

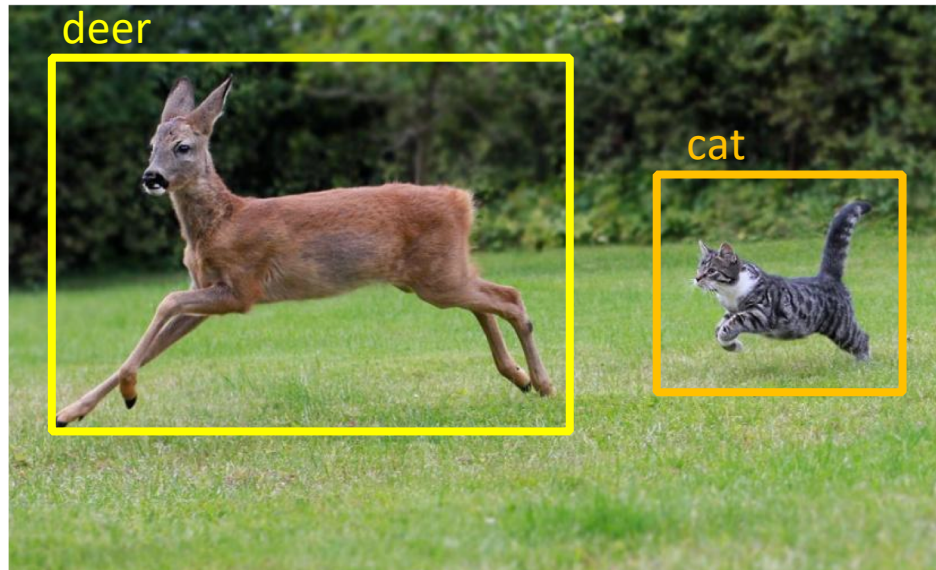
CS6501: Deep Learning for Visual Recognition

Detection, Segmentation

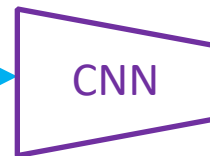
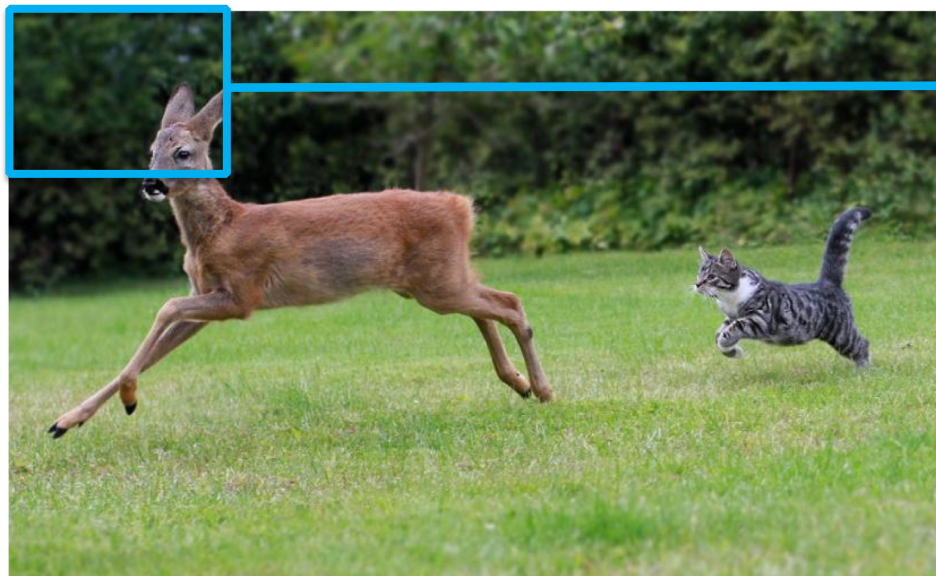
Overview



Object Detection

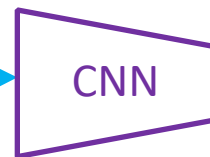
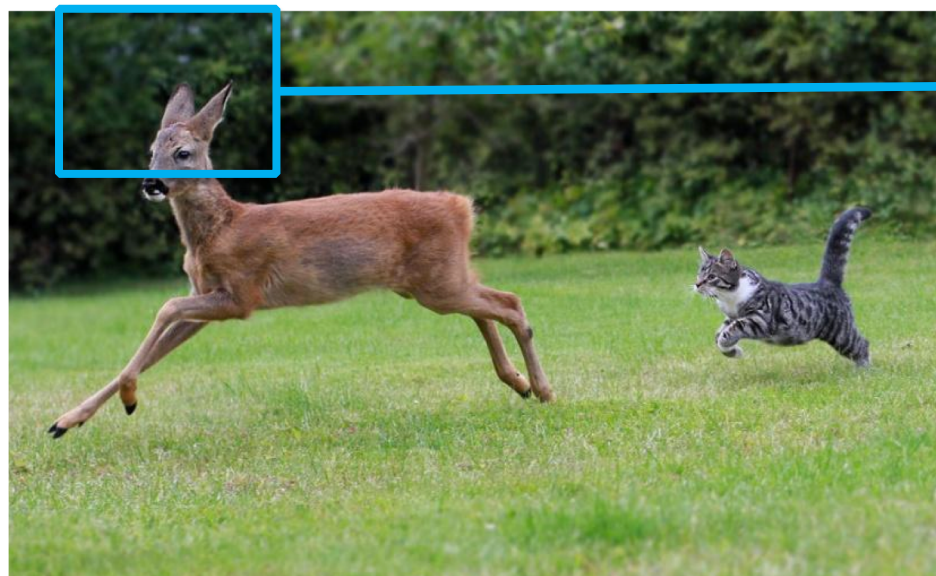


Object Detection as Classification



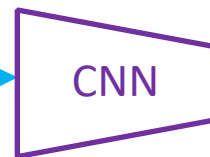
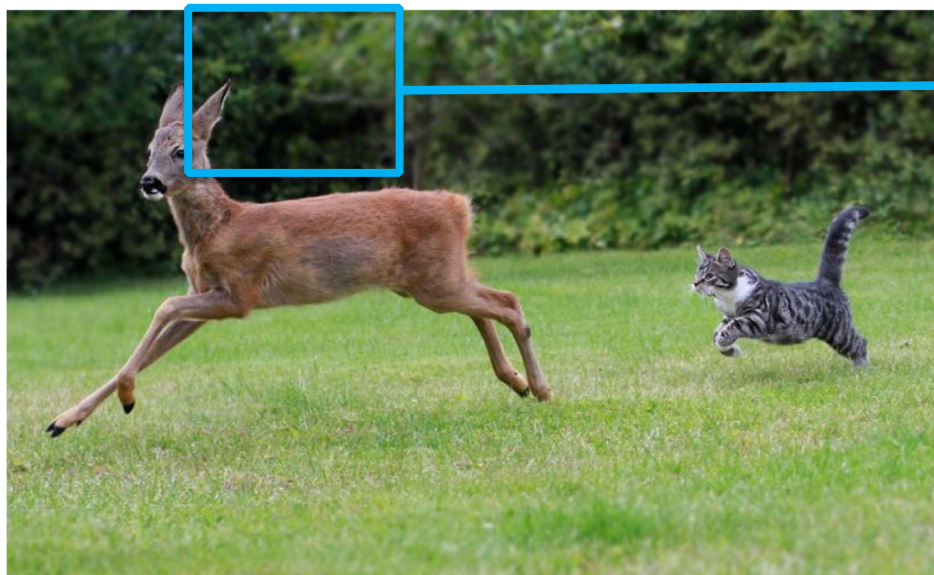
deer?
cat?
background?

Object Detection as Classification



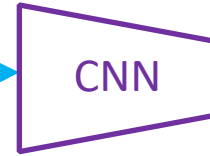
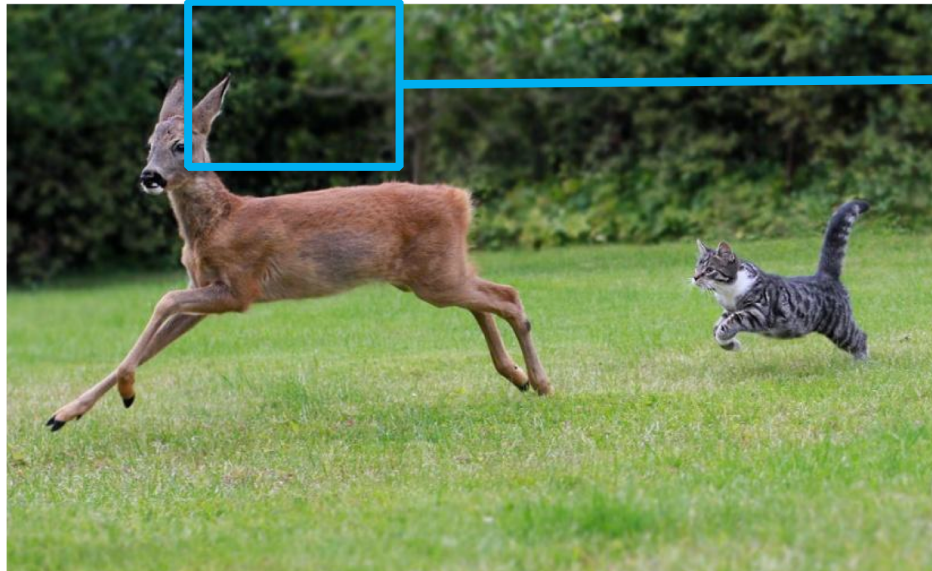
deer?
cat?
background?

Object Detection as Classification



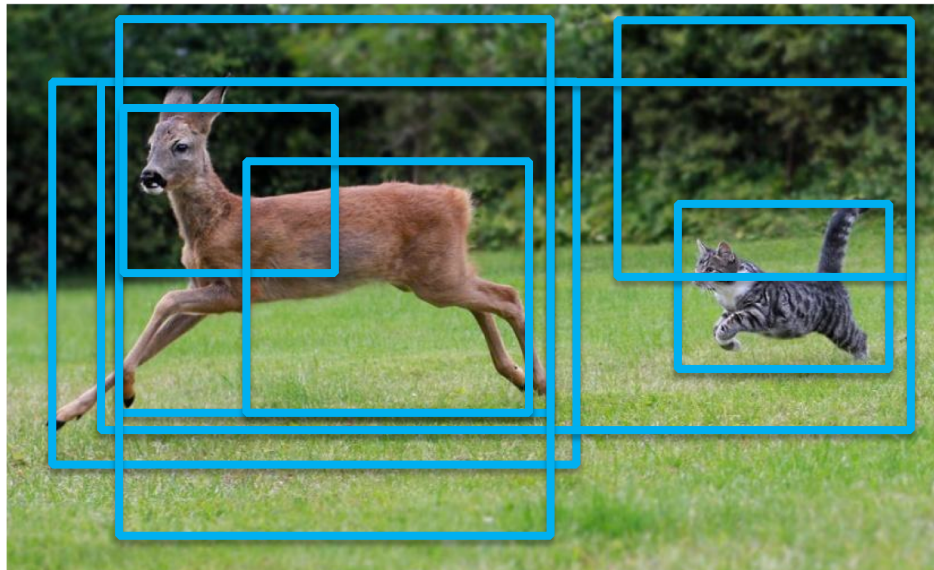
deer?
cat?
background?

Object Detection as Classification with Sliding Window

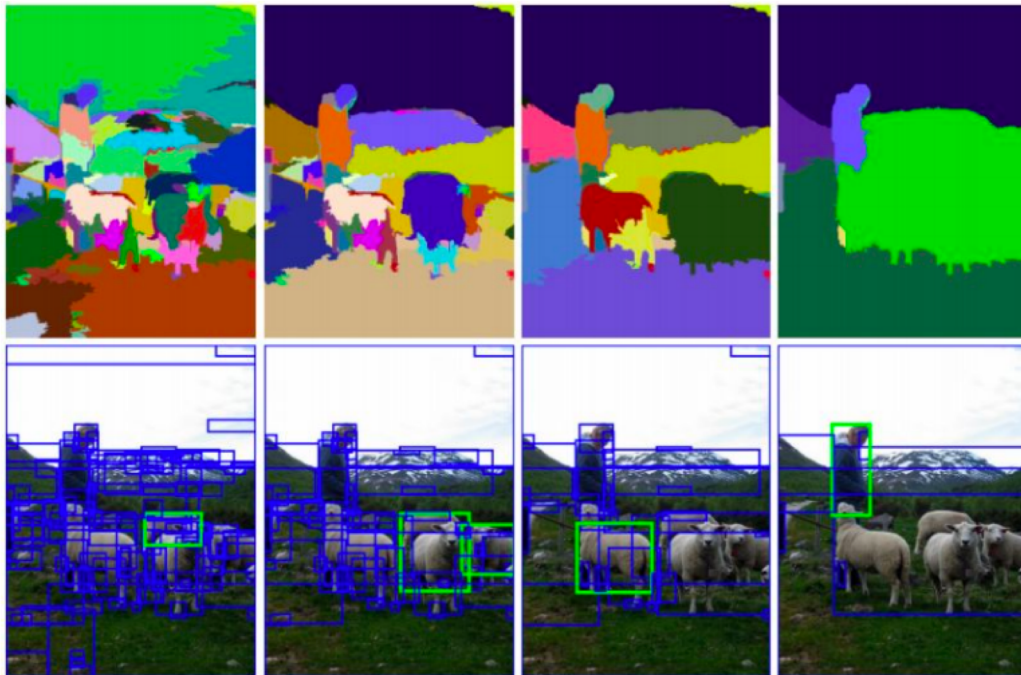


deer?
cat?
background?

Object Detection as Classification with Box Proposals



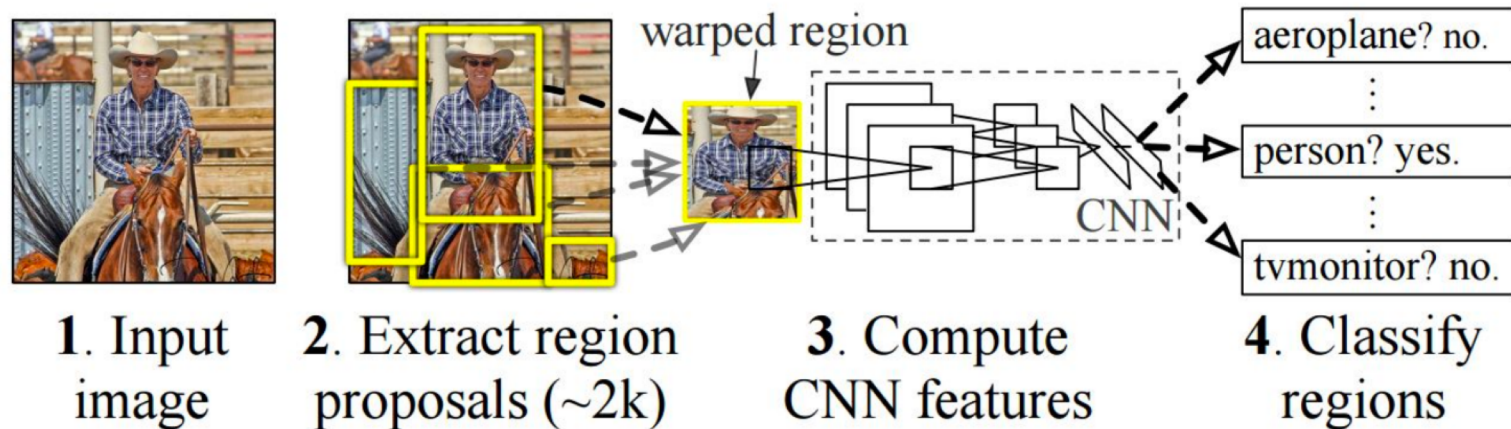
Box Proposal Method – SS: Selective Search



Segmentation As
Selective Search for
Object Recognition. van
de Sande et al. ICCV
2011

RCNN

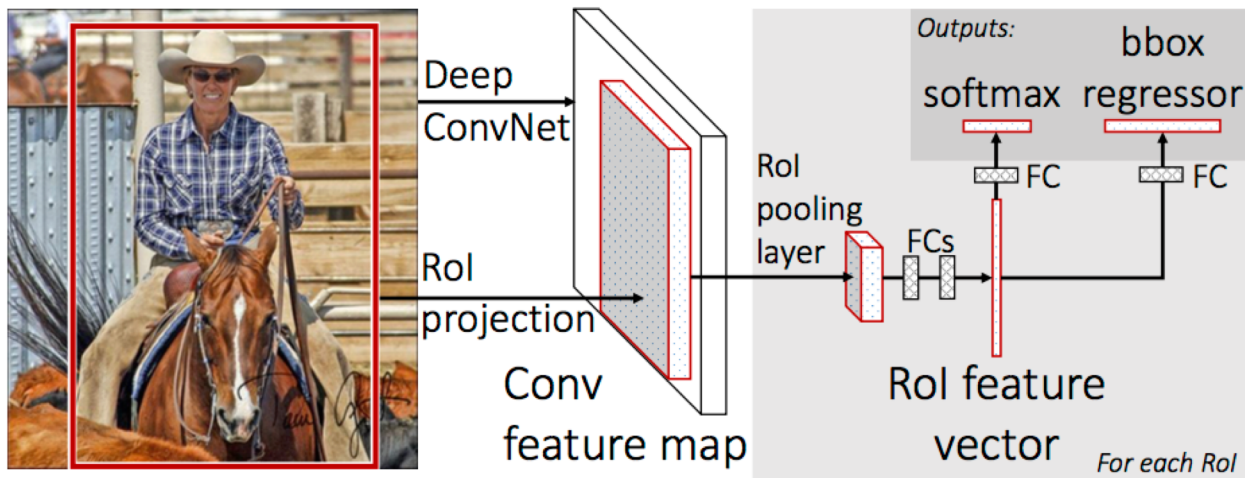
R-CNN: *Regions with CNN features*



<https://people.eecs.berkeley.edu/~rbg/papers/r-cnn-cvpr.pdf>

Rich feature hierarchies for accurate object detection and semantic segmentation. Girshick et al. CVPR 2014.

Fast-RCNN



Idea: No need to recompute features for every box independently,
Regress refined bounding box coordinates.

<https://arxiv.org/abs/1504.08083>

Fast R-CNN. Girshick. ICCV 2015.

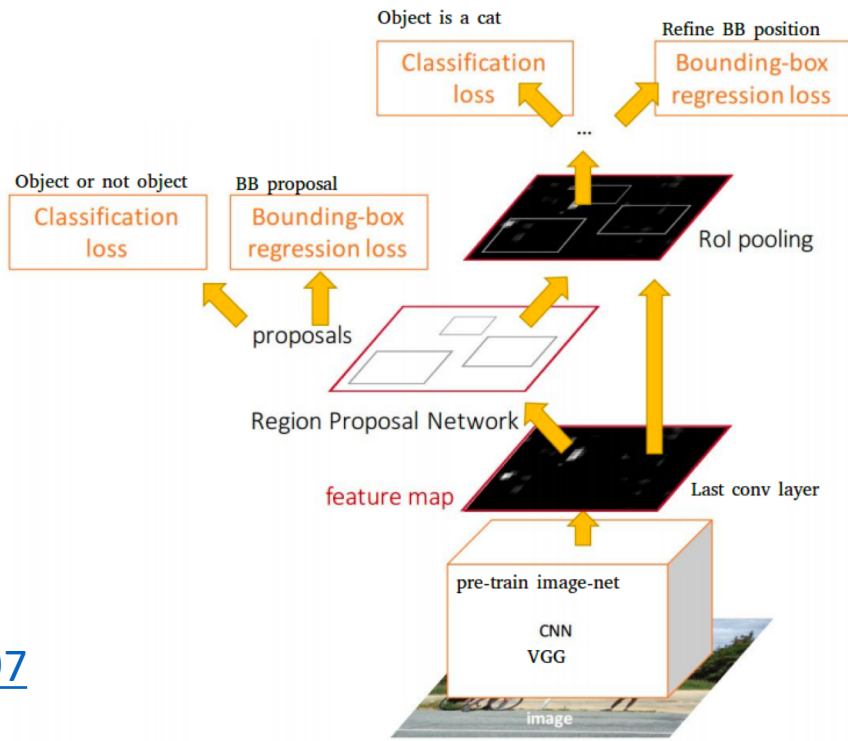
<https://github.com/sunshineatnoon/Paper-Collection/blob/master/Fast-RCNN.md>

Faster-RCNN

Idea: Integrate the Bounding Box Proposals as part of the CNN predictions

<https://arxiv.org/abs/1506.01497>

Ren et al. NIPS 2015.

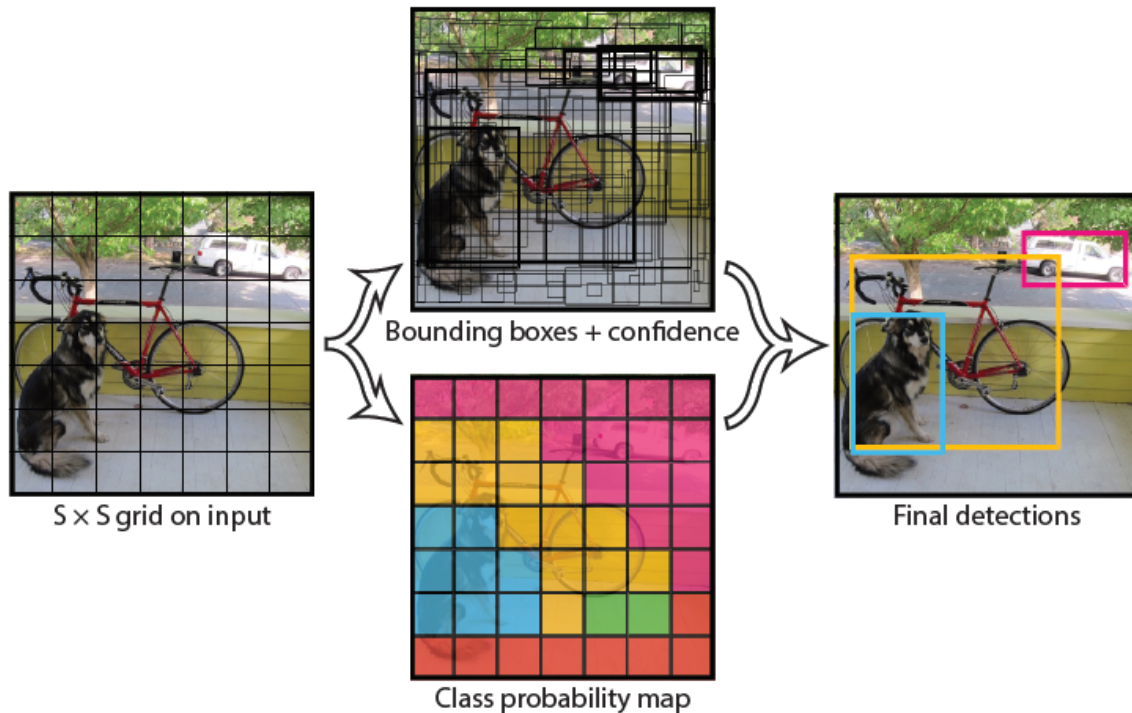


Single-shot Object Detectors

- No two-steps of box proposals + Classification
- Anchor Points for predicting boxes

YOLO- You Only Look Once

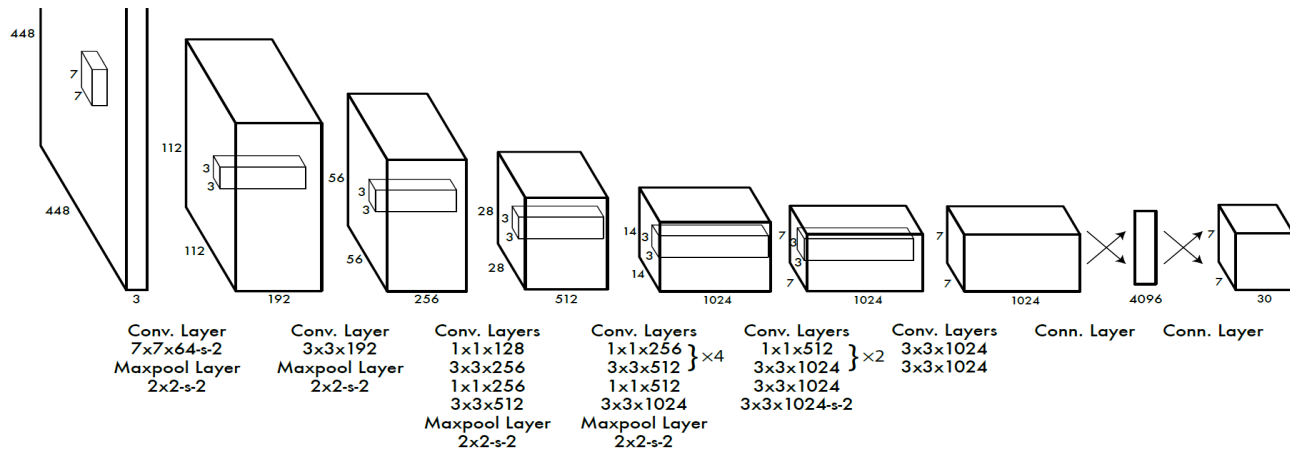
Idea: No bounding box proposals.
Predict a class and a box for every location in a grid.



<https://arxiv.org/abs/1506.02640>

Redmon et al. CVPR 2016.

YOLO- You Only Look Once



Divide the image into 7x7 cells.

Each cell trains a detector.

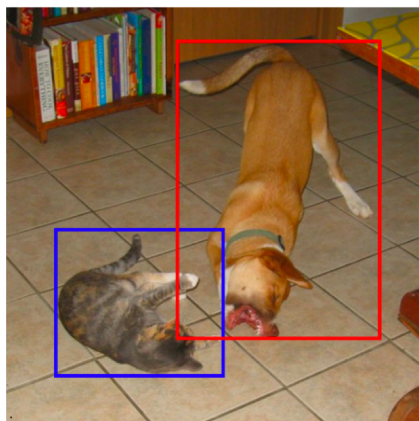
The detector needs to predict the object's class distributions.

The detector has 2 bounding-box predictors to predict bounding-boxes and confidence scores.

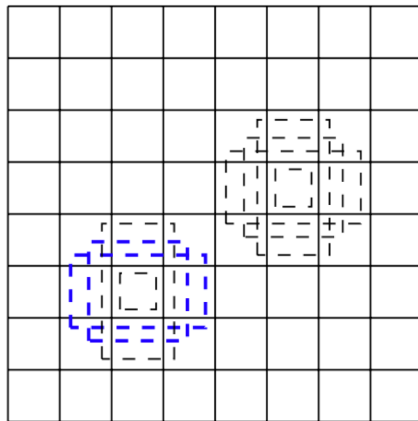
<https://arxiv.org/abs/1506.02640>

Redmon et al. CVPR 2016.

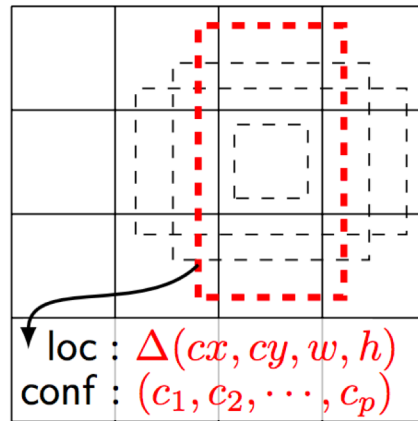
SSD: Single Shot Detector



(a) Image with GT boxes



(b) 8×8 feature map

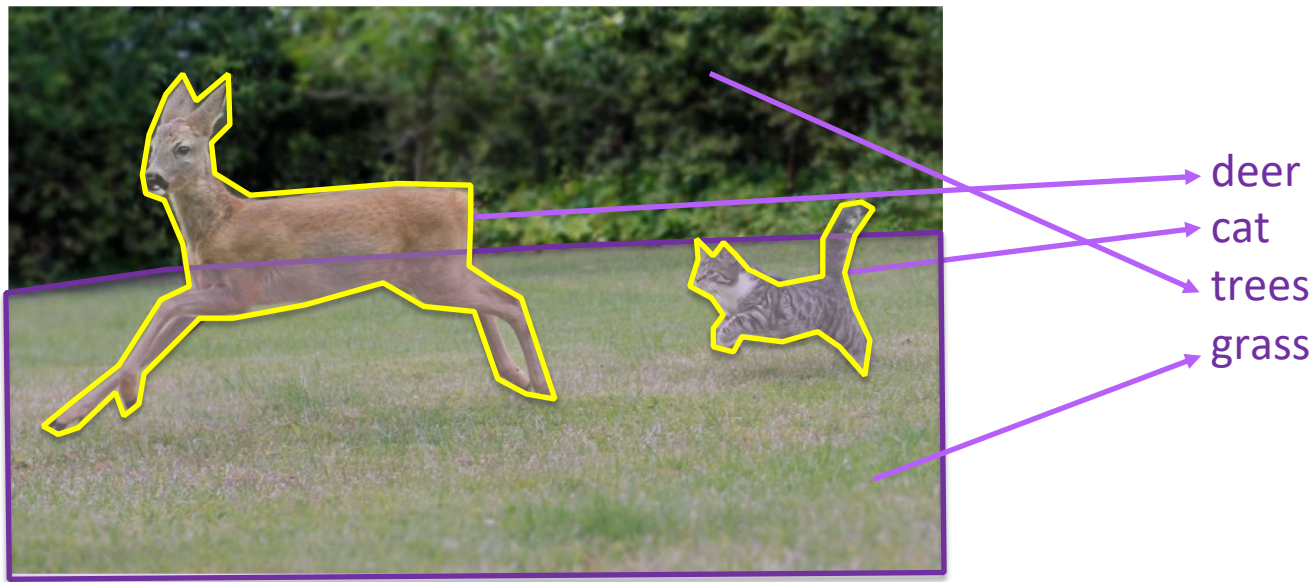


(c) 4×4 feature map

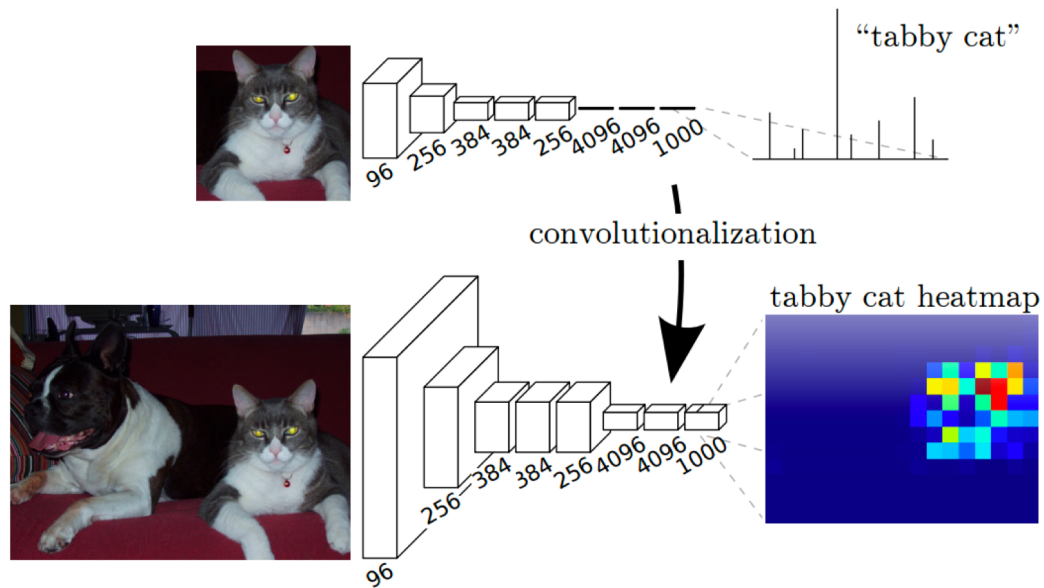
Idea: Similar to YOLO, but denser grid map, multiscale grid maps. + Data augmentation + Hard negative mining + Other design choices in the network.

Liu et al. ECCV 2016.

Semantic Segmentation / Image Parsing



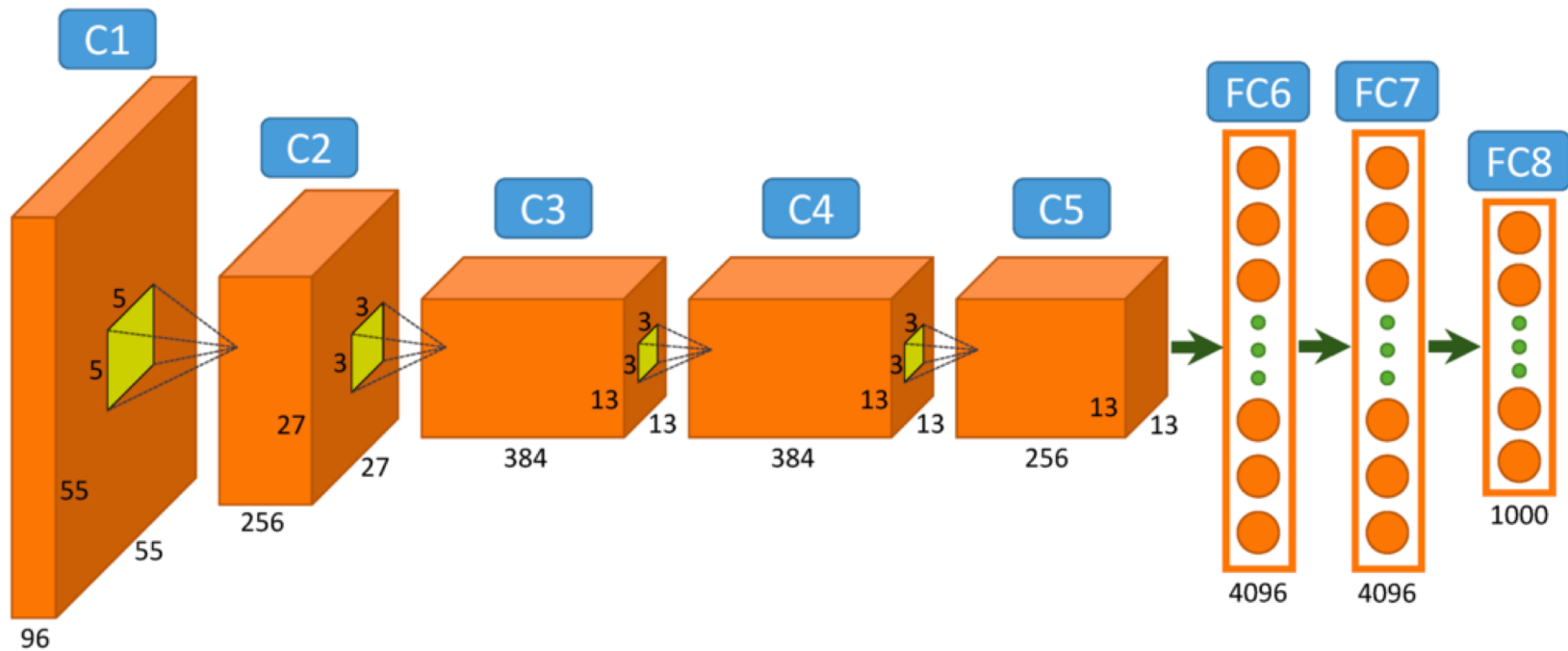
Idea 1: Convolutionalization



However resolution of the segmentation map is low.

https://people.eecs.berkeley.edu/~jonlong/long_shelhamer_fcn.pdf

Alexnet



Idea 1: Convolutionalization

```
nn.Linear(n_inputs, n_outputs) == nn.SpatialConvolution(n_inputs, n_outputs, 1, 1, 1, 1)
```

input tensor:
4096



Linear-layer

W: 4096 x 1000
b: 1000



output tensor:
1000



input tensor:
4096x1x1



SpatialConv

W: 1000x4096x1x1
b: 1000



output tensor:
1000x1x1



Fully Convolutional Networks (CVPR 2015)

Fully Convolutional Networks for Semantic Segmentation

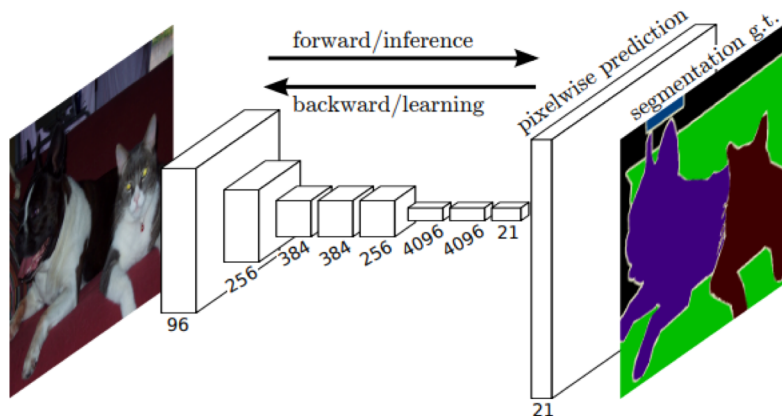
Jonathan Long*

Evan Shelhamer*

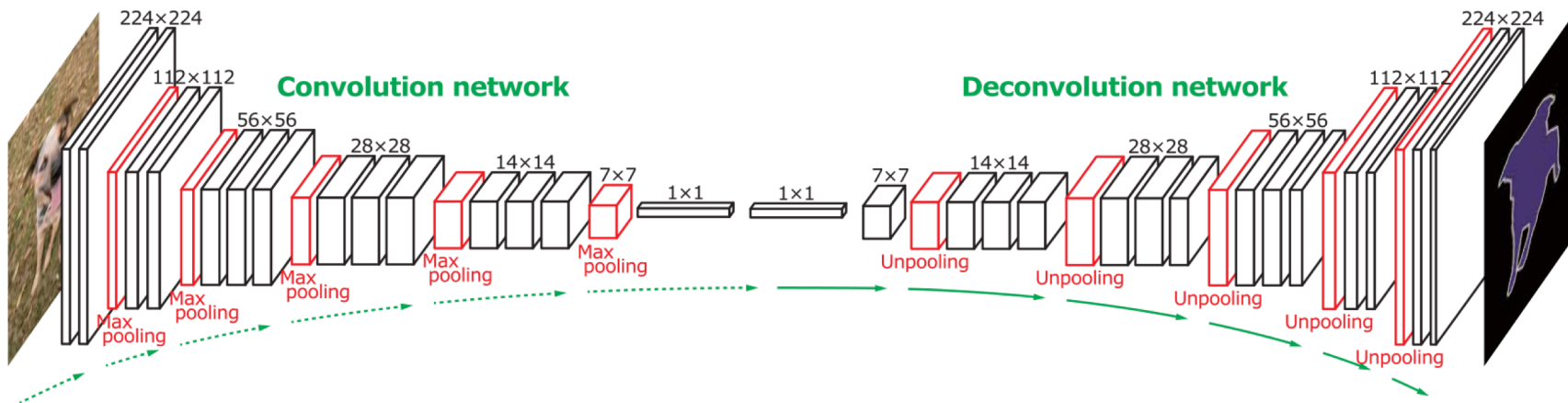
Trevor Darrell

UC Berkeley

`{jonlong, shelhamer, trevor}@cs.berkeley.edu`



Idea 2: Up-sampling Convolutions or "Deconvolutions"

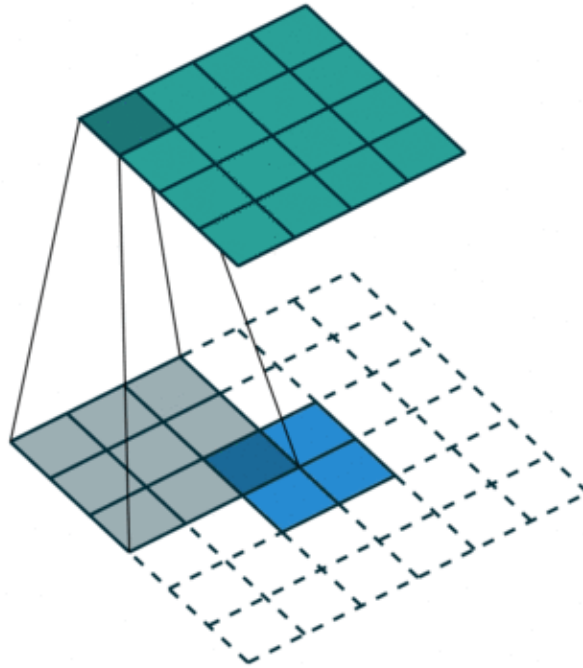


Learning Deconvolution Network for Semantic Segmentation

Hyeonwoo Noh Seunghoon Hong Bohyung Han
Department of Computer Science and Engineering, POSTECH, Korea
{hyeonwoonoh-, maga33, bhhan}@postech.ac.kr

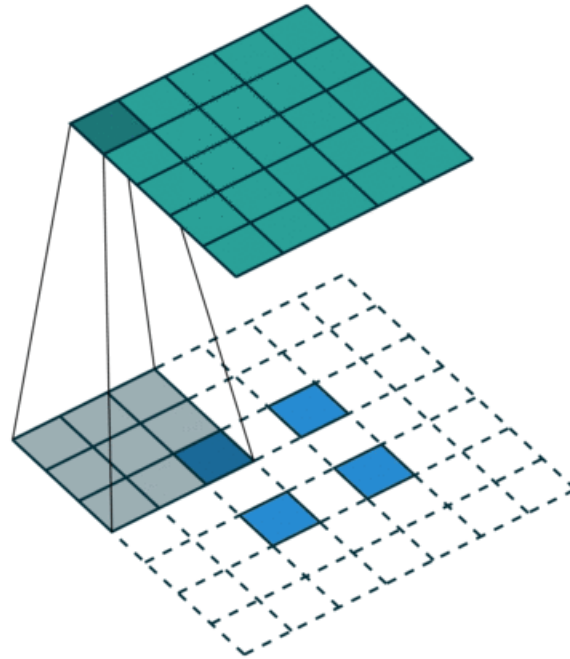
<http://cvlab.postech.ac.kr/research/deconvnet/>

Idea 2: Up-sampling Convolutions or "Deconvolutions"



https://github.com/vdumoulin/conv_arithmetic

Idea 2: Up-sampling Convolutions or "Deconvolutions"



https://github.com/vdumoulin/conv_arithmetic

Idea 2: Up-sampling Convolutions or "Deconvolutions"

Deconvolutional Layers

Upconvolutional Layers

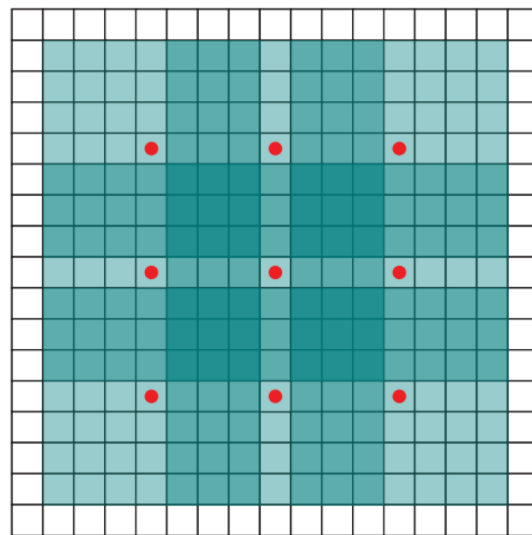
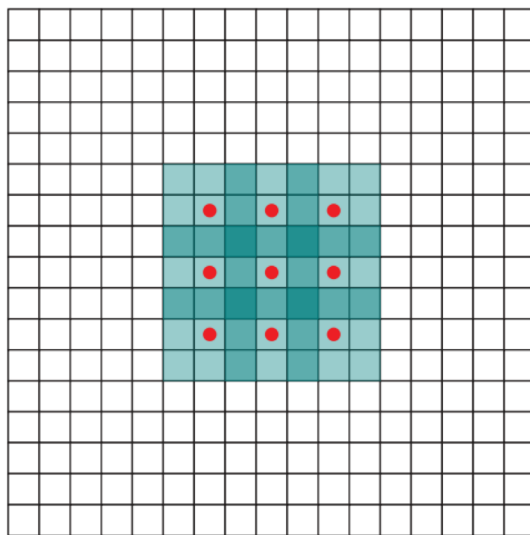
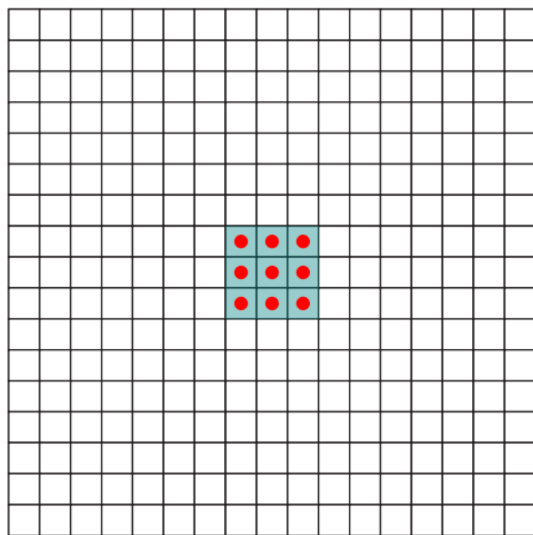
Backwards Strided
Convolutional Layers

Fractionally Strided
Convolutional Layers

Transposed
Convolutional Layers

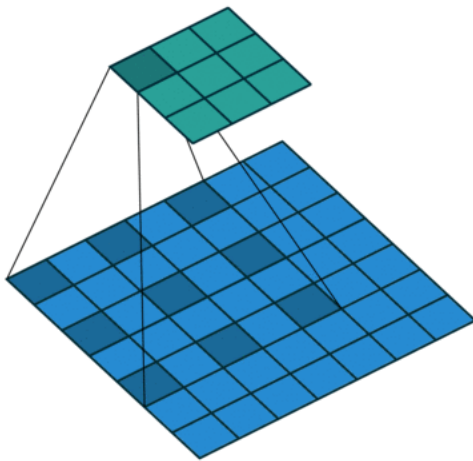
Spatial Full
Convolutional Layers

Idea 3: Dilated Convolutions



MULTI-SCALE CONTEXT AGGREGATION BY
DILATED CONVOLUTIONS

Idea 3: Dilated Convolutions



MULTI-SCALE CONTEXT AGGREGATION BY
DILATED CONVOLUTIONS

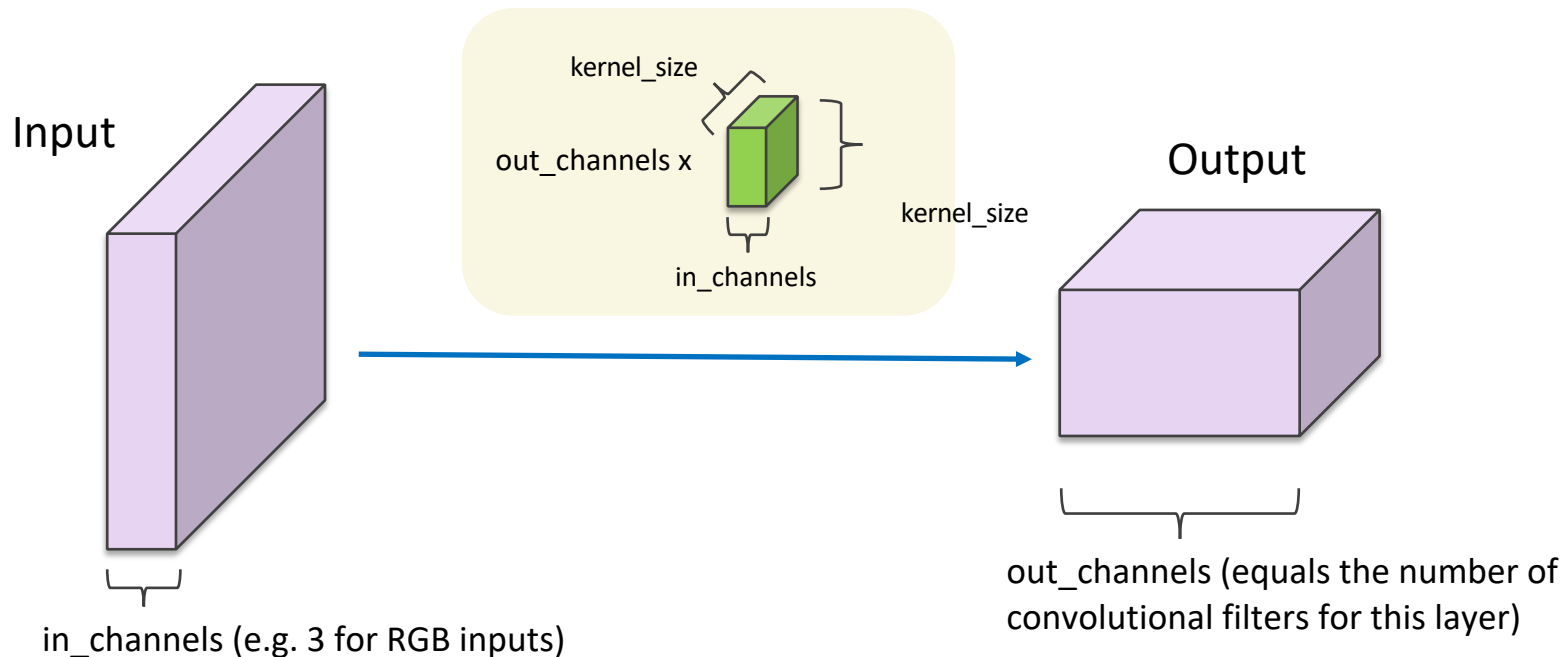
Fisher Yu
Princeton University

Vladlen Koltun
Intel Labs

ICLR 2016

Convolutional Layer in pytorch

```
class torch.nn.Conv2d(in_channels, out_channels, kernel_size, stride=1, padding=0, dilation=1, groups=1, bias=True) \[source\]
```



Questions?