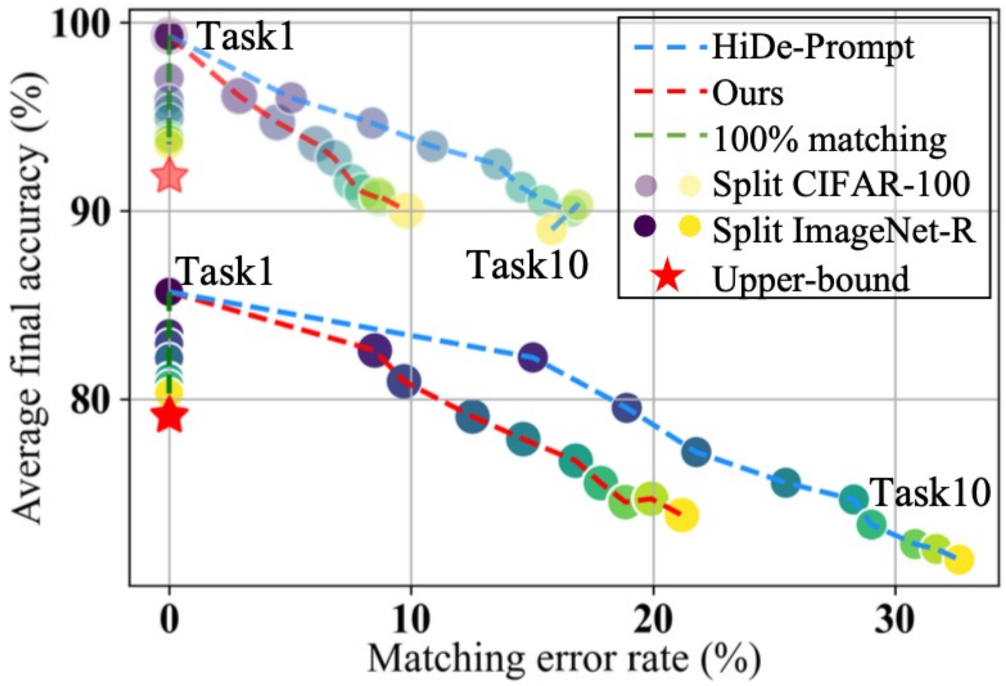
Hybrid Re-matching for Continual Learning with Parameter-efficient Tuning

NeurIPS2025

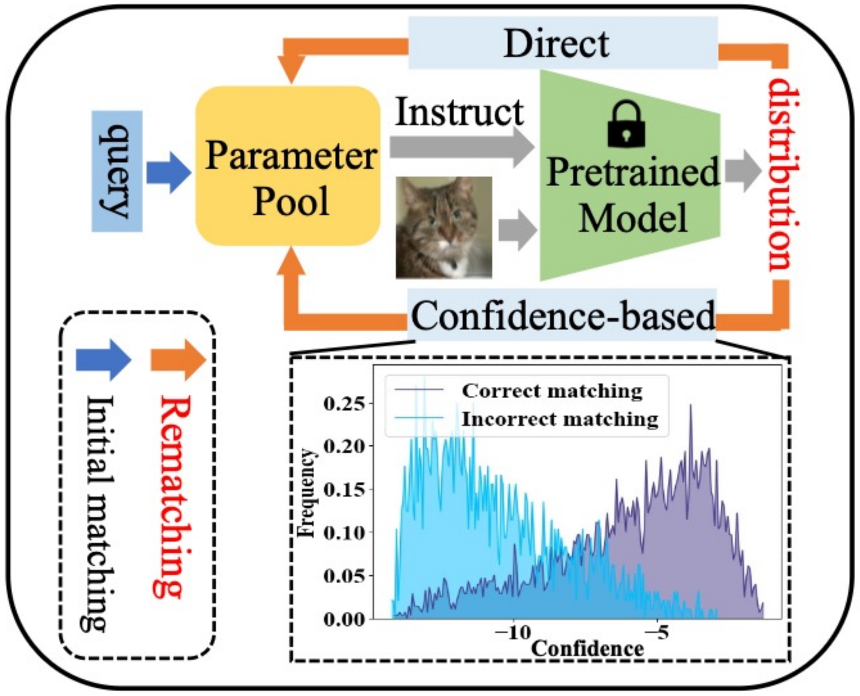
Weicheng Wang

南開大學
Nankai University

Motivation



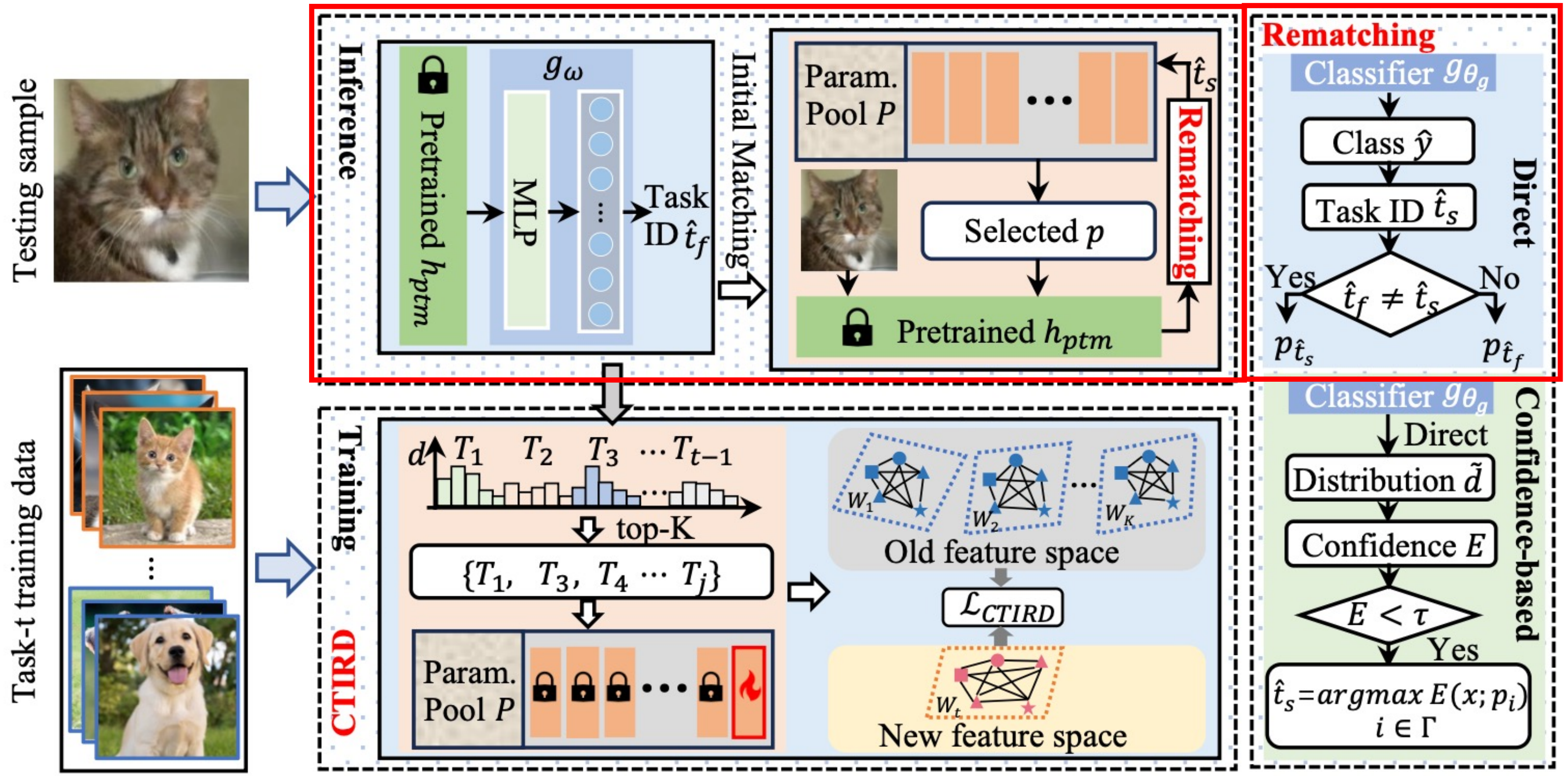
(a)



(b)

- Continual learning based on parameter-efficient tuning achieves superior performance without storing historical exemplars.
- However, reliance solely on pre-trained features for parameter matching exacerbates the inconsistency between the training and inference phases, thereby constraining the overall performance.
- Previous methods determine the task identity through a one-shot matching, which performs sub-optimal. Consequently, we propose a re-matching mechanism to calibrate the matching results.

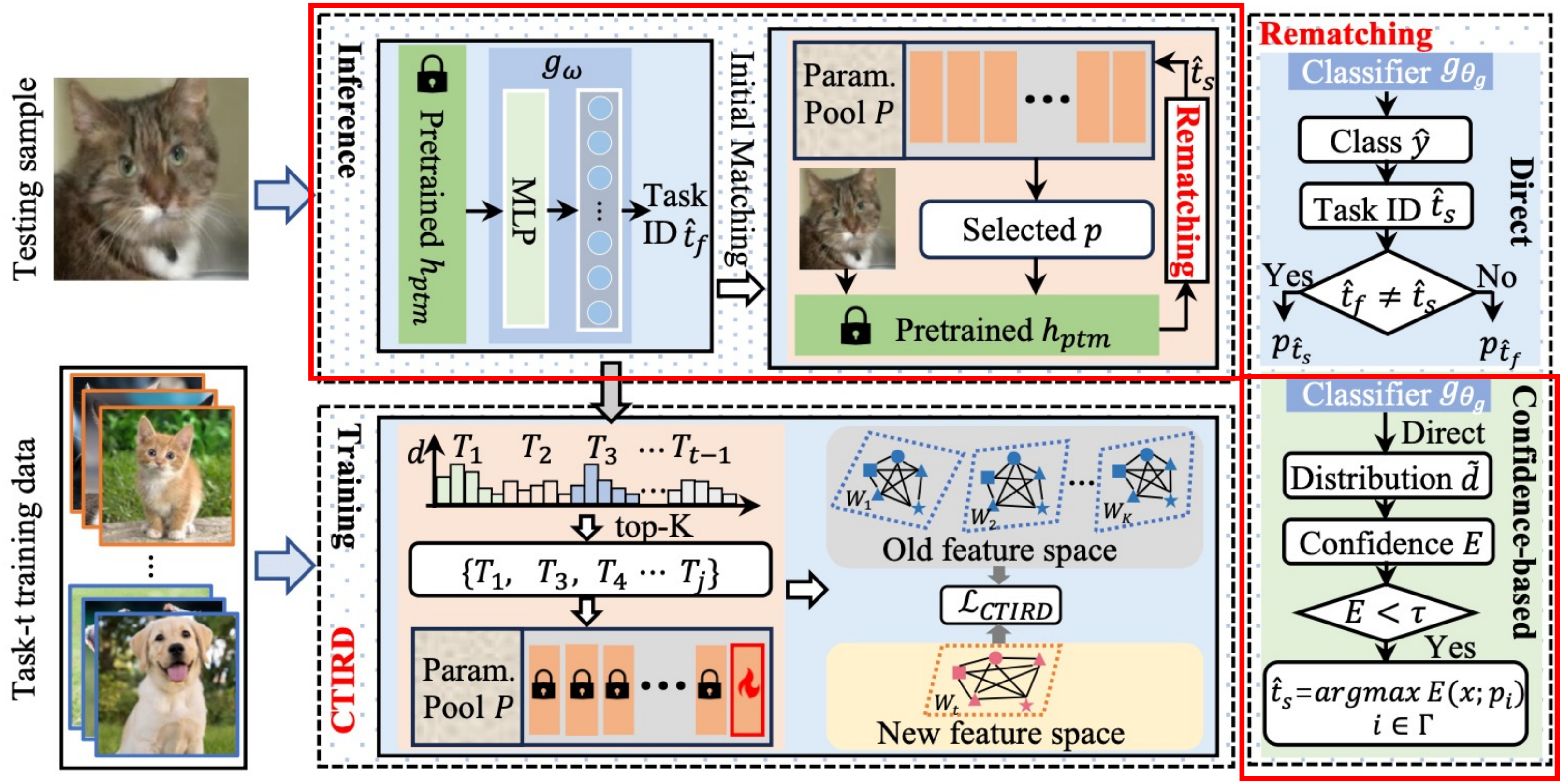
Direct Re-matching



- Owing to the shared knowledge in the parameter pool, the task identity of the predicted class for some samples may be correct, despite the incorrect initial matching.
- These samples can be detected when the predicted class does not belong to the matching task identity. The task identity is directly replaced with the one obtained from the prediction.

We propose a Hybrid Re-Matching Parameter-Efficient Tuning (HRM-PET) method, which contains two re-matching operations.

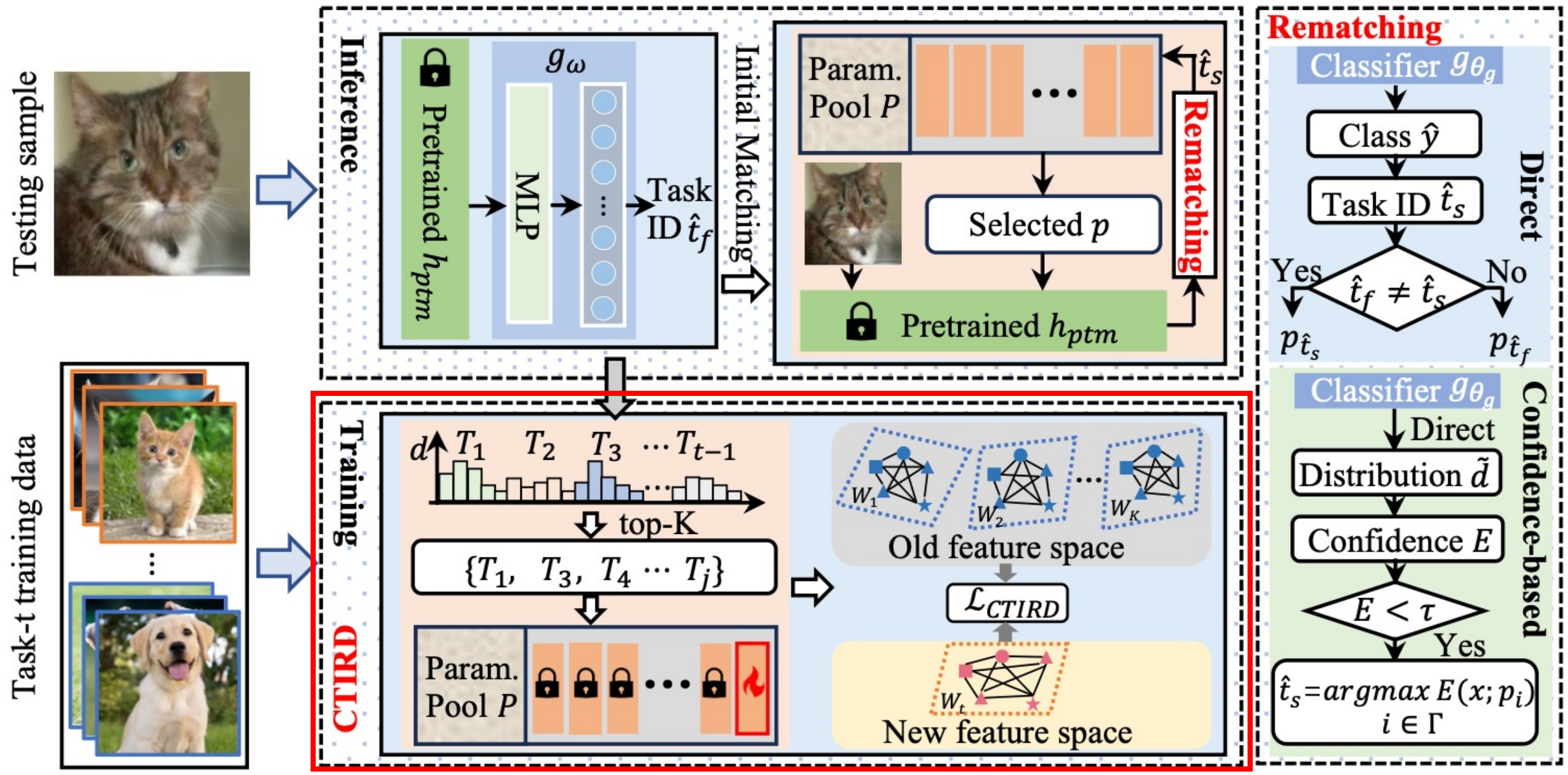
Confidence-based Re-matching



- For the parameters and the class of data mismatched, they have never been trained together. Therefore, the confidence of these predictions may be lower than the correct matching.
- We detect incorrect matching based on the confidence of the predicted distribution, and obtain the corrected task identity from the class with the sub-highest score.

We design confidence-based re-matching for all samples with incorrect initial matching.

Cross-Task Instance Relationship Distillation



- We aspire to directly re-match more samples in the first scenarios. Therefore, we integrate cross-task instance relationship distillation into the PET-based methods, which encourages aligning the shared feature space from different task parameters and improves the task-invariant knowledge.

We further introduce cross-task instance relationship distillation to align the shared feature space, which promotes the learning of task-invariant knowledge.

Comparative experiments

PTM	Method	Split CIFAR-100		Split ImageNet-R		ImageNet-A		5-Datasets	
		$A_N \uparrow$	$F_N \downarrow$	$A_N \uparrow$	$F_N \downarrow$	$A_N \uparrow$	$F_N \downarrow$	$A_N \uparrow$	$F_N \downarrow$
Sup-21K	DualPrompt [88]	86.60±0.19	4.45±0.16	68.79±0.31	4.49±0.14	39.76±0.18	5.85±1.14	76.09±0.91	15.70±1.13
	S-Prompt++ [86]	88.81±0.18	3.87±0.05	69.68±0.12	3.29±0.05	39.10±0.30	5.44±0.98	83.38±0.29	4.11±0.29
	CODA-Prompt [73]	86.94±0.63	4.04±0.18	70.03±0.47	5.17±0.22	30.96±0.14	5.58±0.45	62.54±1.34	18.58±1.67
	LAE [19]	85.10±0.24	5.01±0.35	72.25±0.62	3.43±0.65	40.36±1.26	9.33±0.45	64.16±0.10	10.25±0.22
	CPrompt [21]	83.18±1.89	6.98±1.56	69.24±1.31	5.42±1.78	37.60±1.01	8.12±0.99	71.72±0.65	18.87±1.12
	InfLoRA [47]	85.28±0.10	4.28±0.91	72.24±0.54	**3.20±1.27**	42.12±1.97	6.01±1.71	71.97±0.67	12.23±1.05
	HiDe-LoRA [80]	88.43±0.38	**3.16±0.16**	71.90±0.22	4.33±0.24	42.18±0.10	6.15±0.17	93.25±0.03	1.04±0.03
	HRM-PET (Ours)	**89.45±0.23**	3.83±0.13	**73.86±0.14**	3.60±0.15	**44.28±0.12**	**5.41±0.32**	**93.99±0.12**	**0.61±0.14**
iBOT-21K	DualPrompt [88]	78.76±0.23	9.84±0.24	54.55±0.53	5.38±0.70	21.60±1.70	6.92±0.84	67.60±1.47	26.24±1.11
	S-Prompt++ [86]	79.14±0.65	9.17±1.33	55.16±0.83	4.07±0.16	23.89±0.86	5.86±0.68	71.46±2.94	19.38±4.28
	CODA-Prompt [73]	80.83±0.27	7.50±0.25	61.22±0.35	9.66±0.20	28.10±0.28	6.33±0.76	56.12±3.18	27.23±2.51
	LAE [19]	78.87±0.58	10.55±0.40	60.68±0.12	4.67±0.87	29.76±0.16	9.77±1.06	63.86±1.00	13.75±1.12
	CPrompt [21]	82.34±0.81	6.98±1.16	64.64±0.87	7.06±0.71	31.66±0.61	10.26±0.79	67.38±1.90	16.48±0.81
	InfLoRA [47]	87.95±0.39	4.88±0.46	70.13±0.24	4.90±0.37	33.87±1.55	9.04±1.46	75.64±0.17	8.83±0.59
	HiDe-LoRA [80]	88.44±0.17	4.53±0.34	73.40±0.22	3.77±0.41	38.55±0.43	7.49±0.19	93.53±0.31	1.69±0.10
	HRM-PET (Ours)	**89.70±0.13**	**3.75±0.11**	**75.23±0.21**	**3.68±0.17**	**40.88±0.32**	**4.98±0.61**	**94.38±0.11**	**1.07±0.15**
iBOT-1K	DualPrompt [88]	76.63±0.05	8.41±0.40	61.51±1.05	5.02±0.52	21.60±0.39	6.43±0.64	65.35±1.26	27.62±1.75
	S-Prompt++ [86]	77.53±0.56	8.07±0.97	60.82±0.68	4.16±0.14	21.12±1.12	6.52±0.37	71.71±1.50	6.64±0.14
	CODA-Prompt [73]	79.11±1.02	7.69±1.57	66.56±0.68	7.22±0.38	28.69±0.73	6.34±1.23	44.34±3.33	21.89±4.06
	LAE [19]	75.45±0.43	10.55±0.36	67.95±0.55	5.56±0.41	27.18±0.21	9.77±0.67	67.00±1.46	13.75±0.98
	CPrompt [21]	76.32±0.79	12.49±1.36	68.25±1.43	8.25±1.19	29.70±0.91	8.23±1.11	75.21±0.84	9.86±1.12
	InfLoRA [47]	85.51±0.10	6.28±0.16	71.90±0.10	4.91±2.70	32.01±0.70	7.45±0.67	75.33±0.61	9.12±0.48
	HiDe-LoRA [80]	86.43±0.23	5.23±0.34	72.79±0.28	4.39±0.16	35.63±0.25	6.25±0.13	93.48±0.06	1.70±0.03
	HRM-PET (Ours)	**87.11±0.15**	**4.57±0.14**	**74.64±0.28**	**3.92±0.19**	**37.89±0.34**	**5.19±0.41**	**94.15±0.15**	**0.87±0.20**
DINO-1K	DualPrompt [88]	74.90±0.21	10.26±0.62	58.57±0.45	5.80±0.21	21.27±0.52	5.16±0.81	68.21±1.11	24.04±0.86
	S-Prompt++ [86]	74.97±0.46	7.78±0.66	57.64±0.16	5.08±0.31	20.58±1.52	6.22±0.68	75.19±0.80	6.76±1.76
	CODA-Prompt [73]	77.50±0.64	8.10±0.01	63.15±0.39	6.86±0.11	24.28±0.56	6.66±1.22	51.52±2.10	27.55±3.01
	LAE [19]	73.62±0.44	12.56±0.29	63.70±0.84	5.44±0.24	23.10±0.38	8.68±1.01	65.33±1.55	10.25±1.23
	CPrompt [21]	74.64±1.69	11.79±1.99	63.67±0.57	9.85±0.91	26.49±0.81	7.80±0.87	73.40±0.95	17.67±1.41
	InfLoRA [47]	82.37±0.65	6.40±0.33	68.51±0.20	5.01±0.69	31.67±0.28	5.12±0.49	79.09±0.23	11.23±0.05
	HiDe-LoRA [80]	84.85±0.21	5.36±0.22	70.42±0.22	4.60±0.19	31.72±0.33	4.98±0.34	93.25±0.04	1.34±0.01
	HRM-PET (Ours)	**85.93±0.14**	**5.14±0.24**	**72.32±0.20**	**3.49±0.27**	**33.37±0.35**	**4.78±0.55**	**93.90±0.05**	**0.67±0.07**
MoCo-1K	DualPrompt [88]	77.77±0.68	6.61±1.08	52.57±0.82	**2.73±0.49**	18.07±1.03	**4.11±0.65**	68.17±0.35	23.56±0.46
	S-Prompt++ [86]	76.30±0.54	14.67±0.64	53.15±1.10	4.11±1.84	17.96±0.66	4.50±1.96	71.75±0.57	13.60±0.73
	CODA-Prompt [73]	76.83±0.34	12.60±0.02	55.75±0.26	10.46±0.04	13.75±0.21	8.34±0.19	52.81±2.41	33.58±2.99
	LAE [19]	78.31±0.32	15.24±0.19	54.11±0.16	10.21±0.34	21.42±0.15	12.43±1.11	57.44±0.18	17.11±0.36
	CPrompt [21]	77.38±0.99	12.24±1.56	59.47±2.56	8.61±1.85	22.16±0.91	6.94±0.69	73.17±0.84	17.67±1.34
	InfLoRA [47]	82.70±0.39	8.38±0.66	65.03±0.87	5.69±1.10	26.00±0.81	6.50±0.64	74.00±0.38	10.49±0.42
	HiDe-LoRA [80]	85.37±0.21	5.42±0.19	68.01±0.59	4.57±0.86	28.30±0.47	6.74±0.35	92.72±0.07	1.71±0.06
	HRM-PET (Ours)	**86.00±0.14**	**5.32±0.19**	**69.32±0.22**	4.10±0.35	**29.73±0.52**	6.23±0.61	**93.31±0.08**	**1.23±0.15**

- To evaluate the effectiveness, we perform extensive comparative experiments with 7 state-of-the-art PET-based methods. This comparison involves five pre-trained models across four datasets.

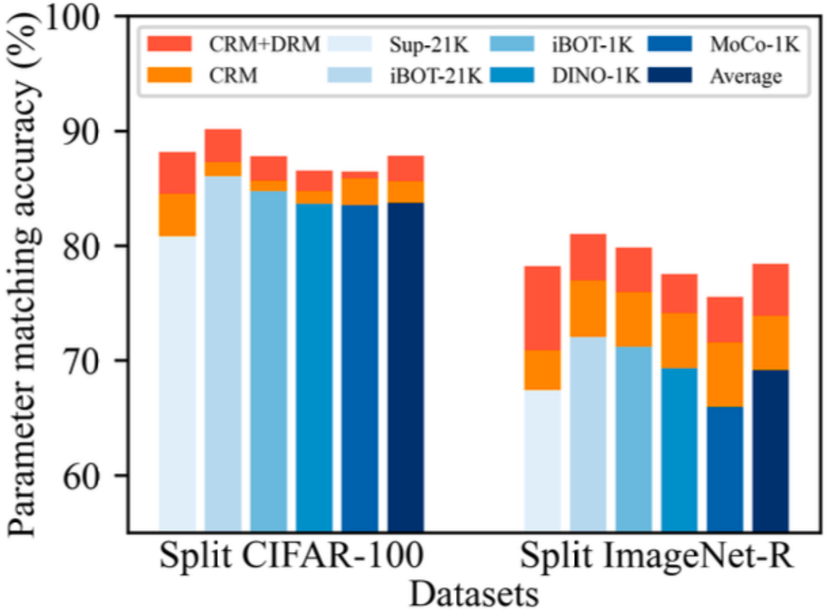
- We empirically validate the effectiveness of HRM-PET on extensive datasets using five distinct pre-trained models.

Ablation experiments

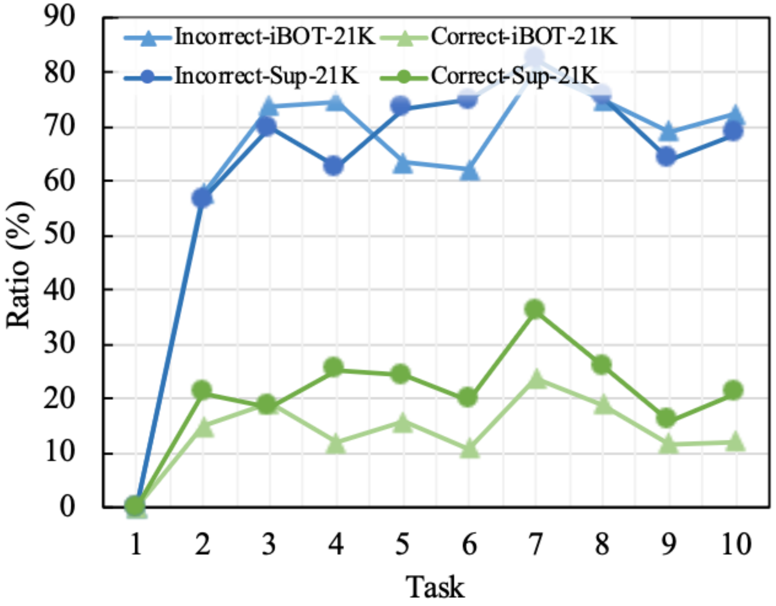
Nankai University 南開大學

Method	Sup-21K	iBOT-21K	iBOT-1K	DINO-1K	MoCo-1K
Baseline	71.47±0.26	73.49±0.31	72.99±0.56	70.40±0.62	67.89±0.29
Baseline+DRM	72.55±0.10	74.11±0.22	73.48±0.35	71.05±0.27	68.70±0.42
Baseline+CRM	72.80±0.15	74.35±0.26	74.01±0.30	71.32±0.23	68.80±0.20
Baseline+DRM+CRM	73.38±0.21	74.80±0.15	74.18±0.18	71.85±0.45	69.10±0.15
Baseline+DRM+CRM+CTIRD	73.86±0.14	75.23±0.21	74.64±0.28	72.32±0.20	69.32±0.22

- Ablation experiment to prob DRM, CRM, and CTIRD in ImageNet-R.



- Parameter matching accuracy



- Proportion of detected incorrect and correct matching

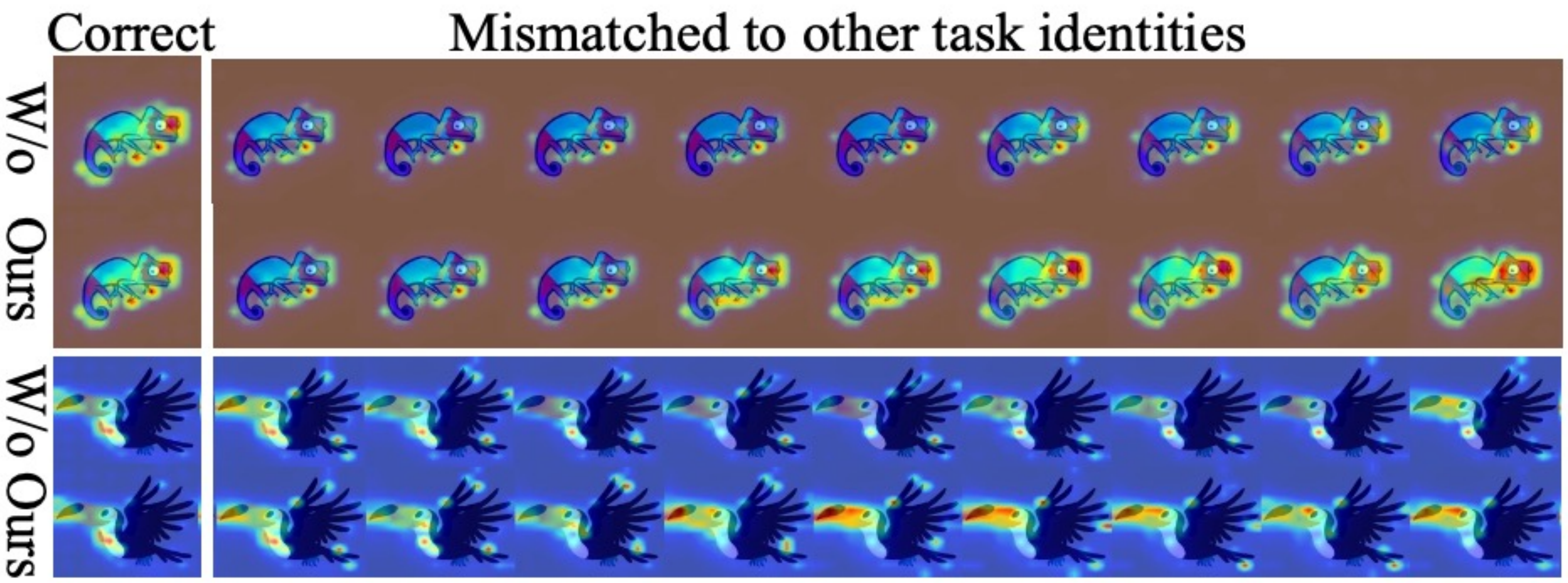
Ablation experiments

Method	CIFAR-100		ImageNet-R	
	$A_N \uparrow$	Time↓	$A_N \uparrow$	Time↓
Baseline	88.40	2.69	71.60	2.81
Baseline+DRM	88.80	2.81	72.60	2.91
Baseline+CRM+DRM	89.45	3.01	73.86	3.24

- The additional computational time during the inference process.

Distillation	CIFAR-100		ImageNet-R	
	Sup-21.	iBOT-21.	Sup-21.	iBOT-21.
Logits [45]	87.66	88.02	72.88	74.51
Features [105]	87.40	83.72	71.15	71.64
IRD* [7]	88.59	88.98	73.21	74.58
CTIRD	**89.45**	**89.70**	**73.86**	**75.23**

- Comparison of different knowledge distillation methods



- Visualization of attention regions with (Ours) and without CTIRD (W/o).

Thank you