Modern Machine Vision from Basics to Advanced Deep Learning

Zoran Sevarac
University Of Belgrade /
Deep Netts

DEEP NETTS

What are you going to learn

- Machine learning basic ideas
- Deep learning, and what's so special about it
- How to use deep learning for image classification
- Advanced deep learning models for object detection

© 2021 Deep Netts

Machine Learning Basics

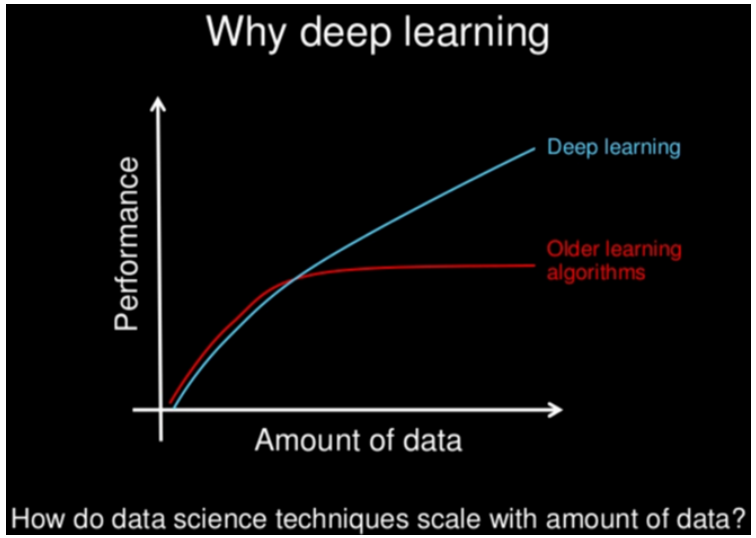
What is Machine Learning?

- Type of computer algorithm that is capable
  to **automatically configure** its internal parameters (learn)
  by looking at a set of **examples** given as a data set,
  in order to perform **specific task** on similar data with usable accuracy.
- **Examples of what it can do :**
  - Learn to assign items to category - **classification**
  - Learn relationship between variables,
    in order to estimate a numeric value - **regression**

© 2021 Deep Netts

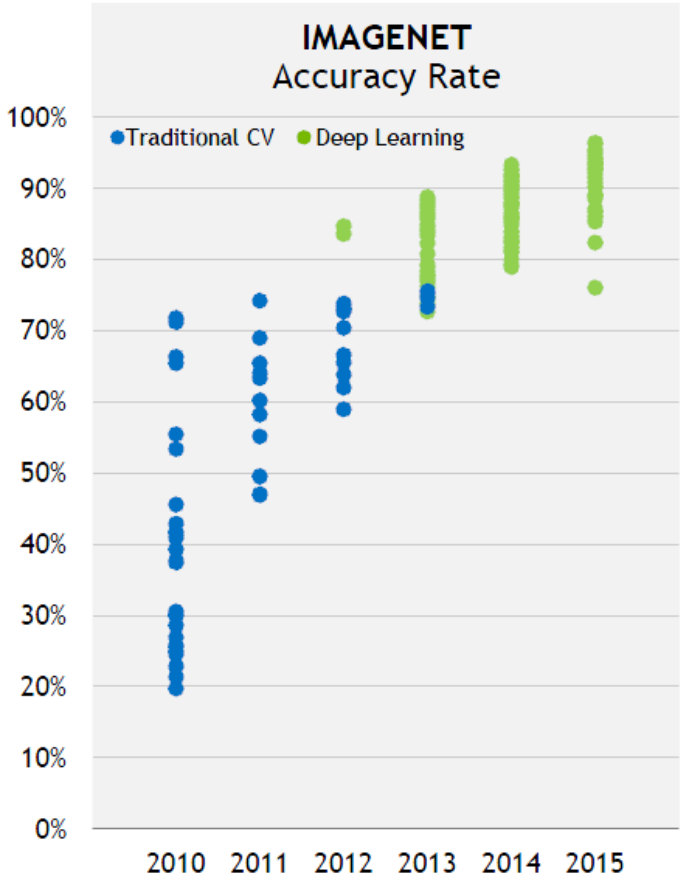
Deep Learning

- Advanced machine learning technique
- Higher accuracy with more data
- Successful in solving vision-based problems, and in general problems with lots of inputs.
- Automates manual feature extraction commonly used for image classification in traditional machine learning
- Breakthrough in image recognition, and natural language processing



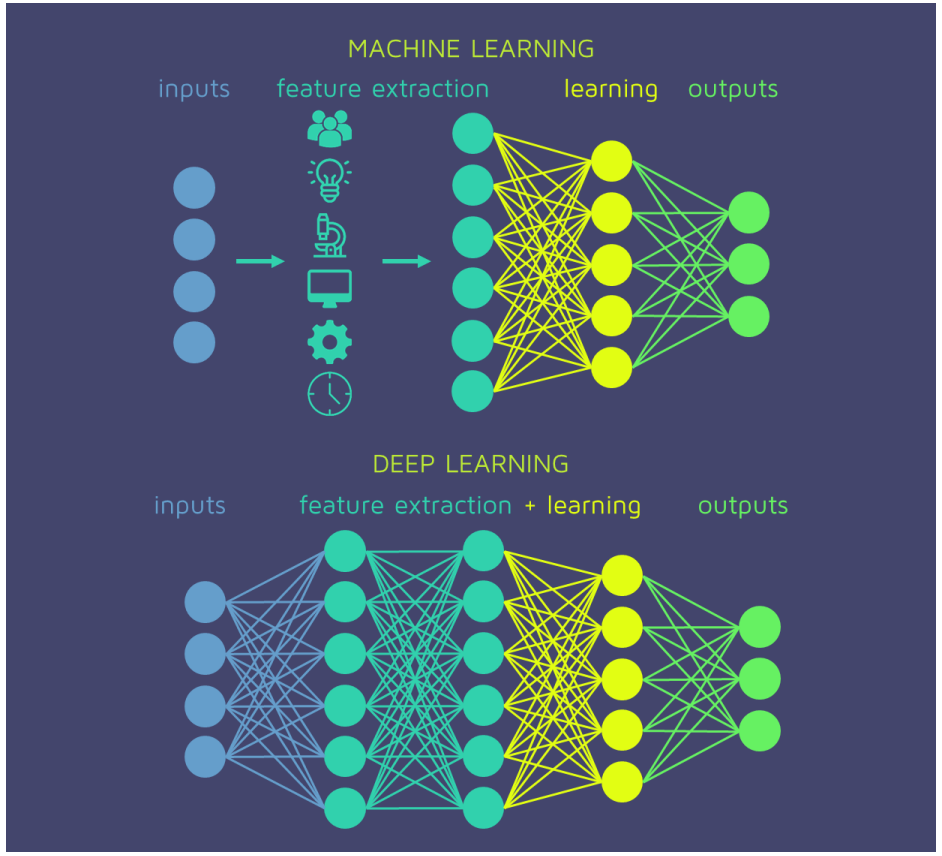
Source: https://machinelearningmastery.com/what-is-deep-learning/

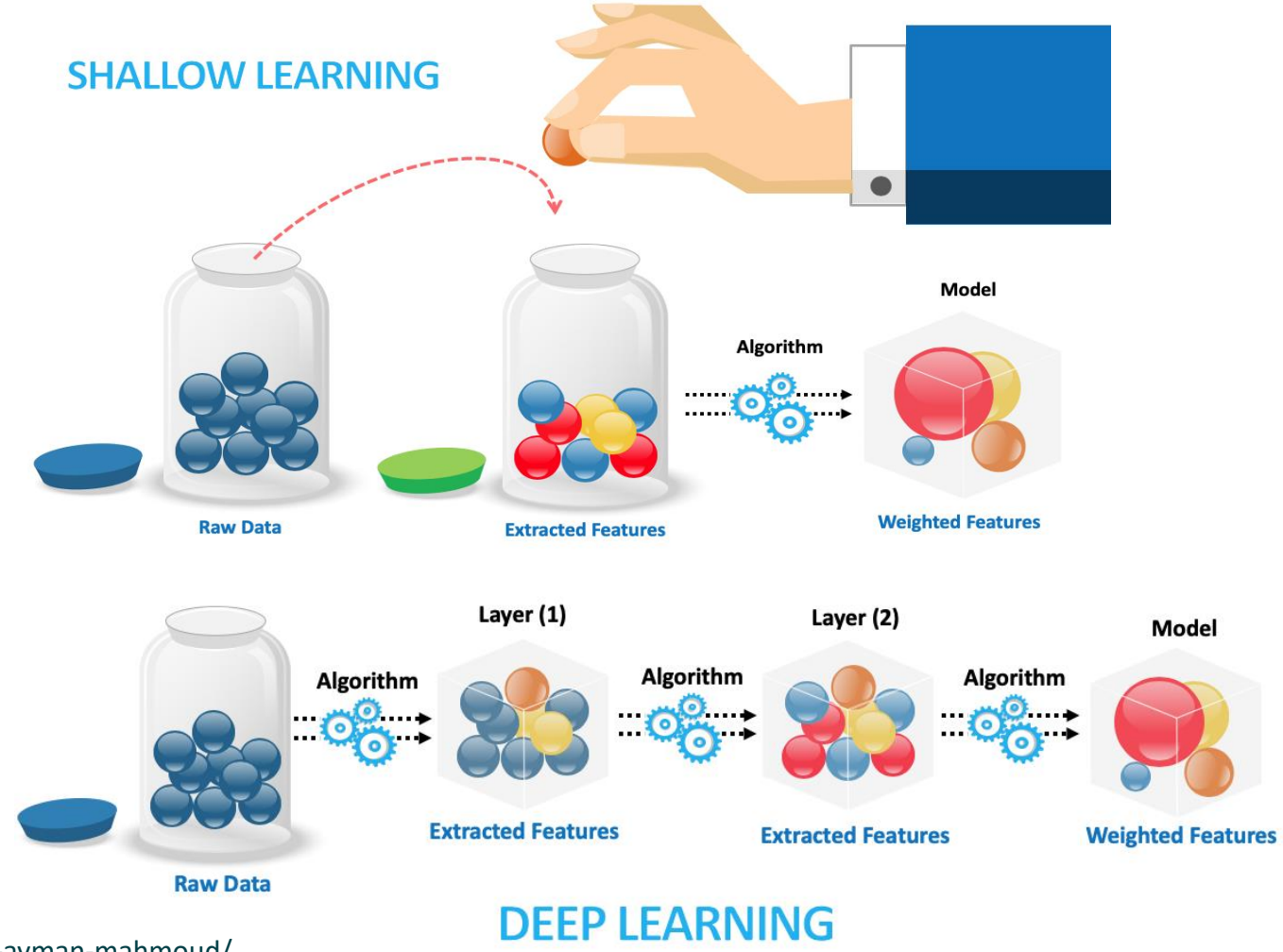
https://www.slideshare.net/NVIDIA/nvidia-ces-2016-press-conference



© 2021 Deep Netts

Shallow vs Deep Learning





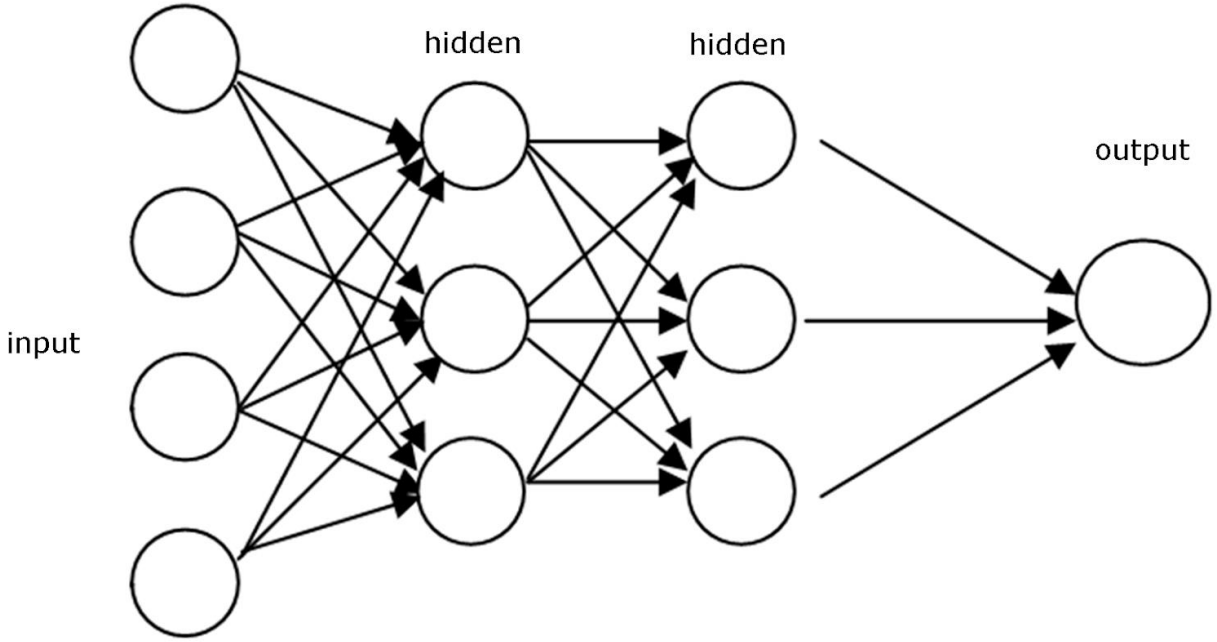
https://www.frontiersin.org/articles/10.3389/fninf.2019.00053/full

https://www.linkedin.com/pulse/introduction-shallow-machine-learning-ayman-mahmoud/

© 2021 Deep Netts

Deep Learning / Neural Networks

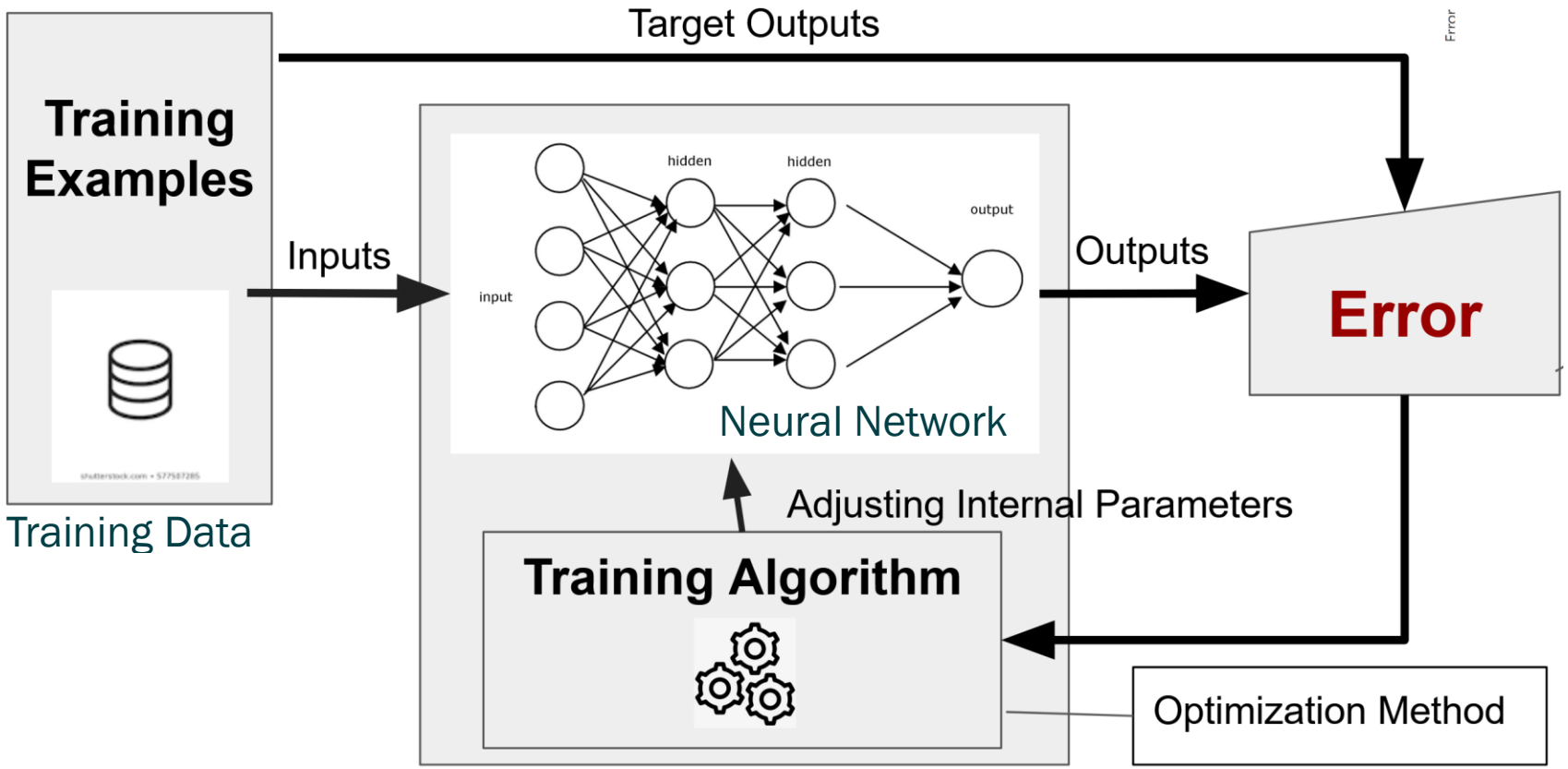
Feed Forward Neural Network



- Nodes are grouped into layers that determine the order of execution/computation.
- Each node performs computation based on its inputs and set of weight coefficients.
- Learning is based on error minimization of total network error for given data set.
- Can be use for both classification and regression tasks

© 2021 Deep Netts

Supervised Training Procedure



Components for training a network:

- Network
- Training data
- Optimizer
- Testing data

© 2021 Deep Netts

From Image Classification to Object Detection

- **Image Classification** – assign an image into one of many different categories
- **Object Localization** – determine the location of a single object within image
- **Object Detection** – localizing and classifying multiple objects within image

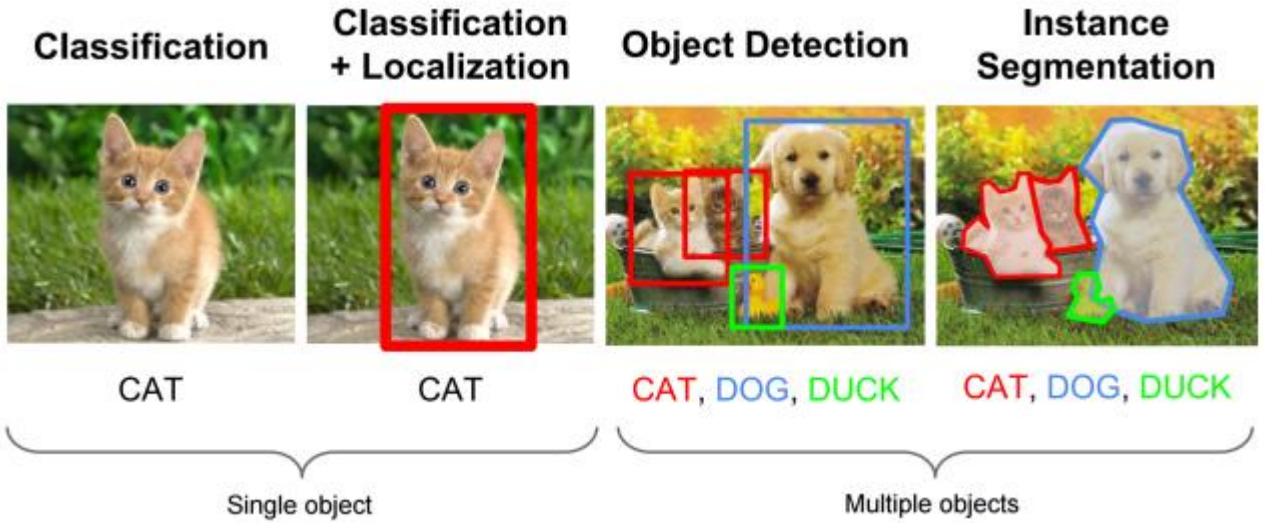
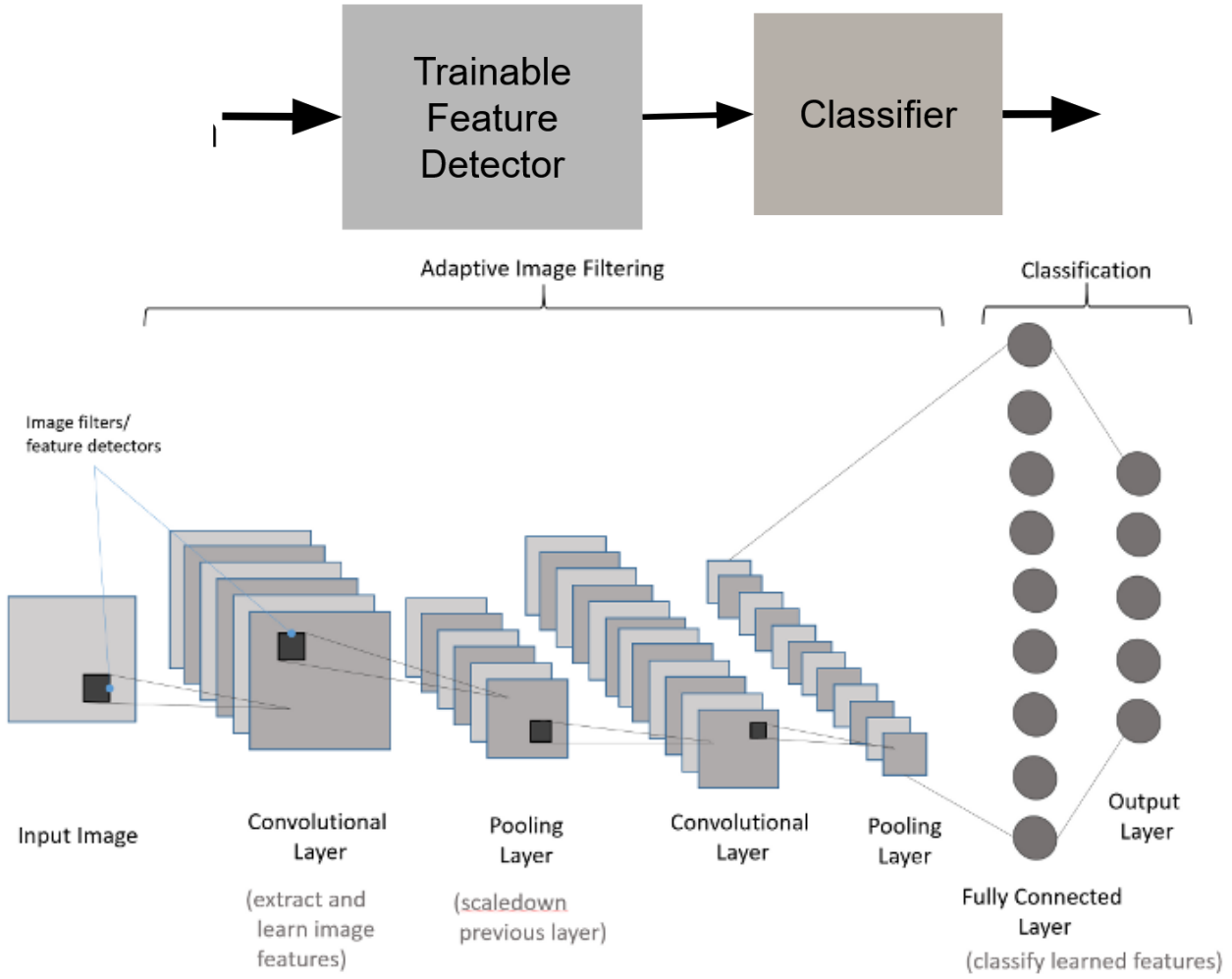


Image Source: https://leonardoaraujosantos.gitbook.io

© 2021 Deep Netts

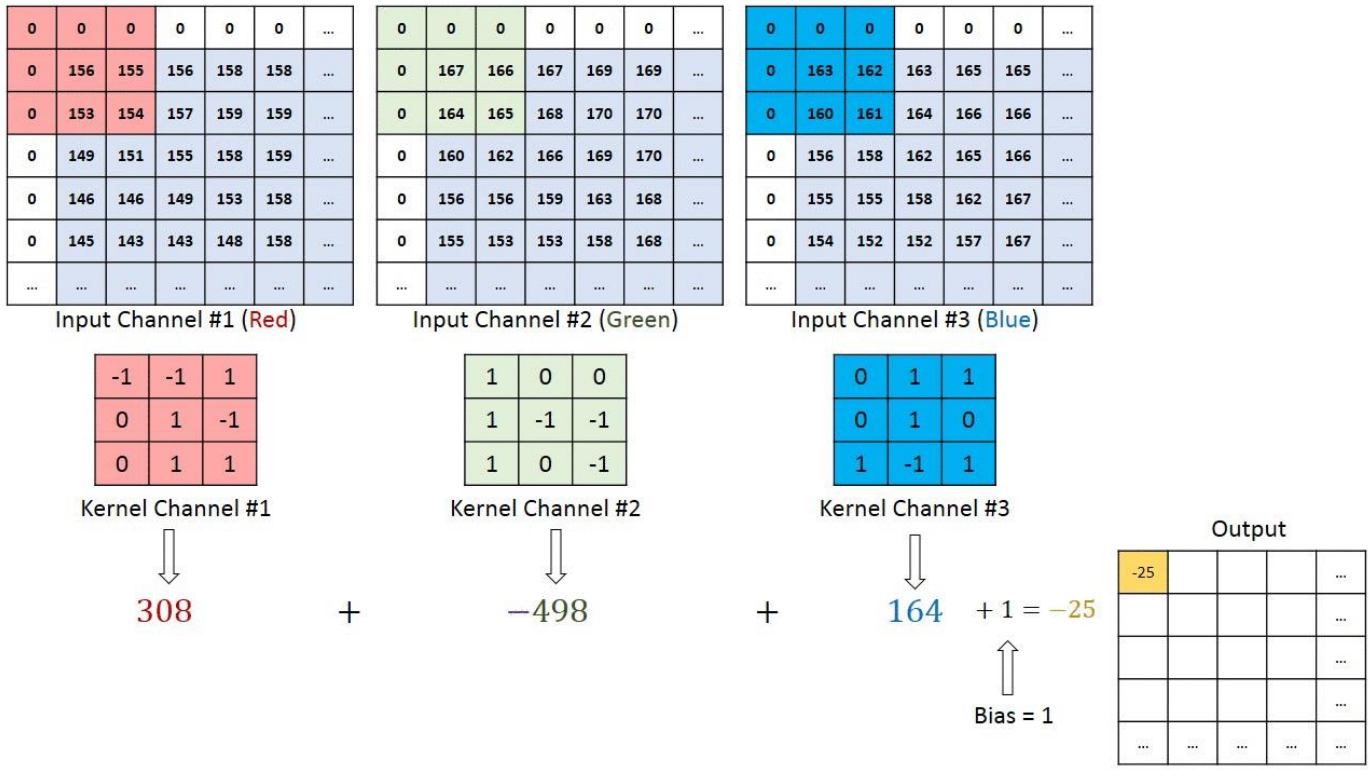
Convolutional Neural Networks and Image Classification

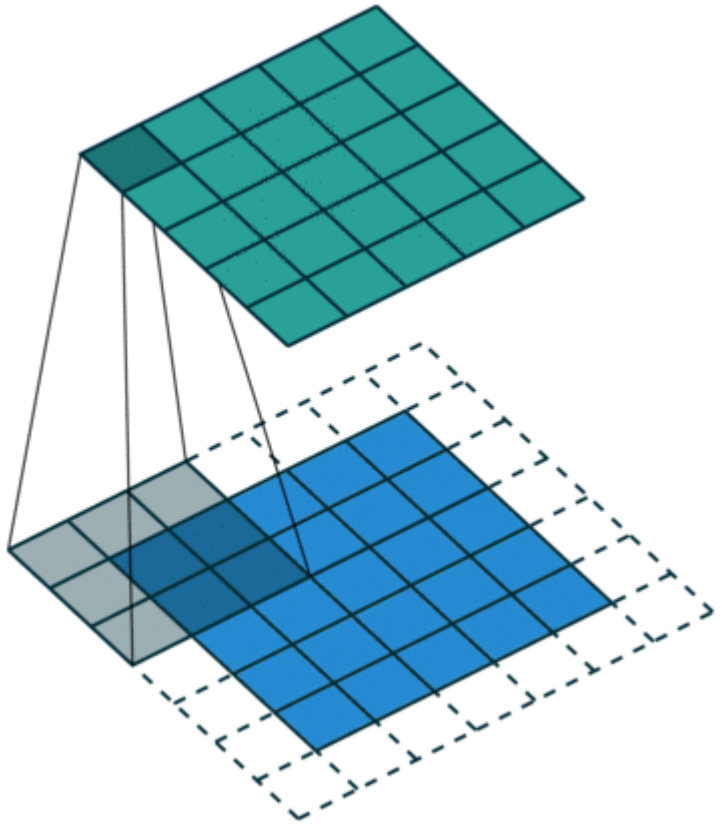
- Extension of a feed forward network specialized for image classification/recognition
- Introduces convolutional layers with adaptive image filters capable for learning and detecting shape and color features
- Reduces image preprocessing and feature extraction – now it's learned during the training



© 2021 Deep Netts

Convolution - the Magic of Convolutional Layer





Convolution operation is sliding small matrix (filter) over input pixels

Image source: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

© 2021 Deep Netts

Types of Layers in CNN

- Convolutional Layer – performs feature detection
- Max Pooling Layer – scales down input images
- Fully Connected Layer – performs classification
- Softmax Layer – estimates probability for categories
- Linear Layer – performs regression, estimates object position
- Dropout Layer – reduces overfitting

© 2021 Deep Netts

Multiple Layers Combined : Image Classification Using CNN

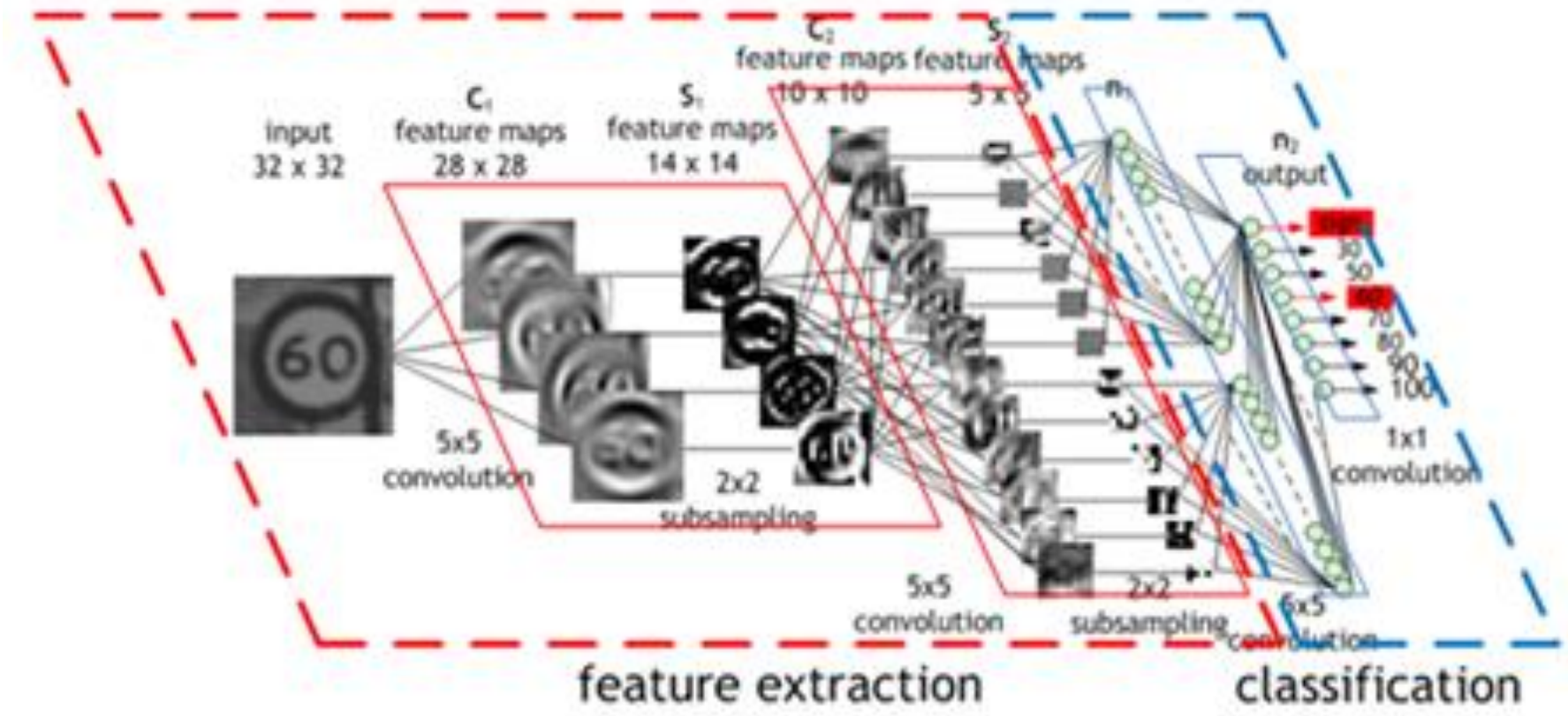
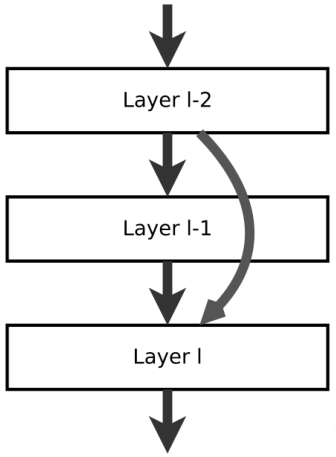


Image source: https://developer.nvidia.com/discover/convolutional-neural-network

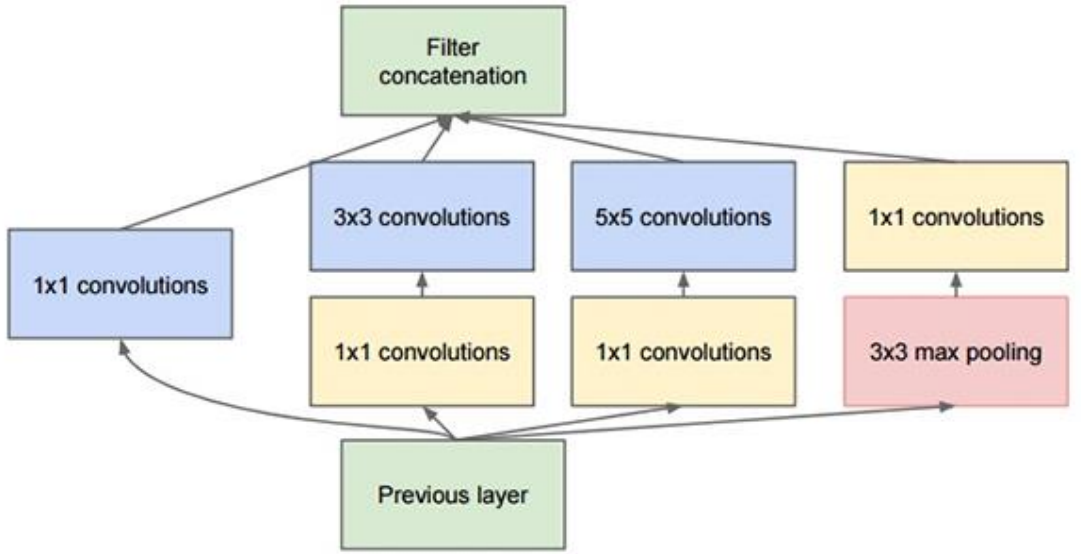
© 2021 Deep Netts

Common CNN Architectures

- VGG Net: 16 or 19 layers, 3x3 conv filters
- ResNet: skip connections enable deeper networks
- InceptionNet: concatenates filters of different sizes 3x3, 5x5, 1x1
- Mobile Net: depthwise separable convolutions, and thinner more efficient networks



Residual/skip connections



Filter concatenation

© 2021 Deep Netts

Quick Overview of the Evolution of Object Detection

- Convolutional Network for Object Detection
  - Region-based Convolutional Neural Network (RCNN)
  - Fast RCNN
  - Faster RCNN
  - You Look Only Once - YOLO
  - Single Shot Detector – SSD

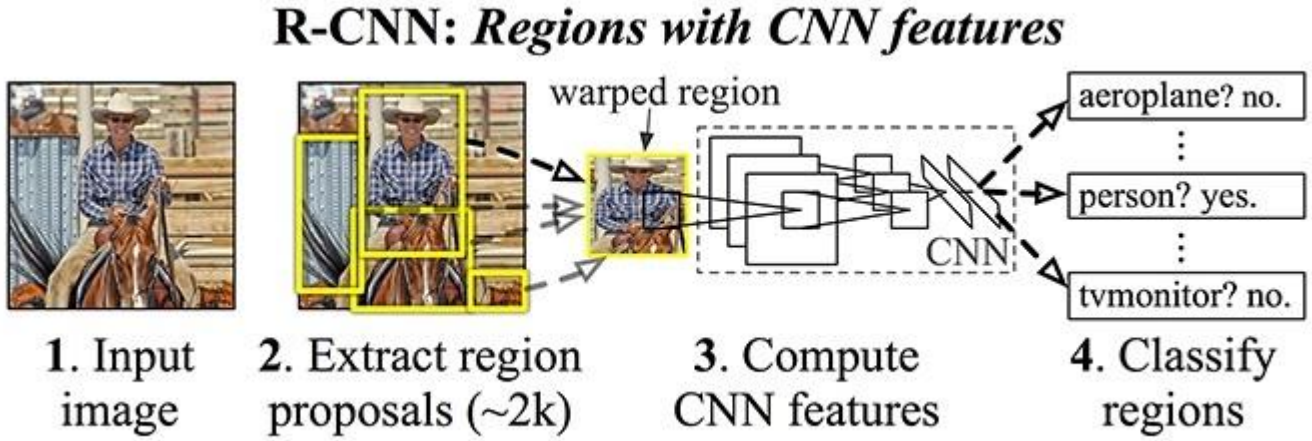
- Convolutional network for image classification +

  Regression for determining the position

© 2021 Deep Netts

Region-based Convolutional Neural Network

- Generate about 2000 **region proposals** from input image using the Selective Search
- Run a convolutional net to classify each **region proposal**
- Intuitive but slow



© 2021 Deep Netts

Fast R-CNN

- In order to reduce computation region proposals are generated from the last convolutional layer, not from the original image.
- The computationally expensive convolutional layers are calculated only once.
- Still using selective search for region proposals

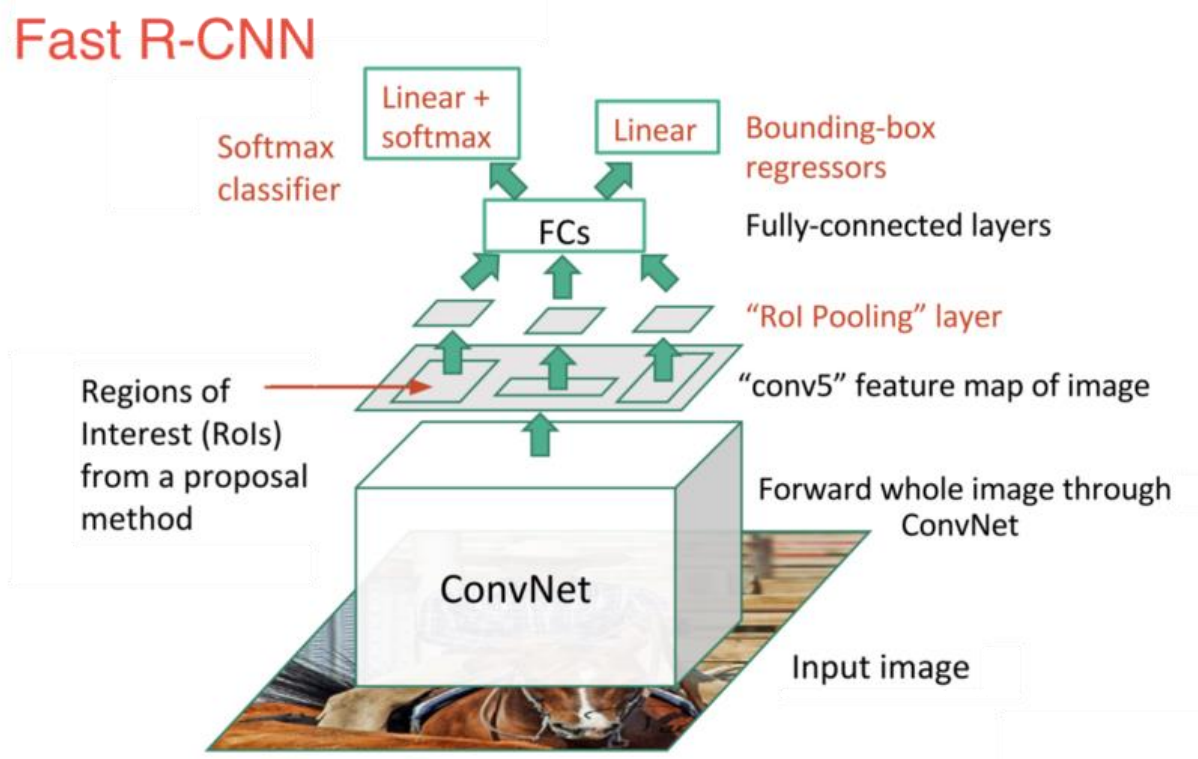
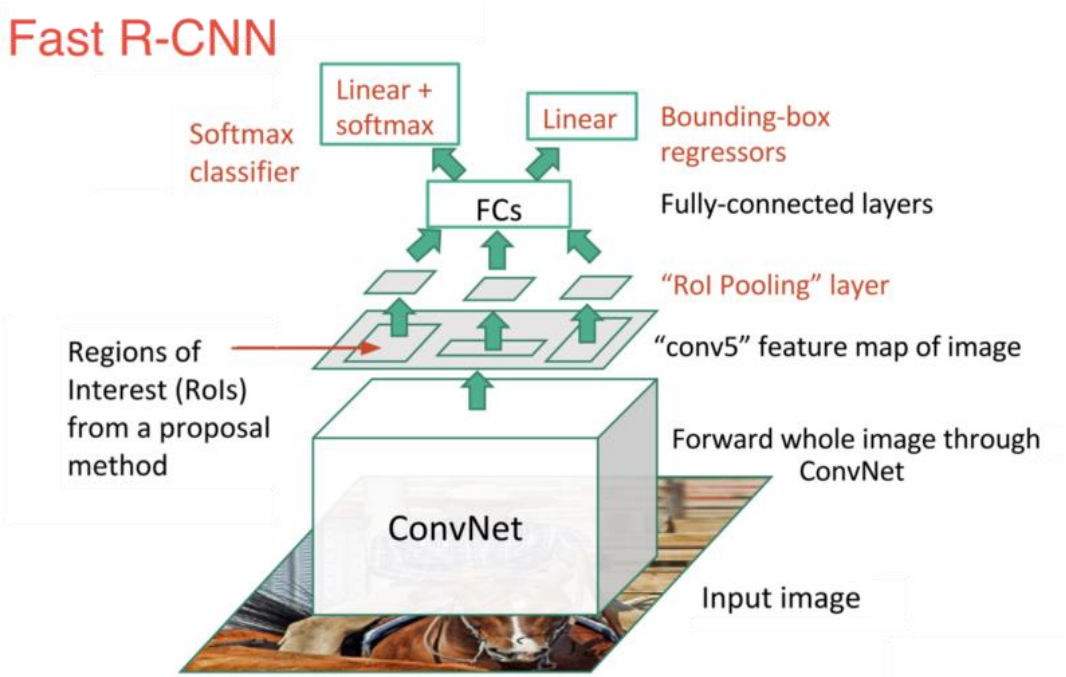


Image source: https://www.kdnuggets.com/2017/10/deep-learning-object-detection-comprehensive-review.html

© 2021 Deep Netts

Faster RCNN

- Introduced region proposal network instead of selective search: does the region contain object?



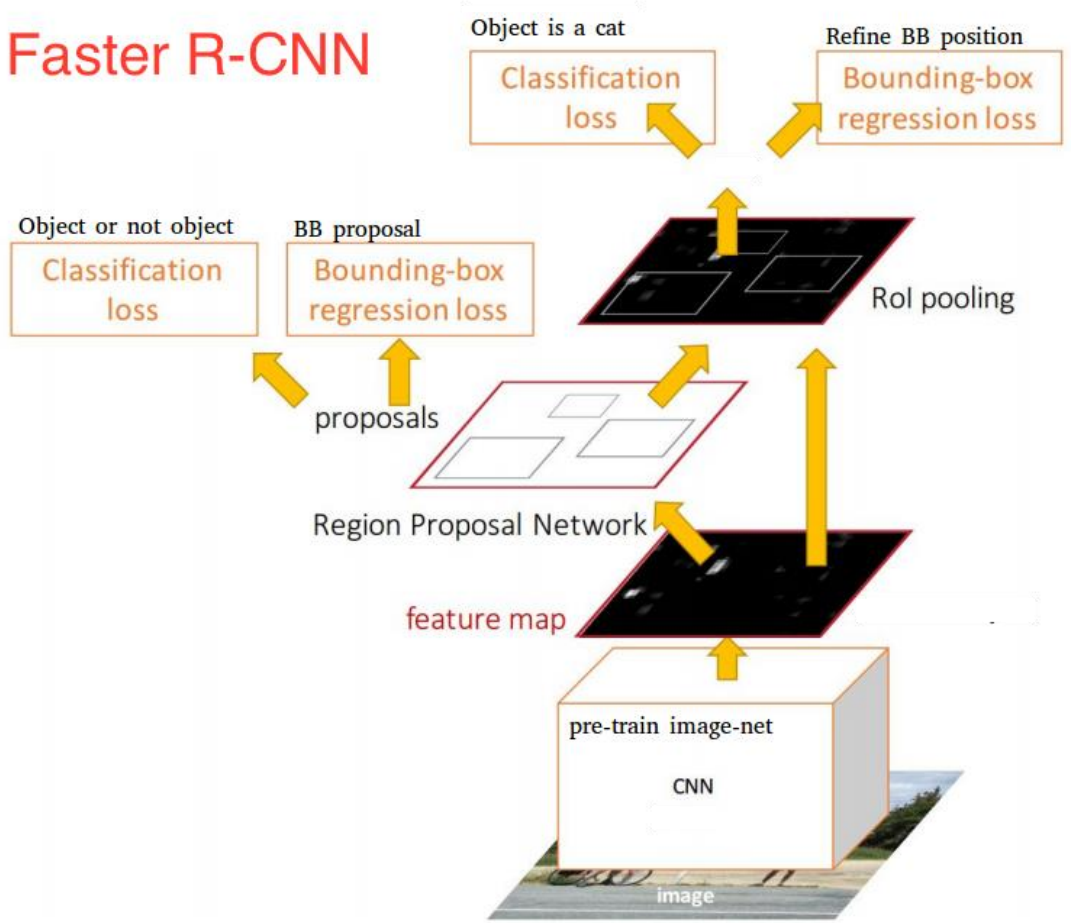
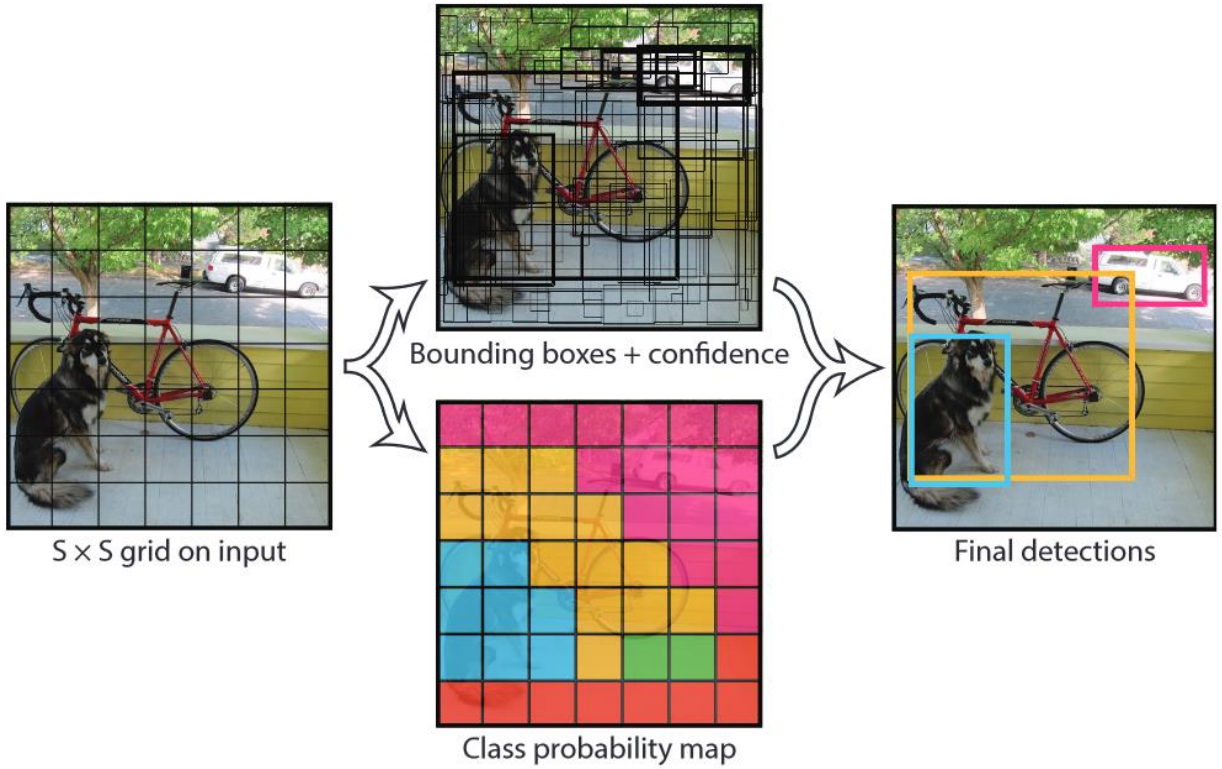


Image source: https://www.kdnuggets.com/2017/10/deep-learning-object-detection-comprehensive-review.html

© 2021 Deep Netts

YOLO – You Look Only Once

- Divides input image into SxS grid
- Each grid cell predicts a specific number of bounding boxes (B) and confidence scores for those boxes.
- If the center of an object falls into a grid cell, that grid cell is responsible for detecting that object.
- Used pretrained convolutional features from (typically ImageNet)
- Faster, and end-to-end training without region proposals, compared to faster RCNN
- Has problems with detecting small grouped objects

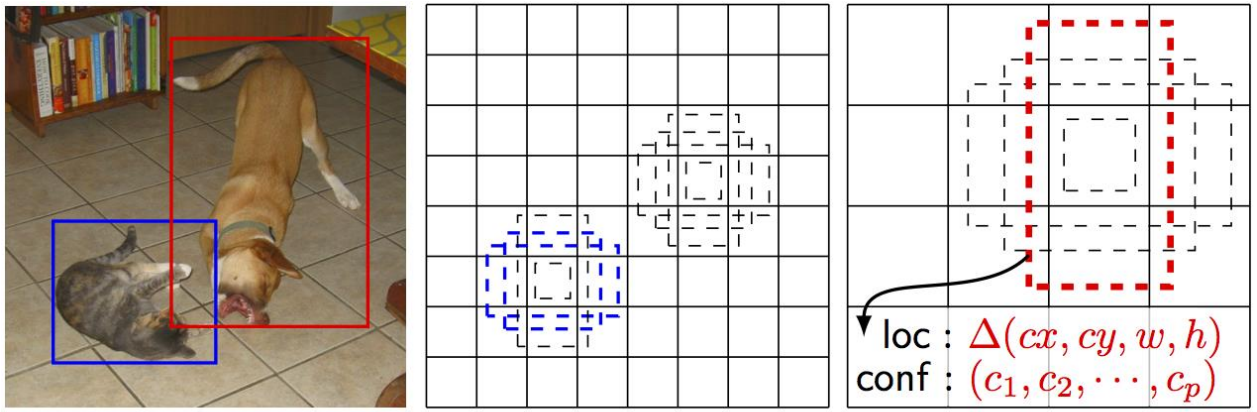


See more at: https://pjreddie.com/media/files/papers/yolo_1.pdf

© 2021 Deep Netts

Single Shot Detector - SSD

- Uses a fixed set of default bounding boxes with small convolutional filters
- For each box, both the shape position and the confidences for all categories are predicted.
- Uses convolutional network (VGG-16, ResNet, MobileNet) as feature extractor, and additional conv layers for object detection
- Improves speed vs accuracy tradeoff
- Compared to YOLO it is faster but less accurate for some problems



(a) Image with GT boxes (b) 8×8 feature map (c) 4×4 feature map

© 2021 Deep Netts

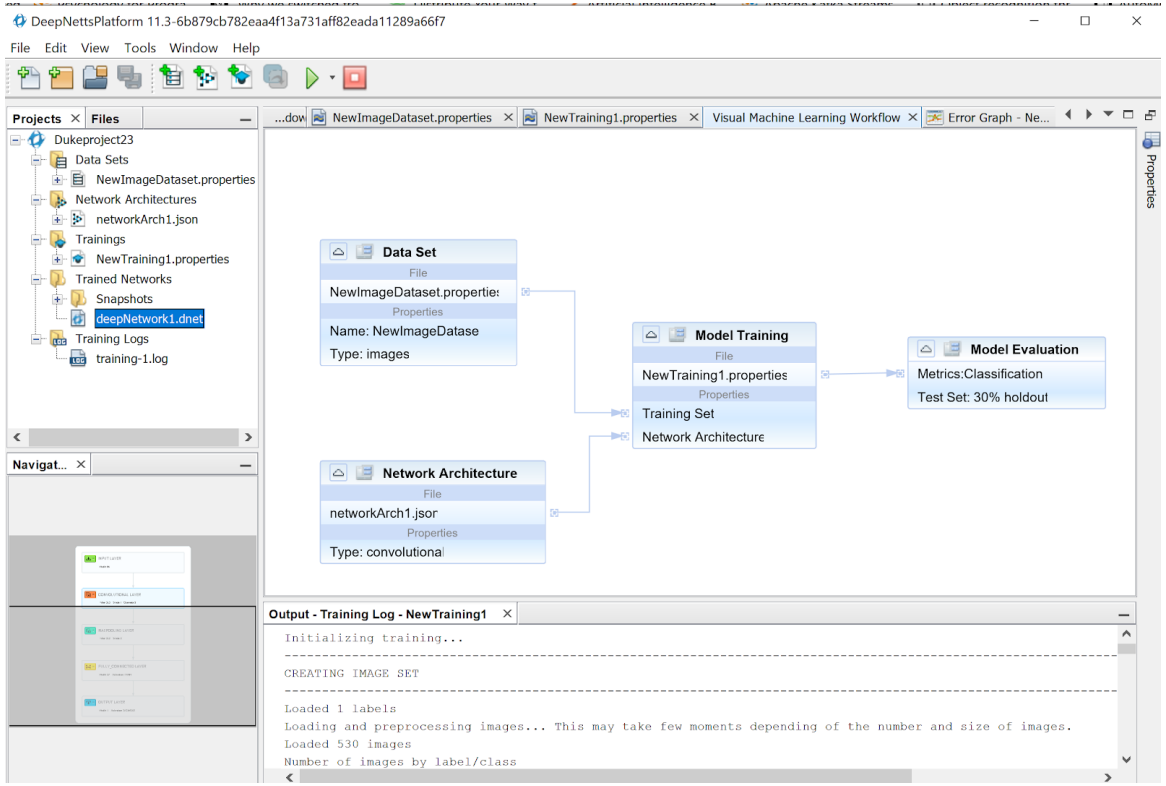
Summary

- Deep learning is enabling modern computer vision
- The fundamental models to do this are convolutional neural networks
- Image classification is basic computer vision
- Image classification can be extended to object detection
- Tensorflow provides various implementations and pretrained convolutional networks that can be used for computer vision problems

© 2021 Deep Netts

About Deep Netts

- Deep Learning Made Easy for Non-Experts
- Deep Learning IDE and Java Library
- Free For Development
  https://www.deepnetts.com/download



- Easy to use, integrate and maintain (lower cost)
- Good accuracy with less data
- No requirements for specialized hardware (GPU)
- Highly portable

© 2021 Deep Netts

Resources to get started

Deep Learning and image recognition book

https://leonardoaraujosantos.gitbook.io/artificial-inteligence/machine_learning/deep_learning/object_localization_and_detection

Deep Learning for Object Detection: A Comprehensive Review https://www.kdnuggets.com/2017/10/deep-learning-object-detection-comprehensive-review.html

Image classification with Convolutional Neural Network in Tensorflow

https://www.tensorflow.org/tutorials/images/cnn

Tensorflow Object Detection API

https://github.com/tensorflow/models/tree/master/research/object_detection

© 2021 Deep Netts