

• Define the generator

• Generate a sample image

• Training a GAN

• Training difficulties

• Other resources

• End

ON OUR RADAR

AI

DATA

DESIGN

ECONOMY

JUPYTER

OPERATIONS

SOFTWARE ARCHITECTURE

SOFTWARE ENGINEERING

IDEAS

Learning Platform

Conferences

Shop

SIGN IN

ARTIFICIAL INTELLIGENCE

Generative Adversarial Networks for beginners

Build a neural network that learns to generate handwritten digits.

By Jon Bruner and Adit Deshpande. June 7, 2017

- Define the generator

- Generate a sample image

- Training a GAN

- Training difficulties

- Other resources

- End

Practical Generative Adversarial Networks for Beginners

You can download and modify the code from [this tutorial on GitHub here](#).

According to Yann LeCun, “adversarial training is the coolest thing since sliced bread.” Sliced bread certainly never created this much excitement within the deep learning community. Generative adversarial networks—or GANs, for short—have dramatically sharpened the possibility of AI-generated content, and have drawn active research efforts since they were first described by Ian Goodfellow et al. in 2014.

GANs are neural networks that learn to create synthetic data similar to some known input data. For instance, researchers have generated convincing images from photographs of everything from bedrooms to album covers, and they display a remarkable ability to reflect higher-order semantic logic.

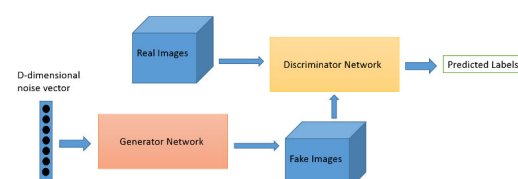
Those examples are fairly complex, but it's easy to build a GAN that generates very simple images. In this tutorial, we'll build a GAN that analyzes lots of images of handwritten digits and gradually learns to generate new images from scratch—*essentially, we'll be teaching a neural network how to write*.



Sample images from the generative adversarial network that we'll build in this tutorial. During training, it gradually refines its ability to generate digits.

GAN architecture

Generative adversarial networks consist of two models: a generative model and a discriminative model.



-
- Define the generator
-
-
-
-
-
-
- Generate a sample image
-
-
- Training a GAN
-
-
-
-
-
-
-
-
-
-
-
- Training difficulties
-
-
- Other resources
- End

The discriminator model is a classifier that determines whether a given image looks like a real image from the dataset or like an artificially created image. This is basically a binary classifier that will take the form of a normal convolutional neural network (CNN).

The generator model takes random input values and transforms them into images through a deconvolutional neural network.

Over the course of many training iterations, the weights and biases in the discriminator and the generator are trained through backpropagation. The discriminator learns to tell "real" images of handwritten digits apart from "fake" images created by the generator. At the same time, the generator uses feedback from the discriminator to learn how to produce convincing images that the discriminator can't distinguish from real images.

Getting started

We're going to create a GAN that will generate handwritten digits that can fool even the best classifiers (and humans too, of course). We'll use [TensorFlow](#), a deep learning library open-sourced by Google that makes it easy to train neural networks on GPUs.

This tutorial expects that you're already at least a little bit familiar with TensorFlow. If you're

"Hello,
not, we recommend reading [TensorFlow!](#)"
"Hello,
or watching the [Tensorflow!](#)" interactive
tutorial on Safari before proceeding.

Loading MNIST data

We need a set of real handwritten digits to give the discriminator a starting point in distinguishing between real and fake images. We'll use [MNIST](#), a benchmark dataset in deep learning. It consists of 70,000 images of handwritten digits compiled by the U.S. National Institute of Standards and Technology from Census Bureau employees and high school students.

Adit

Adit D
Presid
intere:
passio
educa'

- Define the generator
- Generate a sample image
- Training a GAN
- Training difficulties
- Other resources
- End

Let's start by importing TensorFlow along with a couple of other helpful libraries. We'll also import our MNIST images using a TensorFlow convenience function called `read_data_sets`.

```

1 | import tensorflow as tf
2 | import numpy as np
3 | import datetime
4 | import matplotlib.pyplot as plt
5 | %matplotlib inline
6 |
7 | from tensorflow.examples.tutorials.mnist import input_data
8 | mnist = input_data.read_data_sets("MNIST_data/")

```

< Run >

The MNIST variable we created above contains both the images and their labels, divided into a training set called `train` and a validation set called `validation`. (We won't need to worry about the labels in this tutorial.) We can retrieve batches of images by calling `next_batch` on `mnist`. Let's load one image and look at it.

The images are initially formatted as a single row of 784 pixels. We can reshape them into 28 x 28 pixel images and view them using PyPlot.

```

1 | sample_image = mnist.train.next_batch(1)[0]
2 | print(sample_image.shape)
3 |
4 | sample_image = sample_image.reshape([28, 28])
5 | plt.imshow(sample_image, cmap='Greys')

```

< Run >

If you run the cell above again, you'll see a different image from the MNIST training set.

Discriminator network

Our discriminator is a convolutional neural network that takes in an image of size 28 x 28 x 1 as input and returns a single scalar number that describes whether or not the input image is "real" or "fake"—that is, whether it's drawn from the set of MNIST images or generated by the generator.



The structure of our discriminator network is based closely on

- Define the generator

- Generate a sample image

- Training a GAN

- Training difficulties

- Other resources

- End

TensorFlow's sample CNN classifier

`model`. It features two convolutional layers that find 5x5 pixel features, and two "fully connected" layers that multiply weights by every pixel in the image.

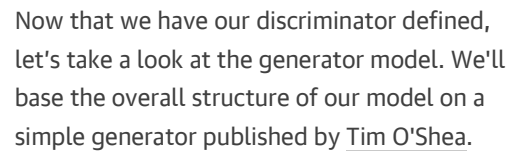
To set up each layer, we start by creating weight and bias variables through `tf.get_variable`. Weights are initialized from a truncated normal distribution, and biases are initialized at zero.

`tf.nn.conv2d()` is TensorFlow's standard convolution function. It takes 4 arguments. The first is the input volume (our `28 x 28 x 1` images in this case). The next argument is the filter/weight matrix. Finally, you can also change the stride and padding of the convolution. Those two values affect the dimensions of the output volume.

If you're already comfortable with CNNs, you'll recognize this as a simple binary classifier—nothing fancy.

```
1 def discriminator(images, reuse=False):
2     if (reuse):
3         tf.get_variable_scope().reuse_variables()
4
5     # First convolutional and pool layers
6     # This finds 32 different 5 x 5 pixel features
7     d_w1 = tf.get_variable('d_w1', [5, 5, 1, 32],
8                             initializer=tf.truncated_normal_initializer(stddev=0.02))
9     d_b1 = tf.get_variable('d_b1', [32],
10                            initializer=tf.constant_initializer(0))
11     d1 = tf.nn.conv2d(input=images, filter=d_w1, strides=[1, 1, 1, 1], padding='SAME')
12     d1 = d1 + d_b1
13     d1 = tf.nn.relu(d1)
14     d1 = tf.nn.avg_pool(d1, ksize=[1, 2, 2, 1], strides=[1, 2, 2, 1], padding='SAME')
15
16     # Second convolutional and pool layers
17     # This finds 64 different 5 x 5 pixel features
18     d_w2 = tf.get_variable('d_w2', [5, 5, 32, 64],
19                             initializer=tf.truncated_normal_initializer(stddev=0.02))
20     d_b2 = tf.get_variable('d_b2', [64],
21                            initializer=tf.constant_initializer(0))
22     d2 = tf.nn.conv2d(input=d1, filter=d_w2, strides=[1, 1, 1, 1], padding='SAME')
23     d2 = d2 + d_b2
24     d2 = tf.nn.relu(d2)
25     d2 = tf.nn.avg_pool(d2, ksize=[1, 2, 2, 1], strides=[1, 2, 2, 1], padding='SAME')
26
27     # First fully connected layer
28     d_w3 = tf.get_variable('d_w3', [7 * 7 * 64, 1024],
29                             initializer=tf.truncated_normal_initializer(stddev=0.02))
30     d_b3 = tf.get_variable('d_b3', [1024],
31                            initializer=tf.constant_initializer(0))
32     d3 = tf.reshape(d2, [-1, 7 * 7 * 64])
33     d3 = tf.matmul(d3, d_w3)
34     d3 = d3 + d_b3
35     d3 = tf.nn.relu(d3)
```

- Training a GAN



You can think of the generator as a kind of reverse convolutional neural network. A typical CNN like our discriminator network transforms a 2- or 3-dimensional matrix of pixel values into a single probability. A generator, however, takes a d -dimensional vector of noise and upsamples it to become a 28 x 28 image. ReLU and batch normalization are used to stabilize the outputs of each layer.

In our generator network, we use three convolutional layers along with interpolation until a 28×28 pixel image is formed. (Actually, as you'll see below, we've taken care to form $28 \times 28 \times 1$ images; many TensorFlow tools for dealing with images anticipate that the images will have some number of *channels*—usually 1 for greyscale images or 3 for RGB color images.)

At the output layer we add a `tf.sigmoid()` activation function; this squeezes pixels that would appear grey toward either black or white, resulting in a crisper image.

11.12.2018

Run

Now we've defined both the generator and discriminator functions. Let's see what a sample output from an untrained generator looks like.

- Define the generator
- Generate a sample image
- Training a GAN
- Training difficulties
- Other resources
- End

keyword, we don't have to specify `batch_size` until later.

```

1 | z_dimensions = 100
2 | z_placeholder = tf.placeholder(tf.float32, [None,
  | z_dimensions])

```

Run

Now, we create a variable (`generated_image_output`) that holds the output of the generator, and we'll also initialize the random noise vector that we're going to use as input. The `np.random.normal()` function has three arguments. The first and second define the mean and standard deviation for the normal distribution (0 and 1 in our case), and the third defines the the shape of the vector (`1 x 100`).

```

1 | generated_image_output = generator(z_placeholder, 1,
  | z_dimensions)
2 | z_batch = np.random.normal(0, 1, [1, z_dimensions])

```

Run

Next, we initialize all the variables, feed our `z_batch` into the placeholder, and run the session.

The `sess.run()` function has two arguments. The first is called the "fetches" argument; it defines the value you're interested in computing. In our case, we want to see what the output of the generator is. If you look back at the last code snippet, you'll see that the output of the generator function is stored in `generated_image_output`, so we'll use `generated_image_output` for our first argument.

The second argument takes a dictionary of inputs that are substituted into the graph when it runs. This is where we feed in our placeholders. In our example, we need to feed our `z_batch` variable into the `z_placeholder` that we defined earlier. As before, we'll view the image by reshaping it to `28 x 28` pixels and show it with PyPlot.

```

1 | with tf.Session() as sess:
2 |     sess.run(tf.global_variables_initializer())
3 |     generated_image = sess.run(generated_image_output,
4 |                               feed_dict={z_placeholder:
  | z_batch})
5 |     generated_image = generated_image.reshape([28, 28])
6 |     plt.imshow(generated_image, cmap='Greys')

```

Run

- Define the generator

- Generate a sample image

- Training a GAN

- Training difficulties

- Other resources

- End

That looks like noise, right? Now we need to train the weights and biases in the generator network to convert random numbers into recognizable digits. Let's look at loss functions and optimization!

Training a GAN

One of the trickiest parts of building and tuning GANs is that they have two loss functions: one that encourages the generator to create better images, and the other that encourages the discriminator to distinguish generated images from real images.

We train both the generator and the discriminator simultaneously. As the discriminator gets better at distinguishing real images from generated images, the generator is able to better tune its weights and biases to generate convincing images.

Here are the inputs and outputs for our networks.

```

1 |tf.reset_default_graph()
2 |batch_size = 50
3 |
4 |z_placeholder = tf.placeholder(tf.float32, [None,
5 |# z_placeholder is for feeding input noise to the
6 |generator
7 |x_placeholder = tf.placeholder(tf.float32, shape =
8 |[None,28,28,1], name='x_placeholder')
9 |# x_placeholder is for feeding input images to the
10 |discriminator
11 |Gz = generator(z_placeholder, batch_size, z_dimensions)
12 |# Gz holds the generated images
13 |Dx = discriminator(x_placeholder)
14 |# Dx will hold discriminator prediction probabilities
15 |for the real MNIST images
16 |
17 |Dg = discriminator(Gz, reuse=True)
18 |# Dg will hold discriminator prediction probabilities for
19 |generated images

```

< Run >

So, let's first think about what we want out of our networks. The discriminator's goal is to correctly label real MNIST images as real (return a higher output) and generated images as fake (return a lower output). We'll calculate two losses for the discriminator: one loss that compares Dx and 1 for real images from the MNIST set, as well as a loss that compares Dg

- Define the generator

and 0 for images from the generator. We'll do this with TensorFlow's

```
tf.nn.sigmoid_cross_entropy_with_logits()
```

function, which calculates the cross-entropy losses between `Dx` and 1 and between `Dg` and 0.

- Generate a sample image

`sigmoid_cross_entropy_with_logits` operates on unscaled values rather than probability values from 0 to 1. Take a look at the last line of our discriminator: there's no softmax or sigmoid layer at the end. GANs can fail if their discriminators "saturate," or become confident enough to return exactly 0 when they're given a generated image; that leaves the discriminator without a useful gradient to descend.

The `tf.reduce_mean()` function takes the mean value of all of the components in the matrix returned by the cross-entropy function. This is a way of reducing the loss to a single scalar value, instead of a vector or matrix.

- Training difficulties

- Other resources

- End

```
1 | d_loss_real = tf.reduce_mean
   | (tf.nn.sigmoid_cross_entropy_with_logits(Dx, tf.ones_like
   | (Dx)))
2 | d_loss_fake = tf.reduce_mean
   | (tf.nn.sigmoid_cross_entropy_with_logits(Dg, tf.zeros_like
   | (Dg)))
```

< Run >

Now let's set up the generator's loss function. We want the generator network to create images that will fool the discriminator: the generator wants the discriminator to output a value close to 1 when it's given an image from the generator. Therefore, we want to compute the loss between `Dg` and 1.

```
1 | g_loss = tf.reduce_mean
   | (tf.nn.sigmoid_cross_entropy_with_logits(Dg, tf.ones_like
   | (Dg)))
```

< Run >

Now that we have our loss functions, we need to define our optimizers. The optimizer for the generator network needs to only update the generator's weights, not those of the discriminator. Likewise, when we train the discriminator, we want to hold the generator's weights fixed.

In order to make this distinction, we need to create two lists of variables, one with the

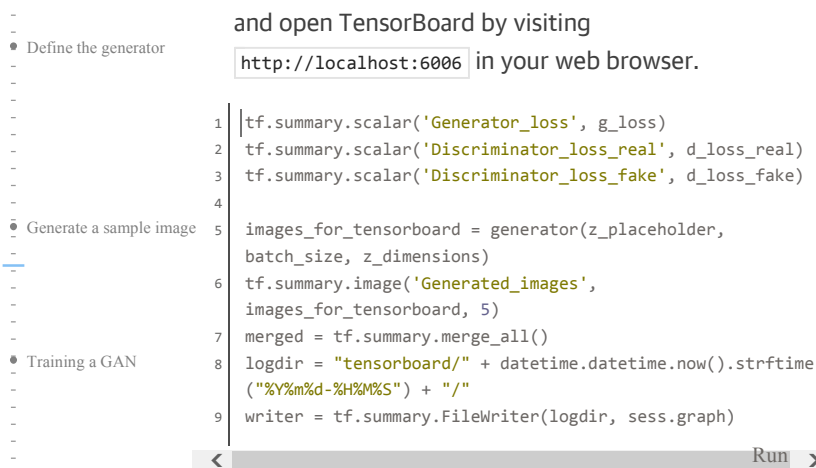
- Define the generator
- Generate a sample image
- Training a GAN
- Training difficulties
- Other resources
- End

```
1 | tvars = tf.trainable_variables()
2 |
3 | d_vars = [var for var in tvars if 'd_' in var.name]
4 | g_vars = [var for var in tvars if 'g_' in var.name]
5 |
6 | print([v.name for v in d_vars])
7 | print([v.name for v in g_vars])
```

We're setting up two different training operations for the discriminator here: one that trains the discriminator on real images and one that trains the discriminator on fake images. It's sometimes useful to use different learning rates for these two training operations, or to use them separately to regulate learning in other ways.

```
1 | # Train the discriminator
2 | d_trainer_fake = tf.train.AdamOptimizer(0.0003).minimize
  | (d_loss_fake, var_list=d_vars)
3 | d_trainer_real = tf.train.AdamOptimizer(0.0003).minimize
  | (d_loss_real, var_list=d_vars)
4 |
5 | # Train the generator
6 | g_trainer = tf.train.AdamOptimizer(0.0001).minimize
  | (g_loss, var_list=g_vars)
```

If you run this script on your own machine, include the cell below. Then, in a terminal window, run `tensorboard --logdir=tensorboard/`



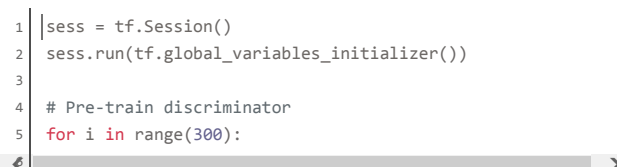
And now we iterate. We begin by briefly giving the discriminator some initial training; this helps it develop a gradient that's useful to the generator.

Then we move on to the main training loop. When we train the generator, we'll feed a random z vector into the generator and pass its output to the discriminator (this is the `Dg` variable we specified earlier). The generator's weights and biases will be updated in order to produce images that the discriminator is more likely to classify as real.

To train the discriminator, we'll feed it a batch of images from the MNIST set to serve as the positive examples, and then train the discriminator again on generated images, using them as negative examples. Remember that as the generator improves its output, the discriminator continues to learn to classify the improved generator images as fake.

Because it takes a long time to train a GAN, **we recommend not running this code block if you're going through this tutorial for the first time**. Instead, follow along but then run the following code block, which loads a pre-trained model for us to continue the tutorial.

If you want to run this code yourself, prepare to wait: it takes about 3 hours on a fast GPU, but could take ten times that long on a desktop CPU.



Run

11.12.2018

- Define the generator
- Generate a sample image
- Training a GAN

- Training difficulties
- Other resources
- End

- Training difficulties
-
-
-
- Other resources
-
- End

- Other resources
-
- End

- Define the generator

The field is still very young, and the next great GAN discovery could be yours!

Other resources

- Generate a sample image

- [The original GAN paper](#) by Ian Goodfellow and his collaborators, published in 2014
- [A more recent tutorial by Goodfellow](#) that explains GANs in somewhat more accessible terms
- [A paper by Alec Radford, Luke Metz, and Soumith Chintala](#) that introduces deep convolutional GANs, whose basic structure we use in our generator in this tutorial. Also see their DCGAN code on [GitHub](#).

This post is part of a collaboration between O'Reilly and [TensorFlow](#). See our [statement of editorial independence](#).

Article image: Generative Adversarial Networks for Beginners (source: O'Reilly).

- Training difficulties

- [Share](#) Other resources

Tweet

Share 295

Share

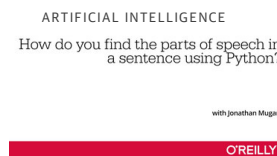
- End



The dynamic forces shaping AI

By Beau Cronin

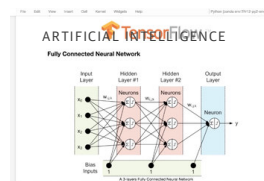
It's time to debate scenarios that will shift the balance of data, compute resources, algorithms, and talent.



How do you find the parts of speech in a sentence using Python?

By Jonathan Mugan

Learn how to use spaCy to parse a sentence to return the parts of speech (noun, verb, etc.) and dependencies.



How do I remove boilerplate code with TensorFlow-Slim's meta-operator?

By Marvin Bertin

Learn how to use TensorFlow-Slim's meta-operators to build deep learning models with a substantially reduced amount of code.



How can I create a neural network training routine with TensorFlow-Slim?

By Marvin Bertin

Learn how to create an automated training routine for any of your deep learning models.

ABOUT US

Our Company
Teach/Speak/Write
Careers
Customer Service
Contact Us

SITE MAP

Ideas
Learning
Topics
All

© 2018 O'Reilly Media, Inc. All trademarks and registered trademarks appearing on oreilly.com are the property of their respective owners.

[Terms of Service](#) • [Privacy Policy](#) • [Editorial Independence](#)

