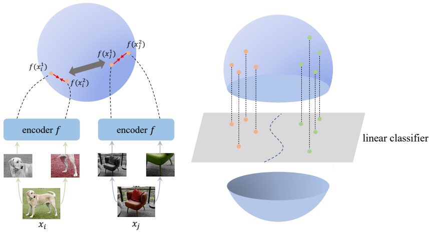
Hyperspherical Consistency Regularization

Cheng Tan, Zhangyang Gao, Lirong Wu, Siyuan Li, Stan Z. Li
AI Lab, School of Engineering, Westlake University.
Institute of Advanced Technology, Westlake Institute for Advanced Study

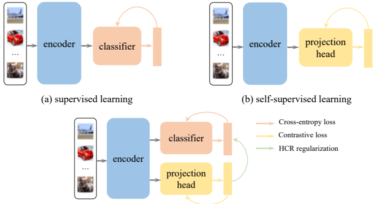
Section 1 Contrastive learning on the hypersphere

- Contrastive learning usually projects feature embeddings on a hypersphere while maximizing distances between negative pairs and minimizing distances between positive pairs.
- Linear classifier learns to separate the hypersphere through the hyperplane.

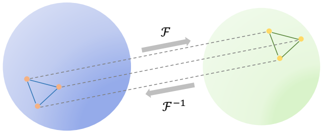


Section 2 Hypersphere consistency regularization

- Conceptual illustration of different learning paradigms.
- Supervised learning and self-supervised learning learn proper representations through different pretext tasks relying on label-dependent and feature-dependent information respectively.
- HCR takes supervised learning as the primary task and forces self-supervised learning to assist it from another perspective.



(a) supervised learning
(b) self-supervised learning
(c) learning with HCR



projection head's hypersphere
classifier's hypersphere

- Define the objective of HCR:

$$HCR\left(p(d_g), q(d_h)\right) = -p(d_g)\log\left(q(d_h)\right) - (1 - p(d_g))\log(1 - q(d_h))$$

where,

$$p(d_g) = C_g \frac{1}{\sigma_g\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(d_g-\mu_g)^2}{\sigma_g^2}\right]$$

$$q(d_h) = C_h \frac{1}{\sigma_h\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(d_h-\mu_h)^2}{\sigma_h^2}\right]$$

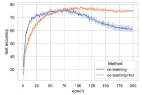
Section 3 Experimental results

- For semi-supervised learning on Stanford Cars, HCR significantly improves the performance of Self-Tuning by an average of 2.30% in different label proportions.
- In fine-grained classification, HCR performs consistently better than the baselines.
- For noisy label learning on CIFAR-100 with symmetric noises, HCR further improves Co-learning by averagely 3.92% under different noise ratios.

Method	Label Proportion 15%	30%	50%
Pseudo-Labeling	40.93±0.23	67.02±0.19	78.71±0.30
Π-model	45.19±0.21	57.29±0.26	64.18±0.29
Mean Teacher	54.28±0.14	66.02±0.21	74.24±0.23
UDA	39.90±0.43	64.16±0.40	71.86±0.56
FixMatch	49.86±0.27	77.54±0.29	84.78±0.33
SimCLRv2	45.74±0.16	61.70±0.18	77.49±0.24
Self-Tuning	74.99±0.11	85.87±0.04	89.83±0.01
Self-Tuning+HCR	**78.76**±0.08	**87.70**±0.07	**91.14**±0.06

Method	Stanford Cars	Aircraft	CUB200
Fine-Tuning	87.20±0.19	81.13±0.21	78.01±0.16
L^2-SP	86.58±0.26	80.98±0.29	78.44±0.17
DELTA	86.32±0.20	80.44±0.20	78.63±0.18
BSS	87.63±0.27	81.48±0.18	78.85±0.31
Co-Tuning	89.53±0.09	83.87±0.09	81.24±0.14
Self-Tuning	92.33±0.10	88.96±0.21	81.60±0.11
Self-Tuning+HCR	**93.03**±0.06	**90.41**±0.03	**82.63**±0.19

Method	sym-20%	sym-50%	sym-80%
Standrad CE	57.79±0.44	33.75±0.46	8.64±0.22
Decoupling	56.18±0.32	31.58±0.54	7.71±0.23
Co-teaching	64.28±0.32	32.62±0.51	6.65±0.71
Co-teaching+	55.40±0.71	26.49±0.45	8.57±1.55
JoCoR	62.29±0.71	30.19±0.60	6.84±0.92
Co-learning	66.58±0.15	55.54±0.43	35.45±0.79
Co-learning+HCR	**70.27**±0.32	**59.93**± 0.25	**39.14**±0.47



- More experiments can be found in our paper.

Summary

- We propose hyperspherical consistency regularization (HCR), to encourage the pairwise distance distribution of the classifier to be similar to the distribution of the projection head in the latent space.
- Through extensive experiments on semi-supervised learning, fine-grained classification and noisy label learning, HCR shows consistent improvements on these tasks.