Transfer, Average, and Last metrics are shown in color.

	Aircraft	Caltech101	CIFAR100	DTD	EuroSAT	Flowers	Food	MNIST	OxfordPet	Cars	SUN397	
Transfer		93.2	69.9	45.4	50.9	71.5	85.8	63.7	90.8	66.1	66.2	70.3
Aircraft	57.9	93.2	69.7	44.4	48.9	71.6	86.3	64.1	90.7	65.9	64.2	
Caltech101	57.9	96.8	70.2	45.5	49.5	71.6	86.5	63.4	90.8	66.0	65.2	
CIFAR100	57.9	96.7	87.5	46.4	52.5	71.2	85.5	63.6	90.6	66.1	66.5	
DTD	57.8	96.8	87.5	77.1	53.0	71.5	85.5	63.6	90.7	66.1	66.6	
EuroSAT	57.9	96.8	87.5	77.1	98.0	71.5	85.6	63.7	90.7	66.1	66.6	
Flowers	57.7	96.7	87.5	77.1	98.0	97.9	85.7	63.6	90.9	66.2	66.5	
Food	57.9	96.8	87.5	77.1	98.0	97.9	89.7	63.7	90.9	66.1	66.6	
MNIST	57.9	96.8	87.5	77.0	98.0	97.9	89.8	99.3	90.9	66.2	66.6	
OxfordPet	57.9	96.8	87.5	77.0	98.0	97.9	89.8	99.3	94.8	66.2	66.6	
Cars	57.9	96.8	87.5	77.0	98.0	97.9	89.8	99.3	94.7	86.2	66.6	
SUN397	57.8	96.8	87.5	77.0	98.0	97.9	89.8	99.3	94.7	86.2	78.7	87.6
Average	57.8	96.4	84.3	68.4	80.9	85.9	87.6	76.6	91.9	69.7	67.3	78.8

Table 15. Accuracy (%) of ours (RA) on the MTIL benchmark with order-II. Each row represents the performance on every dataset of the model trained after the corresponding task. Transfer, Average, and Last metrics are shown in color.

	Cars	Food	MNIST	OxfordPet	Flowers	SUN397	Aircraft	Caltech101	DTD	EuroSAT	CIFAR100	Average
Transfer		86.0	64.9	89.4	70.9	64.5	24.9	93.7	44.4	49.8	70.1	65.9
Cars	86.1	86.0	65.9	89.4	70.8	64.0	25.0	93.2	44.2	49.1	70.1	
Food	86.1	89.3	63.8	89.4	71.1	64.6	25.0	93.6	44.7	49.3	69.6	
MNIST	86.1	89.2	99.5	89.4	71.1	64.6	25.9	93.7	44.8	49.0	69.6	
OxfordPet	86.1	89.3	99.5	94.1	70.7	64.5	24.9	93.8	43.9	49.3	69.8	
Flowers	86.1	89.2	99.5	94.1	97.9	64.7	24.4	93.7	44.6	49.9	69.9	
SUN397	86.1	89.2	99.5	94.1	97.9	77.5	24.1	94.0	44.4	50.4	70.1	
Aircraft	86.1	89.2	99.5	94.1	97.9	77.5	55.6	94.1	44.3	50.9	70.6	
Caltech101	86.1	89.2	99.5	94.1	97.9	77.5	55.5	96.6	44.2	50.8	70.5	
DTD	86.1	89.3	99.5	94.1	97.9	77.5	55.5	96.5	75.9	50.0	70.6	
EuroSAT	86.1	89.3	99.5	94.1	97.9	77.5	55.3	96.6	76.0	98.8	70.6	
CIFAR10	86.1	89.2	99.5	94.1	97.9	77.5	55.4	96.6	76.0	98.8	87.3	87.1
Average	86.1	88.9	93.2	92.8	88.1	71.6	38.8	94.7	53.0	58.7	71.7	76.2

Table 16. Accuracy (%) of ours (LoRA) on the MTIL benchmark with order-II. Each row represents the performance on every dataset of the model trained after the corresponding task. Transfer, Average, and Last metrics are shown in color.

	Cars	Food	MNIST	OxfordPet	Flowers	SUN397	Aircraft	Caltech101	DTD	EuroSAT	CIFAR100	Average
Transfer		85.7	64.6	90.9	71.8	64.7	24.3	93.2	44.8	50.7	70.6	66.1
Cars	86.7	85.7	65.3	90.7	71.3	64.1	24.0	92.7	43.8	51.0	69.5	
Food	86.7	89.9	63.9	91.0	71.9	64.9	24.2	93.1	44.1	50.6	70.4	
MNIST	86.7	89.9	99.4	91.0	71.9	64.9	24.3	93.1	44.4	49.6	70.2	
OxfordPet	86.7	89.9	99.4	94.8	72.0	64.9	24.3	93.1	44.4	49.5	70.2	
Flowers	86.7	89.9	99.4	94.9	97.8	64.9	24.3	93.1	44.7	49.5	70.2	
SUN397	86.7	89.9	99.4	94.8	97.8	78.3	24.6	93.6	45.6	51.4	71.0	
Aircraft	86.7	89.9	99.4	94.8	97.8	78.3	58.1	93.6	45.8	51.5	71.1	
Caltech101	86.7	89.9	99.4	94.8	97.8	78.3	58.2	96.0	45.9	51.5	71.1	
DTD	86.7	89.9	99.4	94.8	97.8	78.3	58.1	96.1	75.8	51.4	71.1	
EuroSAT	86.7	89.9	99.4	94.8	97.8	78.3	58.0	96.1	75.8	97.8	71.4	
CIFAR100	86.7	89.9	99.4	94.9	97.8	78.3	58.2	96.0	75.7	97.8	87.4	87.4
Average	86.7	89.5	93.1	93.8	88.3	72.1	39.7	94.2	53.3	59.2	72.1	76.5

Table 17. Accuracy (%) of ours (RA) on the MTIL-FS setting with order-I. Each row represents the performance on every dataset of the model trained after the corresponding task. Transfer, Average, and Last metrics are shown in color.

	Aircraft	Caltech101	CIFAR100	DTD	EuroSAT	Flowers	Food	MNIST	OxfordPet	Cars	SUN397	
Transfer		92.9	69.5	44.6	50.5	71.2	86.2	66.9	90.4	65.8	64.6	70.3
Aircraft	36.5	92.9	69.5	43.8	49.2	70.6	86.1	63.4	89.7	66.0	64.1	
Caltech101	36.5	94.5	69.5	44.4	49.3	71.0	86.3	63.5	89.7	66.0	64.4	
CIFAR100	36.4	94.5	76.3	45.6	51.8	71.4	86.2	68.4	90.5	65.7	64.8	
DTD	36.4	94.5	76.3	65.0	51.8	71.6	86.2	68.3	90.6	65.8	64.7	
EuroSAT	36.3	94.5	76.3	65.1	84.5	71.6	86.2	68.3	90.5	65.8	64.7	
Flowers	36.4	94.6	76.3	65.1	84.5	93.3	86.3	68.3	90.7	65.8	64.7	
Food	36.4	94.5	76.3	65.1	84.5	93.3	85.2	68.3	90.7	65.7	64.7	
MNIST	36.4	94.6	76.3	65.0	84.5	93.4	85.2	86.7	90.6	65.8	64.7	
OxfordPet	36.5	94.5	76.3	65.1	84.5	93.5	85.2	86.7	91.1	65.8	64.7	
Cars	36.4	94.6	76.3	65.1	84.5	93.3	85.1	86.7	91.2	77.9	64.7	
SUN397	36.5	94.4	76.3	65.0	84.5	93.4	85.2	86.7	91.2	77.9	72.7	78.5
Average	36.4	94.4	75.1	59.5	72.1	83.3	85.7	74.1	90.6	68.0	65.3	73.1

Table 18. Accuracy (%) of ours (LoRA) on the MTIL-FS setting with order-I. Each row represents the performance on every dataset of the model trained after the corresponding task. Transfer, Average, and Last metrics are shown in color.

	Aircraft	Caltech101	CIFAR100	DTD	EuroSAT	Flowers	Food	MNIST	OxfordPet	Cars	SUN397	
Transfer		93.4	69.9	45.4	51.9	71.9	86.5	62.1	90.8	65.4	64.8	70.2
Aircraft	37.9	93.4	69.5	45.3	49.7	71.7	86.4	61.9	90.1	65.2	63.8	
Caltech101	38.0	95.3	70.3	45.9	50.8	71.7	86.5	61.1	90.3	65.3	64.6	
CIFAR100	38.0	95.3	77.8	44.8	53.9	72.2	86.4	62.3	90.8	65.4	65.1	
DTD	38.0	95.3	77.8	63.2	53.3	71.9	86.4	62.4	90.9	65.7	64.9	
EuroSAT	38.0	95.2	77.8	63.1	81.7	72.2	86.5	62.4	90.8	65.6	64.9	
Flowers	38.0	95.3	77.8	63.2	81.7	94.6	86.6	62.3	91.1	65.5	65.0	
Food	38.0	95.2	77.8	63.2	81.7	94.7	86.4	62.4	91.3	65.5	65.0	
MNIST	38.0	95.2	77.8	63.2	81.7	94.7	86.3	93.4	91.2	65.4	64.9	
OxfordPet	38.0	95.3	77.8	63.1	81.7	94.6	86.3	93.3	91.2	65.4	64.9	
Cars	38.0	95.3	77.8	63.1	81.7	94.7	86.3	93.4	91.2	78.0	65.0	
SUN397	38.1	95.2	77.8	63.2	81.7	94.6	86.3	93.4	91.2	78.0	73.4	79.4
Average	38.0	95.1	76.3	58.3	70.9	84.3	86.4	73.5	90.9	67.7	65.6	73.4

Table 19. Accuracy (%) of ours (RA) on the MTIL-FS setting with order-II. Each row represents the performance on every dataset of the model trained after the corresponding task. Transfer, Average, and Last metrics are shown in color.

	Cars	Food	MNIST	OxfordPet	Flowers	SUN397	Aircraft	Caltech101	DTD	EuroSAT	CIFAR100	Average
Transfer		86.0	64.9	89.8	71.5	65.3	24.7	93.1	44.5	48.0	69.6	65.8
Cars	77.8	86.0	65.7	89.6	71.5	65.9	24.9	93.0	44.3	46.9	69.2	
Food	77.8	85.0	64.2	89.8	71.7	65.0	24.7	92.7	44.4	46.1	68.9	
MNIST	77.8	85.0	86.3	89.8	71.7	65.1	24.7	92.6	44.4	47.0	69.8	
OxfordPet	77.8	85.0	86.3	91.8	71.2	65.1	24.6	92.7	44.5	47.0	69.8	
Flowers	77.8	85.0	86.3	91.8	94.3	65.4	24.5	92.7	44.5	46.9	69.8	
SUN397	77.8	85.0	86.3	91.9	94.4	72.9	24.5	94.2	44.8	49.9	69.6	
Aircraft	77.8	85.0	86.3	91.8	94.2	72.9	37.6	94.0	44.7	49.8	69.9	
Caltech101	77.8	85.0	86.3	91.8	94.3	72.9	37.6	94.2	44.7	49.8	69.9	
DTD	77.8	85.0	86.4	91.8	94.4	72.9	37.6	94.2	63.4	49.0	69.8	
EuroSAT	77.8	85.0	86.3	91.9	94.3	72.9	37.6	94.3	63.4	83.9	69.9	
CIFAR10	77.8	85.0	86.3	91.9	94.3	72.9	37.6	94.2	63.4	83.9	74.5	78.3
Average	77.8	85.1	82.4	91.3	86.0	69.5	30.5	93.5	49.7	54.6	70.1	71.9

Table 20. Accuracy (%) of ours (LoRA) on the MTIL-FS setting with order-II. Each row represents the performance on every dataset of the model trained after the corresponding task. Transfer, Average, and Last metrics are shown in color.

	Cars	Food	MNIST	OxfordPet	Flowers	SUN397	Aircraft	Caltech101	DTD	EuroSAT	CIFAR100	Average
Transfer		85.5	60.5	91.2	71.5	65.3	24.0	93.5	45.4	50.0	70.4	65.7
Cars	77.6	85.5	60.1	90.9	70.9	64.8	23.6	93.2	44.9	48.5	69.1	
Food	77.6	86.1	60.9	91.4	71.5	65.2	24.2	93.4	45.3	48.1	69.2	
MNIST	77.6	86.1	94.3	91.3	71.9	65.7	24.2	93.5	45.6	50.2	70.7	
OxfordPet	77.6	86.1	94.3	92.0	71.9	65.4	24.1	93.6	45.6	50.1	70.7	
Flowers	77.6	86.1	94.3	92.0	94.1	65.6	24.1	93.4	45.6	49.2	70.9	
SUN397	77.6	86.2	94.3	92.0	94.0	73.3	23.7	93.6	45.5	52.1	70.8	
Aircraft	77.6	86.1	94.3	92.0	94.0	73.3	39.9	93.6	45.4	52.1	70.7	
Caltech101	77.6	86.2	94.3	91.9	94.1	73.3	39.9	95.0	45.4	52.3	70.7	
DTD	77.6	86.1	94.4	92.0	94.0	73.3	39.8	95.0	62.9	47.0	70.7	
EuroSAT	77.6	86.1	94.3	91.9	94.0	73.3	39.8	95.0	62.9	86.0	70.8	
CIFAR100	77.6	86.1	94.3	92.0	94.0	73.3	39.9	95.0	62.9	86.0	76.7	79.8
Average	77.6	86.1	88.2	91.8	85.9	69.7	31.2	94.0	50.2	56.5	71.0	72.9

Table 21. Accuracy of ours (RA) on the X-TAIL setting with order-I. Each row represents the performance on every dataset of the model trained after the corresponding task. Transfer, Average, and Last metrics are shown in color.

	Aircraft	Caltech101	DTD	EuroSAT	Flowers	Food	MNIST	Pets	Cars	SUN397	
Transfer		75.5	38.4	47.6	70.7	85.2	44.2	88.9	65.9	63.2	64.4
Aircraft	53.3	75.5	37.9	47.2	70.7	85.3	43.6	89.3	66.0	62.2	
Caltech101	53.3	95.3	38.8	48.0	70.8	85.3	43.1	88.8	66.2	63.2	
DTD	53.4	95.3	74.4	44.2	70.5	85.2	45.1	88.7	66.1	63.5	
EuroSAT	53.4	95.2	74.4	97.9	70.5	85.1	44.7	88.6	66.0	63.4	
Flowers	53.4	95.2	74.4	97.9	96.7	85.2	44.6	88.9	65.7	63.3	
Food	53.4	95.3	74.4	97.9	96.7	89.3	45.2	89.1	65.5	63.4	
MNIST	53.4	95.2	74.4	97.9	96.7	89.3	99.3	89.1	65.6	63.4	
Pets	53.4	95.2	74.4	97.9	96.7	89.3	99.3	94.3	65.8	63.3	
Cars	53.4	95.3	74.4	97.9	96.7	89.3	99.3	94.3	81.9	63.4	
SUN397	53.4	95.3	74.4	97.9	96.7	89.3	99.3	94.3	82.0	76.3	85.9
Average	53.4	93.3	67.2	82.5	86.3	87.3	66.4	90.5	69.1	64.5	76.0

Table 22. Accuracy of ours (LoRA) on the X-TAIL setting with order-I. Each row represents the performance on every dataset of the model trained after the corresponding task. Transfer, Average, and Last metrics are shown in color.

	Aircraft	Caltech101	DTD	EuroSAT	Flowers	Food	MNIST	Pets	Cars	SUN397	
Transfer		75.7	38.7	47.4	70.4	85.2	45.9	88.8	65.1	62.8	64.4
Aircraft	55.1	75.7	38.4	47.1	70.3	85.3	43.7	89.0	65.8	61.7	
Caltech101	55.1	95.0	39.1	47.7	70.3	85.2	42.7	88.3	65.0	63.1	
DTD	55.1	95.0	75.5	44.5	70.5	85.2	43.5	88.6	65.0	63.1	
EuroSAT	55.2	94.9	75.5	97.9	70.7	85.2	49.9	88.6	65.0	63.0	
Flowers	55.2	95.0	75.5	97.9	97.6	85.2	49.8	88.9	65.2	62.7	
Food	55.1	94.9	75.5	97.9	97.6	89.3	50.4	89.1	64.8	62.7	
MNIST	55.1	94.9	75.6	97.9	97.6	89.3	99.3	89.2	64.8	62.9	
Pets	55.2	94.9	75.5	97.9	97.6	89.3	99.3	94.8	64.9	62.8	
Cars	55.1	94.9	75.5	97.9	97.6	89.3	99.3	94.7	84.1	62.8	
SUN397	55.2	94.9	75.5	97.9	97.7	89.3	99.3	94.8	84.1	77.4	86.6
Average	55.1	93.0	68.2	82.4	86.7	87.2	67.7	90.6	68.9	64.2	76.4

Table 23. Accuracy of ours (RA) on the X-TAIL setting with order-II. Each row represents the performance on every dataset of the model trained after the corresponding task. Transfer, Average, and Last metrics are shown in color.

	Cars	Aircraft	OxfordPet	Food	SUN397	MNIST	Flowers	DTD	Caltech101	EuroSAT	
Transfer		24.7	89.3	85.3	62.8	46.3	71.2	39.2	76.2	44.9	60.0
Cars	85.9	24.7	89.2	85.3	62.4	45.8	70.9	38.9	75.1	45.5	
Aircraft	85.9	48.3	89.3	85.3	62.9	46.5	70.9	38.8	76.0	46.6	
OxfordPet	85.9	48.2	94.0	85.5	63.0	46.3	71.6	39.0	76.1	46.2	
Food	85.9	48.2	94.1	89.0	63.3	46.7	71.3	39.2	76.4	46.3	
SUN397	85.9	48.2	94.1	89.0	76.6	47.5	71.3	39.6	76.5	46.3	
MNIST	85.9	48.3	94.1	89.1	76.6	99.3	71.3	39.8	76.4	46.1	
Flowers	85.9	48.0	94.0	89.0	76.6	99.3	97.2	40.0	76.6	45.6	
DTD	85.9	48.1	94.0	89.0	76.6	99.3	97.2	73.8	76.9	40.4	
Caltech101	85.9	48.1	94.0	89.0	76.6	99.3	97.2	73.7	93.1	40.8	
EuroSAT	86.0	48.2	94.1	89.0	76.6	99.3	97.2	73.8	93.1	98.1	85.5
Average	85.9	45.8	93.1	87.9	71.1	73.0	81.6	49.6	79.6	50.2	71.8

Table 24. Accuracy of ours (LoRA) on the X-TAIL setting with order-II. Each row represents the performance on every dataset of the model trained after the corresponding task. Transfer, Average, and Last metrics are shown in color.

	Cars	Aircraft	OxfordPet	Food	SUN397	MNIST	Flowers	DTD	Caltech101	EuroSAT	
Transfer		24.8	89.3	85.6	62.1	42.8	71.4	38.7	75.5	46.0	59.6
Cars	86.7	24.8	89.1	85.5	61.9	42.5	71.4	37.8	74.3	45.6	
Aircraft	86.7	50.1	89.5	85.7	62.2	42.6	71.4	38.4	75.0	46.5	
OxfordPet	86.6	50.1	94.6	85.7	62.3	42.4	71.5	38.5	75.2	46.3	
Food	86.7	50.1	94.6	89.2	62.8	43.5	71.5	38.7	75.8	46.4	
SUN397	86.6	50.1	94.5	89.2	77.6	47.3	71.3	39.3	76.1	47.2	
MNIST	86.7	50.2	94.5	89.2	77.6	99.5	71.3	39.4	76.1	47.0	
Flowers	86.7	50.1	94.6	89.2	77.6	99.5	97.9	39.4	76.0	46.8	
DTD	86.7	50.1	94.5	89.2	77.6	99.5	97.9	75.7	76.2	43.9	
Caltech101	86.7	50.1	94.6	89.2	77.6	99.5	97.9	75.7	94.8	43.9	
EuroSAT	86.7	50.0	94.6	89.2	77.6	99.5	97.9	75.8	94.8	98.2	86.4
Average	86.7	47.6	93.5	88.1	71.5	71.6	82.0	49.9	79.4	51.2	72.1