Generative Adversarial Networks (GANs)

Ian Goodfellow, OpenAI Research Scientist
Re-Work Deep Learning Summit
San Francisco, 2017-01-26
Generative Modeling

• Density estimation

  \[ \ldots \ldots \ldots \ldots \ldots \ldots \]

• Sample generation

Training examples  \[ \ldots \ldots \ldots \ldots \ldots \ldots \]

  \[ \ldots \ldots \ldots \ldots \ldots \ldots \]

Model samples

(Goodfellow 2016)
CHAPTER 15. REPRESENTATION LEARNING

Figure 15.6: Predictive generative networks provide an example of the importance of learning which features are salient. In this example, the predictive generative network has been trained to predict the appearance of a 3-D model of a human head at a specific viewing angle.

(Left) Ground truth. This is the correct image, that the network should emit.

(Center) Image produced by a predictive generative network trained with mean squared error alone. Because the ears do not cause an extreme difference in brightness compared to the neighboring skin, they were not sufficiently salient for the model to learn to represent them.

(Right) Image produced by a model trained with a combination of mean squared error and adversarial loss. Using this learned cost function, the ears are salient because they follow a predictable pattern. Learning which underlying causes are important and relevant enough to model is an important active area of research. Figures graciously provided by Lotter et al. (2015).

A benefit of learning the underlying causal factors, as pointed out by Schölkopf et al. (2012), is that if the true generative process has x as an effect and y as a cause, then modeling \( p(x|y) \) is robust to changes in \( p(y) \). If the cause-effect relationship was reversed, this would not be true, since by Bayes' rule, \( p(x|y) \) would be sensitive to changes in \( p(y) \). Very often, when we consider changes in distribution due to different domains, temporal non-stationarity, or changes in the nature of the task, the causal mechanisms remain invariant (the laws of the universe are constant) while the marginal distribution over the underlying causes can change. Hence, better generalization and robustness to all kinds of changes can be achieved.

(Lotter et al 2016)
iGAN

(youtube)

(Zhu et al 2016)

IAN

(youtube)

(Brock et al 2016)
Image to Image Translation

(Isola et al 2016)
Fully Visible Belief Nets

• Explicit formula based on chain (Frey et al, 1996) rule:
  \[ p_{\text{model}}(\mathbf{x}) = p_{\text{model}}(x_1) \prod_{i=2}^{n} p_{\text{model}}(x_i \mid x_1, \ldots, x_{i-1}) \]

• Disadvantages:
  • \( O(n) \) sample generation cost
  • Generation not controlled by a latent code

PixelCNN elephants
(van den Ord et al 2016)
WaveNet

Amazing quality
Sample generation slow

Two minutes to synthesize one second of audio

(Goodfellow 2016)
Adversarial Nets Framework

- $D(x)$ tries to be near 1
- $G(z)$ tries to make $D(G(z))$ near 0
- $D$ tries to make $D(G(z))$ near 1

$x$ sampled from data

$D$ differentiable function

$G$ differentiable function

Input noise $z$ sampled from model
Vector Space Arithmetic

Figure 15.9: A generative model has learned a distributed representation that disentangles the concept of gender from the concept of wearing glasses. If we begin with the representation of the concept of a man with glasses, then subtract the vector representing the concept of a man without glasses, and finally add the vector representing the concept of a woman without glasses, we obtain the vector representing the concept of a woman with glasses. The generative model correctly decodes all of these representation vectors to images that may be recognized as belonging to the correct class. Images reproduced with permission from Radford et al. (2015).

A common is that one could imagine learning about each of them without having to see all the configurations of all the others. Radford et al. (2015) demonstrated that a generative model can learn a representation of images of faces, with separate directions in representation space capturing different underlying factors of variation. Figure 15.9 demonstrates that one direction in representation space corresponds to whether the person is male or female, while another corresponds to whether the person is wearing glasses. These features were discovered automatically, not fixed a priori. There is no need to have labels for the hidden unit classifiers: gradient descent on an objective function of interest naturally learns semantically interesting features, so long as the task requires such features. We can learn about the distinction between male and female, or about the presence or absence of other features by examples covering all of these combinations of values. This form of statistical separability is what allows one to generalize to new configurations of a person's features that have never been seen during training.

(Radford et al, 2015)
3D GAN

Figure 7: Qualitative results of single image 3D reconstruction on the IKEA dataset

(Wu et al, 2016)
OpenAI GAN-created images
Problems with Counting
Problems with Perspective
Problems with Global Structure
This one is real
Semi-Supervised Classification

### CIFAR-10

<table>
<thead>
<tr>
<th>Model</th>
<th>Test error rate for a given number of labeled samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1000</td>
</tr>
<tr>
<td>Ladder network [24]</td>
<td></td>
</tr>
<tr>
<td>CatGAN [14]</td>
<td></td>
</tr>
<tr>
<td>Our model</td>
<td>21.83 ± 2.01</td>
</tr>
<tr>
<td>Ensemble of 10 of our models</td>
<td>19.22 ± 0.54</td>
</tr>
</tbody>
</table>

### SVHN

<table>
<thead>
<tr>
<th>Model</th>
<th>Percentage of incorrectly predicted test examples for a given number of labeled samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>500</td>
</tr>
<tr>
<td>DGN [21]</td>
<td>36.02 ± 0.10</td>
</tr>
<tr>
<td>Virtual Adversarial [22]</td>
<td>24.63</td>
</tr>
<tr>
<td>Auxiliary Deep Generative Model [23]</td>
<td>22.86</td>
</tr>
<tr>
<td>Skip Deep Generative Model [23]</td>
<td>16.61 ± 0.24</td>
</tr>
<tr>
<td>Our model</td>
<td>18.44 ± 4.8</td>
</tr>
<tr>
<td>Ensemble of 10 of our models</td>
<td>5.88 ± 1.0</td>
</tr>
</tbody>
</table>

(Salimans et al 2016)
Learning interpretable latent codes / controlling the generation process

InfoGAN (Chen et al 2016)
Plug and Play Generative Networks

(Nguyen et al 2016)
PPGN for caption to image

oranges on a table next to a liquor bottle

(Nguyen et al 2016)
GAN loss is a key ingredient

Figures S8: A comparison of images produced by different generators

Images from Nguyen et al 2016
First observed by Dosovitskiy et al 2016
StackGANs

This small blue bird has a short pointy beak and brown on its wings.

This bird is completely red with black wings and pointy beak.

A small sized bird that has a cream belly and a short pointed bill.

A small bird with a black head and wings and features grey wings.

(Zhang et al 2016)
Conclusion

• GANs produce rich, realistic imagery

• GANs learn to draw samples from a probability distribution

• Applications include learning from very few labeled examples, interactive artwork generation, and differential privacy