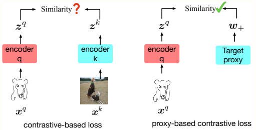
PCL: Proxy-based Contrastive Learning for Domain Generalization

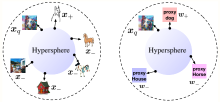
Xufeng Yao, Yang Bai, Xinyun Zhang , Yuechen ZHANG, QI SUN, Ran Chen, Ruiyu Li, Bei Yu. *CUHK & SmartMore*.

Background & Motivation



- ✦ Hard positive pairs may hamper DG
- ✦ Using Proxy to reduce the complexity of contrastive loss

Comparison between two losses



(a) Contrastive-based Loss (b) Proxy-based Loss

- ✦ Contrastive loss: sample-to-sample pairs
- ✦ Proxy loss: proxy-to-sample pairs

Complexity comparsion

Loss function	positive pair	negative pair	relations	category	training complexity
softmax CE loss	(w_y, x_i)	$(w_1, x_i), (w_2, x_i), \ldots, (w_j, x_i)$	proxy-to-sample	proxy-based	$\mathcal{O}(CN)$
Contrastive loss	(x_i, x_i^*)	$(x_i, x_1), (x_i, x_2), \ldots, (x_1, x_n)$	sample-to-sample	pair-based	$\mathcal{O}(N^2)$
MS Loss	$(x_i, x_j)\ldots(x_i, x_m)$	$(x_i, x_1), (x_i, x_2)\ldots(x_1, x_n)$	sample-to-sample	pair-based	$\mathcal{O}(N^2)$
triplet Loss	$(x_i, x_j)\ldots(x_i, x_m)$	$(x_i, x_1), (x_i, x_2)\ldots(x_1, x_n)$	sample-to-sample	pair-based	$\mathcal{O}(N^3)$

- ✦ Pair-based loss: rich sample-to-sample pairs, high complexity
- ✦ Proxy-based loss: low complexity, high generalization

Proxy-based Contrastive Learning

Review Softmax-based CE Loss

- ✦ Advantage: learn a proxy for each class efficently
- ✦ Disadvantage: miss rich sample-to-sample pairs

Review Contrastive-based Loss

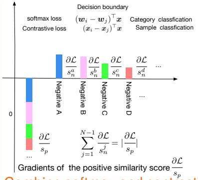
- ✦ Advantage: plentiful sample-to-sample pairs, implicit hard pair mining

$$\mathcal{L}_{\mathrm{CL}} = \lim_{\alpha\to\infty} \frac{1}{\alpha} - \log\left(\frac{\exp(\alpha\cdot s_p)}{\exp(\alpha\cdot s_p) + \sum_{j=1}^{N-1}\exp(\alpha\cdot s_n^j)}\right)$$

$$= \lim_{\alpha\to\infty} \frac{1}{\alpha}\log\left(1 + \sum_{j=1}^{N-1}\exp(\alpha(s_n^j - s_p))\right)$$

$$= \max[s_n^j - s_p]_+.$$

- ✦ Disadvantage: high complexity, hard to optimize



Our Solution Combine softmax and contrastive losses

$$\mathcal{L}_{\mathrm{PCL}} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{\exp(\mathbf{w}_c^\top \mathbf{z}_i\cdot\alpha)}{Z}$$

where Z is given by:

$$Z = \exp(\mathbf{w}_c^\top \mathbf{z}_i\cdot\alpha) + \sum_{k=1}^{C-1}\exp(\mathbf{w}_k^\top \mathbf{z}_j\cdot\alpha) + \sum_{j=1, j\neq i}^{K}\exp(\mathbf{z}_i^\top \mathbf{z}_j\cdot\alpha)$$

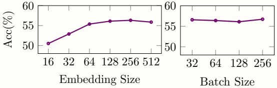
Results

Ablation study on positive loss

Method	PACS	OfficeHome	TerraIncognita
softmax CE	**88.1**	**70.6**	**50.0**
softmax CE w. positive loss	86.7	70.1	48.5

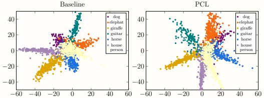
Positive loss is not effective in DG

Abaltion study on different settings

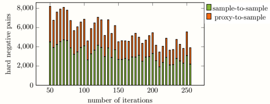


Our method is relative stable with different setting

t-SNE visualization results



Analysis on hard negative pairs selection



Good selection of hard negative pairs

Results

Algorithm	A	C	P	R	Avg
Mixstyle	51.1	53.2	68.2	69.2	60.4
ERM	63.1	51.9	77.2	78.1	67.6
I-Mixup	62.4	54.8	76.9	78.3	68.1
SagNet	63.4	54.8	75.8	78.3	68.1
CORAL	65.3	54.4	76.5	78.4	68.7
SWAD	66.1	57.7	78.4	80.2	70.6
Ours	**67.3**	**59.9**	**78.7**	**80.7**	**71.6**

Our method achieves the sota performance on several benchmarks, e.g., OfficeHome.