Unsupervised Meta-learning via Few-shot Pseudo-supervised Contrastive Learning

Huiwon Jang[A]*, Hankook Lee[B]*†, Jinwoo Shin[A]   [A]Korea Advanced Institute of Science and Technology (KAIST), [B]LG AI Research   *Equal contribution †Work done at KAIST

TL; DR. Constructing online pseudo-tasks via momentum representations and applying contrastive learning improves the pseudo-labeling strategy progressively for meta-learning.

Introduction

Unsupervised meta-learning learns generalizable knowledge about tasks from unlabeled data. It can solve unseen, yet relevant tasks. However, it requires to construct synthetic tasks to meta-learn without labels:

- Assigning pseudo-labels and applying supervised meta-learning (e.g., CACTUs [1], UMTRA [2])

Research Question 1: How to progressively improve a pseudo-labeling strategy during meta-learning?

- Utilizing generative models for generating synthetic tasks (e.g., LASIUM [3], Meta-GMVAE [4], Meta-SVEBM [5])

Research Question 2: How to construct diverse tasks without generative models so that we can scale the methods into large-scale?

Key idea: Construct pseudo-tasks online via momentum representations and apply contrastive learning.

- Momentum network can improve pseudo-labeling progressively
- Momentum queue can construct diverse tasks in an online manner without generative models

Summary of Contribution

We propose Pseudo-supervised Contrast (**PsCo**), an effective unsupervised meta-learning framework for few-shot classification, which constructs diverse pseudo-tasks without labels progressively utilizing the momentum network and the queue of previous batches.

Extensive experiments demonstrate that

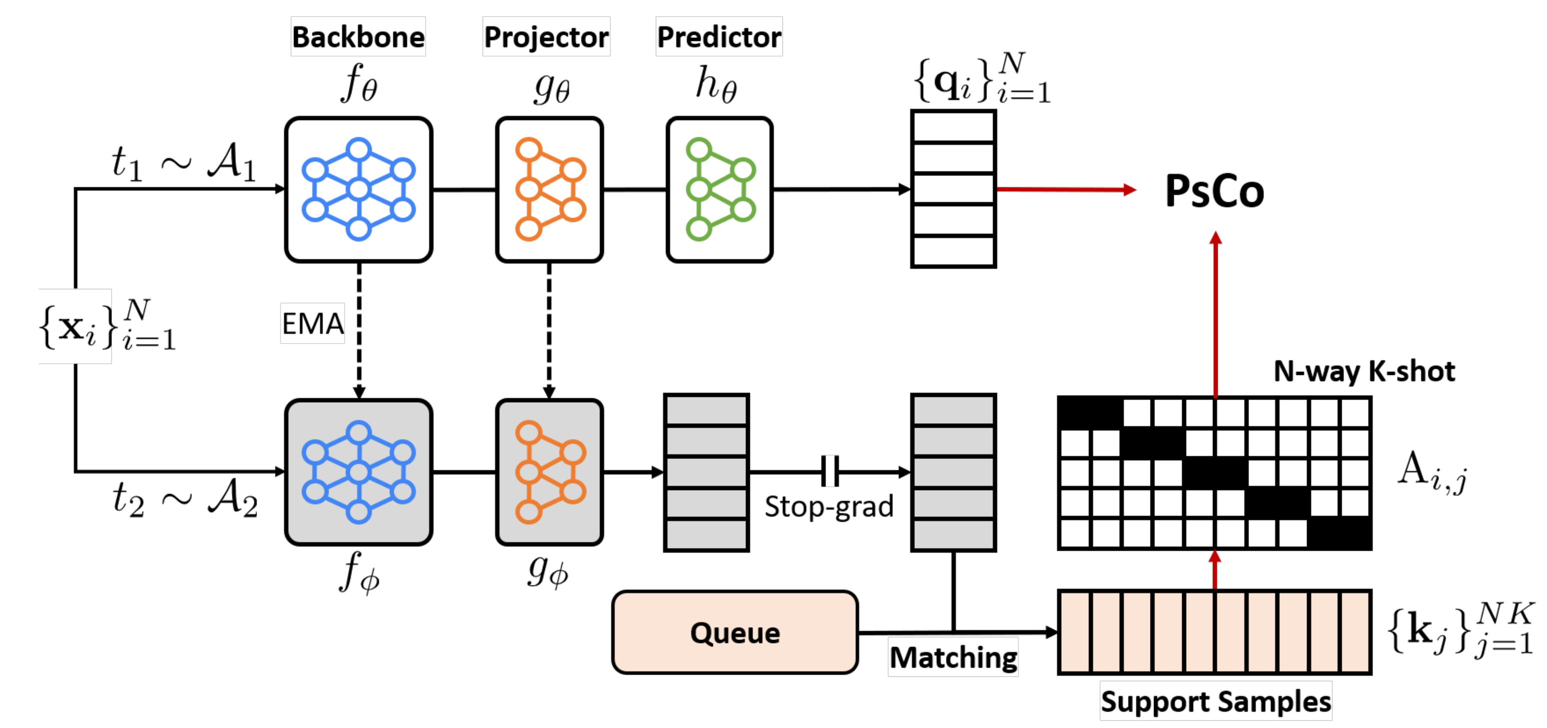
1. **PsCo** achieves SotA performance on standard few-shot tasks
2. **PsCo** shows superiority on cross-domain few-shot tasks
3. **PsCo** is applicable to large-scale architectures/datasets

References

[1] Hsu et al., Unsupervised Learning via Meta-learning, ICLR 2019
[2] Khodadadeh et al., Unsupervised Meta-learning for Few-shot Image Classification, NeurIPS 2019
[3] Khodadadeh et al., Unsupervised Meta-learning through Latent-space Interpolation in Generative Models, ICLR 2019
[4] Lee et al., Meta-GMVAE: Mixture of Gaussian VAE for Unsupervised Meta-learning, ICLR 2021
[5] Kong et al., Unsupervised Meta-learning via Latent Space Energy-based Model of Symbol Vector Coupling, NeurIPSW 2021
[6] He et al., Momentum Contrast for Unsupervised Visual Representation Learning, CVPR 2020
[7] Khosla et al., Supervised Contrastive Learning, NeurIPS 2020

Method

Notation. $\{x_i\}_{i=1}^N$ is a current mini-batch, t_1 is strong augmentation, t_2 is weak augmentation, θ is online encoder, and ϕ is momentum encoder.



How to construct online pseudo-task when meta-training?

- Queries: $\{x_i\}_{i=1}^N$, where query representations are $\{\mathbf{q}_i\}_{i=1}^N$
- Supports: Sample from momentum queue $\{\tilde{\mathbf{z}}_i\}_{j=1}^M$, which are semantically similar to queries while all sampled supports are different. It is obtained by solving the following optimization problem via a matching algorithm:

Step 1: $$\max_{\tilde{A}\in\{0,1\}^{N\times M}} \sum_{i=1}^{N}\sum_{j=1}^{M} \tilde{A}_{i,j}\cdot \mathbf{z}_i^T\tilde{\mathbf{z}}_j \quad \text{such that} \quad \sum_j \tilde{A}_{i,j}=K, \quad \sum_i \tilde{A}_{i,j}\le 1$$

Step 2: $$\boldsymbol{A} \leftarrow \mathrm{TopK}(\widetilde{\boldsymbol{A}}^*)$$

Meta-training. Our objective for learning pseudo-tasks is as follows:

$$\mathcal{L}_{\mathrm{PsCo}} := -\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sum_{j=1}^{NK}\mathrm{A}_{ij}}\sum_{j=1}^{NK}\mathrm{A}_{ij}\log\frac{\exp(\mathbf{q}_i^\top\mathbf{k}_j/\tau)}{\sum_{j=1}^{NK}\exp(\mathbf{q}_i^\top\mathbf{k}_k/\tau)}$$

where $\mathbf{k}_j$ are support representations sampled by the matching algorithm and $\boldsymbol{A}\in\{0,1\}^{N\times NK}$ is pseudo-label assignment matrix.

Meta-test. We use prototypes of supports for prediction:

$$\hat{y} := \mathrm{argmax}_y \mathbf{q}_q^T \boldsymbol{c}_y \quad \text{where} \quad \boldsymbol{c}_y := \mathrm{Normalize}\left(\sum\nolimits_s \mathbf{1}_{y_s=y}\cdot \boldsymbol{z}_s\right)$$

where $\boldsymbol{z}_s, \mathbf{q}_q$ are support and query representations, respectively,

Inspired by the following interpretation of $\mathcal{L}_{\mathrm{PsCo}}$:

$$\mathcal{L}_{\mathrm{PsCo}} = -\frac{1}{N}\sum\nolimits_i \frac{1}{\tau_{\mathrm{PsCo}}}\boldsymbol{q}_i^T\left(\frac{1}{K}\sum\nolimits_j A_{i,j}\boldsymbol{z}_j\right) + \text{term not depending on } \boldsymbol{A}$$

Adaptation scheme for cross-domain few-shot classification:

- Treat each support sample as a query
- Freeze the backbone and fine-tune only the projector and the predictor

Experiment

PsCo achieves state-of-the-art performance on standard (first table) and cross-domain (second table) few-shot benchmarks **(ConvNet).**

Method	Omniglot (way, shot) (5,1)	(5,5)	(20,1)	(20,5)	miniImageNet (way, shot) (5,1)	(5,5)	(5,20)	(5,50)
Training from Scratch	52.50	74.78	24.91	47.62	27.59	38.48	51.53	59.63
Unsupervised meta-learning								
CACTUs-MAML	68.84	87.78	48.09	73.36	39.90	53.97	63.84	69.64
CACTUs-ProtoNets	68.12	83.58	47.75	66.27	39.18	53.36	61.54	63.55
UMTRA	83.80	95.43	74.25	92.12	39.93	50.73	61.11	67.15
LASIUM-MAML	83.26	95.29	-	-	40.19	54.56	65.17	69.13
LASIUM-ProtoNets	80.15	91.10	-	-	40.05	52.53	61.09	64.89
Meta-GMVAE	94.92	97.09	82.21	90.61	42.82	55.73	63.14	68.26
Meta-SVEBM	91.85	97.21	79.66	92.21	43.38	58.03	67.07	72.28
PsCo (Ours)	**96.37**	**99.13**	**89.64**	**97.07**	**46.70**	**63.26**	**72.22**	**73.50**
Self-supervised learning								
SimCLR	92.13	97.06	80.95	91.60	43.35	52.50	61.83	64.85
MoCo v2	92.66	97.38	82.13	92.35	41.92	50.94	60.23	63.45
SwAV	93.13	97.32	82.63	92.12	43.24	52.41	61.36	64.52
Supervised meta-learning								
MAML	94.46	98.83	84.60	96.29	46.81	62.13	71.03	75.54
ProtoNets	98.35	99.58	95.31	98.81	46.56	62.29	70.05	72.04

(a) Cross-domain few-shot benchmarks similar to miniImageNet.

Method	CUB (5, 5)	(5, 20)	Cars (5, 5)	(5, 20)	Places (5, 5)	(5, 20)	Plantae (5, 5)	(5, 20)
Unsupervised meta-learning								
Meta-GMVAE	47.48	54.08	31.39	38.36	57.70	65.08	38.27	45.02
Meta-SVEBM	45.50	54.61	34.27	46.23	51.27	61.09	38.12	46.22
PsCo (Ours)	**57.38**	**68.58**	**44.01**	**57.50**	**63.60**	**73.95**	**52.72**	**64.53**
Self-supervised learning								
SimCLR	52.11	61.89	37.40	50.05	60.10	69.93	43.42	54.92
MoCo v2	53.23	62.81	38.65	51.77	59.09	69.08	43.97	55.45
SwAV	51.58	61.38	36.85	50.03	59.57	69.70	42.68	54.03
Supervised meta-learning								
MAML	56.57	64.17	41.17	48.82	60.05	67.54	47.33	54.86
ProtoNets	56.74	65.03	38.98	47.98	59.39	67.77	45.89	54.29

(b) Cross-domain few-shot benchmarks dissimilar to miniImageNet.

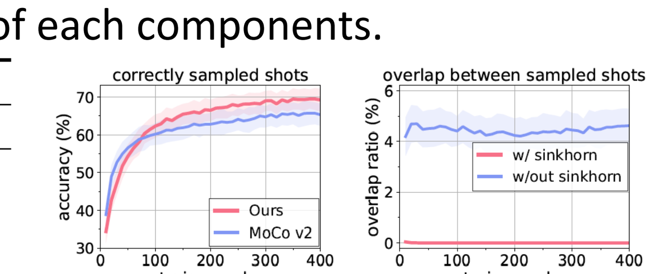
Method	CropDiseases (5, 5)	(5, 20)	EuroSAT (5, 5)	(5, 20)	ISIC (5, 5)	(5, 20)	ChestX (5, 5)	(5, 20)
Unsupervised meta-learning								
Meta-GMVAE	73.56	81.22	73.83	80.11	33.48	39.48	23.23	26.26
Meta-SVEBM	71.82	83.13	70.83	80.21	38.85	48.43	**26.26**	28.91
PsCo (Ours)	**88.24**	**94.95**	**81.08**	**87.65**	**44.00**	**54.59**	24.78	27.69
Self-supervised learning								
SimCLR	79.90	88.73	79.14	85.05	42.83	51.35	25.14	**29.21**
MoCo v2	80.96	89.85	79.94	86.16	43.43	52.14	25.24	29.19
SwAV	80.15	89.24	79.31	85.62	43.21	51.99	24.99	28.57
Supervised meta-learning								
MAML	77.76	83.24	71.48	76.70	47.34	55.09	22.61	24.25
ProtoNets	76.01	83.64	64.91	70.88	40.62	48.38	23.15	25.72

PsCo can be scale into large-scale (**ResNet-50** ImageNet-pretrained).

Method	CUB	Cars	Places	Plantae	CropDiseases	EuroSAT	ISIC	ChestX
MoCo v2	64.16	47.67	81.39	61.36	82.89	76.96	38.26	**24.28**
+PsCo (Ours)	**76.63**	**53.45**	**83.87**	**69.17**	**89.85**	**83.99**	**41.64**	23.60
BYOL	67.45	45.74	75.43	56.86	80.82	77.70	37.27	24.15
+PsCo (Ours)	**82.13**	**56.19**	**83.80**	**71.14**	**92.92**	**85.33**	**42.90**	**26.05**
Supervised	89.13	75.15	84.41	72.91	90.96	85.64	43.34	25.35

Component ablation shows the importance of each components.

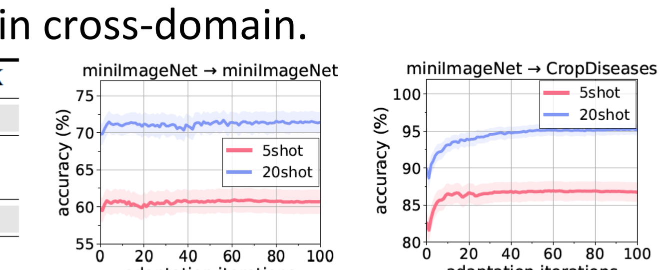
Momentum	Predictor	Sinkhorn	Top-K sampling	$\mathcal{L}_{\mathrm{MoCo}}$	(5, 1)	(5, 5)	(20, 1)	(20, 5)
✓	✓	✓	✓	✓	**96.37**	**99.13**	**89.64**	**97.07**
✗	✓	✓	✓	✓	90.32	96.78	76.17	90.41
✓	✗	✓	✓	✓	90.21	96.86	76.15	90.53
✓	✓	✗	✓	✓	95.81	98.94	88.25	96.57
✓	✓	✓	✗	✓	94.95	98.81	86.32	96.05
✓	✓	✓	✓	✗	93.16	97.40	81.03	91.33



(a) Pseudo-label quality (b) Shot overlap ratio

New **adaptation scheme** is more useful in cross-domain.

Adaptation	miniImageNet	CUB	Cars	Places	Plantae	CropDiseases	EuroSAT	ISIC	ChestX
5-way 5-shot									
✗	63.26	55.15	42.27	62.98	48.31	79.75	74.73	41.18	24.54
✓	**63.30**	**57.38**	**44.01**	**63.60**	**52.72**	**88.24**	**81.08**	**44.00**	**24.78**
5-way 20-shot									
✗	72.22	62.35	51.02	70.85	55.91	84.72	78.96	48.53	27.60
✓	**73.00**	**68.58**	**57.50**	**73.95**	**64.53**	**94.95**	**87.65**	**54.59**	**27.69**



(c) In-domain adaptation (d) Cross-domain adaptation