Training vision models with public transportation datasets

O'Reilly AI Conference
San Francisco 2017

Mo Patel
Director, Artificial Intelligence
Think Big Analytics
Laura Froelich
Data Scientist
Think Big Analytics

Computer Vision
Convolutional Neural Networks
Object Detection
Image Segmentation

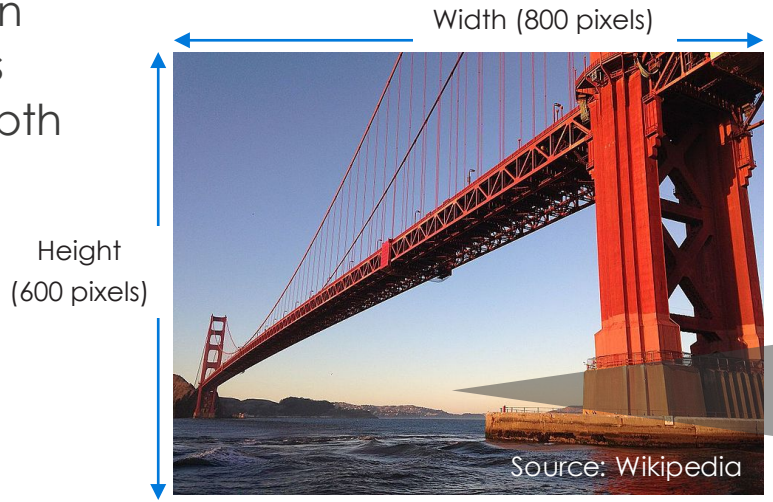
TERADATA.

Agenda

- Computer Vision 101
- History of Computer Vision
- Deep Learning Basics
- Convolutional Neural Networks
- Image Classification Walk Through

Computer Vision 101

- Image Properties
  - Pixels
  - Dimension
  - Channels
  - Color Depth



8-bit Image

Color Depth	Channel Range
1-Bit	0-1
8-Bit	0-255
16-Bit	0-65,535
24-Bit	0-16,777,215

One Pixel

Channel	Red	Green	Blue
Value	0-255	0-255	0-255

TERADATA

Computer Vision 101

Segmentation: Pixel-wise segments of objects in any image

Classification:
Label Item(s) in Image

Description:
Relationship between objects

Detection/Localization:
Bounding Boxes around objects

Goal: Build models that can perform visual tasks

Classification & Description

What is the man doing?

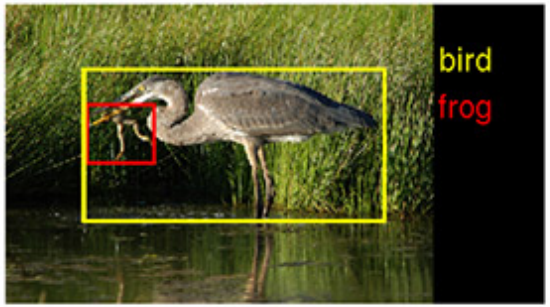
Submit

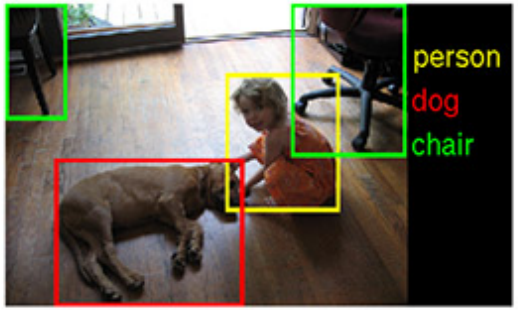
Predicted top-5 answers with confidence:

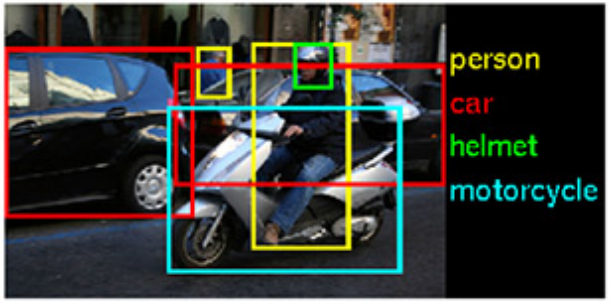
playing	99.126%
tennis	
tennis	0.553%
skateboarding	0.139%
serving	0.075%
jumping	0.033%

TERADATA.

Localization







Source: Kaggle ImageNet Object Detection Challenge

Segmentation

Source: The Cityscapes Dataset - Semantic, instance-wise, dense pixel annotations of 30 classes

Agenda

- Computer Vision 101
- History of Computer Vision
- Deep Learning Basics
- Convolutional Neural Networks
- Image Classification Walk Through

History of Computer Vision

- Objectives
  - Extract generalized features (patterns) from images
  - Determine generalized descriptors for feature
  - Use generalized features to perform visual tasks based on descriptors
- What are features?
  - Edges
  - Shapes
  - Change in Colors/Intensity

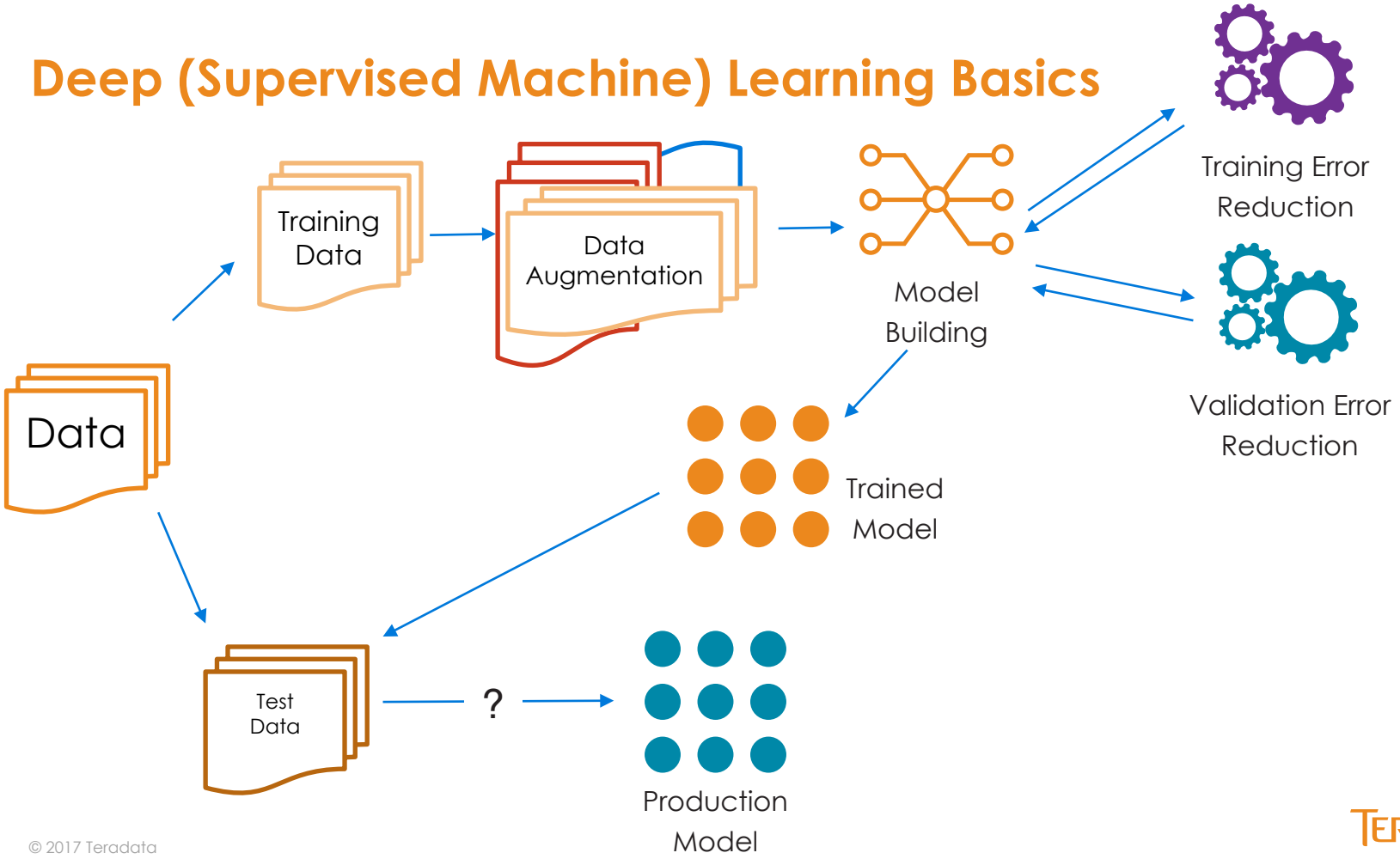
Traditional Machine Learning → Deep Learning

- Featuring engineering challenges
  - Human scaling: Time consuming to create rules for each object
  - Computational scaling: Translating human curated features to algorithms that machines can perform at scale is tedious
- Feature Learning
  - Design computational models to learn features
- Feature Learning Challenges
  - Need large datasets with annotations
  - Model architectures become complex both to design and run

Agenda

- Computer Vision 101
- History of Computer Vision
- Deep Learning Basics
- Convolutional Neural Networks
- Image Classification Walk Through

Deep (Supervised Machine) Learning Basics



Computer Vision Datasets: Video

- Extract Frames and Compose Video
  - FFMPEG
  - https://www.ffmpeg.org/

Extract `ffmpeg -i <video.file> image%03d.jpg`

Compose `ffmpeg -framerate 24 -i img%03d.jpg <output_video.file>`

Computer Vision Datasets: Images

- Data Augmentation in Keras
  - keras.preprocessing.image.ImageDataGenerator

```
featurewise_center
samplewise_center
featurewise_std_normalization
samplewise_std_normalization
zca_whitening
zca_epsilon
rotation_range
width_shift_range
height_shift_range
```

```
shear_range
zoom_range
channel_shift_range
fill_mode
cval
horizontal_flip
vertical_flip
rescale
```

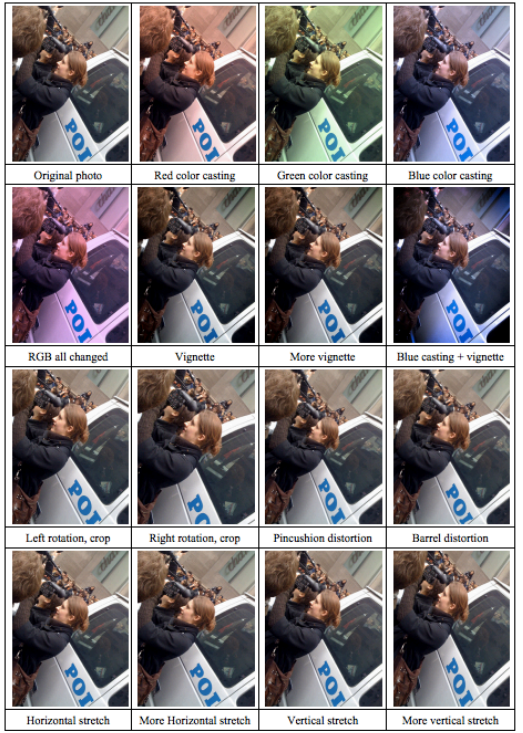


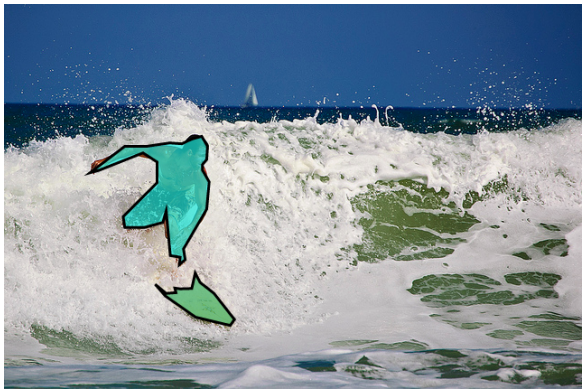
Image Source: Deep Image: Scaling up Image Recognition (Baidu) https://arxiv.org/abs/1501.02876

Computer Vision Datasets: Annotations

{"coco_url":"http://mscoco.org/images/405628",
"date_captured":"2013-11-21 00:26:09",
"file_name":"COCO_train2014_000000405628.jpg",
"flickr_url":"http://farm5.staticflickr.com/4130/5133418813_61f33c9633_z.jpg",
"height":426,
"id":405628,
"license":5,
"width":640}

{"area":7581.427150000003,
"bbox":[95.28,146.43,131.92,142.0],
"category_id":1,"id":188293,
"image_id":405628,
"iscrowd":0,
"segmentation":[[219.87,177.58,224.45,172.08,225.37,157.42,218.95,146.43,208.88,149.18,208.88,149.18,202.46,151.92,198.8,153.76,198.8,153.76,196.97,154.67,184.14,153.76,183.23,153.76,158.49,157.42,133.75,157.42,95.28,186.74,105.35,183.99,153.91,164.75,169.48,174.83,169.48,197.73,135.59,231.63,131.92,235.29,133.75,247.2,152.99,247.2,155.74,247.2,175.9,241.71,179.56,228.88,184.14,278.35,195.14,280.18,194.22,288.43,200.63,282.93,198.8,271.02,224.45,227.05,224.45,227.05,227.2,196.82,219.87,183.99]]},

http://cocodataset.org/#explore?id=405628



TERADATA.

Computer Vision Datasets: Annotations

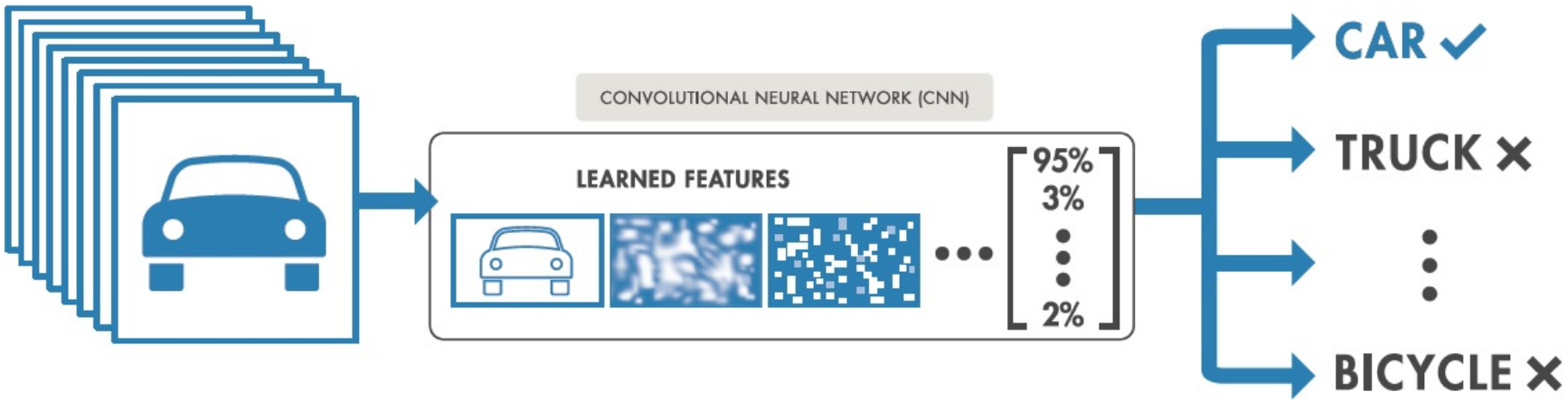
	BOUNDING BOX	DOTS	POLYGON	SEMANTIC SEGMENTATION
COST	Least expensive	More expensive	More expensive	Most expensive
TIME COMMITMENT	Lowest	Medium	Medium	Highest
PRECISION	Good	Great	Great	Excellent
INSTANCE-BASED (OUTPUT CONTAINS DISCRETE OBJECTS)	Yes	Yes	Yes	No
POSSIBLE TO LABEL SEVERAL OBJECTS?	Yes	Yes	Yes	Yes
POSSIBLE TO LABEL SEVERAL CLASSES OF OBJECT IN A SINGLE JOB?	No	No	No	Yes
OUTPUT*	X,Y coordinate, width and length of each box	Series of x,y coordinates	Series of x,y coordinates, with shapes resolving	Coded RGB pixels as an image

Source: CrowdFlower WHAT WE LEARNED LABELING 1 MILLION IMAGES
A practical guide to image annotation for computer vision

Agenda

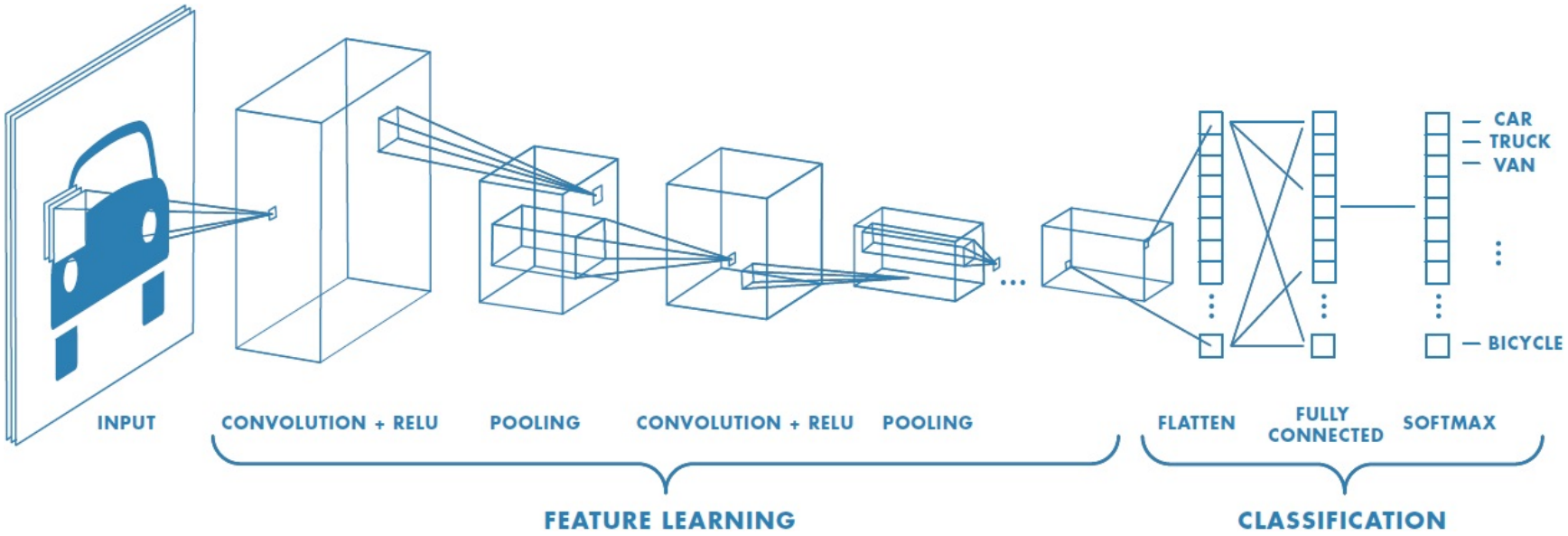
- Computer Vision 101
- History of Computer Vision
- Deep Learning Basics
- Convolutional Neural Networks
- Image Classification Walk Through

Convolutional Neural Networks



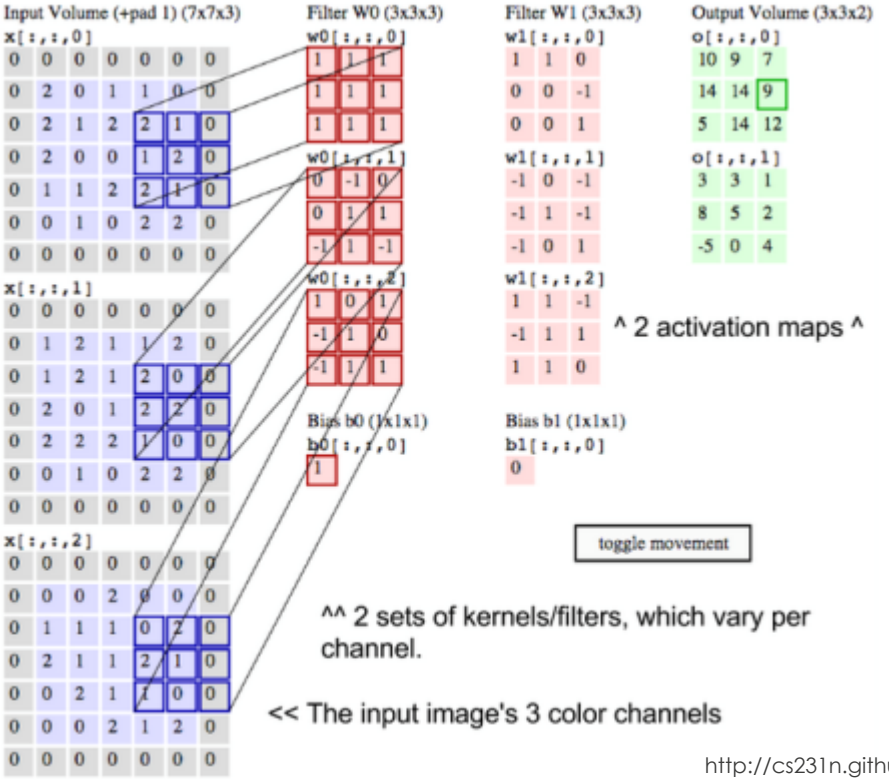
Source: Mathworks

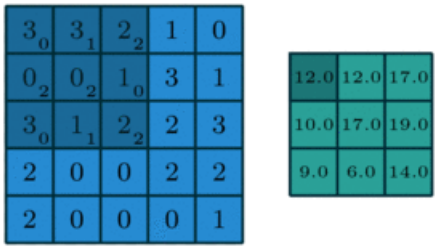
Convolutional Neural Networks



Source: Mathworks

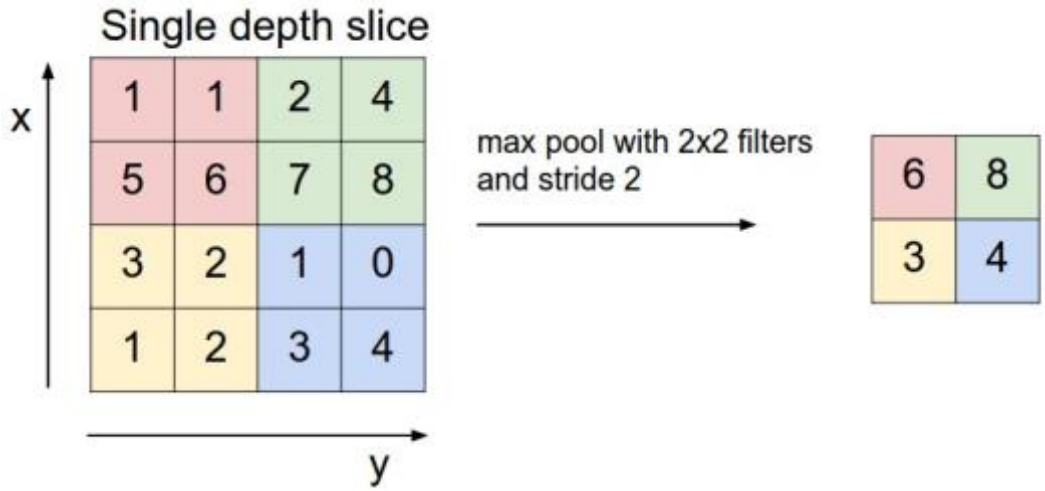
Convolution





http://cs231n.github.io/convolutional-networks/
https://m2dsupsdlclass.github.io/lectures-labs/slides/03_conv_nets/index.html
https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/

Pooling



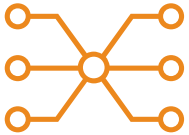
Source: Stanford University CS 231 Course:
http://cs231n.github.io/convolutional-networks/#pool

Fully Connected

- Traditional Machine Learning: Feature Engineering
- Deep Learning: Feature Learning
- Learned Features are fed to traditional classifier such as Multivariate Logistic Regression (Softmax)

Low Level Features

Flatten
10
24
2
3
…
…
…



Classifier

Class	Probability
A	0.01
B	0.07
C	0.13
D	0.60
E	0.03
F	0.05
G	0.11

Loss Reduction

Class	Probability	Truth	Loss/ Error
A	0.01	0	0
B	0.07	0	0
C	0.13	0	0
D	0.60	1	0.22
E	0.03	0	0
F	0.05	0	0
G	0.11	0	0

Cross Entropy Loss

One-Hot Encoded Vector of Class Label

Back Propagation

- Account for loss across the weights and biases in the network
- Use cost function provided by Back Propagation to optimize the weights and biases to reduce loss
- Variety of Stochastic Gradient Descent techniques used for optimization

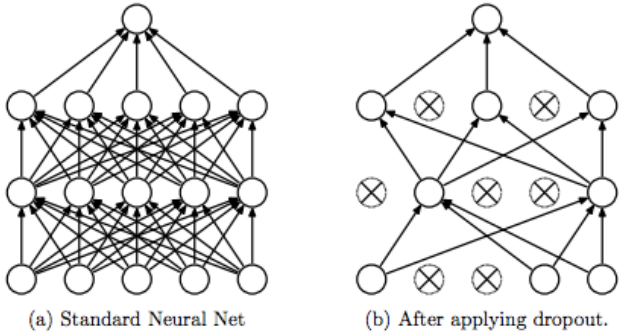
Overview of Gradient Descent Techniques:
http://ruder.io/optimizing-gradient-descent/

TERADATA.

Generalization

Dropout

- Randomly remove learned nodes from network



(a) Standard Neural Net (b) After applying dropout.

Dropout: A Simple Way to Prevent Neural Networks from Overfitting
http://www.jmlr.org/papers/volume15/srivastava14a/srivastava14a.pdf

Batch Normalization

- During training input data shifts as weights and parameter adjust values
- To prevent internal covariate shift, normalize each batch by mean and variance

Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift
https://arxiv.org/abs/1502.03167

Agenda

- Computer Vision 101
- History of Computer Vision
- Deep Learning Basics
- Convolutional Neural Networks
- Image Classification Walk Through

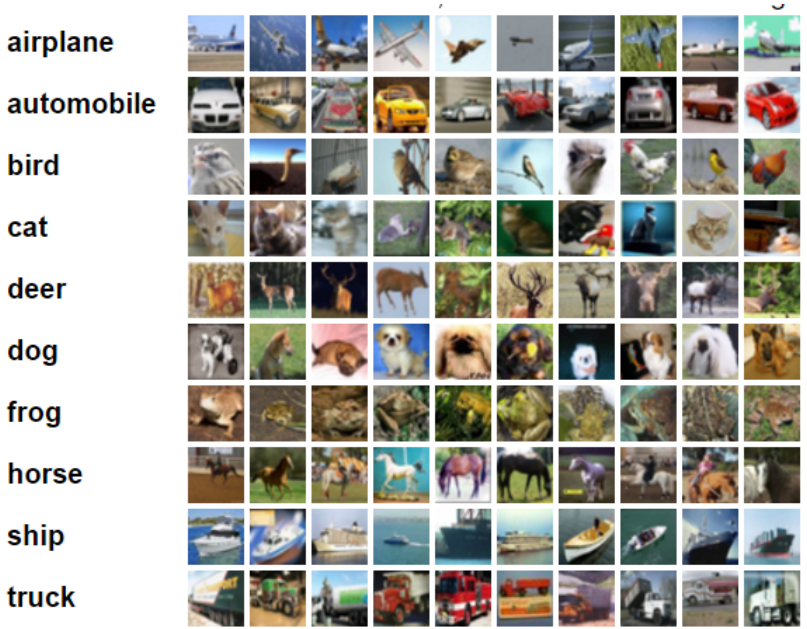
Deep Learning Software Quick-Start

- Get Anaconda Python Distribution
  - https://www.continuum.io/downloads
- Install Keras*
  - Open Anaconda Navigator
  - Go to Environments
  - Click on Channels, Click on Add, Add conda-forge and click on Update Channels
  - Search for TensorFlow and check the box in Name column
  - Search for Keras and check the box in Name column
  - Click Apply
- Launch Jupyter Notebook
  - Open Anaconda Navigator
  - From Home, click on Launch button for Jupyter Notebook

*your mileage may vary

CIFAR10 Image Classification

- Ten classes:
  - airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck
- Dataset:
  - 60,000 32x32 colour images in 10 classes
  - 6000 images per class.
  - 50,000 training images
  - 10,000 test images



https://www.cs.toronto.edu/~kriz/cifar.html