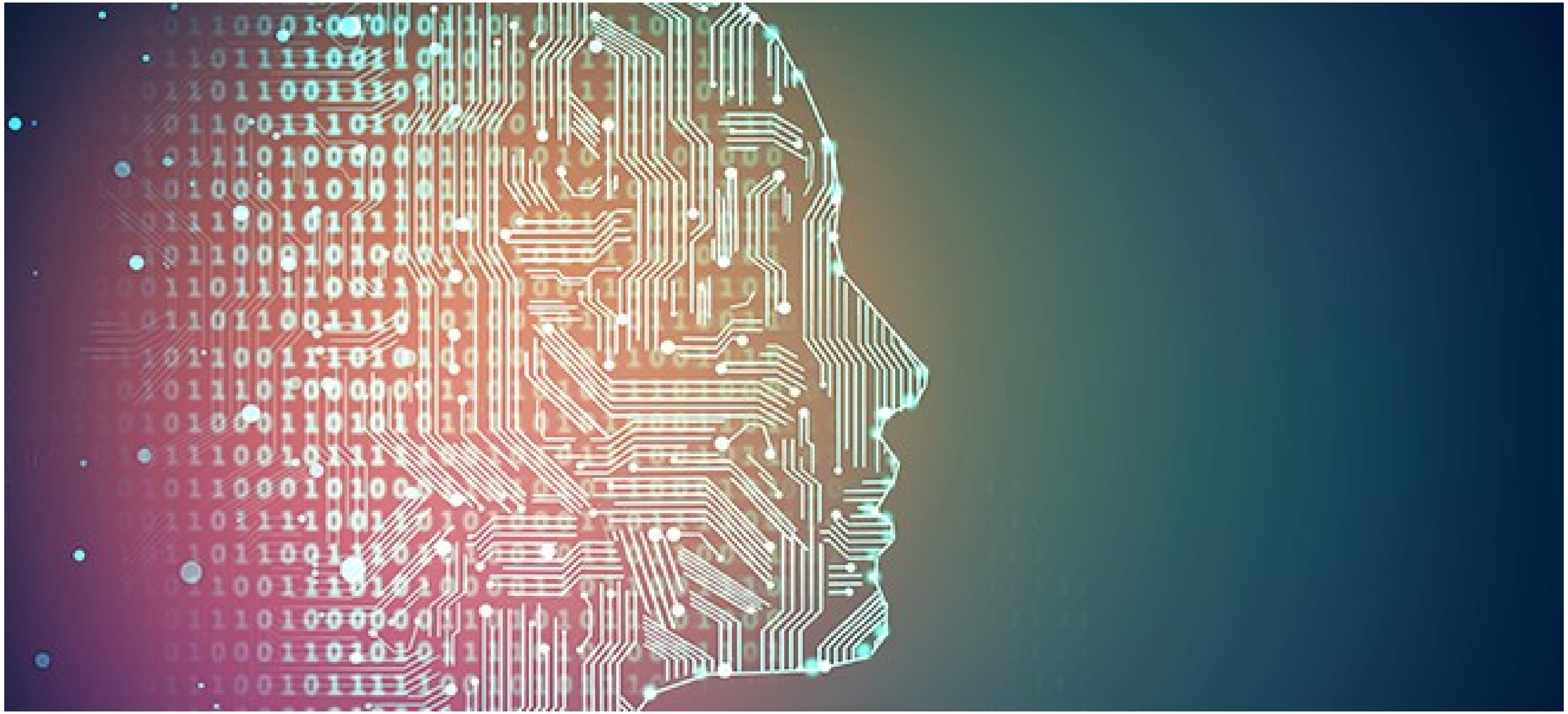
AI Deep Learning: Generative Adversarial Networks

Thuan L Nguyen, PhD

AI Deep Learning: Convolutional Neural Networks (CNN)



AI Deep Learning: Generative Adversarial Networks (GAN)

GAN: Overview

- Generative adversarial networks (GANs) are deep neural network architectures comprised of two networks, pitting one against the other (thus the “adversarial”).

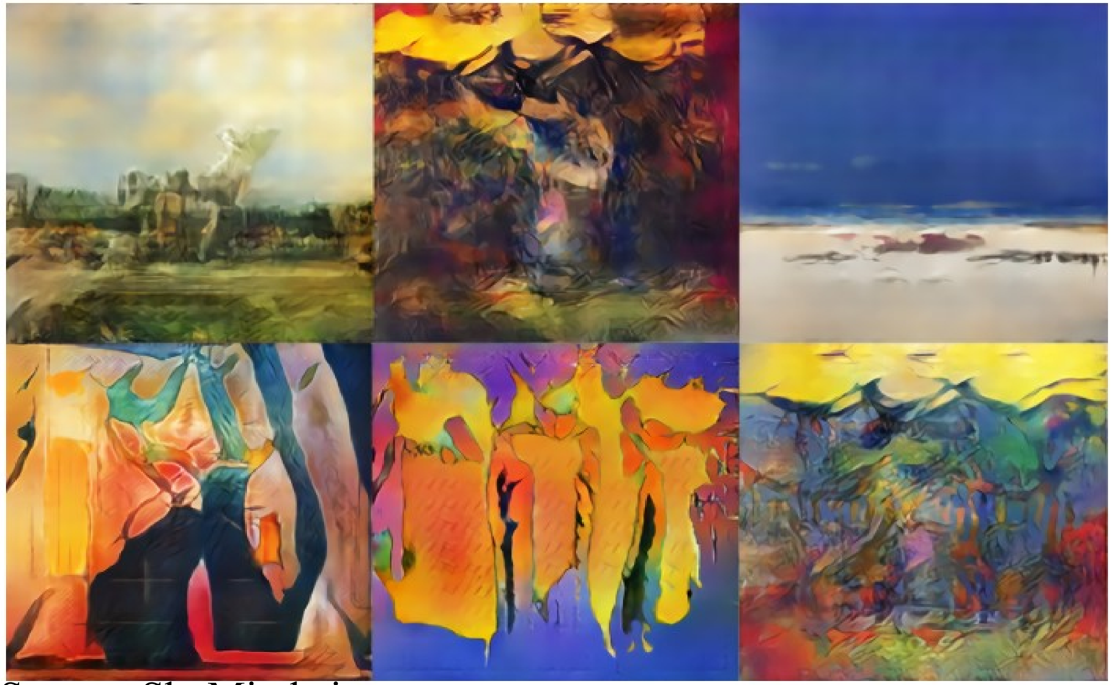
- Ian Goodfellow and other researchers at the University of Montreal proposed GANs in 2014.
  - Referring to GANs, Facebook’s AI research director Yann LeCun called adversarial training “the most interesting idea in the last 10 years in ML.”

- GANs’ potential is huge:
  - GANs can learn to mimic any distribution of data.
  - GANs can be taught to create worlds eerily similar to human ones in any domain: images, music, speech, prose, and even arts.

AI Deep Learning: Generative Adversarial Networks (GAN)

GAN: Overview

- GANs are robot artists:
  - Christie's sold a portrait for $432,000 that had been generated by a GAN, based on open-source code written by Robbie Barrat of Stanford.



Source: SkyMind.ai

AI Deep Learning: Generative Adversarial Networks (GAN)

GAN: Generative vs Discriminative Algorithms

- Discriminative algorithms:
  - Try to **classify** input data
    - Given the features of a data instance
    - They predict a label or category to which that data belongs.
    - In other words, Discriminative algorithms map features to labels.

- For example:
  - Given all the words in an email
  - A discriminative algorithm could predict whether the message is spam or not spam.
    - Spam is one of the labels
    - The bag of words gathered from the email are the features that constitute the input data.
    - This problem can be expressed mathematically:
      - The label is called y and the features are called x.
      - The formulation p(y|x) is used to mean "the probability of y given x"
        - i.e., "the probability that an email is spam given the words it contains."

AI Deep Learning: Generative Adversarial Networks (GAN)

GAN: Generative vs Discriminative Algorithms

- Generative algorithms
  - do the **opposite** to what Discriminative algorithms do:
  - They attempt to predict features **given** a certain label.
    - Instead of predicting a label given certain features like discriminative ones

- About the example of the email spam,
  - The question a generative algorithm tries to answer is:
    - Assuming this email is spam, how likely are these features?

- Generative models care about “how you get x.”
  - While discriminative models care about the relation between y and x

- Generative algorithms allow you to capture p(x|y):
  - The probability of x given y, or the probability of features given a class.

IMPORTANT NOTES:

- Generative algorithms can also be used as classifiers.
- It just so happens that they can do more than categorize input data.

AI Deep Learning: Generative Adversarial Networks (GAN)

GAN: Generative vs Discriminative Algorithms

- Another way to distinguish discriminative from generative:
  - Discriminative models learn the boundary between classes
  - Generative models model the distribution of individual classes

Source: Alec Radford’s research paper (https://arxiv.org/abs/1511.06434)

AI Deep Learning: Generative Adversarial Networks (GAN)

GAN: Generative Network vs Discriminative Network

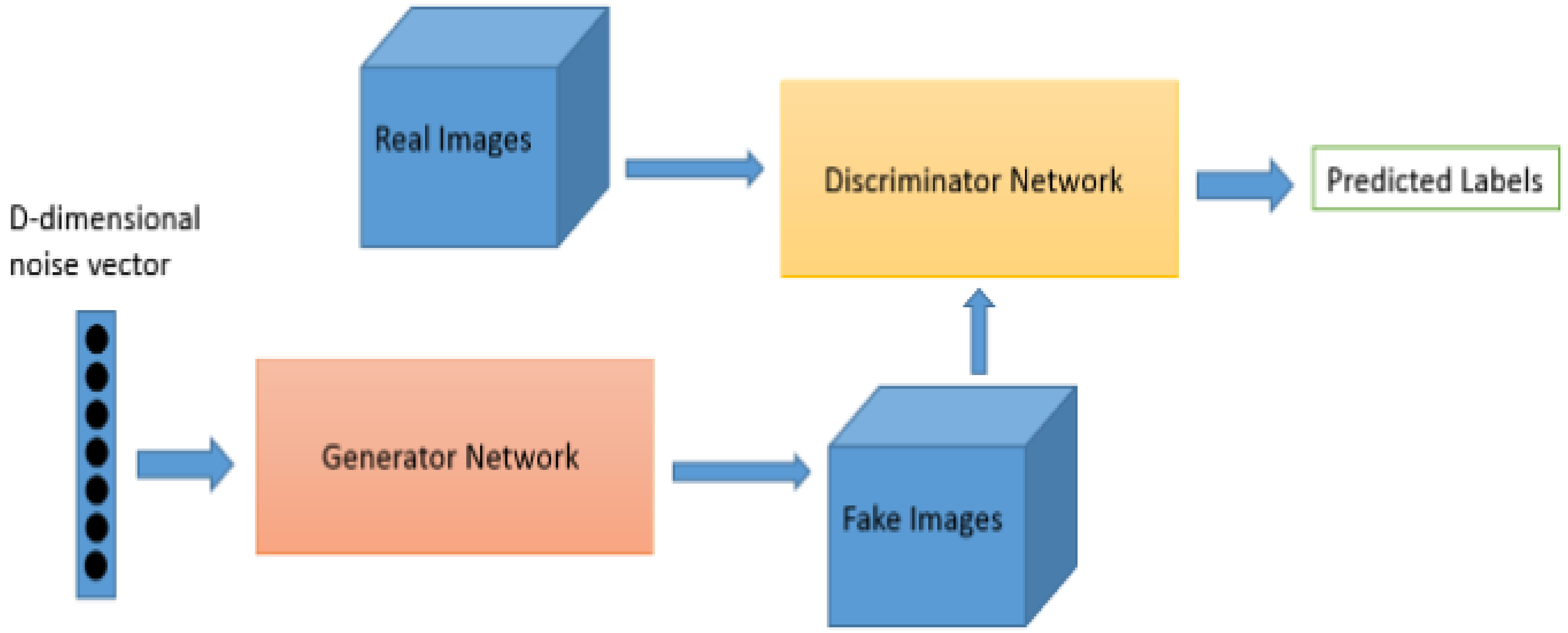
- Generative adversarial network: A combination of two networks
  - One neural network, called the **Generator**, generates new data instances,
  - The other, the **Discriminator**, evaluates them for authenticity.

- **GAN** can be viewed as a cat-and-mouse game between a counterfeiter and a cop.
  - The counterfeiter (**Generator**) is learning to create fake money.
  - The cop (**Discriminator**) is learning to detect the fake money.
  - Both of them are learning and improving.
    - The counterfeiter is constantly learning to create better fakes.
    - The cop is constantly getting better at detecting them.
  - The **end result** will be that the counterfeiter (**Generator**) is now trained to create ultra-realistic money!

AI Deep Learning: Generative Adversarial Networks (GAN)

GAN: How Does It Work?

- A diagram to illustrate how generative adversarial networks work.



Source: O’Reilly

AI Deep Learning: Generative Adversarial Networks (GAN)

GAN: A Network of Two Neural Networks of Opposite Goals

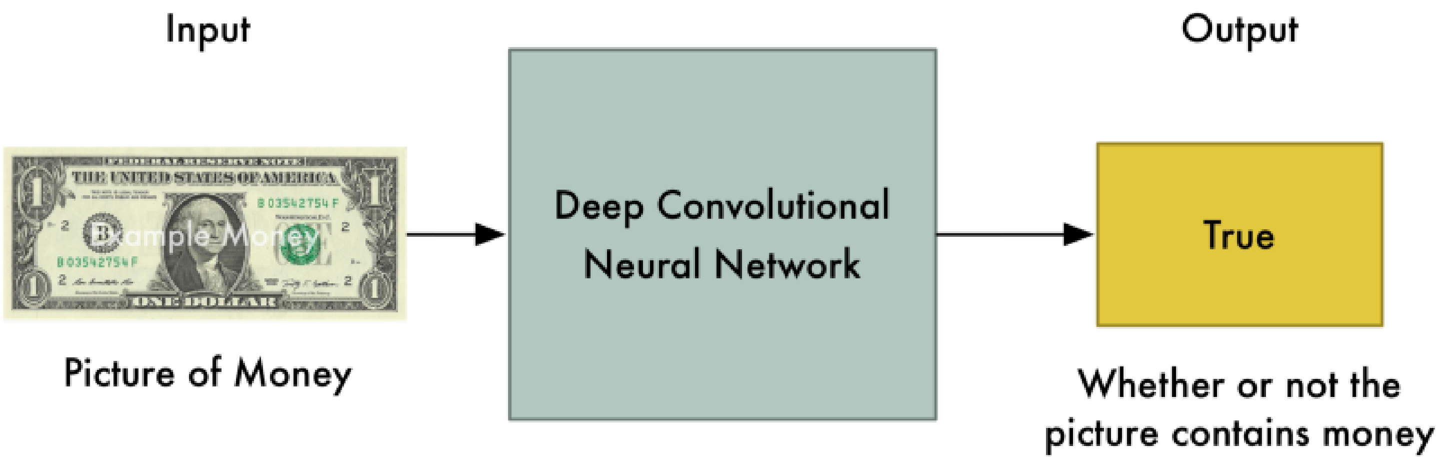
- GAN: A network of two neural networks:
  - The **discriminator network** is a standard convolutional neural network
    - That can categorize the images fed to it, a binomial classifier labeling images as real or fake.
  - The **generator network** is an inverse convolutional network, in a sense:
    - While a standard convolutional classifier takes an image and down-samples it to produce a probability, the generator takes a vector of random noise and **up-samples** it to an image.
    - The first throws away data through down-sampling techniques like max-pooling, and the second generates new data.

- Both networks are trying to **optimize** a different and opposing objective function, or **loss function**, in a zero-sum game.
  - This is essentially an actor-critic model. As the discriminator changes its behavior, so does the generator, and vice versa. Their losses push against each other.

AI Deep Learning: Generative Adversarial Networks (GAN)

GAN: A Network of Two Neural Networks of Opposite Goals

- This first neural network is called the **Discriminator**:
  - It is a standard convolutional network
  - It can categorize the images fed to it, a binomial classifier labeling images as real or fake.

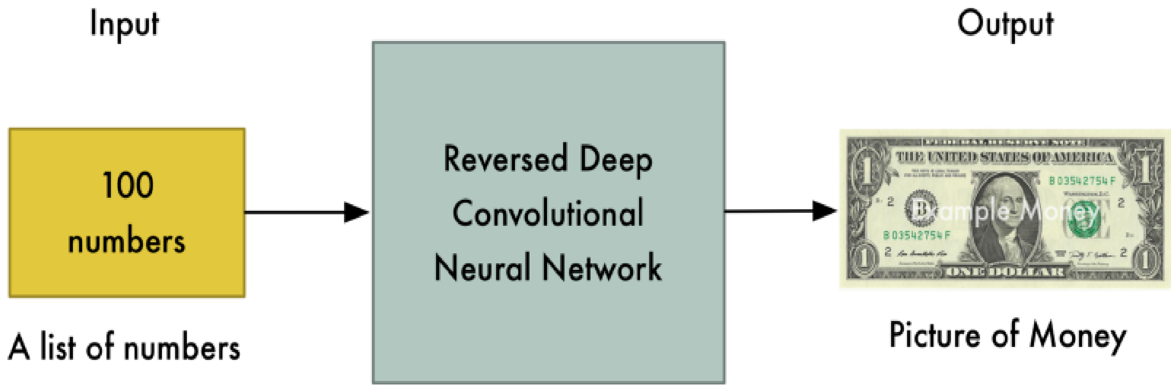

Input
Example Money
Picture of Money
Deep Convolutional Neural Network
Output
True
Whether or not the picture contains money

Source: Adam Geitgey

AI Deep Learning: Generative Adversarial Networks (GAN)

GAN: A Network of Two Neural Networks of Opposite Goals

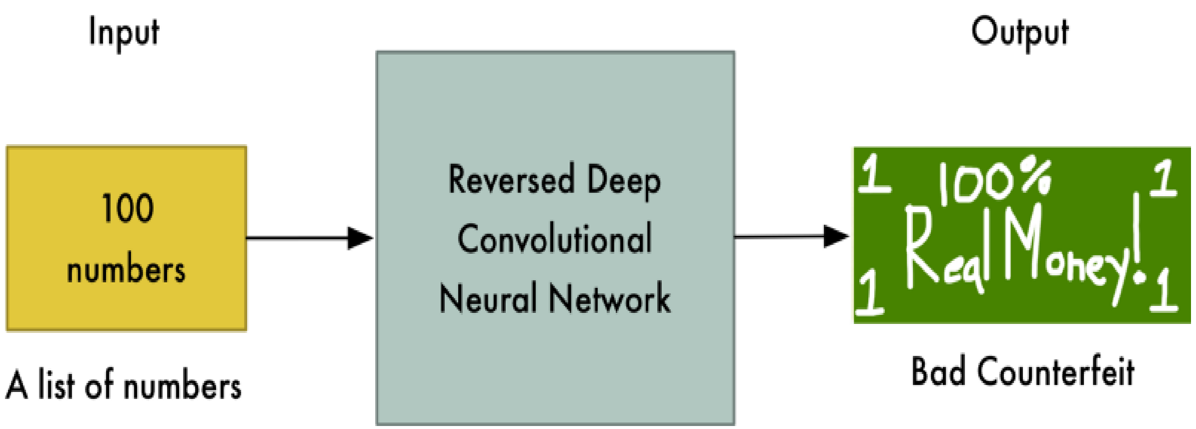
- This second neural network is called the Generator:
    - Let’s imagine that the second neural network is a brand new counterfeiter:
        - Who is just learning how to create fake money.
    - For this second neural network, let’s reverse the layers in a normal convolutional neural network so that everything runs backwards.
    - It takes in a list of values and outputs a picture
        - Instead of taking in a picture and outputting a value.



Source: Adam Geitgey

AI Deep Learning: Generative Adversarial Networks (GAN)

GAN: A Network of Two Neural Networks of Opposite Goals



Source: Adam Geitgey

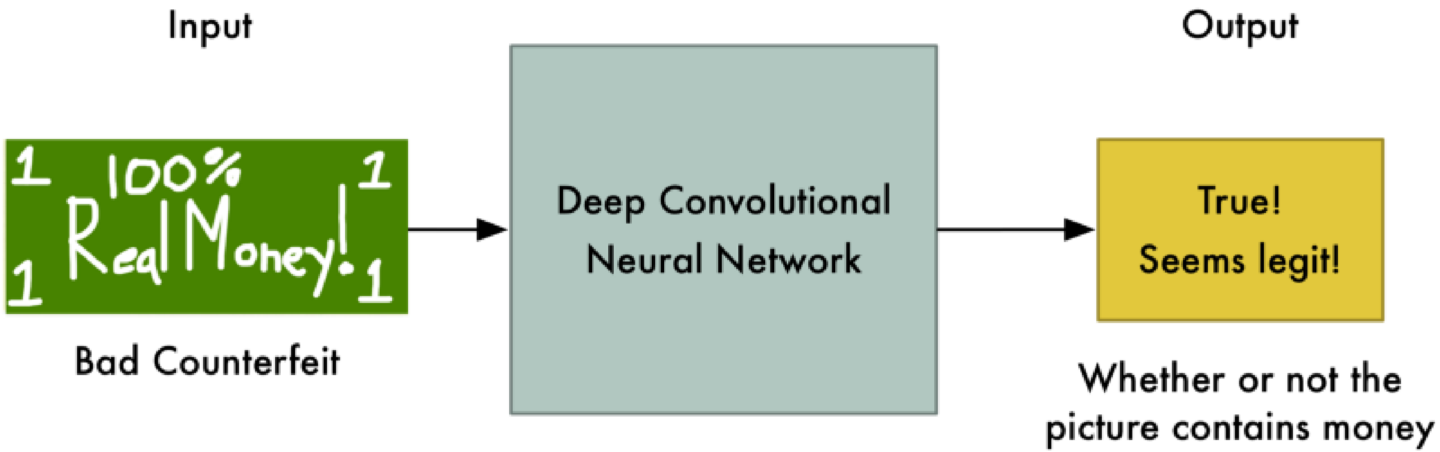
- The **battle**:
  - A police officer (the Discriminator) is looking for fake money.
  - A counterfeiter (the Generator) is printing fake money.
- Let's them battle!
- In the first round:
  - The Generator will create pathetic forgeries that barely resemble money at all because it knows absolutely nothing about what money is supposed to look like in the figure to the left.

AI Deep Learning: Generative Adversarial Networks (GAN)

GAN: A Network of Two Neural Networks of Opposite Goals

- At the start, as a learner, the Discriminator is not doing good jobs at all to detect whether money is real or not

- It won't know the difference:

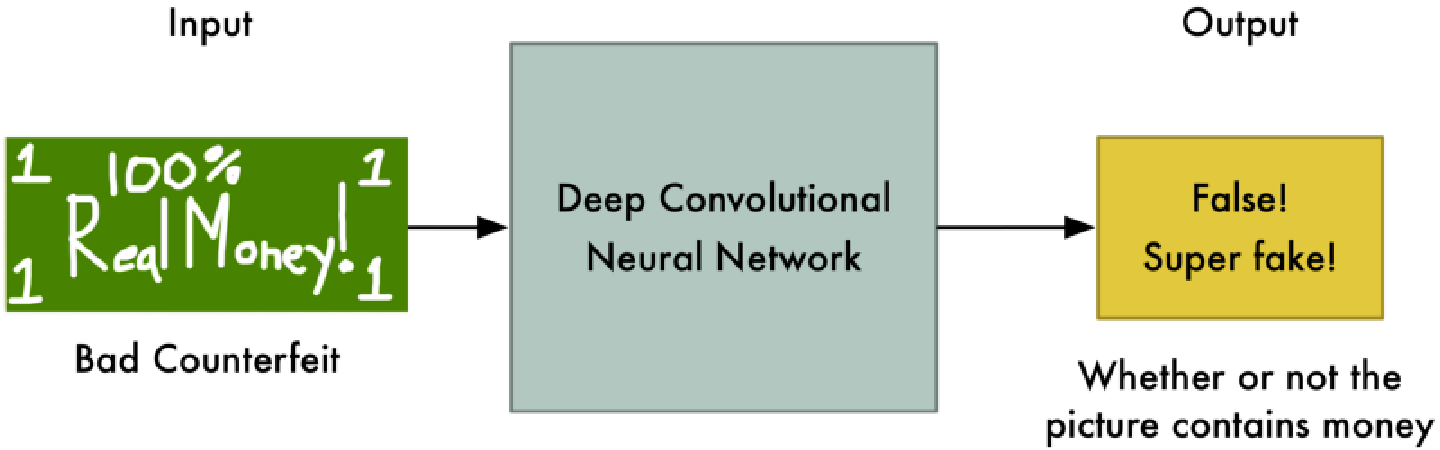


Source: Adam Geitgey

AI Deep Learning: Generative Adversarial Networks (GAN)

GAN: A Network of Two Neural Networks of Opposite Goals

- At this moment, we need to step in and tell the Discriminator that this dollar bill is actually fake.
- Then we show it a real dollar bill and ask it how it looks different from the fake one.
- The Discriminator looks for a new detail to help it separate the real one from the fake one.
  - For example, the Discriminator might notice that real money has a picture of a person on it and the fake money doesn't.
- Using this knowledge, the Discriminator learns how to tell the fake from the real one.
  - It gets better at its job.

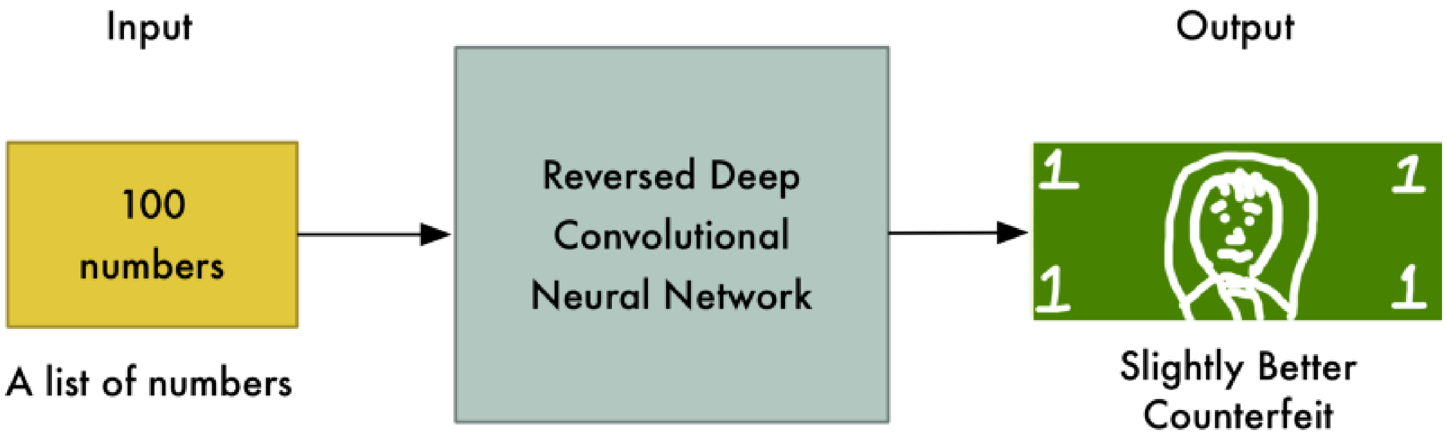

Input
1 100% 1
1 Real Money! 1
Bad Counterfeit
Deep Convolutional Neural Network
Output
False! Super fake!
Whether or not the picture contains money

Source: Adam Geitgey

AI Deep Learning: Generative Adversarial Networks (GAN)

GAN: A Network of Two Neural Networks of Opposite Goals

- Now comes Round 2:
  - We tell the Generator that it’s money images are rejected as fake.
  - So it needs to step up it’s game.
  - We also tell the Generator that:
    - The Discriminator is now looking for faces.
    - So the best way to confuse the Discriminator is to put a face on the bill:


Input
100 numbers
A list of numbers
Reversed Deep Convolutional Neural Network
Output
1
1
1
1
Slightly Better Counterfeit

Source: Adam Geitgey

AI Deep Learning: Generative Adversarial Networks (GAN)

GAN: A Network of Two Neural Networks of Opposite Goals

- Now comes Round 2 (cont.):
  - The new fake bills are being accepted as valid again!
  - So now the Discriminator has to look again at the real dollar and find a new way to tell it apart from the fake one.

- And the battle continues.

- This back-and-forth game between the Generator and the Discriminator continues thousands of times until both networks are experts.
  - Eventually the Generator is producing near-perfect counterfeits and the Discriminator has turned into a Master Detective looking for the slightest mistakes.

- When both networks are sufficiently trained so that humans are impressed by the results produced by both of them, the networks can be used for whatever purpose humans want.

AI Deep Learning: Generative Adversarial Networks (GAN)

GAN: In Summary

- Two neural networks working hand-in-hand:
  - The generator generates new data instances.
  - The discriminator evaluates them for authenticity.
    - i.e. the discriminator decides whether each instance of data it reviews belongs to the actual training dataset or not.

- When both networks are sufficiently trained so that humans are impressed by the results produced by both of them, the networks can be used for whatever purpose humans want.