Adaptive Weighted Co-Learning for Cross-Domain Few-Shot Learning

Abdullah Alchihabi*, Marzi Heidari*, Yuhong Guo*†
*Carleton University, Ottawa, Canada †Canada CIFAR AI Chair, Amii, Canada

Carleton University

BMVC 2024

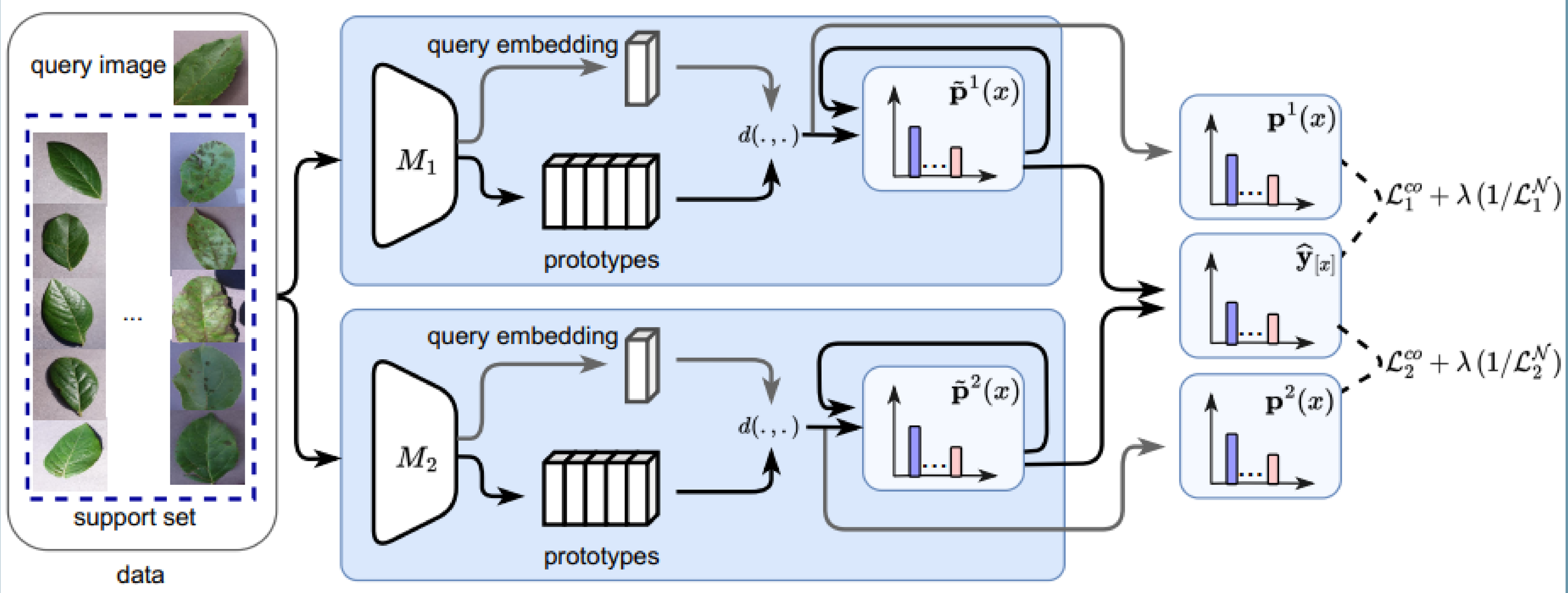
Problem Definition & Method Overview

Cross-Domain Few-Shot Learning (CDFSL)

- **Goal**: Train model in an annotation-rich source domain, then adapt it for a novel few-shot prediction task in target domain.
- Source and target domains have:
  - Different input feature distributions ($\mathcal{P}_s \neq \mathcal{P}_t$).
  - Disjoint output label classes ($\mathcal{Y}_s \cap \mathcal{Y}_t = \emptyset$).
- **Source Domain**: Large labeled dataset $\mathcal{D}_s = (x_i, y_i)_{i=1}^{N_s}$.
- **Target Domain**: Few-Shot task with
  - Support set $S = \{(x_i, y_i)\}_{i=1}^{N_s}$: N classes, K samples each.
  - Query set $Q = \{(x_i, y_i)\}_{i=1}^{N_q}$.
- **Challenges**:
  - Limited few-shot labeled instances for target task.
  - Significant domain shift between source and target domains.

Adaptive Weighted Co-Learning (AWCoL)

- AWCoL consists of 2 prototypical models pre-trained independently in source domain.
- In target domain, both models generate probabilistic predictions for query instances using a WMA strategy.
- Predictions from both models are combined to determine positive pseudo-labels, negative pseudo-labels, and adaptive weights.
- The models are then fine-tuned in alternating fashion using adaptive co-learning.



Source-Domain Pre-Training

- Pre-train 2 prototypical models on different N-way K-shot FSL tasks.
- Compute prototype embedding for class n using feature extractor f:

$$\mathbf{c}_n = \frac{1}{K} \sum_{(x,y) \in S_n} f_\theta(x)$$

- Predict class probability vector for each query instance x:

$$\mathbf{p}_j(x) = \frac{\exp(-d(f_\theta(x), \mathbf{c}_j))}{\sum_{n=1}^{N} \exp(-d(f_\theta(x), \mathbf{c}_n))}$$

- Train each model by minimizing cross-entropy loss over query set:

$$\mathcal{L}_{CE}(Q) = \sum_{(x,y) \in Q} \ell_{CE}(\mathbf{1}_y, \mathbf{p}(x))$$

- **Output:** Independently trained models, M_1 and M_2, with feature encoders f_1 and f_2.

Target Domain Fine-Tuning

Weighted Adaptive Co-Learning

- Integrate the predictions of the models into a co-prediction probability vector $\tilde{\mathbf{p}}^{co}(x)$:

$$\tilde{\mathbf{p}}^{co}(x) = \text{softmax}(\text{avg}(\tilde{\mathbf{p}}^1(x), \tilde{\mathbf{p}}^2(x)))$$

- Determine pseudo-labels for query instances:

$$\hat{y}_{[x]} = \text{argmax}_{n \in \{1,\cdots,N\}} \tilde{\mathbf{p}}_n^{co}(x)$$

- Calculate adaptive weight to mitigate pseudo-label noise:

$$w_{[x]} = \max_{n \in \{1,\cdots,N\}} \tilde{\mathbf{p}}_n^{co}(x)$$

- Update each model m in a given iteration by minimizing:

$$\mathcal{L}_m^{co} = \frac{1}{\sum_{x \in Q} w_{[x]}} \sum_{x \in Q} w_{[x]} \ell_{CE}(\hat{\mathbf{y}}_{[x]}, \mathbf{p}^m(x))$$

WMA based Probabilistic Prediction Generation

- Use WMA to generate probabilistic predictions from both models independently.
- Maintain a WMA prediction probability vector $\tilde{\mathbf{p}}^m(x)$ for each query instance, updated with new predictions $\mathbf{p}^m(x)$:

$$\tilde{\mathbf{p}}^m(x) = (1 - \alpha_m)\, \tilde{\mathbf{p}}^m(x) + \alpha_m\, \mathbf{p}^m(x)$$

- Apply a rectified annealing schedule to α_m to stabilize WMA:

$$\alpha_m = \max(\alpha_{\min}, \gamma \alpha_m)$$

Negative Pseudo-Label Regularization

- Use negative pseudo-labels to regularize the co-learning process.
- Generate negative pseudo-label:

$$\hat{y}_{[x]}^{\mathcal{N}} = \text{Rand}(\{1,..,N\} \setminus \{\hat{y}_{[x]}\})$$

- Maximize adaptive weighted cross-entropy loss for each model m:

$$\mathcal{L}_m^{\mathcal{N}} = \frac{1}{\sum_{x \in Q} w_{[x]}} \sum_{x \in Q} w_{[x]} \ell_{CE}\left(\hat{\mathbf{y}}_{[x]}^{\mathcal{N}}, \mathbf{p}^m(x)\right)$$

- Apply adaptive co-learning loss to fine-tune each model m:

$$\mathcal{L}_m = \mathcal{L}_m^{co} + \lambda\,(1/\mathcal{L}_m^{\mathcal{N}})$$

Comparison Results

- Classification accuracy on 8 target domain datasets for cross-domain 5-way 5-shot classification.

	ChestX	CropDisea.	ISIC	EuroSAT	Places	Planatae	Cars	CUB
MatchingNet*[34]	$22.40_{(0.70)}$	$66.39_{(0.78)}$	$36.74_{(0.53)}$	$64.45_{(0.63)}$	–	–	–	–
MAML*[5]	$23.48_{(0.96)}$	$78.05_{(0.68)}$	$40.13_{(0.58)}$	$71.70_{(0.72)}$	–	–	–	–
ProtoNet*[27]	$24.05_{(1.01)}$	$79.72_{(0.67)}$	$39.57_{(0.57)}$	$73.29_{(0.71)}$	$58.54_{(0.68)}$	$46.80_{(0.65)}$	$41.74_{(0.72)}$	$55.51_{(0.68)}$
MetaOpt*[16]	$22.53_{(0.91)}$	$68.41_{(0.73)}$	$36.28_{(0.50)}$	$64.44_{(0.73)}$	–	–	–	–
RelationNet†[28]	$24.07_{(0.20)}$	$72.86_{(0.40)}$	$38.60_{(0.30)}$	$65.56_{(0.40)}$	$64.25_{(0.40)}$	$42.71_{(0.30)}$	$40.46_{(0.40)}$	$56.77_{(0.40)}$
GNN†[25]	$23.87_{(0.20)}$	$83.12_{(0.40)}$	$42.54_{(0.40)}$	$78.69_{(0.40)}$	$70.91_{(0.50)}$	$48.51_{(0.40)}$	$43.70_{(0.40)}$	$62.87_{(0.50)}$
TPN †[20]	$22.17_{(0.20)}$	$81.91_{(0.50)}$	$45.66_{(0.30)}$	$77.22_{(0.40)}$	$71.39_{(0.40)}$	$50.96_{(0.40)}$	$44.54_{(0.40)}$	$63.52_{(0.40)}$
ATA†[36]	$24.43_{(0.20)}$	$90.59_{(0.30)}$	$45.83_{(0.30)}$	$83.75_{(0.40)}$	$75.48_{(0.40)}$	$55.08_{(0.40)}$	$49.14_{(0.40)}$	$66.22_{(0.50)}$
STARTUP [23]	$26.94_{(0.44)}$	$93.02_{(0.45)}$	$47.22_{(0.61)}$	$82.29_{(0.60)}$	–	–	–	–
HVM[4]	$\mathbf{27.15}_{(0.45)}$	$87.65_{(0.35)}$	$42.05_{(0.34)}$	$74.88_{(0.45)}$	–	–	–	–
ConFeSS [2]	$\underline{27.09}$	88.88	48.85	84.65	–	–	–	–
RDC-FT[17]	$25.48_{(0.20)}$	$93.55_{(0.30)}$	$49.06_{(0.30)}$	$84.67_{(0.30)}$	$74.65_{(0.40)}$	$60.63_{(0.40)}$	$53.75_{(0.50)}$	$67.77_{(0.40)}$
LDP-net[41]	$26.88_{(0.46)}$	$91.89_{(0.50)}$	$48.44_{(0.67)}$	$84.05_{(0.66)}$	$75.47_{(0.73)}$	$59.64_{(0.77)}$	$53.06_{(0.82)}$	$\underline{73.34}_{(0.75)}$
StyleAdv[6]	$26.24_{(0.35)}$	$\underline{96.51}_{(0.28)}$	$\underline{53.05}_{(0.54)}$	$\underline{91.64}_{(0.43)}$	$\underline{79.35}_{(0.61)}$	$\underline{64.10}_{(0.64)}$	$\underline{56.44}_{(0.68)}$	$70.90_{(0.63)}$
AWCoL	$24.50_{(0.33)}$	$\mathbf{99.59}_{(0.10)}$	$\mathbf{58.75}_{(0.60)}$	$\mathbf{96.76}_{(0.27)}$	$\mathbf{92.56}_{(0.36)}$	$\mathbf{67.31}_{(0.64)}$	$\mathbf{62.94}_{(0.75)}$	$\mathbf{86.23}_{(0.52)}$