

# How to recognize fake AI-generated images



Kyle McDonald Dec 5, 2018 · 7 min read



In 2014 machine learning researcher Ian Goodfellow introduced the idea of generative adversarial networks or GANs. “Generative” because they output things like images rather than predictions about input (like “hotdog or not”); “adversarial networks” because they use two neural networks competing with each other in a “cat-and-mouse game”, like a cashier and a counterfeiter: one trying to fool the other into thinking it can generate real examples, the other trying to distinguish real from fake.

The first GAN images were easy for humans to identify. Consider these faces from 2014.



# How to recognize fake AI-generated images



Kyle McDonald Dec 5, 2018 · 7 min read



In 2014 machine learning researcher Ian Goodfellow introduced the idea of generative adversarial networks or GANs. “Generative” because they output things like images rather than predictions about input (like “hotdog or not”); “adversarial networks” because they use two neural networks competing with each other in a “cat-and-mouse game”, like a cashier and a counterfeiter: one trying to fool the other into thinking it can generate real examples, the other trying to distinguish real from fake.

The first GAN images were easy for humans to identify. Consider these faces from 2014.

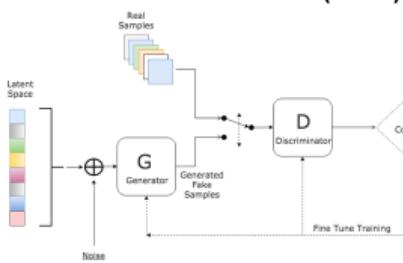


# 8 Machine Learning III

## - Specialized Areas in Machine Learning

### (4) Generative Adversarial Networks GAN

#### Generative Adversarial Networks (GAN)



A Generative Adversarial Network (GAN) is a class of machine learning frameworks designed by Goodfellow and his colleagues in 2014.

Two neural networks contest with each other in a game. Given a training set, this technique learns to generate new data with the same statistics as the training set.

The core idea of a GAN is based on the "indirect" training through the discriminator, which itself is also being updated dynamically. This basically means that the generator is not trained to minimize the distance to a specific image, but rather to fool the discriminator. This enables the model to learn in an unsupervised manner. (Wikipedia)

#### StyleGAN: Motivation Style Transfer



Intuitive Guide to Neural Style Transfer, 2019

[towardsdatascience.com/light-on-math-machine-learning-intuitive-guide-to-neural-style...](https://towardsdatascience.com/light-on-math-machine-learning-intuitive-guide-to-neural-style...)

# GAN Playground in Browser

## GAN Playground - Explore Generative Adversarial Nets in your Browser

### GAN Playground - Explore Generative Adversarial Nets in your Browser

#### DATA

|                               |                      |
|-------------------------------|----------------------|
| Dataset                       | MNIST                |
| Model - discriminator         | Convolutional (disc) |
| Model - generator             | Convolutional (gen)  |
| <b>Hyperparameters</b>        |                      |
| Learning Rate - discriminator | 0.08                 |
| Optimizer - discriminator     | sgd                  |
| Learning Rate - generator     | 0.08                 |
| Beta1 - generator             | 0.9                  |
| Beta2 - generator             | 0.999                |
| Optimizer - generator         | adam                 |
| Batch Size                    | 15                   |
| <b>TRAIN</b>                  | <b>STOP</b>          |
| Normalization                 | [-1, 1]              |
| <b>Statistics</b>             |                      |
| Examples                      | 65000                |
| Input shape                   | [28,28,1]            |
| Label shape                   | [10]                 |
| <b>TRAIN</b>                  | <b>STOP</b>          |
| Normalization                 | [-1, 1]              |
| <b>Statistics</b>             |                      |
| Examples                      | 65000                |
| Input shape                   | [28,28,1]            |
| Label shape                   | [10]                 |

#### DISCRIMINATOR

Input Image  
[28,28,1]

Op type  
Convolution

Field size  
5      Stride  
1      Zero pad  
2      Output dep...  
8

[28,28,8]

Op type  
ReLU

[28,28,8]

Op type  
Max pool

Field size  
2      Stride  
2      Zero pad  
0

[14,14,8]

Op type  
Convolution

Field size  
5      Stride  
1      Zero pad  
2      Output dep...  
16

[14,14,16]

#### GENERATOR

Generator Random Vector  
[100]

Op type  
Fully connected

Hidden units  
784

[784]

Op type  
ReLU

[784]

Op type  
Reshape

Shape (comma separated)  
28, 28, 1

[28,28,1]

Op type  
Convolution

Field size  
3      Stride  
1      Zero pad  
1      Output dep...  
10

[28,28,10]

#### REAL IMAGES

Inferences/sec:  
136

Inference duration: 7.07ms



1 84.1%  
0 15.9%



0 99.9%  
1 0.1%



1 99.9%  
0 0.1%



0 97.3%  
1 2.7%



1 100.0%  
0 0.0%



0 100.0%  
1 0.0%



1 93.5%  
0 6.5%



0 96.3%  
1 3.7%



0 94.5%  
1 5.5%



0 99.7%  
1 0.3%



1 100.0%  
0 0.0%



0 99.9%  
1 0.1%

#### GENERATED IMAGES

Generations/sec:  
136

Generation duration: 7.07ms



0 99.9%  
1 0.1%



0 97.3%  
1 2.7%



0 100.0%  
1 0.0%



0 96.3%  
1 3.7%



0 99.7%  
1 0.3%



0 99.9%  
1 0.1%

# 8 Machine Learning III

## - Specialized Areas in Machine Learning

### Content:

1. Transfer Learning & Teachable Machine
2. YOLO & Real-Time Object Detection
3. Autoencoder & Super Sampling
4. Generative Adversarial Networks (GAN)
5. Reinforcement Learning
6. The Human Role in Machine Learning **Live**
7. Summary



# Michael Amberg

## Todays Content:

1. Transfer Learning & Teachable Machine
2. YOLO & Real-Time Object Detection
3. Autoencoder & Super Sampling
4. Generative Adversarial Networks (GAN)
5. Reinforcement Learning
6. The Human Role in Machine Learning **Live**
7. Summary

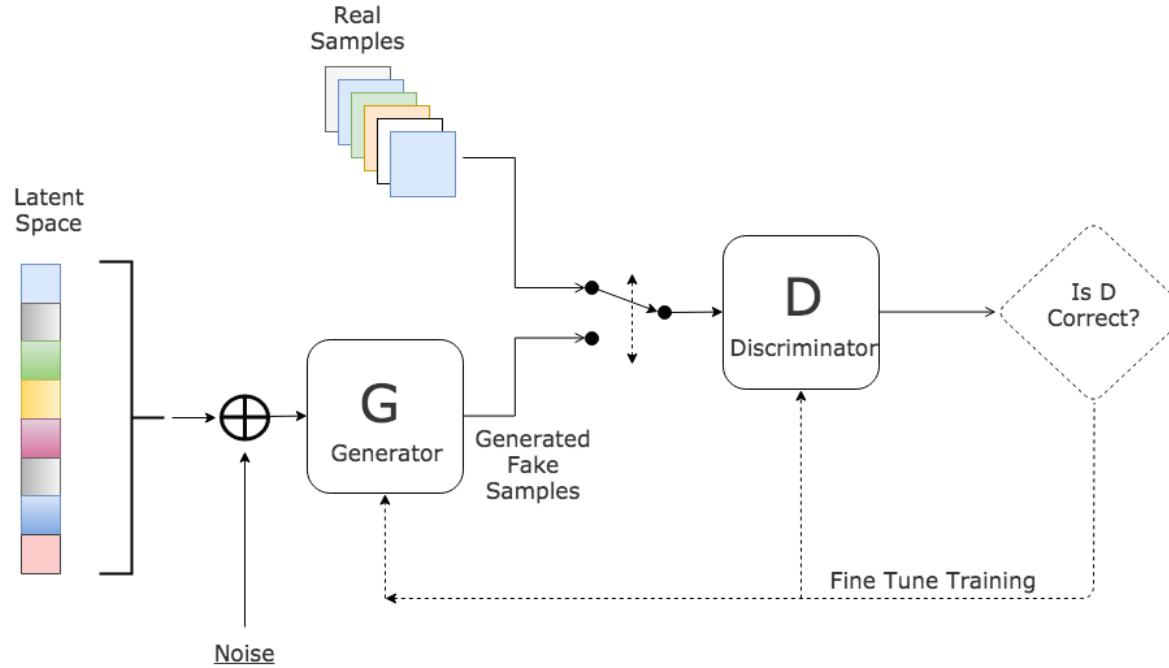


# 4. Generative Adversarial Networks (GAN)



**StyleGAN2 Interpolation Loop, 2019, 15min**  
[www.youtube.com/watch?v=6E1\\_dgYlfc](https://www.youtube.com/watch?v=6E1_dgYlfc)

# Generative Adversarial Networks (GAN)



A **Generative Adversarial Network (GAN)** is a class of **machine learning** frameworks designed by **Goodfellow** and his colleagues in **2014**.

**Two neural networks** contest with each other in a game. Given a **training set**, this technique learns to **generate new data** with the same statistics as the **training set**.

The **core idea** of a GAN is based on the "**indirect**" training through the **discriminator**, which itself is also being **updated dynamically**. This basically means that the **generator** is **not trained to minimize the distance to a specific image**, but rather to **fool the discriminator**. This enables the model to learn in an **unsupervised manner**. (Wikipedia)<sup>18</sup>

# Generative Adversarial Networks (GAN)



As training progresses, the generator gets closer to producing output that can fool the discriminator:



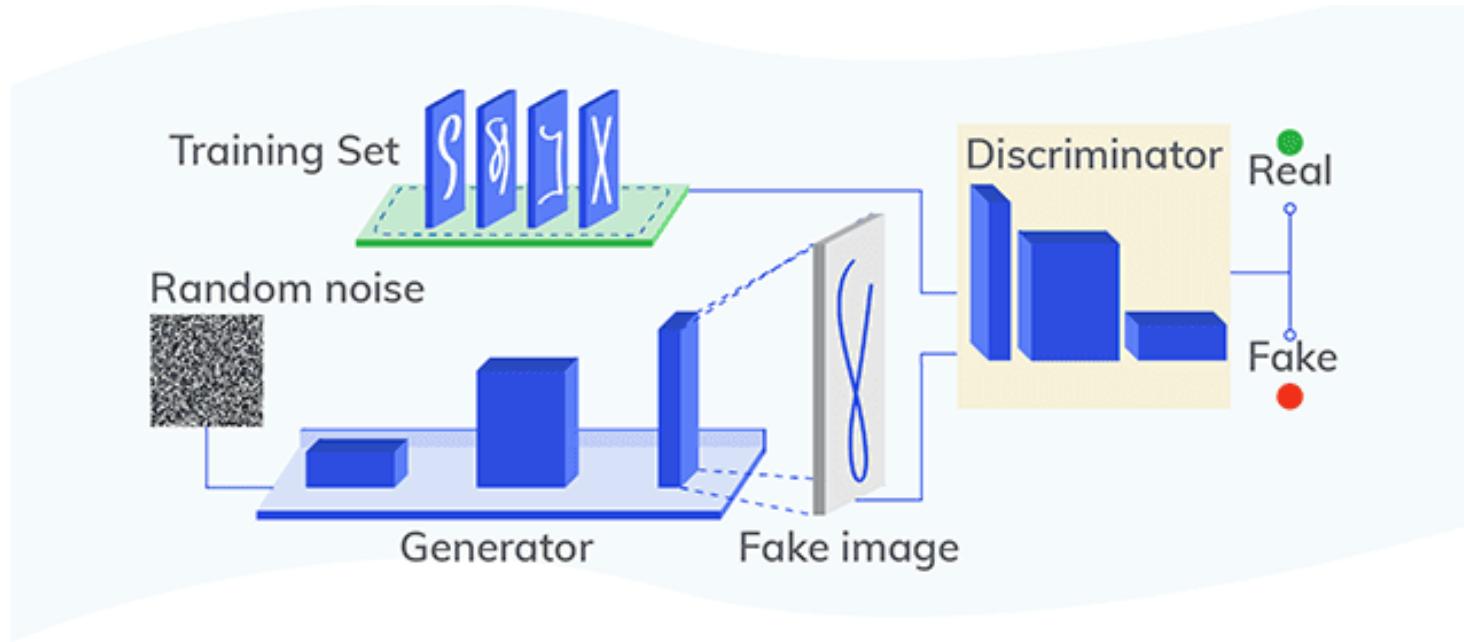
Finally, if generator training goes well, the discriminator gets worse at telling the difference between real and fake. It starts to classify fake data as real, and its accuracy decreases.



## Overview of GAN Structure

[developers.google.com/machine-learning/gan/gan\\_structure](https://developers.google.com/machine-learning/gan/gan_structure)

# Generative Adversarial Networks (GAN)



**GANs' potential for both good and evil is huge**, because they can **learn to mimic any distribution of data**. That is, **GANs can be taught to create worlds eerily similar to our own in any domain: images, videos, music, speech**.

They are **robot artists** in a sense, and their **output is impressive**.  
But they can **also be used to generate fake media content**,  
and are the technology underpinning **Deep Fakes**.

# GAN Playground in Browser

## GAN Playground - Explore Generative Adversarial Nets in your Browser

### GAN Playground - Explore Generative Adversarial Nets in your Browser

#### DATA

|                               |                      |
|-------------------------------|----------------------|
| Dataset                       | MNIST                |
| Model - discriminator         | Convolutional (disc) |
| Model - generator             | Convolutional (gen)  |
| <b>Hyperparameters</b>        |                      |
| Learning Rate - discriminator | 0.08                 |
| Optimizer - discriminator     | sgd                  |
| Learning Rate - generator     | 0.08                 |
| Beta1 - generator             | 0.9                  |
| Beta2 - generator             | 0.999                |
| Optimizer - generator         | adam                 |
| Batch Size                    | 15                   |
| <b>TRAIN</b>                  | <b>STOP</b>          |
| Normalization                 | [-1, 1]              |
| <b>Statistics</b>             |                      |
| Examples                      | 65000                |
| Input shape                   | [28,28,1]            |
| Label shape                   | [10]                 |
| <b>TRAIN</b>                  | <b>STOP</b>          |
| Normalization                 | [-1, 1]              |
| <b>Statistics</b>             |                      |
| Examples                      | 65000                |
| Input shape                   | [28,28,1]            |
| Label shape                   | [10]                 |

#### DISCRIMINATOR

Input Image  
[28,28,1]

Op type  
Convolution

Field size  
5      Stride  
1      Zero pad  
2      Output dep...  
8

[28,28,8]

Op type  
ReLU

[28,28,8]

Op type  
Max pool

Field size  
2      Stride  
2      Zero pad  
0

[14,14,8]

Op type  
Convolution

Field size  
5      Stride  
1      Zero pad  
2      Output dep...  
16

[14,14,16]

#### GENERATOR

Generator Random Vector  
[100]

Op type  
Fully connected

Hidden units  
784

[784]

Op type  
ReLU

[784]

Op type  
Reshape

Shape (comma separated)  
28, 28, 1

[28,28,1]

Op type  
Convolution

Field size  
3      Stride  
1      Zero pad  
1      Output dep...  
10

[28,28,10]

#### REAL IMAGES

Inferences/sec:  
136

Inference duration: 7.07ms



1 84.1%  
0 15.9%



0 99.9%  
1 0.1%



1 99.9%  
0 0.1%



0 97.3%  
1 2.7%



1 100.0%  
0 0.0%



0 100.0%  
1 0.0%



1 93.5%  
0 6.5%



0 96.3%  
1 3.7%



0 94.5%  
1 5.5%



0 99.7%  
1 0.3%



1 100.0%  
0 0.0%



0 99.9%  
1 0.1%

#### GENERATED IMAGES

Generations/sec:  
136

Generation duration: 7.07ms



0 99.9%  
1 0.1%



0 97.3%  
1 2.7%



0 100.0%  
1 0.0%



0 96.3%  
1 3.7%



0 99.7%  
1 0.3%



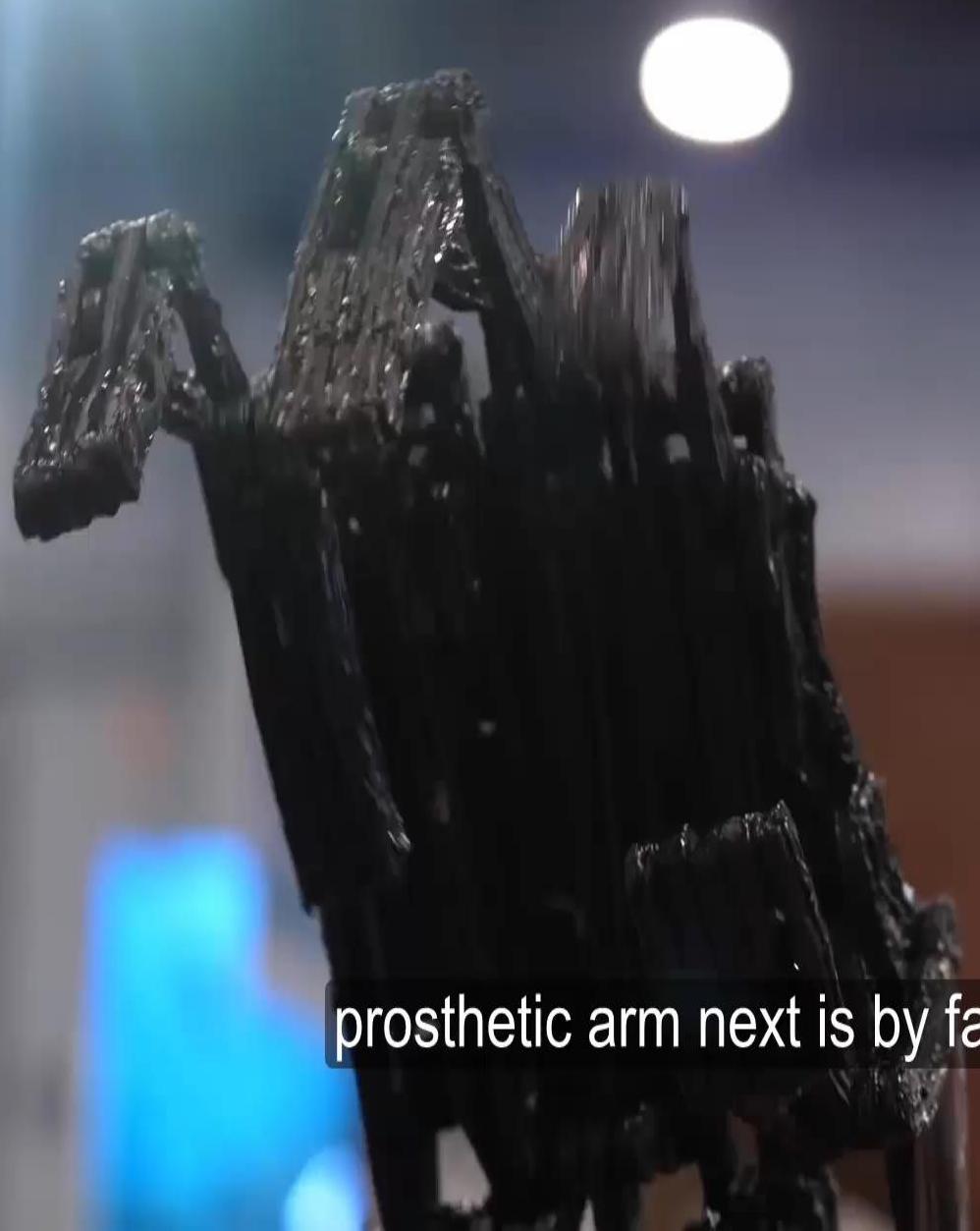
0 99.9%  
1 0.1%

# StyleGAN2: Playground ThisPersonDoesNotExist



# 7 Craziest Tech Products Of CES 2020!

## - Air Taxi



prosthetic arm next is by far the