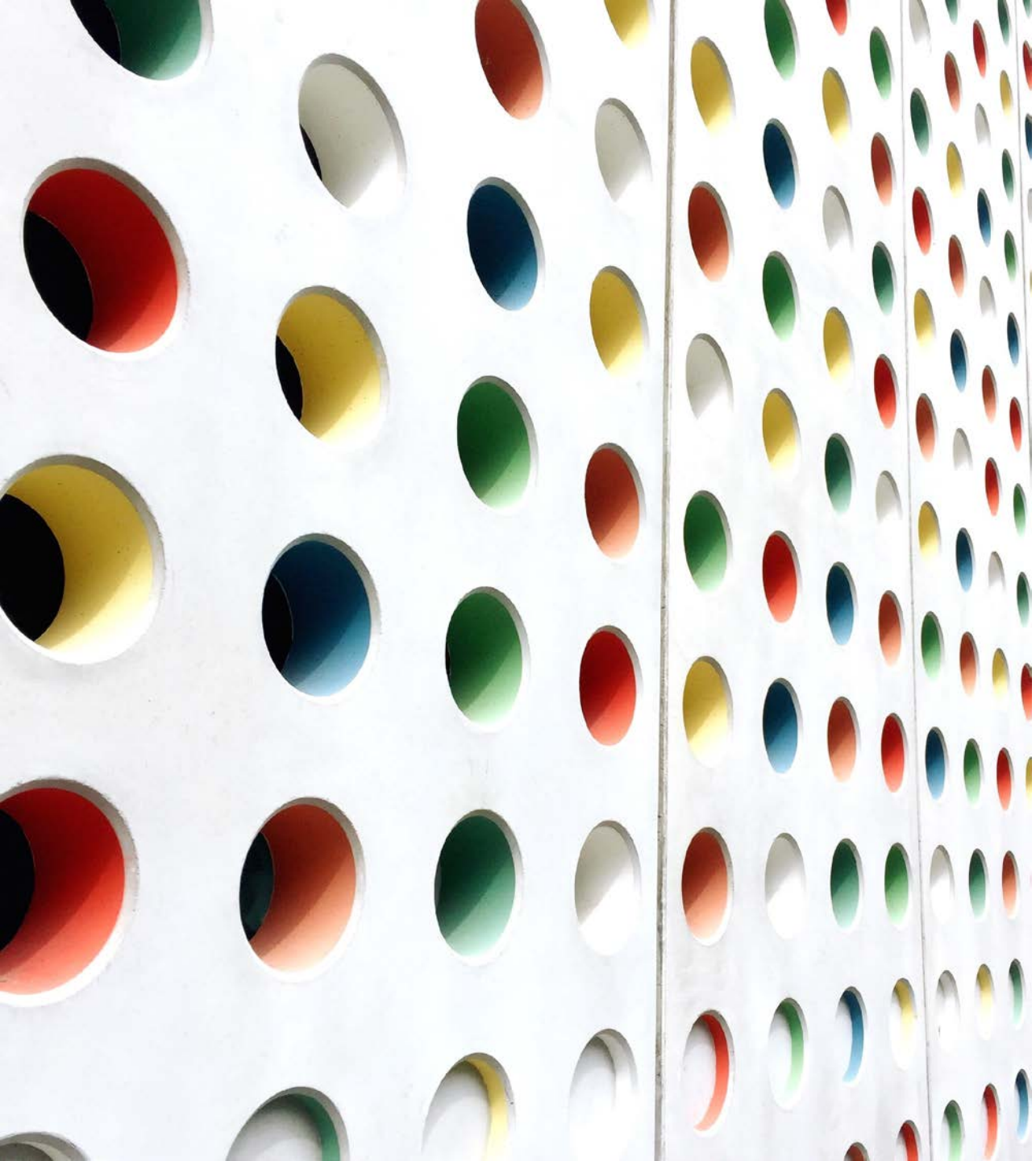


Deep Learning-based Semantic Segmentation for Autonomous Driving

Aleksandar Milosavljević

Associate Professor at University of Niš, Faculty of Electronic Engineering

aleksandar.milosavljevic@elfak.ni.ac.rs

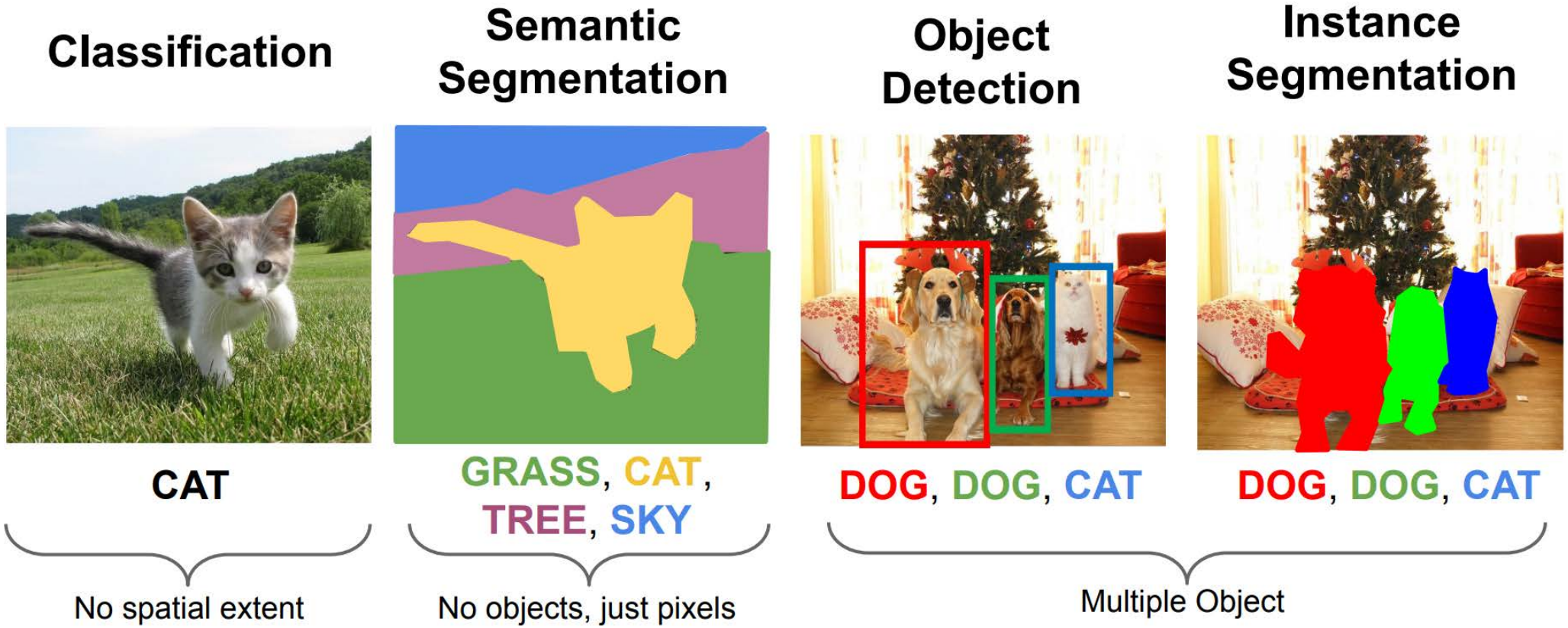
Contents

- Scene understanding and semantic segmentation
- Convolutional neural networks (CNNs) for image classification
- CNN architectures for semantic segmentation
- Semantic segmentation datasets for autonomous driving
- Implementation of semantic segmentation on CamVid dataset

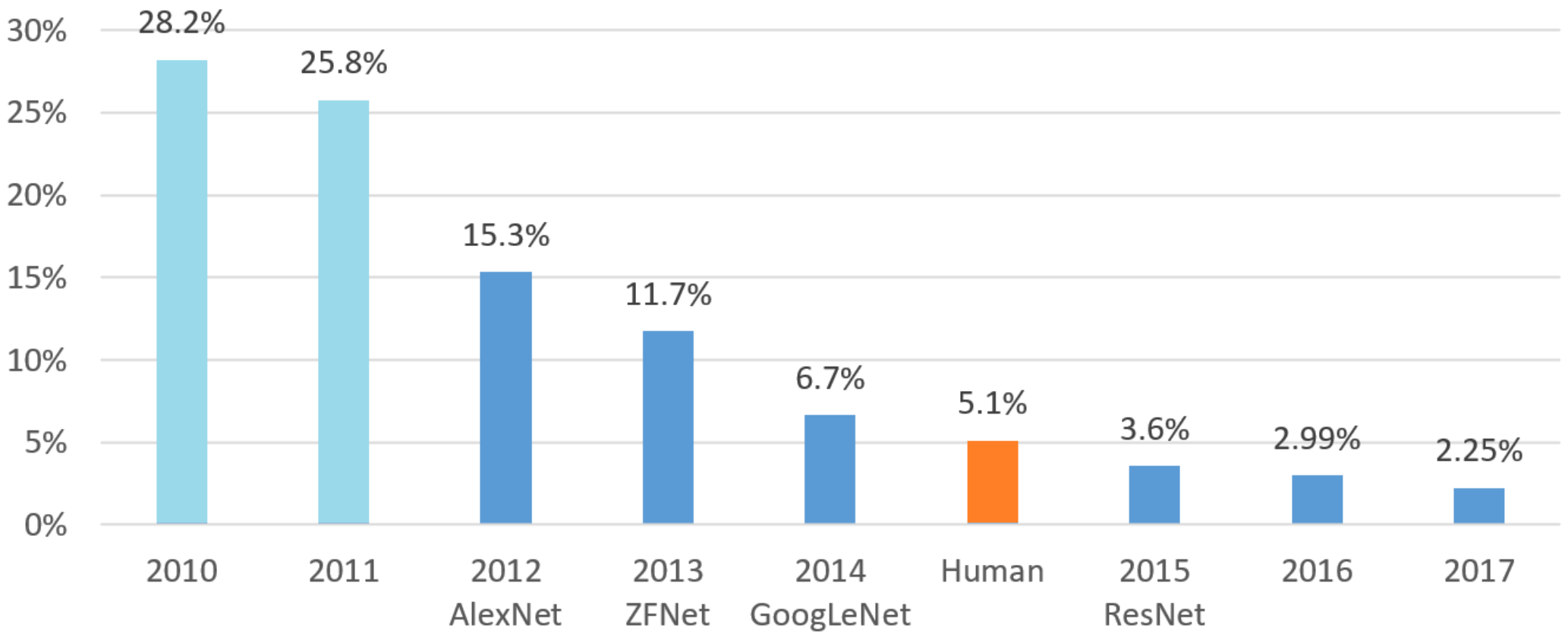
Scene Understanding

- The goal of computer vision is to program a computer to "understand" a scene
- This is obviously very complicated!
- So how to come closer to achieving it?
- By solving simpler computer vision tasks:
  - Detection, segmentation, localization, and recognition of objects in images
  - Object tracking through a sequence of images
  - Mapping a scene to a 3D model of the scene
- CNNs and Deep Learning brought us much closer to that task

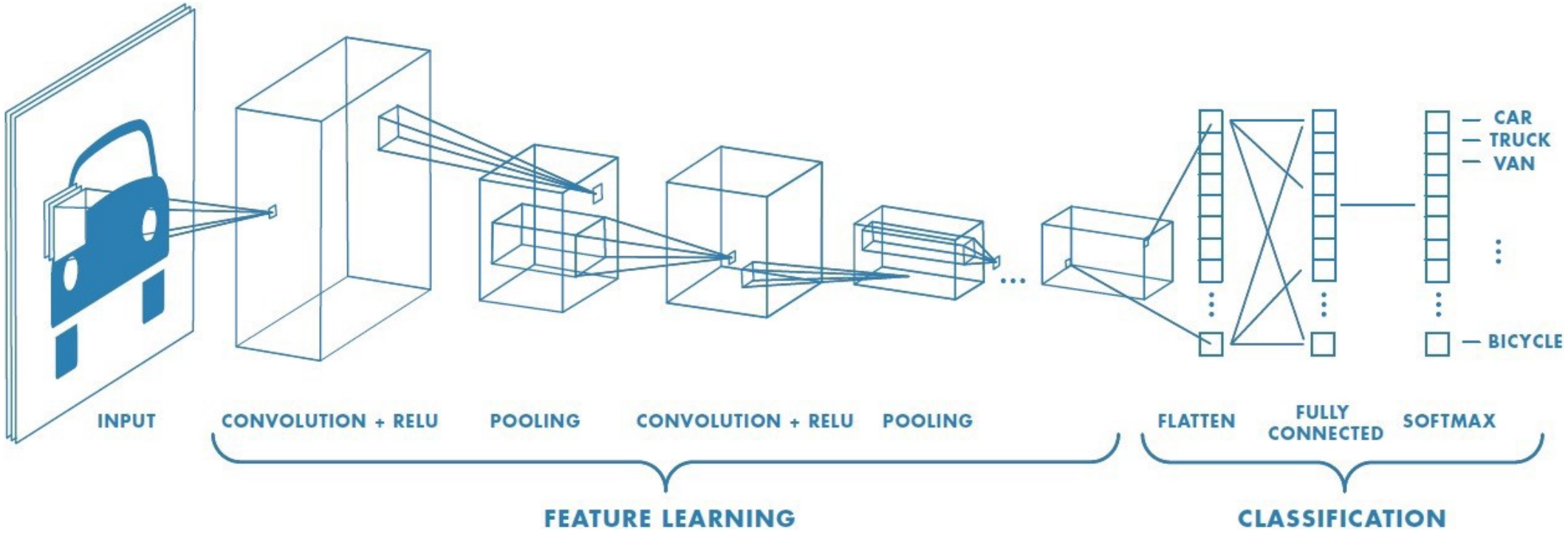
Computer Vision Tasks



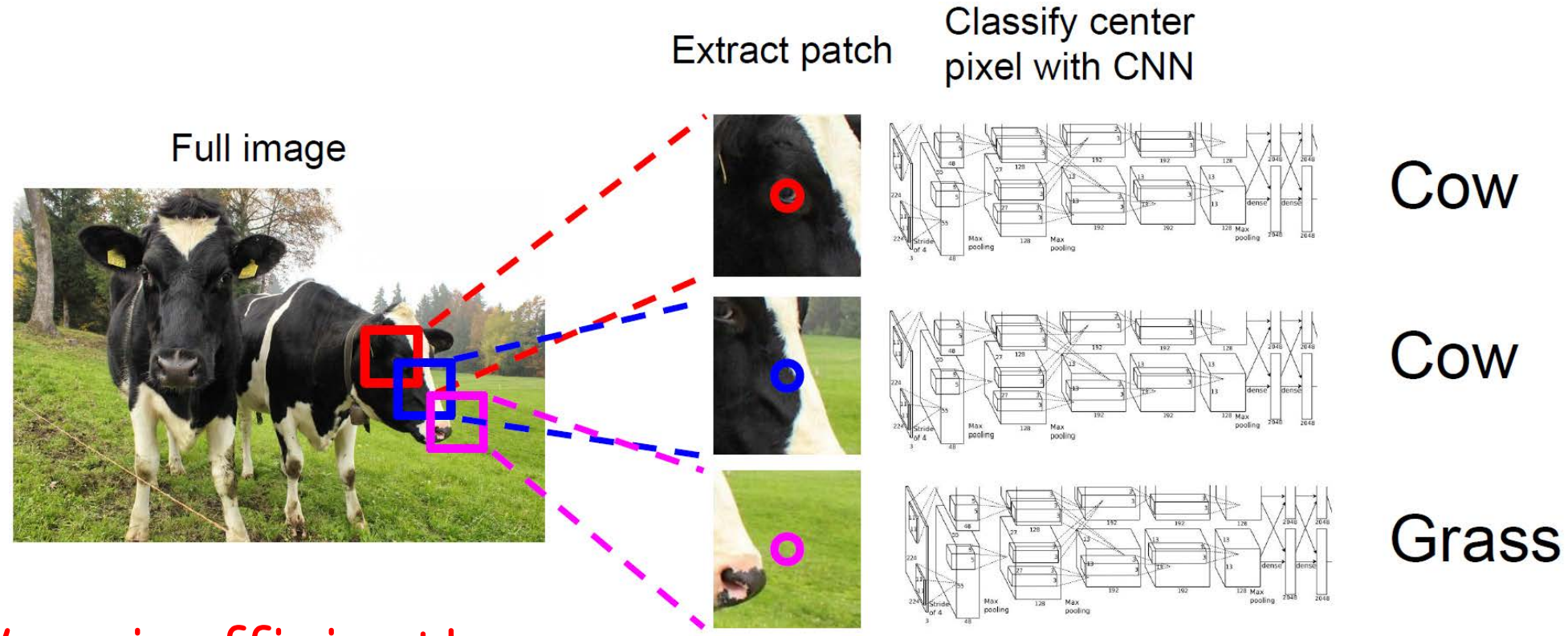
ImageNet Large Scale Visual Recognition Challenge (ILSVRC)



Convolutional Neural Network (CNN)



Semantic Segmentation Idea: Sliding Window

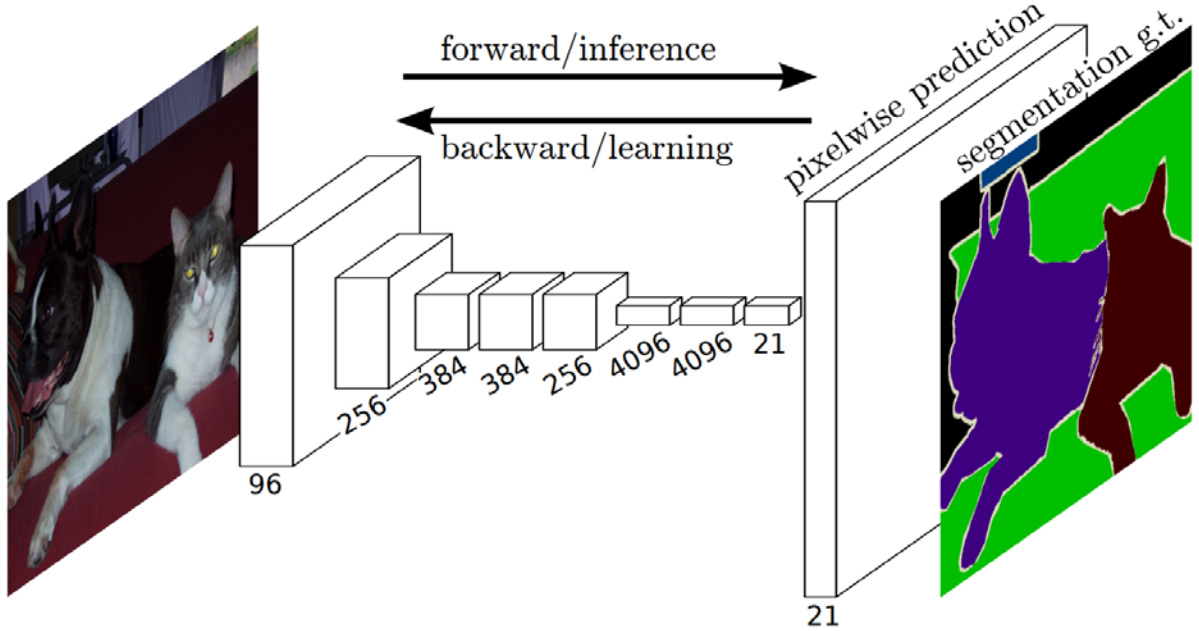


Very inefficient!

Farabet et al, "Learning Hierarchical Features for Scene Labeling," TPAMI 2013
Pinheiro and Collobert, "Recurrent Convolutional Neural Networks for Scene Labeling", ICML 2014

Semantic Segmentation Architectures: Fully Convolutional Network (FCN)

Idea: Encode the entire image with CNN and do pixelwise prediction on top

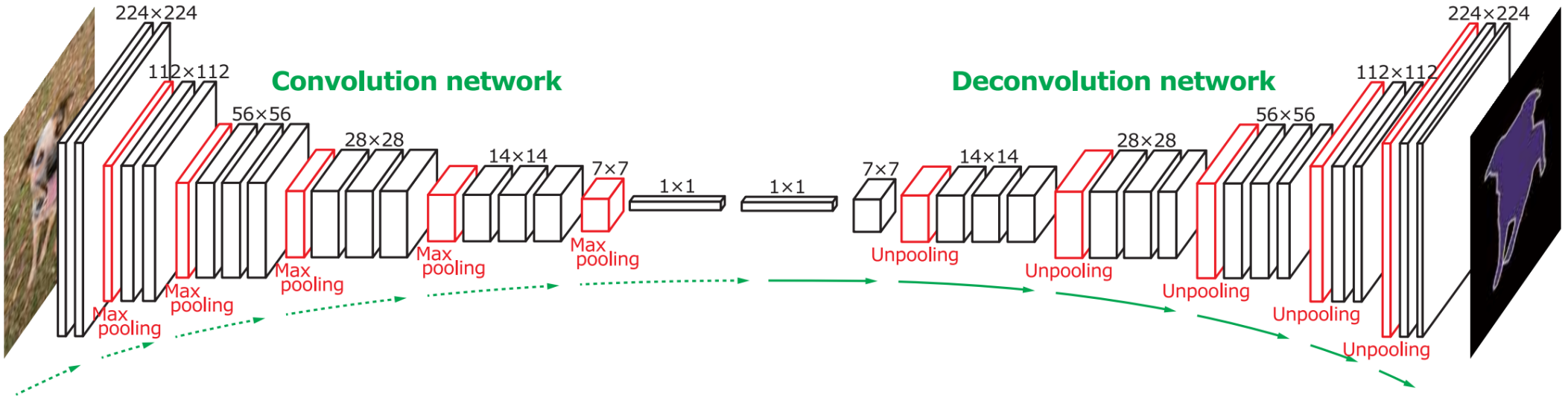


Long, Shelhamer, and Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR 2015

Problem: Classification CNNs reduce spatial dimension and the output needs to be the same size as input

Semantic Segmentation Architectures: DeconvNet

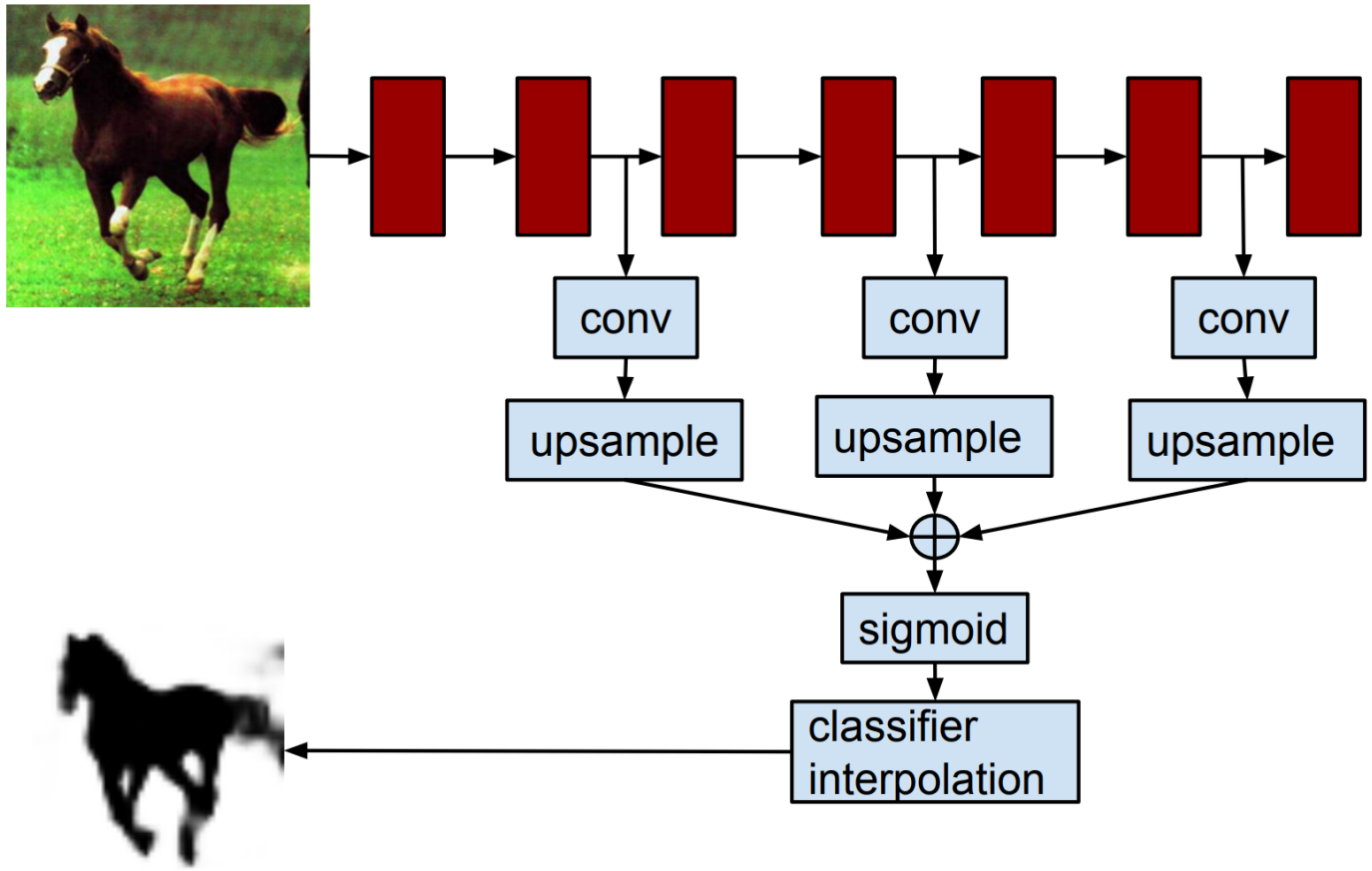
Idea: Use deconvolution network to create predictions



Noh, Hong, and Han, "Learning Deconvolution Network for Semantic Segmentation", ICCV 2015

Problem: More "power" dedicated to creating predictions, but they are still based on spatially reduced information

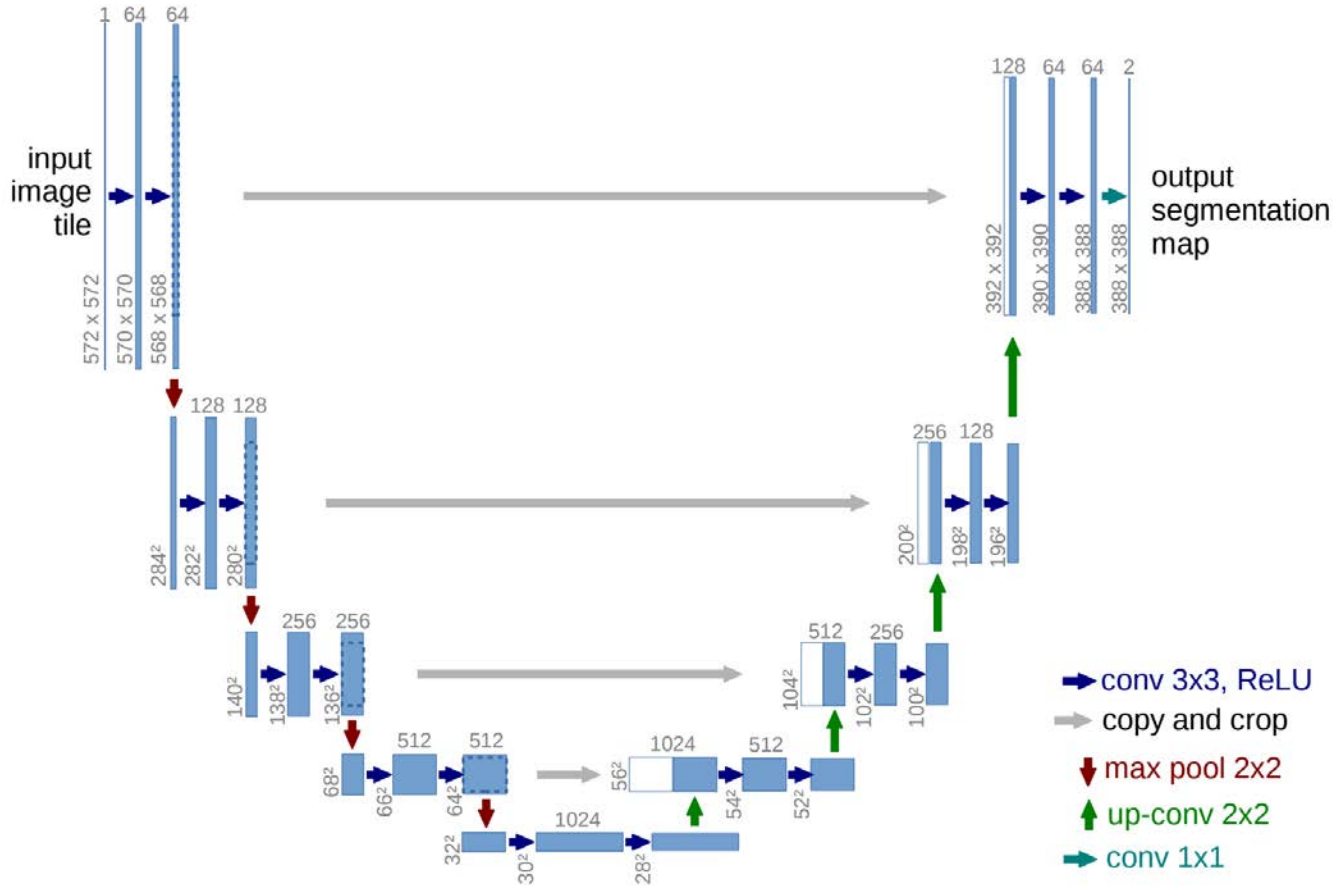
Semantic Segmentation Architectures: Hypercolumn



Idea: Use feature maps from several stages of CNN to create final prediction

Hariharan, Arbeláez, Girshick, and Malik, "Hypercolumns for Object Segmentation and Fine-grained Localization", CVPR 2015

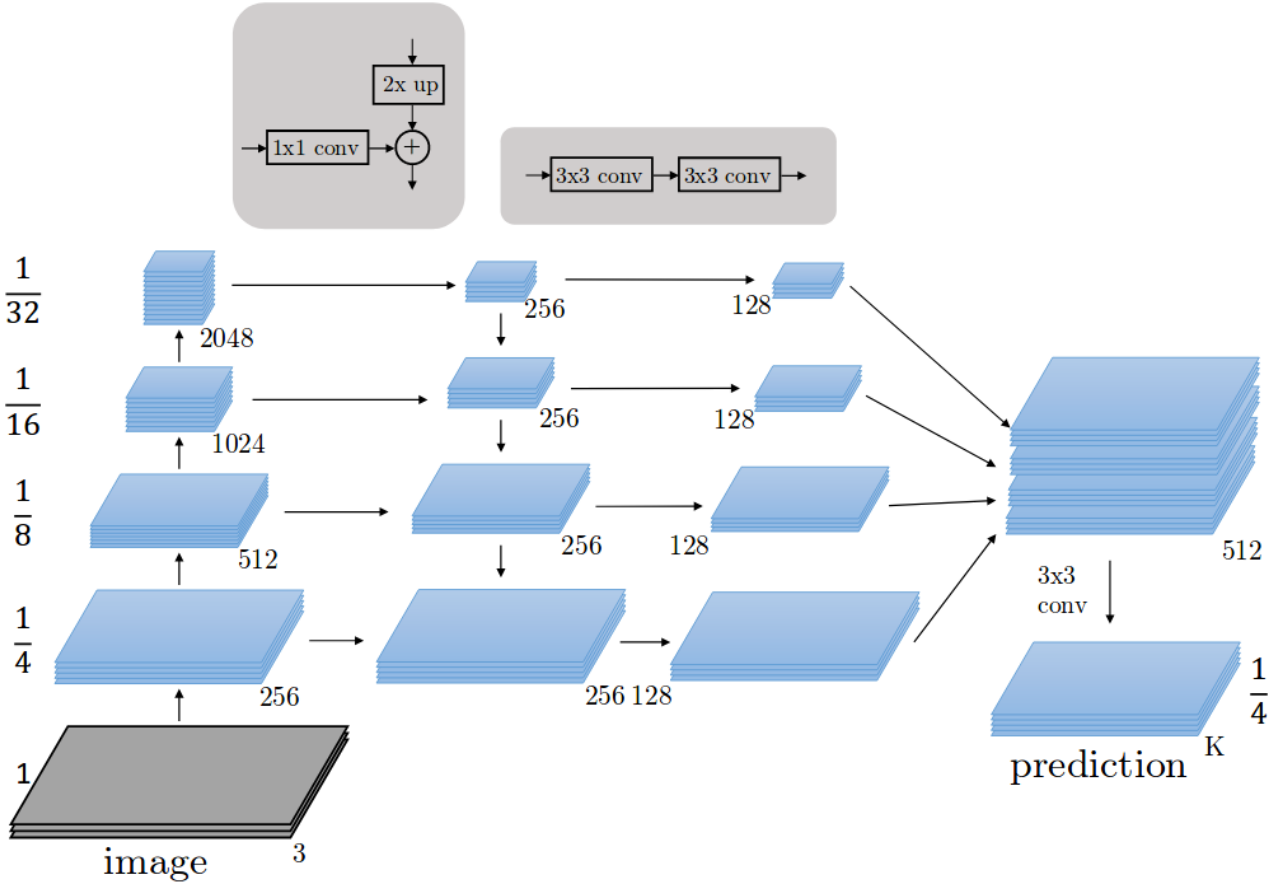
Semantic Segmentation Architectures: U-Net



Idea: Introduce skip connections between encoder and decoder blocks of the same size

Ronneberger, Fischer, and Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", MICCAI 2015

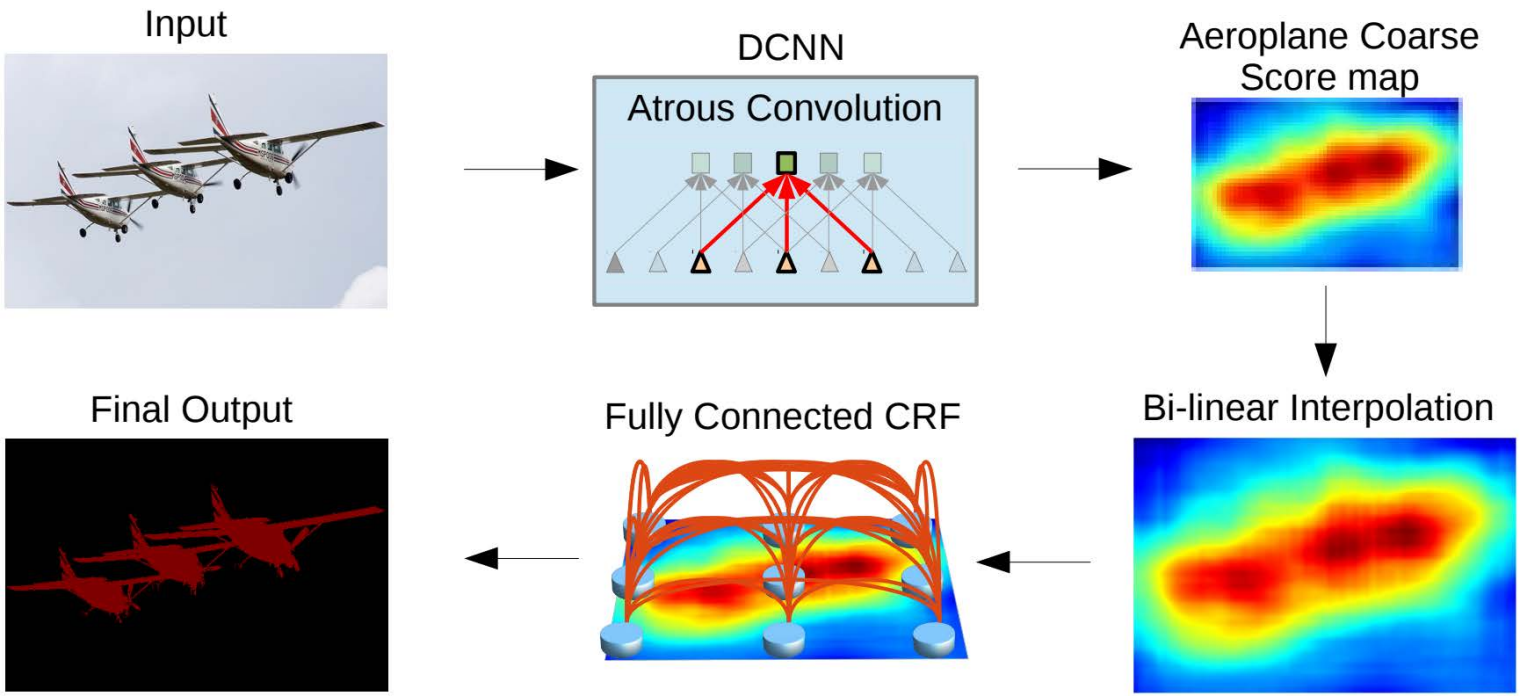
Semantic Segmentation Architectures: Feature Pyramid Network (FPN)



Idea: Make predictions on multiple levels

Lin, Dollar, Girshick, He, Hariharan, and Belongie, "Feature Pyramid Networks for Object Detection", CVPR 2017

Semantic Segmentation Architectures: DeepLab



Ideas:

Use Atrous Convolution in CNN to preserve spatial dimension (from 32x reduction to 8x reduction)

Use Fully Connected Conditional Random Field (CRF) to improve localization performance

Chen, Papandreou, Kokkinos, Murphy, and Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 40, No. 4, 2017

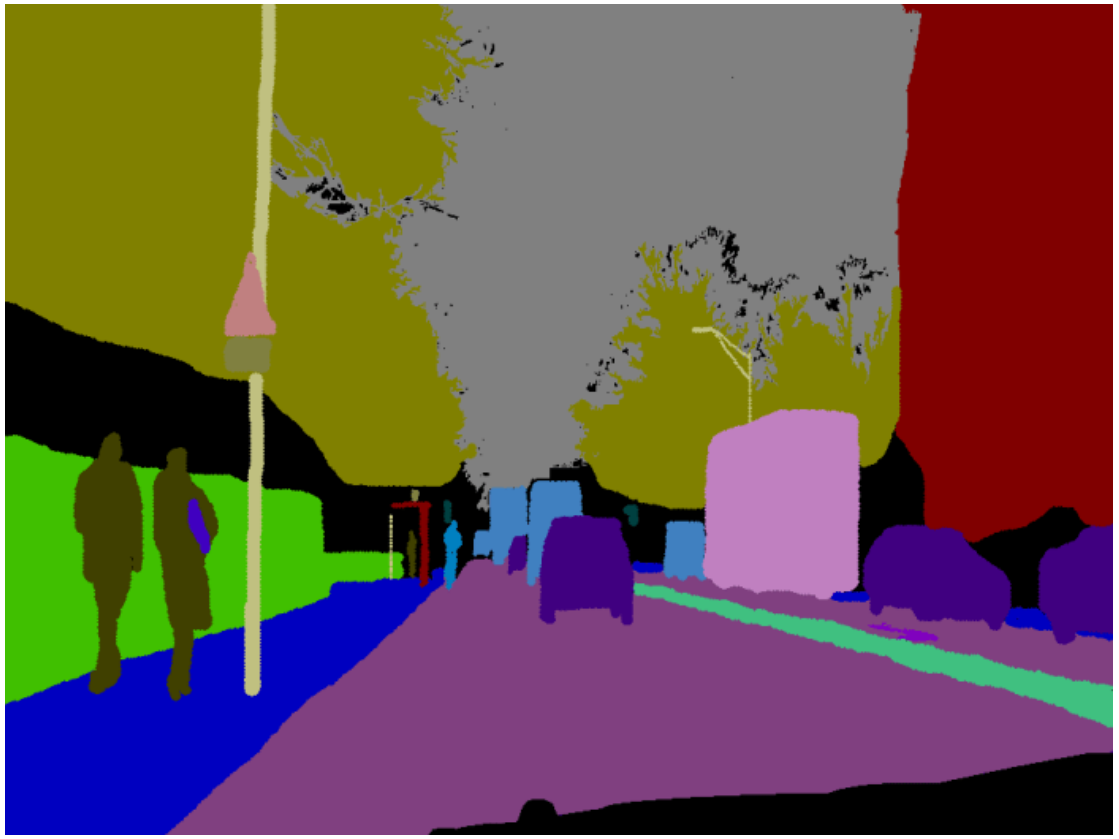
Semantic Segmentation Datasets for Autonomous Driving

Name	Year	Classes	Images	Location	Environment
CamVid	2007	32	700	Cambridge	daylight
KITTI	2012	N/A	N/A	Karlsruhe	daylight
DUS	2013	5	500	Heidelberg	daylight
CityScapes	2016	30	5000	Germany, Switzerland, France	spring, summer, fall
Mapillary Vistas	2017	66	25000	North and South America, Europe, Africa, Asia, Oceania	sun, rain, snow, fog, haze – dawn, daylight, dusk, night

Source: https://medium.com/hackernoon/semantic-segmentation-datasets-for-autonomous-driving-1182ebd2aff0

CamVid

Cambridge-driving Labeled Video Database



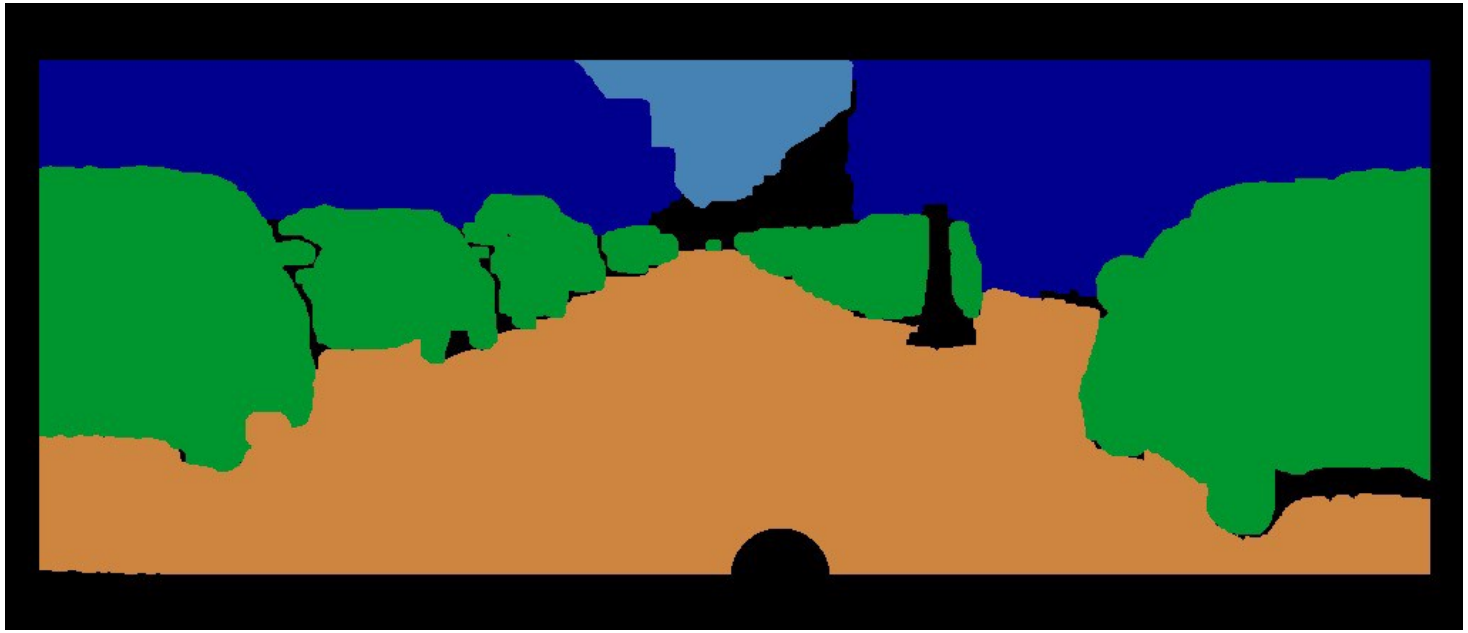
- 960x720 pixels
- 700 images from video sequence of 10 minutes
- One of the first semantically segmented datasets
- Released in late 2007

KITTI

Karlsruhe Institute of Technology and Toyota Technological Institute

- 1242x375 pixels
- Variety of sensors included
- Not with semantically segmented images
- Road and lane detection

DUS
Daimler Urban Segmentation

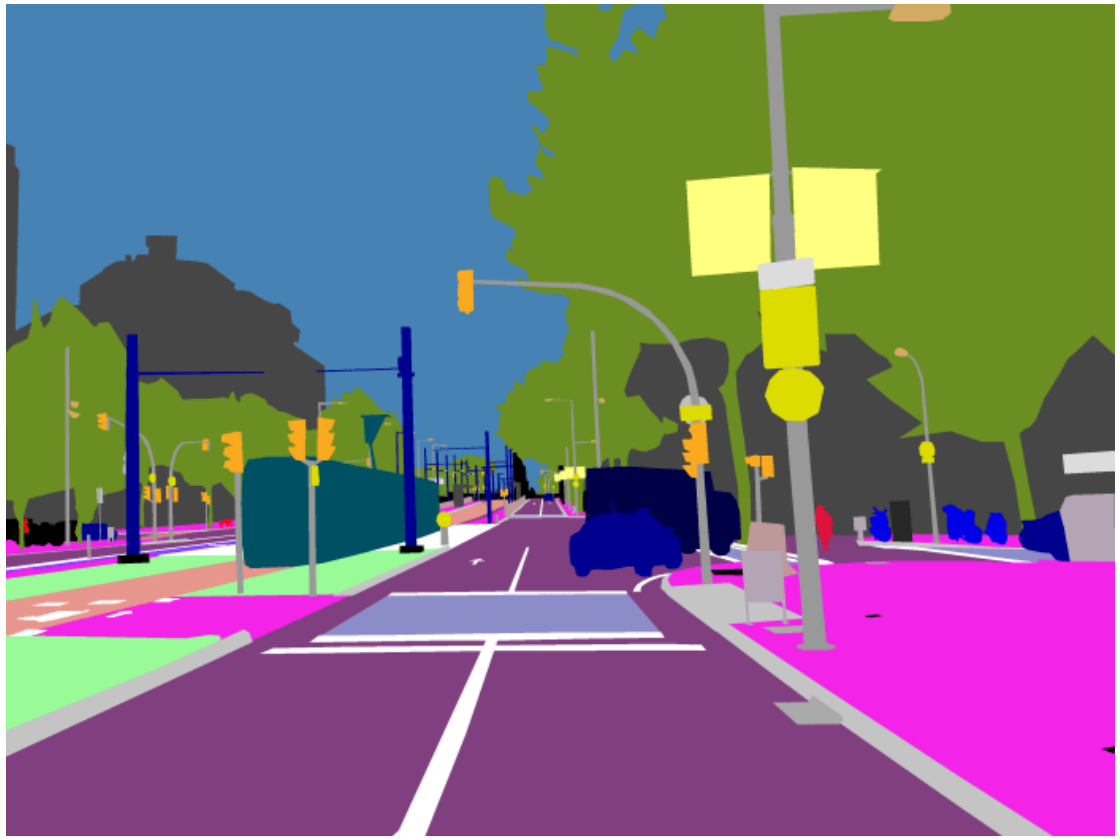


- 1024x440 pixels
- 5000 grayscale images
- 500 semantically segmented images
- No “nature” class

CityScapes
Continuation of DUS

- 2048x1024 pixels
- 30 classes, 8 higher level categories
- 20000 images with coarse segmentation
- IoU class 85.8%, IoU category 93.2%
  https://www.cityscapes-dataset.com/benchmarks/

Mapillary Vitas Dataset



- 4000x3000 pixels
- 25000 hi-res images
- Collaboratively collected
- Instance segmentation for 37 out of 66 classes
- Different viewing angles

Implementation of Semantic Segmentation on CamVid Dataset

https://github.com/a-milosavljevic/camvid-segmentation