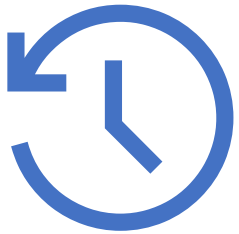
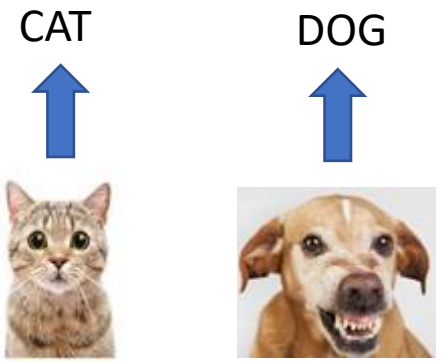


Brief History of Modern Neural Networks

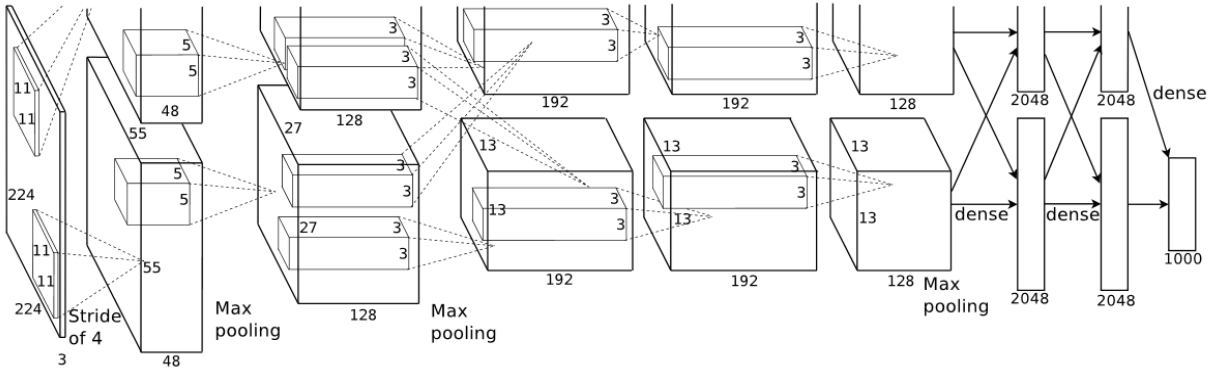
Ing. Marek Hrúz PhD.

Image Classification

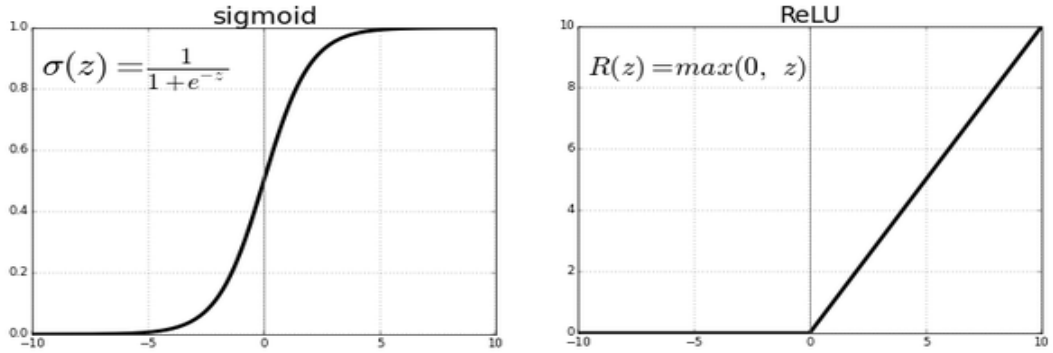


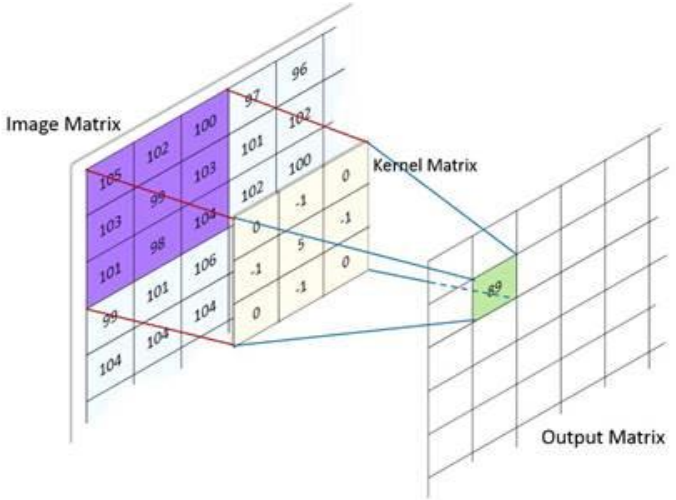
- Taks of assigning a class label to an image
- Historically, people described the images by handcrafted features and trained a classifier
- Works reasonably well for small, non-complex, well-defined data
- To allow a next big leap in computer vision a new dataset was developed
- ImageNet - an image dataset organized according to the WordNet hierarchy
  - Total number of non-empty synsets: 21,841
  - Total number of images: 14,197,122
  - Number of images with bounding box annotations: 1,034,908
- SIFT + FVs in 2011 achieved Top-1 accuracy of 50.9% (1000 classes)

AlexNet 2012



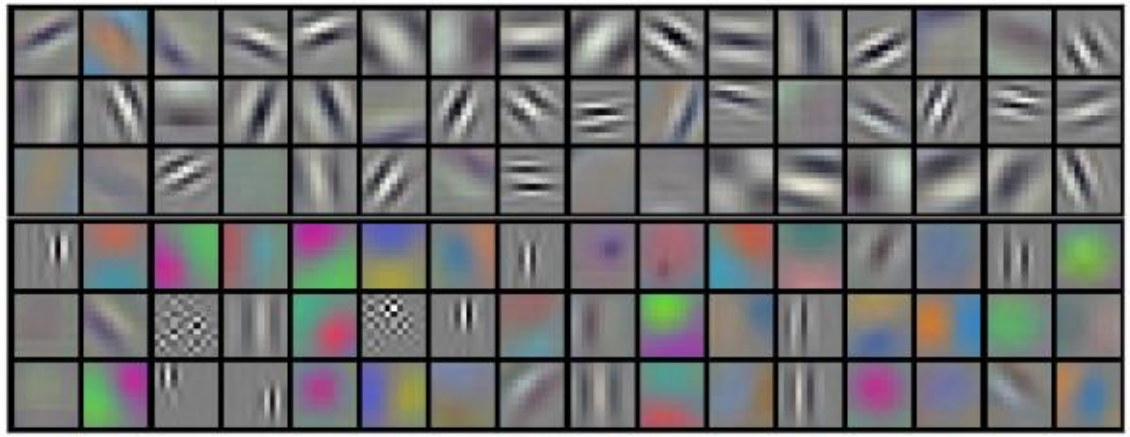
- Introduces Convolutional Neural Networks to the task of ImageNet classification
- The CNN is distributed on 2 GPUs
- Main features:
  - Novel, deep CNN architecture
  - ReLU non-linearity
  - Overlapping max-pooling
  - Data augmentation for overfitting reduction
  - Dropout
- Achieves Top-1 accuracy of 63.3%



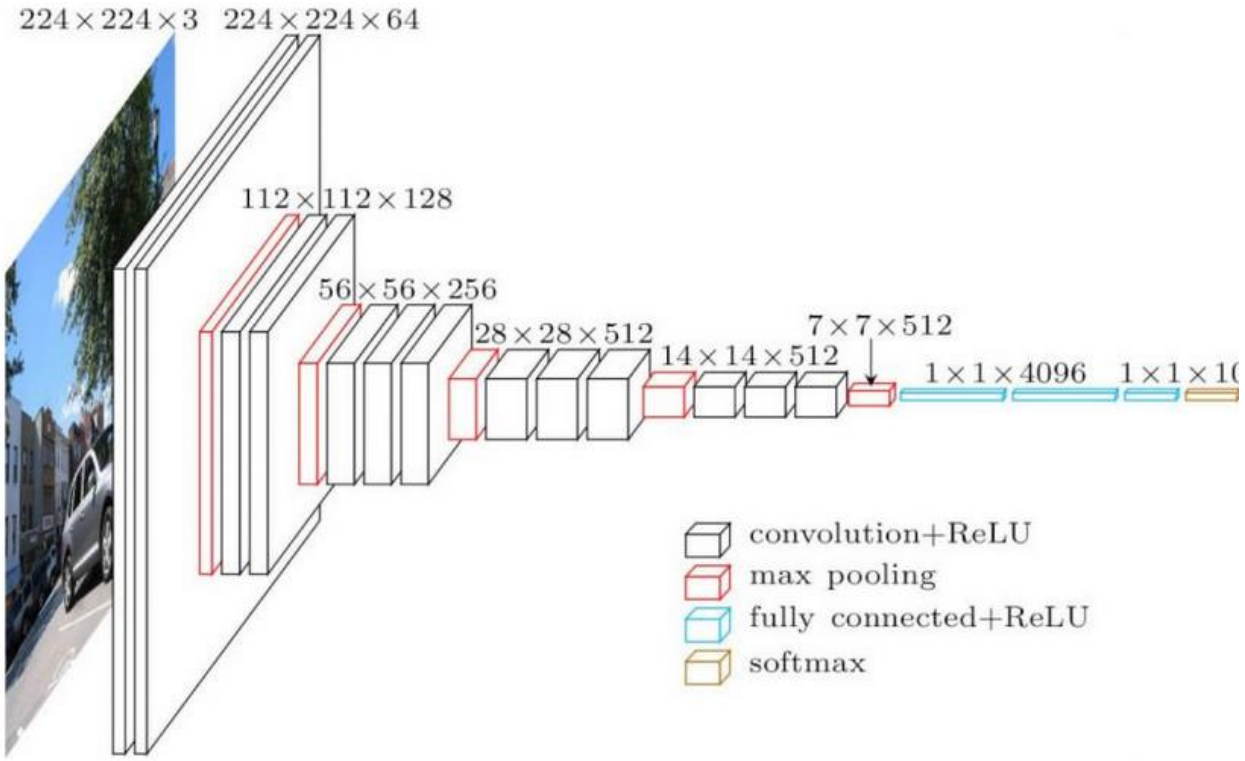


CNN implications

- Convolutional filters can be visualized and explained
- The shift from algorithm design to data preparation
- Semantic representation in deep layers



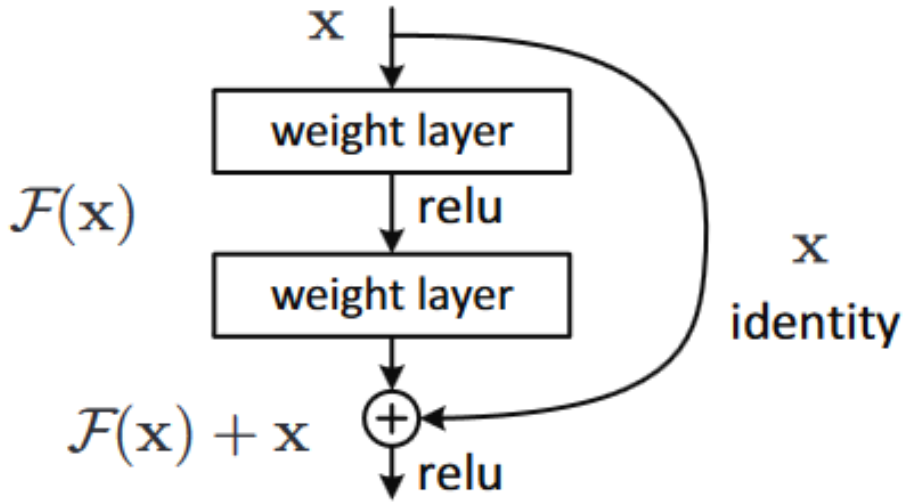
VGG 2014

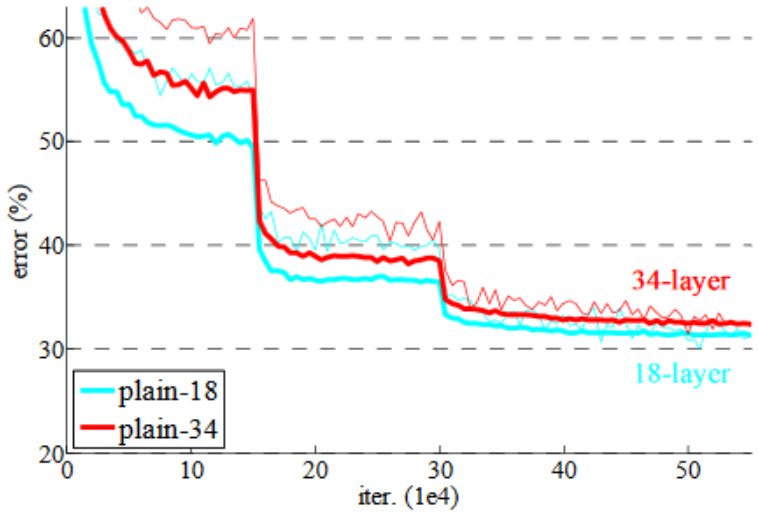


- From Google
- Scheme for deeper models
- Small sizes of convolution kernels (3x3 instead of 11x11)
- Back to non-overlapping pooling
- Achieves Top-1 accuracy of 74.4%

ResNet 2015

- Addresses the problem of **vanishing** and **exploding** gradients
- Introduces residual (skip) connections
- They allow a better flow of the gradient
- Achieves Top-1 accuracy of 78.6%





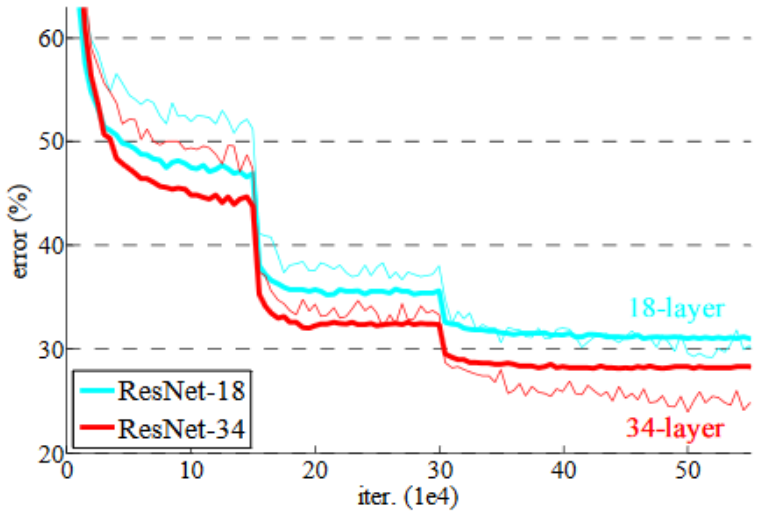
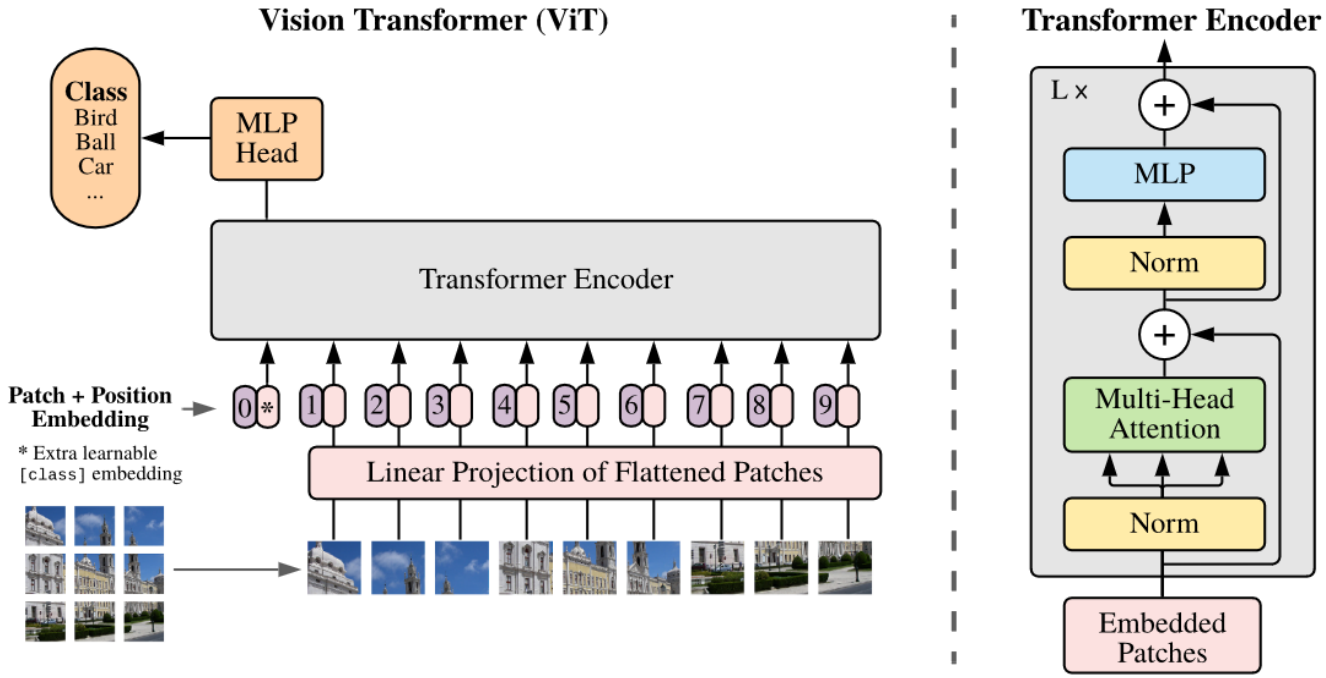


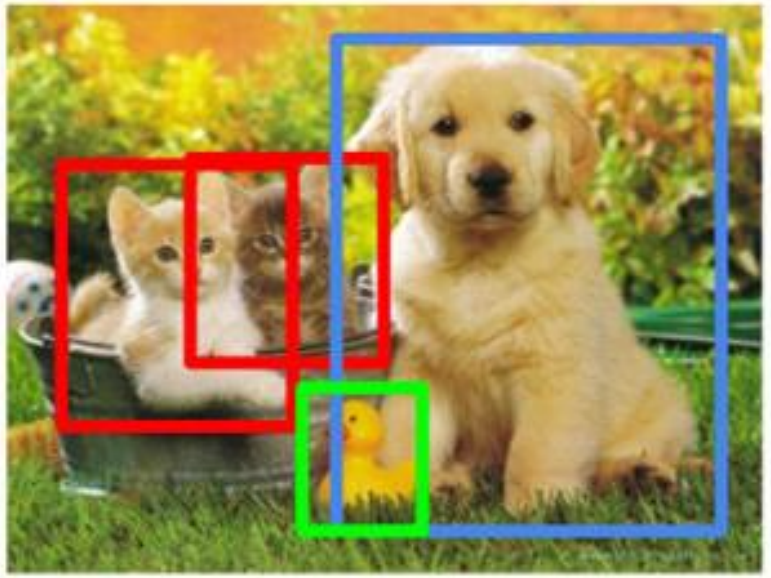
Image Classification in 2020

- In 4 years of time the Top-1 acc increased from 50.9% to 78.6%
- View results on paperswithcode.com
- Best result in 2020 produced by a CNN is 88.5% accuracy
- Transformers are overtaking?
- Top-1 acc 88.55%
- Tensorflow playground
- Keras
- PyTorch



Object Detection

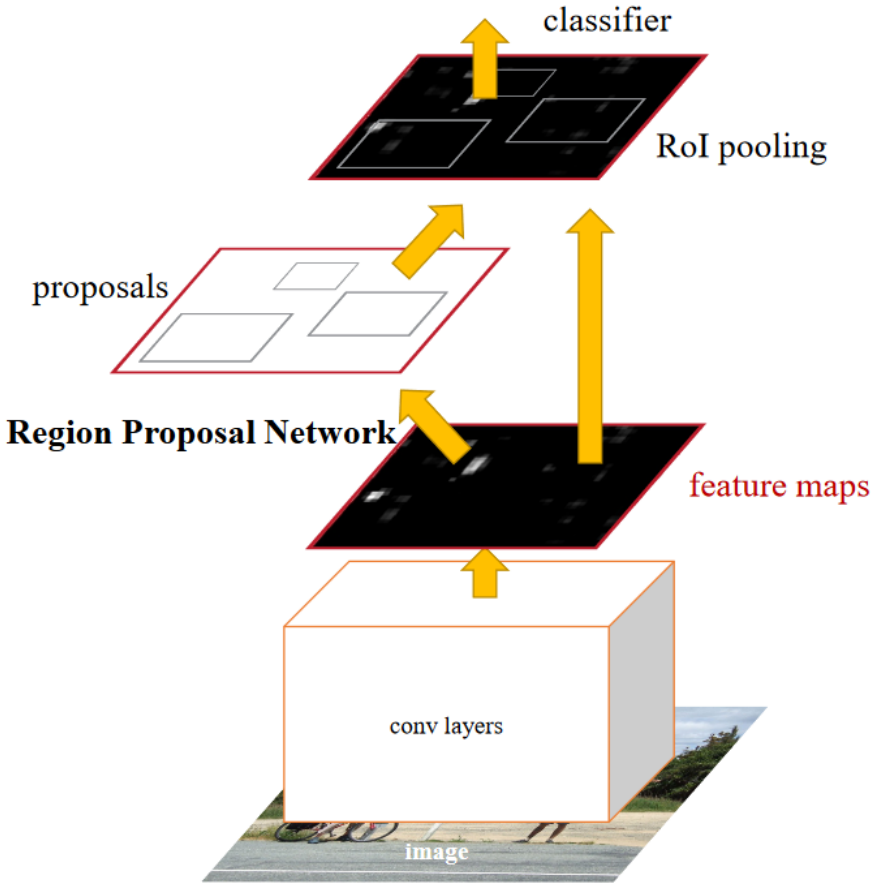
- Task of recognizing and localizing an object
- Historically, two stages:
  - Region proposal
  - Region classification
- MS COCO
  - Object segmentation
  - Recognition in context
  - Superpixel stuff segmentation
  - 330K images (>200K labeled)
  - 1.5 million object instances
  - 80 object categories
  - 91 stuff categories
  - 5 captions per image
  - 250,000 people with keypoints
- Pascal VOC
  - 9,993 annotated images

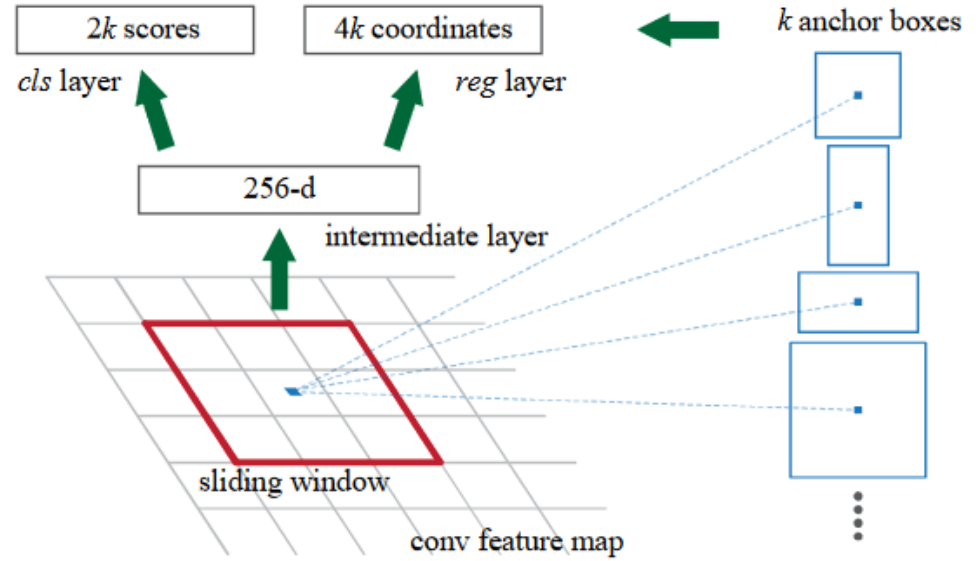


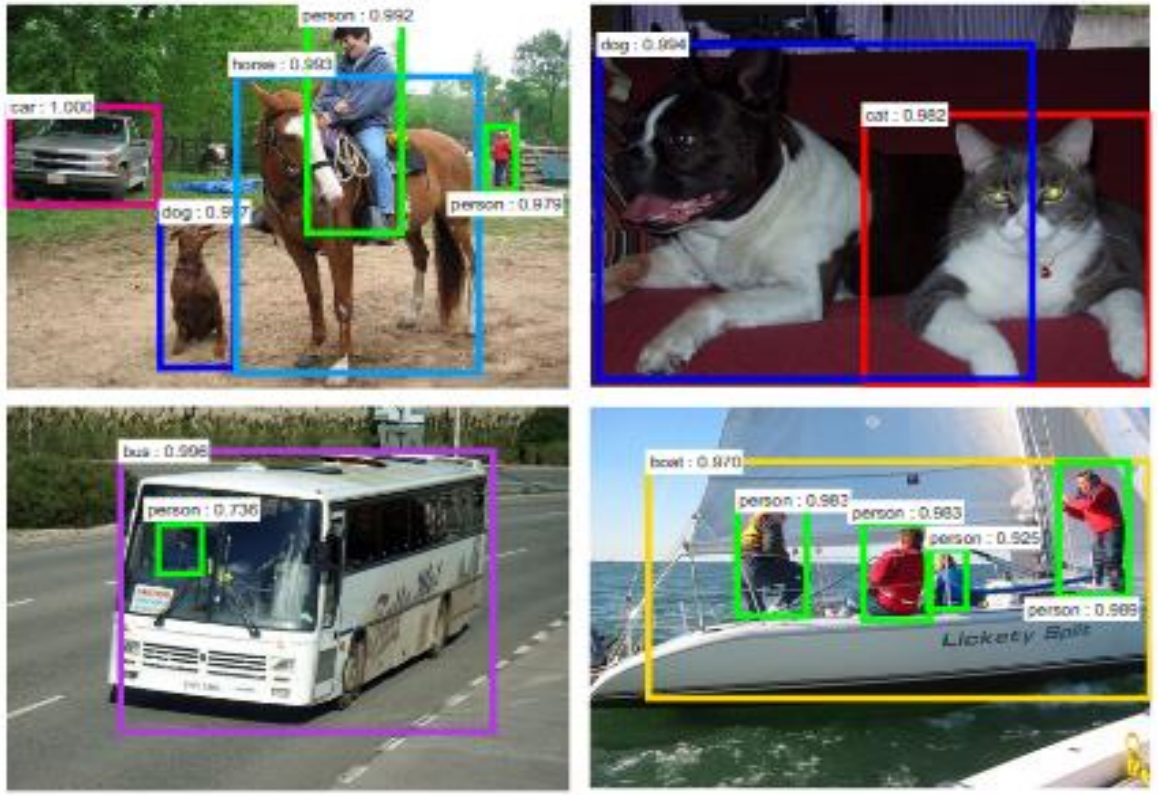
CAT, DOG, DUCK

Faster R-CNN 2015

mAP@0.5 = 42.7% for COCO

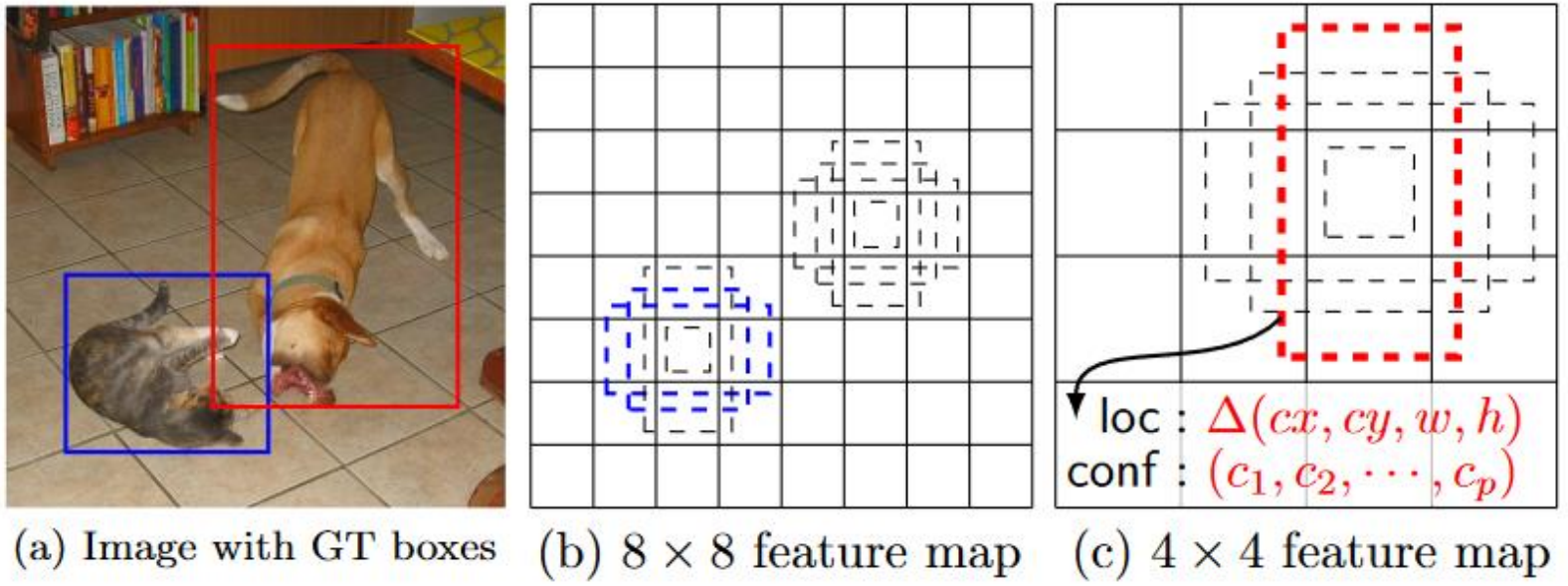


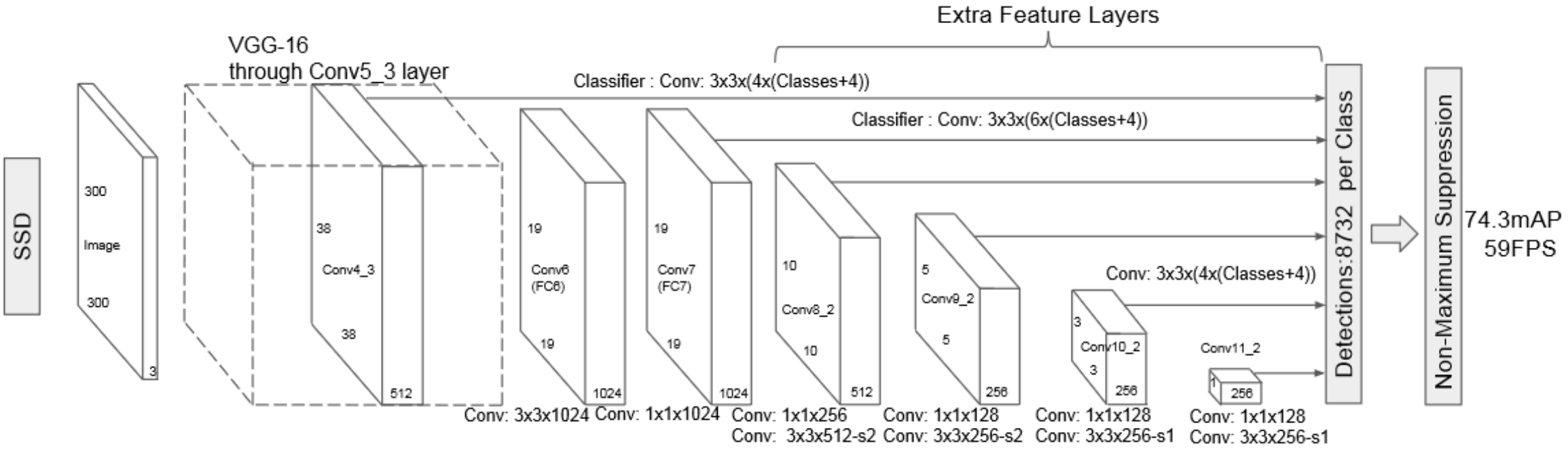




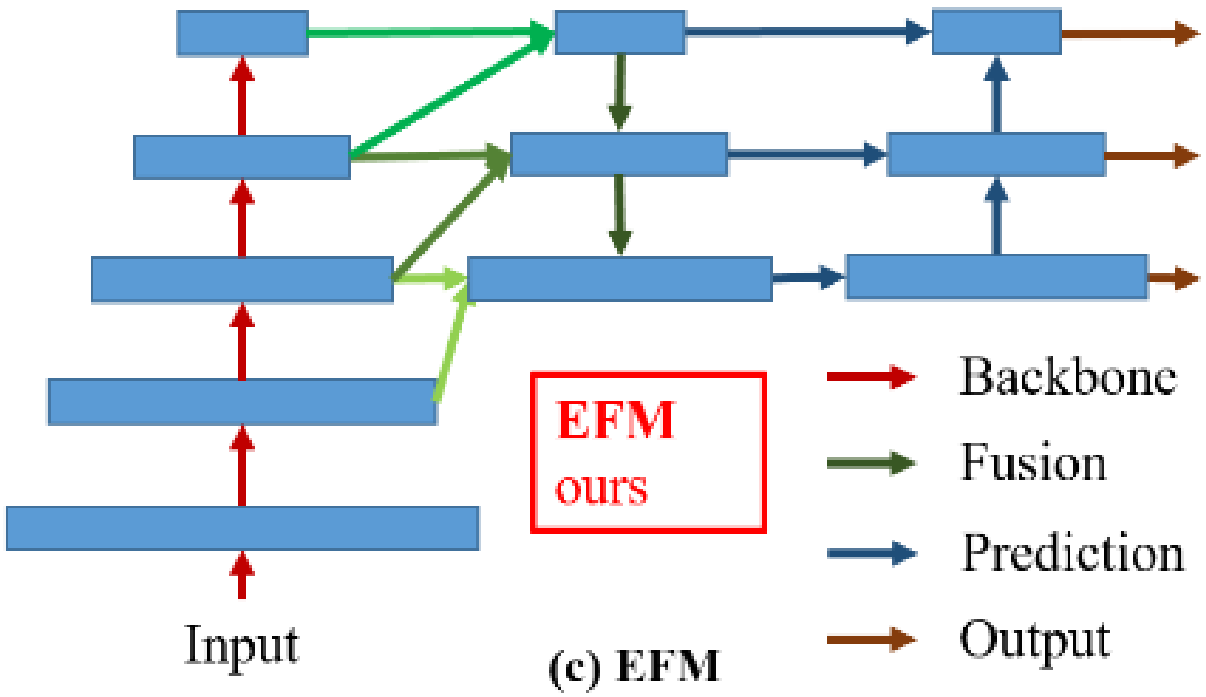
SSD early 2016

mAP@0.5 = 46.5% for COCO





Object Detection and Tracking

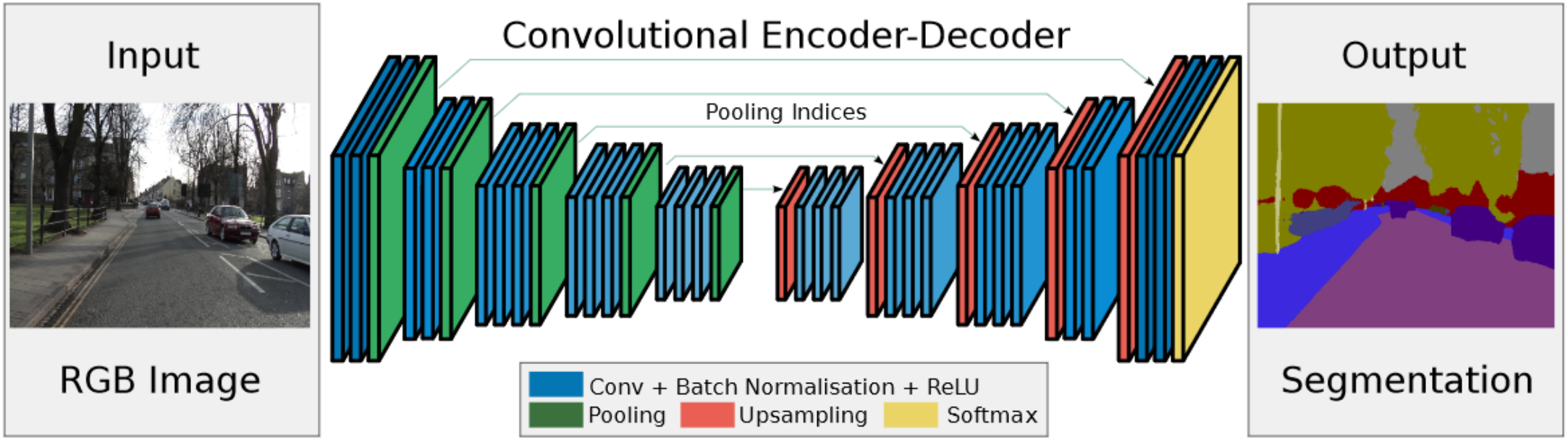


- In 2020 the best performing detector on COCO has **73.2%** mAP@0.5
- Example video of tracking by detection

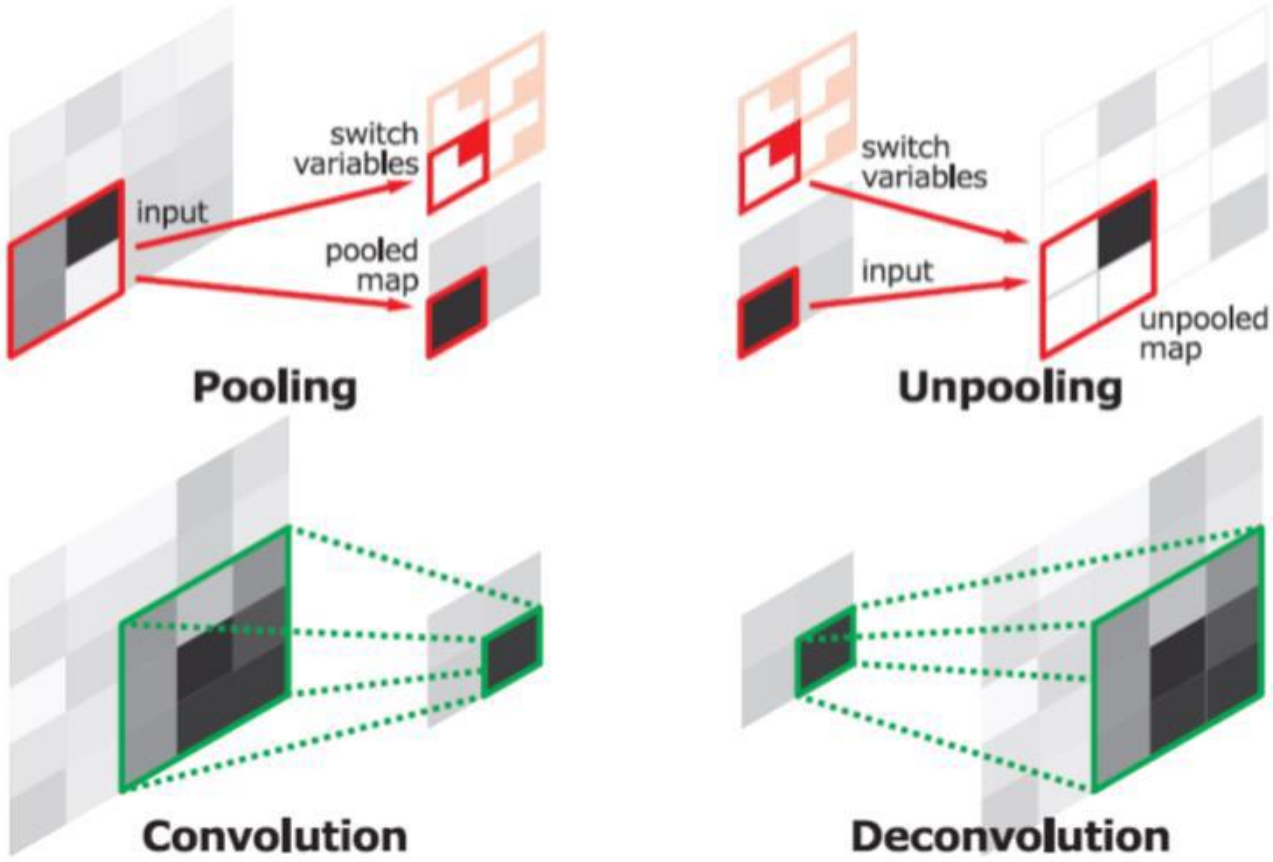
Semantic Segmentation

- CityScapes
- Pascal VOC, MS COCO
- Task of assigning each pixel a class label

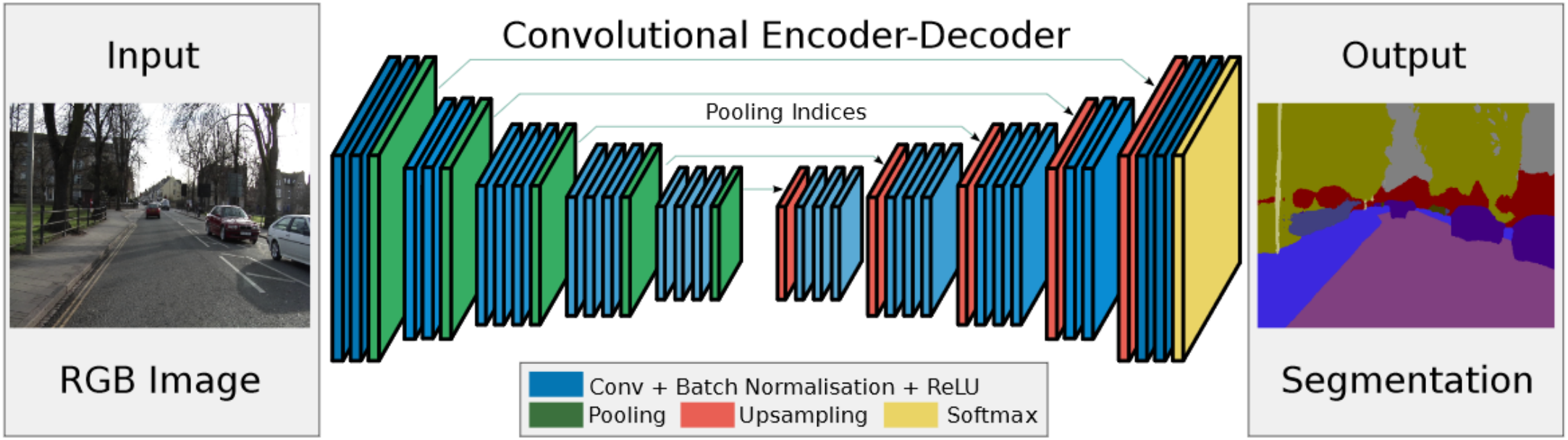
SegNet late 2015



How to upsamle?



SegNet late 2015

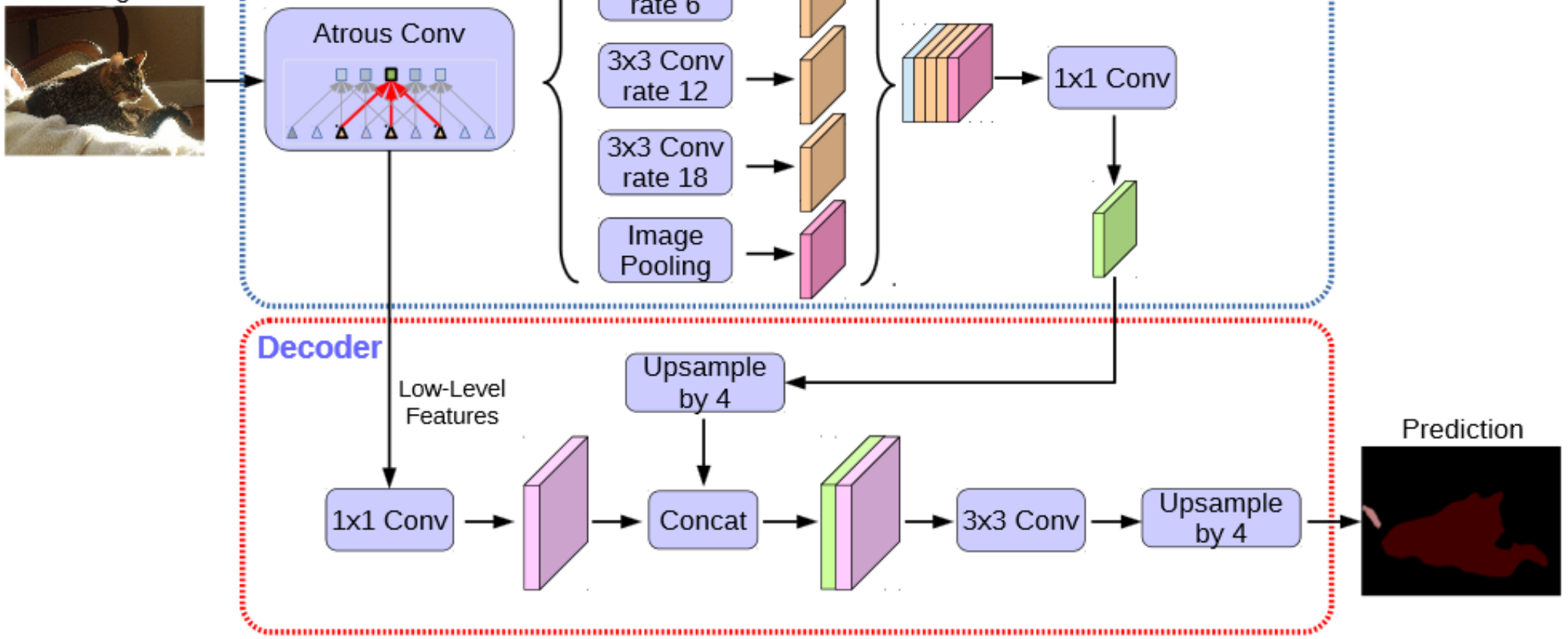


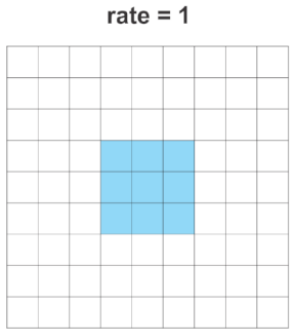
- Online DEMO
- Mean IoU **57.0%** CityScapes

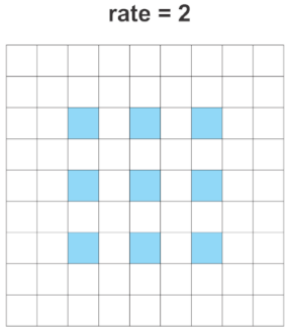
DeepLab 2014 - 2018

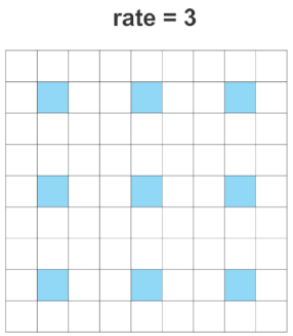
Mean IoU **82.1%**
CityScapes

DEMO



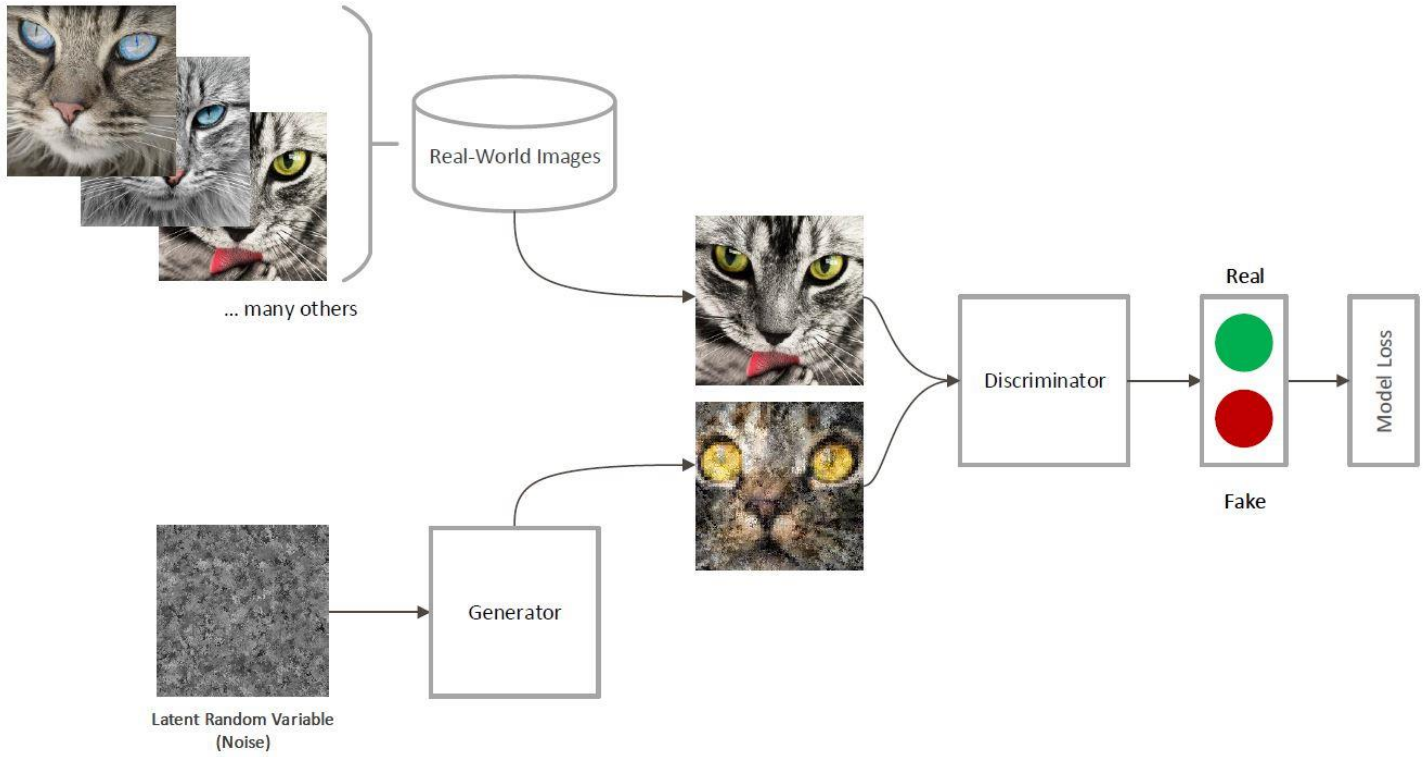






Generative Adversarial Networks 2014

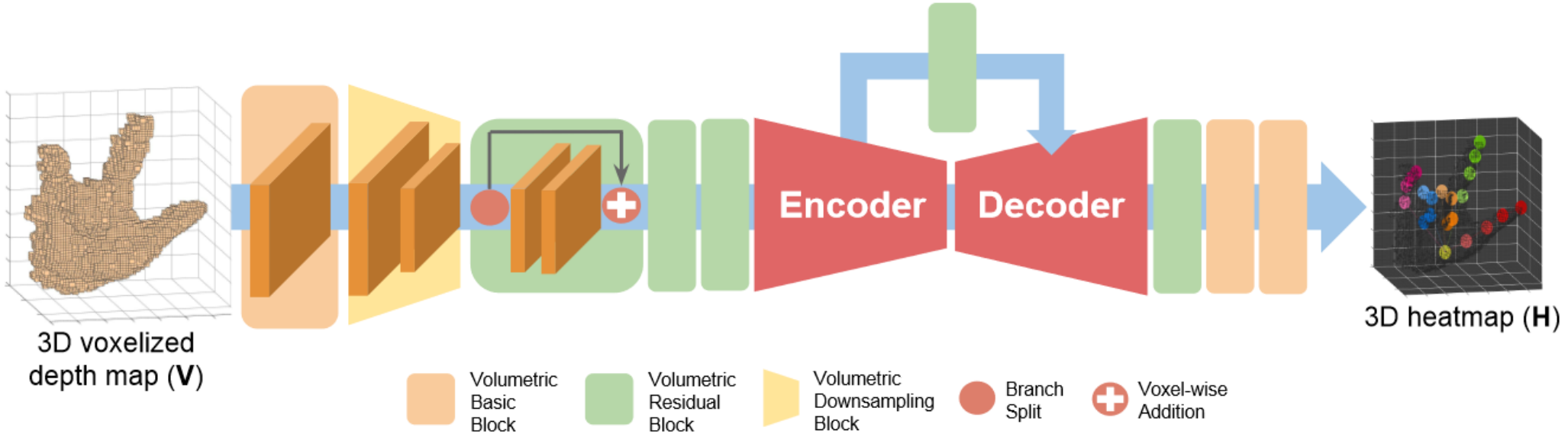
- Models for generating data from noise or other provided input



- http://nvidia-research-mingyuliu.com/gaugan/
- https://www.whichfaceisreal.com/

Regression

- Instead of assingning a class to data, we want to approximate a function
- Given values of *x* and *f(x)*, we train a model of NN
- Hand Pose Estimation



Use Papers With Code