Cross-Domain Cross-Set Few-Shot Learning via Learning Compact and Aligned Representations

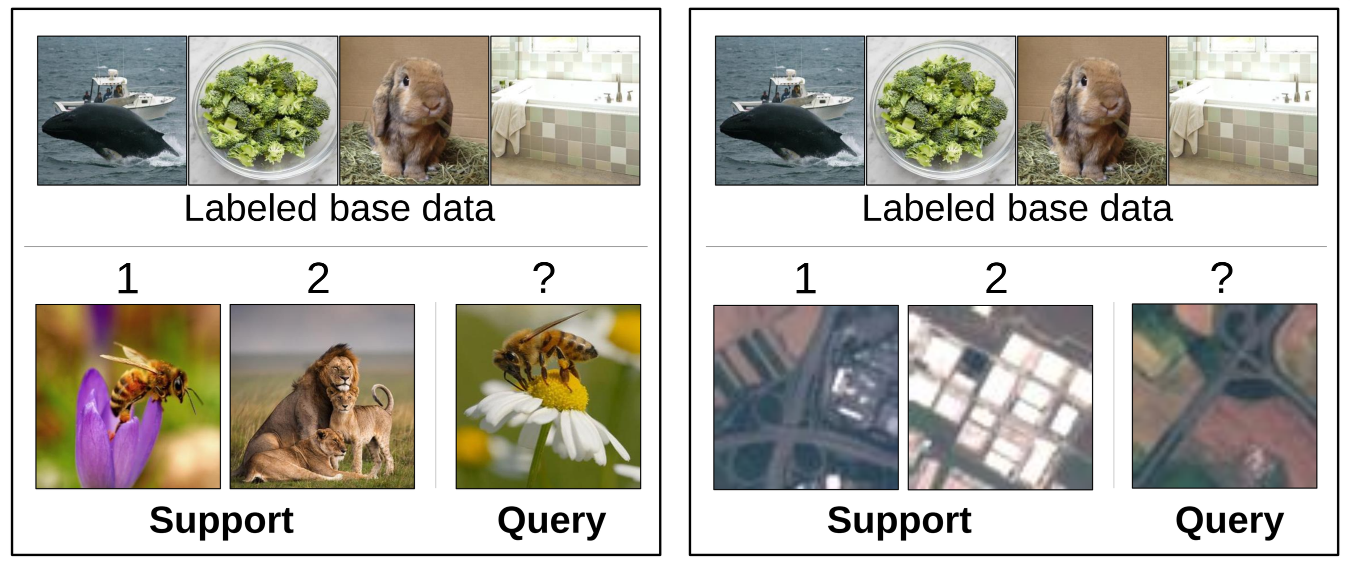
ECCV TEL AVIV 2022

Wentao Chen[1,2] Zhang Zhang[2] Wei Wang[2] Liang Wang[2] Zilei Wang[1] Tieniu Tan[1,2]

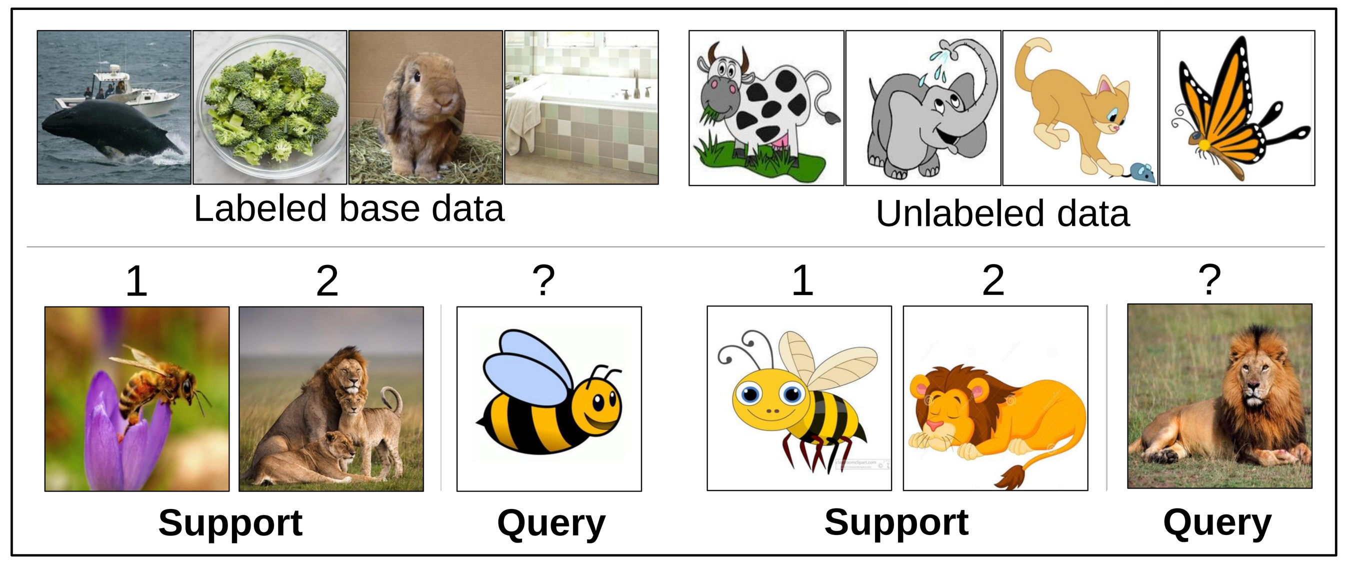
[1]University of Science and Technology of China, Hefei, China
[2]Center for Research on Intelligent Perception and Computing, NLPR, CASIA, Beijing, China

Problem Setup

Few-Shot Learning (FSL) **Cross-Domain FSL (CD-FSL)**



Cross-Domain Cross-Set FSL (CDCS-FSL)



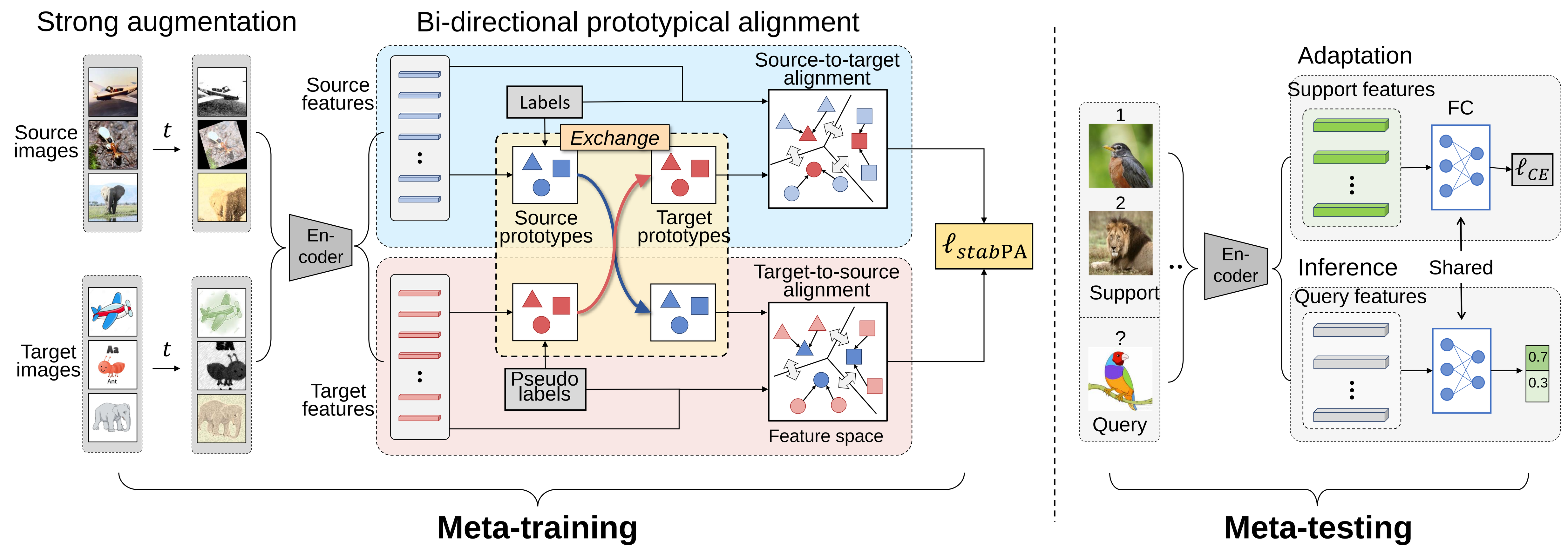
- FSL aims to recognize novel queries based on a few support images with the help of a large base dataset for pre-training/meta-training. All images are assumed to derive from a single domain.
- CD-FSL extends FSL by assuming a domain gap between the base classes and the novel classes.
- We propose CDCS-FSL to further address the domain gap within the novel classes (i.e., between the support and the query). Unlabeled data from the new domain (aka. target domain) are allowed for meta-training.

Motivation

- To address CDCS-FSL, we first need an aligned feature space to alleviate the domain gap between two domains. (**Aligned**)
- Second, compact representations are desired to learn a center-clustered feature space, so that a small support set can better represent a new class. (**Compact**)

Framework

- **Strongly Augmented Bi-directional Prototypical Alignment (stabPA)**



- During **meta-training:** we perform stabPA on source and target domain images to learn a compact and aligned feature space.
- The key of **stabPA:** align samples in one domain to the prototypes in the other domain. The alignment is performed bi-directionally, such that the domain distance and the intra-class variance can be reduced simultaneously.
- During **meta-testing**: the learned feature encoder is fixed and a new FC head is learned on support images and tested on query images.

Benchmark

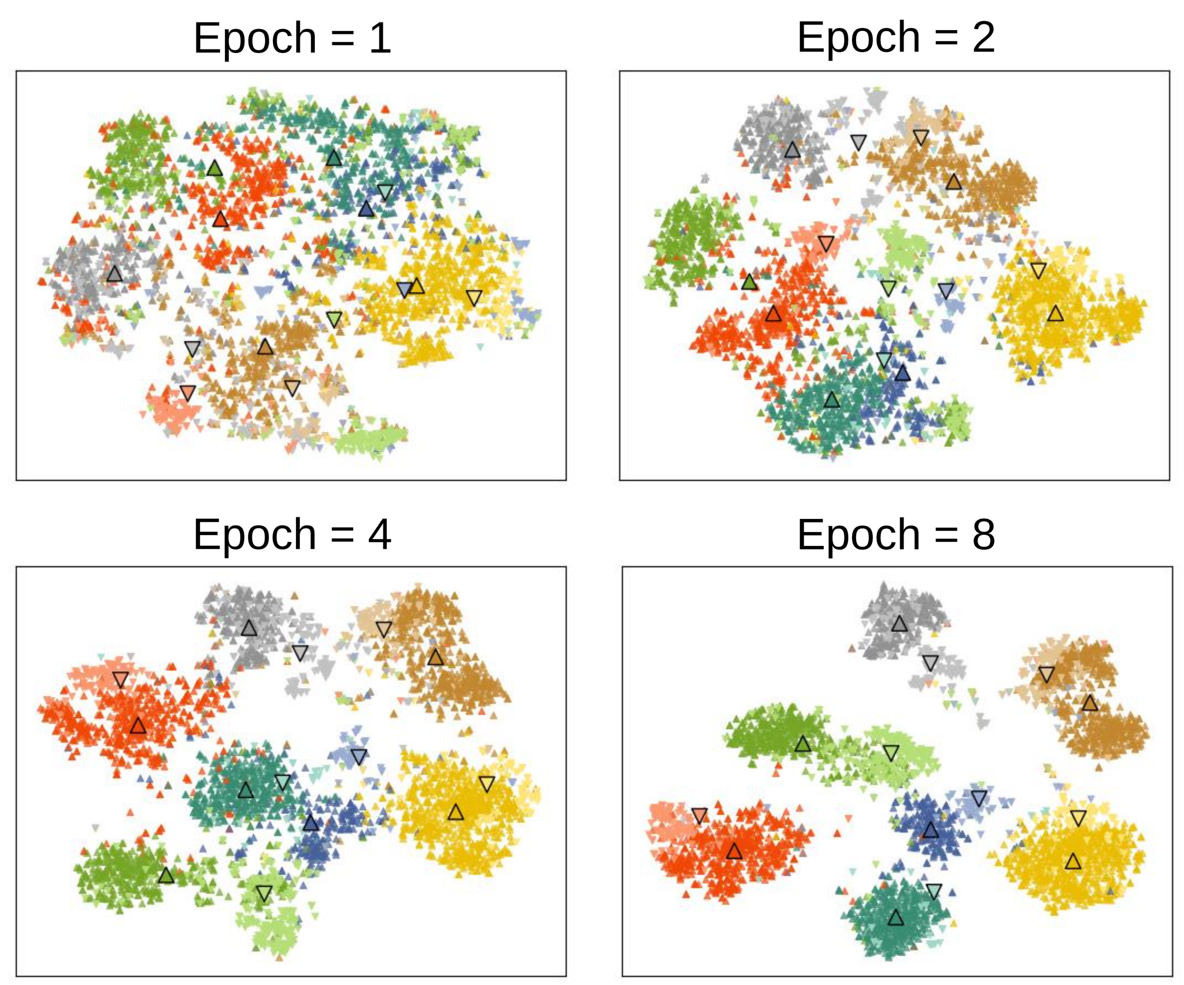
- **DomainNet**

Method	5-way 1-shot						
	r-r	r-p	p-r	r-c	c-r	r-s	s-r
ProtoNet [32]	63.43±0.90	45.36±0.81	45.25±0.97	44.65±0.81	47.50±0.95	39.28±0.77	42.85±0.89
RelationNet [34]	59.49±0.91	42.69±0.77	43.04±0.97	44.12±0.81	45.86±0.95	36.52±0.73	41.29±0.96
MetaOptNet [19]	61.12±0.89	44.02±0.77	44.31±0.94	42.46±0.80	46.15±0.98	36.37±0.72	40.27±0.95
Tian et al. [37]	67.18±0.87	46.69±0.86	46.57±0.99	48.30±0.85	49.66±0.98	40.23±0.73	41.90±0.86
DeepEMD [46]	67.15±0.87	47.60±0.87	47.86±1.04	49.02±0.83	50.89±1.00	42.75±0.79	46.02±0.93
ProtoNet+FWT [39]	62.38±0.89	44.40±0.80	45.32±0.97	43.95±0.80	46.32±0.92	39.28±0.74	42.18±0.95
ProtoNet+ATA [43]	61.97±0.87	45.59±0.84	45.90±0.94	44.28±0.83	47.69±0.90	39.87±0.81	43.64±0.95
S2M2 [23]	67.07±0.84	46.84±0.82	47.03±0.95	47.75±0.83	48.27±0.91	39.78±0.76	40.11±0.91
Meta-Baseline [6]	**69.46**±0.91	**48.76**±0.85	48.90±1.12	**49.96**±0.85	**52.67**±1.08	**43.08**±0.80	**46.22**±1.04
*stab*PA⁻ (*Ours*)	68.48±0.87	48.65±0.89	**49.14**±0.88	45.86±0.85	48.31±0.92	41.74±0.78	42.17±0.95
DANN [11]	-	45.94±0.84	46.85±0.97	47.31±0.86	50.02±0.94	42.44±0.79	43.66±0.92
PCT [35]	-	47.14±0.89	47.31±1.04	**50.04**±0.85	49.83±0.98	39.10±0.76	39.92±0.95
Mean Teacher [36]	-	46.92±0.83	46.84±0.96	48.48±0.81	49.60±0.97	43.39±0.81	44.52±0.89
FixMatch [33]	-	**48.86**±0.87	**49.15**±0.93	48.70±0.82	49.18±0.93	**44.48**±0.80	**45.97**±0.95
STARTUP [26]	-	47.53±0.88	47.58±0.98	49.24±0.87	**51.32**±0.98	43.78±0.82	45.23±0.96
DDN [18]	-	48.83±0.84	48.11±0.91	48.25±0.83	48.46±0.93	43.60±0.79	43.99±0.91
*stab*PA (*Ours*)	-	**53.86**±0.89	**54.44**±1.00	**56.12**±0.83	**56.57**±1.02	**50.85**±0.86	**51.71**±1.01

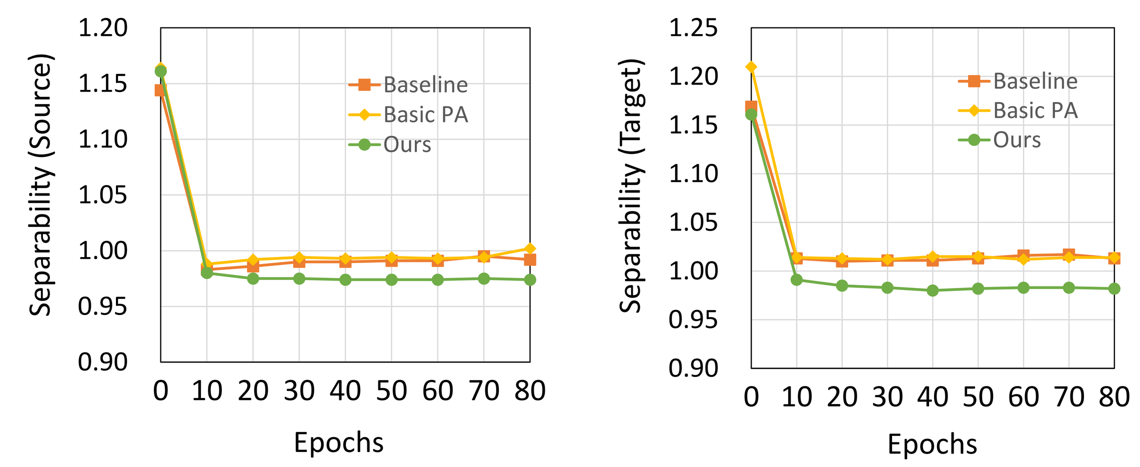
- We build a CDCS-FSL benchmark on the DomainNet dataset, where 'r-p' means the support is from 'real' and the query is from 'painting'.
- When domain shift occurs, the the performance of conventional FSL methods drops very fast (↓21.19% on average for the Meta-Baseline).
- Our method learning a compact and aligned feature space can effectively alleviate the domain shift problem and reduce the performance drop (↑5.66% compared with the Meta-baseline).

Visualization

- **t-SNE**



- **Class Separability (ADR)**



Ablation Study

ℓ_{s-t}	ℓ_{t-s}	aug	real-sketch 1-shot	real-sketch 5-shot	sketch-real 1-shot	sketch-real 5-shot
×	×	×	40.23±0.73	50.41±0.80	41.90±0.86	56.95±0.84
×	×	✓	41.74±0.78	51.03±0.85	42.17±0.95	57.11±0.93
✓	×	×	42.86±0.78	52.16±0.78	44.83±0.95	60.87±0.91
×	✓	×	44.20±0.77	54.83±0.79	44.45±0.92	61.97±0.90
✓	✓	×	47.01±0.84	56.68±0.81	47.59±1.00	64.32±0.86
✓	✓	✓	**50.85**±0.86	**61.37**±0.82	**51.71**±1.01	**68.93**±0.87

Table 1: Ablation studies on DomainNet with 95% confidence interval.

	FixMatch [33]	number of samples 10%	40%	70%	100%	number of base classes 0%	10%	40%	70%	100%
1-shot	47.72	51.76	52.97	53.42	53.92	50.74	51.59	52.48	53.24	53.92
5-shot	62.58	65.96	67.56	67.96	68.55	65.04	65.68	67.07	67.87	68.55

Table 2: The influence of the number of unlabeled data and the number of base classes that the unlabeled data contains. We report average accuracy on DomainNet over 6 situations.