Improving Cross-Domain Few-Shot Classification with Multilayer Perceptron

Shuanghao Bai [1], Wanqi Zhou [1], Zhirong Luan [2], Donglin Wang [3*], Badong Chen [1*]

{baishuanghao, zwq785915792}@stu.xjtu.edu.cn; chenbd@mail.xjtu,edu.cn; luanzhirong@xaut.edu.cn; wangdonglin@westlake.edu.cn;

1. Xi'an Jiaotong University

2. Xi'an University of Technology

3. Westlake University

Background

Motivation The introduction of MLP projector after the encoder:

Unsupervised Learning

- SimCLR and MoCo v2: MLP is adopted after the encoder to improve the models' transferability.

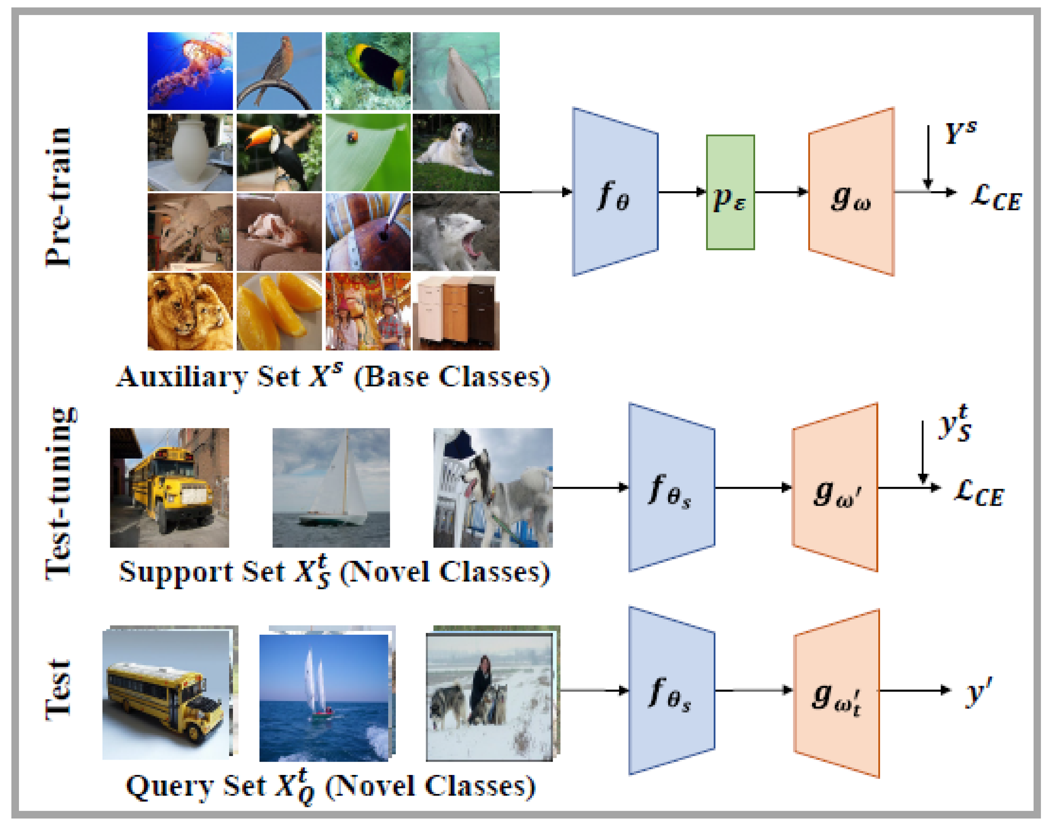
Supervised Learning

- SupCon and a study of SupCon: extend the self-supervised batch contrastive approach to the fully-supervised setting, including the MLP after the encoder.

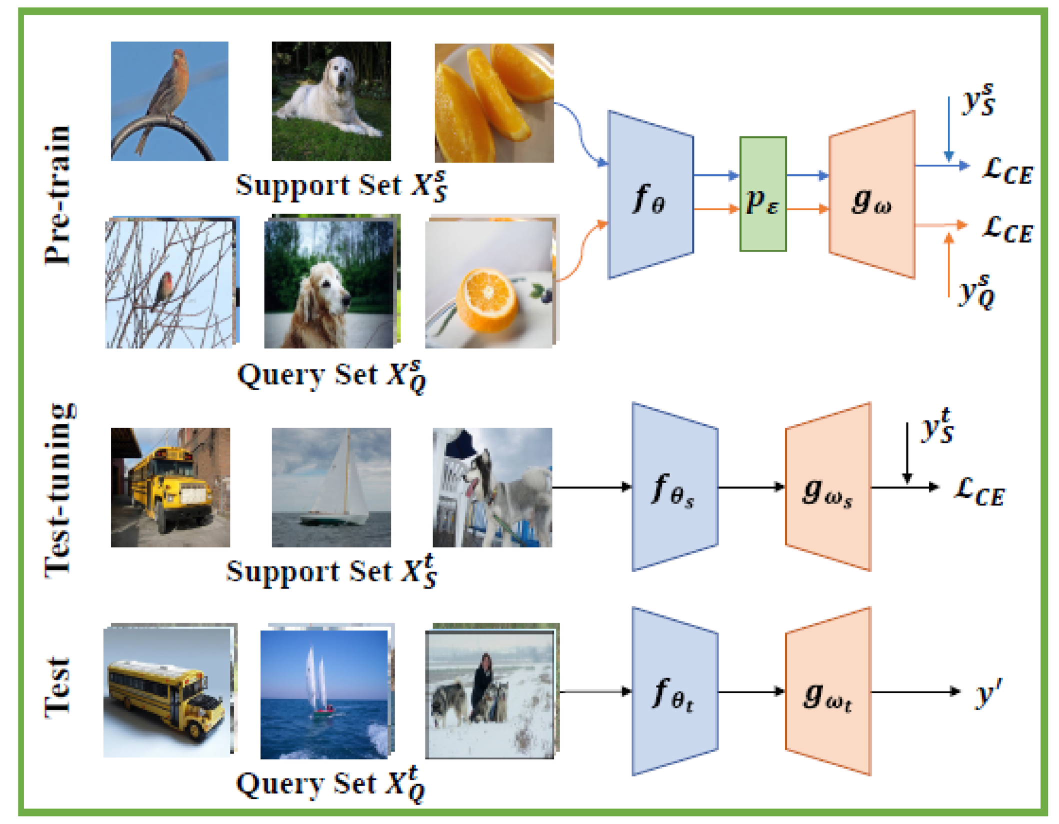
However, Wang et al. argue that previous works overlooked the ablation of the MLP and incorrectly attributed the enhanced transfer performance solely to the contrastive mechanism within the loss function. They demonstrate the effectiveness of MLP.

Paper Code

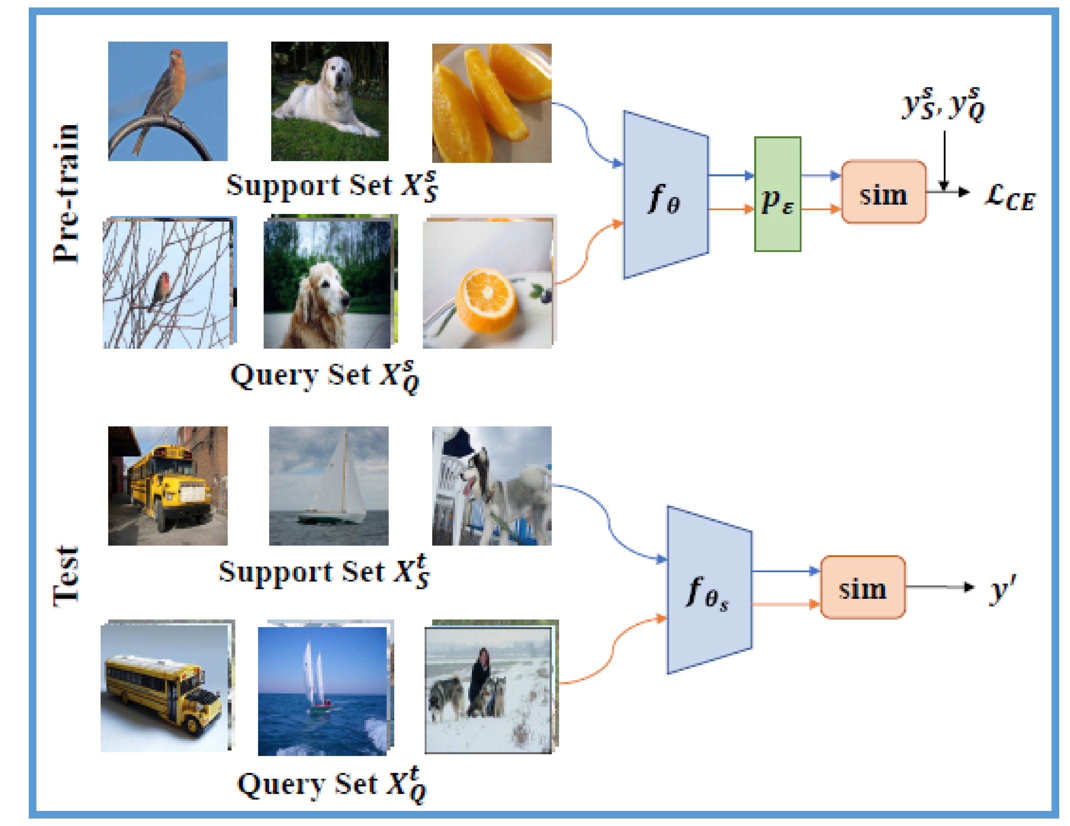
Approach



a) Paradigm for non-episodic based methods with MLP



b) Paradigm for meta-learning based methods with MLP



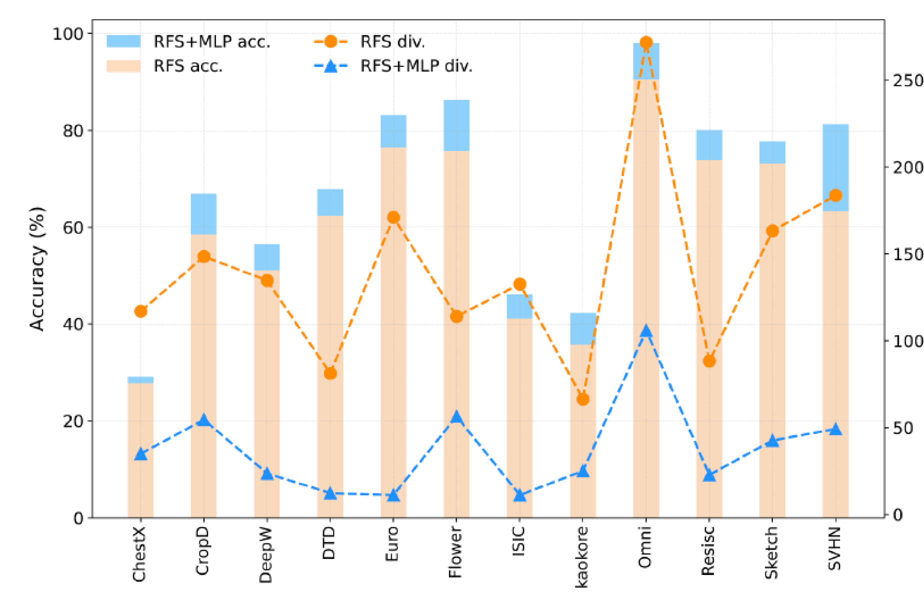
c) Paradigm for metric-learning based methods with MLP

- We initiate the first known and comprehensive effort to study MLP in CDFSC, and further introduce three distinct frameworks in accordance with three types of few-shot classification methods to verify the effectiveness of MLP.
- We empirically demonstrate that MLP helps existing few-shot classification algorithms significantly improve cross-domain generalization performance on 12 datasets and even compare favorably against state-of-the-art CDFSC algorithms.
- Our analyses indicate that MLP helps obtain better discriminative ability and mitigate the distribution shift. Additionally, we find that batch normalization plays the most crucial role in improving transferability.

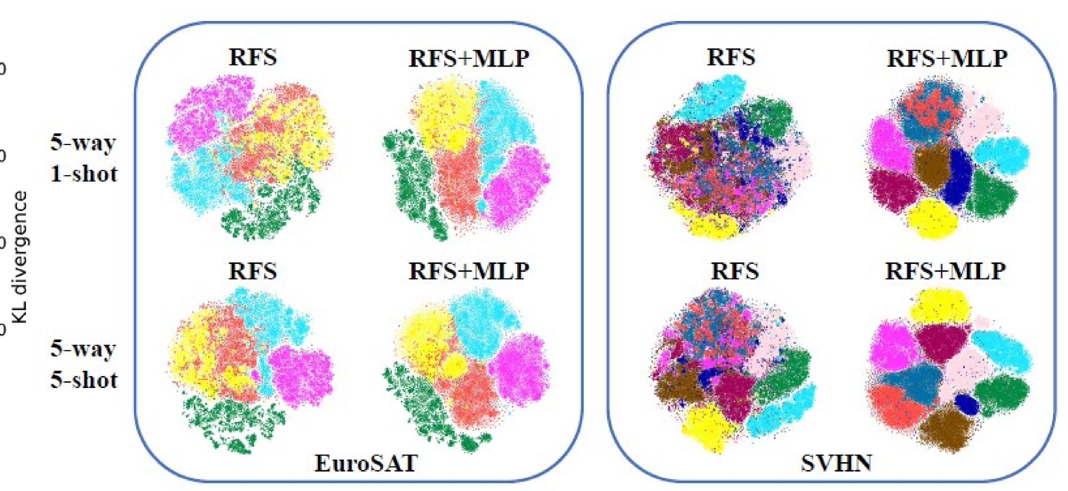
Experimental Results

Type	Method	ChestX	CropD	DeepW	DTD	Euro	Flower	ISIC	Kaokore	Omni	Resisc	Sketch	SVHN	Average
Non-episodic	BL [11]	26.98	65.44	55.27	59.58	83.81	77.59	42.69	37.24	96.24	71.34	72.60	77.49	63.86
	BL+MLP	28.62	70.48	57.54	61.96	84.26	81.49	42.55	40.06	98.12	75.62	75.15	84.19	**66.67**+2.81
	BL++ [11]	25.49	48.22	50.86	51.76	75.79	66.15	40.73	32.64	87.87	63.16	63.72	64.01	55.87
	BL+++MLP	28.09	65.62	57.65	64.33	85.83	81.47	44.32	38.14	97.47	74.76	75.85	82.30	**66.32**+10.45
	RFS [6]	27.68	58.38	51.02	62.31	76.33	75.62	41.05	35.70	90.55	73.82	73.05	63.22	60.73
	RFS+MLP	29.09	66.87	56.42	67.78	83.14	86.24	46.02	42.31	97.95	80.04	77.64	81.27	**67.90**+7.17
Meta-learning	ANIL [8]	24.41	48.69	46.93	46.55	63.96	61.27	37.57	31.50	84.53	58.92	61.90	58.29	52.04
	ANIL+MLP	25.02	58.10	49.35	53.30	75.73	66.45	39.19	32.61	83.02	64.62	61.70	51.97	**55.09**+3.05
	MTL [16]	24.15	33.27	43.14	49.43	54.27	58.18	35.56	31.36	72.77	57.93	60.53	52.72	47.78
	MTL+MLP	25.19	51.23	46.06	49.41	65.19	51.34	34.74	31.59	78.84	53.97	51.62	52.01	**49.27**+1.49
Metric-learning	PN [10]	26.38	55.59	47.50	50.55	70.96	63.56	32.95	33.58	92.68	58.70	54.53	64.38	54.28
	PN+MLP	27.07	60.76	47.83	50.49	73.22	62.63	33.76	32.00	87.06	59.02	51.85	66.69	**54.37**+0.09
	DN4 [17]	27.34	53.62	50.94	58.67	76.52	75.01	42.68	37.67	97.43	70.30	72.81	84.76	62.31
	DN4+MLP	28.38	67.48	56.40	62.15	82.61	80.78	39.41	40.15	98.17	75.78	75.77	88.18	**66.27**+3.96
	CAN [18]	27.46	52.82	53.14	56.85	73.77	71.40	42.32	36.55	83.03	69.96	66.03	58.32	57.64
	CAN+MLP	28.57	67.63	58.93	64.31	83.66	79.81	42.73	38.00	97.08	75.68	73.64	71.54	**65.13**+7.49

- All few-shot classification methods with MLP outperform the vanilla methods.



- MLP can mitigate the distribution shift between the pre-training and evaluation datasets



Method	RFS	RFS+MLP	RFS	RFS+MLP
Dataset	ChestX		CropD	
D_1 (↓)	6.85	**6.78**	6.37	**5.83**
V (↓)	2.73	**2.69**	2.64	**2.61**
r (↓)	1.21	**1.20**	0.85	**0.80**
Dataset	Euro		ISIC	
D_1 (↓)	3.83	**3.57**	6.09	**5.66**
V (↓)	1.73	**1.51**	2.35	**2.05**
r (↓)	0.49	**0.43**	1.04	**0.92**

- MLP help the model obtain better discriminative ability of cluster compactness.
- T-SNE visualization.
- The lower intra-class distance, intra-class variance, and the ratio between the average intra-class distance and inter-class distance.

ICASSP 2024 KOREA

IEEE International Conference on Acoustics, Speech and Signal Processing