code available

Unsupervised Selective Labeling for More Effective Semi-supervised Learning

Xudong Wang* Long Lian* Stella X. Yu

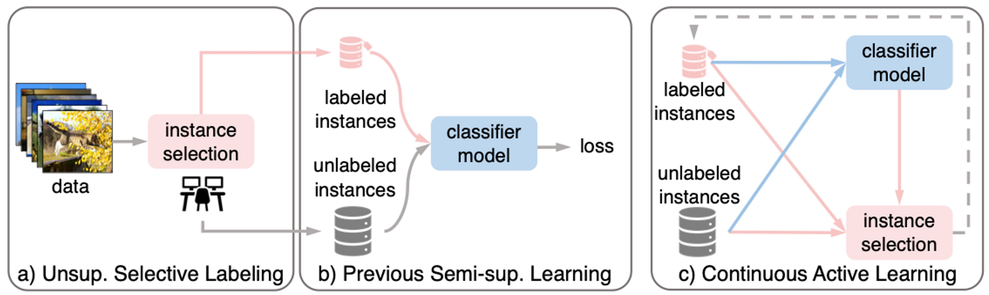
ECCV TEL AVIV 2022

Comparisons with Existing Works on Label-Efficient Learning

USL (Unsup. Selective Labeling)
unsupervisedly selects instances to annotate for SSL.

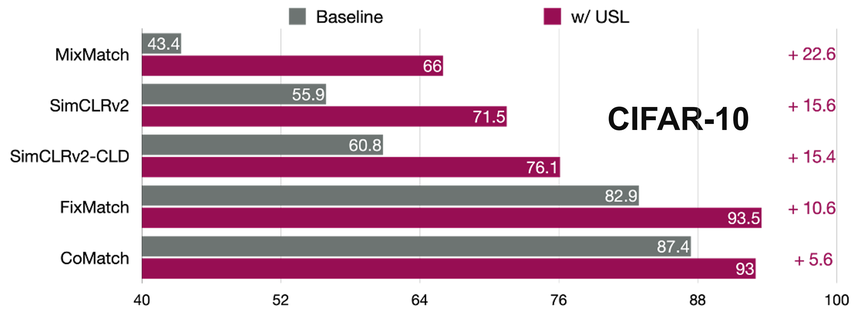
SSL (Previous Semi-sup. Learning)
trains model given a fixed labeled dataset and an unlabeled dataset.

AL (Active Learning)
alternates between model training and instance selection with initial labels.

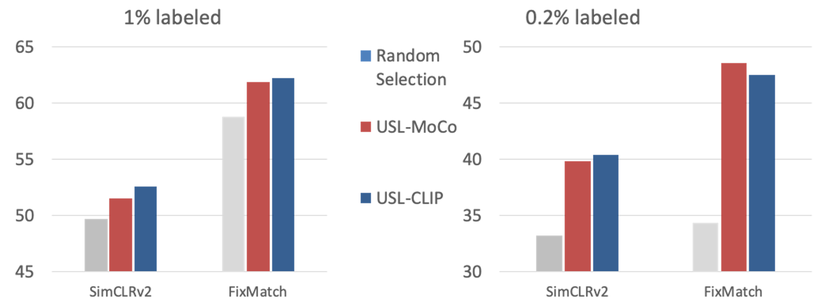


Property	Semi-supervised Learning	Active Learning	Semi-supervised Active Learning	Ours
Uses no initial random labels	✗	✗	✗	✓
Actively queries for labels	✗	✓	✓	✓
Requires annotation only once	✓	✗	✗	✓
Leverages unlabeled data	✓	✗	✓	✓
Allows label reuse across runs	✓	✗	✗	✓

USL as a Universal Add-on to SSL



Scales Well to Large Dataset ImageNet



Training-Free USL: Pick Diverse and Representative Instances in a Given Feature Space

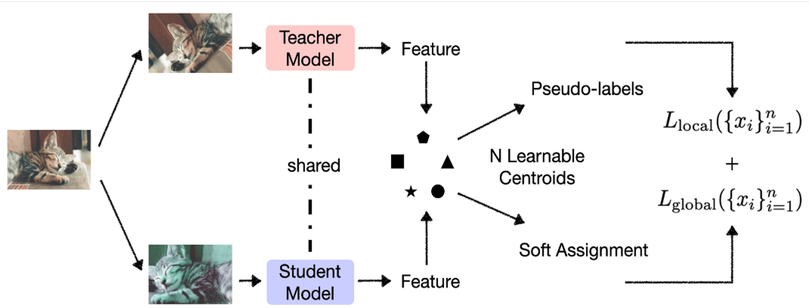
Given an annotation budget of *m* instances, partition data into *m* clusters, select one instance per cluster by instance utility scores:

1. high density value for representativeness
2. large separation across clusters through our regularization algorithm

$$\hat{p}_{\mathrm{KNN}}(V_i,k)=\frac{k}{n}\frac{1}{A_d\cdot\bar{D}^d(V_i,k)}\ ,\quad \min_{\$}\sum_{i=1}^{m}|S_i|\mathrm{Var}(S_i)$$

Try USL out on your dataset without training!

Training-Based USL-T: Joint Feature Learning and Instance Selection



Local constraint:
avoids cluster boundaries passing high density regions

Global constraint:
optimizes a deep variant of k-Means by placing each sample close to its own centroid

Select samples with the maximum softmax confidence scores in each cluster

Our Low- and High-Scored Instances

Class-imbalanced, Ambiguous, Truncated

Class-balanced, Representative, Complete