Deep Learning

Abdelhak Mahmoudi
abdelhak.mahmoudi@um5.ac.ma

INPT- 2020

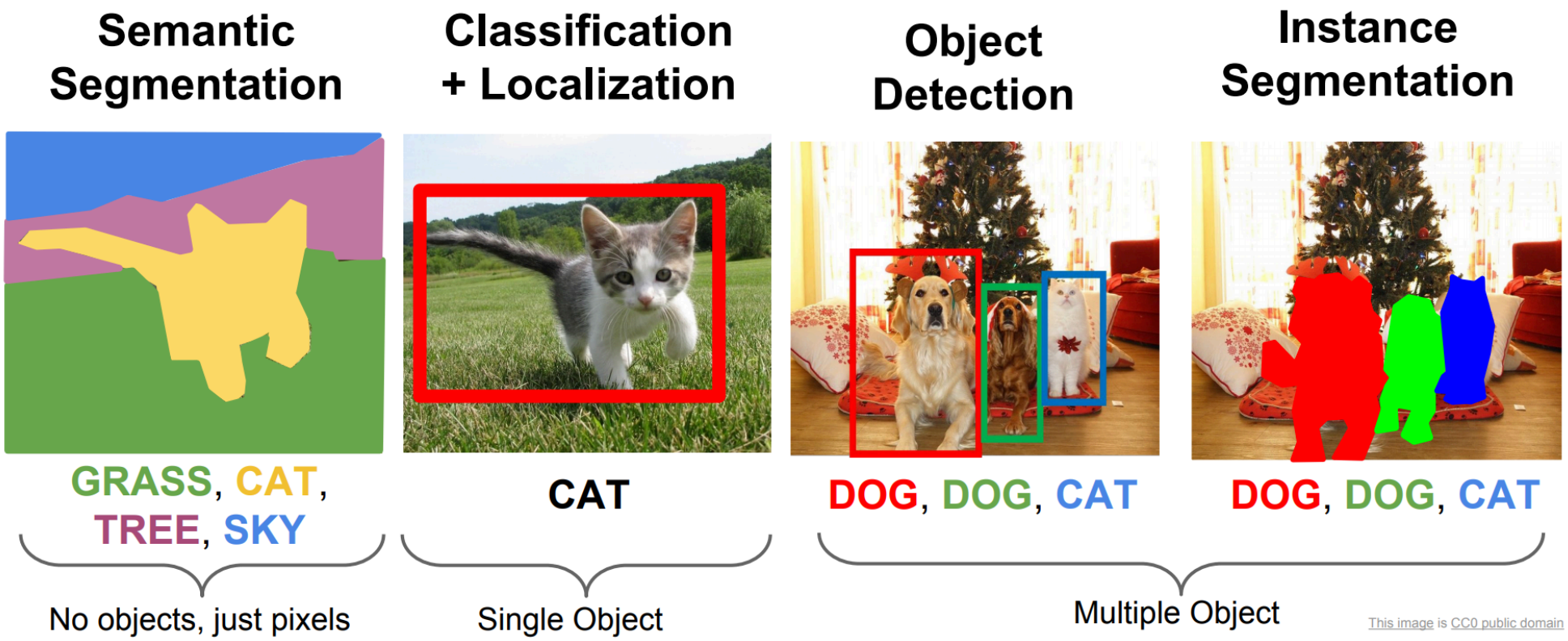
Content

1. Deep Artificial Neural Networks
2. **Convolutional Neural Networks**
3. Sequence Models
4. Generative Models

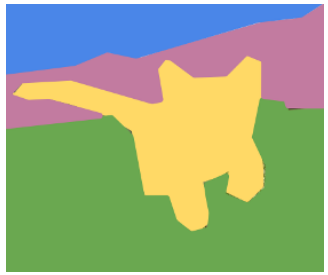
CNN for Computer Vision

- Semantic Segmentation
- Classification + Localization
- Object Detection
- Instance Segmentation

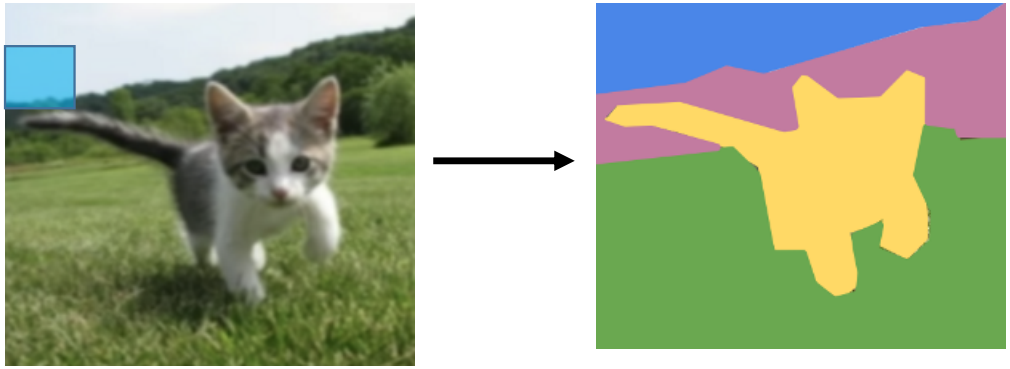
CNN for Computer Vision



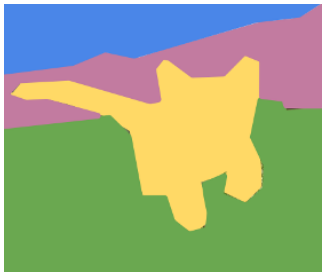
Semantic Segmentation



- Sliding window ?
  - Classify center pixel with CNN
  - Very expensive
  - Not sharing features



Semantic Segmentation

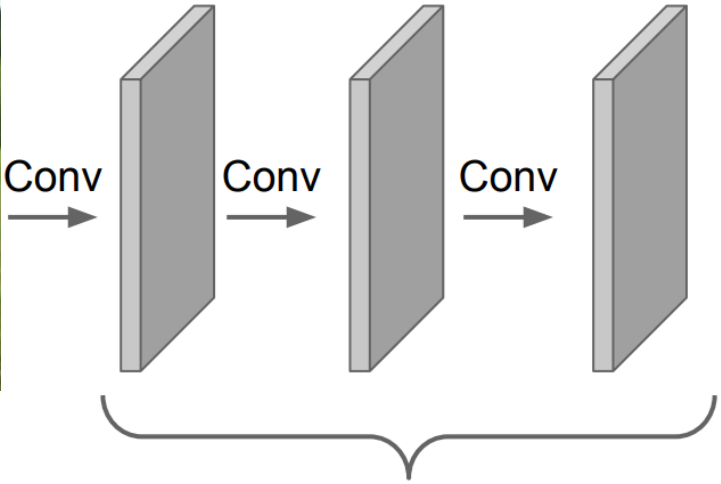


Fully Conv Nets

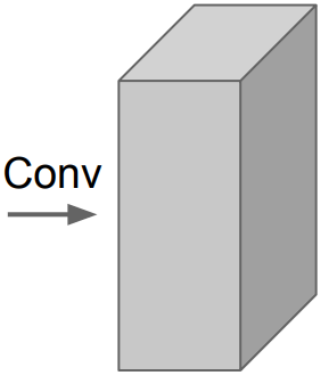
Preserve the size of the input

But still expensive

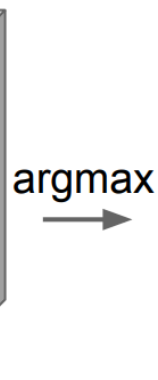
Input:
3 x H x W

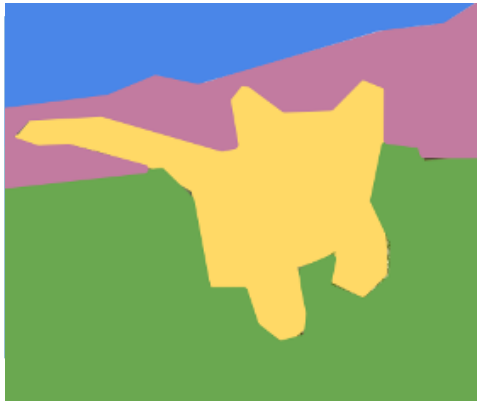


Convolutions:
D x H x W



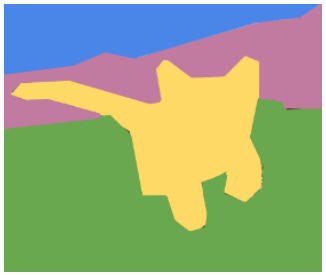
Scores:
C x H x W





Predictions:
H x W

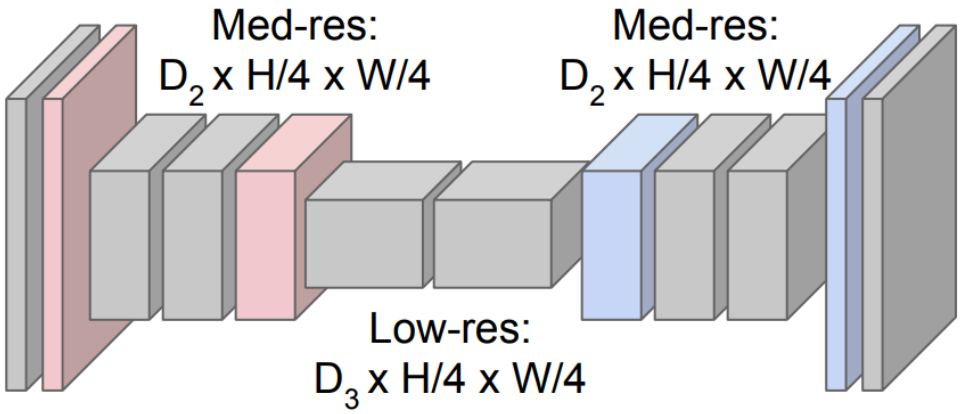
Semantic Segmentation



Fully Conv Nets

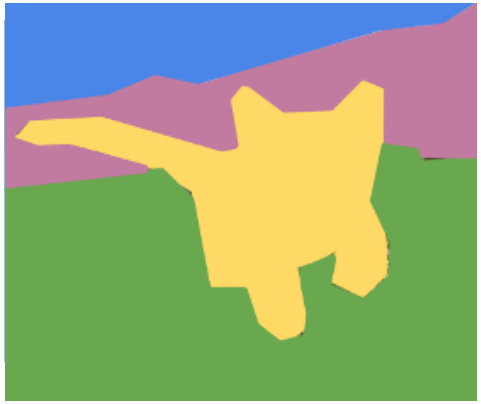
Symmetric with bottleneck

Input:
3 x H x W



High-res:
D_1 x H/2 x W/2

High-res:
D_1 x H/2 x W/2



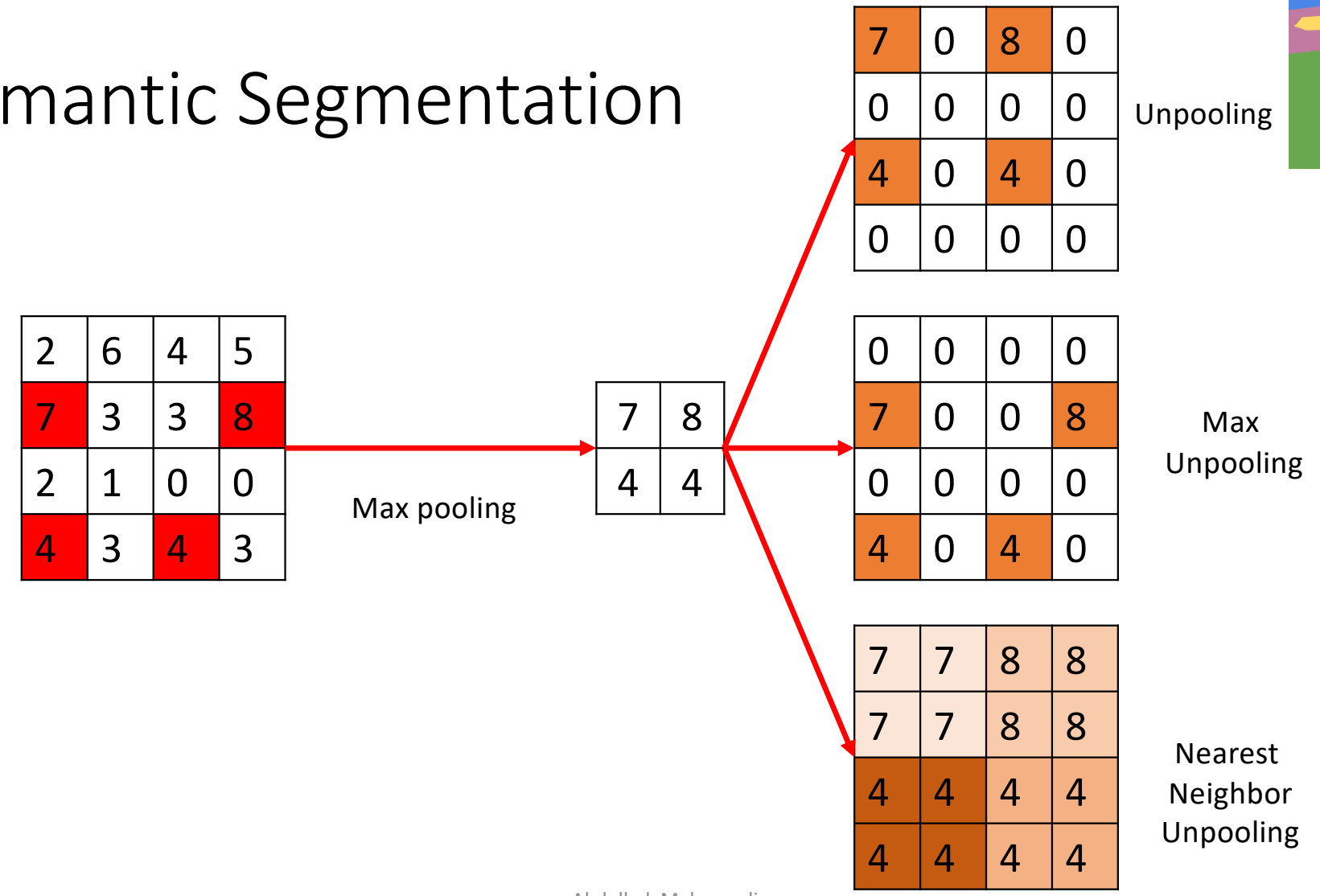
Predictions:
H x W

Convolution
(Encoding, downSampling, pooling)

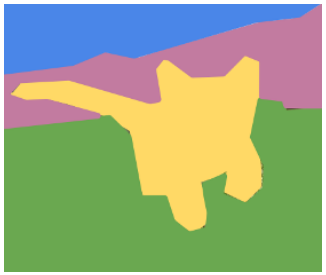
Transpose Convolution
(Decoding, UpSampling, Unpooling)

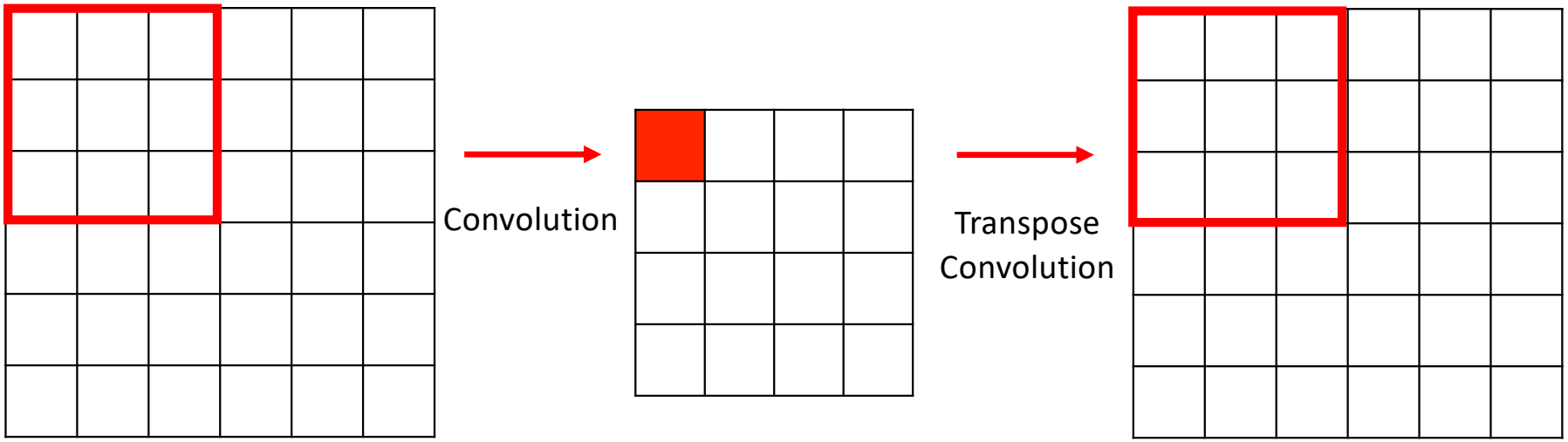
Long, Shelhamer, and Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR 2015

Semantic Segmentation

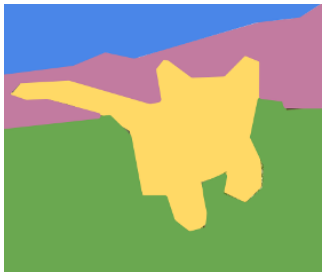


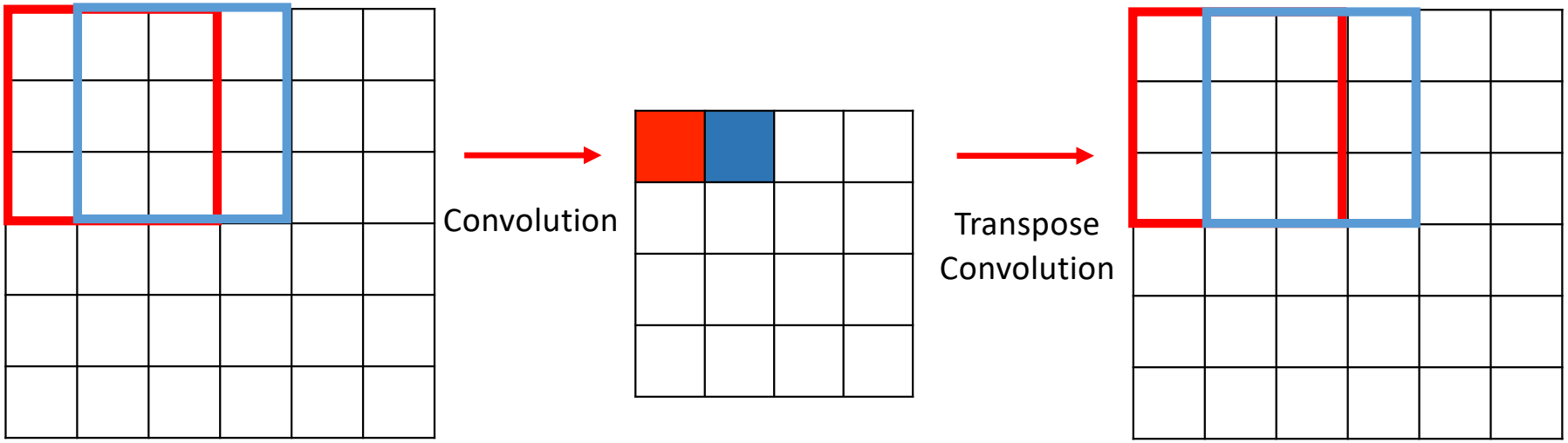
Semantic Segmentation



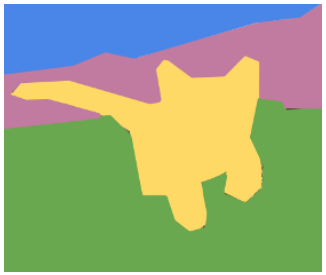


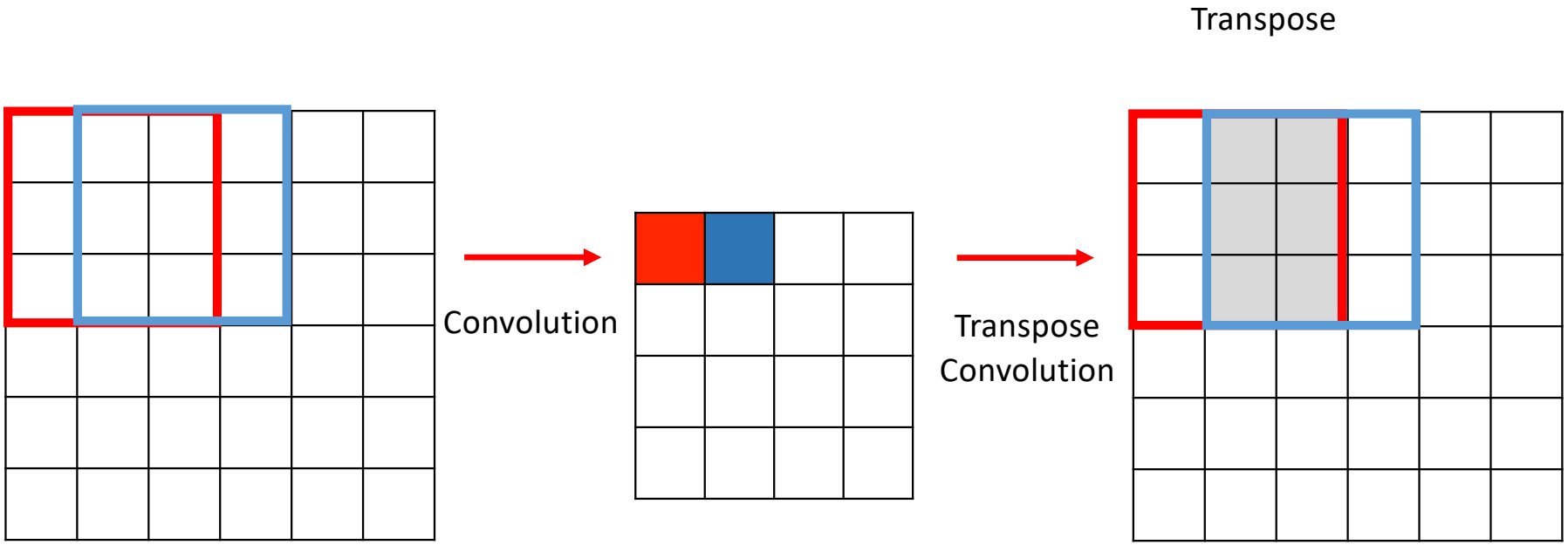
Semantic Segmentation



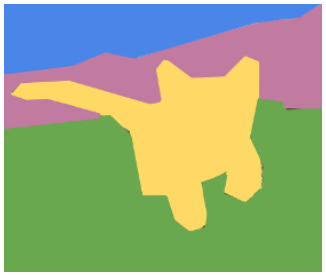


Semantic Segmentation

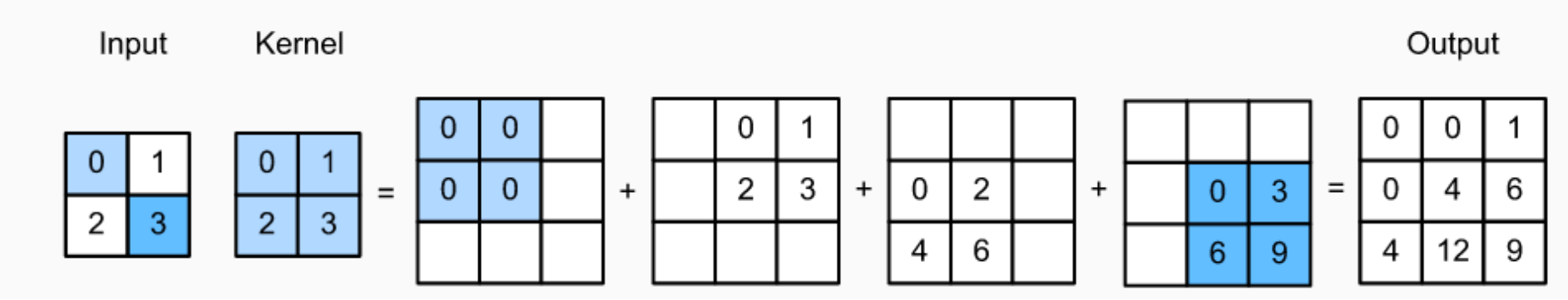




Semantic Segmentation

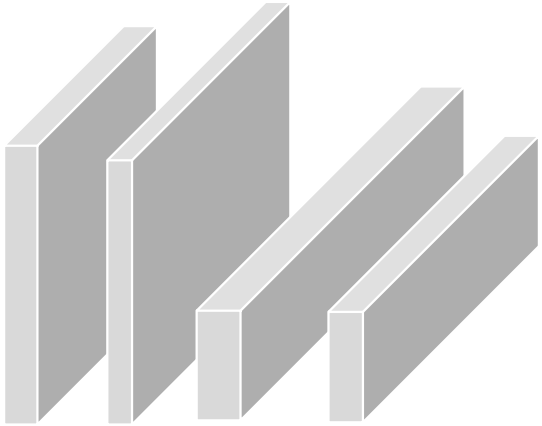


Example of Transpose Convolution



Classification + Localization

- Classification + Rregression



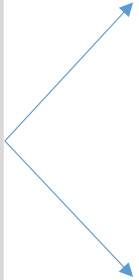


Image class
(Classification)

Bounding box,
(Regression)

CNN

FC

Classification + Localization

Object Recognition

Find a bounding box

b_x, b_y, b_h, b_w

Face Recognition

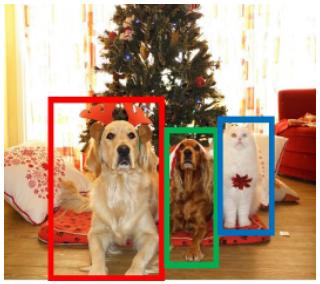
Find landmarks

$l_{1x}, l_{1y},; l_{2x}, l_{2y}; l_{4x}, l_{4y} \ldots; l_{64x}, l_{64y}$

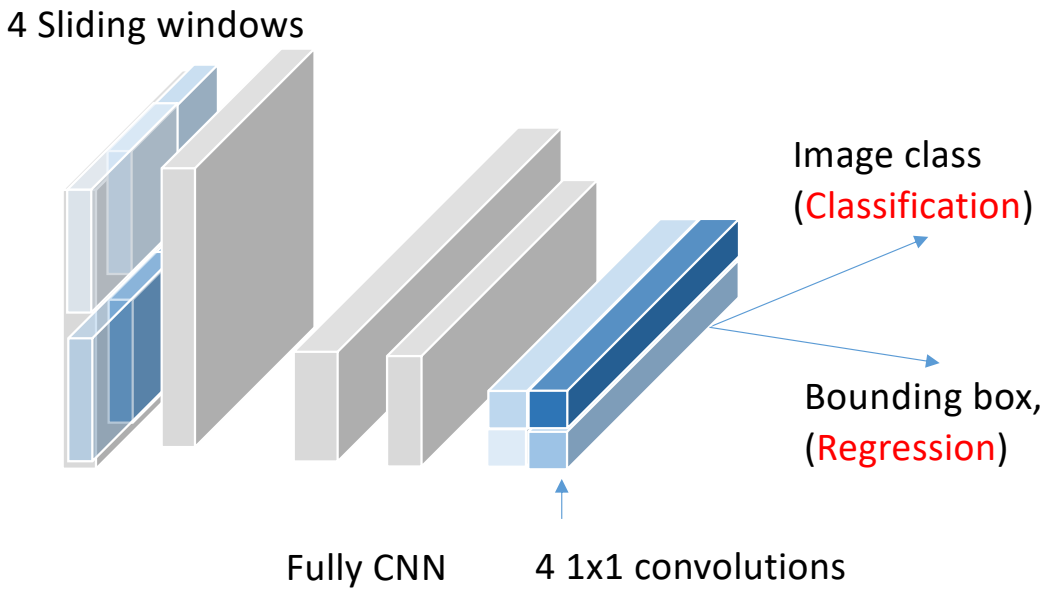
Pose Recognition

Find bones

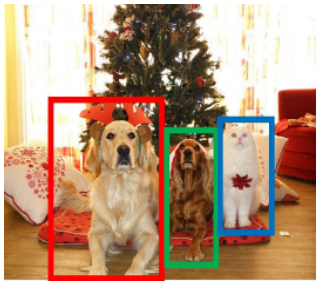
Object Localization



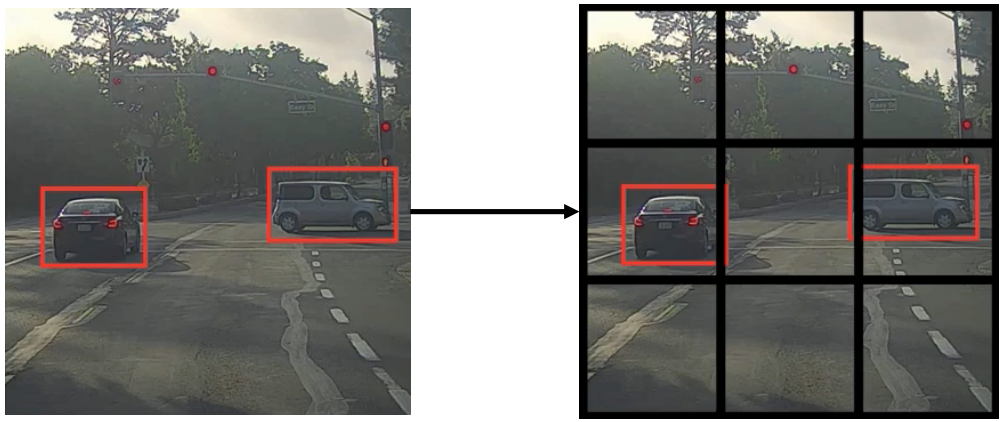
- Multiple Objects
- Different Objects
- Problem
  - Need to apply CNN to huge number of sliding windows to all locations and with different scales. Very computationally expensive!
- 1st Solution
  - Fully ConvNets



Object Localization



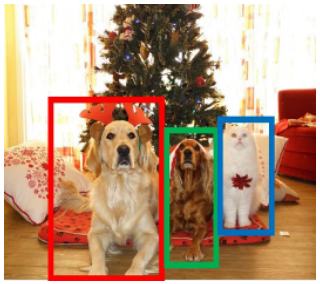
- Multiple Objects
- Different Objects
- Problem
  - Non accurate bounding boxes
- 2nd Solution
  - YOLO (You Only Look Once)



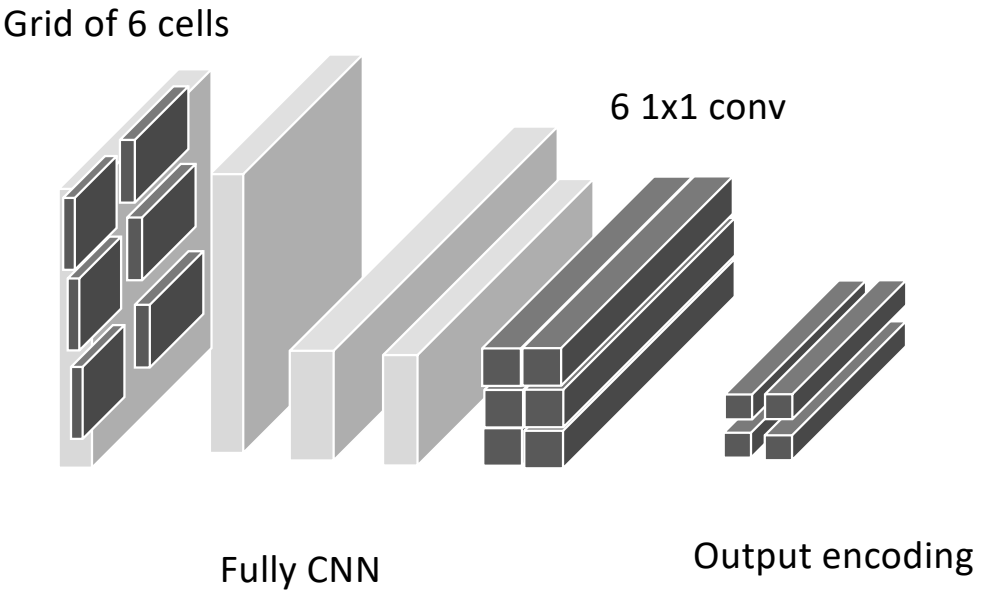
Assign to the appropriate cell of the grid

Redmon et al, “You Only Look Once: Unified, Real-Time Object Detection”, CVPR 2016

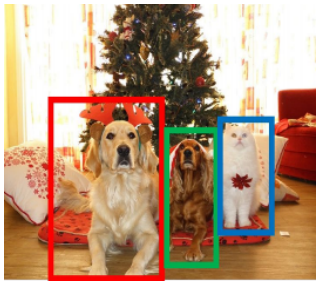
Object Localization



- Multiple Objects
- Different Objects
- Problem
  - Non accurate bounding boxes
- 2nd Solution
  - YOLO (You Only Look Once)
  - Last YOLO-V4 (April 2020)

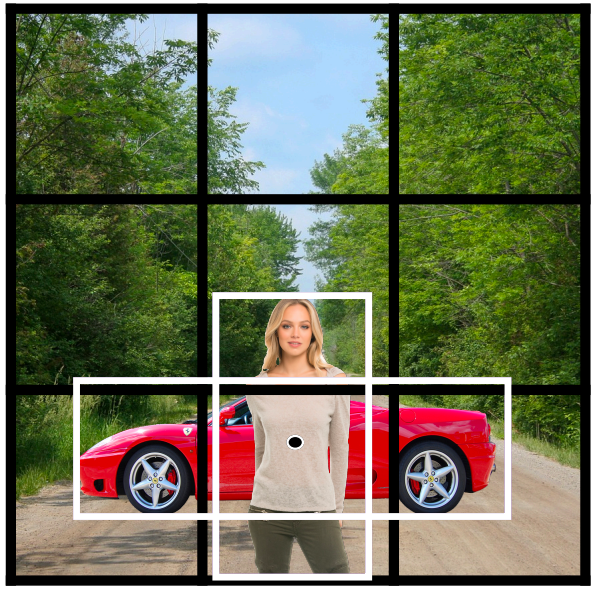


Object Localization



Pedestrian Car motorcycle

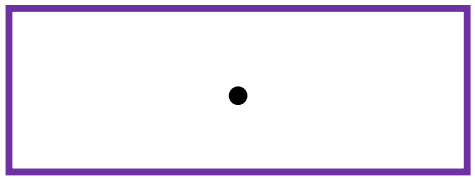
$$\begin{bmatrix}1\\0\\0\end{bmatrix} \quad \begin{bmatrix}0\\1\\0\end{bmatrix} \quad \begin{bmatrix}0\\0\\1\end{bmatrix}$$



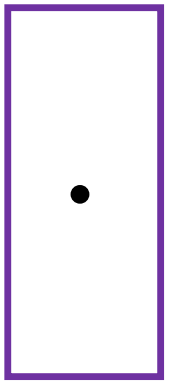
$$y = \begin{bmatrix} p_c \\ b_x \\ b_y \\ b_h \\ b_w \\ c_1 \\ c_2 \\ c_3 \end{bmatrix}$$

$$y = \begin{bmatrix} p_c \\ b_x \\ b_y \\ b_h \\ b_w \\ c_1 \\ c_2 \\ c_3 \\ p_c \\ b_x \\ b_y \\ b_h \\ b_w \\ c_1 \\ c_2 \\ c_3 \end{bmatrix}$$

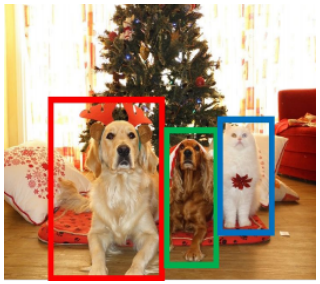
Anchor box 1



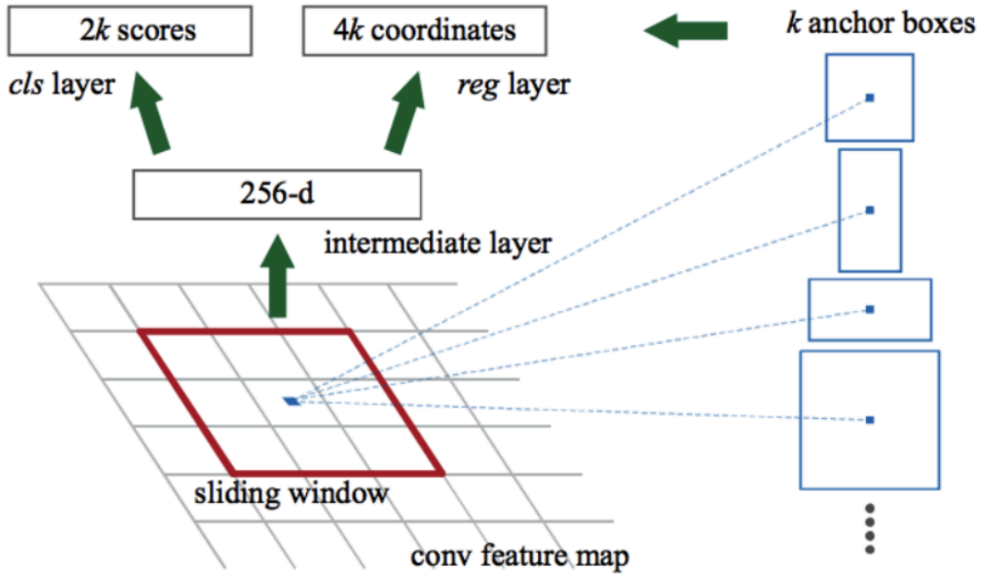
Anchor box 2



Object Localization

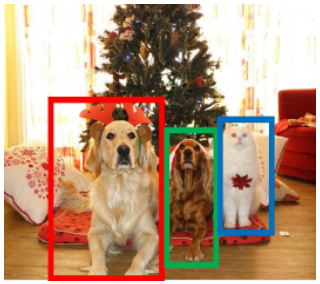


- Multiple Objects
- Different Objects
- Problem
  - Need to apply CNN to huge number of sliding windows to all locations and with different scales. Very computationally expensive!
- 3rd Solution
  - Use Region Proposal Network (RPN)

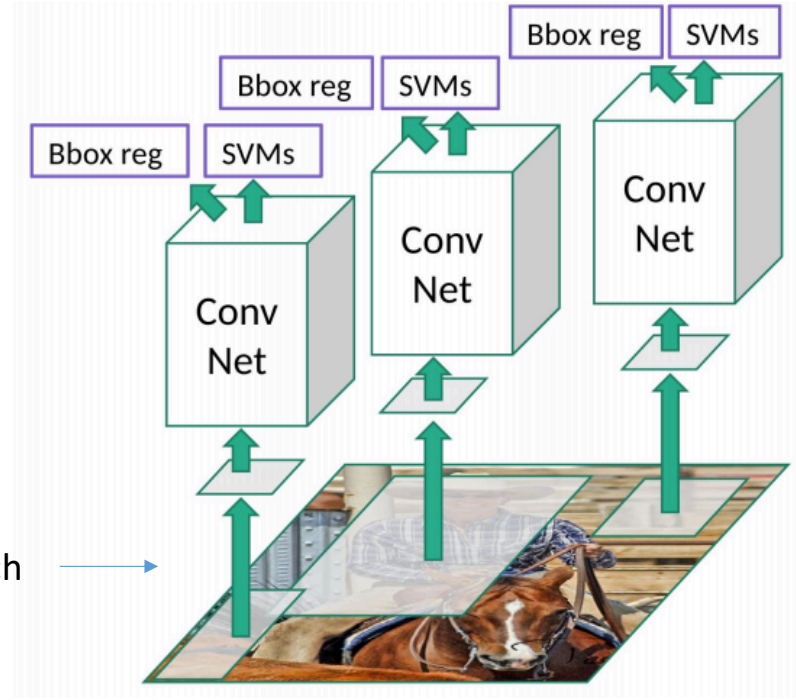


RPN

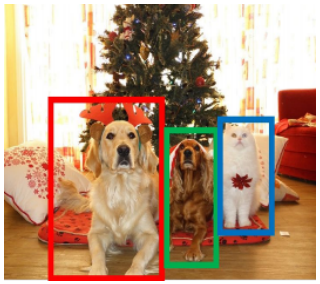
Object Localization



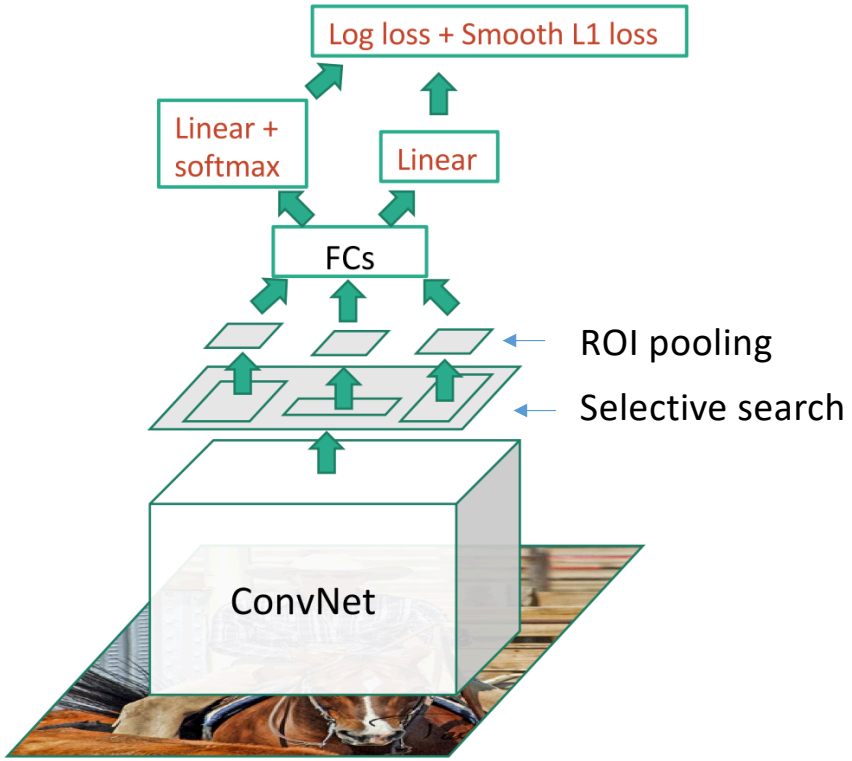
- RCNN
  - 1) Selective Search Proposal method
  - 2) CNN
  - 3) Classification + Regression



Object Localization



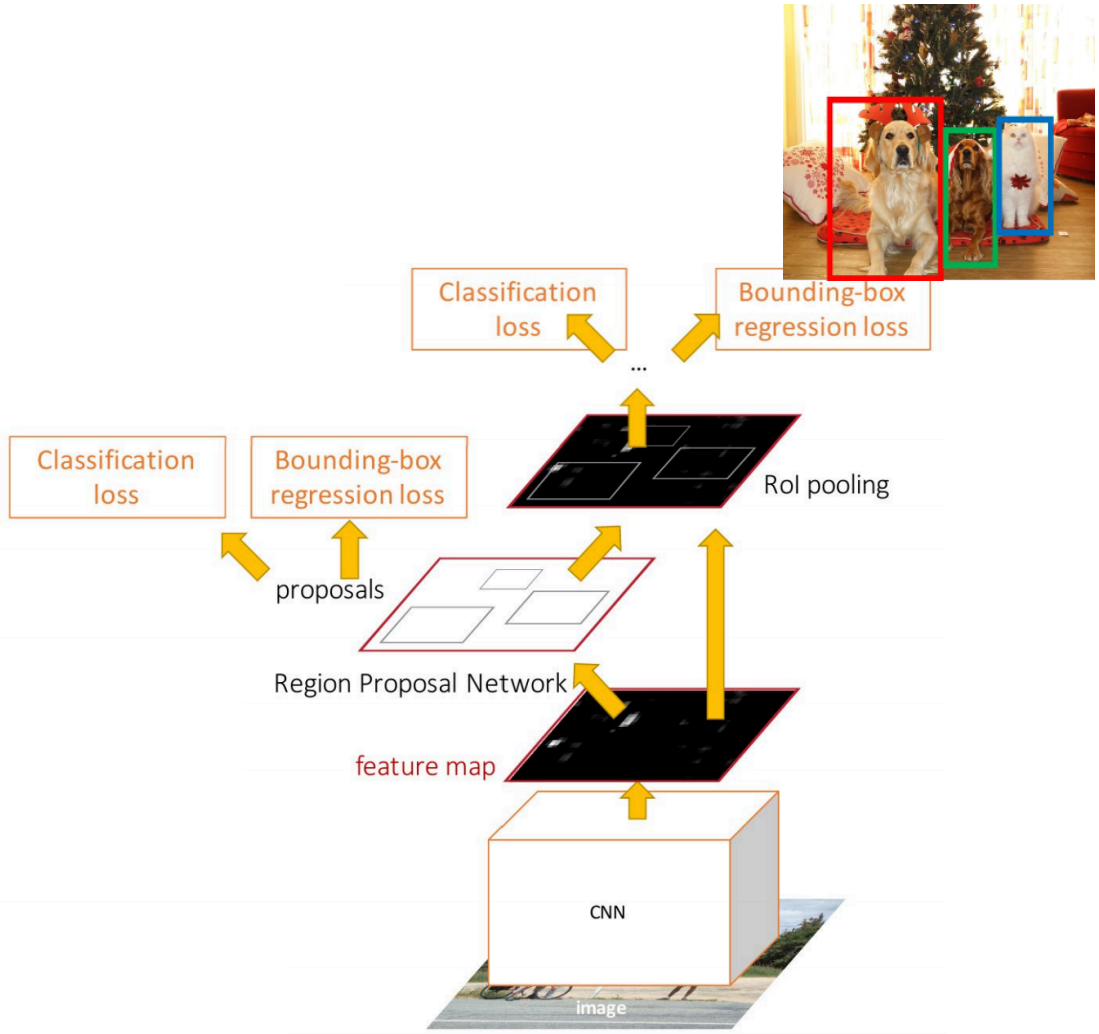
- Fast-R-CNN
  - 1) CNN
  - 2) Selective Search Proposal method
  - 3) ROI pooling
  - 4) Classification + Regression



Grshick, "Fast R-CNN", ICCV 2015.

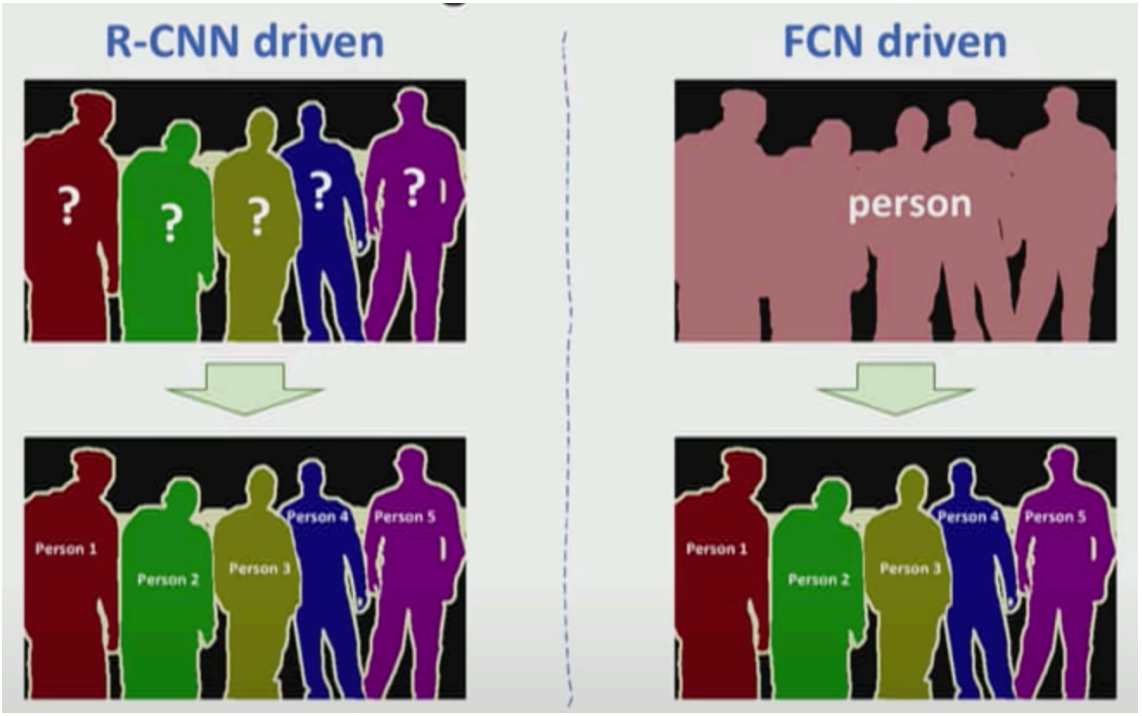
Object Localization

- Fast**er**-R-CNN
  - 1) CNN
  - 2) Region Proposal Network (RPN)
  - 3) ROI pooling
  - 4) Classification + Regression



Instance Segmentation

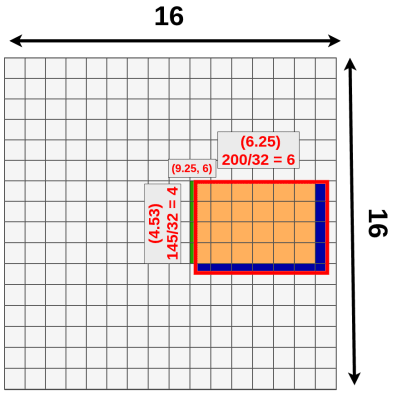
- Mask-RCNN
  - Combine Faster RCNN with FCNN



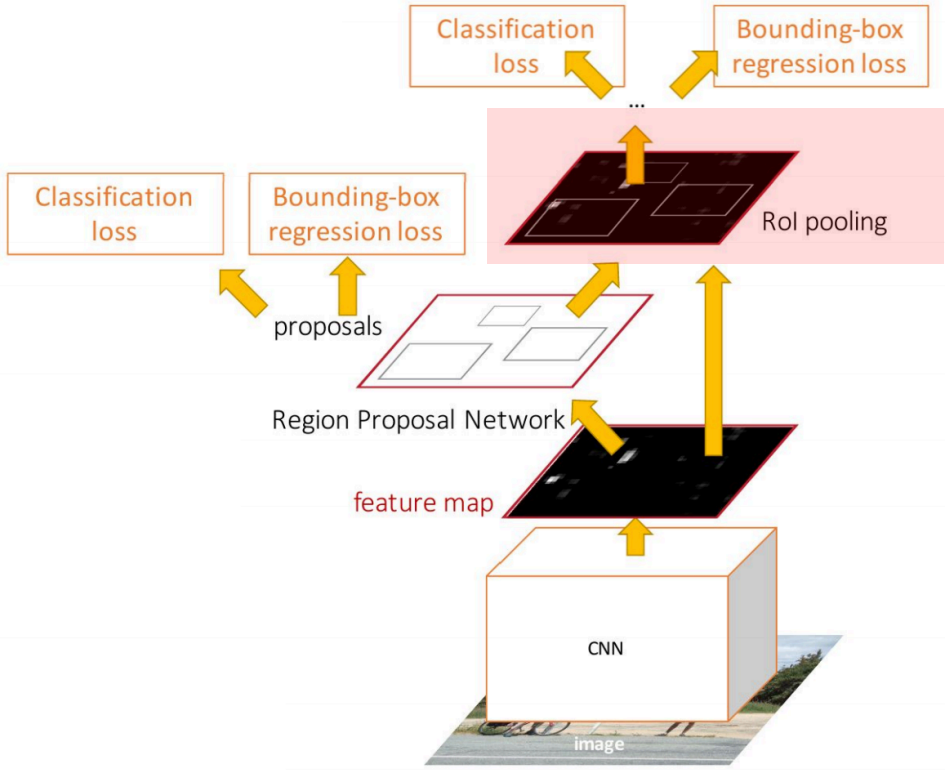
He et al, “Mask R-CNN”, arXiv 2017
Figures copyright Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick, 2017.

Instance Segmentation

- Mask-RCNN
  - Combine Faster RCNN with FCNN
  - Replace ROI Pooling by ROI alignment or ROI Warping



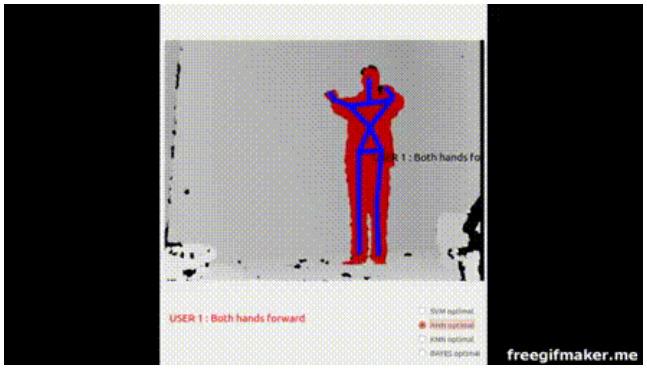
https://towardsdatascience.com/@kemalpiro



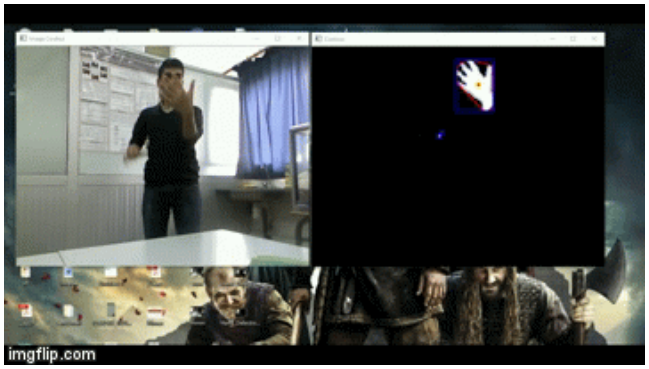
He et al, "Mask R-CNN", arXiv 2017
Figures copyright Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick, 2017.

CNN for Computer Vision at LIMIARF Lab

Pose recognition



Hand Gestures recognition



Hand Signs recognition