Difficulty-based Sampling for Debiased Contrastive Representation Learning

Taeuk Jang[1] Xiaoqian Wang[1]

CVPR 2023

Poster Tag: THU-PM-328

Purdue University[1]

JUNE 18-22, 2023

CVPR

VANCOUVER, CANADA

Overview

Motivation

- Due to unsupervised nature, it is not trivial to find *legitimate* negative samples in contrastive learning, *e.g., false negative problem*.
- Previous works proposed statistical approaches to address the problem such as false negative debiasing and hard negative mining.

Contributions

- Propose a novel debiased contrastive learning method that addresses the problem from a new perspective by incorporating relative difficulty with data bias.
- Introduce triplet loss as bias-amplifying contrastive loss, which serves as an effective surrogate for learning biased representation.
- Theoretically show that the triplet loss amplifies the bias in self-supervised representation learning.

Motivation

Contrastive Learning[1]: Learn representation that samples with same class are gathered and different class to be apart.

- $\mathbf{x}^a \sim p(\mathbf{x})$: anchor
- $\mathbf{x}^+ \sim p(\mathbf{x}^+|\mathbf{x})$: positive samples
- $\mathbf{x}^- \sim p(\mathbf{x})$: negative samples

$$\mathbb{E}_{\mathbf{x}^a,\mathbf{x}^+,\mathbf{x}^-}\left[-\log \frac{e^{E(\mathbf{x}^a)^\intercal E(\mathbf{x}^+)}}{e^{E(\mathbf{x}^a)^\intercal E(\mathbf{x}^+)} + \sum_{j=1}^{M} e^{E(\mathbf{x}^a)^\intercal E(\mathbf{x}^{-(j)})}}\right]$$

Finding legitimate negatives is critical

- Negative samples are drawn from the same sample space as anchor.
  - True negative vs False negative [2] : negatives can have same class as anchor.
  - Easy negative vs Hard negative [3] : hard negative samples are informative.

Both require domain knowledge about distribution τ^-, τ^- and assume label distribution is uniform.

[1] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In ICML, 2020.
[2] Ching-Yao Chuang, Joshua Robinson, Lin Yen-Chen, Antonio Torralba, and Stefanie Jegelka. Debiased contrastive learning. arXiv preprint arXiv:2007.00224, 2020.
[3] Joshua Robinson, Ching-Yao Chuang, Suvrit Sra, and Stefanie Jegelka. Contrastive learning with hard negative samples. arXiv preprint arXiv:2010.04592, 2020

Motivation

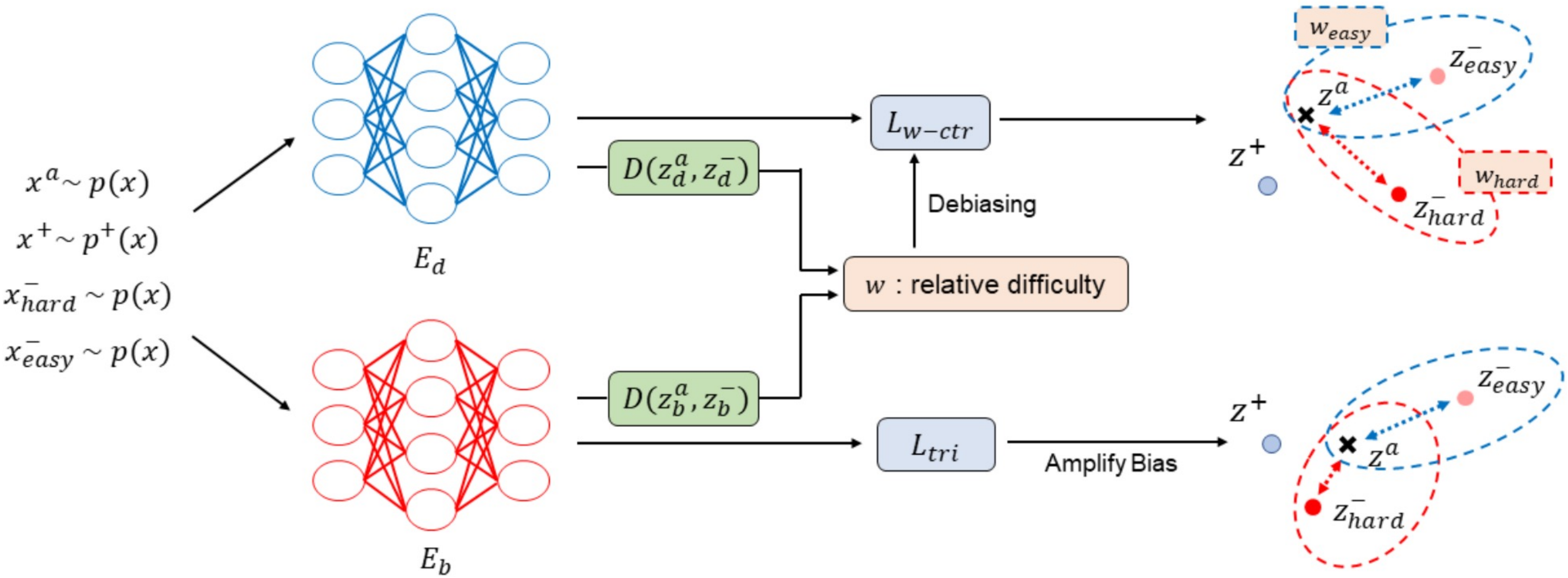
Supervised Learning

- Difficulty of samples are related to data bias. For instance,
  - Texture, color, and background in image classification[1].
  - Race and gender in face recognition[2].
- Samples against the data bias are likely to be hard samples.
  - *e.g., bird in the water vs. bird in the forest*

- Emphasize bias-conflicting samples for better performance and generalization as they are more informative[3,4].

[1] Hyojin Bahng, Sanghyuk Chun, Sangdoo Yun, Jaegul Choo, and Seong Joon Oh. Learning de-biased representations with biased representations. In ICML, 2020.
[2] Taeuk Jang, Feng Zheng, and Xiaoqian Wang. Constructing a fair classifier with generated fair data. In AAAI, 2021
[3] Jungsoo Lee, Eungyeup Kim, Juyoung Lee, Jihyeon Lee, and Jaegul Choo. Learning debiased representation via disentangled feature augmentation. In NeurIPS, 2021
[4] Evan Z Liu, Behzad Haghgoo, Annie S Chen, Aditi Raghunathan, Pang Wei Koh, Shiori Sagawa, Percy Liang, and Chelsea Finn. Just train twice: Improving group robustness without training group information. In ICML, 2021.

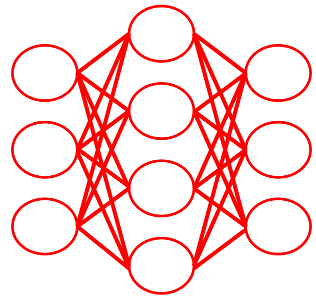
Difficulty-based Sampling for Debiased Contrastive Learning



- We employ two encoders:
  - Bias-amplifying encoder E_b: intentionally amplify bias that focuses on easy samples.
  - Debiased encoder E_d: emphasize hard negative samples leveraging relative difficulty by referencing representation from E_b.

Difficulty-based Sampling for Debiased Contrastive Learning

Learning bias-amplifying representation



E_b

- We employ triplet loss[1] in self-supervised manner to learn bias- amplifying representation.

$$\mathcal{L}_{tri} = \mathbb{E}\Big[||E_b(\mathbf{x}^a) - E_b(\mathbf{x}^+)||_2^2 - ||E_b(\mathbf{x}^a) - E_b(\mathbf{x}^-)||_2^2\Big]$$

- The derivative of triplet loss for optimization:

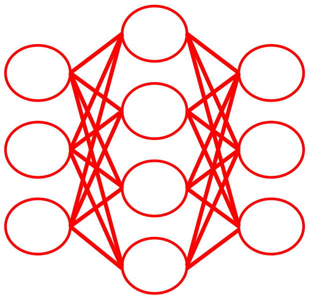
$$\nabla_{\theta_b}\mathcal{L}_{tri} = \mathbb{E}\Big[2{\Delta^+}^{\intercal}\nabla\big(E_b(\mathbf{x}^a) - E_b(\mathbf{x}^+)\big) - 2{\Delta^-}^{\intercal}\nabla\big(E_b(\mathbf{x}^a) - E_b(\mathbf{x}^-)\big)\Big],$$

where $\Delta^+ = E_b(\mathbf{x}^a) - E_b(\mathbf{x}^+), \quad \Delta^- = E_b(\mathbf{x}^a) - E_b(\mathbf{x}^-)$

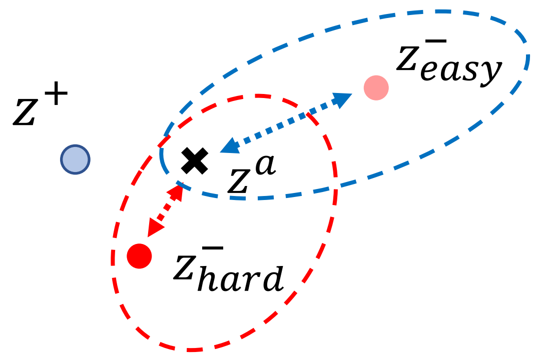
[1] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In CVPR, 2015.

Difficulty-based Sampling for Debiased Contrastive Learning

Learning bias-amplifying representation



E_b



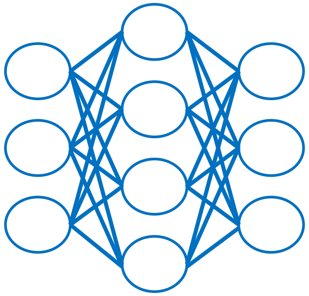
$$\nabla_{\theta_b}\mathcal{L}_{tri} = \mathbb{E}\left[2\Delta^{+\intercal}\nabla\big(E_b(\mathbf{x}^a) - E_b(\mathbf{x}^+)\big) - 2\Delta^{-\intercal}\nabla\big(E_b(\mathbf{x}^a) - E_b(\mathbf{x}^-)\big)\right],$$

where $\Delta^+ = E_b(\mathbf{x}^a) - E_b(\mathbf{x}^+), \quad \Delta^- = E_b(\mathbf{x}^a) - E_b(\mathbf{x}^-)$

- The gradient on negative sample is weighted by Δ^-.
  - Samples distinguishable from anchor ($\Delta^- \gg 0$), *i.e., easy negatives*.
  - Samples similar to anchor ($\Delta^- \approx 0$), *i.e., hard negatives*.
  - Triplet loss **amplifies bias** in the representation.

Difficulty-based Sampling for Debiased Contrastive Learning

Learning debiased representation



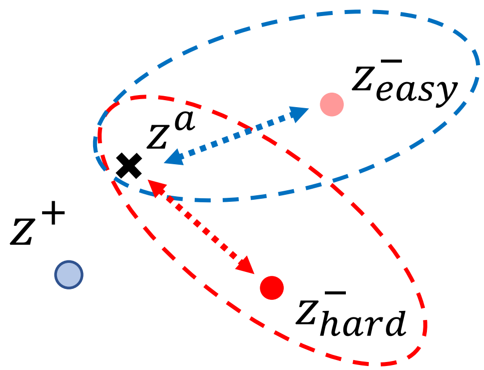
E_d

- We want to learn debiased encoder E_d by referencing biased encoder E_b.
- Weight each negative sample differently by relative difficulty of negative sample $\mathbf{x}^-$ given an anchor $\mathbf{x}^a$
- Relative difficulty: $w\big((\mathbf{z}_d^a, \mathbf{z}_d^-), (\mathbf{z}_b^a, \mathbf{z}_b^-)\big) = 1 + \dfrac{\tilde{D}(\mathbf{z}^a, \mathbf{z}_d^-)}{\tilde{D}(\mathbf{z}^a, \mathbf{z}_d^-) + \tilde{D}(\mathbf{z}^a, \mathbf{z}_b^-)}$,

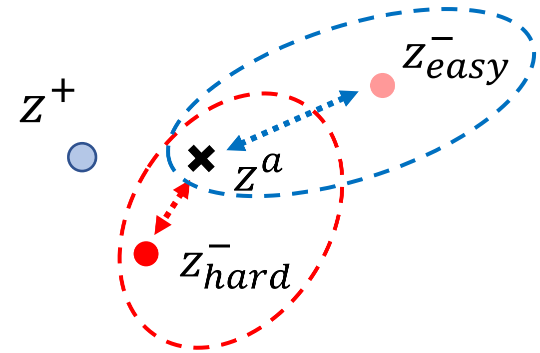
where $\tilde{D}(\mathbf{z}_i^a, \mathbf{z}_i^-) = \dfrac{D(\mathbf{z}_i^a, \mathbf{z}_i^-)}{\max_{(\mathbf{x}^a, \mathbf{x}^-)\in\mathcal{B}} D(E_i(\mathbf{x}^a), E_i(\mathbf{x}^-))}$

Difficulty-based Sampling for Debiased Contrastive Learning

Learning debiased representation



Representation by E_d



Representation by E_b

- $w \in [1,2]$
  - $w \approx 2$ (hard negatives): $\widetilde{D}(\mathbf{z}_b^a, \mathbf{z}_b^-) \ll \widetilde{D}(\mathbf{z}_d^a, \mathbf{z}_d^-)$
  - $w \approx 1$ (easy negatives): $\widetilde{D}(\mathbf{z}_b^a, \mathbf{z}_b^-) \gg \widetilde{D}(\mathbf{z}_d^a, \mathbf{z}_d^-)$
- Emphasize negative samples projected closer to anchor by E_b as

$$\mathbb{E}\left[-\log \frac{e^{E(\mathbf{x}^a)^\intercal E(\mathbf{x}^+)}}{e^{E(\mathbf{x}^a)^\intercal E(\mathbf{x}^+)} + w(\mathbf{z}^a, \mathbf{z}_b^-, \mathbf{z}_d^-)e^{E(\mathbf{x}^a)^\intercal E(\mathbf{x}^-)}}\right]$$

- We can also apply statistical debiasing as DCL [1] and HCL [2].

[1] Ching-Yao Chuang, Joshua Robinson, Lin Yen-Chen, Antonio Torralba, and Stefanie Jegelka. Debiased contrastive learning. arXiv preprint arXiv:2007.00224, 2020.
[2] Joshua Robinson, Ching-Yao Chuang, Suvrit Sra, and Stefanie Jegelka. Contrastive learning with hard negative samples. arXiv preprint arXiv:2010.04592, 2020

Quantitative Results

		CIFAR-10			CIFAR-100		
Method	Y	ACC (top-1)	ACC (top-5)	ACC (worst)	ACC (top-1)	ACC (top-5)	ACC (worst)
JTT [26]	O	85.67 ± 0.7	99.65 ± 0.2	72.33± 0.5	61.66 ± 0.6	83.53 ± 0.9	24.00 ± 1.5
SimCLR [4]	×	89.12 ± 0.6	99.74 ± 0.1	75.7 ± 0.4	64.86 ± 0.6	89.67 ± 0.3	20.00 ± 0.2
DCL [8]	×	91.66 ±0.3	99.78 ±0.1	81.2±0.2	68.26 ±0.3	91.19 ±0.1	20.00 ±0.2
HCL [36]	×	91.25 ±0.2	99.78 ±0.1	81.5 ±0.2	68.73 ±0.4	91.19 ±0.1	29.00 ±0.8
WCL (E_d)	×	**92.71±0.3**	**99.84±0.1**	**83.3±0.8**	**69.09±0.2**	**91.63±0.3**	**31.00±0.7**
WCL (E_b)	×	75.61 ±0.7	98.61 ± 0.4	52.6 ± 0.5	41.61± 0.3	69.26 ± 0.2	1.0 ± 0.5

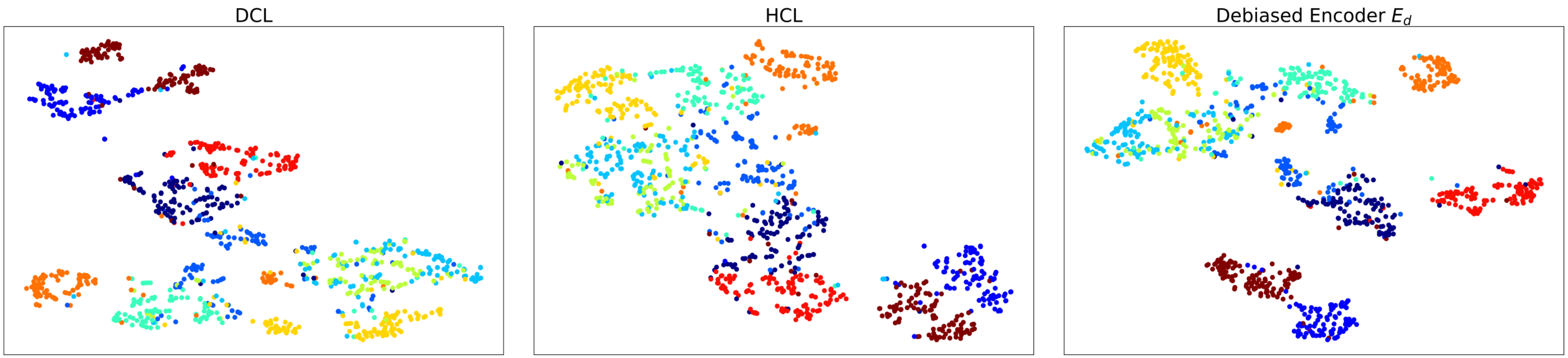
Table 1. Performance evaluation on CIFAR-10 and CIFAR-100.

		Waterbirds [37]		CelebA [27]	
Method	Y	ACC (top-1)	ACC (worst)	ACC (top-1)	ACC (worst)
JTT [26]	O	77.81 ± 2.3	70.00 ± 1.5	76.83± 1.3	67.66 ± 0.5
SimCLR [4]	×	77.80 ± 1.5	0.00	78.61 ± 1.5	44.30 ± 0.7
DCL [8]	×	65.80 ± 1.7	4.51 ± 1.2	77.12 ± 1.6	44.95 ± 0.3
HCL [36]	×	69.31 ± 1.2	5.26 ± 1.1	76.13 ± 2.1	52.13 ± 0.8
WCL (E_d)	×	76.92 ± 0.3	**31.58 ± 3.5**	78.11 ± 2.3	**57.40 ± 1.2**
WCL (E_b)	×	73.64 ± 1.4	14.29 ± 1.5	58.84 ± 2.5	39.79 ± 1.3

Table 2. Performance evaluation on Waterbirds and CelebA dataset. Note thet JTT is supervised learning method. Among the self-supervised learning methods, WCL (ours) achieves the best worst group accuracy with comparable overall performance.

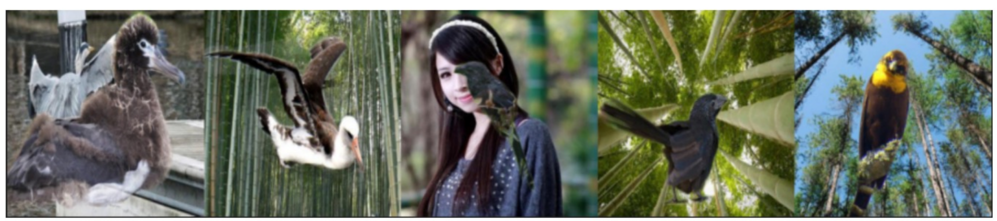
Qualitative Results

- t-SNE visualization on CIFAR-10

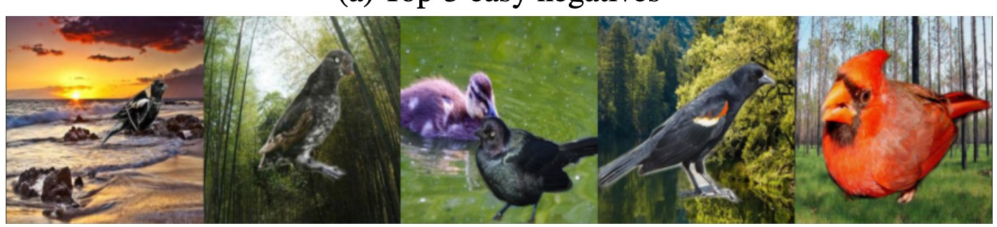


(a) DCL [8] (b) HCL [36] (c) WCL (Ours)

- Visualization of top-5 easy/hard negative on CUB dataset



(a) Top-5 easy negatives



(b) Top-5 hard negatives

Thank you for watching and see you by our poster

Difficulty-based Sampling for Debiased Contrastive Representation Learning

Poster Tag: THU-PM-328

Taeuk Jang[1] Xiaoqian Wang[1]

Purdue University[1]

JUNE 18-22, 2023

CVPR

VANCOUVER, CANADA