COMPUTER SCIENCE & ENGINEERING
TEXAS A&M UNIVERSITY

TEXAS
The University of Texas at Austin

Self-Damaging Contrastive Learning

Ziyu Jiang[1], Tianlong Chen[2], Bobak Mortazavi[2], Zhangyang Wang[2],

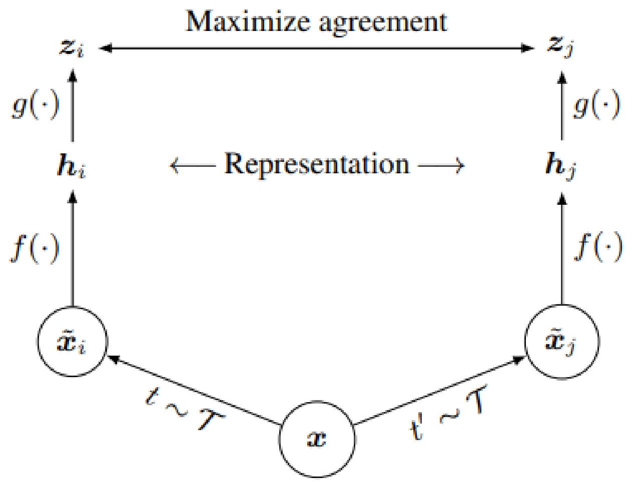
[1]Texas A&M University [2]University of Texas at Austin

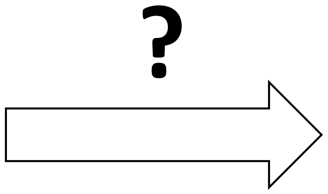
Code VITA Homepage

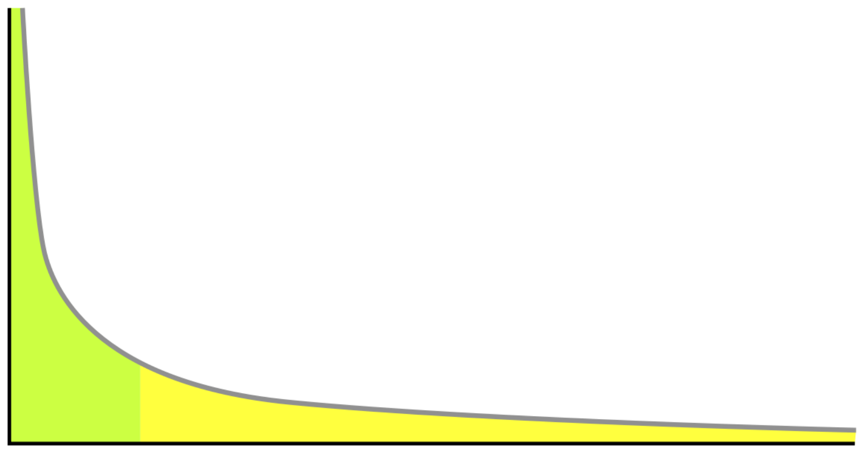
VITA

Background

- Contrastive learning achieved great success to learn powerful visual representation
- How would contrastive learning perform on uncontrolled real-world data with long tail distribution?

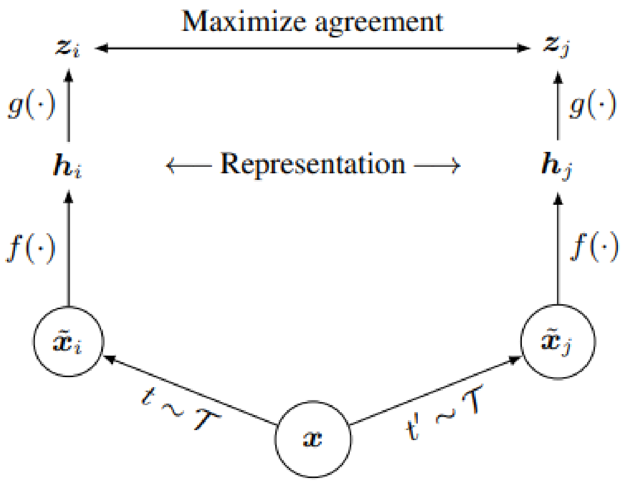


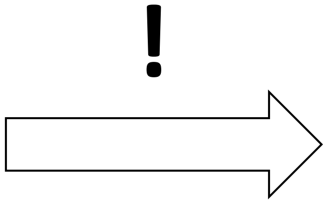


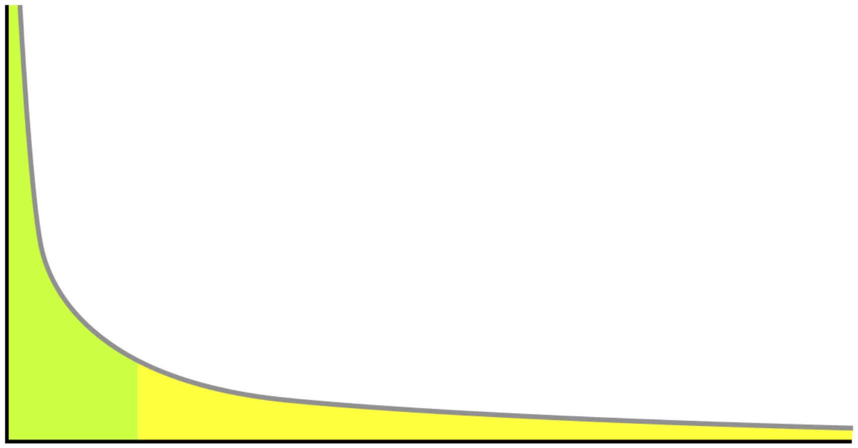


Observation

- Short answer: Contrastive Learning is not immune to long tail distribution.
- The long tail distribution would make contrastive learning biased to major classes, especially for few-shot downstream tasks.







How to resolve

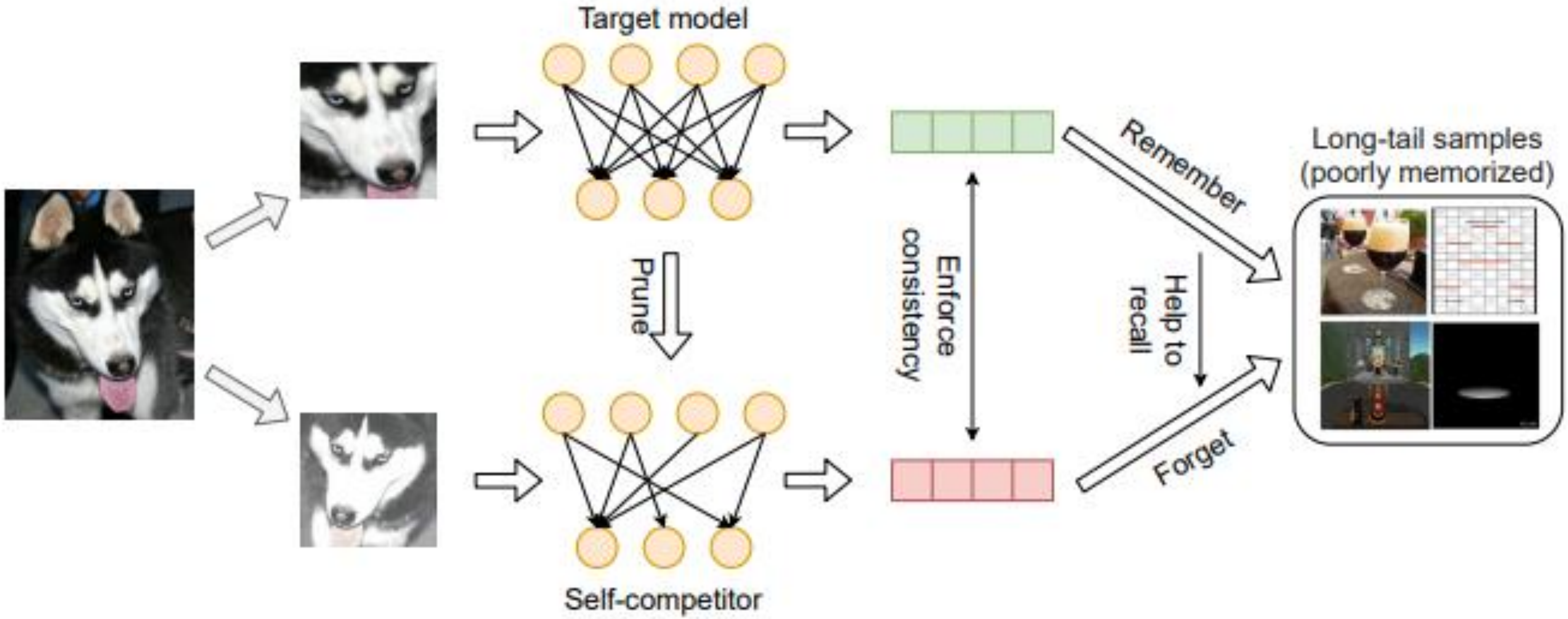
- Challenge: the absence of class information.
- The traditional strategies like re-sampling and re-balancing cannot be straightforwardly made to work here.
- We need to find tail class first.

The blessing from PIE samples

- Pruning Identified Exemplars (PIEs) are images where there is a high level of disagreement between the predictions of pruned and non-pruned models. PIEs can be thought as images that are forgotten after pruning.
- They are high likely to be rare and atypical samples, which probably comes from tail class. [1]

[1] Hooker, Sara, et al. "What Do Compressed Deep Neural Networks Forget?." arXiv preprint arXiv:1911.05248 (2019).

Framework



Target model
Prune
Self-competitor
Enforce consistency
Remember
Help to recall
Forget
Long-tail samples (poorly memorized)

Experiment

- ***Linear separability performance***

Dataset	Framework	*Many* ↑	*Medium* ↑	*Few* ↑	*Std* ↓	All ↑
CIFAR10-LT	SimCLR	78.18 ± 4.18	76.23 ± 5.33	71.37 ± 7.07	5.13 ± 3.66	75.55 ± 0.66
	SDCLR	86.44 ± 3.12	81.84 ± 4.78	76.23 ± 6.29	5.06 ± 3.91	82.00 ± 0.68
CIFAR100-LT	SimCLR	50.10 ± 1.70	47.78 ± 1.46	43.36 ± 1.64	3.09 ± 0.85	47.11 ± 0.34
	SDCLR	58.54 ± 0.82	55.70 ± 1.44	52.10 ± 1.72	2.86 ± 0.69	55.48 ± 0.62
ImageNet-100-LT	SimCLR	69.54	63.71	59.69	4.04	65.46
	SDCLR	70.10	65.04	60.92	3.75	66.48

Experiment

- ***Few-shot performance***

Dataset	Framework	*Many* ↑	*Medium* ↑	*Few* ↑	*Std* ↓	All ↑
CIFAR10	SimCLR	76.07 ± 3.88	67.97 ± 5.84	54.21 ± 10.24	9.80 ± 5.45	67.08 ± 2.15
	SDCLR	76.57 ± 4.90	70.01 ± 7.88	62.79 ± 7.37	6.99 ± 5.20	70.47 ± 1.38
CIFAR100	SimCLR	30.72 ± 2.01	21.93 ± 2.61	15.99 ± 1.51	6.27 ± 1.20	22.96 ± 0.43
	SDCLR	29.72 ± 1.52	25.41 ± 1.91	20.55 ± 2.10	3.98 ± 0.98	25.27 ± 0.83
Imagenet-100-LT	SimCLR	48.36	39.00	35.23	5.52	42.16
	SDCLR	48.31	39.17	36.46	5.07	42.38

COMPUTER SCIENCE & ENGINEERING
TEXAS A&M UNIVERSITY

TEXAS
The University of Texas at Austin

Thank you for listening

Code | VITA | Homepage

VITA