Learning with instance-dependent label noise: A sample sieve approach

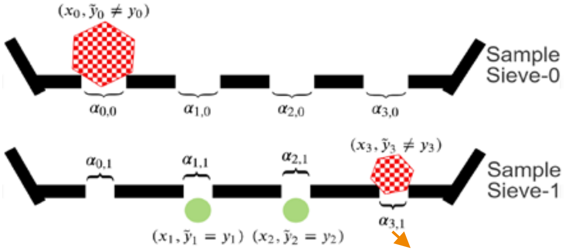
Hao Cheng*, **Zhaowei Zhu***, Xingyu Li, Yifei Gong, Xing Sun and Yang Liu#

zwzhu, yangliu@ucsc.edu

Code

REsponsible & Accountable Learning (REAL)
@ University of California, Santa Cruz

https://github.com/UCSC-REAL



UNIVERSITY OF CALIFORNIA SANTA CRUZ

Tencent YouTu Lab

Background

- Problems: $Y \to \tilde{Y}$

1. **Wrong** correlation patterns
2. **Expensive** human-efforts to fix errors

Background

- Challenges:

1. **Unknown** noise rates $\mathbb{P}(\tilde{Y}|X,Y)$
2. **Instance-dependent** label noise $\mathbb{P}(\tilde{Y}|X,Y) \neq \mathbb{P}(\tilde{Y}|Y)$

   - while noise existing works [1-3] **assume feature independency** $\mathbb{P}(\tilde{Y}|X,Y) = \mathbb{P}(\tilde{Y}|Y)$
3. Loss-correction/reweighting [1-3]: **hard to estimate** $\mathbb{P}(\tilde{Y}|X,Y), \forall X$

- Solutions:

1. Confidence regularizer (learn ***clean*** distributions) - CR
2. Dynamic sample sieve (separate ***clean/corrupted*** examples) - CORES^2
3. **Regular** training (sieved ***clean*** examples) +

   Consistency training (features of sieved ***corrupted*** examples) - CORES^{2*}

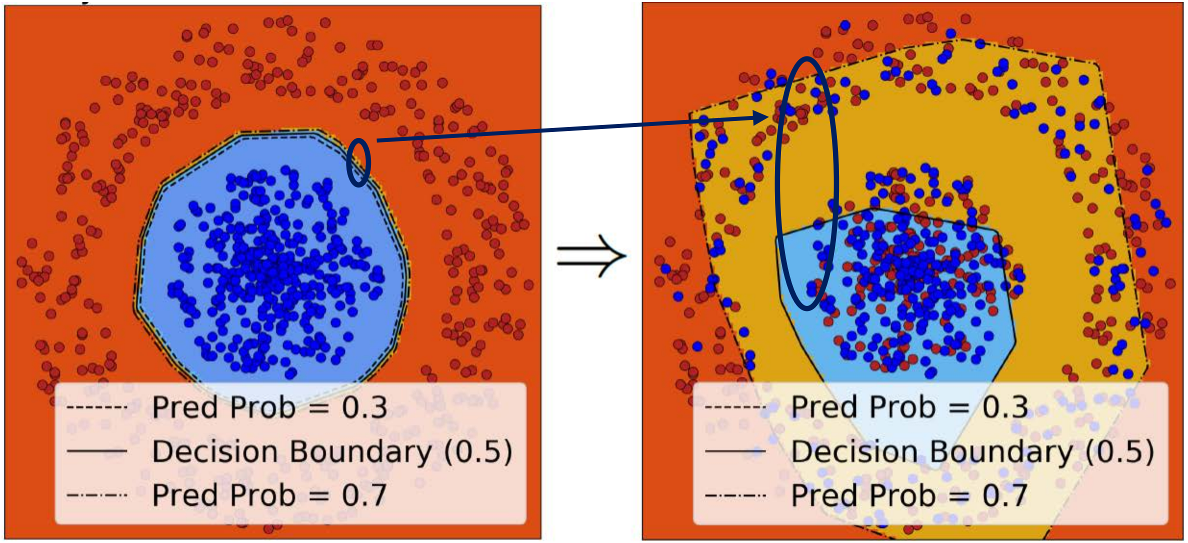
[1] N. Natarajan, et al. “Learning with noisy labels." NeurIPS'13.
[2] T. Liu & D. Tao. “Classication with noisy labels by importance reweighting." TPAMI'15.
[3] G. Patrini, et al. “Making deep neural networks robust to label noise: A loss correction approach." CVPR'17.

Confidence Regularizer (CR)

- Motivation

Clean

40% Noise



Observation:
Label noise **reduces** the **confidence** of predictions

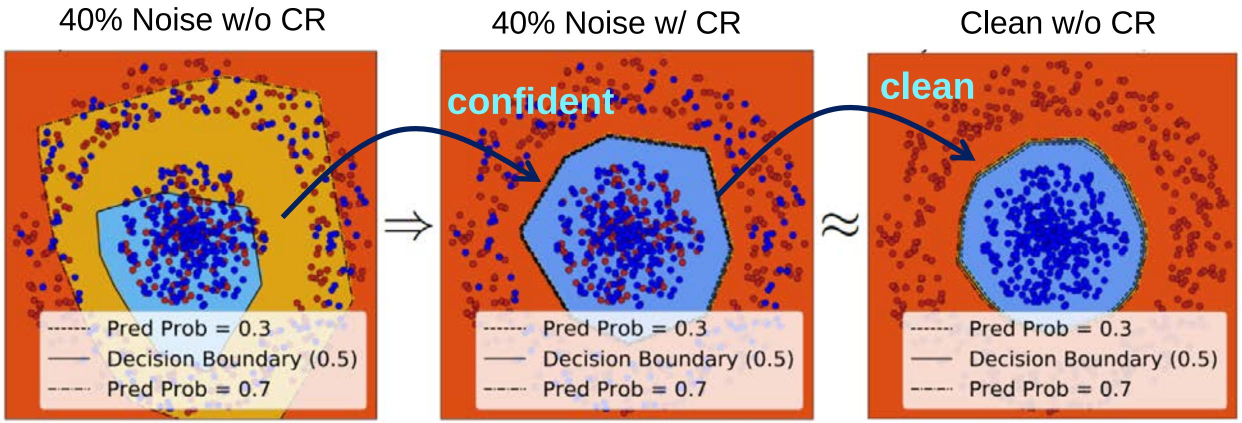
Our idea:
Encourage **confident** prediction to remove corrupted examples

Confidence Regularizer (CR)

- Solution:

$$\textbf{Confidence Regularizer:}\quad \ell_{\text{CR}}(f(x_n)) := -\beta \cdot \mathbb{E}_{\mathcal{D}_{\widetilde{Y}|\widetilde{D}}}[\ell(f(x_n), \widetilde{Y})]$$

- 2-D visualization:



CR helps:

1. Make **confident** predictions
2. Learn **clean** distributions

Dynamic Sample Sieve

- **CO**nfidence **RE**gularized **S**ample **S**ieve (**CORES**[2])

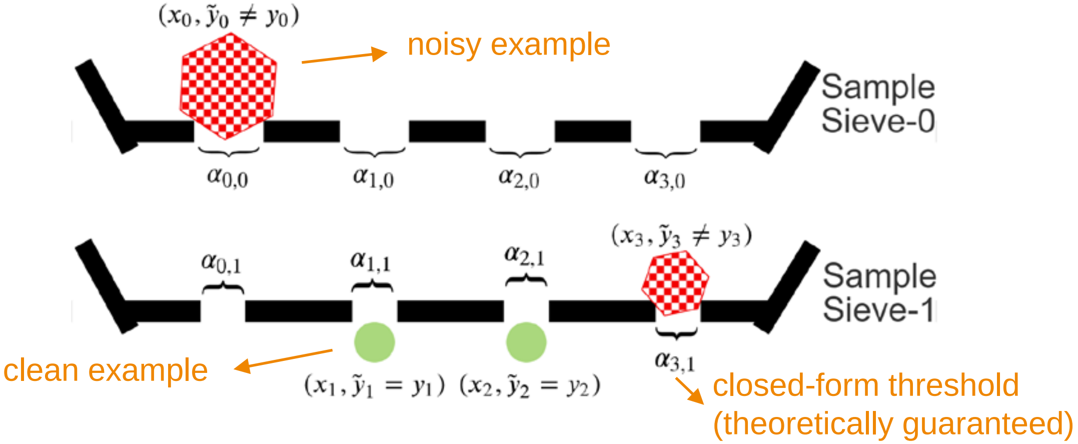
Intuition:
Sieve out **large-loss** examples

Confidence Regularized Sample Sieve

$$\min_{\substack{f\in\mathcal{F},\\ v\in\{0,1\}^N}} \sum_{n\in[N]} v_n\left[\ell(f(x_n),\tilde{y}_n)+\ell_{\text{CR}}(f(x_n))-\alpha_n\right]$$

$$\text{s.t.}\quad \ell_{\text{CR}}(f(x_n)) := -\beta\cdot\mathbb{E}_{\mathcal{D}_{\widetilde{Y}|\widetilde{D}}}\,\ell(f(x_n),\widetilde{Y}),$$

$$\alpha_n := \frac{1}{K}\sum_{\tilde{y}\in[K]}\ell(\bar{f}(x_n),\tilde{y})+\ell_{\text{CR}}(\bar{f}(x_n)).$$



Theoretical Results

- **Theorem:** CORES^2 sieves out the corrupted examples:

1. Condition: Classifier predicts better than *random guess* on the example
2. Conditions:
    - clean labels = Bayes optimal
    - noisy labels are *informative*
    - infinite model capacity and sufficiently many examples
    - minimize CR-regularized CE loss

1. clean $\nrightarrow$ corrupted, corrupted $\rightarrow$ clean

2. clean $\nrightarrow$ corrupted, corrupted $\nrightarrow$ clean

- **Why this is true?**

1. Decoupling the expected CR-regularized CE loss:

noisy loss with CR = clean loss + label shift + noise effect (β)

2.CR helps learn the clean distribution:

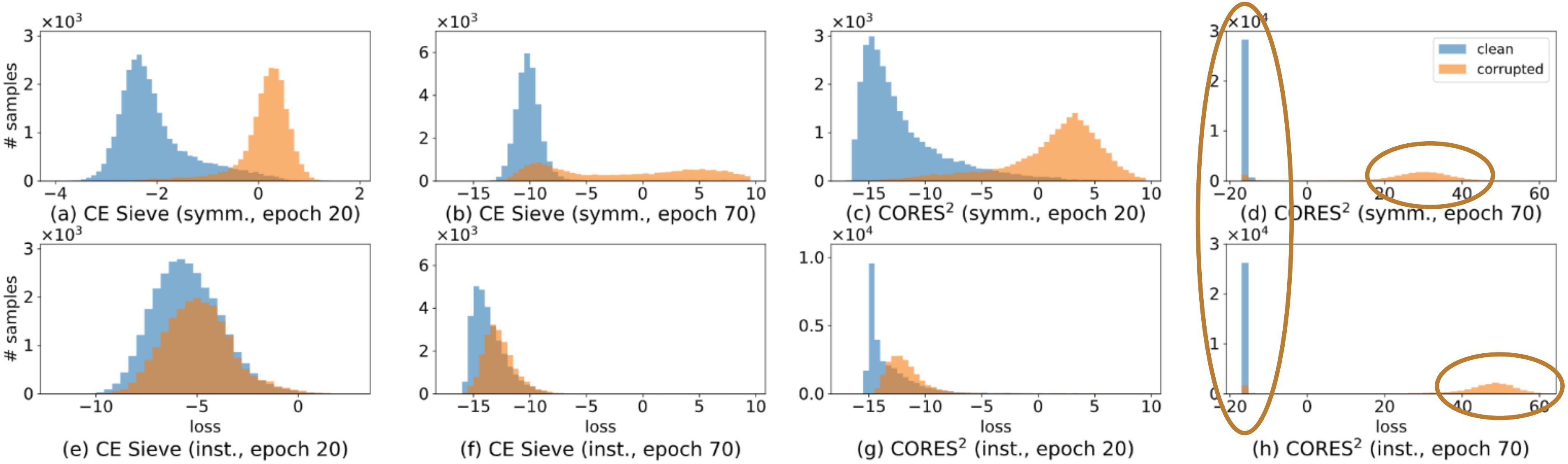
noise effect can be **canceled** or **reversed** by proper β

3. Proper setup of threshold α (guaranteed closed-form)

Experiment

Symm.: feature-independent
Inst.: instance-dependent
Goal: split **clean vs. corrupted**

- Loss distributions of training w/ or w/o CR



After sample sieve, we can **treat "clean" and "corrupted" examples differently**

UC SANTA CRUZ

Experiment

Method	*Inst. CIFAR10*			*Inst. CIFAR100*		
	$\varepsilon = 0.2$	$\varepsilon = 0.4$	$\varepsilon = 0.6$	$\varepsilon = 0.2$	$\varepsilon = 0.4$	$\varepsilon = 0.6$
Cross Entropy	87.16	75.16	44.64	58.72	41.14	25.29
Forward T (Patrini et al., 2017)	88.08	82.67	41.57	58.95	41.68	22.83
L_{DMI} (Xu et al., 2019)	88.80	82.70	70.54	58.66	41.77	28.00
L_q (Zhang & Sabuncu, 2018)	86.45	69.02	32.94	58.18	40.32	23.13
SCE (Wang et al., 2019)	89.11	72.04	44.83	59.87	41.76	23.41
Co-teaching (Han et al., 2018)	88.66	69.50	34.61	43.03	23.13	7.07
Co-teaching+ (Yu et al., 2019)	89.04	69.15	33.33	41.84	24.40	8.74
JoCoR (Wei et al., 2020)	88.71	68.97	30.27	44.28	22.77	7.54
Peer Loss (Liu & Guo, 2020)	89.33	81.09	73.73	59.92	45.76	33.61
CORES^2	**89.50**	**82.84**	**79.66**	**61.25**	**47.81**	**37.85**
(Regular + Consistency) CORES^{2*}	**95.42**	**88.45**	**85.53**	**72.91**	**70.66**	**63.08**

Table: Comparison of test accuracies under *instance-dependent* label noise

Thank you !

Join our poster session!

@ Poster Session 9, May 5, 2021, 5 p.m. - 7 p.m. (PDT)