Generative Adversarial Networks

Ian Goodfellow, Staff Research Scientist, Google Brain
NVIDIA Distinguished Lecture Series in Machine Learning
USC, Los Angeles 2017-09-05
Generative Modeling

- Density estimation
  
  ![Diagram showing density estimation](image)

- Sample generation
  
  ![Diagram showing sample generation](image)

Training examples

Model samples

(Goodfellow 2017)
Maximum Likelihood

\[ \theta^* = \arg \max_\theta \mathbb{E}_{x \sim p_{data}} \log p_{model}(x \mid \theta) \]
Adversarial Nets Framework

- $D(x)$ tries to be near 1
  - Differentiable function $D$
  - $x$ sampled from data

- $D$ tries to make $D(G(z))$ near 0,
  - $G$ tries to make $D(G(z))$ near 1
  - $x$ sampled from model

- Differentiable function $G$
- Input noise $z$

(Goodfellow et al., 2014)
What can you do with GANs?

- Simulated environments and training data
- Missing data
  - Semi-supervised learning
- Multiple correct answers
- Realistic generation tasks
- Simulation by prediction
- Solve inference problems
- Learn useful embeddings
Apple’s first research paper tries to solve a problem facing every company working on AI
GANs for simulated training data

(Unlabeled Real Images)

(Shrivastava et al., 2016)
GANs for domain adaptation

(Bousmalis et al., 2016)
What can you do with GANs?

- Simulated environments and training data
- Missing data
  - Semi-supervised learning
- Multiple correct answers
- Realistic generation tasks
- Simulation by prediction
- Solve inference problems
- Learn useful embeddings
What is in this image?

(Yeh et al., 2016)
Generative modeling reveals a face

(Yeh et al., 2016)
What can you do with GANs?

- Simulated environments and training data
- Missing data
  - Semi-supervised learning
- Multiple correct answers
- Realistic generation tasks
- Simulation by prediction
- Solve inference problems
- Learn useful embeddings
Supervised Discriminator

(Odena 2016, Salimans et al 2016)
Semi-Supervised Classification

MNIST: 100 training labels -> 80 test mistakes
SVHN: 1,000 training labels -> 4.3% test error
CIFAR-10: 4,000 labels -> 14.4% test error

(Dai et al 2017)
What can you do with GANs?

- Simulated environments and training data
- Missing data
  - Semi-supervised learning
- Multiple correct answers
- Realistic generation tasks
- Simulation by prediction
- Solve inference problems
- Learn useful embeddings
Next Video Frame Prediction

Ground Truth

What happens next?

(Lotter et al 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.

Next Video Frame Prediction

Ground Truth  MSE  Adversarial

(Lotter et al 2016)
Next Video Frame(s) Prediction

Mean Squared Error
Mean Absolute Error
Adversarial

(Mathieu et al. 2015)
What can you do with GANs?

• Simulated environments and training data
• Missing data
  • Semi-supervised learning
• Multiple correct answers
• Realistic generation tasks
• Simulation by prediction
• Solve inference problems
• Learn useful embeddings
iGAN

youtube

(Zhu et al., 2016)
Introspective Adversarial Networks

(Brock et al., 2016)
Image to Image Translation

(Isola et al., 2016)
Unsupervised Image-to-Image Translation

Day to night

(Liu et al., 2017)
CycleGAN

(Zhu et al., 2017)
This bird has a yellow belly and tarsus, grey back, wings, and brown throat, nape with a black face.

(Zhang et al., 2016)
What can you do with GANs?

• Simulated environments and training data

• Missing data
  • Semi-supervised learning

• Multiple correct answers

• Realistic generation tasks
  • Simulation by prediction

• Solve inference problems

• Learn useful embeddings
Simulating particle physics

Save millions of dollars of CPU time by predicting outcomes of explicit simulations

(de Oliveira et al., 2017)
What can you do with GANs?

- Simulated environments and training data
- Missing data
  - Semi-supervised learning
- Multiple correct answers
- Realistic generation tasks
- Simulation by prediction
- Solve inference problems
- Learn useful embeddings
Adversarial Variational Bayes

\[(\mu, \tau)\]

\[(\tau, \eta_1)\]

AVB

VB (full-rank)

HMC

(Mescheder et al, 2017)
What can you do with GANs?

- Simulated environments and training data
- Missing data
  - Semi-supervised learning
- Multiple correct answers
- Realistic generation tasks
- Simulation by prediction
- Solve inference problems
- Learn useful embeddings
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.

(Radford et al, 2015)
Learning interpretable latent codes / controlling the generation process

InfoGAN (Chen et al 2016)
How long until GANs can do this?

Training examples

Model samples
AC-GANs

(Odena et al., 2016)
Minibatch GAN on ImageNet

(Salimans et al., 2016)
Cherry-Picked Results
Problems with Counting
Problems with Perspective
Problems with Global Structure
This one is real
Challenges

- Non-convergence, especially mode collapse
- Discrete output variables
Non-convergence

- Recent theoretical work argues that existing GAN training algorithms should converge under some reasonable conditions (Nagarajan and Kolter 2017, Heusel et al 2017).

- The convergence may be very slow because the Jacobian of the player’s training gradients with respect to their parameters has unfortunate eigenvalue structure (Roth et al 2017).

- Mode collapse remains poorly understood; no widespread agreement on whether it is primarily a form of non-convergence.
Discrete output variables

• Tasks like text generation for machine translation require a generator that produces discrete outputs

• GAN training requires the output to be differentiable with respect to the generator parameters

• Straightforward approaches like Gumbel-Softmax and REINFORCE have so far been disappointing on NLP tasks
Track updates at the GAN Zoo

https://github.com/hindupuravinash/the-gan-zoo
Conclusion

• GANs are generative models based on game theory

• GANs open the door to a wide range of engineering tasks

• There are still important research challenges to solve before GANs can generate arbitrary data