Adversarial Knowledge Distillation for a Compact Generator

Hideki Tsunashima¹,², Hirokatsu Kataoka², Junji Yamato³, Qiu Chen³, Shigeo Morishima⁴
1) Waseda University, 2) AIST, 3) Kogakuin University, 4) Waseda Research Institute for Science and Engineering

Introduction

- Training SOTA GANs needs large amount computation resources (e.g. StyleGAN needs 7 days on 8 Tesla V100).
- Conventional compression methods for GANs.
  - LIT—only applied to ResNet architecture.
  - MSE-method—aggravate image quality from original GANs.

- Our contributions
  - Close image quality of a student to a teacher.
  - Ours isn’t constrained by network architecture.

Adversarial Knowledge Distillation for Generative models (AKDG)

![Diagram of AKDG](image)

- **AKDG**
  - The key concept is very simple but effective.
  - The key of AKDG is the additional discriminator which judges a teacher or student.

Results

- Confirming the effectiveness of AKDG on a toy dataset
  - 2D Ring
    - The connection between modes is sparse in the student trained with AKDG.
  - 2D Grid
    - The student does not miss the modes in comparison with the vanilla model.

![Images of 2D Ring and 2D Grid](image)

- Compressing GANs on the real datasets.
  - CIFAR-10

<table>
<thead>
<tr>
<th>Methods</th>
<th># layers</th>
<th># params</th>
<th># mult/add</th>
<th>IS</th>
<th>FID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher (PNGAN [2])</td>
<td>15</td>
<td>18.3M</td>
<td>8.9B</td>
<td>21.22</td>
<td></td>
</tr>
<tr>
<td>DCGAN vanilla (batch64) [40]</td>
<td>8</td>
<td>3.7M/79.8%</td>
<td>785M/91.2%</td>
<td>44.72</td>
<td></td>
</tr>
<tr>
<td>DCGAN vanilla (batch256) [40]</td>
<td>8</td>
<td>3.7M/79.8%</td>
<td>785M/91.2%</td>
<td>37.89</td>
<td></td>
</tr>
<tr>
<td>DCGAN-MSE [21]</td>
<td>8</td>
<td>3.7M/79.8%</td>
<td>785M/91.2%</td>
<td>98.96</td>
<td></td>
</tr>
<tr>
<td>DCGAN-AKDG (ours)</td>
<td>8</td>
<td>3.7M/79.8%</td>
<td>785M/91.2%</td>
<td>27.86</td>
<td></td>
</tr>
</tbody>
</table>

- **Why does not AKDG work in case of the small batchsize?**
- Visualize the real images and the images of the teacher using UMAP. Red circle indicates images of the teacher which does not overlap the real images.
- The distribution gap incurs the collapse of training.

![Visualization of real images and images of the teacher](image)

Conclusion and Future Work

Conclusion

We propose the novel compression method for unconditional GANs. In all experiments, our AKDG outperforms the conventional compression method.

Future Work

- We will fix the problem of large batchsize training.
- We will incorporate other network compression methods, such as pruning, quantization and tensor factorization.