Generative Adversarial Networks (GANs)

Ian Goodfellow, OpenAI Research Scientist
Presentation at AI With the Best, 2016-09-24
Generative Modeling

• Density estimation

• Sample generation

[Diagram showing a process with 'Training examples' on the left and 'Model samples' on the right, with intermediate steps in between.]
Adversarial Nets Framework

D tries to output 1

Differentiable function $D$

$x$ sampled from data

D tries to output 0

Differentiable function $D$

$x$ sampled from model

Differentiable function $G$

Input noise $Z$
DCGAN Architecture

Most “deconvs” are batch normalized

(Radford et al 2015)
DCGANs for LSUN Bedrooms

(Radford et al 2015)
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).

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 glasses, without having to characterize all of the configurations of the 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.
Mode Collapse

- Fully optimizing the discriminator with the generator held constant is safe.
- Fully optimizing the generator with the discriminator held constant results in mapping all points to the argmax of the discriminator.
- Can partially fix this by adding nearest-neighbor features constructed from the current minibatch to the discriminator ("minibatch GAN")
  (Salimans et al 2016)
Minibatch GAN on CIFAR

<table>
<thead>
<tr>
<th>Training Data</th>
<th>Samples</th>
</tr>
</thead>
</table>

(Salimans et al 2016)
Minibatch GAN on ImageNet

(Salimans et al 2016)
Cherry-Picked Results
Text to Image with GANs

this small bird has a pink breast and crown, and black primaries and secondaries.

this magnificent fellow is almost all black with a red crest, and white cheek patch.

the flower has petals that are bright pinkish purple with white stigma

this white and yellow flower have thin white petals and a round yellow stamen

(Reed et al 2016)
Generating Pokémon

youtube

(Yota Ishida)
Single Image Super-Resolution

original
bicubic (21.59dB/0.6423)
SRResNet (23.44dB/0.7777)
SRGAN (20.34dB/0.6562)

(Ledig et al 2016)
iGAN

youtube

(Zhu et al 2016)
Introspective Adversarial Networks

youtube
Conclusion

• GANs are generative models based on supervised learning and game theory

• GANs learn to generate realistic samples

• Like other generative models, GANs still need a lot of improvement