Supplementary Material: Calibrating CNNs for Lifelong Learning

Pravendra Singh∗1, Vinay Kumar Verma∗2, Pratik Mazumder1, Lawrence Carin2, Piyush Rai1
1CSE Department, IIT Kanpur, India 2Duke University, USA
psingh@cse.iitk.ac.in, vinaykumar.verma@duke.edu, pratikm@cse.iitk.ac.in, lcarin@duke.edu, piyush@cse.iitk.ac.in

1 Additional Results on SVHN

Table 1 reports the experimental results on the SVHN dataset for ResNet-18 and ResNet-18(1/3) architectures. ResNet-18(1/3) is simply ResNet-18 [1], with the number of filters in each layer reduced by 3 times [2]. We use SGD optimizer in all our experiments. In all cases, we run experiments for 5 random task orders and report the average accuracy. From the results, we can see that even with ResNet-18(1/3), which has lesser parameters than ResNet-18, results are comparable for CCLL<1,1> model. CCLL<4,1> with ResNet-18(1/3) performs even better as compared to CCLL<1,1> with ResNet-18.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Architecture</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Final (Āp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCLL&lt;1,1&gt;</td>
<td>ResNet-18</td>
<td>98.77</td>
<td>98.54</td>
<td>98.44</td>
<td>98.48</td>
<td>98.20</td>
</tr>
<tr>
<td>CCLL&lt;1,1&gt;</td>
<td>ResNet-18(1/3)</td>
<td>98.57</td>
<td>98.25</td>
<td>98.34</td>
<td>98.13</td>
<td>98.15</td>
</tr>
<tr>
<td>CCLL&lt;4,1&gt;</td>
<td>ResNet-18(1/3)</td>
<td>98.77</td>
<td>98.86</td>
<td>98.64</td>
<td>98.61</td>
<td>98.50</td>
</tr>
</tbody>
</table>

Table 1: Experimental results on SVHN dataset with ResNet-18 and ResNet-18(1/3) architectures. There are 5 tasks, and the reported accuracy for each task is the average of all accuracies up to that task.

2 Additional Results on CIFAR-100

Fig. 1 shows the experimental results for CIFAR-100 incremental learning tasks using 10, 20 and 50 classes at a time using ResNet-18(1/3) architecture. CCLL with larger values of α such as 2, 4, 8, performs better as shown in Fig. 1.

3 Additional Results on ImageNet-100/10

The results in Table 2 indicate that our method CCLL<4,1> performs better than CCLL<1,1> for ImageNet-100/10. However, CCLL<1,1> introduces 0.51% more parameters per task and CCLL<4,1> introduces 1.66% more parameters per task.

∗Equal contribution.
Figure 1: Experimental results on CIFAR-100 dataset with tasks containing 10, 20 and 50 classes with ResNet-18(1/3) architecture.

<table>
<thead>
<tr>
<th>Methods</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>Final ($A_{10}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCLL&lt;1,1&gt;</td>
<td>99.8</td>
<td>99.0</td>
<td>99.2</td>
<td>98.6</td>
<td>98.4</td>
<td>98.2</td>
<td>97.7</td>
<td>97.8</td>
<td>97.9</td>
<td></td>
</tr>
<tr>
<td>CCLL&lt;4,1&gt;</td>
<td>99.2</td>
<td>99.2</td>
<td>98.9</td>
<td>98.9</td>
<td>99.0</td>
<td>98.9</td>
<td>98.6</td>
<td>98.5</td>
<td>98.6</td>
<td>98.7</td>
</tr>
</tbody>
</table>

Table 2: Large-scale lifelong learning experiments on ImageNet dataset using ResNet-18 architecture. There are 10 tasks, and the reported accuracy for each task is the average of all accuracies up to that task.

References
