# Co-learning: Learning from Noisy Labels with Self-supervision

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

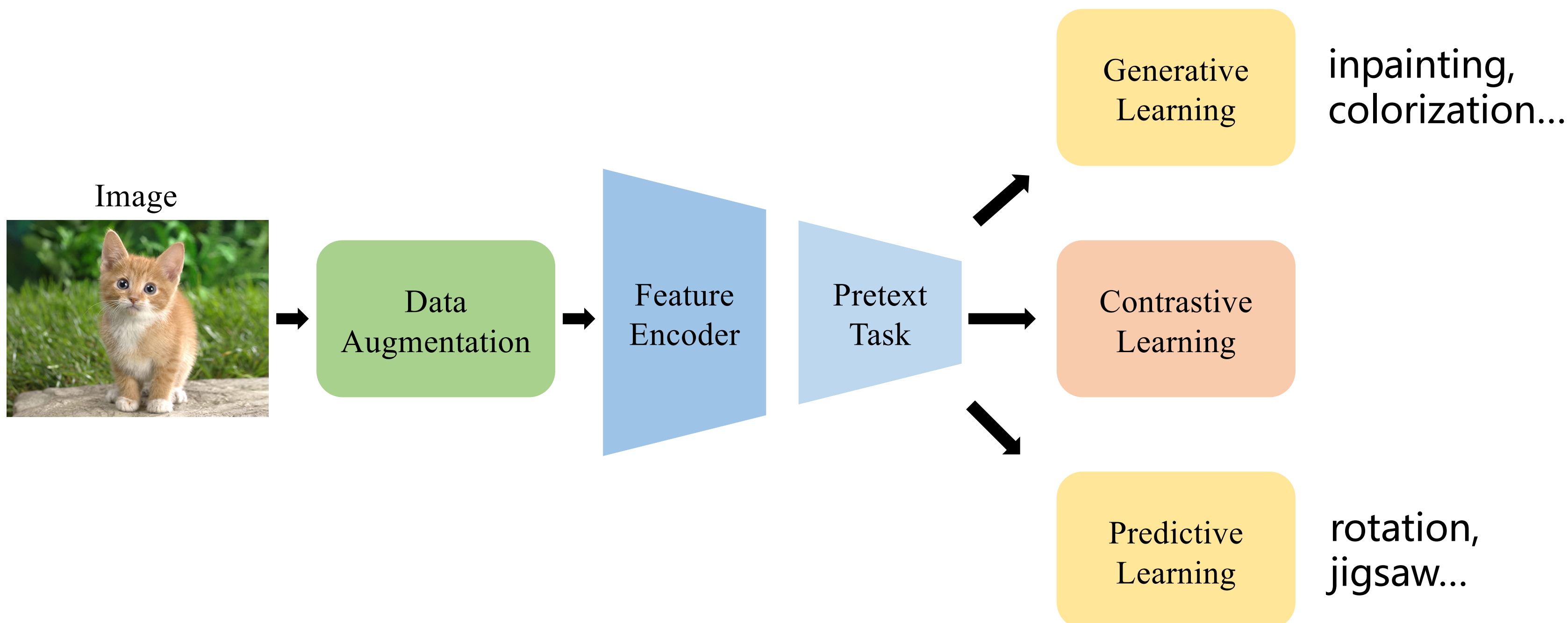
tancheng@westlake.edu.cn
ACM multimedia, 2021

# Noisy label learning – Examples

Manual labeling may produce some wrong labels

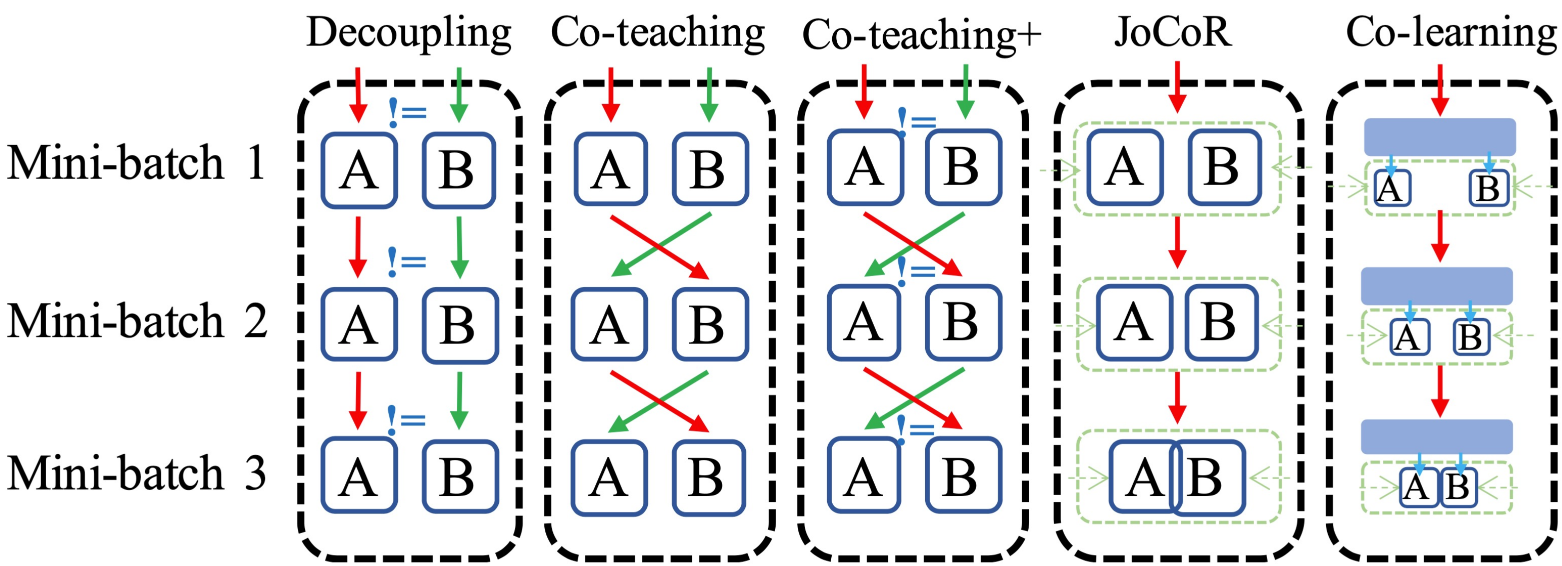
Data collected by web crawlers often contains incorrect labels as well

# Self-supervised Learning



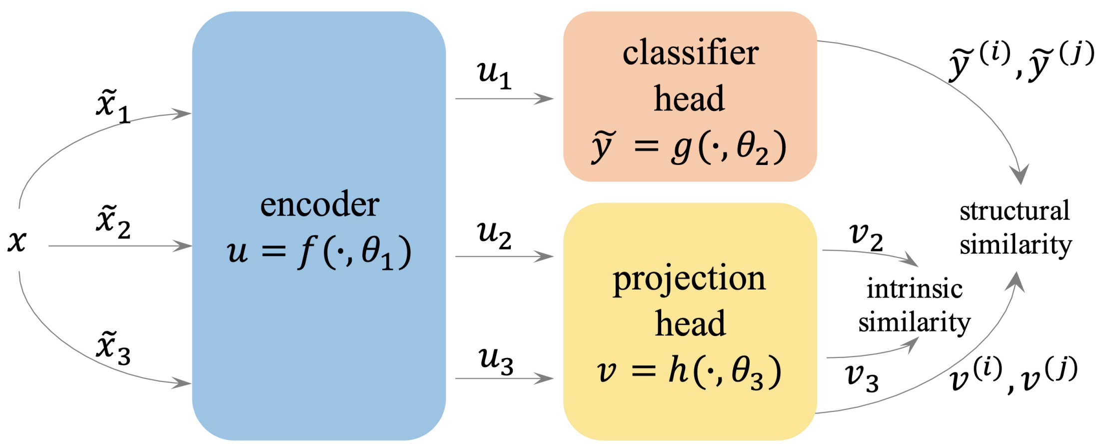

A typical self-supervised learning pipeline

# Noisy label learning – Co-training-based schemes

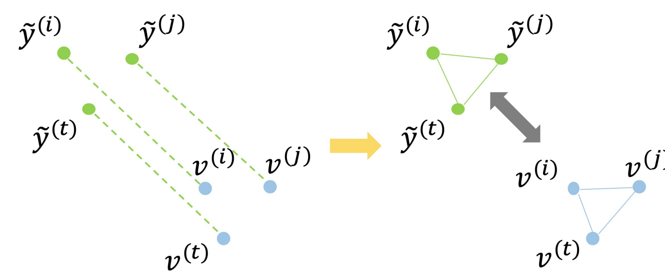



Comparisons of leading methods in dealing with noisy labels

# Co-learning

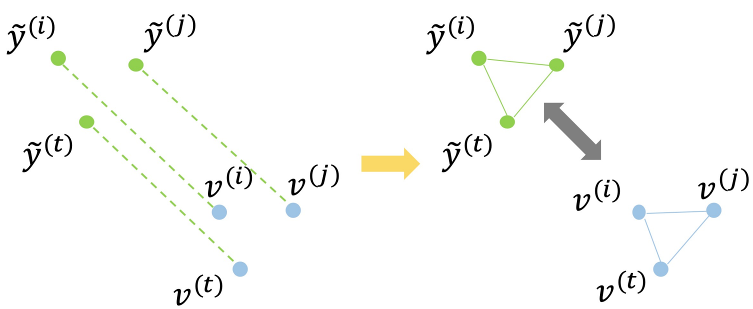



Co-learning framework




Structural similarity

# Structural similarity




Structural similarity

Convert the Euclidean distances $d\left(v^{(i)}, v^{(j)}\right), d(\tilde{y}^{(i)}, \tilde{y}^{(j)})$ into the similarity metrics $p\left(d\left(v^{(i)}, v^{(j)}\right)\right), p\left(d\left(\tilde{y}^{(i)}, \tilde{y}^{(j)}\right)\right)$

We prefer the metrics satisfying $\lim_{d\to+\infty} p(d) = 0$ and $\lim_{d\to 0} p(d) = 1$

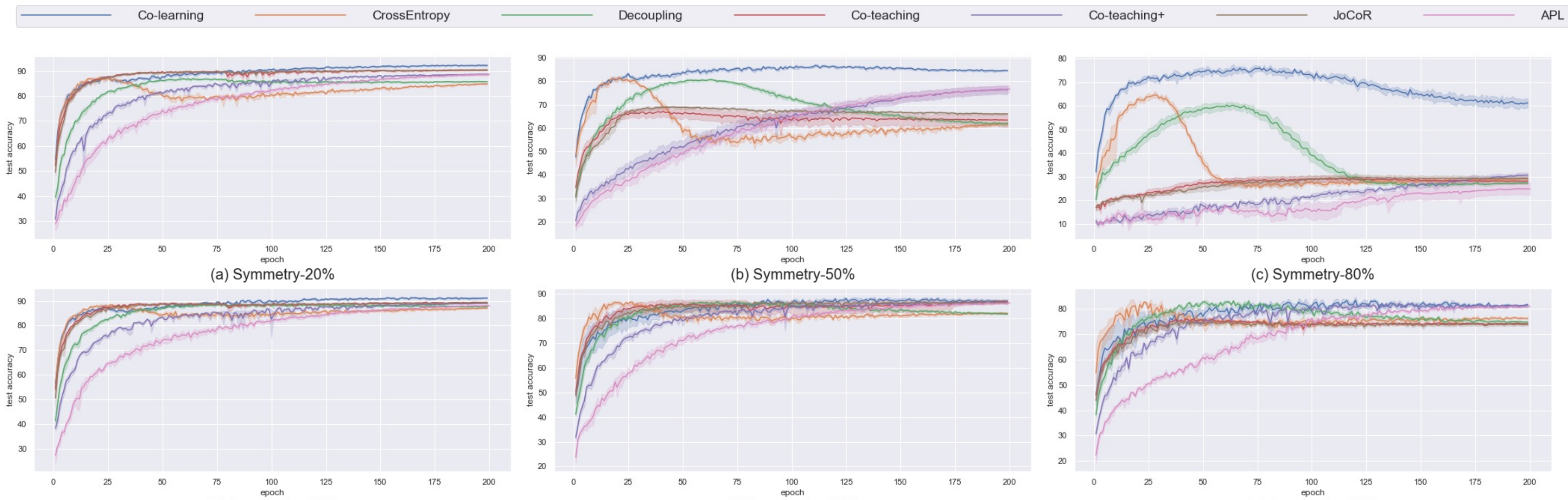
The similarity metric is formulated as:

$$p(d) = C_\sigma \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{d-\mu}{\sigma}\right)^2}$$

Structural similarity loss is defined as the KL-divergence between the similarity metrics of $\tilde{y}$ and $v$:

$$\mathcal{L}_{str} = \sum_{i\neq j} p\left(d\left(v^{(i)}, v^{(j)}\right)\right) \log \frac{p\left(d\left(v^{(i)}, v^{(j)}\right)\right)}{p\left(d\left(\tilde{y}^{(i)}, \tilde{y}^{(j)}\right)\right)}$$

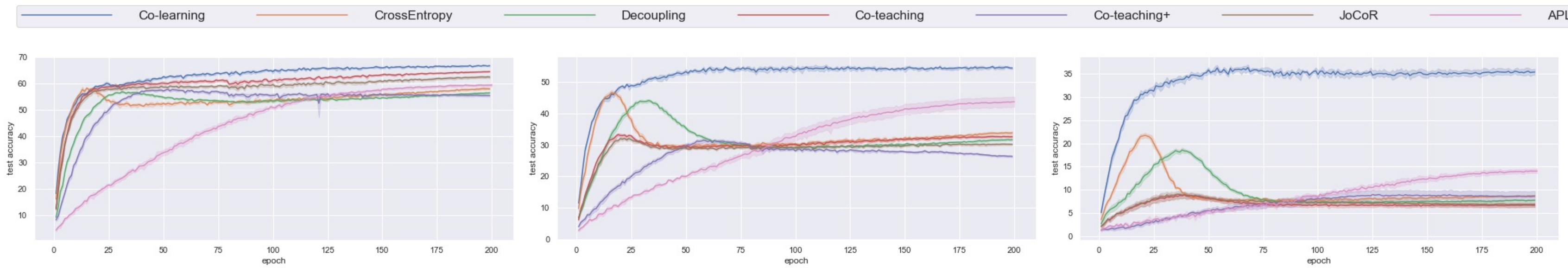
# Experiments – CIFAR-10




(a) Symmetry-20% (b) Symmetry-50% (c) Symmetry-80%
(d) Asymmetry-20% (f) Asymmetry-30% (g) Asymmetry-40%

| Flipping-Rate | Standard | Decoupling | Co-teaching | Co-teaching+ | JoCoR | APL | Co-learning |
|---|---|---|---|---|---|---|---|
| Symmetric-20% | 84.81±0.24 | 85.75±0.31 | 90.29±0.19 | 88.63±0.32 | 90.43±0.25 | 88.54±0.45 | **92.21**±0.31 |
| Symmetric-50% | 61.49±0.58 | 61.93±0.82 | 63.45±3.89 | 76.27±2.80 | 66.00±0.53 | 76.51±1.73 | **84.49**±0.34 |
| Symmetric-80% | 28.98±0.26 | 27.23±0.84 | 28.03±1.67 | 30.37±1.69 | 29.19±1.64 | 24.75±2.87 | **61.20**±2.29 |
| Asymmetric-20% | 87.00±0.20 | 87.66±0.29 | 89.38±0.33 | 89.00±0.18 | 89.20±0.26 | 88.02±0.29 | **91.07**±0.32 |
| Asymmetric-30% | 81.99±0.31 | 81.83±0.26 | 86.58±1.32 | 86.22±0.26 | 86.41±0.45 | 86.03±0.21 | **86.89**±0.87 |
| Asymmetric-40% | 76.30±0.34 | 74.97±0.38 | 74.25±0.38 | 81.25±0.75 | 73.95±1.00 | 80.97±0.19 | **81.42**±0.52 |

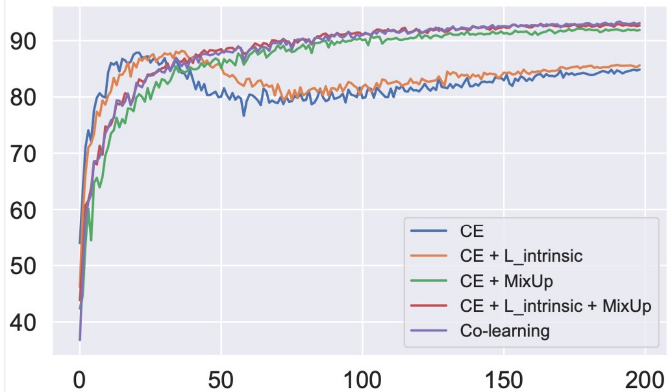
# Experiments – CIFAR-100



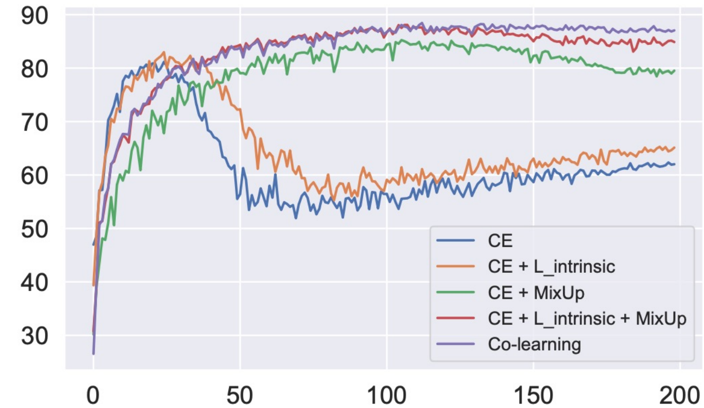
(a) Symmetry-20%
(b) Symmetry-50%
(c) Symmetry-80%
(d) Asymmetry-20%
(f) Asymmetry-30%
(g) Asymmetry-40%

| Flipping-Rate | Standard | Decoupling | Co-teaching | Co-teaching+ | JoCoR | APL | Co-learning |
|---|---|---|---|---|---|---|---|
| Symmetric-20% | 57.79±0.44 | 56.18±0.32 | 64.28±0.32 | 55.40±0.71 | 62.29±0.71 | 59.21±0.50 | **66.58**±0.15 |
| Symmetric-50% | 33.75±0.46 | 31.58±0.54 | 32.62±0.51 | 26.49±0.45 | 30.19±0.60 | 43.53±1.84 | **54.54**±0.43 |
| Symmetric-80% | 8.64±0.22 | 7.71±0.23 | 6.65±0.71 | 8.57±1.55 | 6.84±0.92 | 13.97±0.53 | **35.45**±0.79 |
| Asymmetric-20% | 59.36±0.36 | 57.97±0.24 | 59.76±0.53 | 56.11±0.60 | 58.58±0.51 | 58.89±0.40 | **65.26**±0.76 |
| Asymmetric-30% | 51.06±0.44 | 49.86±0.54 | 49.53±0.79 | 47.12±0.73 | 49.04±0.91 | 51.46±0.15 | **56.97**±1.22 |
| Asymmetric-40% | 42.49±0.23 | 41.51±0.67 | 40.62±0.79 | 38.98±0.54 | 39.72±0.76 | 41.96±0.92 | **47.62**±0.79 |

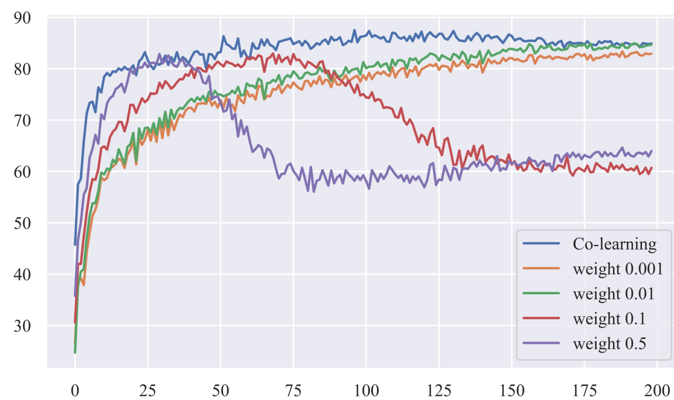
# Experiments – Ablation study on CIFAR-10




(a) CIFAR-10 with Symmetry-20%




(b) CIFAR-10 with Symmetry-50%




(c) weighted supervised loss on CIFAR-10 with Symmetry-20%

# Summary

- Point out the problems of the common co-training paradigm in noisy learning.
- Propose a new noisy learning method known as: Co-learning, which assisted supervised learning through self-supervised learning.
- Reproduce similar methods under a unified framework for fair comparison and obtain the best performance on multiple benchmark data sets. (github.com/chengtan9907/Co-training-based_noisy-label-learning)

Comparison with state-of-the-art methods in techniques used

| | Co-teaching | Co-teaching+ | JoCoR | Co-learning |
|---|---|---|---|---|
| Agreement | ✓ | ✓ | ✓ | ✓ |
| Small-loss | ✓ | ✓ | ✓ | ✗ |
| Double classifiers | ✓ | ✓ | ✓ | ✗ |
| Cross update | ✓ | ✓ | ✗ | ✗ |