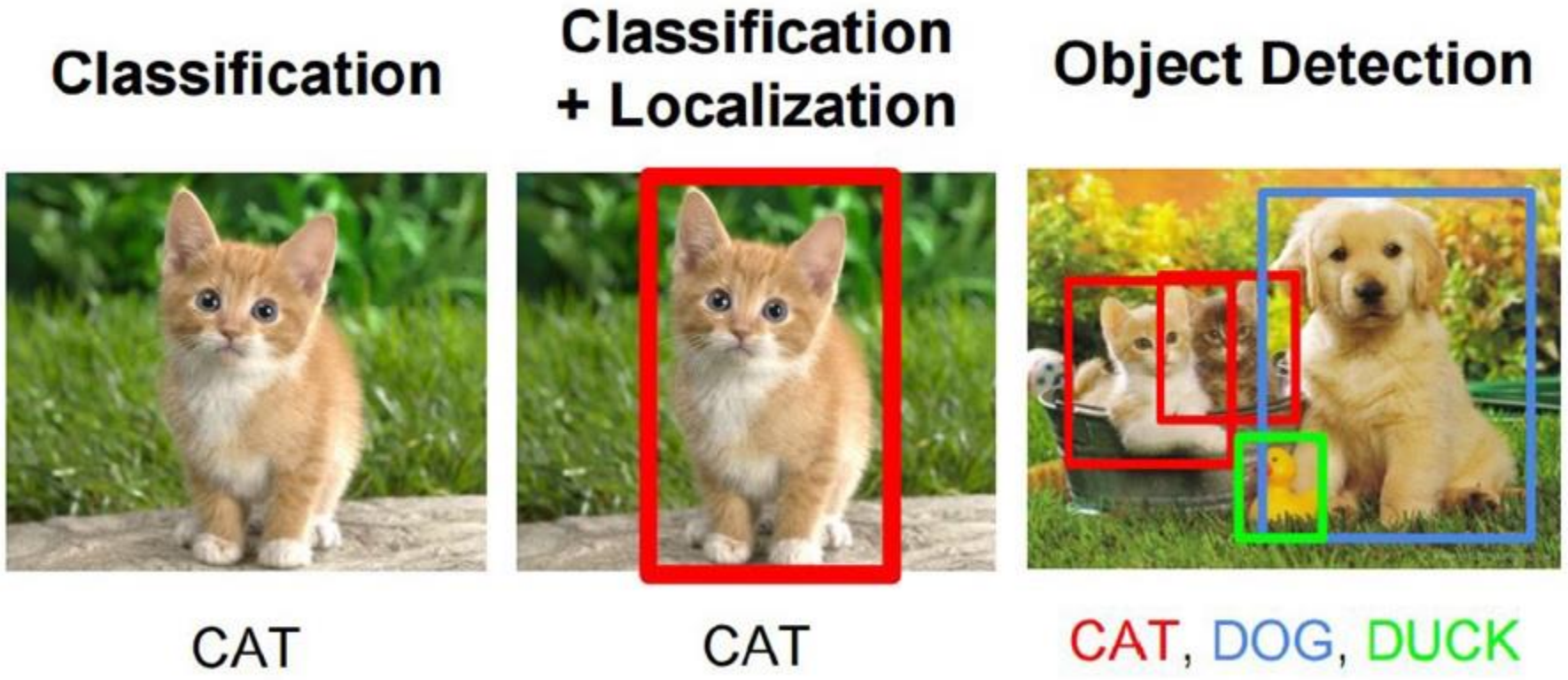
Overfeat :
Integrated Recognition, Localization and Detectionusing Convolutional Networks

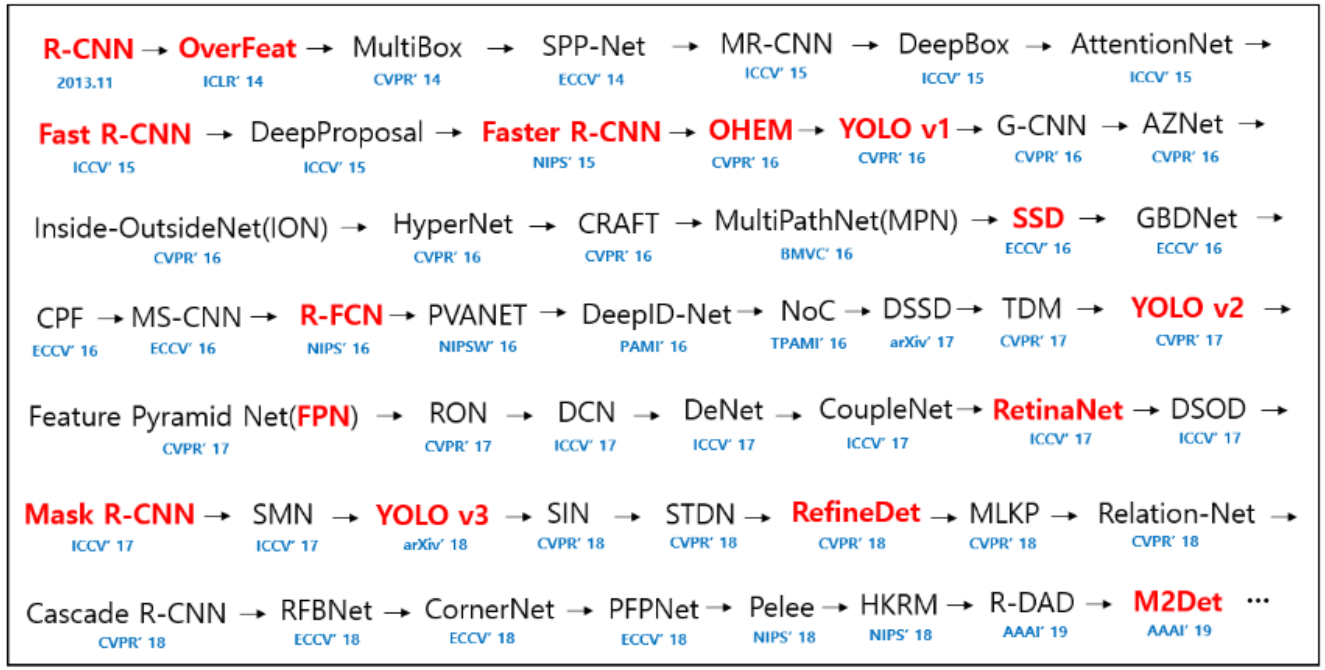
paper seminar

2조

Object detection?

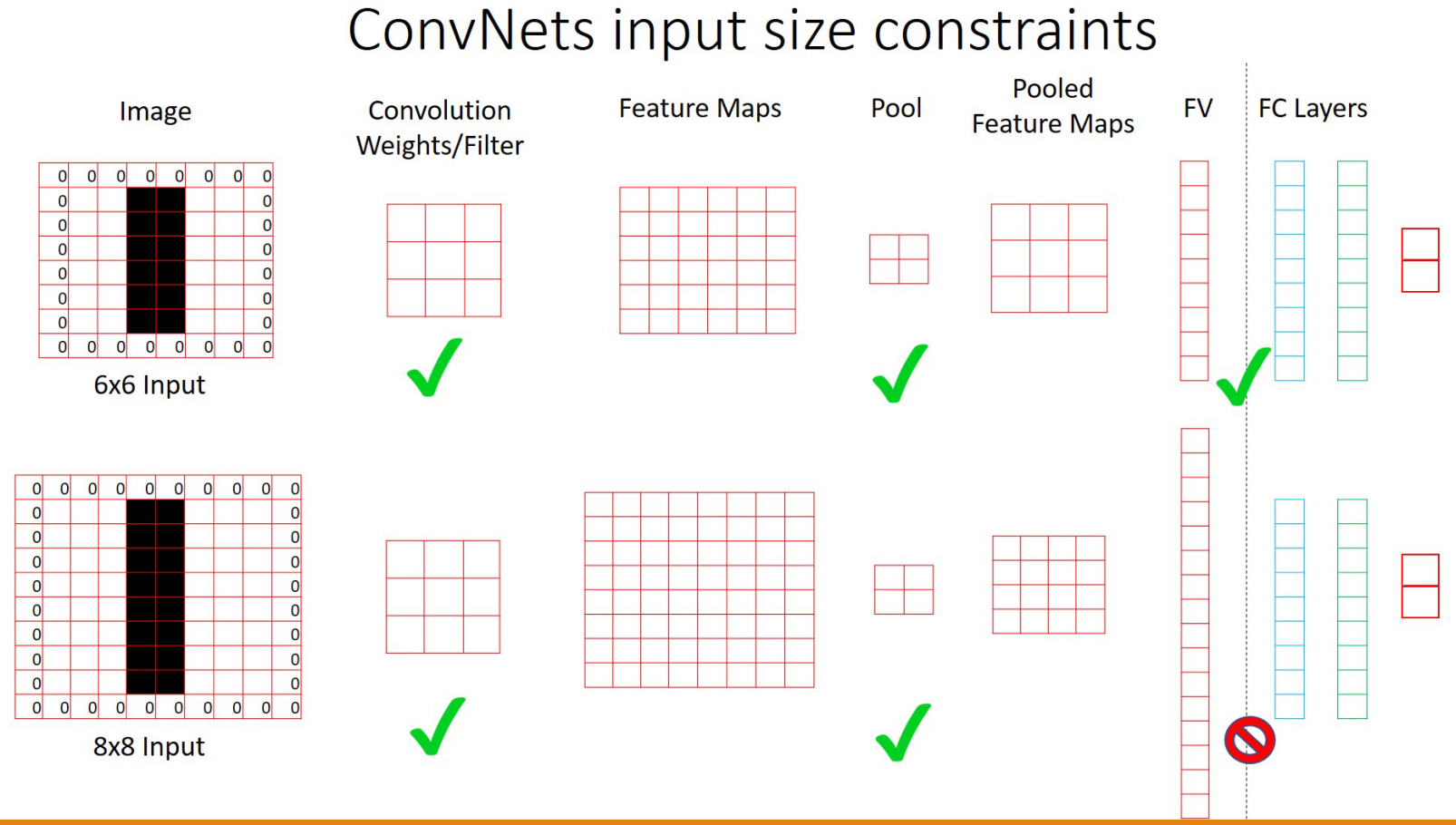


Object detection?

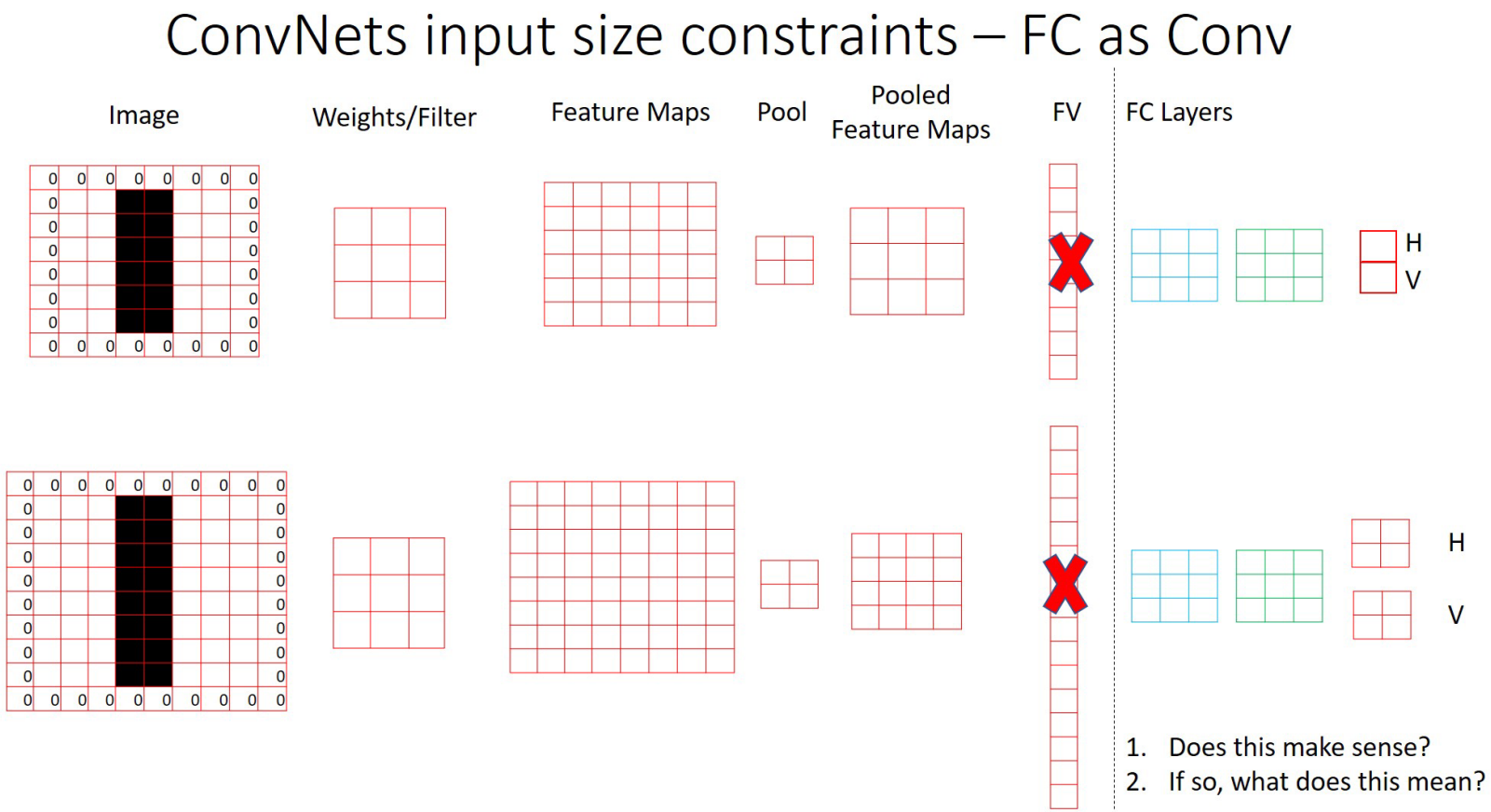


Convnet’s input size constraints?

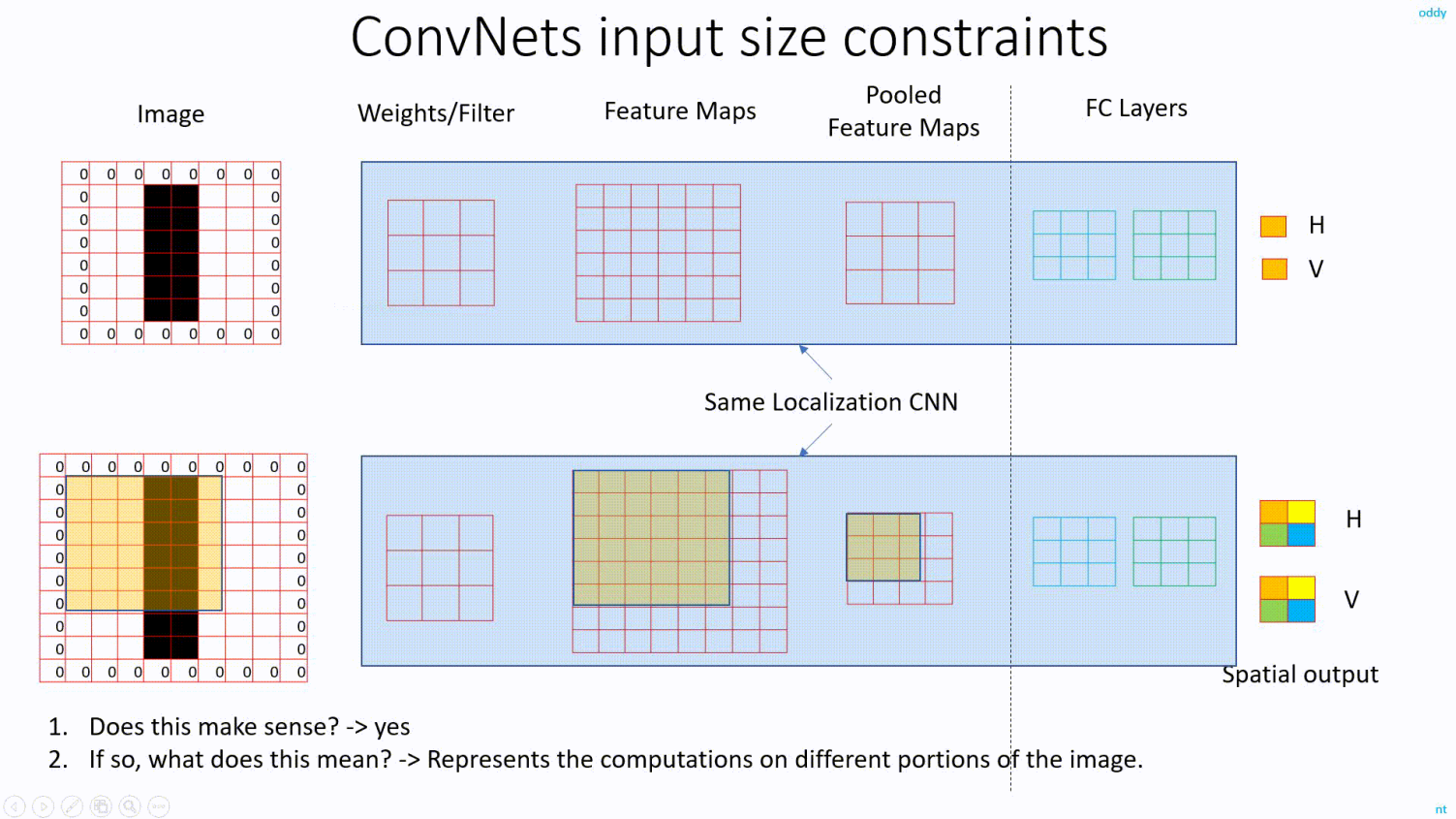
ConvNet의 입력 크기 제약 조건
(Alexnet 224)



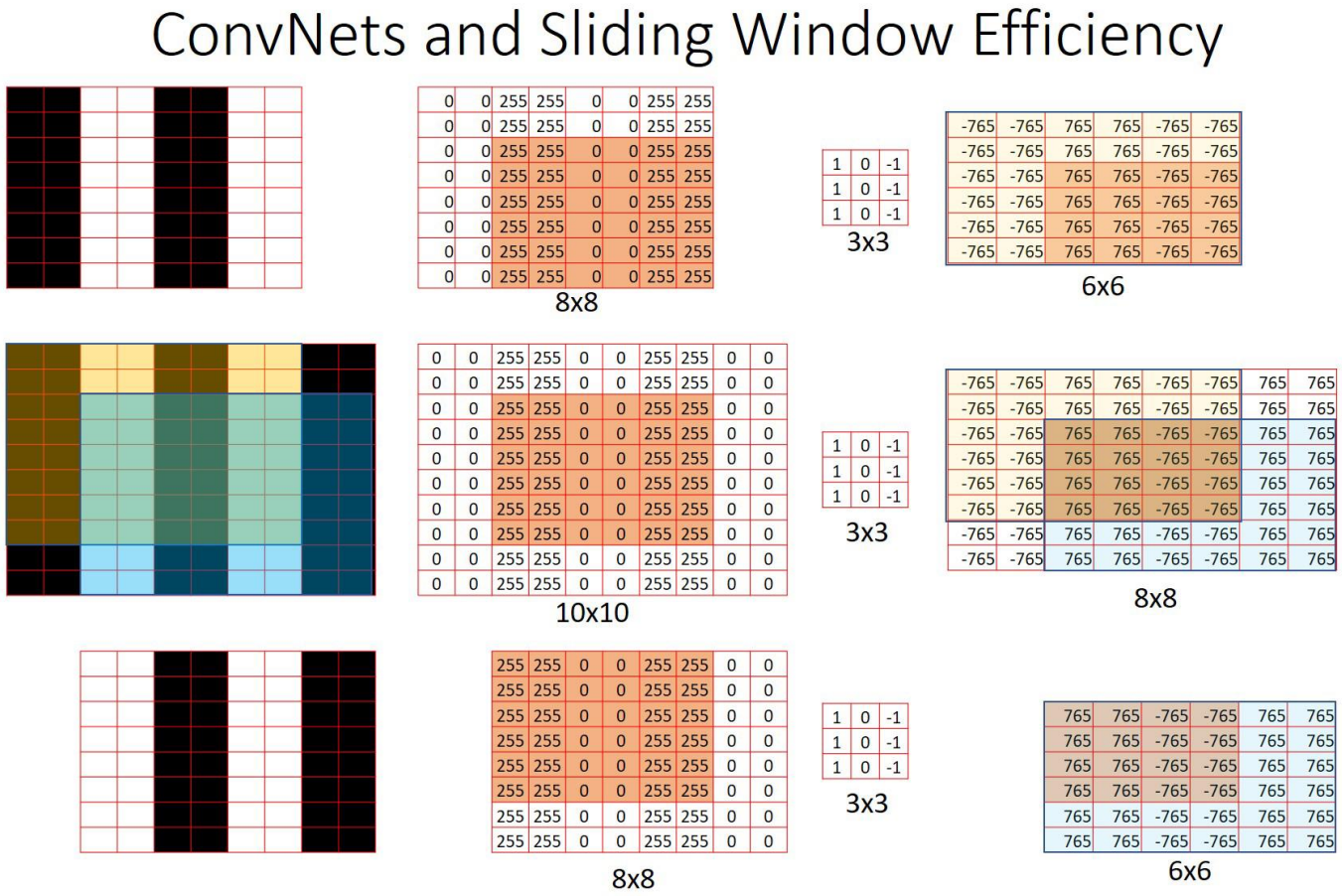
Receptive field and spatial output



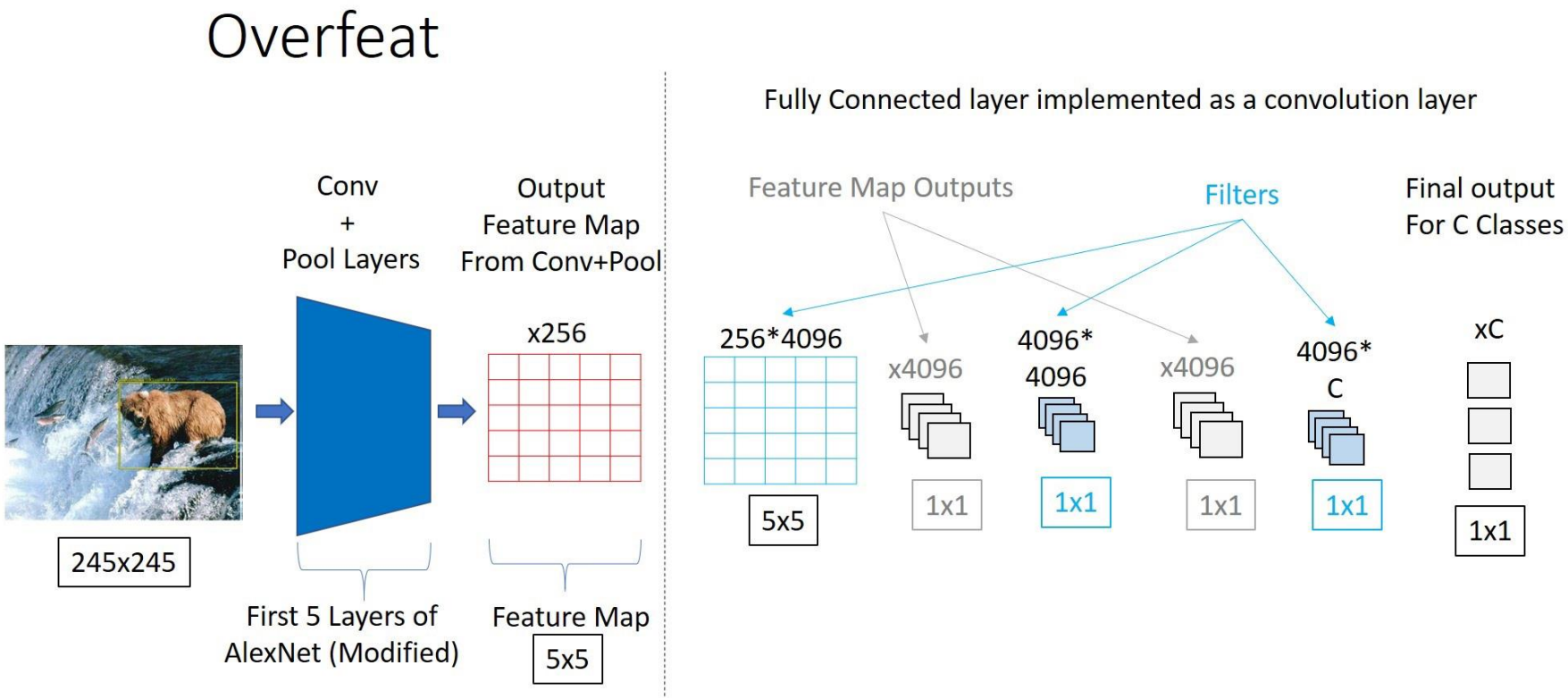
Receptive field and spatial output



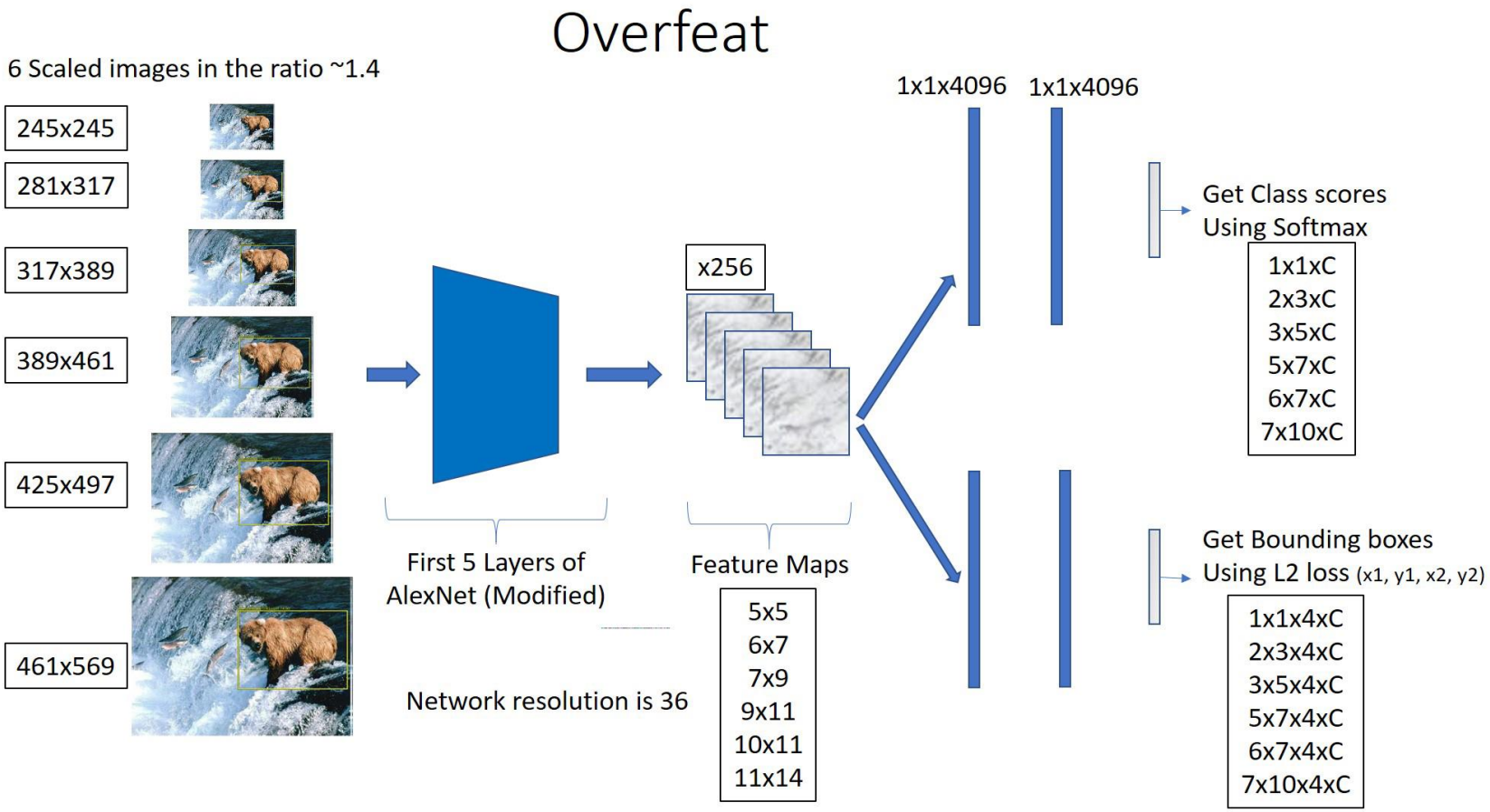
Convnet and sliding window efficiency



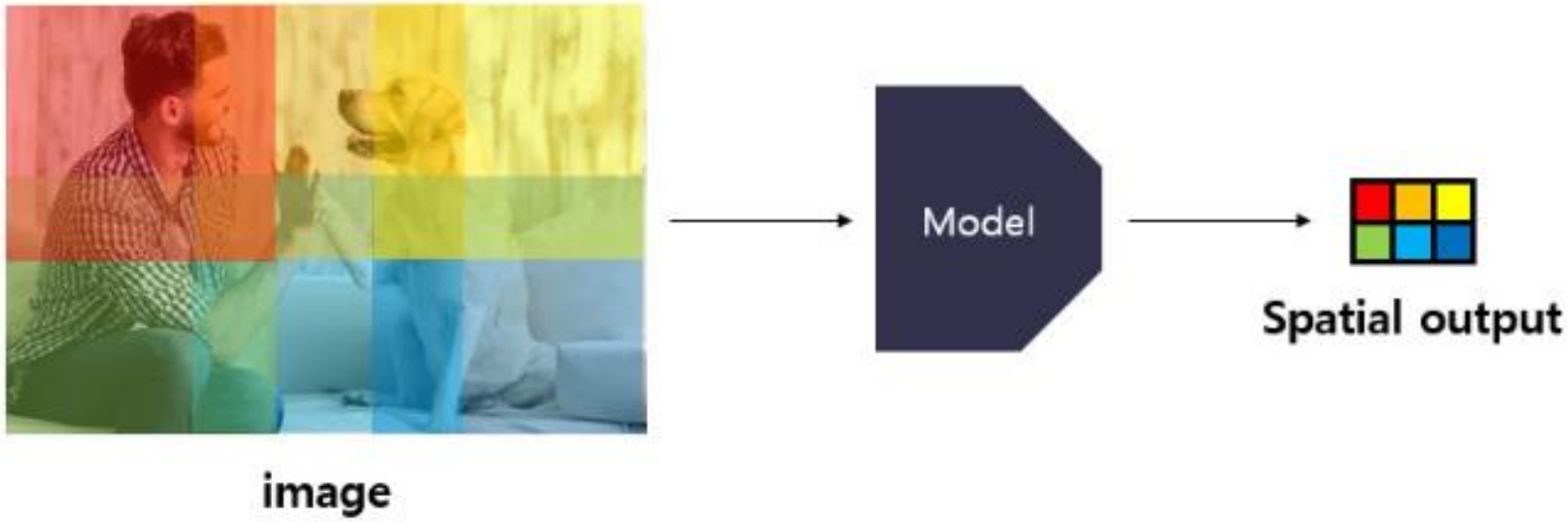
overfeat


Overfeat
Conv
+
Pool Layers
Output
Feature Map
From Conv+Pool
x256
245x245
First 5 Layers of
AlexNet (Modified)
Feature Map
5x5
Fully Connected layer implemented as a convolution layer
Feature Map Outputs
Filters
Final output
For C Classes
256*4096
x4096
4096*
4096
x4096
4096*
C
xC
5x5
1x1
1x1
1x1
1x1
1x1

overfeat

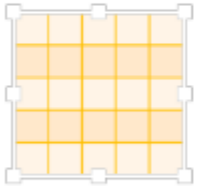

Overfeat
6 Scaled images in the ratio ~1.4
245x245
281x317
317x389
389x461
425x497
461x569
First 5 Layers of AlexNet (Modified)
x256
Feature Maps
5x5
6x7
7x9
9x11
10x11
11x14
Network resolution is 36
1x1x4096
1x1x4096
Get Class scores Using Softmax
1x1xC
2x3xC
3x5xC
5x7xC
6x7xC
7x10xC
Get Bounding boxes Using L2 loss (x1, y1, x2, y2)
1x1x4xC
2x3x4xC
3x5x4xC
5x7x4xC
6x7x4xC
7x10x4xC

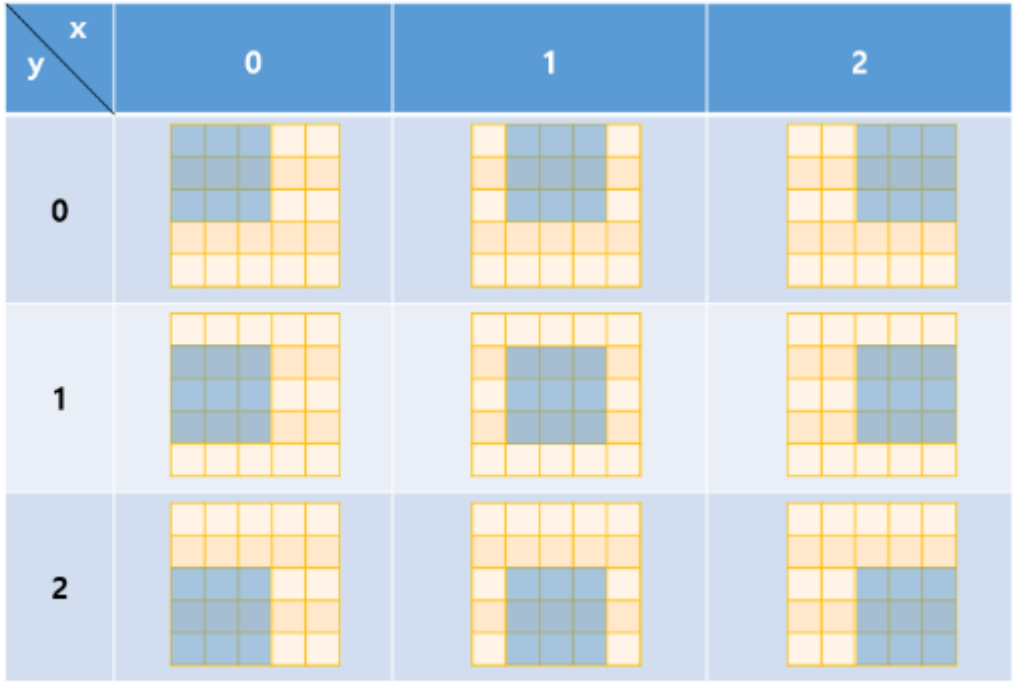
overfeat



Spatial output example

overfeat



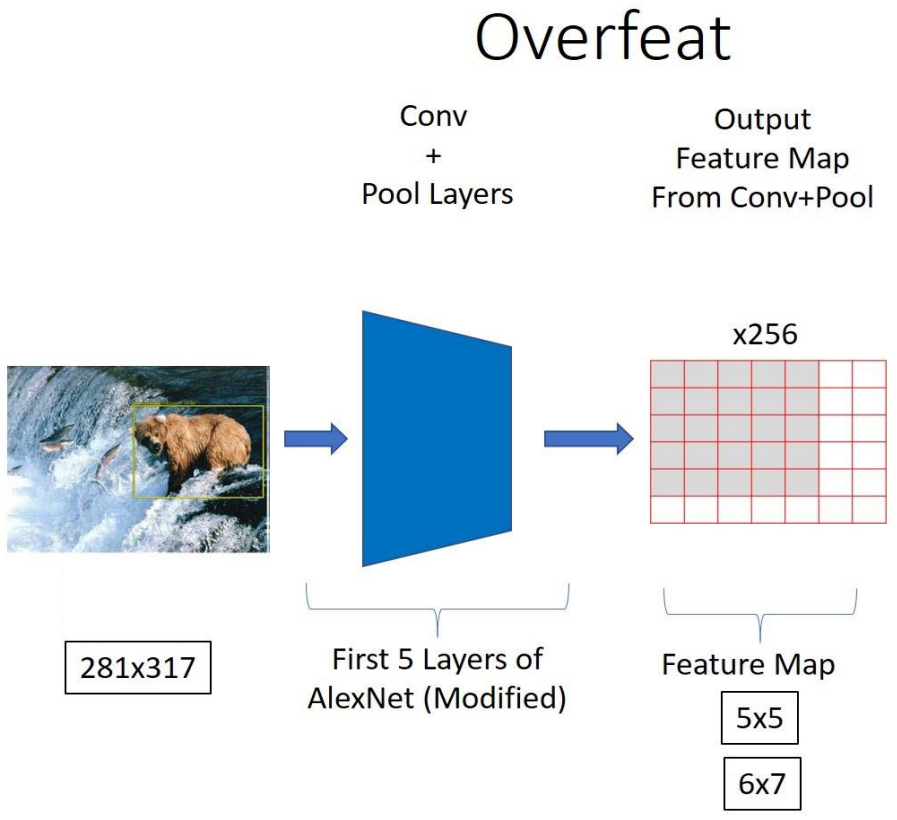


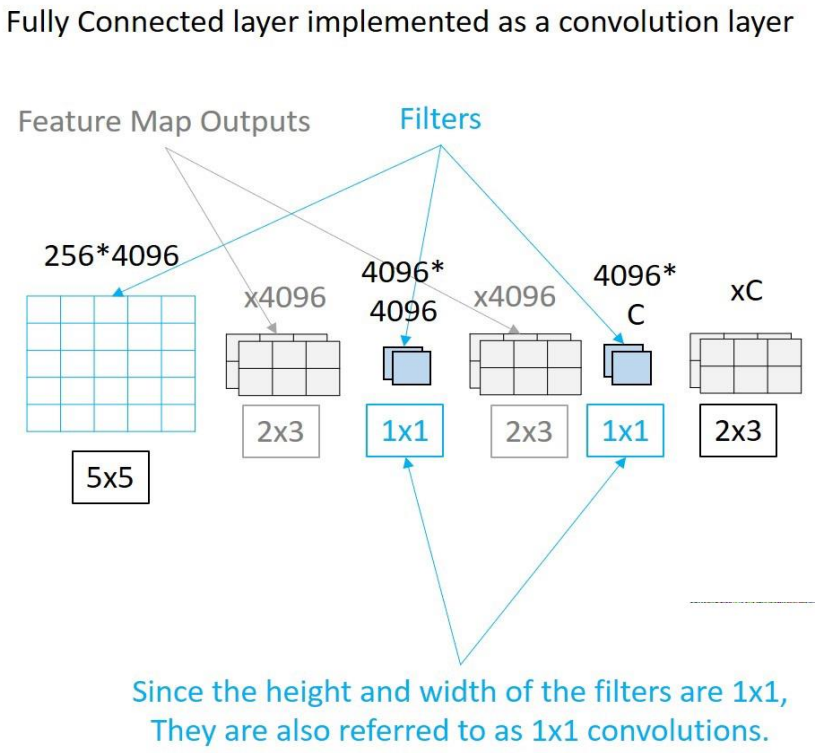
overfeat

Scale	Input size	Layer 5 pre-pool	Layer 5 post-pool	Classifier map (pre-reshape)	Classifier map size
1	245x245	17x17	(5x5)x(3x3)	(1x1)x(3x3)xC	3x3xC
2	281x317	20x23	(6x7)x(3x3)	(2x3)x(3x3)xC	6x9xC
3	317x389	23x29	(7x9)x(3x3)	(3x5)x(3x3)xC	9x15xC
4	389x461	29x35	(9x11)x(3x3)	(5x7)x(3x3)xC	15x21xC
5	425x497	32x35	(10x11)x(3x3)	(6x7)x(3x3)xC	18x24xC
6	461x569	35x44	(11x14)x(3x3)	(7x10)x(3x3)xC	21x30xC

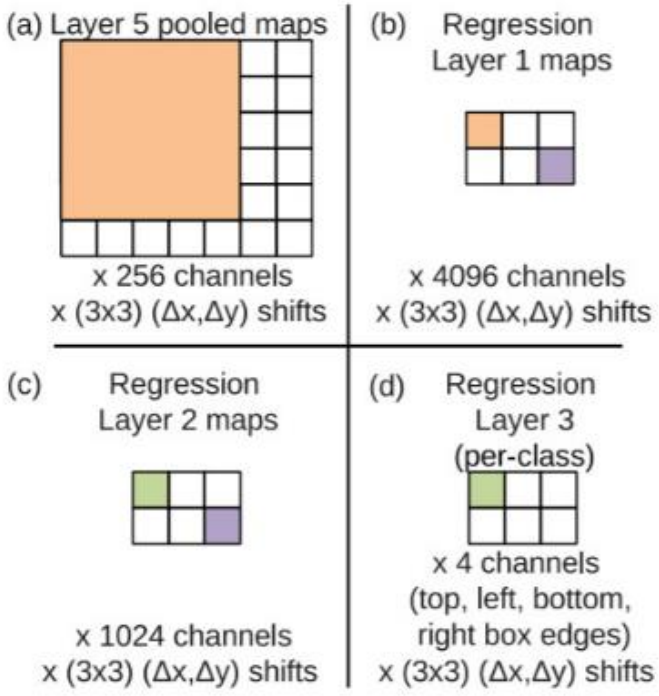
Spatial dimension of multi scale approach

overfeat


Overfeat
Conv
+
Pool Layers
Output
Feature Map
From Conv+Pool
x256
281x317
First 5 Layers of
AlexNet (Modified)
Feature Map
5x5
6x7


Fully Connected layer implemented as a convolution layer
Feature Map Outputs
Filters
256*4096
x4096
4096*
4096
x4096
4096*
C
xC
5x5
2x3
1x1
2x3
1x1
2x3
Since the height and width of the filters are 1x1,
They are also referred to as 1x1 convolutions.

overfeat



Bounding box regressor Inference Process

overfeat

overfeat

1. Receptive Field
2. Implementing FC layers as convolution operation
3. ConvNet’s Sliding Window Efficiency
4. 1x1 Convolution
5. Spatial Output
6. Effective Stride
7. Confidence score thresholding
8. Non Max Suppression and IoU

Thank You

INTEGRATED RECOGNITION, LOCALIZATION AND DETECTION USING CONVOLUTIONAL NETWORKS