Why Prototypes Collapse: Diagnosing and Preventing Partial Collapse in Prototypical Self-Supervised Learning

Gabriel Y. Arteaga, Marius Aasan, Rwiddhi Chakraborty, Martine Hjelkrem-Tan, Thalles Silva, Michael Kampffmeyer and Adín Ramírez Rivera

University of Oslo, UiT The Arctic University of Norway, University of Campinas

Motivation and Proposed Solution

In prototypical SSL, joint encoder–prototype optimization creates shortcuts that lead to prototype collapse.

- **Problem?** Lack of target diversity hurts representation learning, especially in long-tailed settings, and the common remedy of prototype overparameterization leads to substantial redundancy during training.
- **Solution:** Decouple the joint optimization of encoder f_θ and prototypes C by alternating between two updates at itertation t:

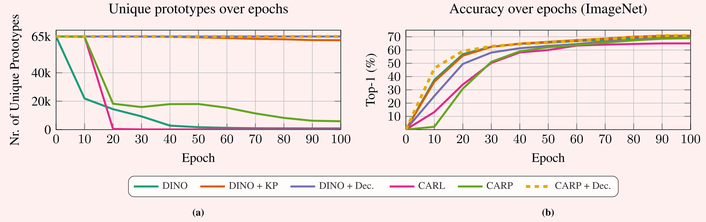
$$\text{(i) Prototype Update} \qquad C^t = \underset{C \in \mathcal{C}}{\operatorname{argmin}}\, \mathcal{L}_C\left(C^{t-1}, h_\phi^t\right) \tag{1}$$

$$\text{(ii) Encoder Update} \qquad \theta^{t+1} = \underset{\theta}{\operatorname{argmin}}\, \mathcal{L}_f\left(h_\theta^t, C^t\right) \tag{2}$$

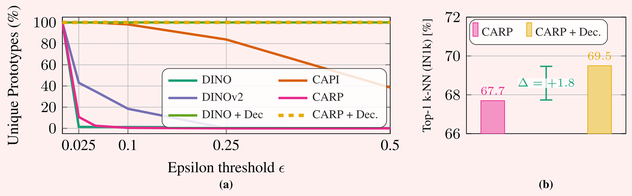
The prototype update (i) is solved via an online Gaussian Mixture Model using an EM-style procedure, independent of the encoder loss $\mathcal{L}_f$.

Training Dynamics

Collapse occurs early



Unique Prototypes with stricter thresholds ϵ



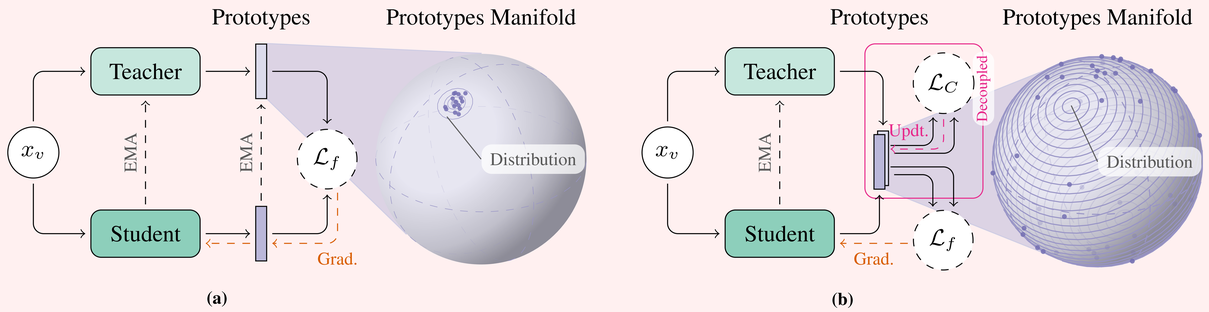
Training Time & Peak GPU Consumption

Reduces memory use

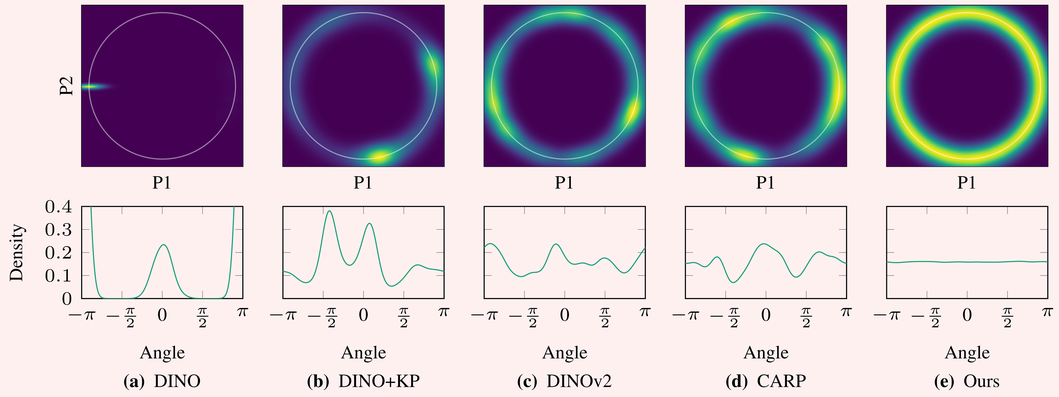
Method	Epochs	Batch Size	Total Crops	Time (h)
CARP	100	1024	12	37.9
CARP + Dec.	100	1024	12	38.1

Batch Size	CARP	CARP + Dec.
128	5.4G	5.3G ↓0.1
2048	62.4G	60.9G ↓1.5

Method Overview

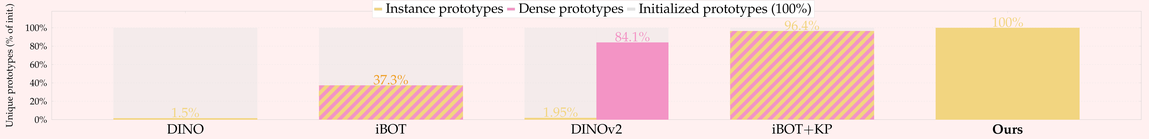


Prototype Distribution using Principal Components



Prototype Collpse: Instance vs Dense Prototypes

Instance Protos. ≪ Dense Protos.



Linear Classification on ImageNet-1k

Method	Backbone	Ep.	k-NN	Linear
DINO	RN-50	400	67.5	75.3
SWAV	RN-50	400	65.0	74.6
DeepCluster-v2	RN-50	800	66.6	75.2
CARP	RN-50	400	67.7	75.3
CARP + Dec.	RN-50	400	69.1	75.3
DINO	ViT-S/16	300	72.8	76.2
CARP	ViT-S/16	300	73.6	76.3
CARP + KP	ViT-S/16	300	73.7	76.1
CARP + Dec.	ViT-S/16	300	74.1	76.2
MSN	ViT-S/16	600	–	76.9
DINO	ViT-S/16	800	74.5	77.0
DINO-vMF	ViT-S/16	800	74.7	77.0
CARP + Dec.	ViT-S/16	800	75.3	76.4
DINO	ViT-B/16	400	76.1	78.2
DINO-vMF	ViT-B/16	400	77.4	78.8
CARP + Dec.	ViT-B/16	400	76.7	78.1

Transfer Learning

ResNet-50 (400 epochs)

Method	Air	C101	Cars	Flw	Food	Pets	SUN	VOC	Avg
DINO	60.0	90.9	65.9	95.6	78.4	89.0	66.1	84.3	78.8
CARP	61.0	91.7	64.2	95.8	78.9	90.3	66.1	84.5	79.1
CARP + Dec.	58.5	92.1	67.3	95.9	78.7	91.1	65.2	84.6	79.2

ViT-S/16 (300 epochs)

Method	Air	C101	Cars	Flw	Food	Pets	SUN	VOC	Avg
CARP	59.2	93.6	64.5	96.4	77.5	93.4	65.9	85.2	79.5
CARP + KP	60.5	93.5	63.6	96.4	78.6	93.4	66.0	85.3	79.7
CARP + Dec.	61.8	92.9	63.9	96.2	78.5	94.1	64.5	85.4	79.7

Linear Classificaiton on iNaturalist2018

Methods	Head (> 100)	Medium (> 20 & ≤ 100)	Tail (≤ 20)	All
DINO	55.2	46.2	41.9	45.3
DINO + KP	59.8 ↑4.6	50.4 ↑4.2	45.0 ↑3.1	49.0 ↑3.7
DINO + Decoupling	44.9 ↓10.3	37.3 ↓8.9	33.1 ↓8.8	36.2 ↓9.1
CARP	56.0	46.9	42.5	45.9
CARP + KP	55.5 ↓0.5	46.4 ↓0.5	41.8 ↓0.7	45.3 ↓0.6
CARP + Dec.	59.1 ↑3.1	49.3 ↑2.4	45.9 ↑3.4	48.9 ↑3.0

Sensitivty Analysis on DINO

Method	Dec.	Cent.	SK	# Prot.	T. temp.	k-NN
DINO		✓		65536	0.04 → 0.07	68.9
DINO	✓	✓		65536	0.04 → 0.07	27.8 ↓41.1
DINO	✓		✓	65536	0.04 → 0.07	62.2 ↓6.7
DINO	✓		✓	65536	0.05 → 0.025	67.7 ↓2.2
DINO	✓		✓	1024	0.05 → 0.025	68.0 ↓0.9
DINO	✓		✓	2048	0.05 → 0.025	68.7 ↓0.2

ICLR 2026 — International Conference on Learning Representations — April, 2026 — Rio de Janeiro, Brazil

Acknowledgements: EuroHPC LUMI (Sigma2 NN8104K) & SFI Visual Intelligence (RCN 309439) & FRIPRO (grants no. 315029 and 359216)