RankMatch: Fostering Confidence and Consistency in Learning with Noisy Labels (Supplementary Material)

Ziyi Zhang$^{1,2}$, Weikai Chen$^3$, Chaowei Fang$^4$, Zhen Li$^5$, Lechao Chen$^6$, Liang Lin$^2$, Guanbin Li$^{2,7,*}$

$^1$National Key Laboratory of Novel Software Technology, Nanjing University, Nanjing, China
$^2$Sun Yat-sen University, Guangzhou, China, $^3$ Tencent America
$^4$ Xidian University, $^5$ The Chinese University of Hong Kong (Shenzhen), $^6$ Zhejiang Lab
$^7$ Research Institute, Sun Yat-sen University, Shenzhen, China

zhangziyi@lamda.nju.edu.cn, liguanbin@mail.sysu.edu.cn

1. Algorithm

Many existing methods $^{[2][4][5][4]}$ alternatively train two networks to combat the confirmation bias. In this paper, we show that simultaneously training two networks still performs well. Specifically, we divide training dataset $\mathcal{D}$ by Eq. 5 and get two clean sets $\mathcal{D}_{\text{cln}}^{(1)}, \mathcal{D}_{\text{cln}}^{(2)}$ by two networks $P(F(x, \theta_i)) (i \in \{1, 2\})$ respectively. The final clean set is:

$$\mathcal{D}_{\text{cln}} = \mathcal{D}_{\text{cln}}^{(1)} \cap \mathcal{D}_{\text{cln}}^{(2)}.$$  (1)

The remaining samples are regarded as noisy samples: $\mathcal{D}_{\text{ny}} = \mathcal{D} \setminus \mathcal{D}_{\text{cln}}$. Every image is augmented twice by the two types of augmentation: $(v^w_i, v'^w_i) = (A_w(x_i), A_s(x_i)), (v'^w_i, v'^w_i) = (A_w(x_i), A_s(x_i))$. Following DivideMix $^{[2]}$, weak augmented images $(v^w_i, v'^w_i)$ are leveraged to “co-guess” the correct labels for noisy samples and guide the learning of each network. The full algorithm for implementing RankMatch are shown below.

2. Additional Ablation Study

2.1. Rank Contrastive Loss

Our proposed Rank Contrastive Loss (RCL) strengthens the consistency of the similar samples while pushes “dissimilar” samples away, which makes features more discriminative and benefits the sample selection. Thus, we first visualize the features of training images using UMAP $^{[3]}$. As shown in Fig. 1 features derived by RCL become more compact, and the density of samples around the decision boundaries is reduced, which implies the representations become more discriminative.

2.2. Sensitivity Analysis

SCV introduces three hyperparameters: threshold $\tau$, number of prototypes in each class $K$ and the number of

*Corresponding author.
Figure 1: Visualization of features learned by RankMatch with or without RCL on CIFAR-100 with 80% label noise. Each variant of RankMatch is trained with 200 epochs. We randomly draw samples of 10 classes from the 100 classes. Compared with representations trained without RCL, full RankMatch generates more discriminative features.

(a) RankMatch without RCL

(b) RankMatch with RCL

Figure 2: Sensitivity to the variance of hyperparameters. Experiments are conducted on CIFAR-10 under 50% and 90% symmetric noises. We vary the threshold \(\tau\) from 0.90 to 0.99 to study the effect of confidence in RankMatch. SCV sets up multiple prototypes in each class as confident voting candidates and ranks k-nearest candidates as voters. We range the number of prototypes \(K\) from 5 to 30, and range the number of voters \(k\) from 2 to 10. The parameter \(r\) is used to rank top-\(r\) channels for RCL. We range it from 3 to 25.

Table 1: List of RankMatch hyperparameters for CIFAR

<table>
<thead>
<tr>
<th>Hyperparameter</th>
<th>CIFAR-10</th>
<th>CIFAR-100</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\lambda_n)</td>
<td>0.2</td>
<td>1</td>
</tr>
</tbody>
</table>
80 epochs. We initialize the learning rate as 0.003 and reduce it by a factor of 10 after 40 epochs. For WebVision, we train the network for 100 epochs, and reduce the initialized learning rate 0.01 by a factor of 10 after 50 epochs.

4. Additional Experimental Results

Table 2: Average test accuracy (%) on CIFAR-10 dataset over the last 10 epochs. We run our method three times with different random seeds and report the mean and the standard deviation.

<table>
<thead>
<tr>
<th>Noise type</th>
<th>Sym.</th>
<th>Asym.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20%</td>
<td>50%</td>
</tr>
<tr>
<td>Noise level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cross-Entropy</td>
<td>82.7</td>
<td>57.9</td>
</tr>
<tr>
<td>DivideMix</td>
<td>95.0</td>
<td>93.7</td>
</tr>
<tr>
<td>RankMatch</td>
<td>96.4 ± 0.08</td>
<td>95.39 ± 0.09</td>
</tr>
</tbody>
</table>

We run our method for three times with different random seeds and report the mean and the standard deviation in Table 2. Compared with the DivideMix [2], our method outperforms most recent methods by a large margin and with small standard deviation, which implies that our experimental results are statistically significant.

References