Twin Contrastive Learning with Noisy Labels
—Supplmentary Material—

Zhizhong Huang[1]   Junping Zhang[1]   Hongming Shan[2,3*]

[1] Shanghai Key Lab of Intelligent Information Processing, School of Computer Science, Fudan University, Shanghai 200433, China

[2] Institute of Science and Technology for Brain-inspired Intelligence and MOE Frontiers Center for Brain Science, Fudan University, Shanghai 200433, China

[3] Shanghai Center for Brain Science and Brain-inspired Technology, Shanghai 200031, China

{zzhuang19, jpzhang, hmshan}@fudan.edu.cn

A. k-NN evaluation

We perform the k-NN classification over the learned representations with $k = 200$. For comparisons, we removed all proposed components and reported the performance on the representations learned in a pure unsupervised manner. The clean labels are involved in testing but excluded in the training phase. The results are shown in Tables A1 and A2.

The representations learned by our method have consistently outperformed the unsupervised learning, regardless of label noise with different ratios. These results indicate that our method has maintained meaningful representations better than the pure unsupervised learning model.

CIFAR-10

Noise type/rate	*Sym.*				*Asym.*	*Avg.*
	20%	50%	80%	90%	40%	
k-NN (**ours**)	94.9	94.0	92.2	90.6	92.8	**92.9**
k-NN (**unsup.**)			—			86.4

Table A1. k-NN evaluation on the learned representations of TCL and unsupervised baseline on CIFAR-10.

CIFAR-100

Noise type/rate	*Sym.*				*Avg.*
	20%	50%	80%	90%	
k-NN (**ours**)	76.7	72.6	67.3	64.1	**70.2**
k-NN (**unsup.**)			—		53.8

Table A2. k-NN evaluation on the learned representations of TCL and unsupervised baseline on CIFAR-100.

B. Asymmetric Label Noise

Table B3 shows the results of TCL and TCL+ for CIFAR-10/100 under different asymmetric ratios, where our method has consistently outperformed the competitors.

We note that, unlike *symmetric* label noise, the classes with above 50% *asymmetric* label noise cannot be distinguished, which makes 40% becomes the most extreme scenario. In addition, we found that the asymmetric label noise would make the dataset imbalance, where the assumption of uniform distribution does not hold.

Here, we employ the class imbalance ratio $r = \max(\{N_z\}_{z=1}^K)/\min(\{N_z\}_{z=1}^K)$ used in long-tailed learning to measure whether the label distribution is uniform, where K and N_z are the numbers of classes and samples in z-th class, respectively. The lower r is, the more uniform the distribution becomes. For CIFAR-10 under the extreme high asymmetric label noise (*i.e.* 40%), $r = 2.40$; that is, the asymmetric label noise makes the dataset non-uniform. However, TCL can still achieve pleasing performance on non-uniform datasets, which suggests that TCL can effectively detect those mislabeled samples to form a uniform distribution. Specifically, for those clean samples (clean probability $w_i > 0.5$), $r = 1.37$, which is much more balanced over noisy labels.

	CIFAR-10			**CIFAR-100**			
	10%	20%	30%	10%	20%	30%	40%
DivideMix [20]	93.8	93.2	92.5	69.5	69.2	68.3	51.0
ELR [25]	94.4	93.3	91.5	75.8	74.8	73.6	70.0
Sel-CL+ [23]	<u>95.6</u>	<u>95.2</u>	<u>94.5</u>	<u>78.7</u>	<u>77.5</u>	<u>76.4</u>	<u>74.2</u>
TCL (**ours**)	95.1	94.7	94.4	78.2	76.8	75.5	73.1
TCL+ (**ours**)	**95.9**	**95.3**	**94.8**	**79.0**	**78.0**	**76.9**	**74.4**

Table B3. Comparisons with SOTAs under *asymmetric* label noise.

*Corresponding author