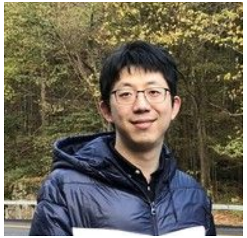
Rethinking the Value of Labels for Improving Class-Imbalanced Learning



Yuzhe Yang

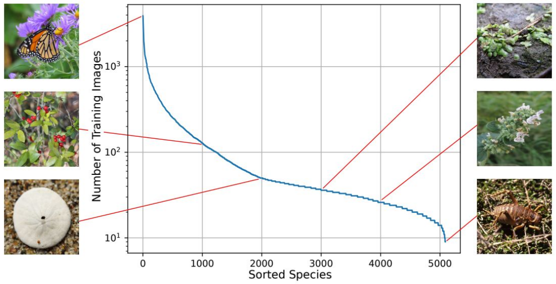
Zhi Xu

NeurIPS 2020

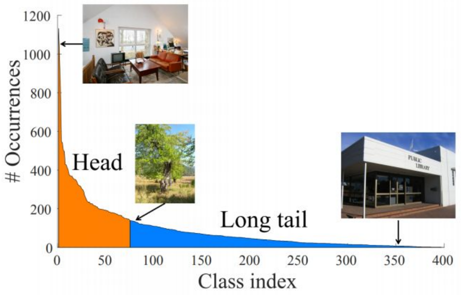
http://www.mit.edu/~yuzhe/imbalanced-semi-self.html

Real-world data are often imbalanced (long-tailed)

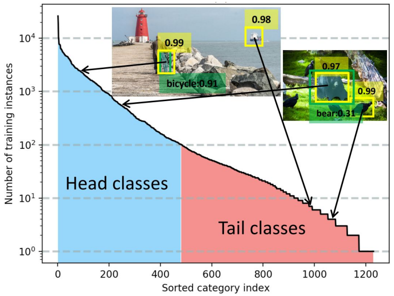
- Image recognition / Object detection / Semantic segmentation ...



Van Horn et al. 2018



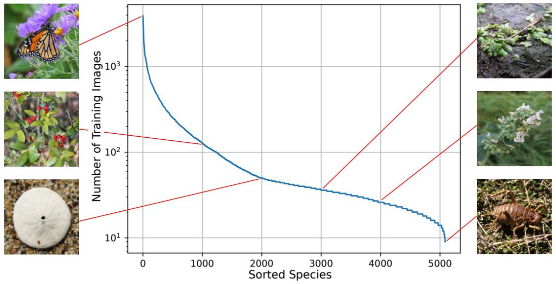
Wang et al. 2017



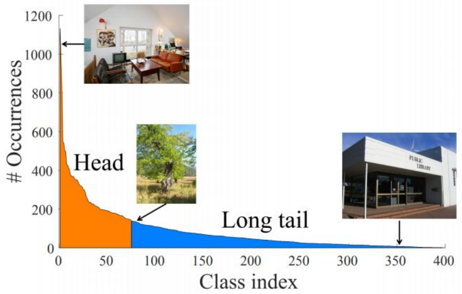
Gupta et al. 2018

Real-world data are often imbalanced (long-tailed)

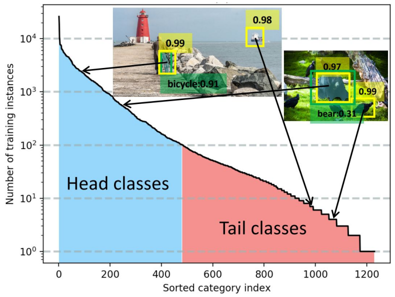
- Image recognition / Object detection / Semantic segmentation ...



Van Horn et al. 2018



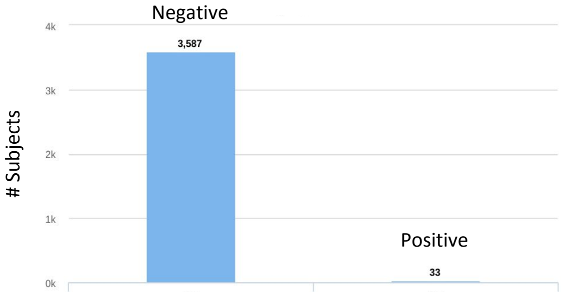
Wang et al. 2017



Gupta et al. 2018

- Critical applications

Autonomous driving

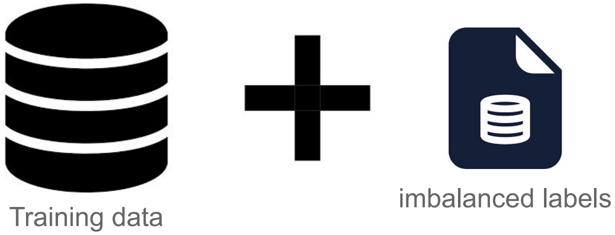


Medical diagnosis

Dilemma: Value of imbalanced labels

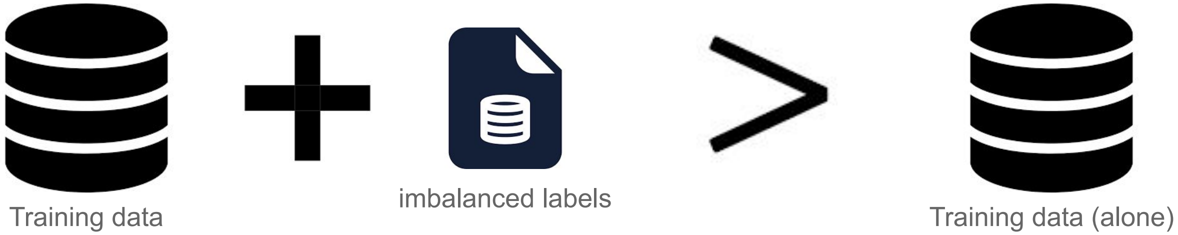
Dilemma: Value of imbalanced labels

- Positive value



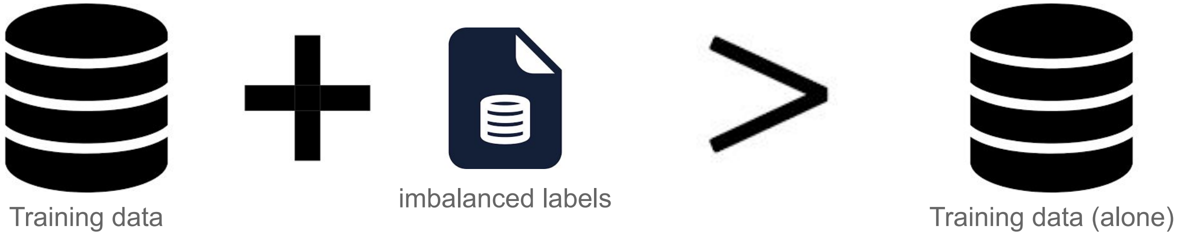
Dilemma: Value of imbalanced labels

- Positive value

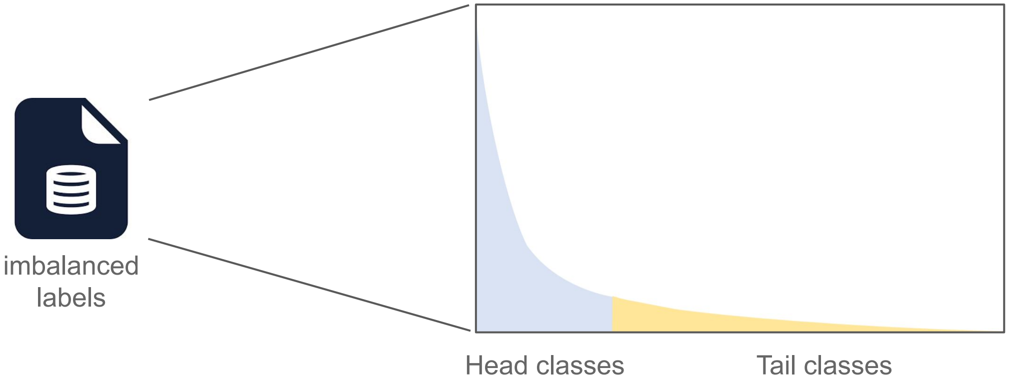


Dilemma: Value of imbalanced labels

- Positive value

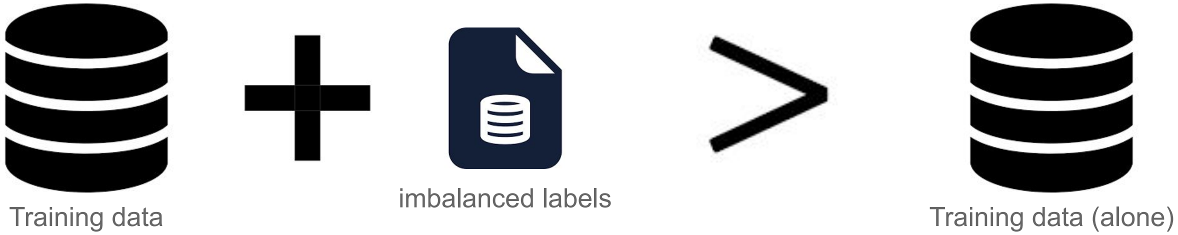


- Negative value

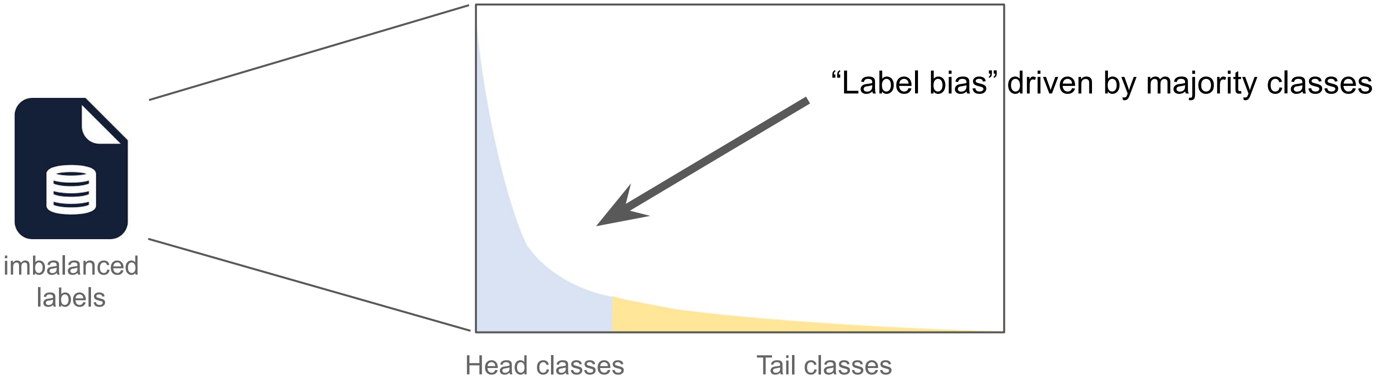


Dilemma: Value of imbalanced labels

- Positive value



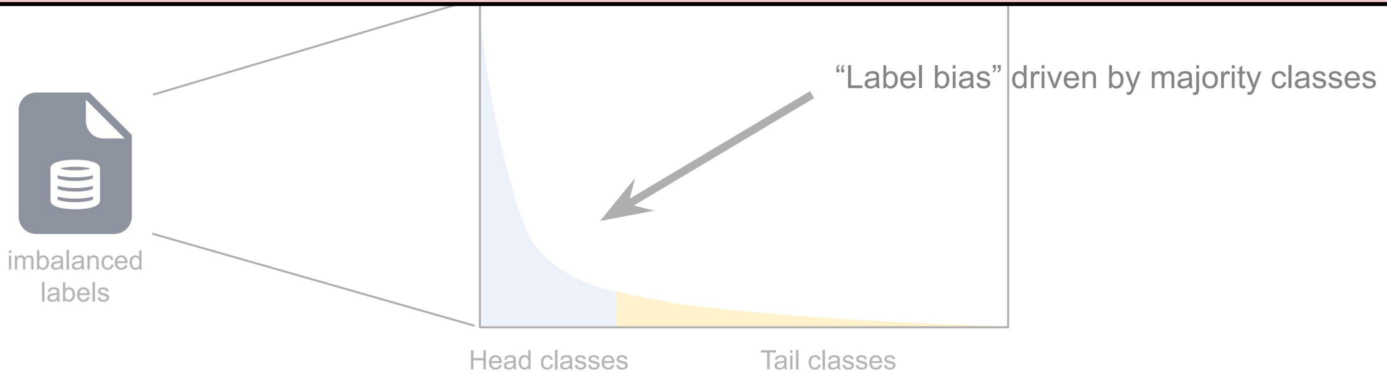
- Negative value



Dilemma: Value of imbalanced labels

- Positive value

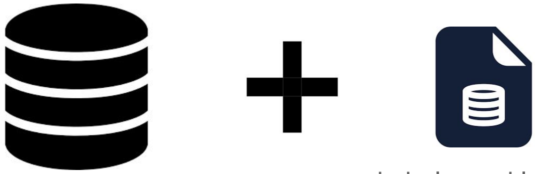
How to exploit the value of imbalanced labels?



Semi-supervision & Self-supervision help!

Semi-supervision & Self-supervision help!

- Positive viewpoint

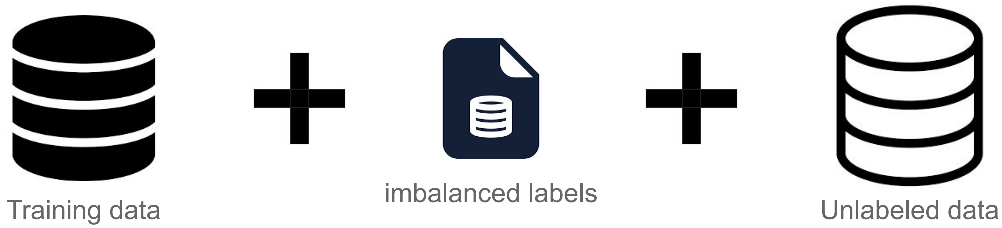


Training data

imbalanced labels

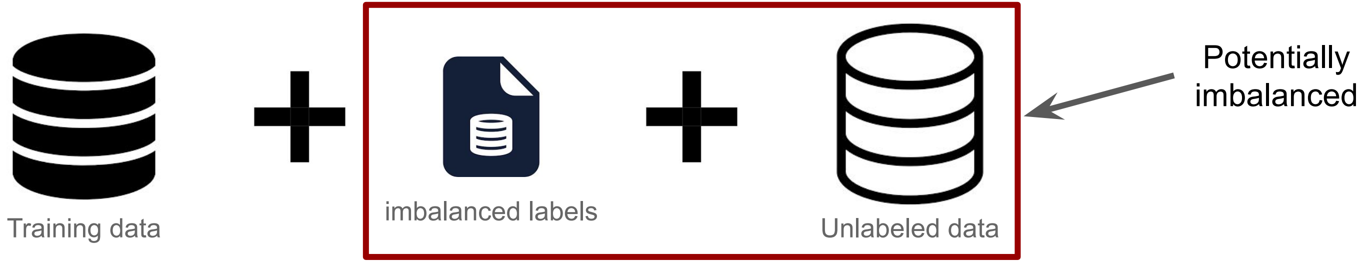
Semi-supervision & Self-supervision help!

- Positive viewpoint



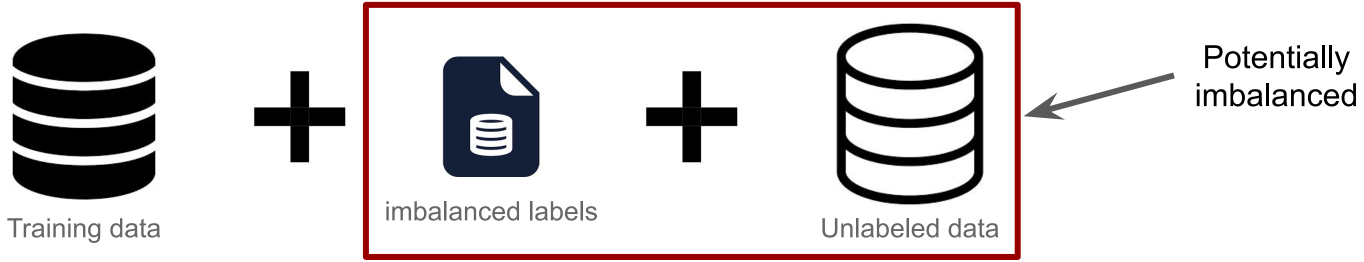
Semi-supervision & Self-supervision help!

- Positive viewpoint: Semi-supervised learning using imbalanced labels

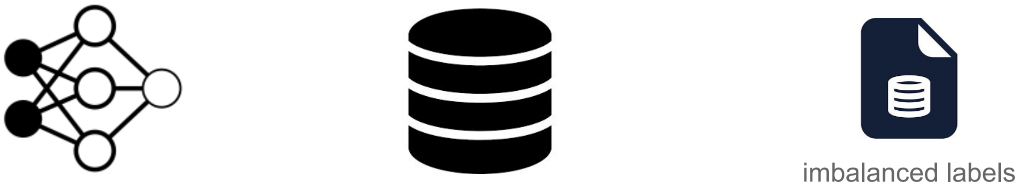


Semi-supervision & Self-supervision help!

- Positive viewpoint: Semi-supervised learning using imbalanced labels

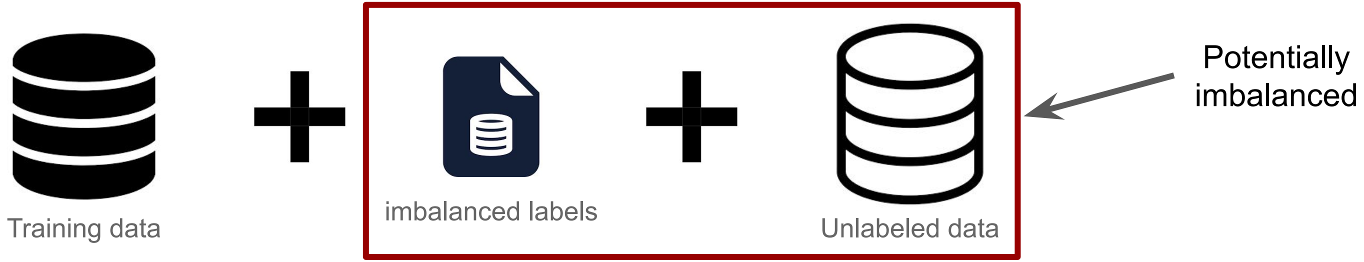


- Negative viewpoint

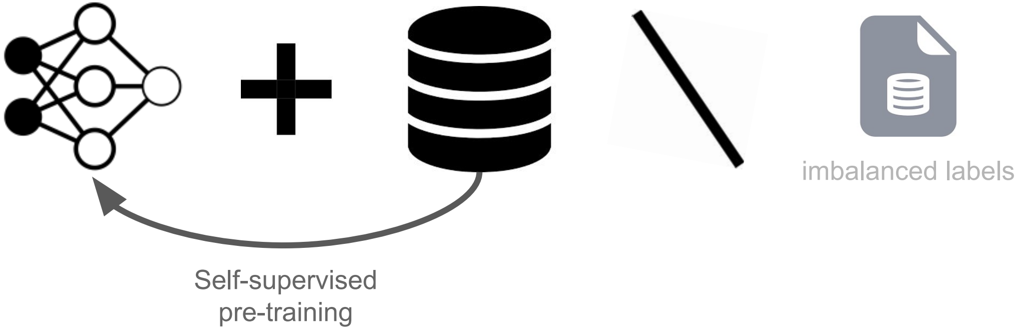


Semi-supervision & Self-supervision help!

- Positive viewpoint: Semi-supervised learning using imbalanced labels

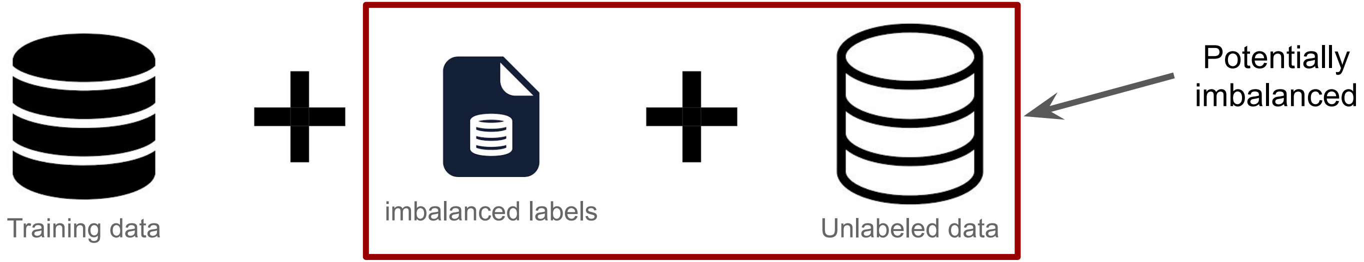


- Negative viewpoint: Self-supervised pre-training in the first learning stage

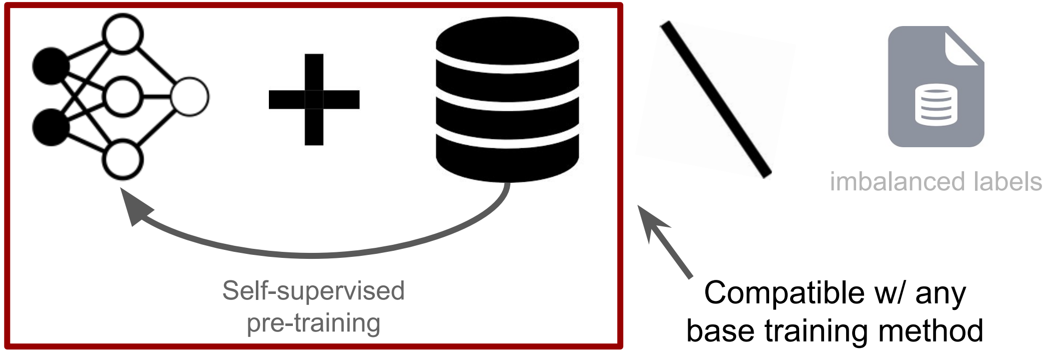


Semi-supervision & Self-supervision help!

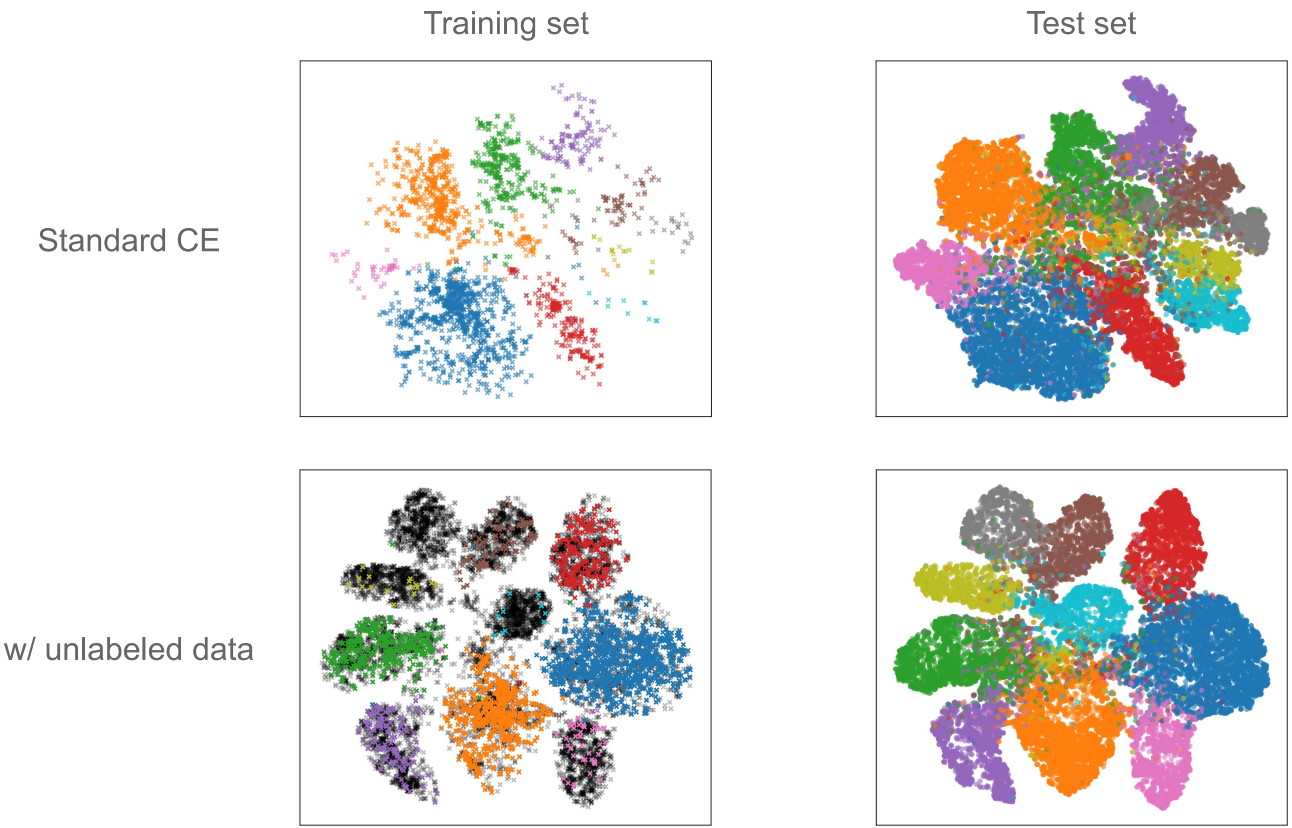
- Positive viewpoint: Semi-supervised learning using imbalanced labels



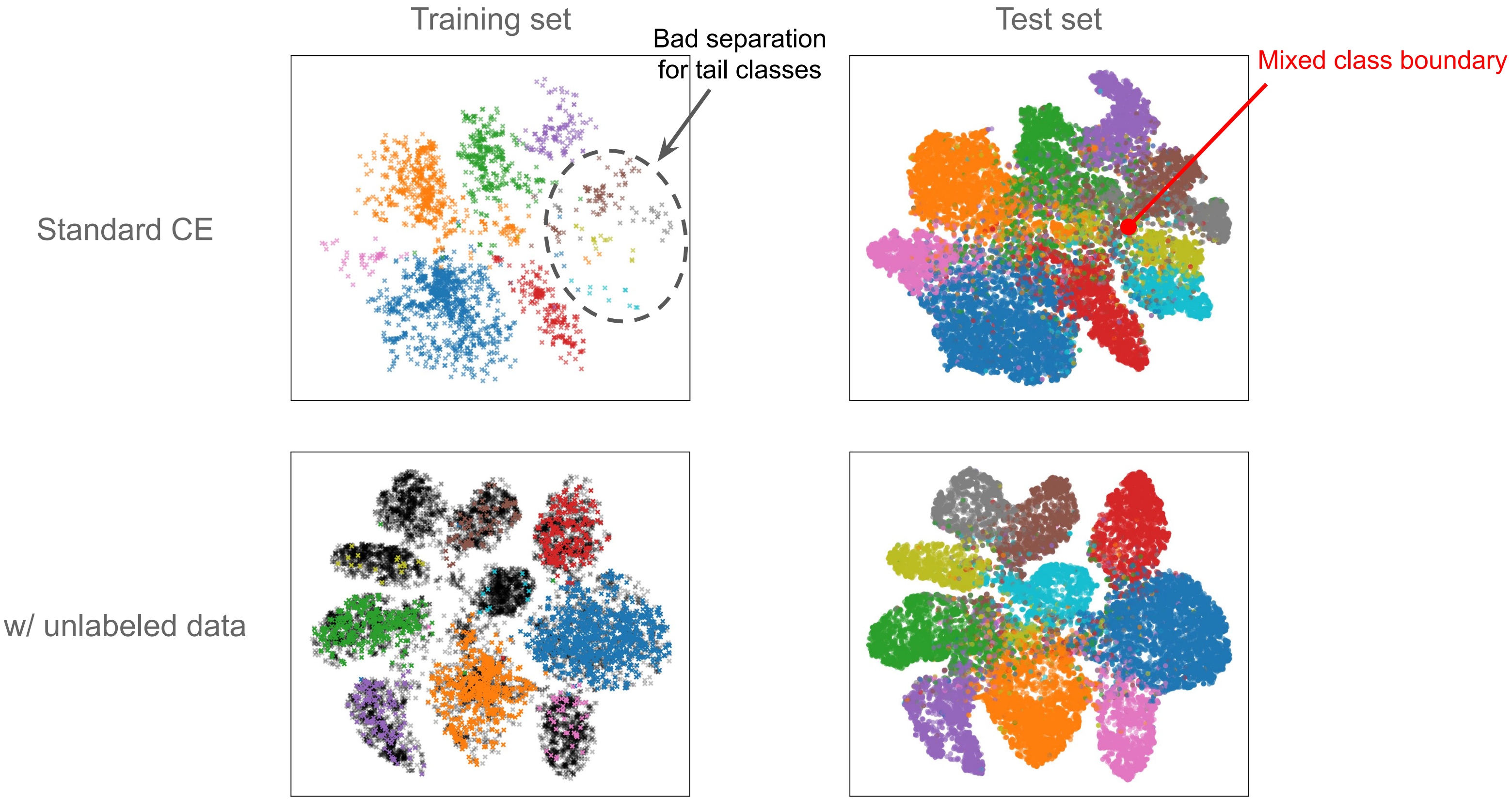
- Negative viewpoint: Self-supervised pre-training in the first learning stage



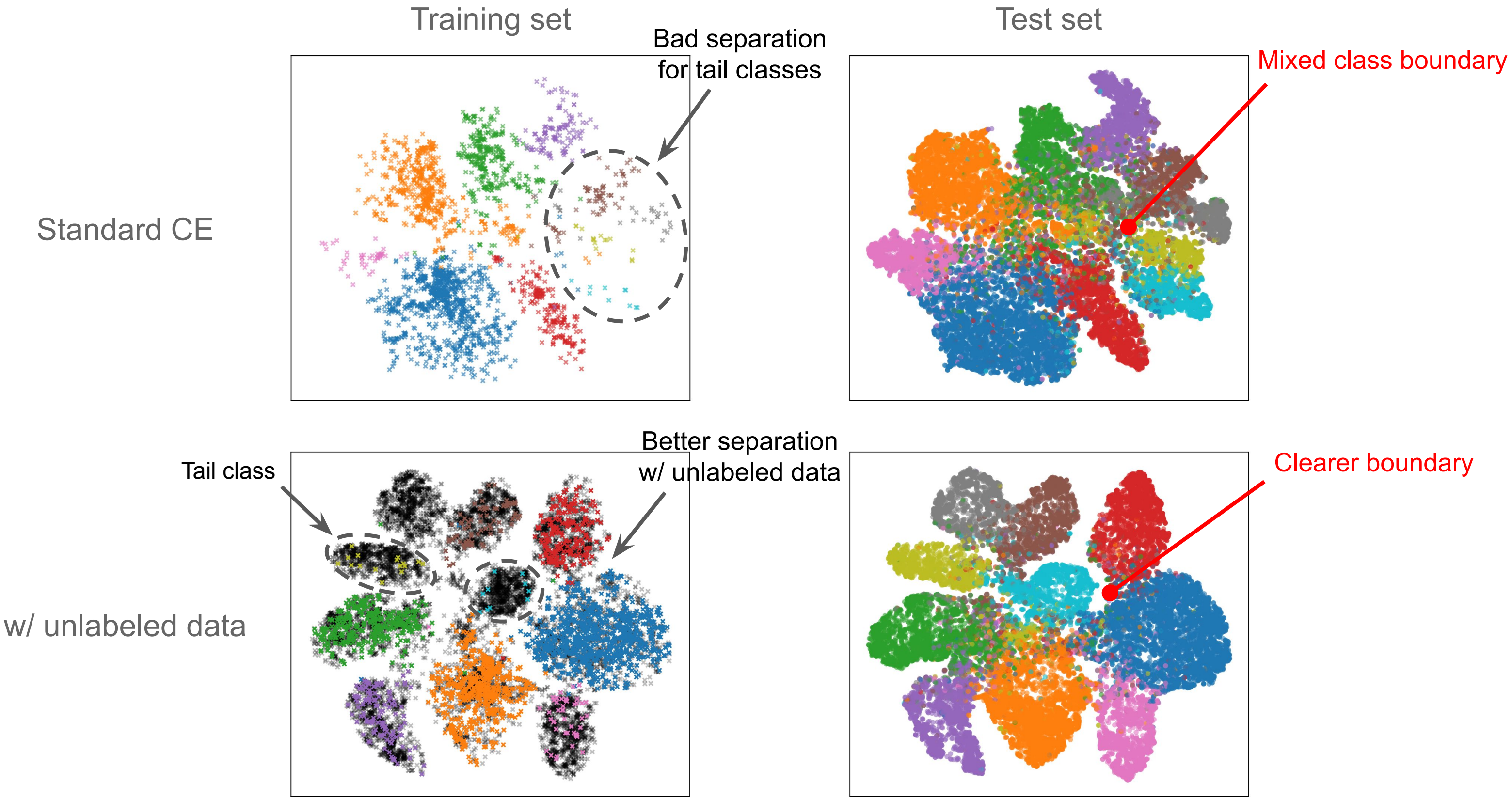
Semi-supervised imbalanced learning



Semi-supervised imbalanced learning



Semi-supervised imbalanced learning



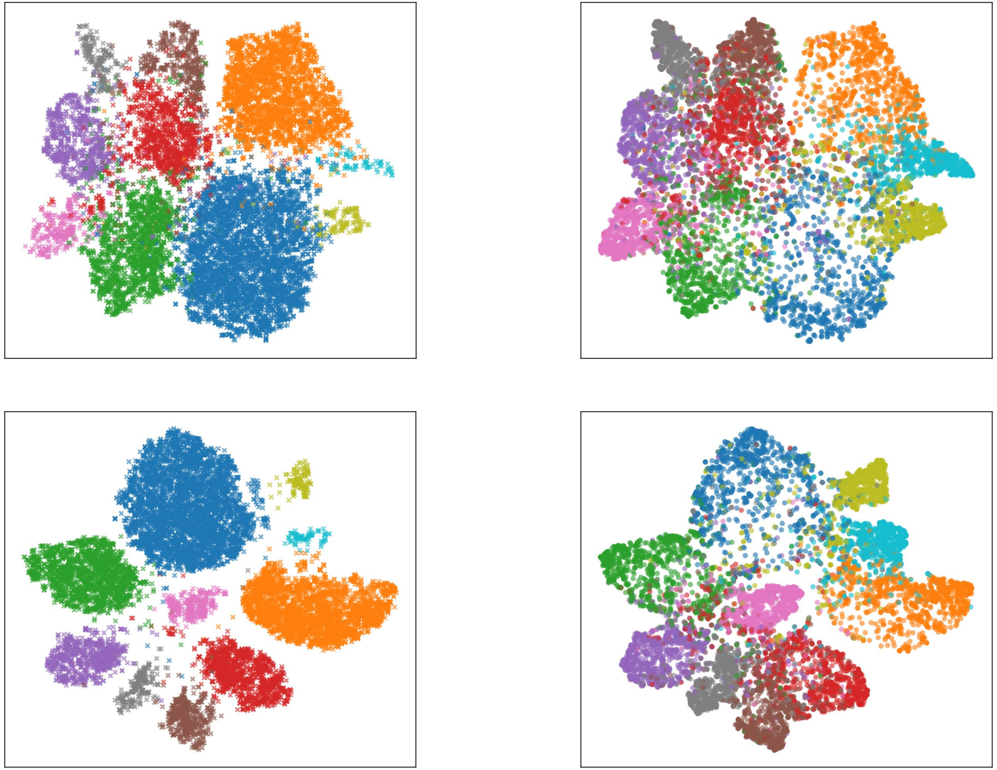
Self-supervised imbalanced learning

Training set

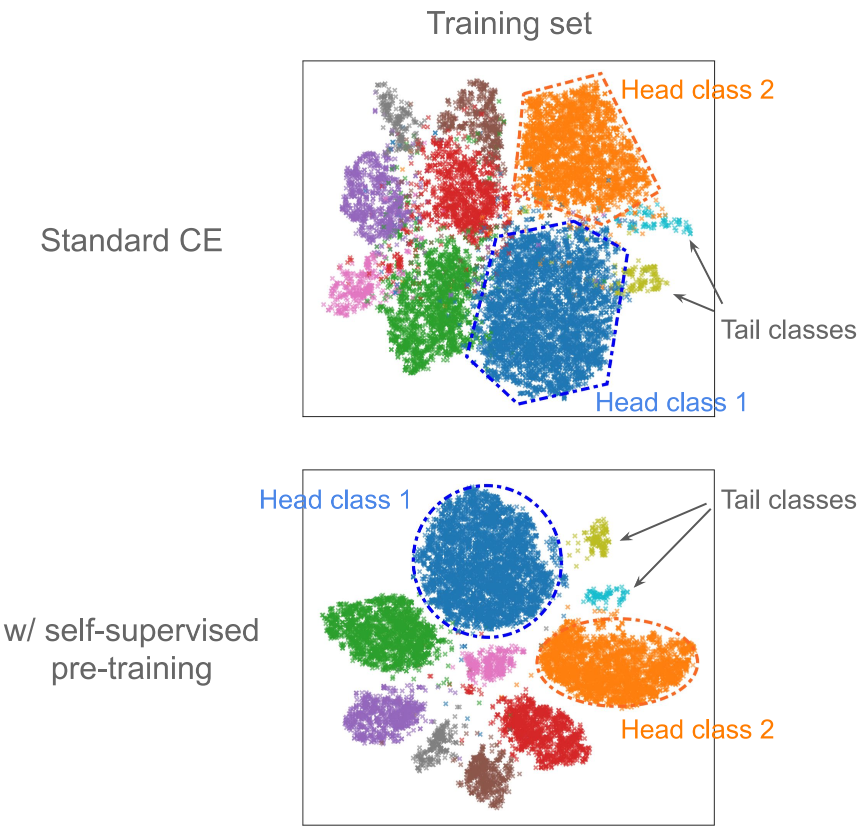
Test set

Standard CE

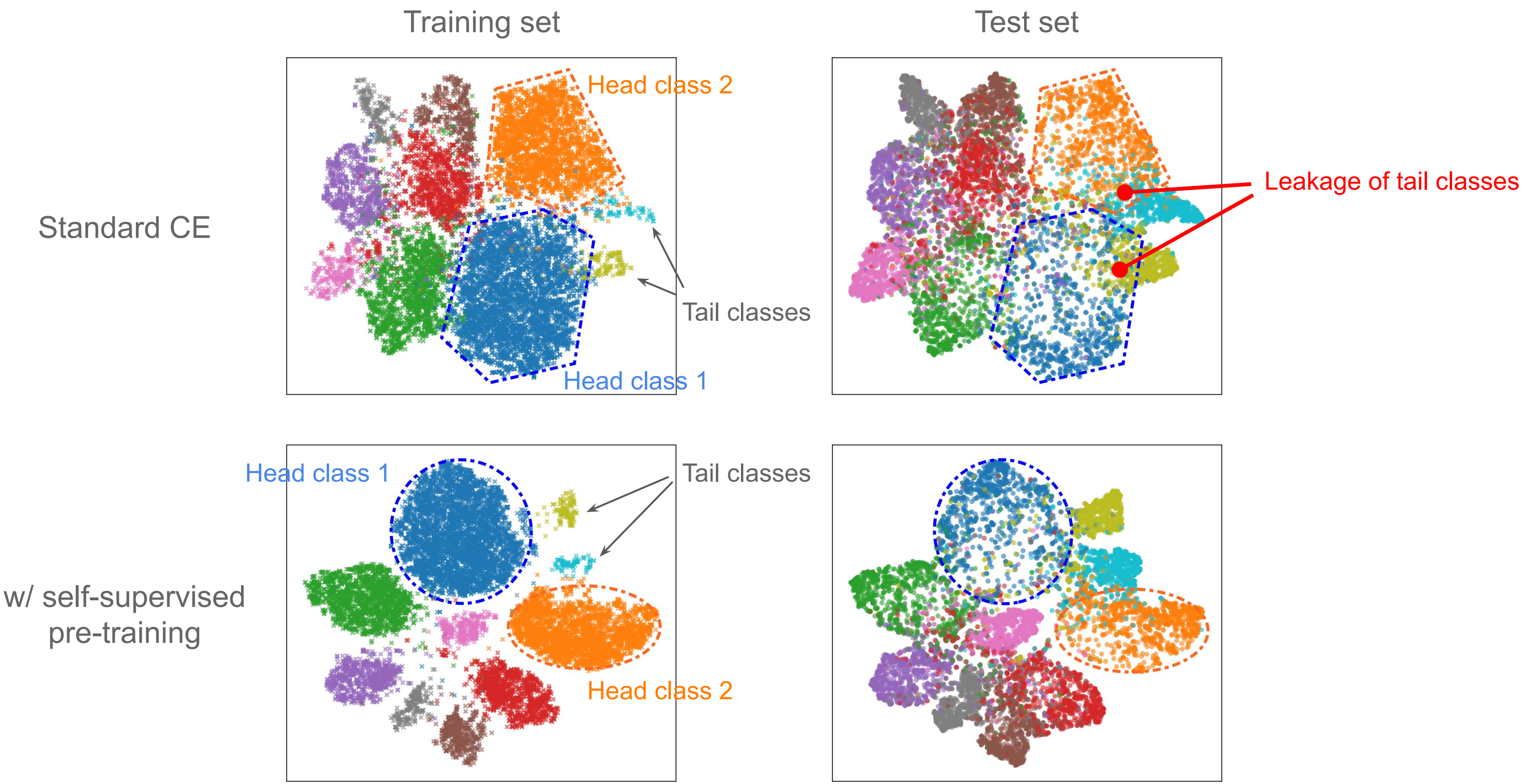
w/ self-supervised pre-training



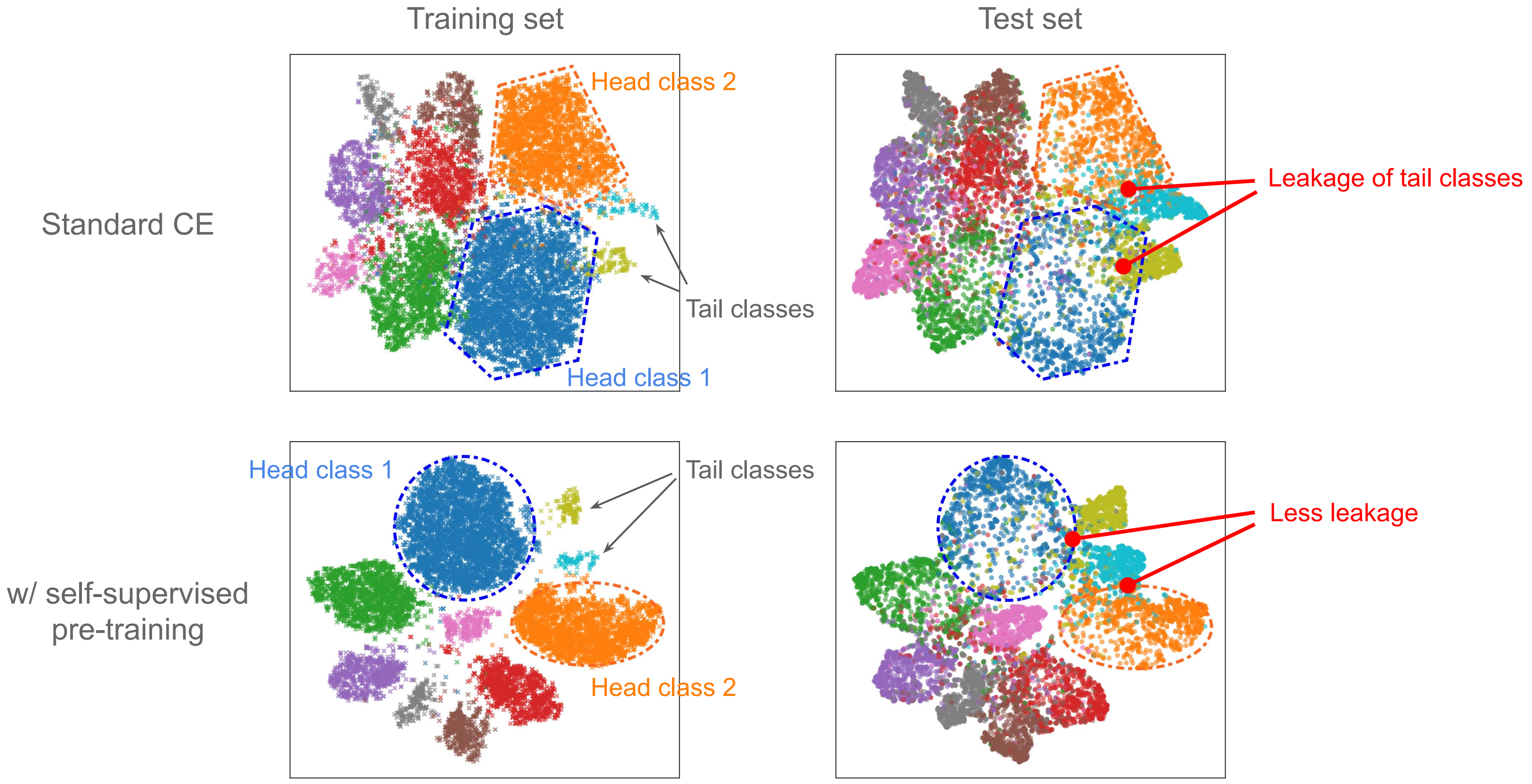
Self-supervised imbalanced learning



Self-supervised imbalanced learning



Self-supervised imbalanced learning



Consistent performance gains

- Semi-supervised imbalanced learning

	CIFAR-10-LT	SVHN-LT
Standard CE	70.36	80.02
w/ unlabeled data	**82.52** (+12.16)	**86.98** (+6.96)

- Self-supervised imbalanced learning

	CIFAR-10-LT	CIFAR-100-LT	ImageNet-LT	iNaturalist 2018
Standard CE	70.36	38.32	38.4	60.7
w/ SSP	**76.53** (+6.17)	**43.06** (+4.74)	**45.6** (+7.2)	**64.4** (+3.7)

Consistent performance gains

- Semi-supervised imbalanced learning

	CIFAR-10-LT	SVHN-LT
Standard CE	70.36	80.02
w/ unlabeled data	**82.52** (+12.16)	**86.08** (+6.06)

Superior improvements across various datasets!
(more results in paper)

	CIFAR-10-LT	CIFAR-100-LT	ImageNet-LT	iNaturalist 2018
Standard CE	70.36	38.32	38.4	60.7
w/ SSP	**76.53** (+6.17)	**43.06** (+4.74)	**45.6** (+7.2)	**64.4** (+3.7)

Summary

- Unlabeled data helps imbalanced learning via a semi-supervised manner
- Self-supervised pre-training can substantially improve imbalanced performance
- Theoretical analysis + large-scale extensive experiments

Check out our code and models at...

- Project website: https://www.mit.edu/~yuzhe/imbalanced-semi-self.html
- Code (relevant data + pretrained models): https://github.com/YyzHarry/imbalanced-semi-self