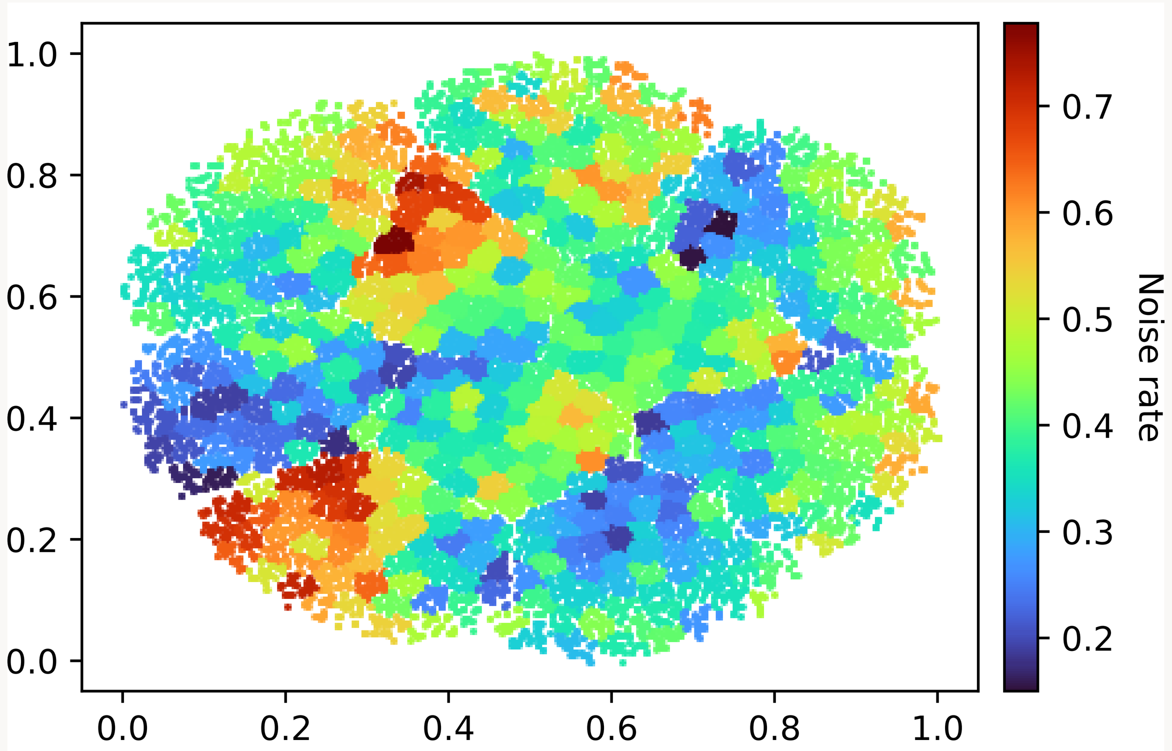
Learning with Non-Uniform Label Noise: A Cluster-Dependent Weakly Supervised Approach

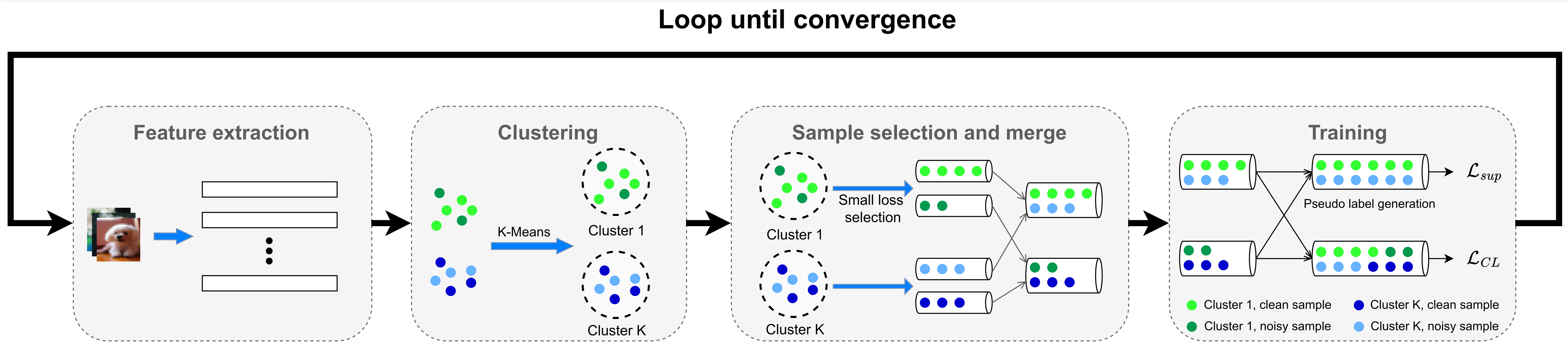
Mengtian Zhang, Bo Jiang, Yuye Ling, Xinbing Wang
Shanghai Jiao Tong University

Motivation & Introduction



- The successes of current DNNs heavily rely on the precisely labeled data. However,
  - The average ratio of corrupted labels in real-world datasets range **from 8.0% to 38.5%** [Song et al., 2022].
  - The local noise rates inside the real dataset can vary greatly, e.g. **from 25% to 70%**.
- Existing works explicitly or implicitly rely on the assumption of uniform label noise.
  - The methods of **class-dependent transition matrix** assume all samples in same class share same noise rate.
  - With the implicit assumption of uniform noise rate, the methods of **small-loss trick** would select simple patterns first and regard most samples in the hard regions as corrupt data.
  - Besides, the methods of **instance-dependent noise** need additional information or extra assumption.
- In this paper, we consider robust learning with non-uniform label noise that requires no additional information or assumption.

Methodology: ClusterCL



- **Feature extraction.** Embedded sample features would be extracted using the current trained model.
- **Clustering.** K-Means (or any other clustering method) is employed to group samples in the feature space.
- **Cluster-dependent sample selection.**
  - 2-dimension Gaussian Mixture Models are employed on the loss distributions.
  - Small-loss criteria are adopted to select clean samples for each cluster separately.
- Note: the above four steps would be looped for epochs.

- **Training** for one epoch with hybrid losses.
  - Supervised cross-entropy loss:

$$\mathcal{L}_{sup} = \mathcal{L}_{mixup}(x', y') + \mathcal{L}_{fmix}(x'', y'') = \mathcal{L}_{CE}(x', y') + \mathcal{L}_{CE}(x'', y'') \tag{1}$$

  - unsupervised contrastive loss:

$$\mathcal{L}_{CL} = \sum_{i=1}^{N} l_{CL}(i) = -\sum_{i=1}^{N} \log \frac{\exp(z_i \cdot z_i^{+}/t)}{\sum_{j=1}^{K}(z_i \cdot z_j/t)} \tag{2}$$

  - Overall training loss:

$$\mathcal{L}_{total} = \mathcal{L}_{sup} + \lambda \mathcal{L}_{CL} \tag{3}$$

Results of Classification Accuracy

Human annotated label noise on CIFAR-N datasets [Wei et al., 2022].

- CIFAR-N datasets provide realistic noisy labels, with noisy rate aggregate 9.03%, random-1 17.23%, Worst 40.21%, and fine 40.20%.

Methods	CIFAR-10N			CIFAR-100N
	Aggr	Rand1	Worst	Fine
CE (Standard)	89.87	84.15	76.86	55.96
T-Revision	89.39	87.99	82.10	54.45
PTD	89.93	89.83	80.16	16.01
ELR+	94.81	94.54	90.89	67.04
DivideMix	95.15	95.12	92.71	71.13
ProMix	96.83	96.17	94.05	70.54
SOP	95.61	95.28	93.24	67.81
ClusterCL(ours)	**96.86**	**96.29**	**94.13**	**71.87**

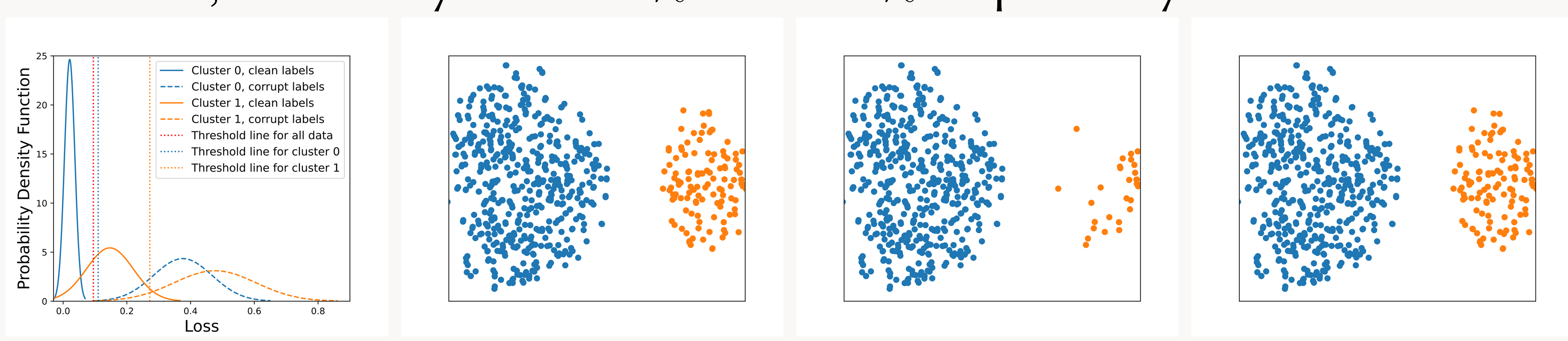
- 500 ~ 50000 training images, sampled from CIFAR-10 Worst dataset.

Methods	CIFAR-10N ($\tau \approx 40\%$)						
	$N = 500$	2000	5000	10000	20000	40000	50000
CE (Standard)	32.54	41.05	49.58	58.61	63.84	74.33	76.86
T-Revision	28.54	29.64	32.69	63.47	77.37	80.66	82.10
PTD	18.99	26.59	39.01	65.85	66.69	70.97	80.16
ELR+	38.39	56.29	67.24	75.26	84.30	89.77	90.89
DivideMix	36.52	58.43	70.03	77.77	87.38	91.83	92.71
ProMix	34.96	58.15	69.75	77.95	88.06	92.71	94.05
SOP	37.21	54.68	67.43	75.15	85.52	91.88	93.24
ClusterCL(ours)	**44.26**	**64.31**	**76.19**	**84.67**	**89.85**	**93.08**	**94.13**

- ClusterCL achieves highest accuracy, especially when the training set is small and the noise rate is high.

Results of Sample Selection

- **Visualization comparison.** Color **blue** and **orange** are two selected clusters, with noisy rate 21.1% and 60.8% respectively.



(a) Loss Distribution (b) Ground Truth (c) Baseline Method (d) ClusterCL(ours)

- Precision (%), Recall (%), and F1-score (%) in the clean sample selection step on CIFAR-10 Worst dataset with 5000 training samples.

Methods	Precision	Recall	F1-score
CE (Standard)	59.02	100.0	74.22
DivideMix	88.34	82.48	85.31
ProMix	**89.19**	82.03	85.46
ClusterCL(ours)	89.04	**88.45**	**88.74**

- ClusterCL achieve slightly lower precision but much higher recall and thus higher F1-score.

Conclusion

- The paper propose a novel weakly supervised method for robust learning with non-uniform label noise.
- ClusterCL achieves the state-of-the-art performance on both synthetic and real-world datasets.

PAPER