Generative Adversarial Networks (GANs)

Josh Zhanson

Two main types:

Generative and **Discriminative**

Discriminative Model

- Discriminates between two different classes of data

- Example: convolutional neural net that outputs 1 if input is a human face, 0 if not

Generative Model

- *Generates* new data that fit the distribution of the training data

- Example: Gaussian Mixture Model (GMM) can generate new random points that pretty much match the training distribution

But finding a generative model for training data is calculation heavy:

Calculations include:

- Maximum-likelihood
- Marginal probabilities
- Partition functions
- Most-likely estimates
- Etc.

Possible if finding a simple GMM, but prohibitive with deep neural nets…

Enter **adversarial training**!

For training data: $X \subset \mathbb{R}^d$, we have

- Generator g: $\mathbb{R}^n \rightarrow \mathbb{R}^d$
- Discriminator d: $\mathbb{R}^d \rightarrow \{0,1\}$

But both g and d are neural nets!

We train them both, alternating between the two

Generating/discriminating can be expressed as loss function

Modeled as a zero-sum or minimax game

i.e. counterfeiter vs. police

Eventually, the generator becomes good at fooling sophisticated trained discriminators

In practice, us!

Advantages:

Generating artificial data indistinguishable from real data by a neural net = generating realistic data

Doesn’t require probability computations like maximum likelihood approach

Disadvantages:

Hard to train, no closed form loss function like log-loss or mean squared error

Therefore, a lot more trial and error, really complex tasks difficult, require a lot of hyperparameter tuning

Versus Variational Autoencoders (VAEs)

"VAEs use log-likelihood to optimize, can use to evaluate quality of model

But injected noise + imperfect reconstruction + standard decoder = blurry images"

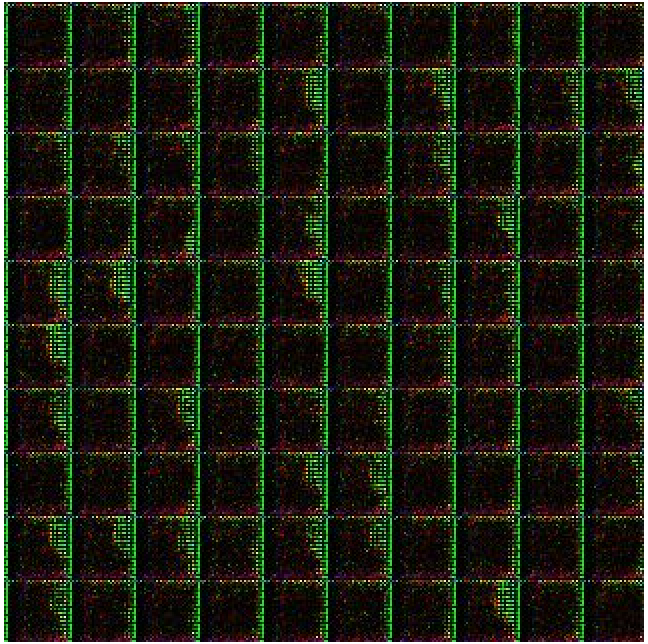
“VAEs optimize likelihood (always probability mass from estimated data manifold) while GANs optimize something else (can be good with very sharp estimated density function even if does not perfectly coincide with data density), both an advantage and a disadvantage”

“GANs tend to be much more finicky to train than VAEs, not to mention that we do not have a clear objective function to optimize, but they tend to yield nicer images.”

-Yoshua Bengio

VAEs vs. GANs (OpenAI)

VAEs, log time



GANs, linear time

Issues with Monte Carlo (GANs don’t need it, Boltzmann machines rely on it)

“Boltzmann machines have never really scaled to realistic tasks like ImageNet. GANs are at least able to learn to draw a few messed up dogs when trained on ImageNet.”
-Ian Goodfellow

Figure 6: Samples generated from the ImageNet dataset. *(Left)* Samples generated by a DCGAN. *(Right)* Samples generated using the techniques proposed in this work. The new techniques enable GANs to learn recognizable features of animals, such as fur, eyes, and noses, but these features are not correctly combined to form an animal with realistic anatomical structure.

“Improved Techniques for Training GANs”

arXiv:1606.03498 [cs.LG]

Do GANs converge?

Ian Goodfellow: “Unclear, but probably most important question”

In practice on small samples, sometimes, large samples, so far, never

Ian Goodfellow’s research direction: the convergence problem

“The basic issue is that all the theory says GANs should be great at the Nash equilibrium, but gradient descent is only guaranteed to get to the Nash equilibrium in the convex case. When both players are represented by neural nets, it’s possible for them to keep adapting their strategies forever without actually arriving at the equilibrium.”

Adequate live demo:

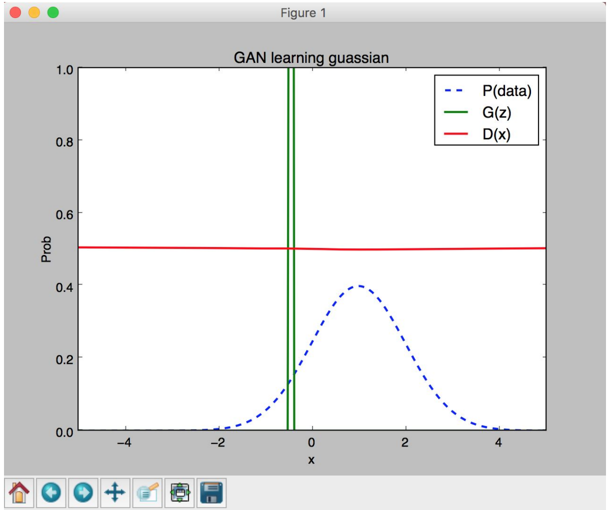
http://cs.stanford.edu/people/karpathy/gan/

Better demo code source:

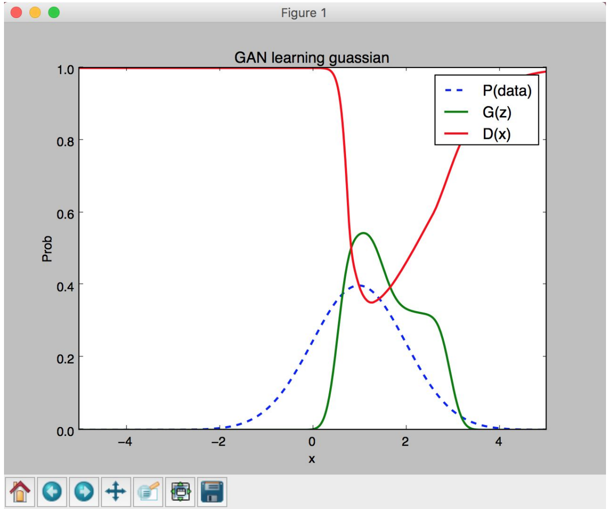
https://gist.github.com/Newmu/4ee0a712454480df5ee3

Run on

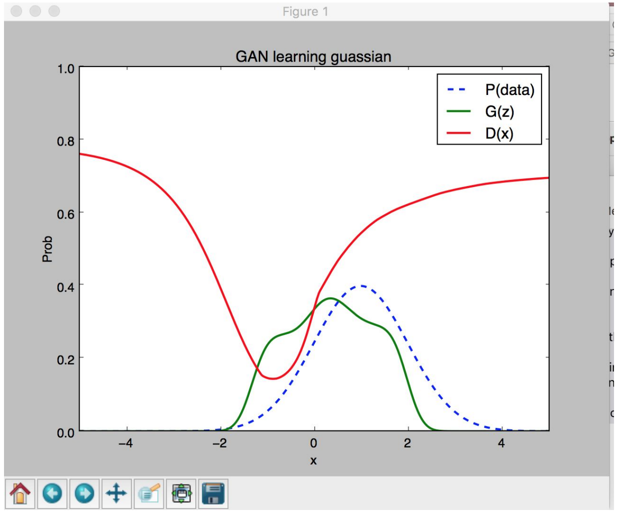
https://github.com/llSourcell/Generative-Adversarial-Network-Demo



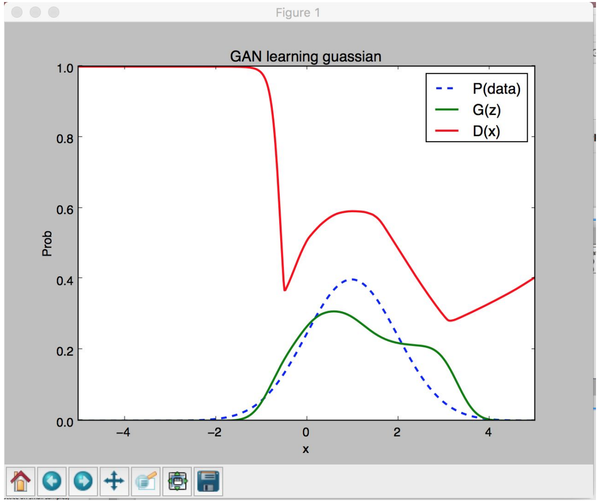
Initialization



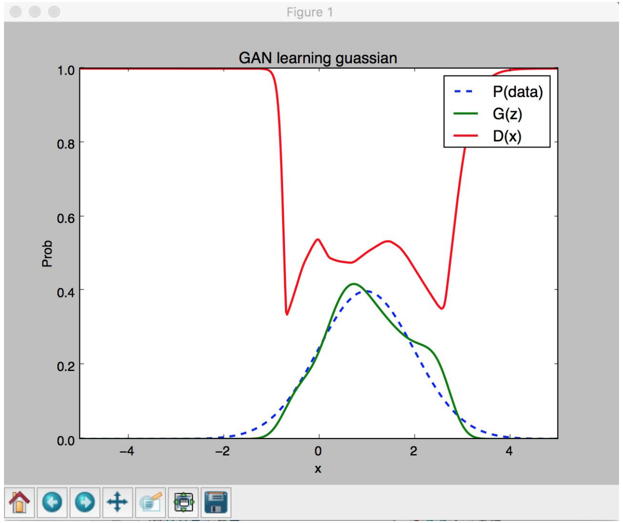
90 epochs



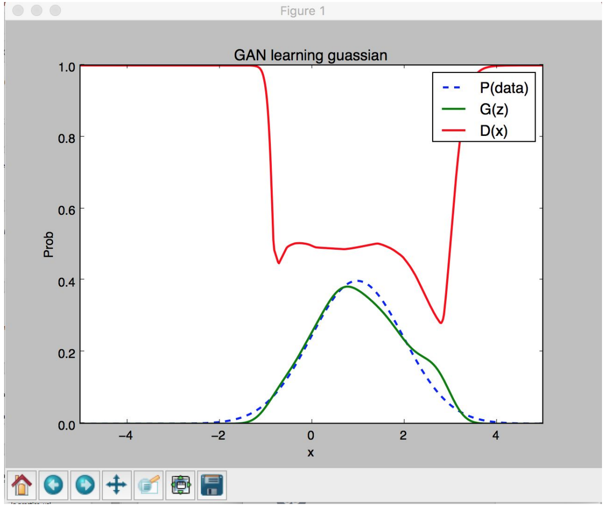
250 epochs



500 epochs



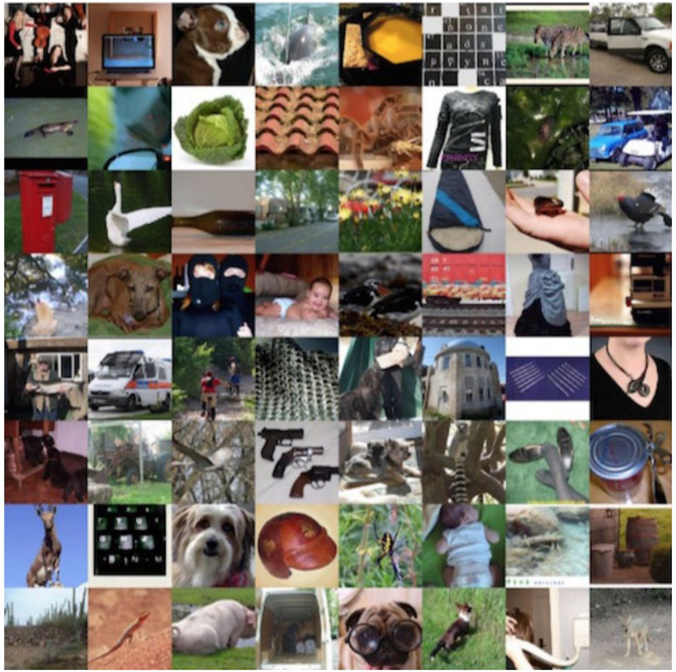
1000 epochs



1500 epochs

Results, Important Papers (further reading)

OpenAI ImageNet Samples

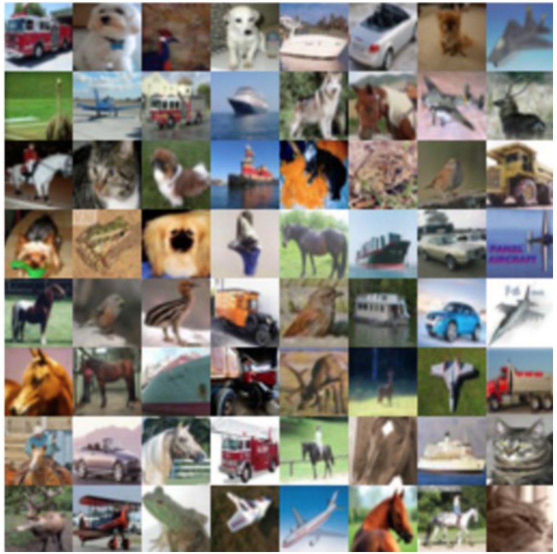


Real images (ImageNet)

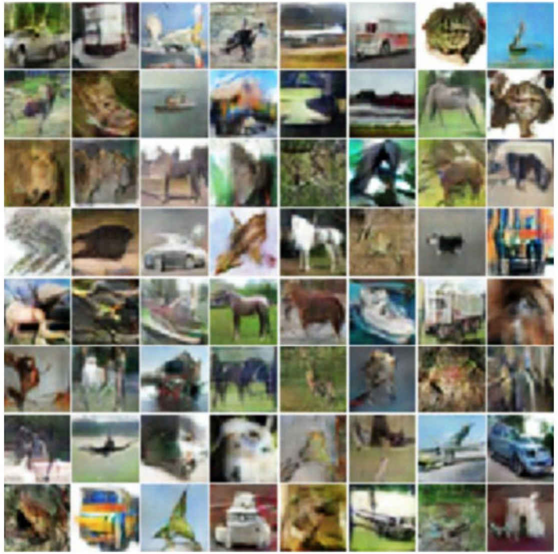
Generated images

OpenAI CIFAR-10

Our CIFAR-10 samples also look very sharp - Amazon Mechanical Turk workers can distinguish our samples from real data with an error rate of 21.3% (50% would be random guessing):



Real images (CIFAR-10)



Generated images

“Generative Adversarial Nets”

Generative Adversarial Nets

Ian J. Goodfellow[*] **Jean Pouget-Abadie**[†] **Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair**[‡] **Aaron Courville, Yoshua Bengio**[§]
Département d’informatique et de recherche opérationnelle
Université de Montréal
Montréal, QC H3C 3J7

Abstract

We propose a new framework for estimating generative models via an adversarial process, in which we simultaneously train two models: a generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than G. The training procedure for G is to maximize the probability of D making a mistake. This framework corresponds to a minimax two-player game. In the space of arbitrary functions G and D, a unique solution exists, with G recovering the training data distribution and D equal to $\frac{1}{2}$ everywhere. In the case where G and D are defined by multilayer perceptrons, the entire system can be trained with backpropagation. There is no need for any Markov chains or unrolled approximate inference networks during either training or generation of samples. Experiments demonstrate the potential of the framework through qualitative and quantitative evaluation of the generated samples.

“Improving Techniques for Training GANs”

Improved Techniques for Training GANs

Tim Salimans tim@openai.com
Ian Goodfellow ian@openai.com
Wojciech Zaremba woj@openai.com
Vicki Cheung vicki@openai.com
Alec Radford alec.radford@gmail.com
Xi Chen peter@openai.com

Abstract

We present a variety of new architectural features and training procedures that we apply to the generative adversarial networks (GANs) framework. We focus on two applications of GANs: semi-supervised learning, and the generation of images that humans find visually realistic. Unlike most work on generative models, our primary goal is not to train a model that assigns high likelihood to test data, nor do we require the model to be able to learn well without using any labels. Using our new techniques, we achieve state-of-the-art results in semi-supervised classification on MNIST, CIFAR-10 and SVHN. The generated images are of high quality as confirmed by a visual Turing test: our model generates MNIST samples that humans cannot distinguish from real data, and CIFAR-10 samples that yield a human error rate of 21.3%. We also present ImageNet samples with unprecedented resolution and show that our methods enable the model to learn recognizable features of ImageNet classes.

“Generating images with recurrent adversarial networks”

Generating images with recurrent adversarial networks

Daniel Jiwoong Im[1]
Montreal Institute for Learning Algorithms
University of Montreal
imdaniel@iro.umontreal.ca

Chris Dongjoo Kim
Department of Engineering and Computer Science
York University
kimdon20@gmail.com

Hui Jiang
Department of Engineering and Computer Science
York University
hj@cse.yorku.ca

Roland Memisevic
Montreal Institute for Learning Algorithms
University of Montreal
memisevr@iro.umontreal.ca

Abstract

Gatys et al. (2015) showed that optimizing pixels to match features in a convolutional network is a way to render images of high visual quality. Unrolling this gradient-based optimization can be thought of as a recurrent computation, that creates images by incrementally adding onto a visual "canvas". Inspired by this view we propose a recurrent generative model that can be trained using adversarial training. In order to quantitatively compare adversarial networks we also propose a new performance measure, that is based on letting the generator and discriminator of two models compete against each other.

bution (by plugging samples into the decoder).

The second class of generative models is based on adversarial sampling [4]. This approach forgoes the need to encourage a particular latent distribution (and, in fact, the use of an encoder altogether), by training a simple feed-forward neural network to generate "data-like" examples. "Data-likeness" is judged by a simultaneously trained, but otherwise separate, discriminator neural network.

For both types of approach, sequential variants were introduced recently, which were shown to work much better in terms of visual quality: The DRAW network [5], for example, is a sequential version of the variational autoencoder, where images are generated by accumulating updates into a canvas using a recurrent network. An example of a sequential adversarial network is the LAPGAN model [1], which

“Generative Adversarial Text to Image Synthesis” — VERY COOL

Generative Adversarial Text to Image Synthesis

Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran REEDSCOT[1], AKATA[2], XCYAN[1], LLAJAN[1]
Bernt Schiele, Honglak Lee SCHIELE[2],HONGLAK[1]
[1] University of Michigan, Ann Arbor, MI, USA (UMICH.EDU)
[2] Max Planck Institute for Informatics, Saarbrücken, Germany (MPI-INF.MPG.DE)

Abstract

Automatic synthesis of realistic images from text would be interesting and useful, but current AI systems are still far from this goal. However, in recent years generic and powerful recurrent neural network architectures have been developed to learn discriminative text feature representations. Meanwhile, deep convolutional generative adversarial networks (GANs) have begun to generate highly compelling images of specific categories, such as faces, album covers, and room interiors. In this work, we develop a novel deep architecture and GAN formulation to effectively bridge these advances in text and image modeling, translating visual concepts from characters to pixels. We demonstrate the capability of our model to generate plausible images of birds and flowers from detailed text descriptions.

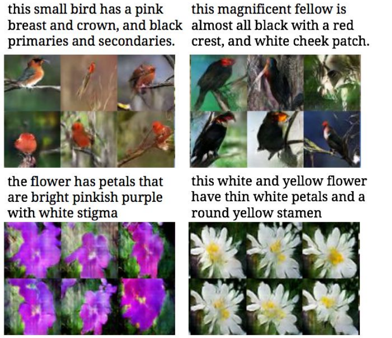


Figure 1. Examples of generated images from text descriptions. Left: captions are from zero-shot (held out) categories, unseen text. Right: captions are from the training set.

Bunch of Quora links (Ian Goodfellow, Yoshua Bengio):

https://www.quora.com/What-are-the-pros-and-cons-of-using-generative-adversarial-networks-a-type-of-neural-network

https://www.quora.com/In-what-way-are-Adversarial-Networks-related-or-different-to-Adversarial-Training/answer/Ian-Goodfellow

https://www.quora.com/Do-generative-adversarial-networks-always-converge

https://www.quora.com/What-is-the-advantage-of-generative-adversarial-networks-compared-with-other-generative-models

https://www.quora.com/What-are-some-exciting-future-applications-of-Generative-Adversarial-Networks

https://www.quora.com/What-are-Generative-Adversarial-Networks

https://www.quora.com/Can-Generative-Adversarial-networks-use-multi-class-labels

This one is crazy:
https://www.quora.com/What-is-missing-from-adversarial-networks-for-them-to-truly-model-a-representation-of-the-world

Ian Goodfellow answered this one himself, which is pretty funny:
https://www.quora.com/What-research-directions-is-Ian-Goodfellow-pursuing-to-improve-Generative-Adversarial-Networks

A Youtube Link: https://www.youtube.com/watch?v=deyOX6Mt_As

Brief (spotty) summary:
https://adeshpande3.github.io/Deep-Learning-Research-Review-Week-1-Generative-Adversarial-Nets

OpenAI (Ian Goodfellow works here now): https://openai.com/blog/generative-models/

Here is the original presentation on Google Drive