Knowledge Distillation via Generative Adversarial Networks

Augmented Intelligent And Interaction (AII) Workshop

李哲榮
2018/05/19
Models for Smaller Devices

- Different forms for different stages
Methods

• Model compression
  – Pruning, quantization, data compression...

• Matrix/tensor decomposition
  – PCA, SVD, CPD, Sparse coding, fast convolution...

• Smaller models
  – SqueezeNet, MobileNet, ...

• Architecture search
  – PPP-net, ...

• Knowledge distillation
Knowledge Distillation

- soften output
- modify softmax

\[ q_i = \frac{\exp(z_i/T)}{\sum_j \exp(z_j/T)} \]
Generative Adversarial Networks

- Vanilla GAN

\[
\min_G \max_D V(D, G) = \log D(x) + \log(1 - D(G(z)))
\]
KDGAN

Input image

Teacher network

Real feature map

G

Fake feature map

C

D

Real / Fake

Soft target

Hard target

Student network
Experiments

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>68.53%</td>
</tr>
<tr>
<td>Logits Mimic Learning</td>
<td>50.95%</td>
</tr>
<tr>
<td>KD</td>
<td>69.14%</td>
</tr>
<tr>
<td>KDGAN</td>
<td>74.10%</td>
</tr>
<tr>
<td>Teacher(DenseNet-40)</td>
<td>74.23%</td>
</tr>
</tbody>
</table>

**Table 1.** Testing accuracy for training the student networks with 8 convolutional layers and 8M parameters by Baseline (typical training process), Logits Mimic Learning, KD, and KDGAN.

<table>
<thead>
<tr>
<th>Model</th>
<th>No. Parameters</th>
<th>Accuracy</th>
<th>Inference time</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 conv-20M (KDGAN)</td>
<td>20.2M</td>
<td>74.36%</td>
<td>4.56ms</td>
</tr>
<tr>
<td>8 conv-28M (KDGAN)</td>
<td>28.1M</td>
<td>75.25%</td>
<td>5.7ms</td>
</tr>
<tr>
<td>MobileNet (KDGAN)</td>
<td>3.5M</td>
<td>77.20%</td>
<td>2.79ms</td>
</tr>
<tr>
<td>MobileNet(Baseline)</td>
<td>3.5M</td>
<td>72.99%</td>
<td>2.79ms</td>
</tr>
<tr>
<td>DenseNet-100(Teacher)</td>
<td>7.2M</td>
<td>77.94%</td>
<td>18.02ms</td>
</tr>
</tbody>
</table>

**Table 8.** Testing accuracy and inference time for training simple CNNs with 8 convolutional layers and 20.2M, 28.1M parameters, and MobileNet as student networks by KDGAN.