Group2: Machine Learning

Shubham Arora, Thuy Pham, Mengting Song,Ningrong Chen, Bowen Qin,

Agenda

1. Computer Vision - General Tasks
2. Object Detection Technologies
3. Unsupervised Learning

Computer Vision

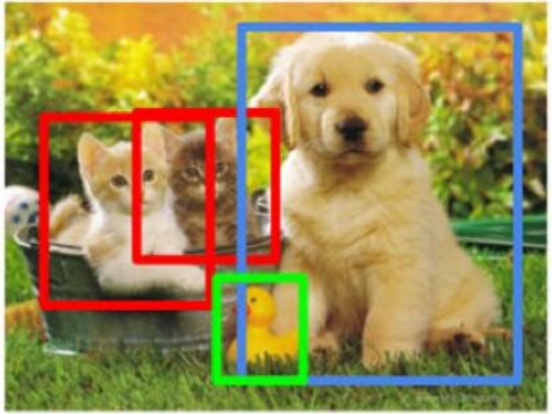
A subfield of Artificial Intelligence and Machine Learning that deals with tasks that require high level understanding of images and videos. The expected benchmark in any such task is to reach human level accuracy. Human level accuracy is anywhere between 2-5% (depending on task)

Tasks in Computer Vision

Computer vision deals with the following set of tasks:

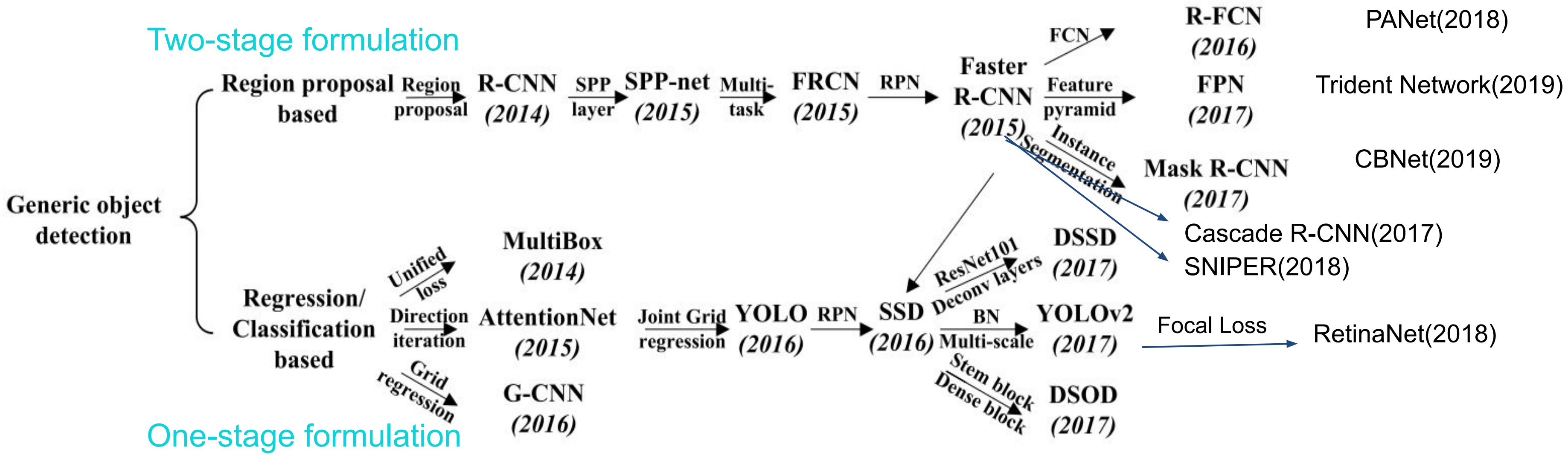
- Recognition
  - Image classification
  - Image captioning
  - Object localization
  - Segmentation
  - Object Detection
- Motion Analysis
  - Tracking
  - Optical flow
- Other
  - Image recolorization
  - Super-Resolution

What is Object Detection?

Classification	Object Detection
	
CAT	CAT, DOG, DUCK

Object Detection

is a computer technology related to computer vision and image processing that deals with detecting instances of semantic objects of a certain class (such as humans, buildings, or cars) in digital images and videos.



Two Types of Object Detection(Partial)

Deep Learning

Artificial Neural Networks that are characterized by having multiple layers

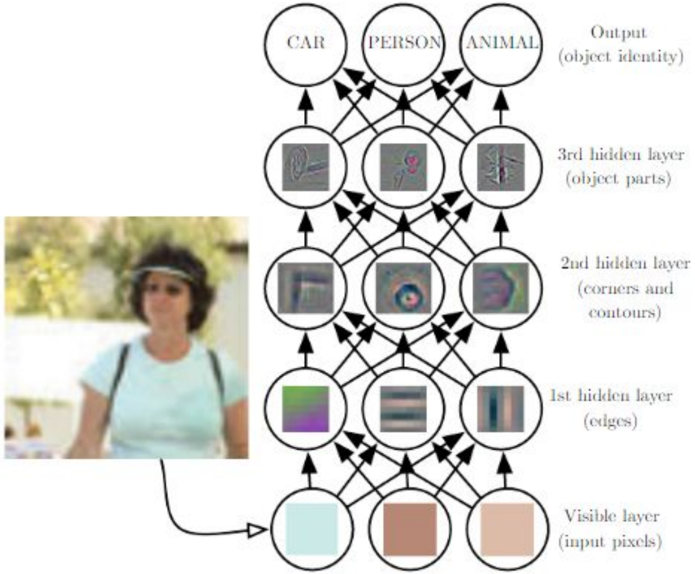


Figure: from DeepLearning book by Goodfellow-et-al-2016

Basic Deep Learning

- The goal of a neural networks is to approximate some function f()
- Intuitively, we want the NN to learn increasing level of abstraction in successive layers

CNN History

- 1998 - CNN’s were first introduced in the LeNet paper by Yann LeCun. LeNet-5 was one of the first Neural Network that utilized backpropagation using Supervised Learning.
- 2012 - AlexNet - Geoff Hinton et al. - Image classifier on the ImageNet database. First NN that beat handcrafted image classifiers on ImageNet dataset
- Dawn of realization that Deep CNN’s were a feasible architecture.
- Followed many CNN enhancements - ResNet, Fast-CNN, Faster-CNN, etc

Why Deep CNNs ?

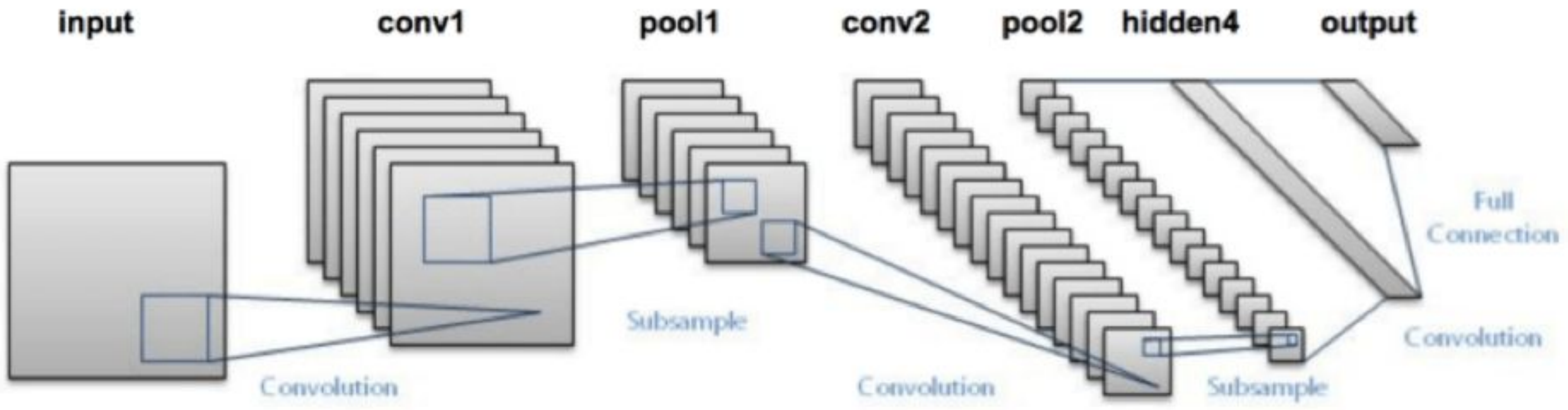
Deep Convolutional Neural Networks are the state of the art today because:

- Increase in size of datasets - we have **cameras** everywhere
- Increase in Computing power - **Moore's** law
- Much better parallel computing - **GPU** for matrix operations
- Reduced code complexity - **CNNs** homogenized the vision stack - no need of complex feature extraction

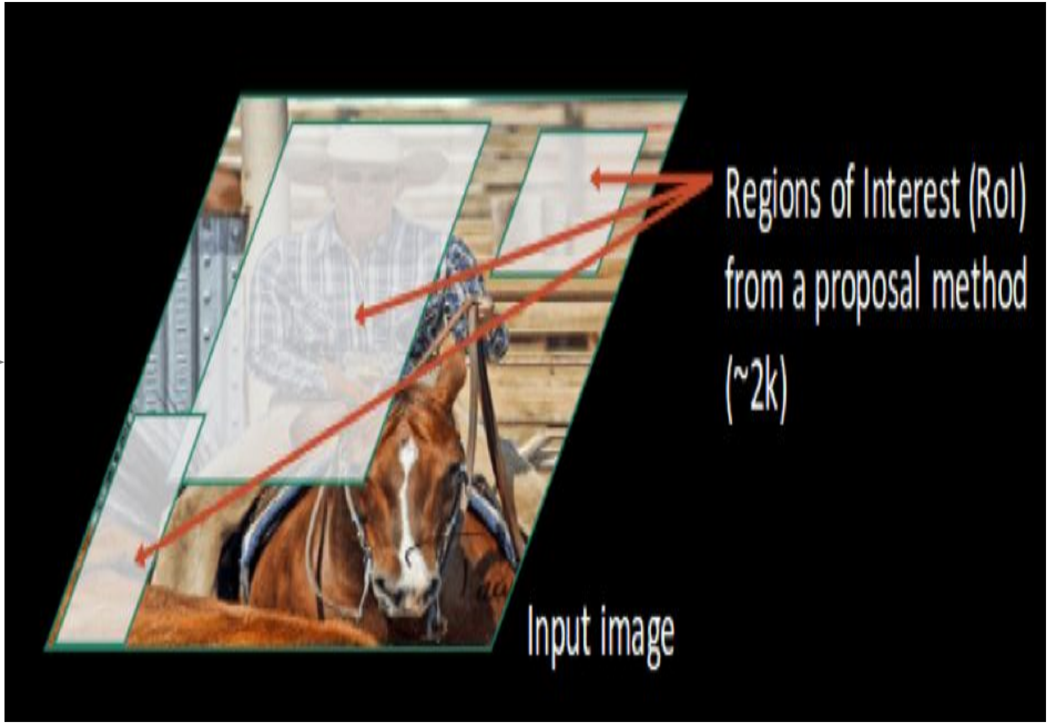
CNN Family - CNN

1. Take an image as input
2. Divide the image into various regions
3. Then consider each region as a separate image.
4. Pass all these regions (images) to the CNN and classify them into various classes.
5. Once we have divided each region into its corresponding class, we can combine all these regions to get the original image with the detected objects

CNN Family - CNN



CNN Family - RNN (Region-based CNN)



Selective Search

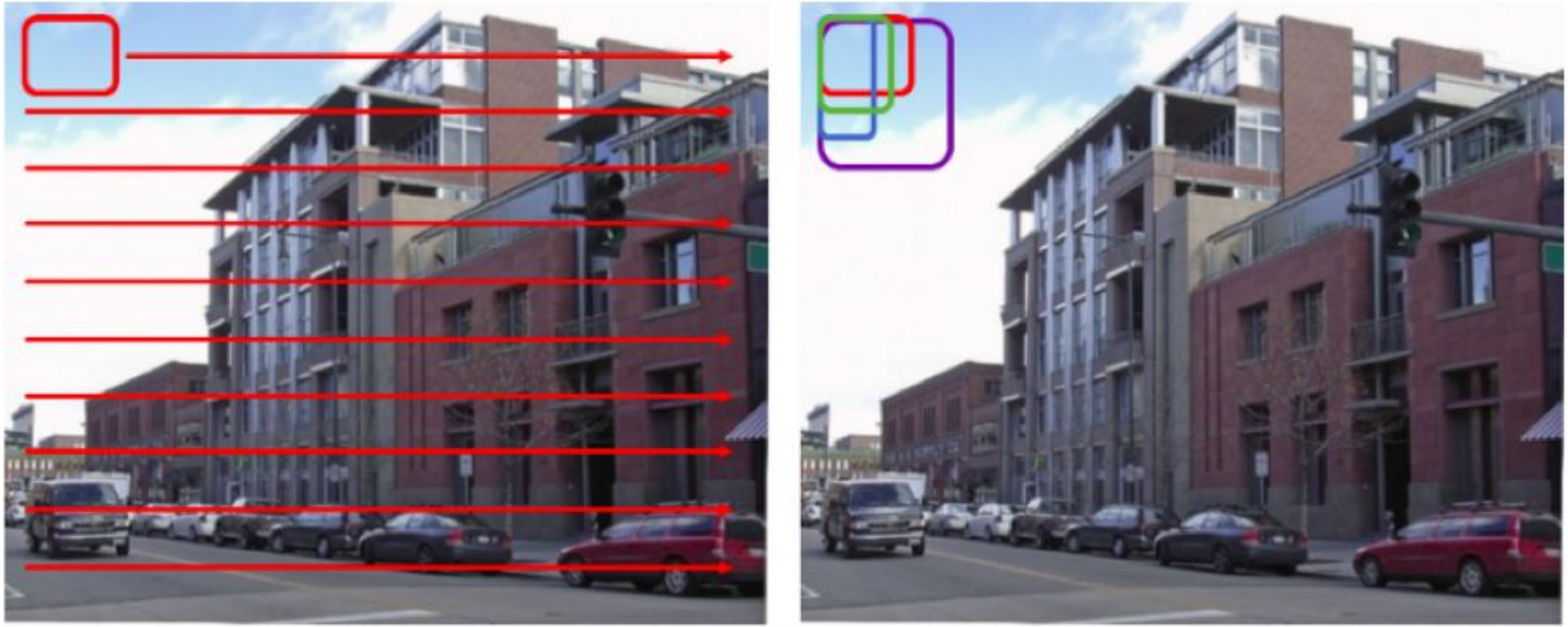
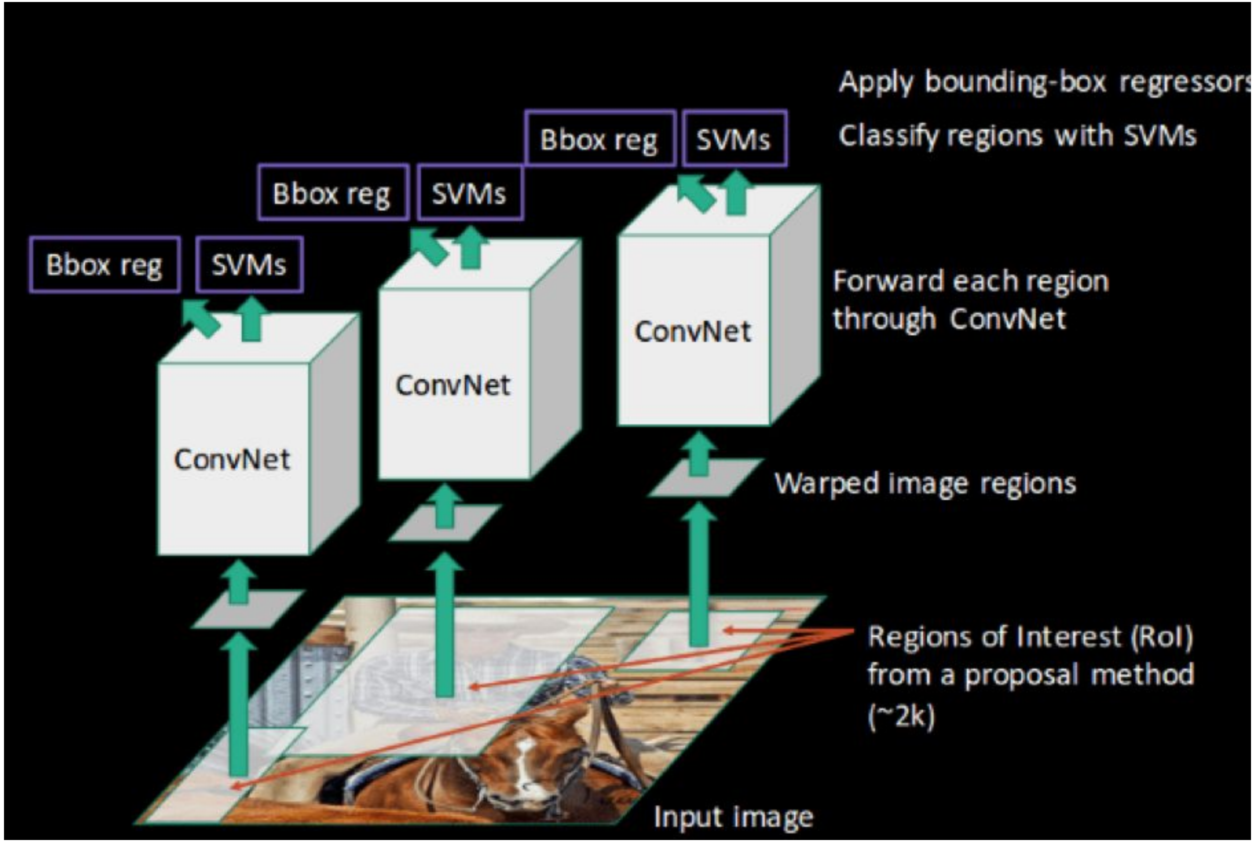


Illustration of Sliding Window (Left) with Different Aspect Ratios and Sizes (Right)

CNN Family - RCNN (Region-based CNN)

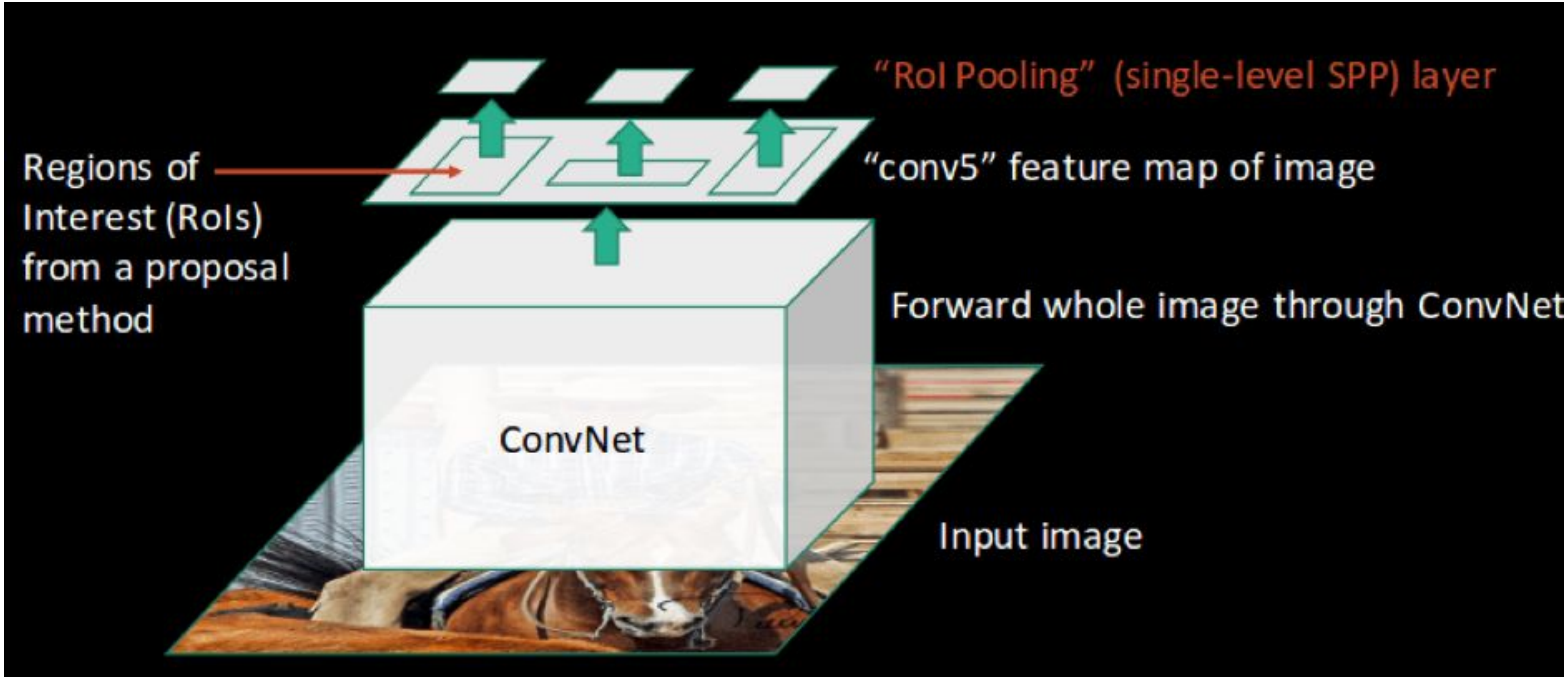


Problem:

Training an RCNN model is expensive and slow:

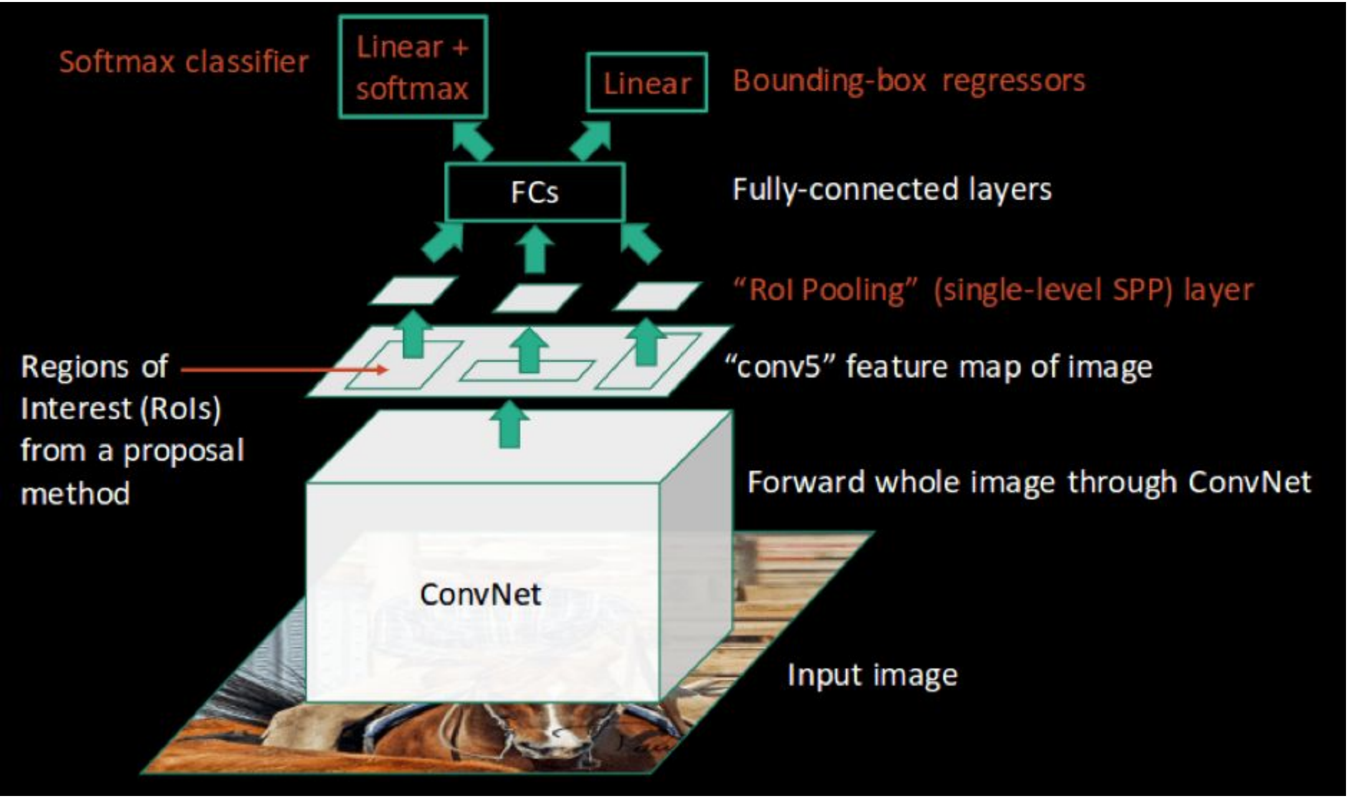
- Extracting **2,000 regions** for each image based on selective search
- It takes around **40-50 seconds** to make predictions for each new image

CNN Family - Fast RCNN



- The **whole image** is passed to a ConvNet which returns the region of interests accordingly
- Apply the RoI pooling layer on the extracted regions of interest to make sure all the regions are of the same size

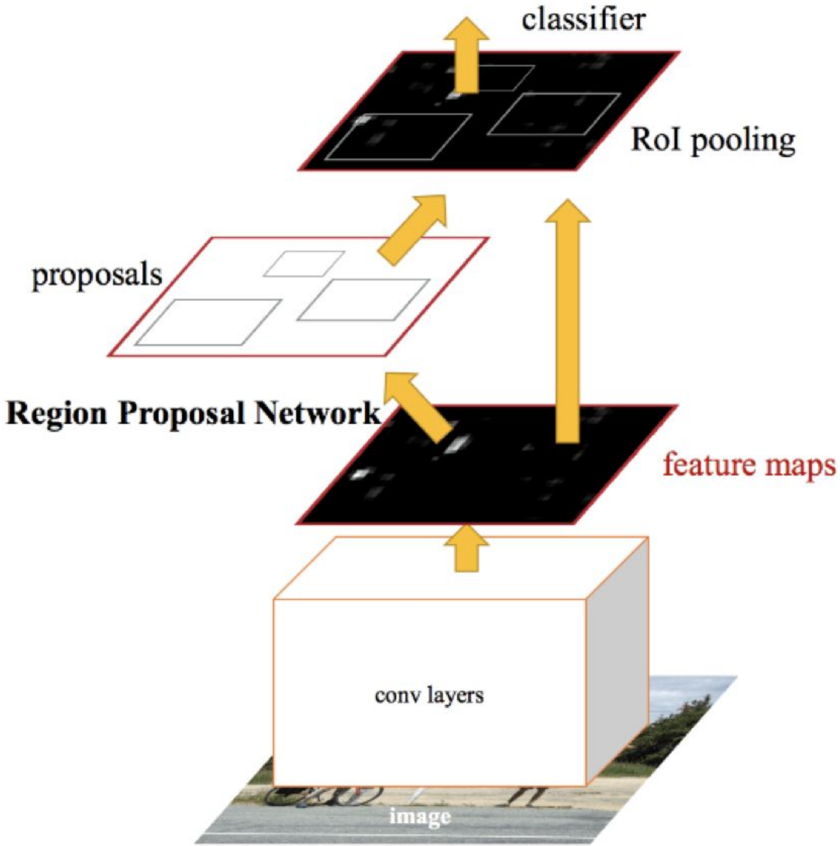
CNN Family - Fast RCNN



It also uses **selective search** as a proposal method to find the Regions of Interest.

It takes around 2 **seconds per image** to detect objects. But when not applicable for large real-life datasets.

CNN Family - Faster RCNN

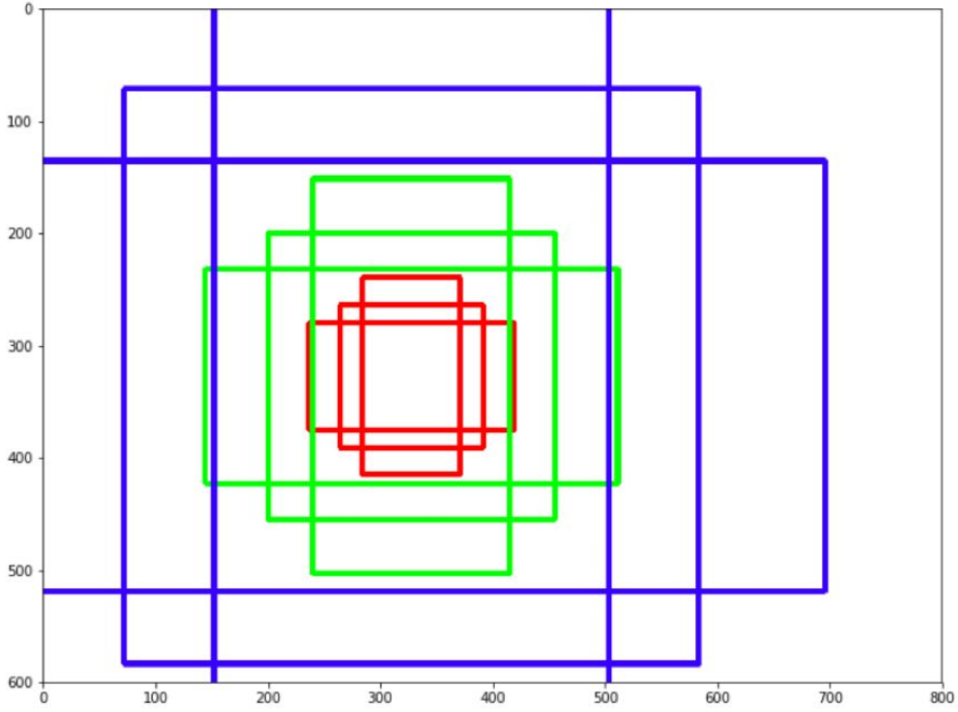


Instead of Selective Search, Faster RNN use **Region Proposal Network** (RPN) on these feature maps. This returns the object proposals along with their objectness score.

1. RPN predicts the probability that an anchor is an object (it does not consider which class the object belongs to).
2. RPN predicts the bounding box regressor for adjusting the anchors to better fit the object.

It takes around 0.2 **seconds per image** to detect objects using Faster RCNN.

How does RPN work?



Anchors at (320, 320)

An example:

9 anchors at the position (320, 320) of an image with size (600, 800)

- Three colors represent three scales or sizes: 128x128, 256x256, 512x512.
- The three boxes of each color have height-width ratios 1:1, 1:2 and 2:1 respectively.
- Choose one position at every stride of 16, there will be 1989 (39x51) positions. This leads to 17901 (1989 x 9) boxes to consider.

YOLO - You Only Look Once

The YOLO framework (You Only Look Once) incredibly fast and can process **45 frames per second**.

1. YOLO first takes an input image.
2. The framework then divides the input image into grids (say a 3 X 3 grid).
3. Apply image classification and localization on each grid.

YOLO - You Only Look Once

	pc
	bx
	by
	bh
y =	bw
	c1
	c2
	c3

For each grid cell, the label y will be an 8-dimensional vector:

- p_c defines whether an object is present in the grid or not (it is the probability).
- b_x, b_y, b_h, b_w specify the bounding box if there is an object.
- c_1, c_2, c_3 represent different classes.

YOLO - You Only Look Once

y =	1
	bx
	by
	bh
	bw
	0
	1
	0

(0,0)

(1,1)

y =	1
	0.4
	0.3
	0.9
	0.5
	0
	1
	0

Since there is an object in this grid, p_c will be equal to 1.

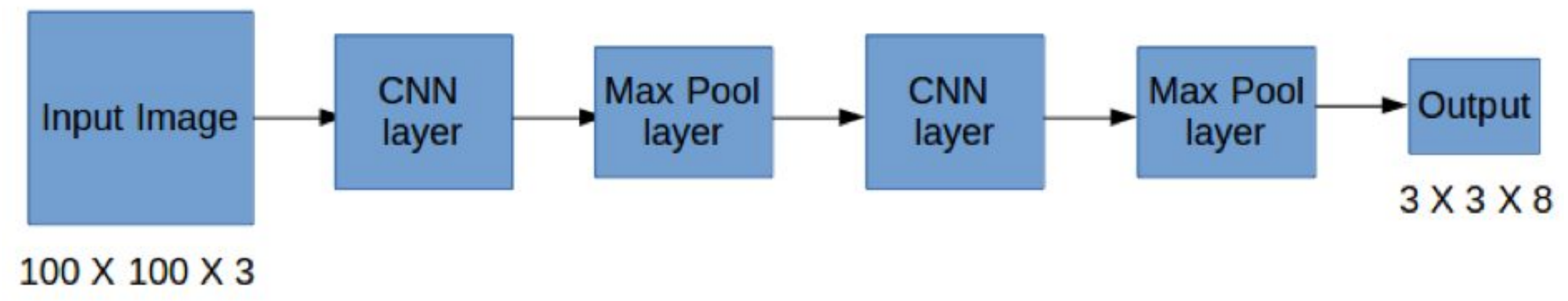
Say c_1 - pedestrian, c_2 - car. c_3 - traffic lights.

Since car is the second class, c_2 = 1 and c_1 and c_3 = 0.

b_x, b_y are the **x and y coordinates of the midpoint of the object** with respect to this grid.

b_h is the **ratio of the height of the bounding box** (red box in the above example) to the height of the corresponding grid cell, b_w is the **ratio of the width of the bounding box** to the width of the grid cell.

YOLO - You Only Look Once



Even if an object spans out to more than one grid, it will only be assigned to a single grid in which its mid-point is located. We can reduce the chances of multiple objects appearing in the same grid cell by increasing the number of grids.

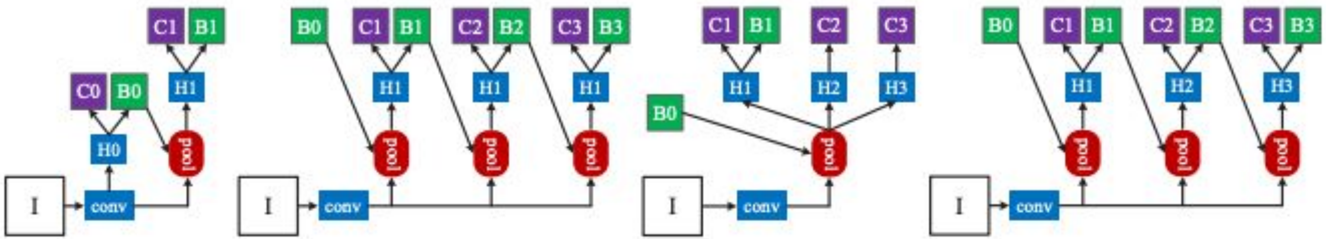
Generally, in real-world scenarios we take larger grids (e.g. 19 X 19).

Cascade R-CNN

Four-Stage

The first model to cascade detectors which upgrade two-stage to four-stage for delving into High Quality Object Detection for addressing the IoU threshold problem:

In object detection, an intersection over union (IoU) threshold is required to define positives and negatives. An object detector, trained with low IoU threshold, e.g. 0.5, usually produces noisy detections. However, detection performance tends to degrade with increasing the IoU thresholds.



(a) Faster R-CNN (b) Iterative BBox at inference (c) Integral Loss (d) Cascade R-CNN

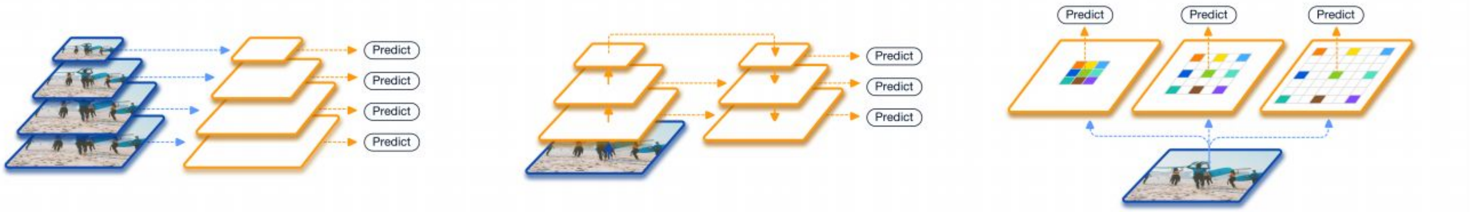
Figure 3. The architectures of different frameworks. "I" is input image, "conv" backbone convolutions, "pool" region-wise feature extraction, "H" network head, "B" bounding box, and "C" classification. "B0" is proposals in all architectures.

SNIPER

Methods for handling scale variation:

Multi-scale image pyramid

SNIPER (Scale Normalization for Image Pyramids with Efficient Resampling) is a multi-scale algorithm for addressing the speed problem of image pyramids developed by B.Singh and al.(2018). Instead of processing every pixel in an image pyramid, SNIPER propose chips (scale specific context-regions that cover maximum proposals at a particular scale). The number of chips generated per image during training adaptively changes based on the scene complexity.



(a) Image Pyramid (b) Feature Pyramid (c) Trident Network

PANet

Methods for handling scale variation:

Multi-level features of different spatial resolutions to alleviate scale variation

Path Aggregation Network (PANet) optimized FCN by improving information flow in proposal-based instance segmentation framework.(S.Liu and al.2018) Specifically, this enhances the entire feature hierarchy in FPN with accurate localization signals in lower layers by bottom-up path augmentation, which shortens the information path between lower layers and topmost feature. The developers present adaptive feature pooling, which links feature grid and all feature levels to make useful information in each feature level propagate directly to following proposal subnetworks. A complementary branch capturing different views for each proposal is created to further improve mask prediction.

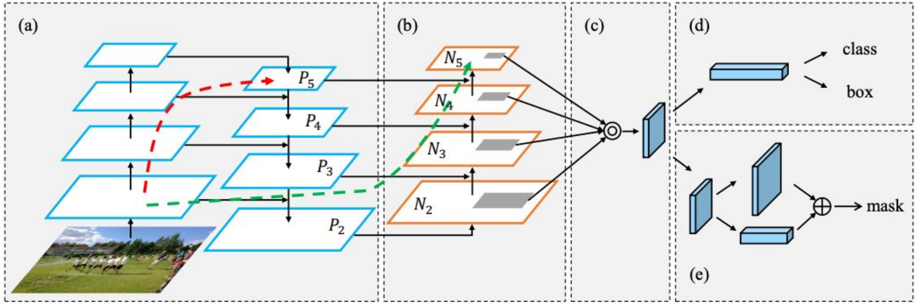


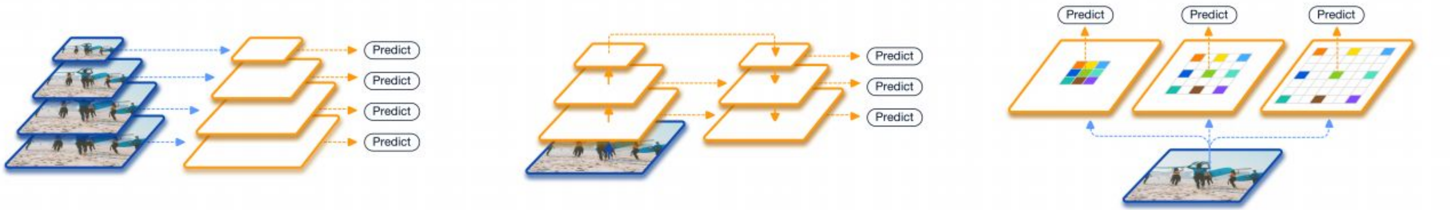
Figure 1. Illustration of our framework. (a) FPN backbone. (b) Bottom-up path augmentation. (c) Adaptive feature pooling. (d) Box branch. (e) Fully-connected fusion. Note that we omit channel dimension of feature maps in (a) and (b) for brevity.

TridentNet

Methods for handling scale variation:

Trident Network

Different from other scale method like image pyramid and feature pyramid, Scale-aware Trident Network (Y.Li and al.2019) solves the scale variation based on COCO dataset in the SimpleDet Framework. Image Pyramid’s testing speed is slower, but scaling performance is pretty good. Feature Pyramid is similar to image pyramid on feature for speeding up, but the performance is not as good as image pyramid. So the writers purposed a network to solve the scale variation problem by combining the advantages of various receptive fields on different sizes’ objects.

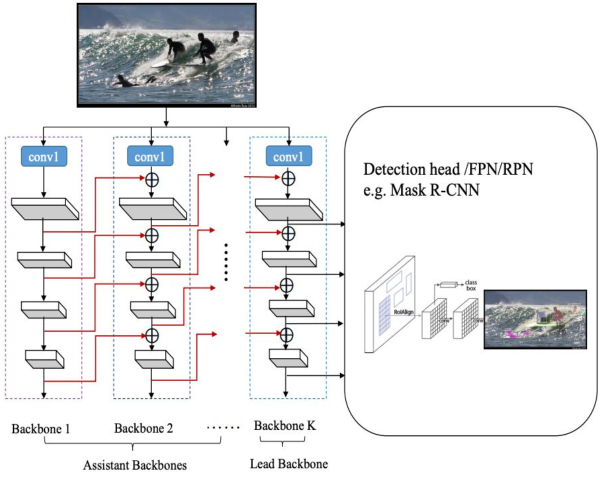


(a) Image Pyramid (b) Feature Pyramid (c) Trident Network

CBNet

A new backbone Network

Y.Liu and al.(2019) purposed a Composite Backbone Network (CBNet) for improving the detection performance from existing backbones. And Recently, the best state-of-the-art model on COCO test-dev is done by CBNet and Cascade Mask R-CNN which combines the Cascade R-CNN for object detection and Mask R-CNN for instance segmentation.



TensorFlow

An end-to-end open source platform for machine learning. It provides a collection of workflows to develop and train models.

Anywhere

TensorFlow allows deep neural network computing to be deployed on any number of CPUs or GPUs on a server, PC or mobile device, using only one TensorFlow API

Any language

It supports C, Python, JavaScript, C++, Java, Go, and Swift. Other support language including C#, Haskell, Julia, Ruby, Rust, and Scala are still under development.

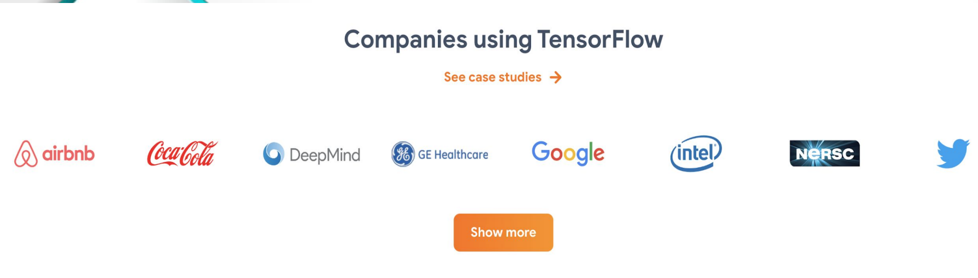
Powerful

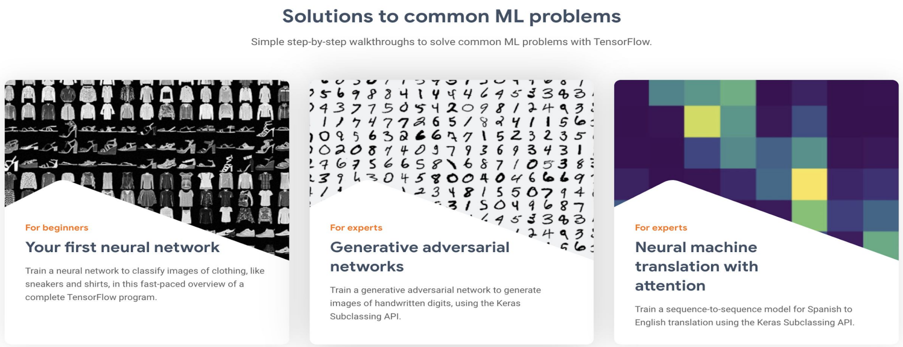
It has been integrated with Keras, a high-level neural network API.

It also support common NN architectures such as recurrent neural networks (RNN) and convolutional neural networks (CNN).) and Deep Trusted Network (DBN).

Community

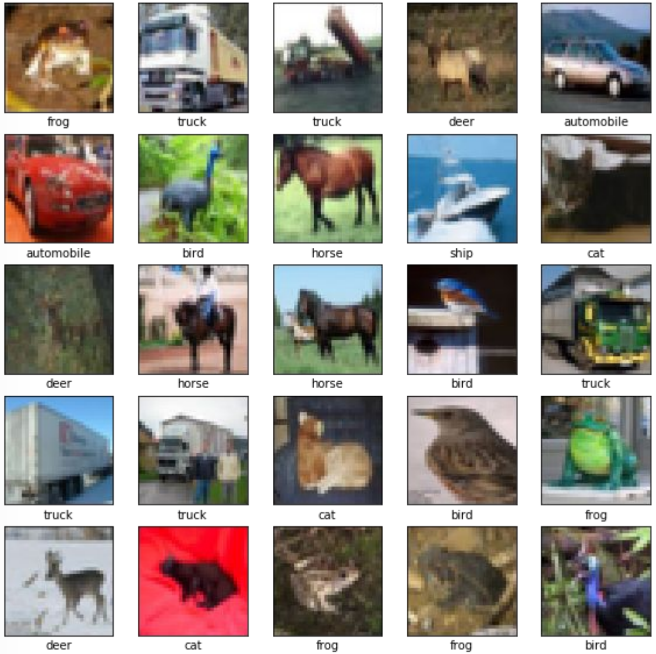
It is more than just a software library. It is a suite of software including TensorFlow, TensorBoard and TensorServing.

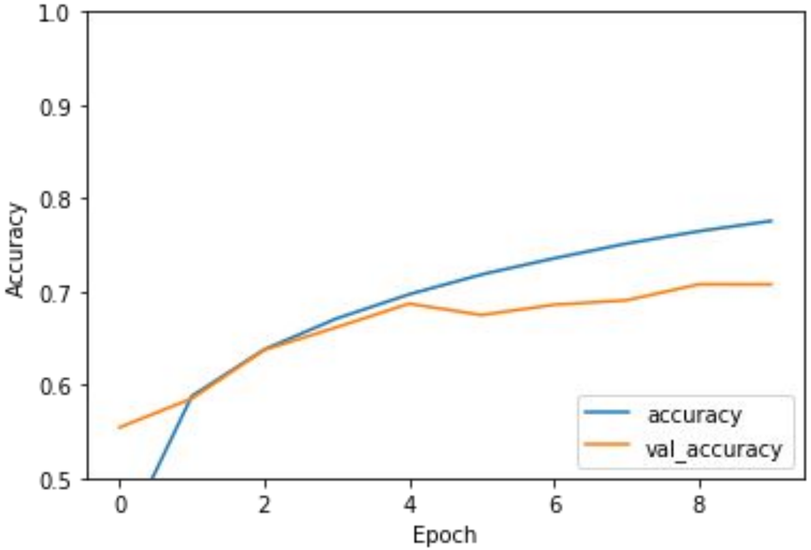




TensorFlow

E.g. Training a simple Convolutional Neural Network (CNN) to classify CIFAR images with test accuracy of over 70%





Unsupervised Learning

is a branch of machine learning wherein algorithms/models are created to infer patterns from a dataset

Unsupervised Learning: Fraud Detection

Applications of Unsupervised Learning

- Clustering
- Anomaly detection

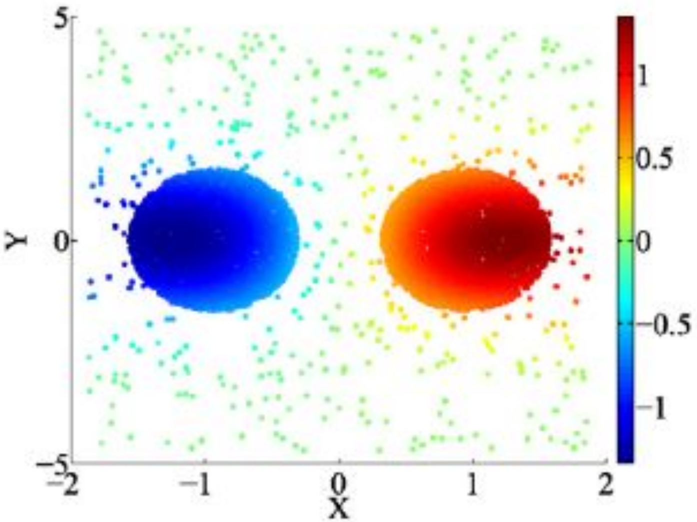
Specific case: **Credit Card Fraud Detection**

- Traditional approach: supervised learning
- Proposed approach: sequential combination of supervised & unsupervised learning
- Results:
  - Worse than supervised learning (random forest model)
  - Potential: augment data with only best outlier score
- Considerations:
  - Performance metrics
  - Limitations
    - Parameter tuning
    - Choice of contextual attributes

Unsupervised Learning: Fraud Detection

Specific case: Auto-Insurance Fraud Detection

- Unsupervised learning preferred to supervised
- Use **Spectral Ranking for Anomaly** (SRA)
  - Ranking beneficial for this case
  - Spectral clustering = techniques that use eigenvalues of a similarity matrix of data to perform dimensionality reduction before clustering in fewer dimensions
    - Calculate "class clustering strength" indicator = z, "inverse" = f
    - Green/yellow points = outliers (have low z value, high f value)



Unsupervised Learning:
Fraud Detection

Specific case: Auto-Insurance Fraud Detection

- **Result**: SRA performed better than One-Class SVM (OC-SVM) and Local Outlier Factor (LOF)
  - More Validation: top 3 most important features same as those discovered by supervised random forest model = base policy, car types, and fault
- **Benefits**:
  - SRA can simultaneously distinguish small clusters and global outliers
  - Proposed SRA can distinguish edges of main clusters from cores
- **Considerations**
  - What measure of “similarity” to use?
  - No mechanism to tune unsupervised learning parameters

Machine Ethics

Is an emerging field in philosophy concerned with the moral behavior of artificially intelligent beings

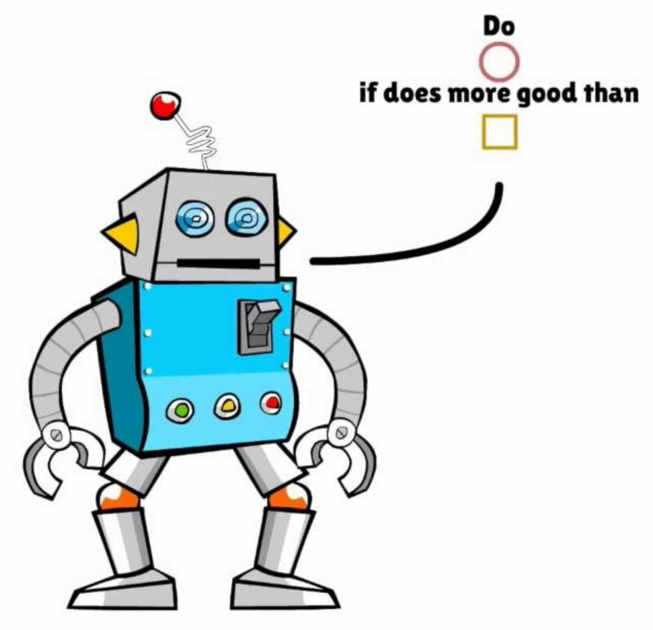
Machine Ethics

Also known as...

- Machine morality
- Computational morality
- Computational Ethics

Not to be confused with...

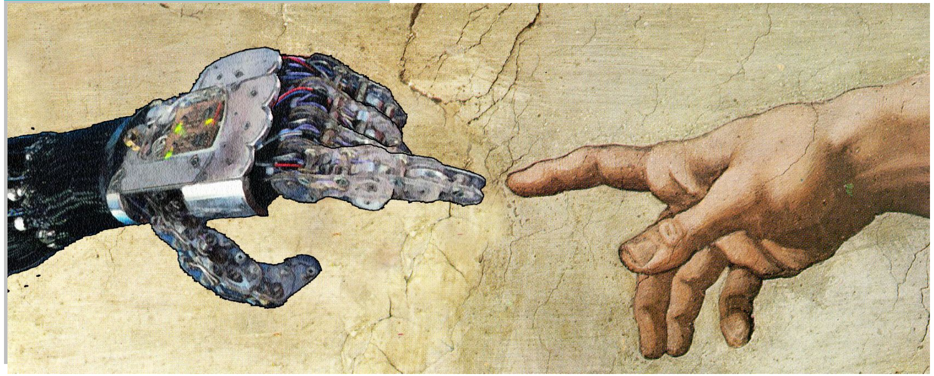
- Roboethics
- Computer ethics
- Philosophy of technology



Machine Ethics:
Creating an Ethical Intelligent Agent

Main Points

- Need for machine ethics
- Interdisciplinary nature
- Explicit ethical agent vs implicit ethical agent



- Concerns
  - Can ethics be computed? - hedonistic act utilitarianism
  - Are machines beings that can behave ethically?
- Ex: EthEl

INTRODUCTION: UNSUPERVISED LEARNING

The various problems in pattern recognition are solved according to training samples whose categories are unknown (not marked)

One of the main methods: Cluster analysis. That is used in unsupervised learning to group, or segment, datasets with shared attributes in order to extrapolate algorithmic relationships.

INTRODUCTION: CLUSTERING DIVIDES

Clustering divides the data set into different categories based on the data characteristics, so that the data within the category is relatively similar/correlated, and the data similarity/correlation between the categories is relatively small. We focus on two properties. The first property is consistency and the second property is association.

ANALYSIS

Pros:

(1) No label & Clustering

Unsupervised learning is density estimation (look for descriptive data statistics), which means that can start working as long as it knows how to calculate similarity.

(2) Reduce dimension

Unsupervised learning need to extract features, or use layer/item clustering to reduce the dimension of data features.

ANALYSIS

(3) Non-independent

For different scenarios, the distribution of positive and negative samples may have offset. Large offsets are likely to cause noise to the classifier, but for unsupervised learning’s situation is much better.

(4) Interpretable

The reason for the classification of supervised algorithms is unclear, because these rules are derived by manual modeling and cannot be self-generated. Unsupervised clustering is well explained. Because the elements in one group have similar features and consistency.

ANALYSIS

(5) Expandability

A n-dimensional model, if is added into a very strong feature, which will break up the original classification. The unsupervised algorithm is scalable. No matter how high the weight of this multi-dimensional data is, it does not affect the original result output. The original result can still be retained, and only needs to process new dimension data.

Cons:

(1) Accuracy and validity

Because researchers can set the label for the supervised learning in advance, which means supervised learning can be run in a more reasonable way under control. But for unsupervised learning, it learns objects in its own logic analysis without any one interfering, so it may have unpleasant results.

APPLIED FIELD

(1) Data mining

Unsupervised learning is often used for data mining to find out what is in a large amount of unlabeled data. Its training data is unlabeled, and the training goal is to classify or distinguish observations.

(2) Abnormal detection

Unsupervised learning classifies the objects according to their features, so if the abnormal objects have little consistency with other groups, they will be detected. For example, it always be applied to anti-fraud in finance industry.

APPLIED FIELD

(3) Detect a segment objects

It can be used in recognizing segment objects. For example, Computer Vision in Unmanned vehicle will use this method to detect the roads, lines and traffic lights.

(4) Advertisement

Applying this method to find users' interest based on what the users usually search or browse and then advertise precisely.

Summary

1

Computer Vision: Object Detection

- Intro to Deep Learning, with focus on CNNs
- State-of-the-art models
    - R-CNN, Fast R-CNN, Faster R-CNN
    - YOLO
    - Cascade R-CNN, SNIPER,PANet,TridentNet, CBNet
- TensorFlow

2

Unsupervised Learning: Fraud Detection

- Hybrid model (unsupervised + supervised) for credit card fraud
- Spectral Ranking for Anomaly (SRA) for auto-insurance fraud

3

Machine Ethics

- Important & necessary
- Interdisciplinary nature
- Should aim to build explicit ethical agents
- Complicated

Thank you

This text can be replaced with your own text