UNSUPERVISED REPRESENTATION LEARNING WITH DEEP CONVOLUTIONAL GENERATIVE ADVERSARIAL NETWORKS

WENJING YANG
Overview

• Background
• DCGAN architecture
• DCGAN training: LSUN data, Imagenet-1k, Faces dataset
• DCGAN evaluation: CIFAR-10
• Contributions of DCGAN
• Future work
Model Architecture

Core approach: modifying 3 recent changes to CNN architectures

1. The all convolutional net
2. Trend towards eliminating fully connected layers on top of convolutional features
3. Batch Normalization
More Guidelines

- Replace pooling layers
- Use batch normalization for the generator and discriminator
- Remove fully connected hidden layers for deeper architectures
- Use ReLU as the activation function for the generator for all layers
- Use Tanh as the activation function for the generator for the output layer
- Use LeakyReLU as the activation function for the discriminator for all layers
Training DCGAN

DCGAN is trained on 3 datasets

1. Large-scale Scene Understanding (LSUN)
2. Imagenet-1k
3. Faces dataset
Details

1. No preprocessing on training images
2. Scaling to the range of the tanh activation function [-1, 1].
3. Train with mini-batch stochastic gradient descent (SGD) with size of 128
4. All weights initialized from N(0,0.2)
5. Slope of the leak was set to 0.2 in all models in the LeakyReLU
6. Adam optimizer is used and learning rate was set to 0.0002.
Example Architecture
LSUN Bedrooms Dataset
Five Epochs
Empirical Validation

• Classifying CIFAR-10 using GANs as a feature extractor
• Classifying SVHN digits using GANs as a feature extractor
## CIFAR-10 Dataset

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>Accuracy (400 per class)</th>
<th>max # of features units</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Layer K-means</td>
<td>80.6%</td>
<td>63.7% (±0.7%)</td>
<td>4800</td>
</tr>
<tr>
<td>3 Layer K-means Learned RF</td>
<td>82.0%</td>
<td>70.7% (±0.7%)</td>
<td>3200</td>
</tr>
<tr>
<td>View Invariant K-means</td>
<td>81.9%</td>
<td>72.6% (±0.7%)</td>
<td>6400</td>
</tr>
<tr>
<td>Exemplar CNN</td>
<td>84.3%</td>
<td>77.4% (±0.2%)</td>
<td>1024</td>
</tr>
<tr>
<td>DCGAN (ours) + L2-SVM</td>
<td>82.8%</td>
<td>73.8% (±0.4%)</td>
<td>512</td>
</tr>
</tbody>
</table>
SVHN Digits Dataset

<table>
<thead>
<tr>
<th>Model</th>
<th>error rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>KNN</td>
<td>77.93%</td>
</tr>
<tr>
<td>TSVM</td>
<td>66.55%</td>
</tr>
<tr>
<td>M1+KNN</td>
<td>65.63%</td>
</tr>
<tr>
<td>M1+TSVM</td>
<td>54.33%</td>
</tr>
<tr>
<td>M1+M2</td>
<td>36.02%</td>
</tr>
<tr>
<td>SWWAE without dropout</td>
<td>27.83%</td>
</tr>
<tr>
<td>SWWAE with dropout</td>
<td>23.56%</td>
</tr>
<tr>
<td>DCGAN (ours) + L2-SVM</td>
<td>22.48%</td>
</tr>
<tr>
<td>Supervised CNN with the same architecture</td>
<td>28.87% (validation)</td>
</tr>
</tbody>
</table>
Other applications

smiling woman

neutral woman

neutral man

= smiling man
Other applications
Contributions

• Improvement on architectural topology of convolutional GANs
• Better trained discriminators for image classification applications
• Ability to visualize filters
• More stable architecture for training GANs
Future work

• Instability problem in the architecture
• Extending applications to other domains
  • Frame prediction
  • Speech synthesis
• Theoretical support
Reference


Thank you!