
CVPR
JUNE
19-24
2022
NEW ORLEANS • LOUISIANA
CVSPORTS

PCL: Proxy-based Contrastive Learning for Domain Generalization

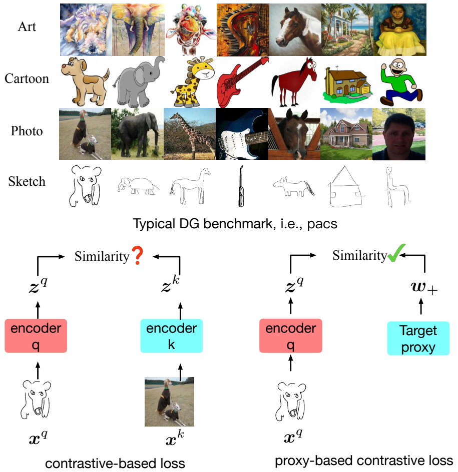
Xufeng Yao[1], Yang Bai[1], Xinyun Zhang[1], Yuechen Zhang[1], Qi Sun[1], Ran Chen[1], RuiYu Li[2], Bei Yu[1]

[1]The Chinese University of Hong Kong
[2]SmartMore
`{xfyao,byu}@cse.cuhk.edu.hk`

May. 27, 2022

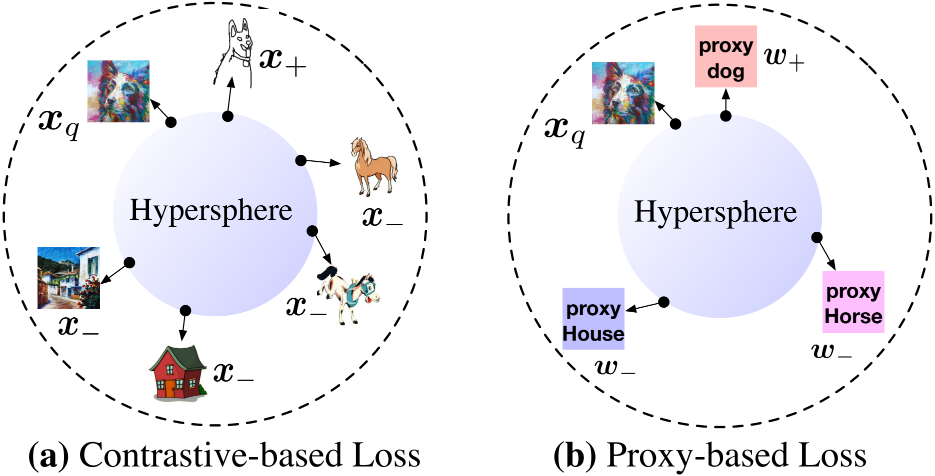
Background and Motivation

Background of Domain Generalization



- DG aims to train a model from multiple source domains that can generalize well on target domain.
- Contrastive learning offers a potential solution, but is not effective in DG.
- We aims to use proxy-based contrastive learning to address the problem.

Comparison between two losses



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

Complexity comparison

Loss function	positive pair	negative pair	relations	category	training complexity
softmax CE loss	$(\boldsymbol{w}_y, \boldsymbol{x}_i)$	$(\boldsymbol{w}_1, \boldsymbol{x}_i), (\boldsymbol{w}_2, \boldsymbol{x}_i), ..., (\boldsymbol{w}_j, \boldsymbol{x}_i)$	proxy-to-sample	proxy-based	$\mathcal{O}(CN)$
Contrastive loss	$(\boldsymbol{x}_i, \boldsymbol{x}_i^*)$	$(\boldsymbol{x}_i, \boldsymbol{x}_1), (\boldsymbol{x}_i, \boldsymbol{x}_2), ..., (\boldsymbol{x}_1, \boldsymbol{x}_n)$	sample-to-sample	pair-based	$\mathcal{O}(N^2)$
MS Loss	$(\boldsymbol{x}_i, \boldsymbol{x}_j)...(\boldsymbol{x}_i, \boldsymbol{x}_m)$	$(\boldsymbol{x}_i, \boldsymbol{x}_1), (\boldsymbol{x}_i, \boldsymbol{x}_2)...(\boldsymbol{x}_1, \boldsymbol{x}_n)$	sample-to-sample	pair-based	$\mathcal{O}(N^2)$
triplet Loss	$(\boldsymbol{x}_i, \boldsymbol{x}_j)...(\boldsymbol{x}_i, \boldsymbol{x}_m)$	$(\boldsymbol{x}_i, \boldsymbol{x}_1), (\boldsymbol{x}_i, \boldsymbol{x}_2)...(\boldsymbol{x}_1, \boldsymbol{x}_n)$	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 of softmax CE loss

- Pros: Learn a proxy for each classes efficiently.
- Pros: Low complexity, safe convergence.
- Cons: Miss rich sample-to-sample pairs.

$$\mathcal{L}_{\mathrm{CE}} = -\log \frac{\exp(\boldsymbol{w}_c^\top \boldsymbol{z}_i)}{\exp(\boldsymbol{w}_c^\top \boldsymbol{z}_i) + \sum_{j=1}^{C-1} \exp(\boldsymbol{w}_j^\top \boldsymbol{z}_i)}, \tag{1}$$

Review of Contrastive loss

- Pros: Leverage dense sample-to-sample pairs.
- Pros: Implicit hard pair mining.
- Cons: High complexity, unstable convergence.

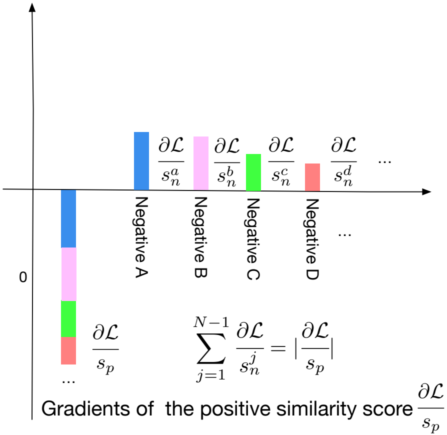
$$\mathcal{L}_{\mathrm{CL}} = -\log \frac{\exp(\mathbf{z}_i^\top \mathbf{z}_+ \cdot \alpha)}{\exp(\mathbf{z}_i^\top \mathbf{z}_+ \cdot \alpha) + \sum \exp(\mathbf{z}_i^\top \mathbf{z}_- \cdot \alpha)}, \tag{2}$$

Implicit hard pair mining in contrastive loss

- By controlling α, contrastive loss implicitly conduct hard pair mining.
- Sufficient pairs guarantee the performance.

$$
\begin{aligned}
\mathcal{L}_{\mathrm{CL}} &= \lim_{\alpha\to\infty} \frac{1}{\alpha} - \log\left(\frac{\exp(\alpha\cdot s_p)}{\exp(\alpha\cdot s_p) + \sum_{j=1}^{N-1}\exp(\alpha\cdot s_n^j)}\right) \\
&= \lim_{\alpha\to\infty} \frac{1}{\alpha}\log\left(1+\sum_{j=1}^{N-1}\exp(\alpha(s_n^j - s_p))\right) \\
&= \max[s_n^j - s_p]_+,
\end{aligned} \tag{3}
$$

High complexity may impede the performance



gradients analysis

Combine Softmax CE and Conrtastive Loss

- Softmax: Low complexity, overlook sample-to-sample pairs
- Contrastive: High complexity, rich pairs, unstable convergence.

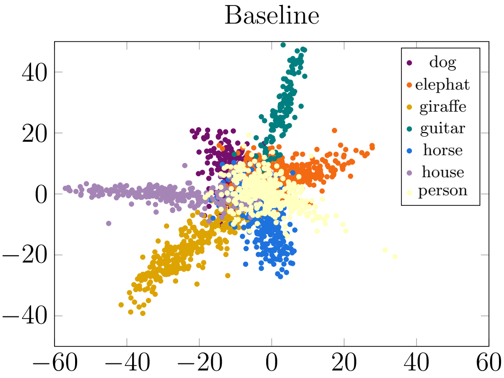
$$\mathcal{L}_{\mathrm{PCL}} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{\exp(\mathbf{w}_c^\top \mathbf{z}_i \cdot \alpha)}{Z}, \tag{4}$$

where Z is given by:

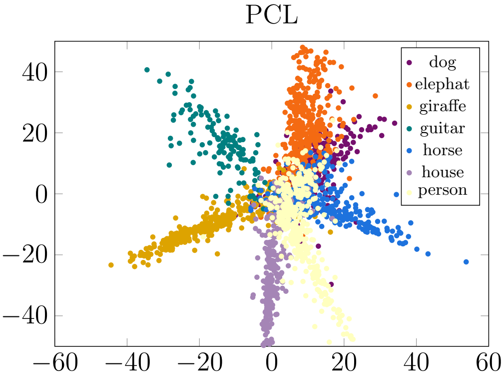
$$Z = \exp(\mathbf{w}_c^\top \mathbf{z}_i \cdot \alpha) + \sum_{k=1}^{C-1} \exp(\mathbf{w}_k^\top \mathbf{z}_j \cdot \alpha) + \sum_{j=1, j\neq i}^{K} \exp(\mathbf{z}_i^\top \mathbf{z}_j \cdot \alpha). \tag{5}$$

Experimental Results

Visualizaiton of learned features



softmax



Ours

Experimental Results

Table: Comparison with state-of-the-art methods on OfficeHome benchmark with ResNet-50 imagenet-pretrained model

Algorithm	A	C	P	R	Avg
Mixstyle[1]	51.1	53.2	68.2	69.2	60.4
SagNet[2]	63.4	54.8	75.8	78.3	68.1
CORAL[3]	65.3	54.4	76.5	78.4	68.7
SWAD[4]	66.1	57.7	78.4	80.2	70.6
Ours	**67.3**	**59.9**	**78.7**	**80.7**	**71.6**

[1] Zhou; Kaiyang; et al. Domain generalization with mixstyle. ICLR 2021..

[2] Nam; Hyeonseob; et al. Reducing domain gap by reducing style bias. CVPR. 2021..

[3] Sun; Baochen;
and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. ECCV; 2016..

[4] Cha; Junbum; et al. Swad: Domain generalization by seeking flat minima. NeurIPS; 2021..

Table: Comparison with state-of-the-art methods on PACS benchmark with ResNet-50 imagenet-pretrained model

Algorithm	A	C	P	S	Avg.
Mixstyle[5]	86.8	79.0	96.6	78.5	85.2
CORAL[6]	88.3	80.0	97.5	78.8	86.2
SagNet[7]	87.4	80.7	97.1	80.0	86.3
SWAD[8]	89.3	83.4	97.3	82.5	88.1
Ours	**90.2**	**83.9**	98.1	82.6	**88.7**

[5] Zhou; Kaiyang; et al. Domain generalization with mixstyle. ICLR 2021..

[6] Sun; Baochen;
and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. ECCV; 2016..

[7] Nam; Hyeonseob; et al. Reducing domain gap by reducing style bias. CVPR. 2021..

[8] Cha; Junbum; et al. Swad: Domain generalization by seeking flat minima. NeurIPS; 2021..

Table: Comparison with state-of-the-art methods on TerraIncognita benchmark with ResNet-50 imagenet-pretrained model

Algorithm	Location100	Location38	Location43	Location46	Avg.
Mixstyle[9]	54.3	34.1	55.9	31.7	44.0
CORAL[10]	51.6	42.2	57.0	39.8	47.7
SagNet[11]	53.0	43.0	57.9	40.4	48.6
SWAD[12]	55.4	44.9	59.7	39.9	50.0
Ours	58.7	**46.3**	**60.0**	**43.6**	**52.1**

[9] **Zhou; Kaiyang; et al. Domain generalization with mixstyle. ICLR 2021..**

[10] **Sun; Baochen;**
and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. ECCV; 2016..

[11] **Nam; Hyeonseob; et al. Reducing domain gap by reducing style bias. CVPR. 2021..**

[12] **Cha; Junbum; et al. Swad: Domain generalization by seeking flat minima. NeurIPS; 2021..**

Table: Comparison with state-of-the-art methods on DomainNet benchmark with ResNet-50 ImageNet pre-trained model

Algorithm	clip	info	paint	quick	real	sketch	Avg
Mixstyle[13]	51.9	13.3	37.0	12.3	46.1	43.4	34.0
SagNet[14]	57.7	19.0	45.3	12.7	58.1	48.8	40.3
CORAL[15]	59.2	19.7	46.6	13.4	59.8	50.1	41.5
SWAD[16]	66.0	22.4	53.5	**16.1**	65.8	55.5	46.5
Ours	**67.9**	24.3	**55.3**	15.7	**66.6**	**56.4**	**47.7**

[13] Zhou; Kaiyang; et al. Domain generalization with mixstyle. ICLR 2021..

[14] Nam; Hyeonseob; et al. Reducing domain gap by reducing style bias. CVPR. 2021..

[15] Sun; Baochen;
and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. ECCV; 2016..

[16] Cha; Junbum; et al. Swad: Domain generalization by seeking flat minima. NeurIPS; 2021..

THANK YOU!