

# DECONFOUNDED NOISY LABELS LEARNING

**Anonymous authors**

Paper under double-blind review

## A APPENDIX

### A.1 ERROR ESTIMATION

To estimate the bound of error in Section 3.3, the following theorem is used, and the proof of the theorem can be found in Wang et al. (2021); Abramovich & Persson (2016); Gao et al. (2017).

**Theorem 1** *Given a non-linear function  $f : G \rightarrow \mathbb{R}$ ,  $G$  is a closed subset of  $\mathbb{R}$ , and a random variable  $X$  with the probability distribution  $p(X)$ , the expectation  $\mu$ , if the three conditions hold,*

- (1)  $f$  is bounded on any compact subset of  $\mathbb{R}$ ;
- (2)  $|f(x) - f(\mu)| = O(|x - \mu|^a)$  at  $x \rightarrow \mu$  for  $a > 0$
- (3)  $|f(x)| = O(|x|^b)$  at  $x \rightarrow \infty$  for  $b \geq a$

then,

$$|E[f(x)] - f(\mu)| \leq \sup_{x \in G, x \neq \mu} \frac{|f(x) - f(\mu)|}{|x - \mu|^a + |x - \mu|^b}$$

### A.2 FIVE NOISY DATASETS.

**Synthetic noise on CIFAR10 and CIFAR-100.** According to previous works Han et al. (2018); Xia et al. (2019; 2020); Bai et al. (2021), the two synthetic noisy datasets contain 50k training images and 10k test images of size  $32 \times 32$  with different levels of symmetric, pairflip, and instance-dependent label noise. Symmetric noise uniformly flips labels to all possible labels. Pairflip noise flips is generated by noisy labels into their adjacent class. Instance noise is generated by image features. More details of synthetic noise can be found in Bai et al. (2021).

**Human-annotated real-world noisy labels CIFAR-10N and CIFAR-100N.** In CIFAR-10N, each training image contains one clean label and three human annotated labels, with five different noisy-label sets. (1) Aggregate: aggregation of three noisy labels by majority voting. If the submitted three labels are different for an image, the aggregated label will be randomly selected among the three labels. (2) Random  $i$  ( $i \in \{1, 2, 3\}$ ): the  $i$ -th submitted label for each image. (3) Worst: dataset with the highest noise rate. For each image, if there exist any wrongly annotated labels in three noisy labels, the worst label is randomly selected from wrong labels. Otherwise, the worst label is equal to the clean label. The noise rates of the above mentioned five noisy label sets are 9.03% (Aggregate), 17.23% (Random 1), 18.12% (Random 2), 17.64% (Random 3) and 40.21% (Worst). In CIFAR-10N, 60.27% of the training images have received unanimous label from three independent labelers. All three random sets have  $\approx 18\%$  noise level. In CIFAR-100N dataset, each image contains a coarse label and a fine label given by a human annotator. Most batches have approximately 40% noisy fine labels and 25% noisy coarse labels. The overall noise level of coarse and fine labels are 25.60% and 40.20%, respectively.

**Clothing1M** Clothing1M consists of 1 million training images collected from online shopping websites with labels generated from surrounding texts. It contains 15k validation images, and 10k test images with clean labels.

### A.3 NETWORK STRUCTURE AND HYPERPARAMETERS.

For experiments without semi-supervised learning, we follow previous works Bai et al. (2021); Li et al. (2019), and use ResNet-18 for CIFAR-10 and ResNet-34 for CIFAR-100. The network is trained for 200 epochs and utilize SGD with a 0.9 momentum. The initial learning rate is set to 0.1

and decayed with a factor of 10 at the 100th and 150th epoch respectively, and a weight decay is set to  $10^{-4}$ . For finetuning, we employ an Adam optimizer with a learning rate of  $10^{-4}$ .

For experiments with semi-supervised learning, we follow the setting of Bai et al. (2021) with PreAct Resnet-18. The network is trained for 300 epochs. For optimization, we use a single cycle of cosine annealing, and the learning rate begins from 0.02 and ends at  $2 * 10^{-4}$ , with a weight decay of  $5 * 10^{-4}$ . An Adam optimizer is adopted with a learning rate of  $10^{-4}$ .

## REFERENCES

- Shoshana Abramovich and Lars-Erik Persson. Some new estimates of the ‘jensen gap’. *Journal of Inequalities and Applications*, 2016(1):1–9, 2016.
- Yingbin Bai, Erkun Yang, Bo Han, Yanhua Yang, Jiatong Li, Yinian Mao, Gang Niu, and Tongliang Liu. Understanding and improving early stopping for learning with noisy labels. *Advances in Neural Information Processing Systems*, 34:24392–24403, 2021.
- Xiang Gao, Meera Sitharam, and Adrian E Roitberg. Bounds on the jensen gap, and implications for mean-concentrated distributions. *arXiv preprint arXiv:1712.05267*, 2017.
- Bo Han, Quanming Yao, Xingrui Yu, Gang Niu, Miao Xu, Weihua Hu, Ivor Tsang, and Masashi Sugiyama. Co-teaching: Robust training of deep neural networks with extremely noisy labels. *Advances in neural information processing systems*, 31, 2018.
- Junnan Li, Richard Socher, and Steven CH Hoi. Dividemix: Learning with noisy labels as semi-supervised learning. In *International Conference on Learning Representations*, 2019.
- Wenjie Wang, Fuli Feng, Xiangnan He, Xiang Wang, and Tat-Seng Chua. Deconfounded recommendation for alleviating bias amplification. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pp. 1717–1725, 2021.
- Xiaobo Xia, Tongliang Liu, Nannan Wang, Bo Han, Chen Gong, Gang Niu, and Masashi Sugiyama. Are anchor points really indispensable in label-noise learning? *Advances in Neural Information Processing Systems*, 32, 2019.
- Xiaobo Xia, Tongliang Liu, Bo Han, Chen Gong, Nannan Wang, Zongyuan Ge, and Yi Chang. Robust early-learning: Hindering the memorization of noisy labels. In *International conference on learning representations*, 2020.