Neighborhood Collective Estimation for Noisy Label Identification and Correction *(Supplementary Material)*

Jichang Li[1,2][0000−0001−5778−2232], Guanbin Li[1*][0000−0002−4805−0926], Feng Liu[3][0000−0002−4811−7828], and Yizhou Yu[2*][0000−0002−0470−5548]

[1] Sun Yat-sen University, Guangzhou 510006, China
[2] The University of Hong Kong, Hong Kong
[3] Deepwise AI Lab, Beijing, China
csjcli@connect.hku.hk, liguanbin@mail.sysu.edu.cn, liufeng@deepwise.com, yizhouy@acm.org

We present additional implementation details and additional analysis of our proposed method, Neighborhood Collective Estimation (NCE), in this supplementary material.

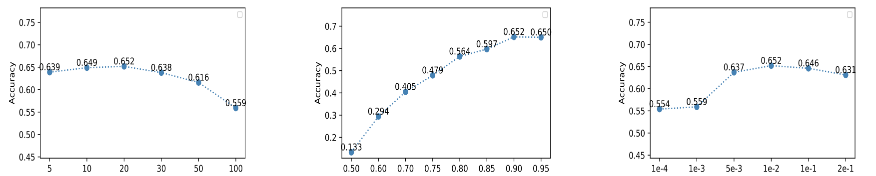
1 Additional Implementation Details

Our experiments on both Clothing-1M [5] and Webvision-1.0 [3] employ similar hyper-parameter settings, *e.g.*, $T_{wu} = 1$, $B = 32$, $B' = 32$, $\alpha = 0.5$ and $K = 20$. τ is set to 0.65 for Clothing-1M and 0.90 for Webvision-1.0. As shown in Table 2 and Table 3 in the main text, on these two datasets, our model only using clean samples in $\mathcal{D}_{\mathbf{clean}}$ for training outperforms previous state-of-the-art methods. Then, we follow the practice in "DivideMix" [2] and also set $\gamma = 0.0$ so that there is no need to set τ'. The learning rate schedule is the same for both datasets, that is, after half training epochs, the initial learning rate is divided by 10. The initial learning rate for Clothing-1M and Webvision-1.0 is set to 0.002 and 0.01 respectively. In addition, we choose Resnet-50 [1] and Inception-Resnet-V2 [4] as the backbones for Clothing-1M and Webvision-1.0, respectively. We train the models using a SGD optimizer with a momentum of 0.9 and a weight decay of 1×10^{-3}. Moreover, the number of training epochs for Clothing-1M and Webvision-1.0 are $T_{tr} = 80$ and $T_{tr} = 100$, respectively.

2 Additional Analysis

Hyper-parameter sensitivity. We also investigate the sensitivity of our proposed method to three key hyper-parameters, *i.e.*, K, τ and τ'. Taking CIFAR-100 with (Noise ratio: 0.80; Noise type: Symmetric) as an example, Fig. 1 shows that the model reaches a significantly high classification performance in this

*Corresponding Authors are Guanbin Li and Yizhou Yu.



(a) Performance variations to varying K **(b)** Performance variations to varying τ **(c)** Performance variations to varying τ'

Fig. 1. Sensitivity with respect to hyper-parameters K, τ and τ'. We conduct these experiments on CIFAR-100 with the same noise profile (Noise ratio: 0.80; Noise type: Symmetric). In this noise profile, our model achieves the best accuracy of 65.2% when we set $K = 20$, $\tau = 0.90$ and $\tau' = 0.01$.

LNL case when we set K, τ and τ' to 20, 0.90, and 0.01 respectively; on the other hand, a probable decrease in accuracy ensues when we change any of those parameters. With achieving fair comparisons, we follow "DivideMix" [2] to set other hyper-parameters that are involved in the training process or network architectures.

References

1. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: European conference on computer vision. pp. 630–645. Springer (2016)
2. Li, J., Socher, R., Hoi, S.C.: Dividemix: Learning with noisy labels as semi-supervised learning. arXiv preprint arXiv:2002.07394 (2020)
3. Li, W., Wang, L., Li, W., Agustsson, E., Gool, L.V.: Webvision database: Visual learning and understanding from web data. Arxiv Preprint (2017)
4. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-first AAAI conference on artificial intelligence (2017)
5. Xiao, T., Xia, T., Yang, Y., Huang, C., Wang, X.: Learning from massive noisy labeled data for image classification. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2691–2699 (2015)