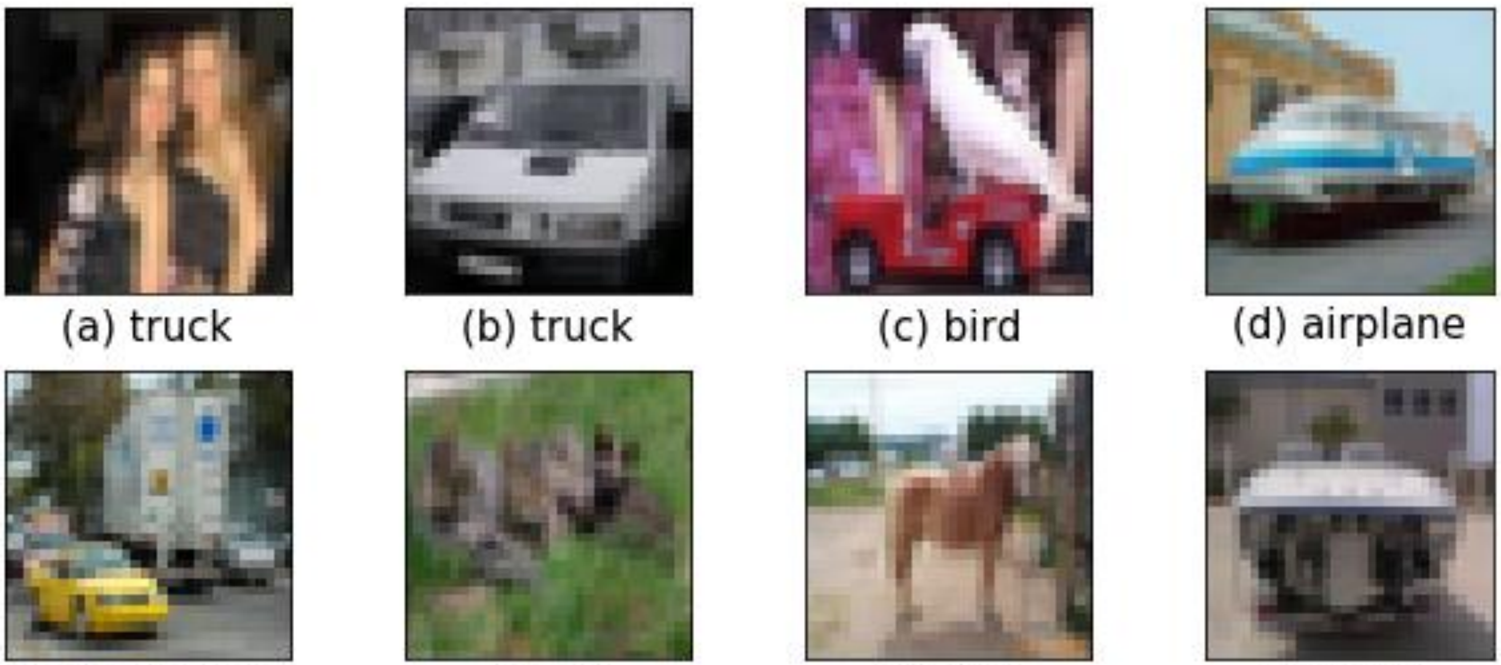
Understanding and Utilizing Deep Neural Networks Trained with Noisy Labels

Pengfei Chen
Supervisor: Prof. Shengyu Zhang, Prof, Shih-Chi Chen
Dept of Computer Science and Engineering
The Chinese University of Hong Kong

Does CIFAR contain noisy labels?

Introduction

Noisy labels exist even in CIFAR-10!



(a) truck (b) truck (c) bird (d) airplane

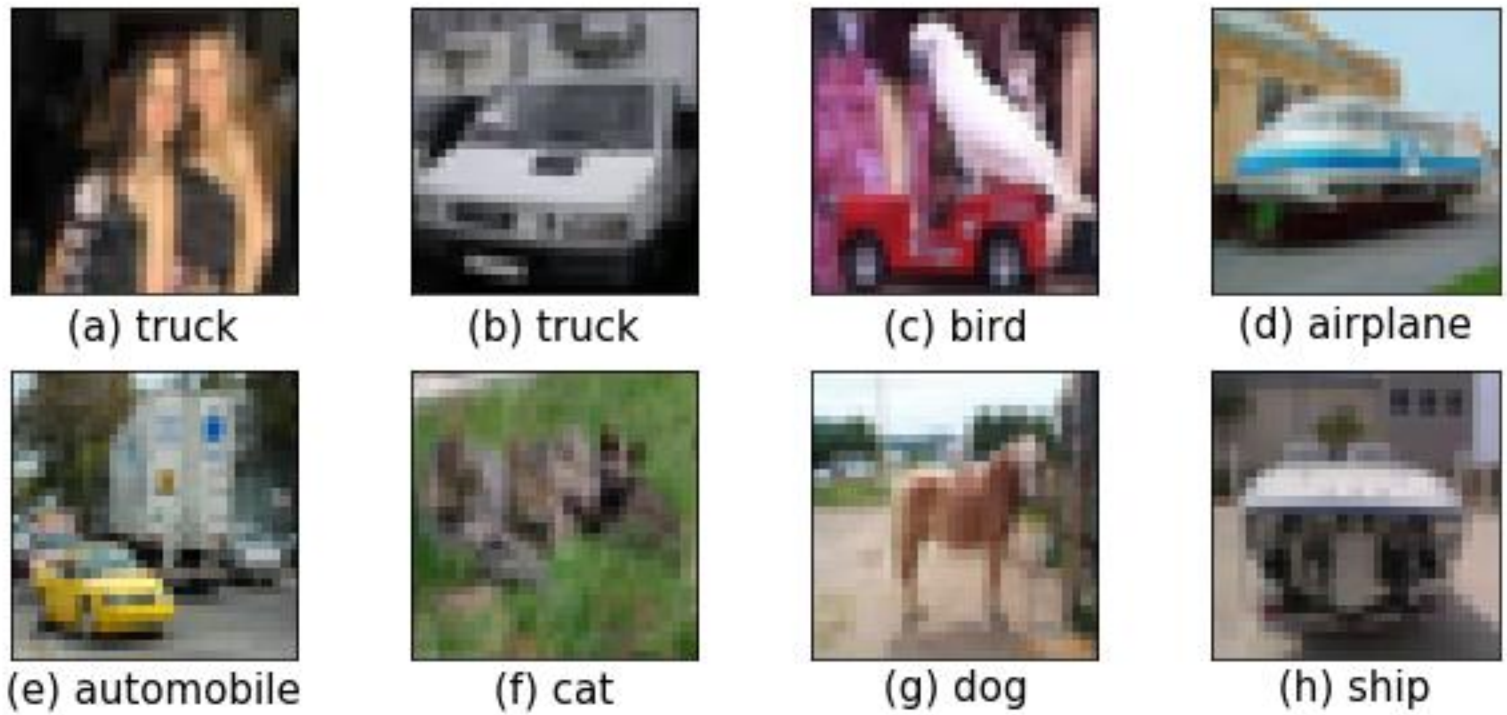
(e) automobile (f) cat (g) dog (h) ship

CIFAR-10, Krizhevsky & Hinton, 2009

Introduction

Noisy labels are ubiquitous

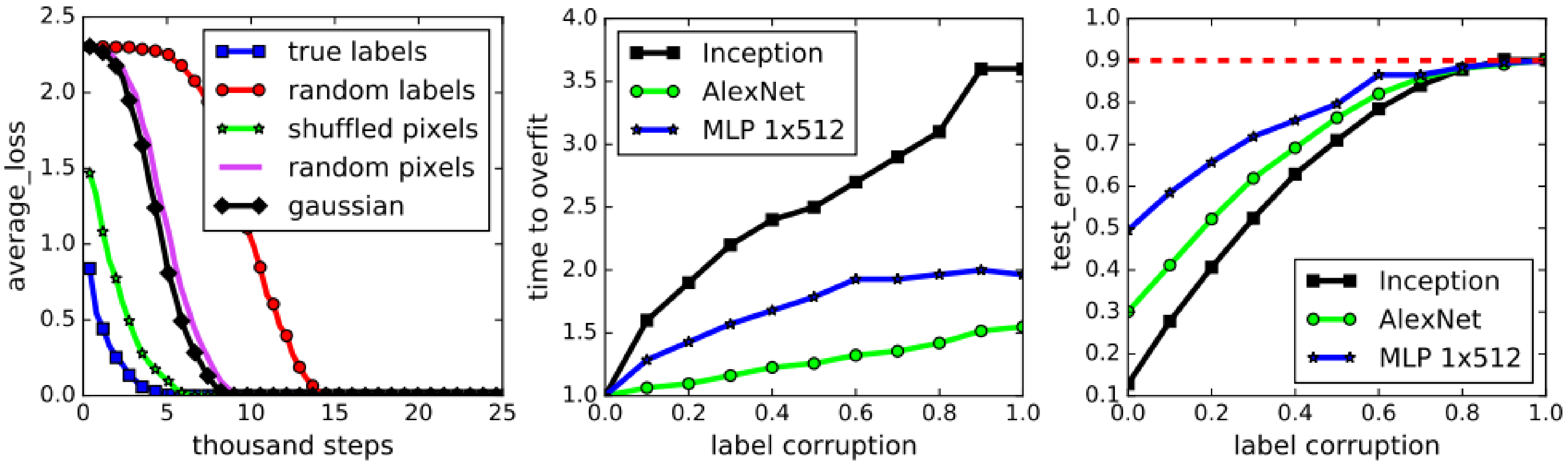
- Online queries (Schroff et al., 2011; Divvala et al., 2014)
- Crowdsourcing (Yan et al., 2014; Chen et al., 2017)



CIFAR-10, Krizhevsky & Hinton, 2009

Noisy labels are devastating

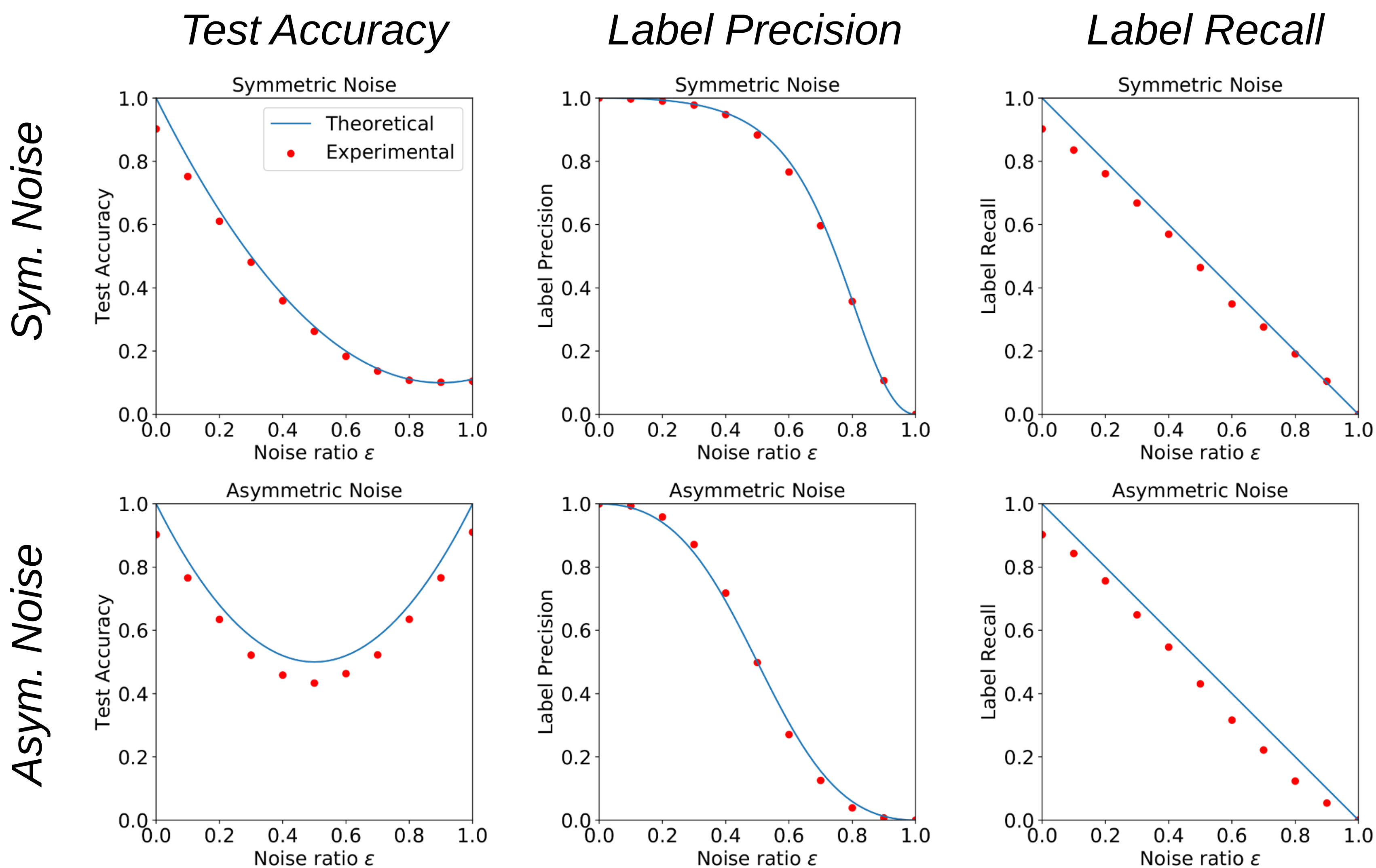
- Memorizing of noisy labels
- Poor generalization performance



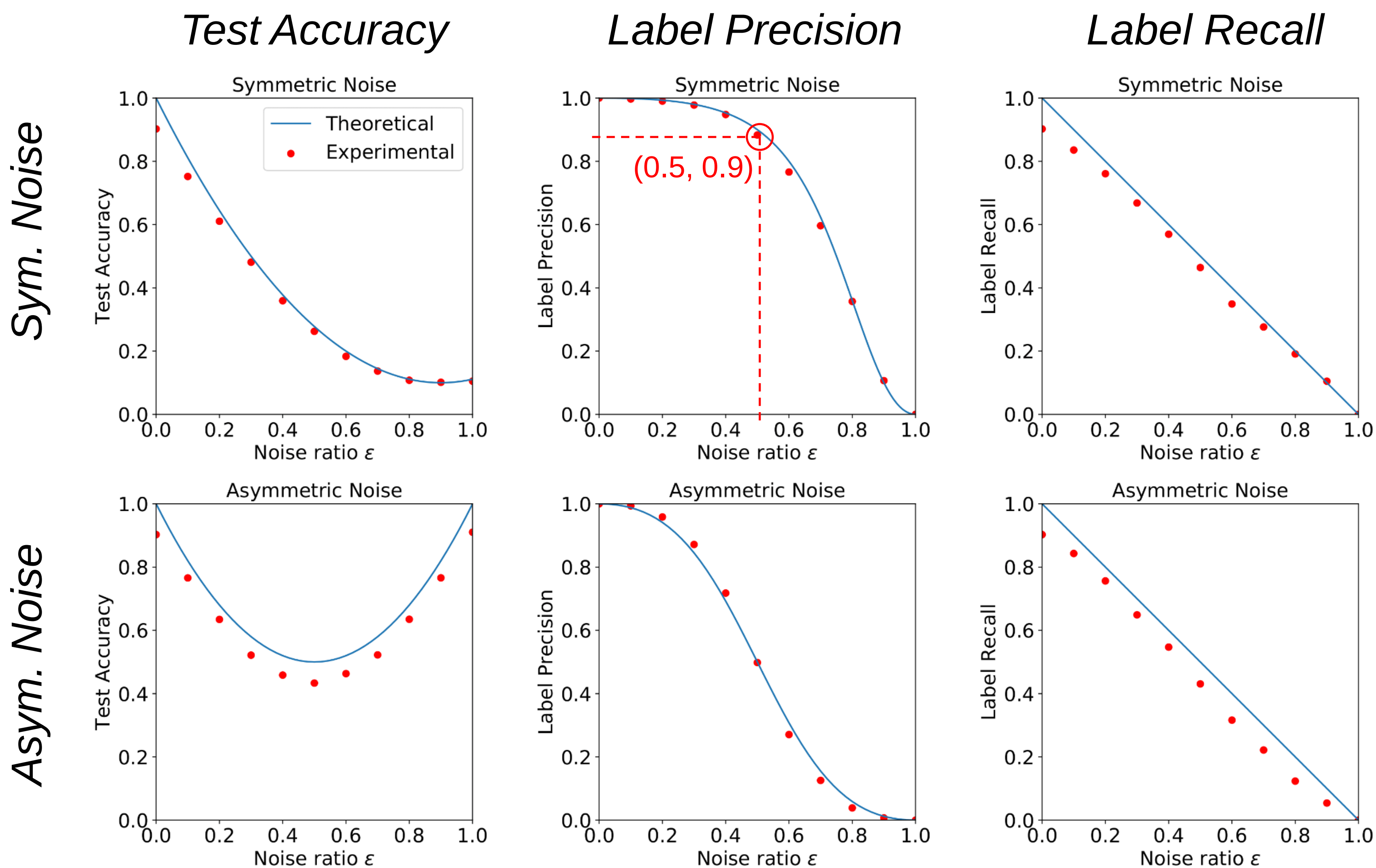
(a) learning curves (b) convergence slowdown (c) generalization error growth

Zhang et al., 2017

Cross-validation



Cross-validation



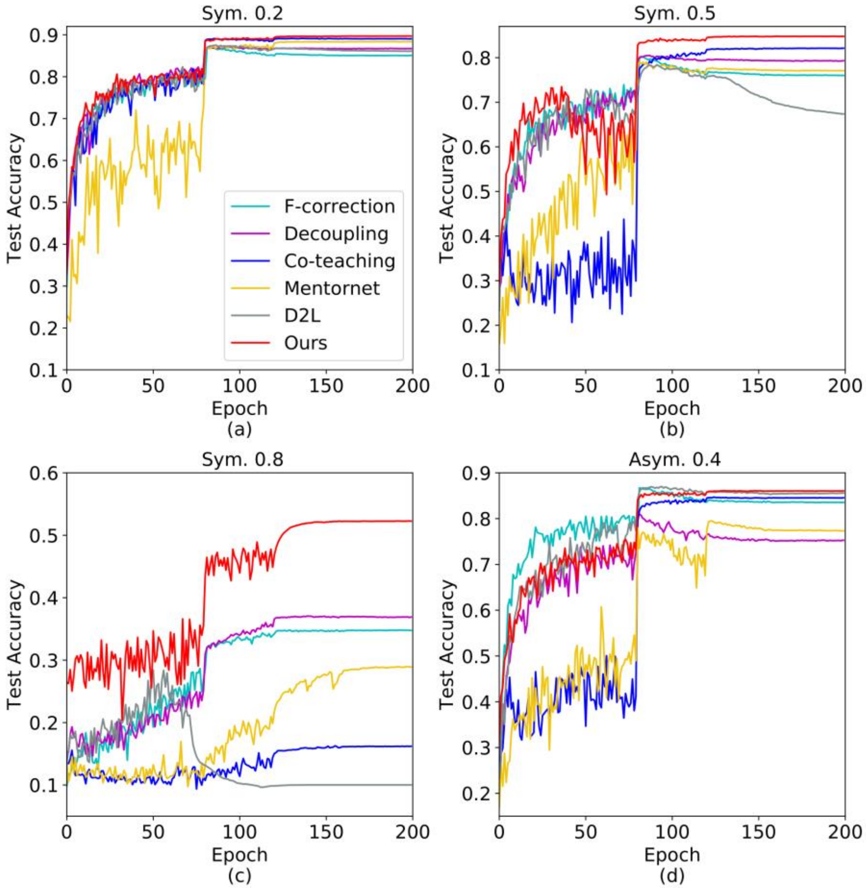
Training

CIFAR10

- Random flipping original labels
- Testing on the clean test set

Table 1. Test accuracy

Method	Sym.			Asym.
	0.2	0.5	0.8	0.4
F-correction	85.08 ±0.43	76.02 ±0.19	34.76 ±4.53	83.55 ±2.15
Decoupling	86.72 ±0.32	79.31 ±0.62	36.90 ±4.61	75.27 ±0.83
Co-teaching	89.05 ±0.32	82.12 ±0.59	16.21 ±3.02	84.55 ±2.81
MentorNet	88.36 ±0.46	77.10 ±0.44	28.89 ±2.29	77.33 ±0.79
D2L	86.12 ±0.43	67.39 ±13.62	10.02 ±0.04	85.57 ±1.21
Ours	**89.71** ±0.18	**84.78** ±0.33	**52.27** ±3.50	**86.04** ±0.54



Test accuracy during training

WebVision

- Crawled from websites using the same 1000 concepts as ImageNet
- Containing real-world noisy labels

Table 2. Test accuracy on WebVision val. and ILSVRC2012 val.

Method	WebVision Val.	ILSVRC2012 Val.
F-correction	61.12 (82.68)	57.36 (82.36)
Decoupling	62.54 (84.74)	58.26 (82.26)
Co-teaching	63.58 (85.20)	61.48 (84.70)
MentorNet	63.00 (81.40)	57.80 (79.92)
D2L	62.68 (84.00)	57.80 (81.36)
Ours	**65.24 (85.34)**	**61.60 (84.98)**

A formal study of noisy labels

- Relationship of noise level and test accuracy
- Mitigating the impact of label noise

Future work

- Structured data (E.g., Graph)
  - Social Networks
  - Molecules
  - Citation graphs

Alchemy Contest (Tencent, Quantum Lab)

- Graph Neural Networks (GNNs)
- Predicting properties of molecules
- 130,000+ molecules
- 12 properties

THANK YOU!

pfchen@cse.cuhk.edu.hk

Reference

1. Yan, Y., Rosales, R., Fung, G., Subramanian, R., and Dy, J. Learning from multiple annotators with varying expertise. Machine learning, 95(3):291–327, 2014.
2. He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. CVPR, 2016a.
3. Chen, G., Zhang, S., Lin, D., Huang, H., and Heng, P. A. Learning to aggregate ordinal labels by maximizing separating width. ICML, 2017.
4. Schroff, F., Criminisi, A., and Zisserman, A. Harvesting image databases from the web. TPAMI, 33(4):754–766, 2011.
5. Divvala, S. K., Farhadi, A., and Guestrin, C. Learning everything about anything: Webly-supervised visual concept learning. CVPR, 2014.
6. Li, W., Wang, L., Li, W., Agustsson, E., and Van Gool, L. Webvision database: Visual learning and understanding from web data. arXiv preprint arXiv:1708.02862, 2017.
7. Krizhevsky, A. and Hinton, G. Learning multiple layers of features from tiny images. Technical report, Citeseer, 2009.

Reference

8. Zhang, C., Bengio, S., Hardt, M., Recht, B., and Vinyals, O. Understanding deep learning requires rethinking generalization. ICLR, 2017.
9. Patrini, G., Rozza, A., Menon, A. K., Nock, R., and Qu, L. Making deep neural networks robust to label noise: A loss correction approach. CVPR, 2017.
10. Goldberger, J. and Ben-Reuven, E. Training deep neuralnetworks using a noise adaptation layer. ICLR, 2017.
11. Malach, E. and Shalev-Shwartz, S. Decoupling" when to update" from" how to update". NeurIPS, 2017.
12. Jiang, L., Zhou, Z., Leung, T., Li, L.-J., and Fei-Fei, L. Mentornet: Learning data-driven curriculum for very deep neural networks on corrupted labels. ICML, 2018.
13. Han, B., Yao, Q., Yu, X., Niu, G., Xu, M., Hu, W., Tsang, I., and Sugiyama, M. Co-teaching: robust training deep neural networks with extremely noisy labels. NeurIPS, 2018.
14. Reed, S., Lee, H., Anguelov, D., Szegedy, C., Erhan, D., and Rabinovich, A. Training deep neural networks on noisy labels with bootstrapping. ICLR, 2015.

Reference

15. Ren, M., Zeng, W., Yang, B., and Urtasun, R. Learning to reweight examples for robust deep learning. ICML, 2018.
16. Tanaka, D., Ikami, D., Yamasaki, T., and Aizawa, K. Joint optimization framework for learning with noisy labels. CVPR, 2018.
17. Ma, X., Wang, Y., Houle, M. E., Zhou, S., Erfani, S. M., Xia, S.-T., Wijewickrema, S., and Bailey, J. Dimensionalitydriven learning with noisy labels. ICML, 2018.
18. Arpit, D., Jastrz¸ebski, S., Ballas, N., Krueger, D., Bengio, E., Kanwal, M. S., Maharaj, T., Fischer, A., Courville, A., Bengio, Y., et al. A closer look at memorization in deep networks. ICML, 2017.