Paper ID: 2492

Detecting Corrupted Labels Without Training a Model to Predict

Zhaowei Zhu, Zihao Dong, Yang Liu

UC Santa Cruz

{**zwzhu**, yangliu}@ucsc.edu

Code

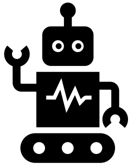
A Data-Centric Method

Noisy Labels Are Everywhere

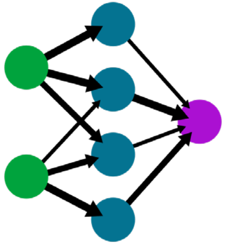
- Noisy labels may come from:

Human [1]

Sensor [2]



Model [3]



- Challenges [4]:
  - Cause disparate impact
  - Require disparate treatment

We need to remove wrong annotations if possible!

[1] J. Wei, Z. Zhu, H. Cheng, T. Liu, G. Niu, Y. Liu. Learning with noisy labels revisited: A study using real-world human annotations. *ICLR 2022*.
[2] J. Wang, H. Guo, Z. Zhu, Y. Liu. Policy Learning Using Weak Supervision. *NeurIPS 2021*.
[3] Z. Zhu, T. Luo, Y. Liu. The Rich Get Richer: Disparate Impact of Semi-Supervised Learning. *ICLR 2022*.
[4] Y. Liu. Understanding Instance-Level Label Noise: Disparate Impacts and Treatments. *ICML 2021*.

Detect Corrupted Labels

Noisy: correct/wrong
Corrupted: wrong annotations

Our Contributions

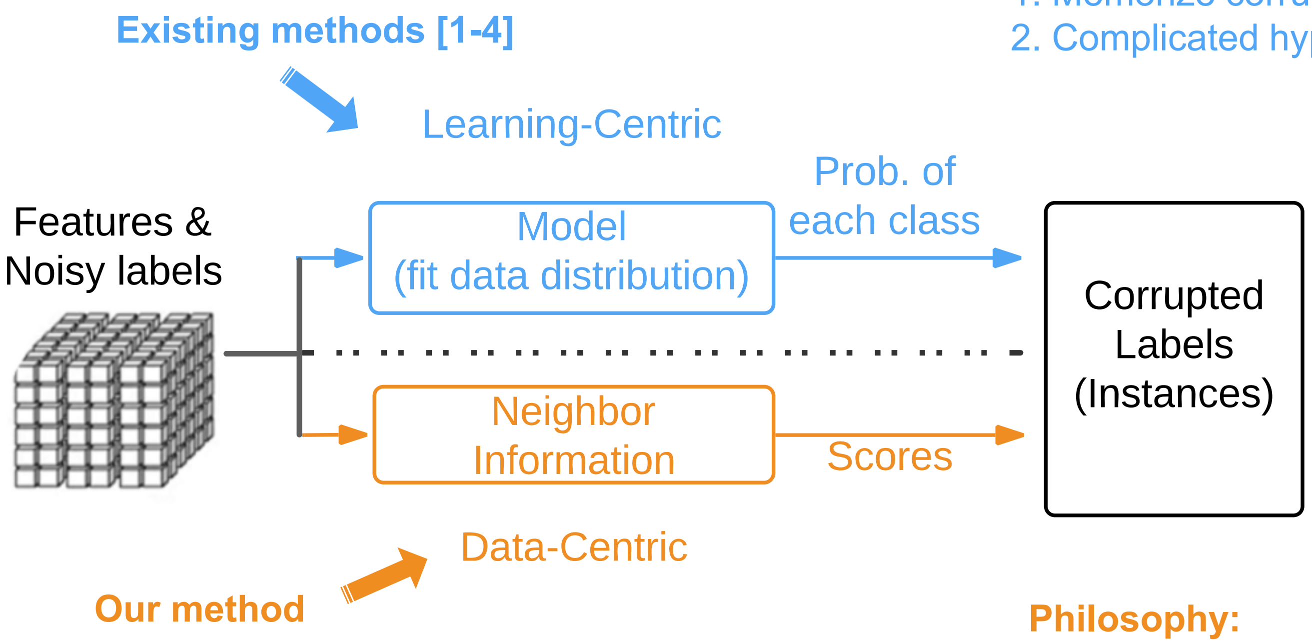
We detect corrupted labels
without training a model!

Highlights:

- ✓ New perspective: *data-centric*
- ✓ Efficient algorithms: Voting/Ranking
- ✓ Theoretical & Numerical Analyses

Existing Methods vs Our Methods

Problems:
1. Memorize corrupted labels
2. Complicated hyperparameter tuning



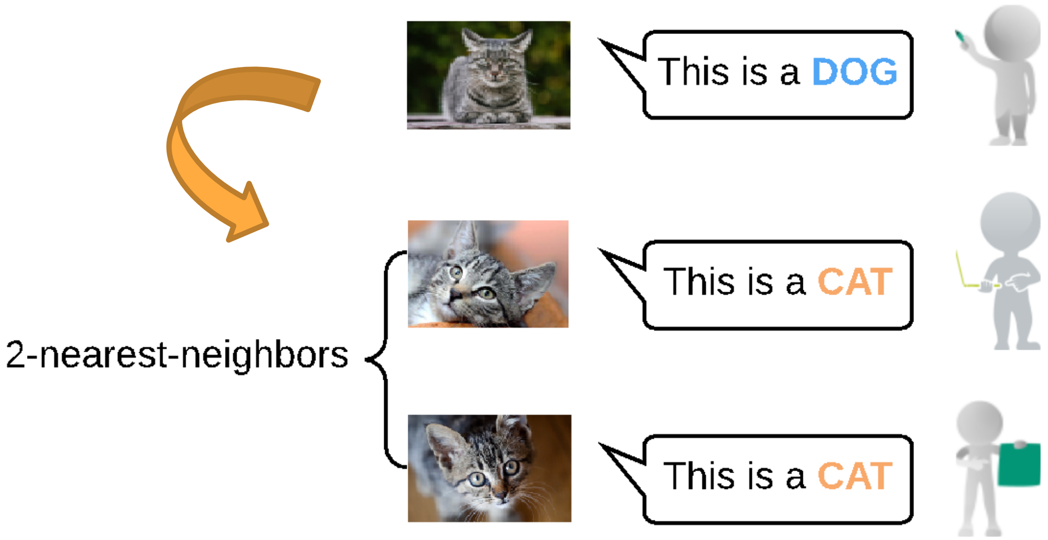
Philosophy:
No learning → no memorization

[1] H. Cheng, Z. Zhu, X. Li, Y. Gong, X. Sun, Y. Liu. Learning with instance-dependent label noise: A sample sieve approach. *ICLR 2021.*
[2] C. Northcutt, L. Jiang, I. Chuang. Confident learning: Estimating uncertainty in dataset labels. *JAIR 2021.*
[3] D. Bahri, H. Jiang, M. Gupta. Deep k-nn for noisy labels. *ICML 2020.*
[4] G. Pruthi, F. Liu, S. Kale, M. Sundararajan. Estimating training data influence by tracing gradient descent. *NeurIPS 2020.*

Motivation: Check with your neighbors

- Key Assumption: k-Nearest-Neighbor (k-NN) label clusterability [1]

Same true label



DOG: 1 votes

CAT: 2 votes

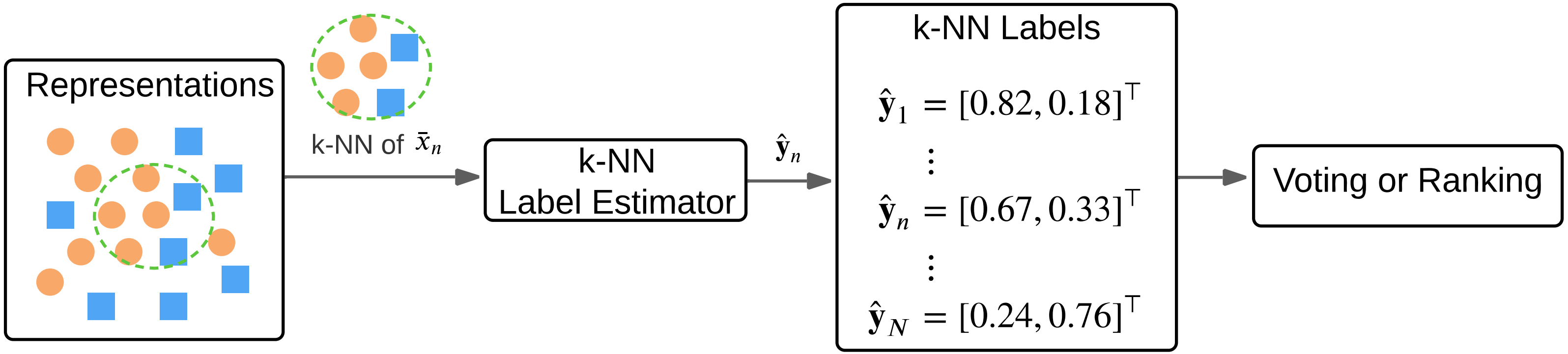
CAT is the **majority**'s choice!

True label should be CAT

Can we do better?

[1] Z. Zhu, Y. Song, Y. Liu. Clusterability as an alternative to anchor points when learning with noisy labels. *ICML 2021.*

Our Method



Representations:
- Raw features
- Other pre-processed features.

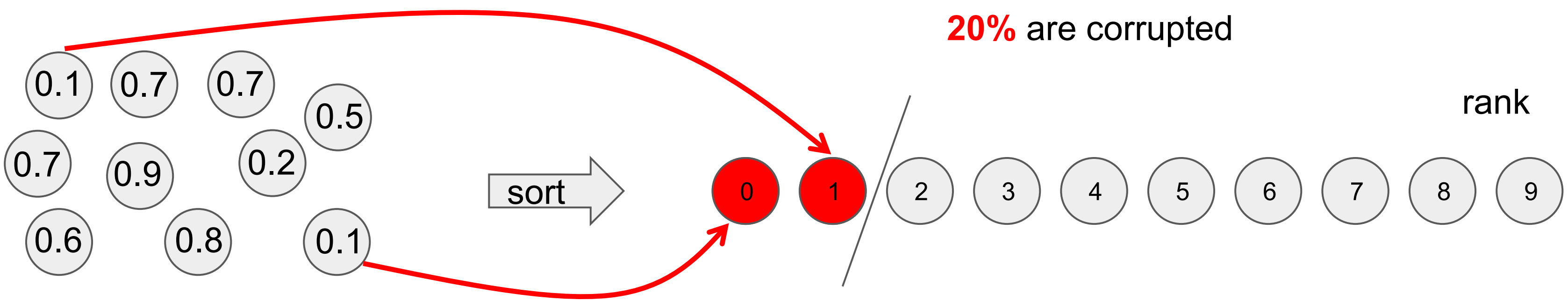
Voting: Majority vote (illustrated before)

Ranking (better in most cases):
- Score function
- Threshold (guaranteed)

UC SANTA CRUZ

Ranking-Based Global Detection

20% are corrupted



Scores of **10** noisy instances

Threshold: **10** * **20%** = **2**

Lower score → Corrupted

Score function:

CosSim(kNN_label, noisy_label)

Threshold: (given by HOC)

Prob. of being corrupted

HOC: Z. Zhu, Y. Song, Y. Liu. Clusterability as an alternative to anchor points when learning with noisy labels. *ICML 2021.*

Experiments

➢ CIFAR. (Clean test accuracy ⬆)

Method	CIFAR10				CIFAR100			
	Human	*Symm. 0.6*	*Asym. 0.3*	*Inst. 0.4*	*Human*	*Symm. 0.6*	*Asym. 0.3*	*Inst. 0.4*
CORES	65.00	92.94	7.68	**87.43**	3.52	**92.34**	0.02	9.67
CL	55.85	80.59	76.45	62.89	64.58	78.98	52.96	50.08
TracIn	55.02	76.94	73.47	58.85	61.75	76.74	48.42	49.89
Deep k-NN	56.21	82.35	75.24	63.08	57.40	70.69	56.75	63.85
SimiFeat-V	**82.30**	**93.21**	**82.52**	81.09	**73.19**	84.48	**65.42**	**74.26**
SimiFeat-R	**83.28**	**95.56**	**83.58**	**82.26**	**74.67**	**88.68**	**62.89**	**73.53**

Ours

-V: Voting
-R: Ranking

CIFAR human annotations:
J. Wei, Z. Zhu, H. Cheng, T. Liu, G. Niu, Y. Liu.
Learning with noisy labels revisited: A study using real-world human annotations. *ICLR 2022*

Experiments

- Clothing1M (Clean test accuracy ⬆)

Data Selection	# Training	Best Epoch	Last 10	Last
(Baseline) None	1M (100%)	70.32	69.44 ± 0.13	69.53
R50-Img	770k (77.0%)	72.37	71.95 ± 0.08	71.89
ViT-B/32-CLIP	700k (70.0%)	72.54	72.23 ± 0.17	72.11
R50-Img Warmup-1	767k (76.7%)	**73.64**	$\mathbf{73.28 \pm 0.18}$	**73.41**

Ours with different representations

Existing methods (Best Epoch):

HOC 73.39%, GCE+SimCLR 73.35%, CORES 73.24%, GCE 69.75%.

UC SANTA CRUZ

Thank you !

Want to know more?

1. Scan the QR code to reproduce
2. Join our **onsite** poster session

Code