Building Real Applications

With Deep Learning

Plan

What is Deep Learning?

Who is using it?

How do we use it?

Deep Learning is in Everything!

Building Intuition

Given a Problem

How to build a deep learning model to solve the problem?

Problem

What does it mean when someone gives you a problem?

Problem

Data (Features + Label)

Comedy?	Length (m)	Female Lead?	Super Hero?	Like?
Yes	90	No	No	No
No	120	Yes	Yes	Yes
...	...	...	...	...

Goal (Given features, predict label)

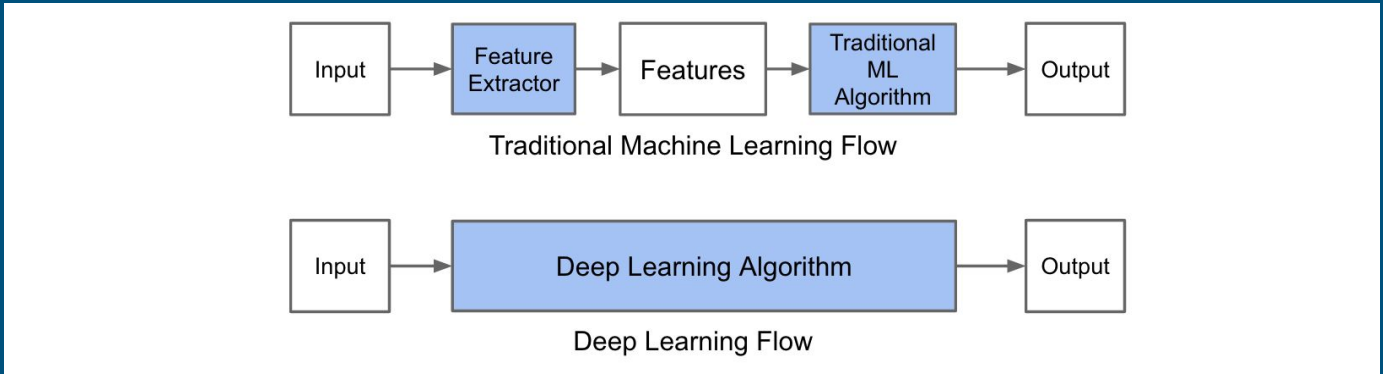
Deep Learning vs. Other Machine Learning

Deep Learning

Learn a representation for the data

Decision Trees, KNN, etc.

Implement an algorithm, given features



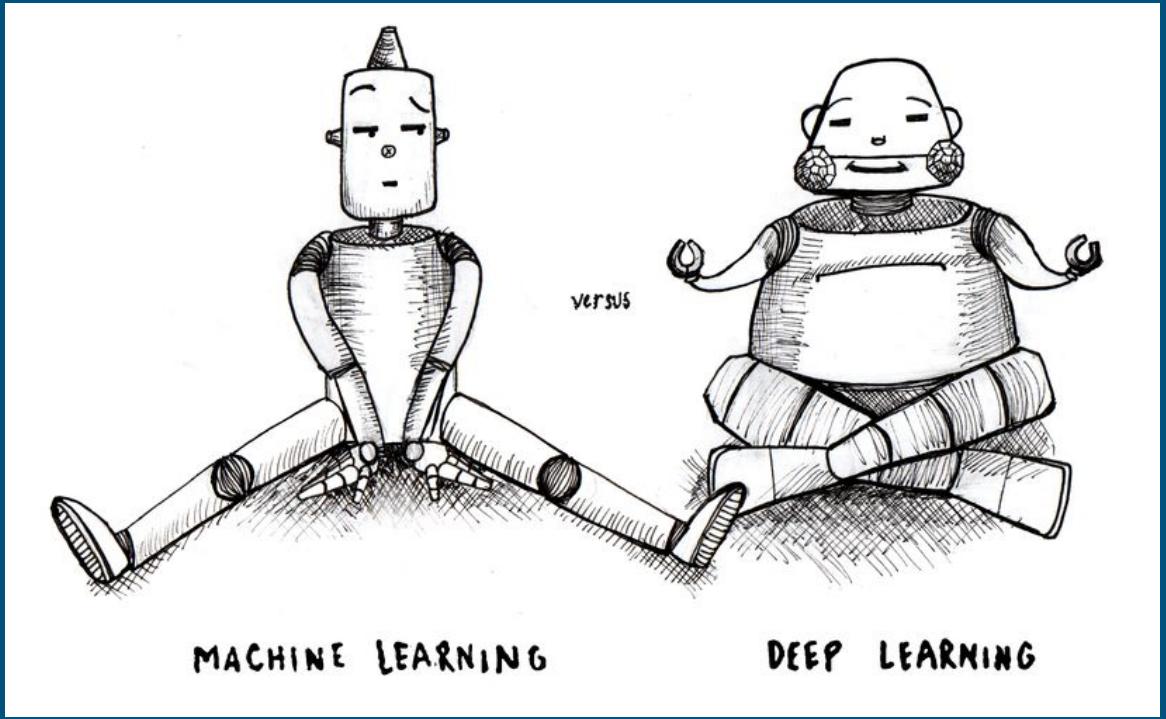
Why Deep Learning

Powerful

Feature Learning

Software availability

Sounds good



Plan

What is Deep Learning?

Who is using it?

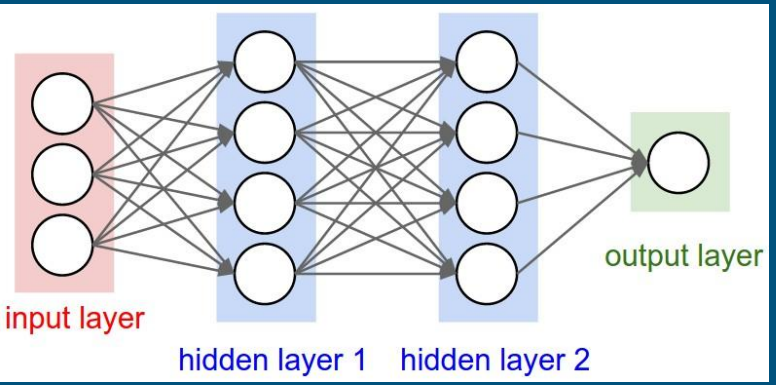
How do we use it?

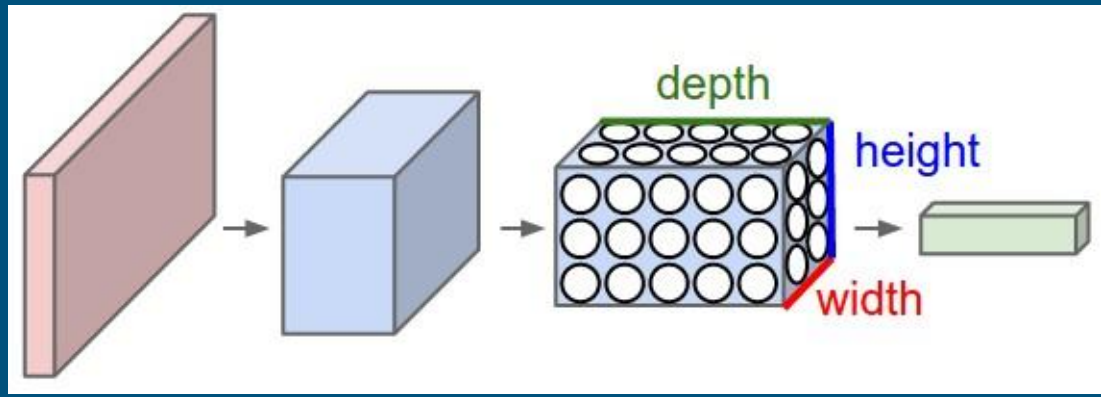
- **Architecture (Layer Types)**
- Loss Functions
- Hyper Parameters
- Challenges
- Tricks

Convolutional Neural Networks (CNN/ConvNet)

Architecture

- Number of layers
- Number of neurons in each layer





Layer Types

Convolutional

Fully Connected

Pooling

Dropout

Activation

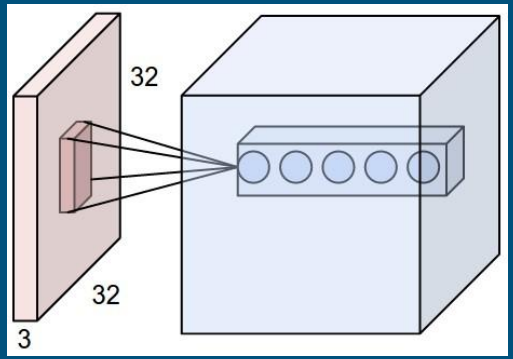
Layer Types: Convolutional

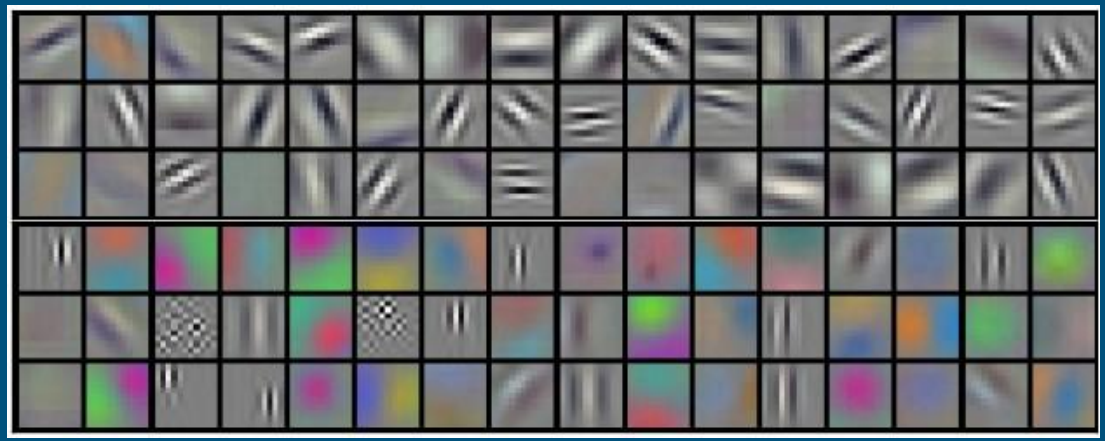
Filters

Size Filters

Stride

Padding





Layer Types: Pooling

In practice, most use Max Pooling.

$$f(x,y) = \max(x,y)$$

$$\delta f/\delta x = 1 \ (x >= y)$$

$$\delta f/\delta y = 1 \ (y >= x)$$

Layer Types: Pooling

$$f(x,y) = \max(x,y)$$

$$\delta f/\delta x = 1 \ (x >= y)$$

$$\delta f/\delta y = 1 \ (y >= x)$$

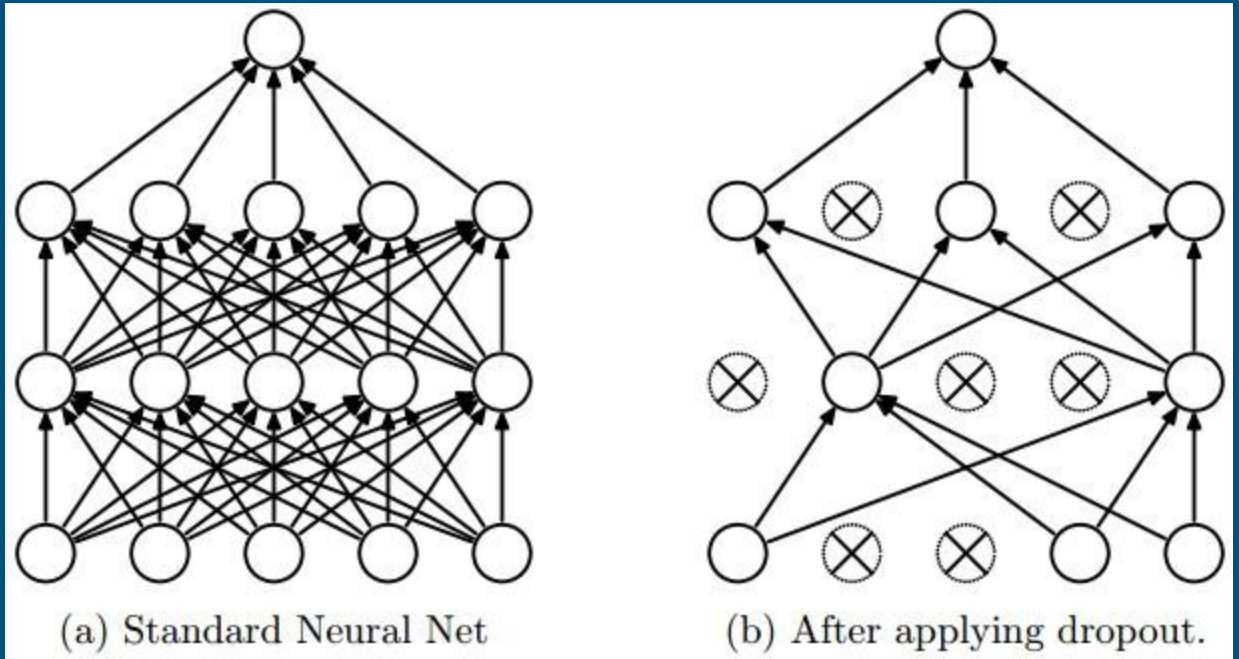
Intuition: x=4, y=2

f(x,y) = 4, and the function is not sensitive to the setting of y.

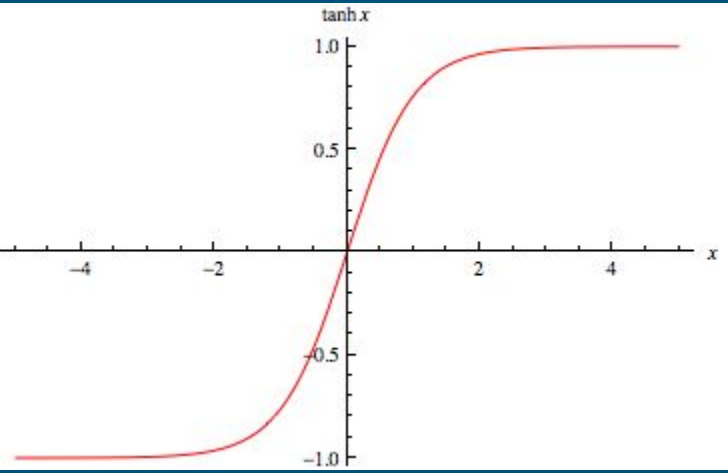
Increase y by small amount h, f(x,y+h) = 4, so there is no effect

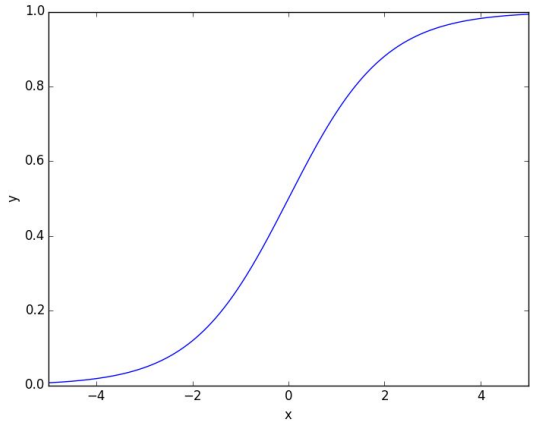
Layer Types: Dropout

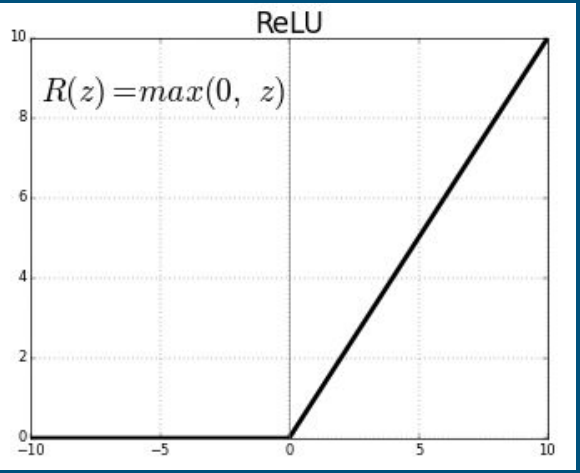
Randomly turn off activations in a layer



Layer Types: Activation



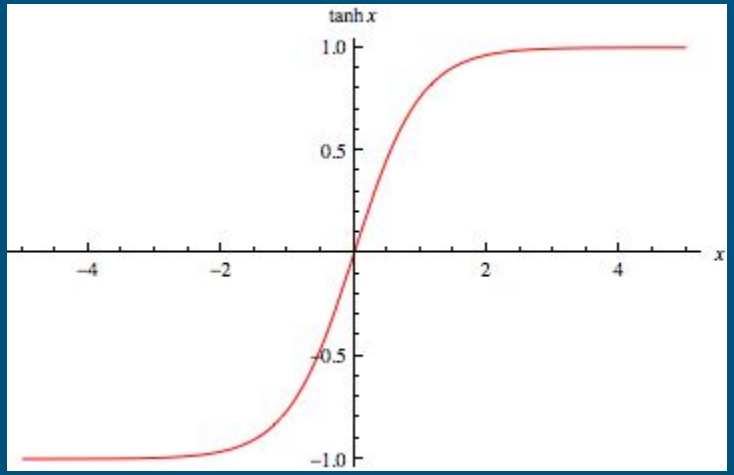




Vanishing/Exploding Gradient

Activation Function: Large input regions mapped to extremely small range.

Large change in input produces small change in output → small gradient



Plan

What is Deep Learning?

Who is using it?

How do we use it?

- Architecture (Layer Types)
- **Loss Functions**
- Hyper Parameters
- Challenges
- Tricks

Loss Function

Common Choices:

- Softmax
- Sigmoid Cross Entropy
- Euclidean

Plan

What is Deep Learning?

Who is using it?

How do we use it?

- Architecture (Layer Types)
- Loss Functions
- **Hyper Parameters**
- Challenges
- Tricks

Hyper Parameters

Batch Size

Updating network parameters with each batch

Learning Rate

Step Size

When to reduce learning rate

Momentum

Add fraction of previous weight update to current update

Weight Decay

Reduce the value of weights

Plan

What is Deep Learning?

Who is using it?

How do we use it?

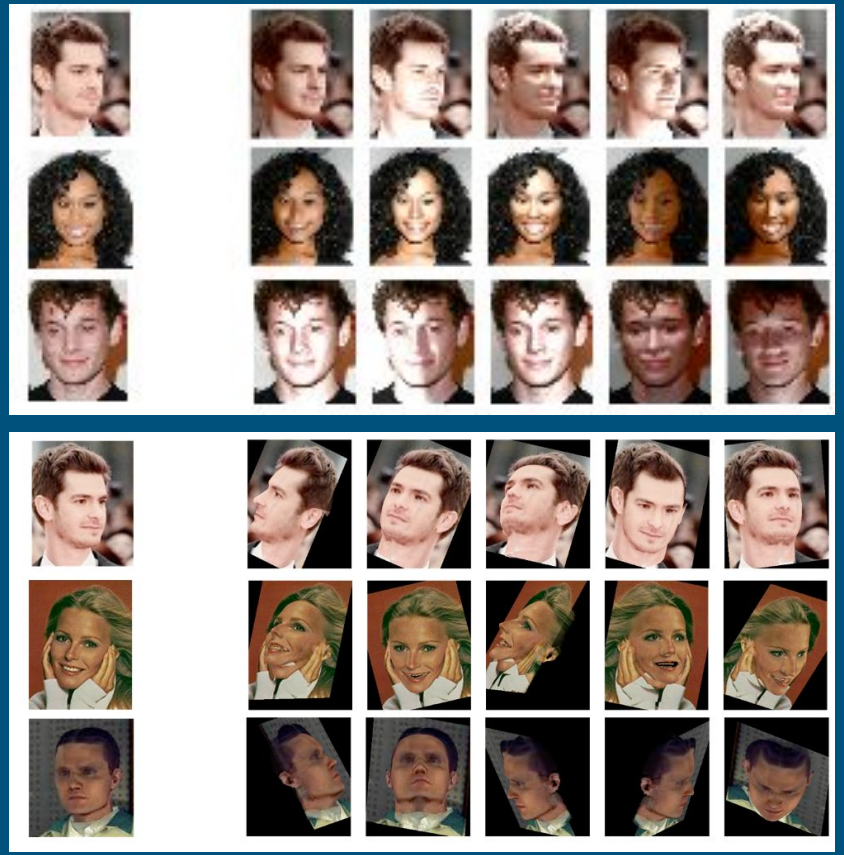
- Architecture (Layer Types)
- Loss Functions
- Hyper Parameters
- **Challenges**
- Tricks

Overfitting

HUGE Problem

What can we do to avoid it?

- Early Stopping
- Reduce # Parameters
- Dropout
- Data Augmentation



Training Time/Memory

GPUs have limited memory (6GB)

Training can take hours, days, or weeks!

Biggest factors in deciding your architecture, and hyper parameters

Plan

What is Deep Learning?

Who is using it?

How do we use it?

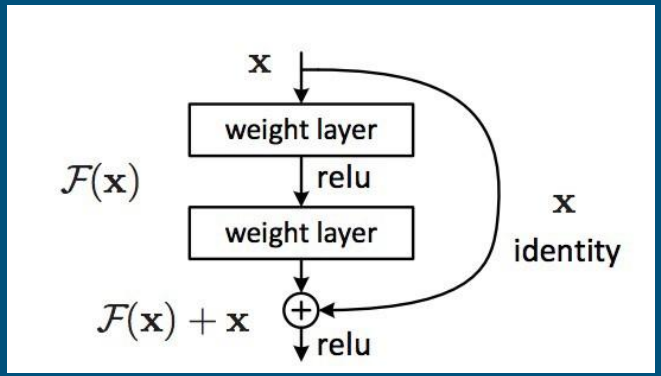
- Architecture (Layer Types)
- Loss Functions
- Hyper Parameters
- Challenges
- **Tricks**

Common Architectures

Alexnet (8 layers) 60M parameters

VGG-Net (19 layers) 138M parameters

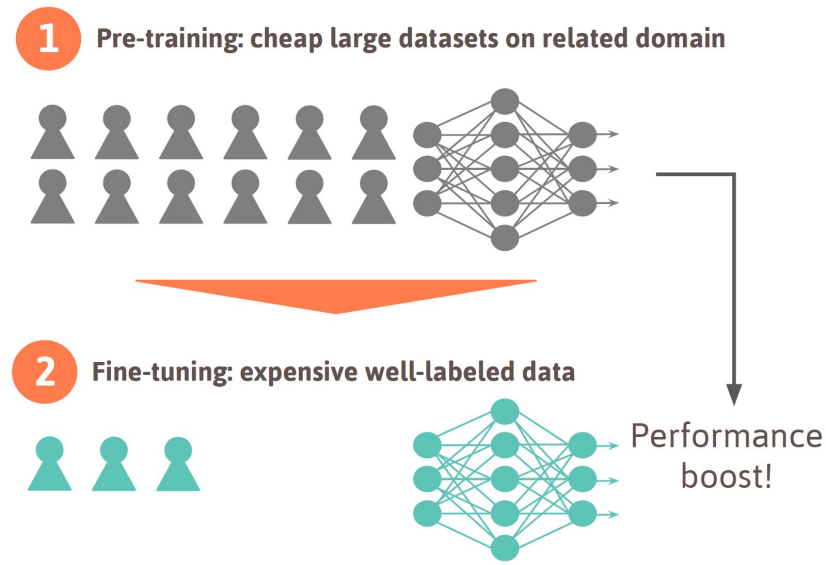
Res-Net (50+ layers) 25M parameters



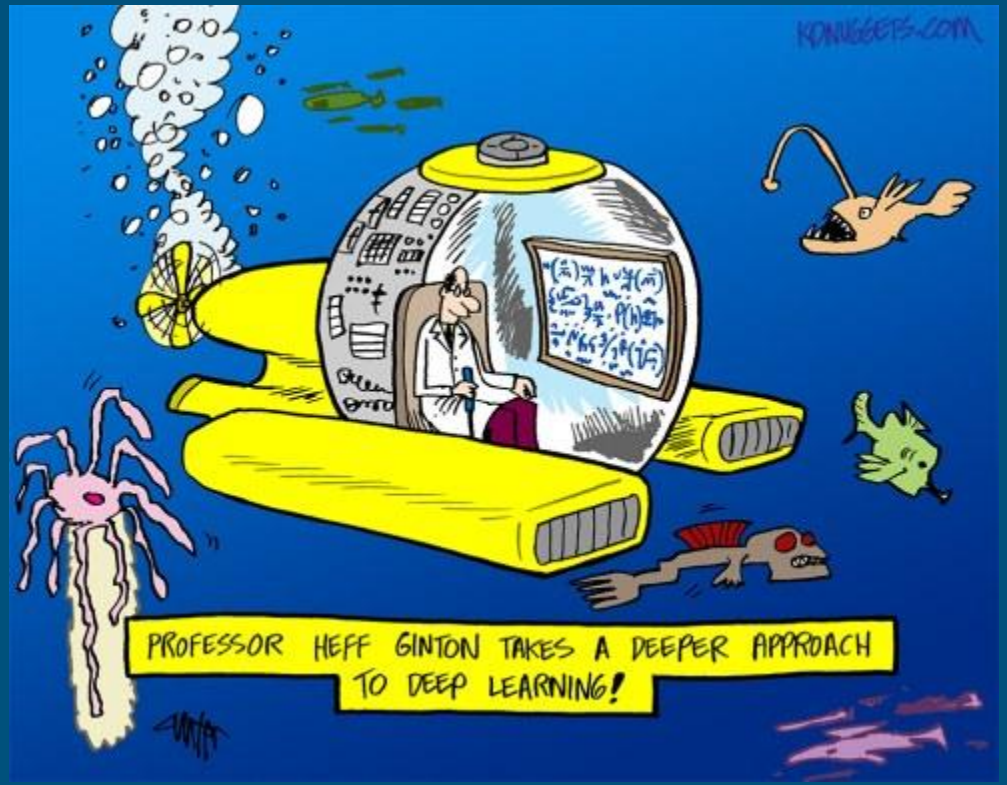
Fine-Tuning

Typically what is done in practice

Most problems do not have enough data to train these networks from scratch



Deeper = Better?



Debugging (What is going wrong?)

Start with a small amount of data and make sure you can overfit it

Looking at activations

Visualizers

Mouth Open

High Cheekbones

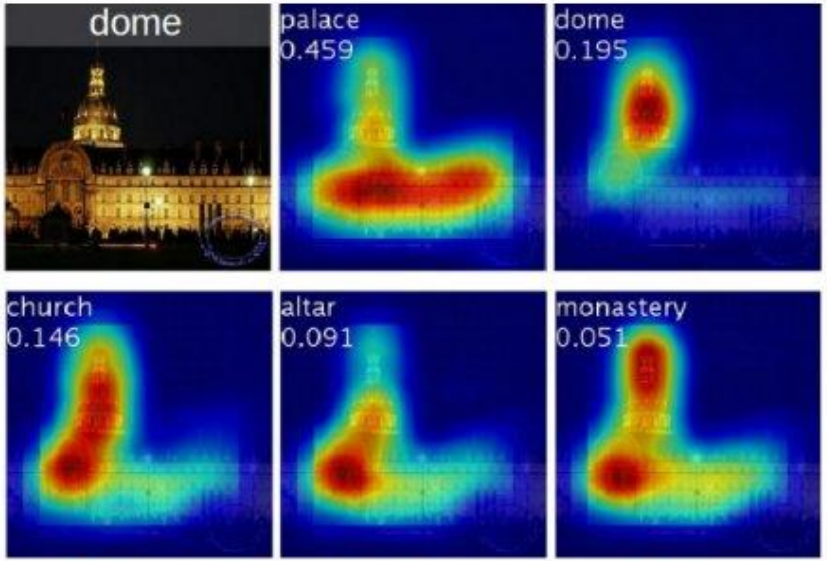
5 o'clock Shadow

Arched Eyebrows

Bushy Eyebrows

Earrings

Activations



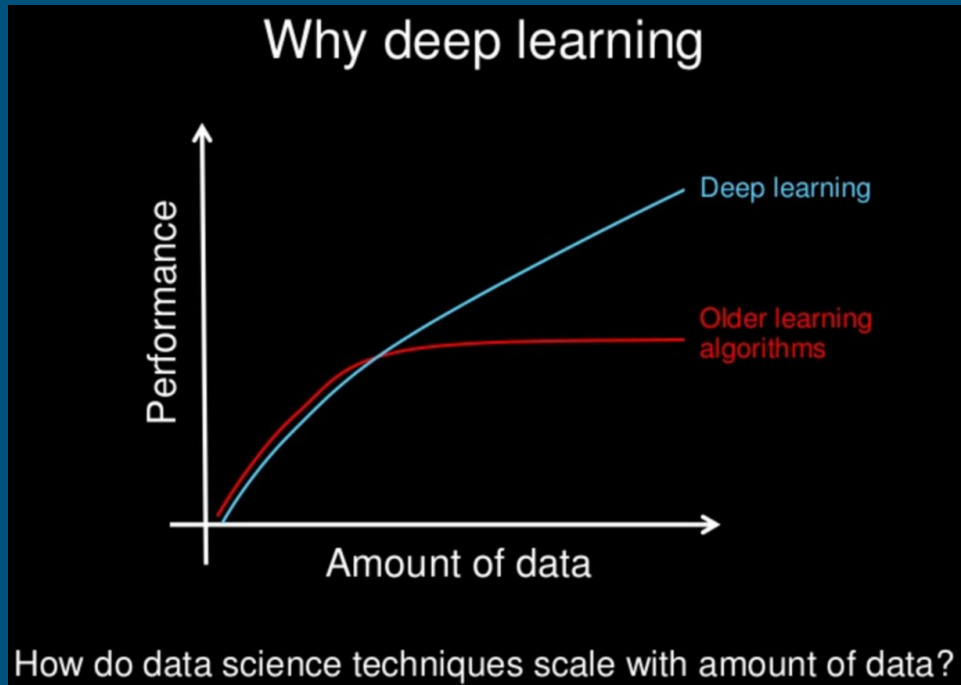
Class activation maps of top 5 predictions

Class activation maps for one object class

Deep Learning Fails

Happens A LOT

Training/Validation/Testing Distribution



Automatic Caption Generation

Deep Learning Fails

Deep Learning Fails


WHEN A USER TAKES A PHOTO, THE APP SHOULD CHECK WHETHER THEY'RE IN A NATIONAL PARK...
SURE, EASY GIS LOOKUP. GIMME A FEW HOURS.
...AND CHECK WHETHER THE PHOTO IS OF A BIRD.
I'LL NEED A RESEARCH TEAM AND FIVE YEARS.
IN CS, IT CAN BE HARD TO EXPLAIN THE DIFFERENCE BETWEEN THE EASY AND THE VIRTUALLY IMPOSSIBLE.

Personal Experience

Getting familiar with software (Caffe, Torch, TensorFlow, MatConvNet)

Most Time:

- Creating/Choosing Loss Function
- Augmenting Data

Less Time:

- Picking Architecture
- Choosing Hyperparameters

Review

Deep Learning is a tool

Be aware of challenges

Combine different ML methods