MDCS: More Diverse Experts with Consistency Self-distillation for Long-tailed Recognition (Supplementary Material)

Qihao Zhao¹,², Chen Jiang¹, Wei Hu¹, Fan Zhang¹*, Jun Liu²
¹ Beijing University of Chemical Technology, China
² Singapore University of Technology and Design, Singapore
{zhaoqh, jiangchen, huwei, zhangf}@mail.buct.edu.cn, jun_liu@sutd.edu.sg

1. The efficiency of our Consistency Self-distillation.

![Diagram](attachment:diagram.png)

As illustrated in Fig. 1, previous methods [3, 1, 4] reduced the model variance only by using an ensemble of multiple experts. In contrast, our approach not only reduces the variance by ensemble but also reduces the model variance by CS for each expert. The effect of CS is not only to reduce the model variance. Each expert gets richer constraint information through weakly augmented images, which enhances the expert’s own recognition ability. As shown in Table 1, experts with stronger recognition abilities also produce more diverse ensemble models.


We implement our method with PyTorch. Following [4, 2], we use ResNeXt-50/ResNet-50 for ImageNet-LT, ResNet-32 for CIFAR100/10-LT, ResNet-152 for Places-LT and ResNet-50 for iNaturalist 2018 as backbones, respectively. Moreover, we adopt the cosine classifier for prediction on all datasets. The details settings for our method are shown in Table 2.

References

Table 1. The efficiency of Consistency Self-distillation. With CS, not only is the model variance reduced, but also the expert recognition ability and the final model diversity are improved.

<table>
<thead>
<tr>
<th>Method</th>
<th>E1 Acc</th>
<th>E2 Acc</th>
<th>E3 Acc</th>
<th>All Acc</th>
<th>σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>w/o CS</td>
<td>38.8</td>
<td>45.2</td>
<td>31.4</td>
<td>50.7</td>
<td>53.4</td>
</tr>
<tr>
<td>w/ CS</td>
<td>40.6(+1.8)</td>
<td>50.8(+5.6)</td>
<td>36.0(+4.6)</td>
<td>56.1(+5.4)</td>
<td>60.4(+7.0)</td>
</tr>
</tbody>
</table>

Table 2. Statistics of the used network architectures and hyper-parameters in our experiments.