1. Datasets and Architectures

We conduct experiments on several benchmark datasets, including CIFAR100 [4], STL-10 [1], TinyImageNet and ImageNet [2]. Four architectures are used for the teacher and student networks, namely ResNet [3], VGG [8], ShuffleNet [10], MobileNet [7].

2. Implementation Details

The CIFAR100 dataset consists of 50,000 images of size 32 × 32 with 500 images per class and 10,000 test images. The TinyImageNet dataset is a subset of ImageNet, consisting of 100,000 images of size 64 × 64 from 200 classes. STL-10 consists of 5000 labeled training images from 10 classes and 100,000 unlabeled images, and a test set of 8,000 images. To keep our cross-modal transfer experiment’s consistency, we down-sample each image to size 32 × 32. We normalized all images by channel means and standard deviations.

Following the same experimental settings of existing works [9, 5, 6], we use the SGD optimizer with momentum for all networks. For MobileNetV2 and ShuffleNet, we use a learning rate of 0.01. For the rest of the networks, the learning rate is initialized with 0.05. All the learning rates are decayed by 0.1 every 30 epochs after the first 150 epochs until the last 240 epoch. We implement the networks and training procedures in Pytorch.

3. Additional Visualization Results

We further provide more visualization results on the CIFAR-100 datasets, which is illustrated in Fig 1 and Fig 2. We observe the similar results that the proposed HKD method have similar topology structure with the teacher network, which demonstrates the effectiveness of the proposed HKD method.

4. Additional Parameter Analysis

We further provide the parameter analysis on the number of layers in Graph Neural Networks. More specifically, we test the one layer(L=1) and two layer(L=2) graph neural networks. We do not set higher number of layers to avoid the over-smoothing problem.

<table>
<thead>
<tr>
<th>Teacher</th>
<th>ResNet32x4</th>
<th>VGG13</th>
<th>ResNet50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student</td>
<td>ResNet8x4</td>
<td>MobileNetV2</td>
<td>VGG8</td>
</tr>
<tr>
<td>L=1</td>
<td>76.13 ± 0.05</td>
<td>70.48±0.25</td>
<td>74.85±0.26</td>
</tr>
<tr>
<td>L=2</td>
<td>76.05±0.11</td>
<td>70.28±0.07</td>
<td>74.82±0.24</td>
</tr>
</tbody>
</table>
References


