

## Contributions

We demonstrate the prototype collapse issue of pre-trained features in a typical Prototypical Contrastive Learning (PCL) framework. ♦ We analyze it with the proposed NEMD score defined on the normalized embedding space.

• We improve the **prototypical** learning of representations through alignment, uniformity, and correlation, mitigating prototype collapse issues.

#### **Preliminaries**

instance-wise contrastive learning

$$\mathcal{L}_{\text{InfoNCE}} = \sum_{i=1}^{n} -\log \frac{\exp(\mathbf{v}_i \cdot \mathbf{v}'_i / \tau)}{\sum_{j=1}^{r} \exp(\mathbf{v}_i \cdot \mathbf{v}'_j / \tau)}$$
prototypical contrastive learning
$$\mathcal{L}_{\text{ProtoNCE}} = \mathcal{L}_{\text{InfoNCE}} + \sum_{i=1}^{n} -\frac{1}{M} \sum_{m=1}^{M} \log \frac{\exp}{\sum_{i=1}^{r} e^{i\pi t}}$$

## **Normalized Earth Moving Distance**

- To quantify the level of prototype collapse in PCL methods, assume prototypes as empirical distributions on a normalized hyper-sphere
- NEMD is then calculated between distributions associated with prototypes to measure the distance of separate prototypes

$$\mathcal{L}_{\text{NEMD}}(p,q) = \inf_{\gamma \in \prod(\mathbf{c}_p^m,\mathbf{c}_q^m)} \mathbb{E}_{(x,y) \sim \gamma} \|$$

## Experimental

## Results on NEMD

#### collapsing solutions $\rightarrow$ high NEMD

• uniform distribution  $\rightarrow$ low NEMD



# Rethinking Prototypical Contrastive Learning through Alignment, Uniformity and Correlation

Shentong Mo<sup>1</sup>, Zhun Sun<sup>2</sup>, Chao Li<sup>3</sup> <sup>1</sup>Carnegie Mellon University & <sup>2</sup>Tohoku University & <sup>3</sup>AIP, RIKEN

 $\mathbf{v}_i \cdot \mathbf{c}_s^m / \boldsymbol{\phi}_s^m$  $\exp(\mathbf{v}_i \cdot \mathbf{c}_j^m / \boldsymbol{\phi}_j^m)$ 

|x-y||

## **Comparison on ImageNet-100**

Method	Arch.	Param.(M)	Batch	Epochs	<b>Top-1(%)</b>	<b>Top-5</b> (%)
CMC[29]	ResNet-50	24	256	200	66.20	87.00
MoCo[17]	ResNet-50	24	256	200	72.80	91.64
Biased CMC [12]	ResNet-50	24	256	200	73.58	92.06
Debiased CMC [12]	ResNet-50	24	256	200	74.60	92.08
MoCo+align/uniform[33]	ResNet-50	24	256	200	74.60	92.74
NPID [35]	ResNet-50	24	256	200	75.30	_
BYOL [15]	ResNet-50	24	4096	1000	75.80	_
PCL-v1 [22]	ResNet-50	24	256	200	76.17	93.52
PCL-v2 [22]	ResNet-50+MLP	28	256	200	78.35	94.25
SwAV [5]	ResNet-50	24	4096	200	80.20	95.02
LooC [36]	ResNet-50	24	256	200	81.10	95.30
CLD [34]	ResNet-50	24	256	200	<u>81.50</u>	<u>95.48</u>
PAUC (ours)	ResNet-50	24	256	200	84.46	97.15

## Alignment

• To pull embeddings from positive prototypes together and push negative prototypes away:

 $\mathcal{L}_{p-align}(\mathbf{c}^m;s) =$ 

## Uniformity

• To alleviate the inter-prototype collapsing issues:

 $\mathcal{L}_{\text{p-uniform}}(\mathbf{c}^m;t) =$ 

## Correlation

avoid inter-prototype collapsing:

 $\mathcal{L}_{p-\mathrm{corr}} = \mathbb{E}_{(p,q)\sim}$ 

## **Qualitative Comparison**



$$\mathbb{E}_{(p,q)\sim p_{\text{pos}}}[\|\mathbf{c}_p^m - \mathbf{c}_q^m\|_2^s]$$

$$= \mathbb{E}_{(p,q)\sim p_{\text{proto}}} \left[ e^{-t \|\mathbf{c}_p^m - \mathbf{c}_q^m\|_2^2} \right]$$

To distinguish the difference between each prototype further to

$$\sum_{p_{\text{proto}}} [\mathbf{c}_p^m \log (\mathbf{c}_q^m \odot (\mathbf{c}_p^m)^{-1})]$$

#### **Comparison on ImageNet-1K**

Method	Arch.	Param.(M)	Batch	Epochs	<b>Top-1(%)</b>			
Instance-wise Contrastive:								
CPC [25]	ResNet-101	28	512	200	48.70			
MoCo [17]	ResNet-50	24	256	200	60.60			
PIRL [23]	ResNet-50	24	1024	800	63.60			
CMC [29]	ResNet-50+MLP{ $L,ab$ }	47	256	200	64.00			
CPCv2 [18]	ResNet-170	303	512	200	65.90			
MoCo+align/uniform [15]	ResNet-50	24	256	200	67.69			
AMDIM [1]	Custom-ResNet	192	1008	150	68.10			
LoCo [37]	ResNet-50	24	4096	800	69.50			
SimCLR [7]	ResNet-50+MLP	28	4096	400	70.00			
InfoMin [30]	ResNet-50	24	256	200	70.10			
MoCHi [21]	ResNet-50+MLP	28	512	200	70.60			
PIC [4]	ResNet-50	24	512	1600	70.80			
SWD [6]	ResNet-50+MLP	28	2048	800	70.90			
MoCov2 [10]	ResNet-50+MLP	28	256	200	71.10			
SimSiam [9]	ResNet-50+MLP	28	256	800	71.30			
SimCLRv2 [8]	ResNet-50+MLP	28	4096	800	71.70			
MoCov3 [11]	ResNet-50+MLP	28	4096	300	72.80			
VICReg [2]	ResNet-50+MLP	28	2048	1000	73.20			
AdCo [27]	ResNet-50	24	256	200	73.20			
Barlow Twins [38]	ResNet-50	24	2048	1000	73.20			
BYOL [15]	ResNet-50+MLP	35	4096	400	73.20			
BYOL [15]	ResNet-50+MLP	35	4096	1000	74.30			
MoCov3 [11]	ResNet-50+MLP	28	4096	1000	<u>74.60</u>			
Prototypical Contrastive:								
PCL [22]	ResNet-50	24	256	200	61.50			
PCLv2 [22]	ResNet-50+MLP	28	256	200	67.60			
CLD [34]	ResNet-50	24	256	200	71.50			
SwAV [5]	ResNet-50+MLP	28	256	200	72.70			
PAUC (ours)	ResNet-50	24	256	200	75.16			



where  $\alpha$ ,  $\beta$ , and  $\gamma$  denote the weights of the alignment, uniformity, and correlation loss.



 $\mathcal{L} = \mathcal{L}_{\text{InfoNCE}} - \sum_{i=1}^{n} \frac{1}{M} \sum_{m=1}^{M} \left( \log \frac{\exp(\mathbf{v}_i \cdot \mathbf{c}_s^m / \phi_s^m)}{\sum_{i=1}^{r} \exp(\mathbf{v}_i \cdot \mathbf{c}_i^m / \phi_i^m)} \right)$ 

 $+ \alpha \mathcal{L}_{p-align} + \beta \mathcal{L}_{p-uniform} + \gamma \mathcal{L}_{p-corr}$