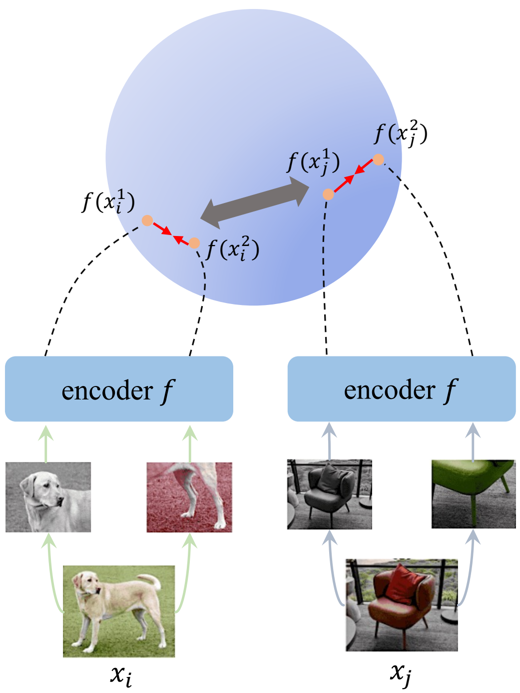
Hyperspherical Consistency Regularization

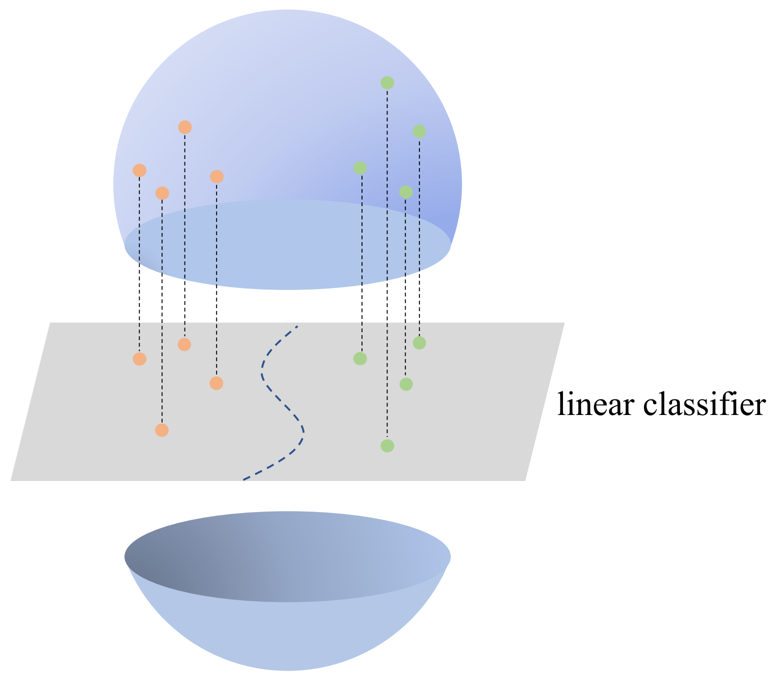
Cheng Tan, Zhangyao Gao, Lirong Wu, Siyuan Li, Stan Z. Li
AI Lab, School of Engineering, Westlake University.
Institute of Advanced Technology, Westlake Institute for Advanced Study.

tancheng@westlake.edu.cn
CVPR, 2022

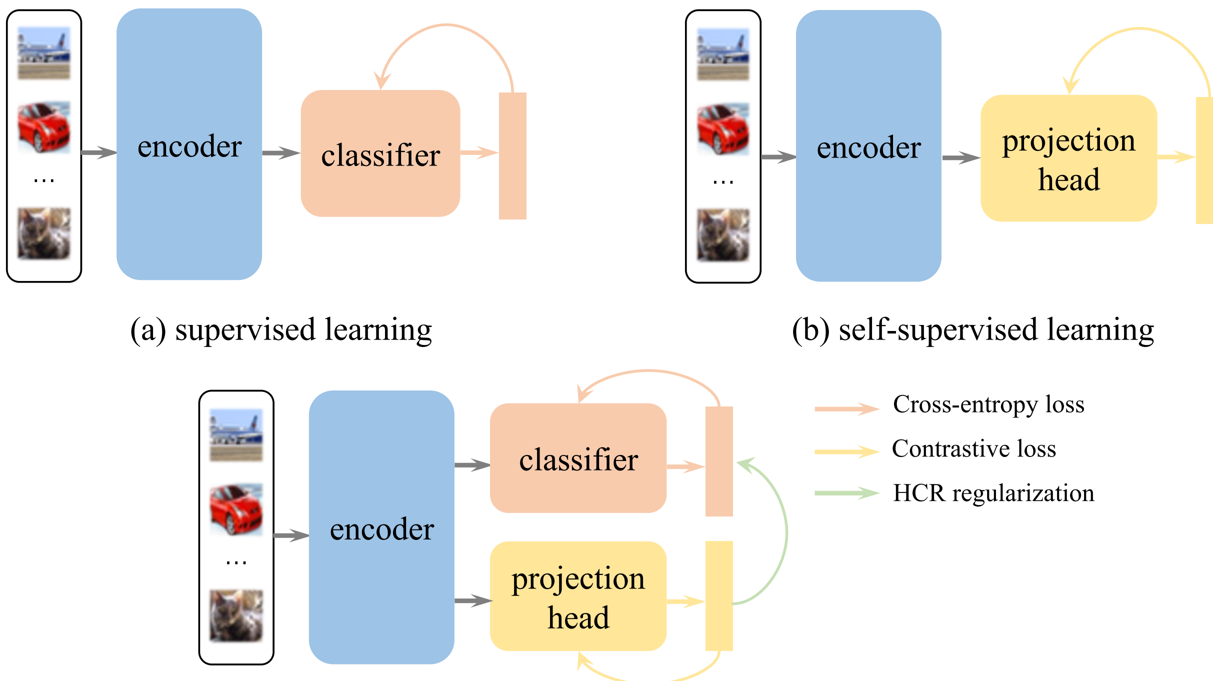
Contrastive learning on the hypersphere

Linear classifier learns to separate the hypersphere through the hypersphere.





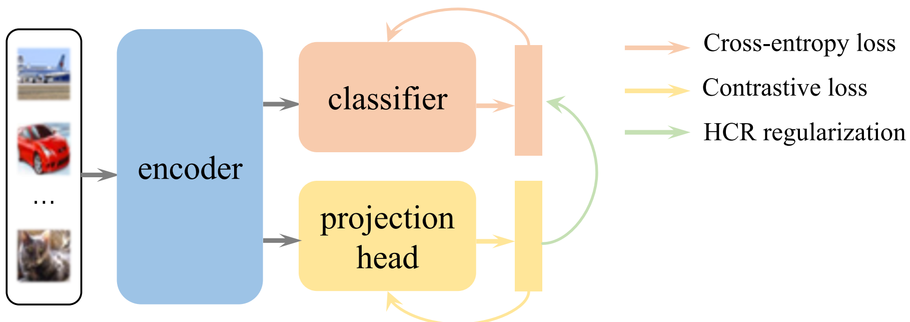
Learning paradigms

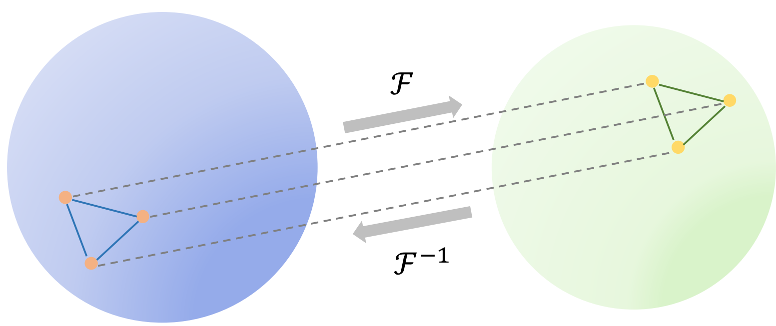


(c) learning with HCR

HCR takes supervised learning as the primary task and forces self-supervised learning to assist it from another perspective.

Hypersphere consistency regularization





projection head's hypersphere classifier's hypersphere

Define the objective of HCR:

$$
\begin{aligned}
&HCR\left(p(d_g), q(d_h)\right) \\
&= -p(d_g)\log\left(q(d_h)\right) - (1 - p(d_g))\log(1 - q(d_h))
\end{aligned}
$$

where,

$$
\begin{aligned}
p(d_g) &= C_g \frac{1}{\sigma_g \sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{\left(d_g - \mu_g\right)^2}{\sigma_g^2}\right], \\
q(d_h) &= C_h \frac{1}{\sigma_h \sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(d_h - \mu_h)^2}{\sigma_h^2}\right].
\end{aligned}
$$

Experiments – Semi-supervised learning

西湖大學 WESTLAKE UNIVERSITY

Classification accuracy (%) on Stanford Cars dataset.

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

Classification accuracy (%) on FGVC Aircraft dataset.

Method	Label Proportion 15%	30%	50%
Pseudo-Labeling	46.83±0.30	62.77±0.31	73.21±0.39
Π-model	37.72±0.25	58.49±0.26	65.63±0.36
Mean Teacher	51.59±0.23	71.62±0.29	80.31±0.32
UDA	43.96±0.45	64.17±0.49	67.42±0.53
FixMatch	55.53±0.26	71.35±0.35	78.34±0.43
SimCLRv2	40.78±0.21	59.03±0.29	68.54±0.30
Self-Tuning	66.68±0.17	79.94±0.09	84.35±0.08
Self-Tuning+HCR	**70.54**±0.02	**82.64**±0.04	**86.89**±0.15

Classification accuracy (%) on CUB-200-2011 dataset.

Method	Label Proportion 15%	30%	50%
Pseudo-Labeling	45.33±0.23	56.20±0.29	64.07±0.32
Π-model	45.20±0.25	58.49±0.26	65.63±0.36
Mean Teacher	53.26±0.19	66.66±0.20	74.37±0.30
UDA	46.90±0.31	61.16±0.35	71.86±0.43
FixMatch	44.06±0.23	63.54±0.18	75.96±0.29
SimCLRv2	45.74±0.15	62.70±0.24	71.01±0.34
Self-Tuning	64.79±0.06	74.31±0.07	78.45±0.31
Self-Tuning+HCR	**66.42**±0.24	**75.06**±0.13	**79.48**±0.16

Classification accuracy (%) of transfer learning methods.

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

Experiments – Noisy label learning

CIFAR-100 with symmetric noises.

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

CIFAR-100 with asymmetric noises.

Method	asym-20%	asym-30%	asym-40%
Standrad CE	59.36±0.36	51.06±0.44	42.49±0.23
Decoupling	57.97±0.24	49.86±0.54	41.51±0.67
Co-teaching	59.76±0.53	49.53±0.79	40.62±0.79
Co-teaching+	56.11±0.60	47.12±0.73	38.98±0.54
JoCoR	58.58±0.51	49.04±0.91	39.72±0.76
Co-learning	65.26±0.76	56.97±1.22	47.62±0.79
Co-learning+HCR	**68.85**±0.22	**61.94**±0.17	**50.29**±0.69

CIFAR-100 with instance-dependent noises.

Method	ins-20%	ins-30%	ins-40%
Standrad CE	55.45±0.54	48.77±0.47	41.30±0.27
Decoupling	52.20±0.48	45.32±0.83	36.33±0.47
Co-teaching	55.16±0.61	45.24±0.37	34.64±1.00
Co-teaching+	50.37±0.85	40.73±0.58	32.15±0.80
JoCoR	54.21±0.34	45.03±0.52	34.08±1.05
Co-learning	69.42±0.42	65.45±0.86	60.40±1.37
Co-learning+HCR	**70.03**±0.31	**66.89**±0.41	**62.91**±0.84

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.