CMPE 258, Deep Learning

Object Detection

April 5, 2018

DMH 149A

Taehee Jeong
Ph.D., Data Scientist

SJSU SAN JOSÉ STATE UNIVERSITY

Assignment_5

Due April 8th, 2018

Deadline for re-submitting is April 15th, 2018

Grading policy:

The code is supposed to be executable without any extra effort and produce reasonable result within 50 minutes.

If the code cannot be executable with any error or taking more than 50 minutes, 50 points will be assigned.

If the code can be executable without any error within 50 minutes, score will be assigned as following formula.

Score = (10 – cost) * 10

Re-submitting is available until March 15th, but 10 point will be deducted every re-submitting after March 8th.

If extra effort is needed to get reasonable result (whatever it is), 5 to 10 points will be deducted.

You may use your trained weights and bias (transfer learning). In this case, please make sure to submit the trained weights and bias as one separate file (*para_yourFirstName_LastName.hdf5)*

Mid-term Exam_2

Start Morning on April 12th .

End the midnight on April 15th

Image classification using CNN

Group Project Proposal

Title submission deadline: April 9th

- Project title
- List of Members
- Preferred presentation day: 4/12 or 4/24

Group Project Proposal

Content during proposal

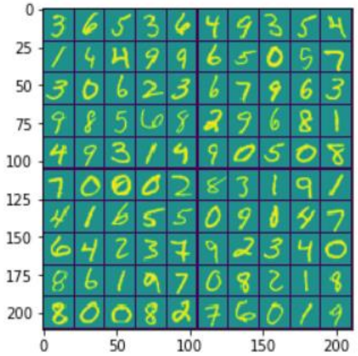
- Justification for the project
- Background: any relevant previous work
- How to collect data set
- Which algorithms / platform will be used
- What is the role for each team member

Today's lesson

Object detection

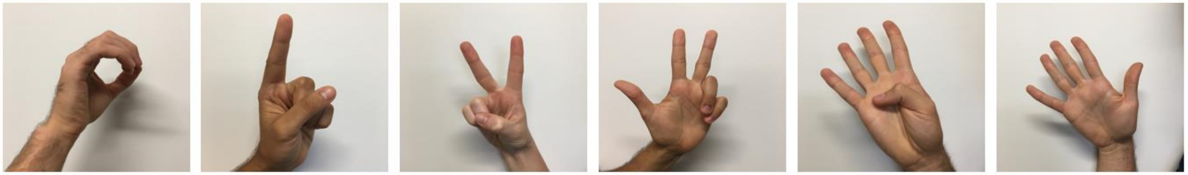
- Sliding windows
- 1 x 1 convolution
- Bounding box
- Intersection over union
- Non-max suppression

Image classification



Images for Hand written digits

Signs images



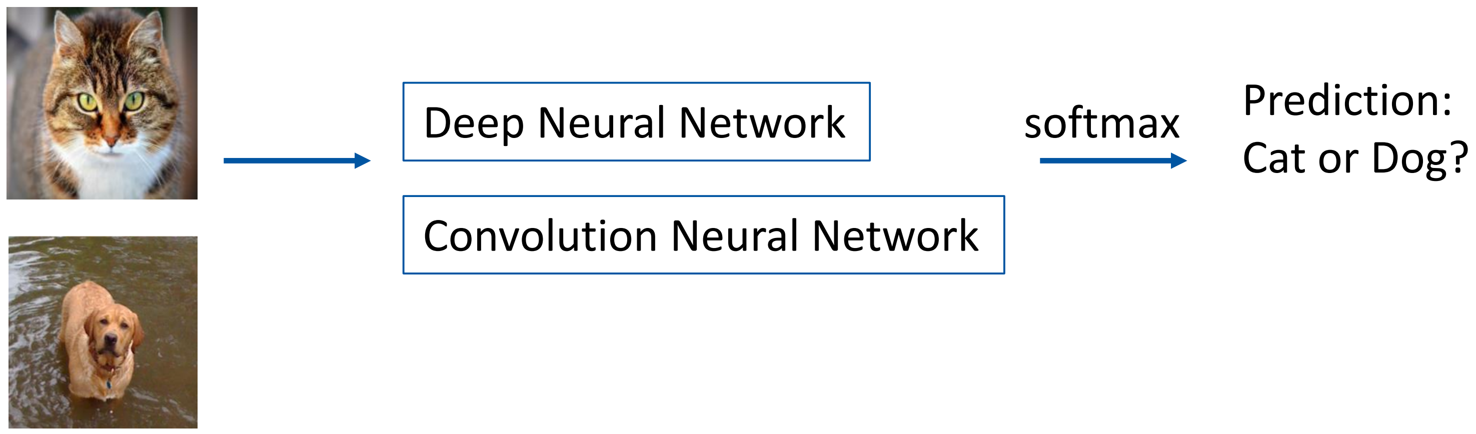
y = 0 y = 1 y = 2 y = 3 y = 4 y = 5

Coursera (Deep Learning specialization)

Image classification

Input image

Output



Deep Neural Network

Convolution Neural Network

softmax

Prediction:
Cat or Dog?

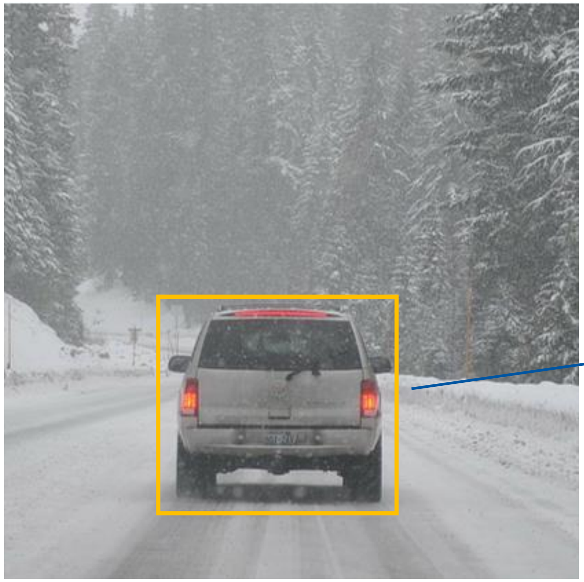
Object localization

Where is the car located in the image?

<deep learning, Andrew Ng>

Pictures taken from a camera while driving around the Silicon Valley [drive.ai](https://www.drive.ai/)

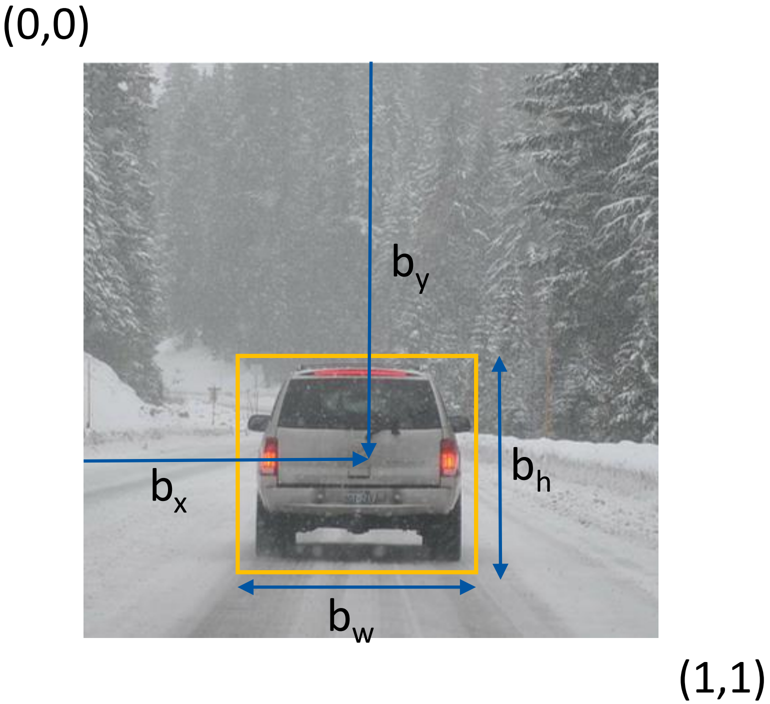
Object localization



Bounding box

<deep learning, Andrew Ng>

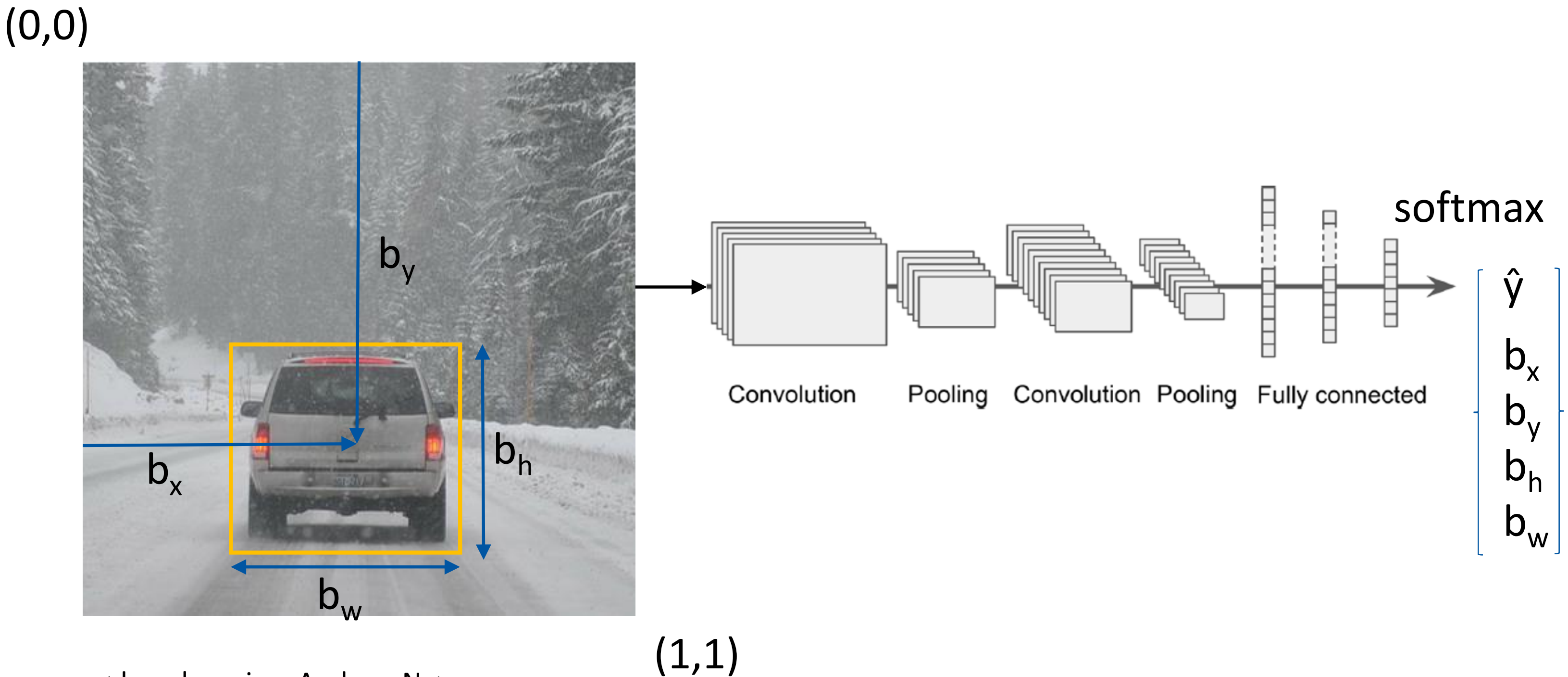
Classification with localization



Bounding box
$b_x = 0.5$
$b_y = 0.7$
$b_h = 0.3$
$b_w = 0.4$

<deep learning, Andrew Ng>

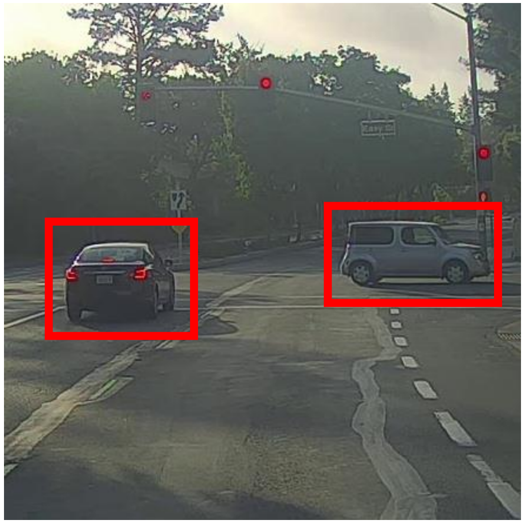
Classification with localization



$$\begin{bmatrix} \hat{y} \\ b_x \\ b_y \\ b_h \\ b_w \end{bmatrix}$$

<deep learning, Andrew Ng>

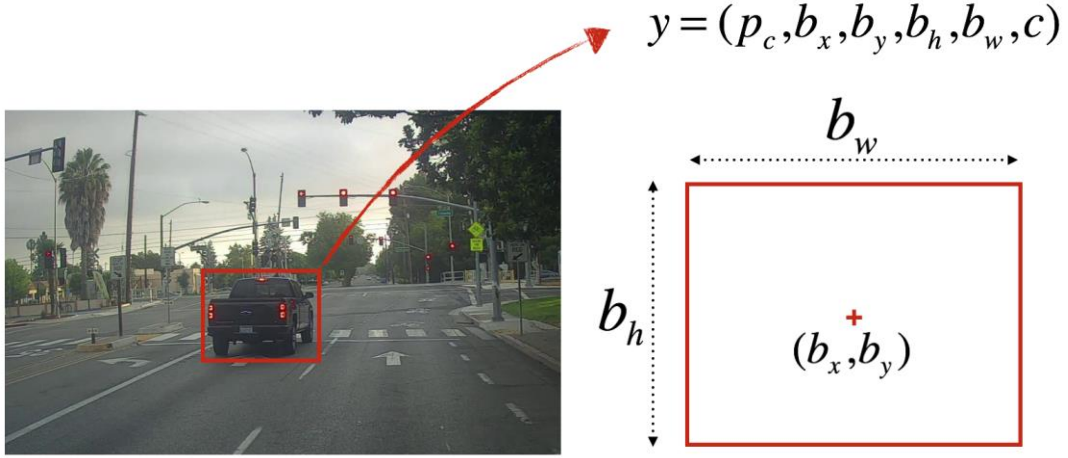
Object detection



How about multiple objects in one image?

<deep learning, Andrew Ng>

Example of bounding box



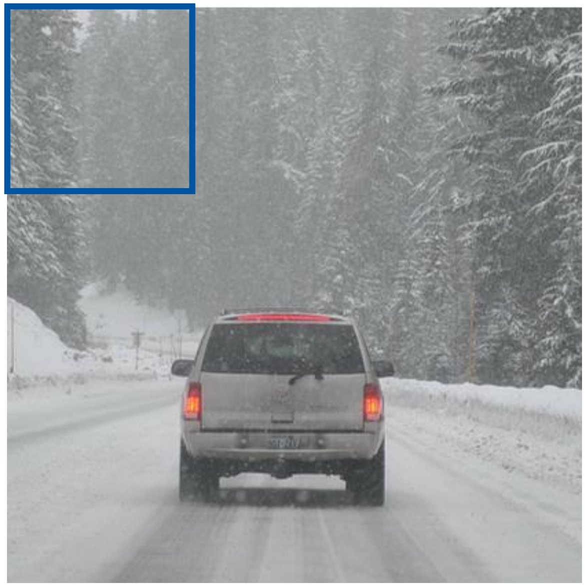
$$y=(p_c,b_x,b_y,b_h,b_w,c)$$

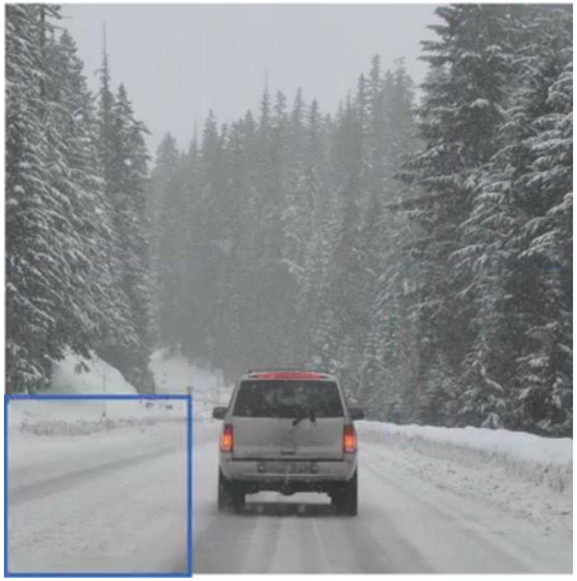
$p_c=1$: confidence of an object being present in the bounding box

$c=3$: class of the object being detected (here 3 for “car”)

<deep learning, Andrew Ng>

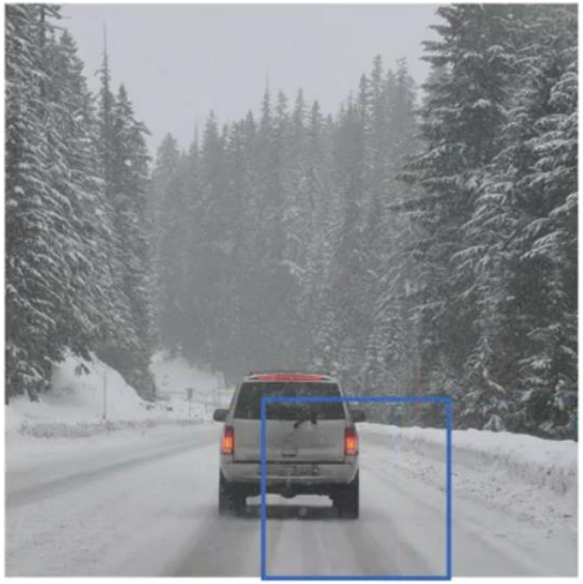
Sliding windows detection





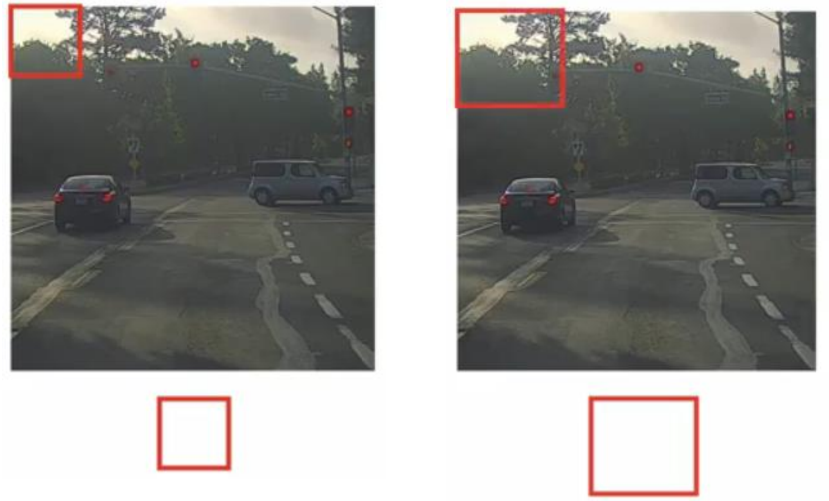
<Deep Learning, Andrew Ng>

Sliding windows detection



<Deep Learning, Andrew Ng>

Sliding windows detection



Instead of one size,
Many different size of windows (from small one to larger one) are needed to apply.

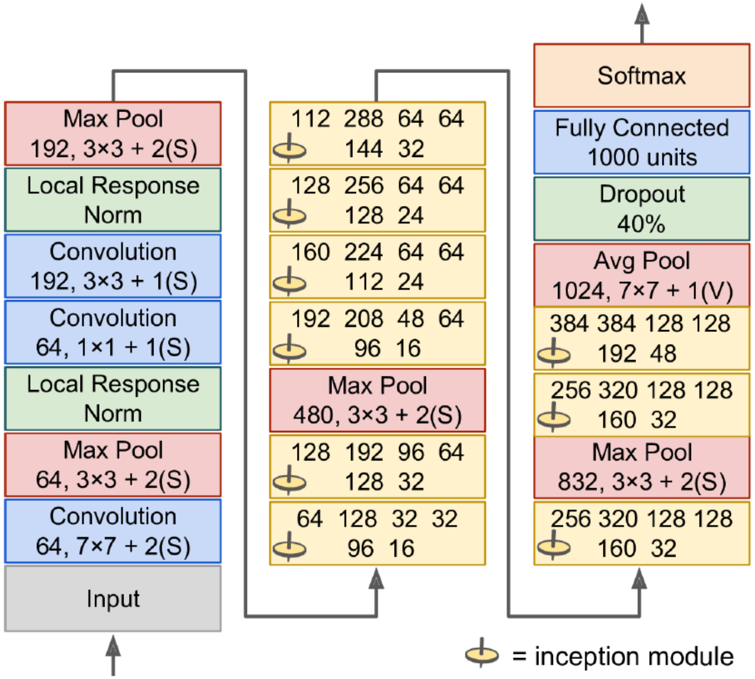
<Deep Learning, Andrew Ng>

Sliding windows detection

Implementation Problem : sliding windows detection takes long time to compute.

GoogLeNet

won the ILSVRC 2014 challenge with 93% accuracy.



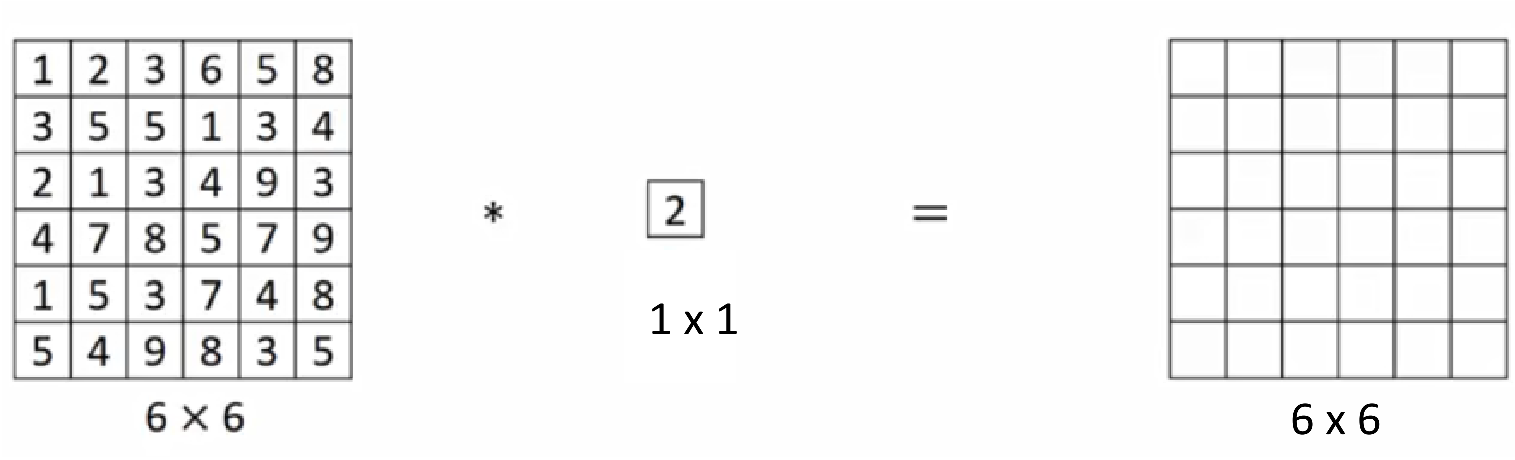
1 x 1 convolution is used.

<Hands-on ML, Aurelien Geron>

“Going Deeper with Convolutions,” C. Szegedy et al. (2015)

1x 1 convolutions

In the case of one channel



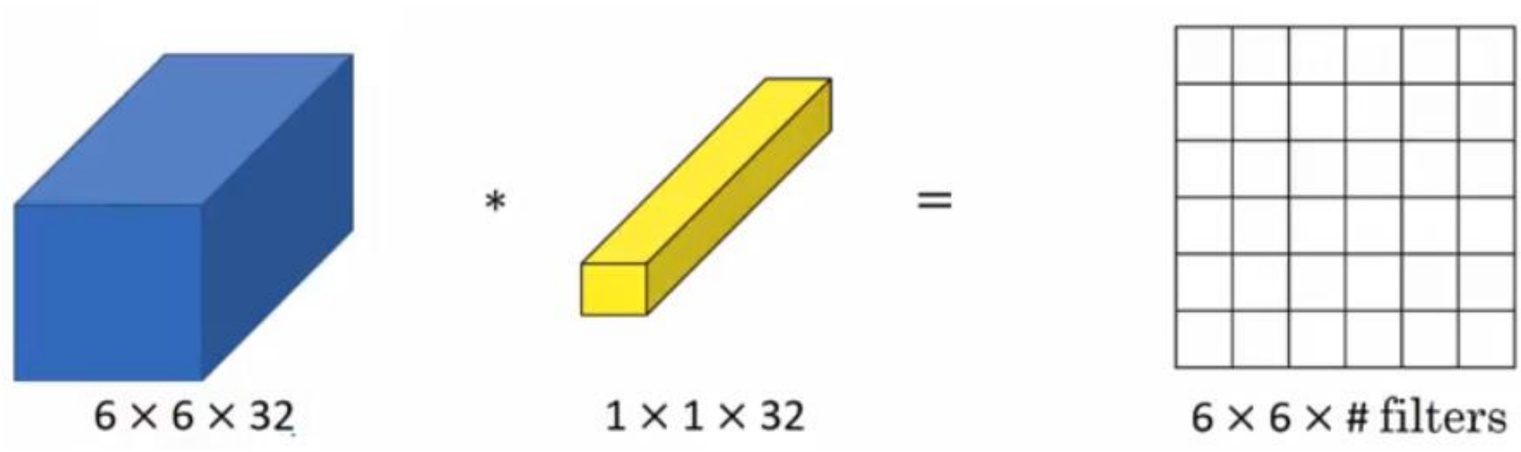
1	2	3	6	5	8
3	5	5	1	3	4
2	1	3	4	9	3
4	7	8	5	7	9
1	5	3	7	4	8
5	4	9	8	3	5

<Deep Learning, Andrew Ng>

Lin et al., 2013. Network in network

1 x 1 convolutions

In the case of multiple channels



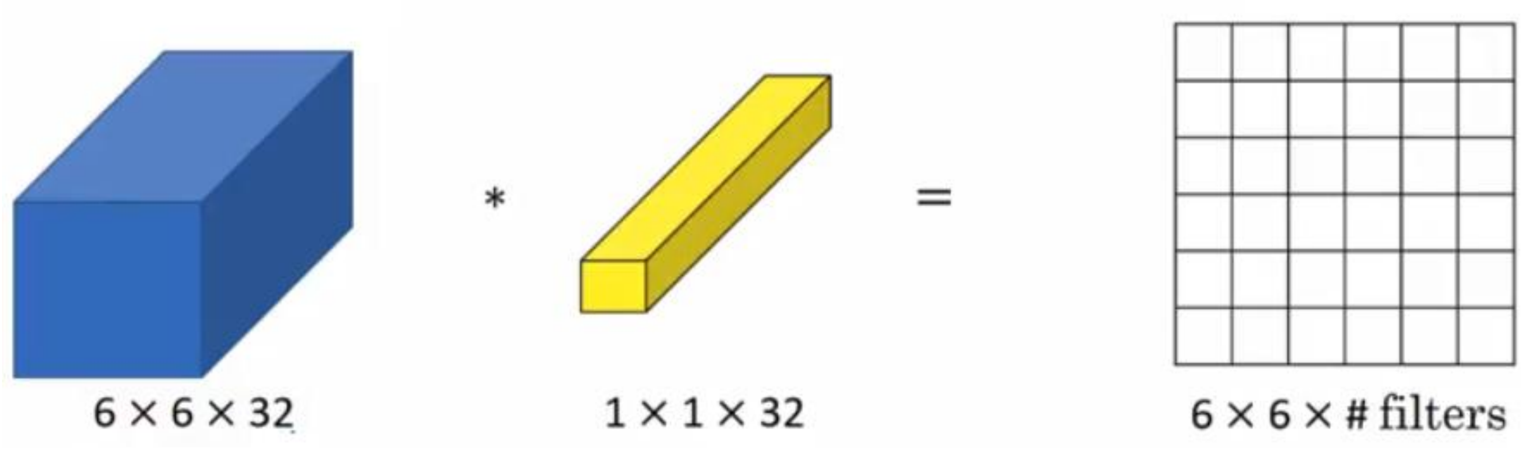
Convolution : sum of individual pixel over all channels

<Deep Learning, Andrew Ng>

Lin et al., 2013. Network in network

1 x 1 convolutions

Network in network



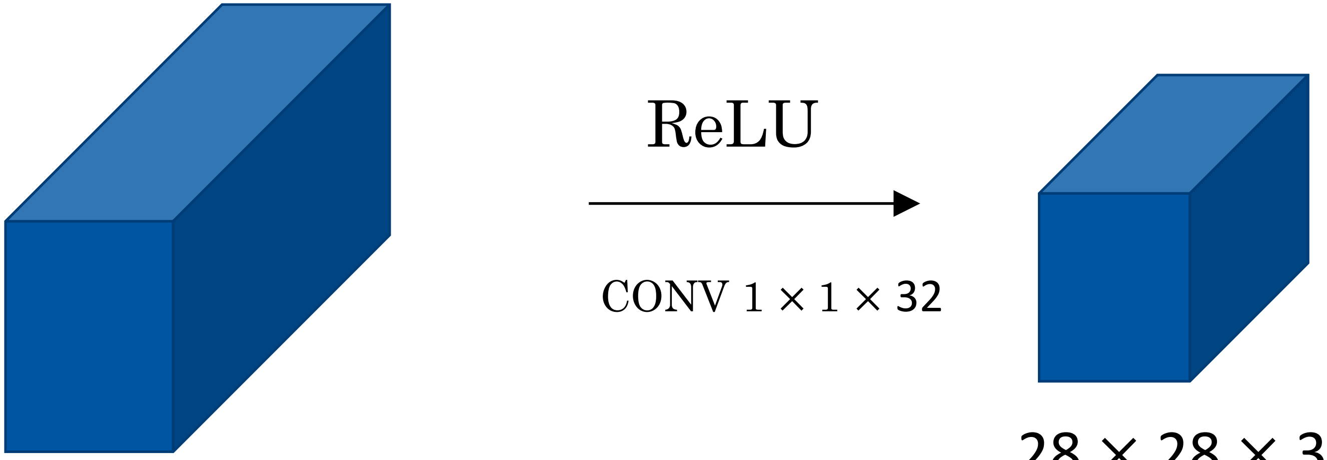
1 x 1 x 32 x n_c (# filters)

Convolution
Activation

<Deep Learning, Andrew Ng>

Lin et al., 2013. Network in network

1×1 convolutions



ReLU

CONV 1 × 1 × 32

28 × 28 × 192

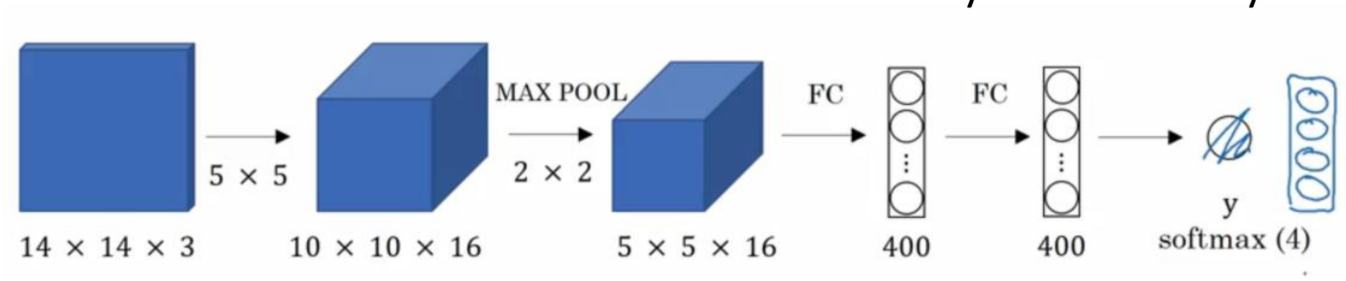
28 × 28 × 32

<Deep Learning, Andrew Ng>

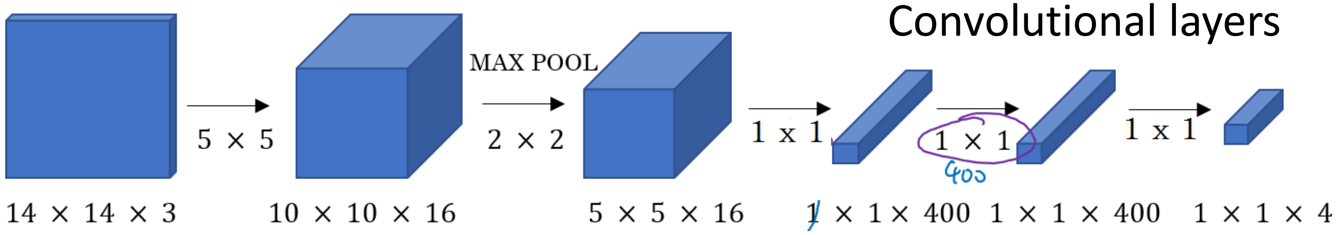
Lin et al., 2013. Network in network

Turning Fully connected layer into convolutional layers

Fully connected layers

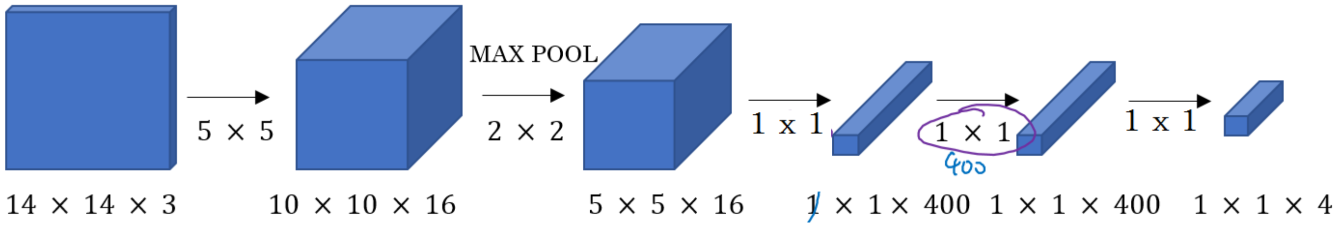


Convolutional layers

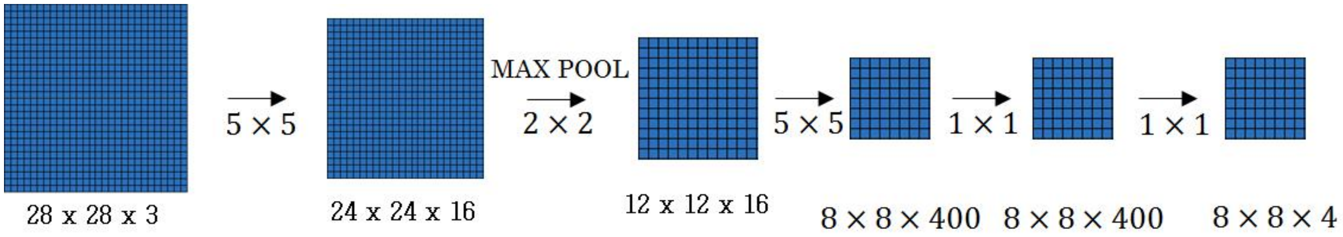


<Deep Learning, Andrew Ng>

Convolution implementation of sliding windows



Sliding windows : 14 x 14, stride: 2

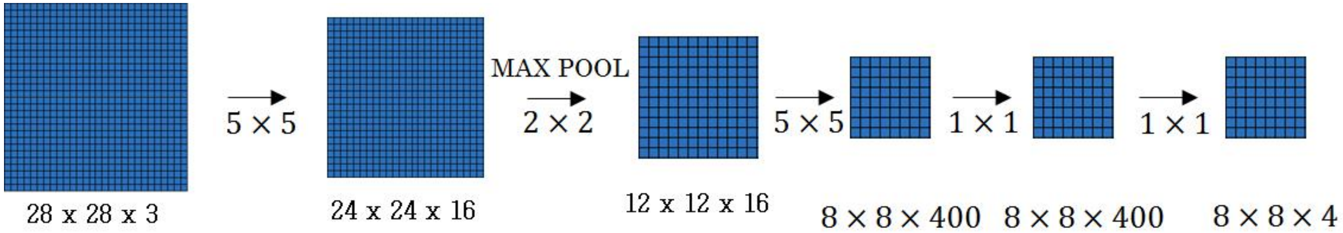


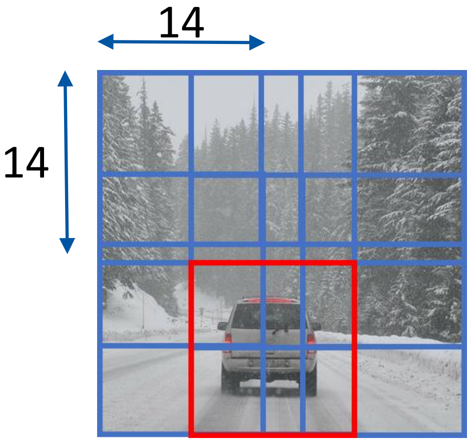
<Deep Learning, Andrew Ng>

Sermanet et al., 2014, OverFeat: Integrated recognition, localization and detection using convolutional networks

Convolution implementation of sliding windows

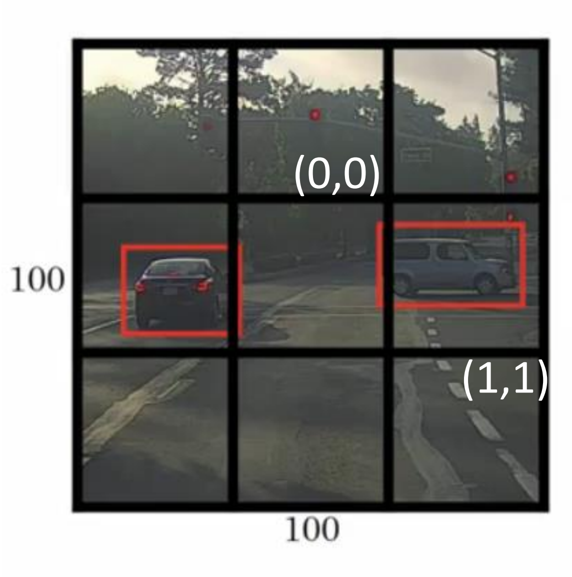
Sliding windows : 14 x 14, stride: 2





<Deep Learning, Andrew Ng>

Specify bounding boxes



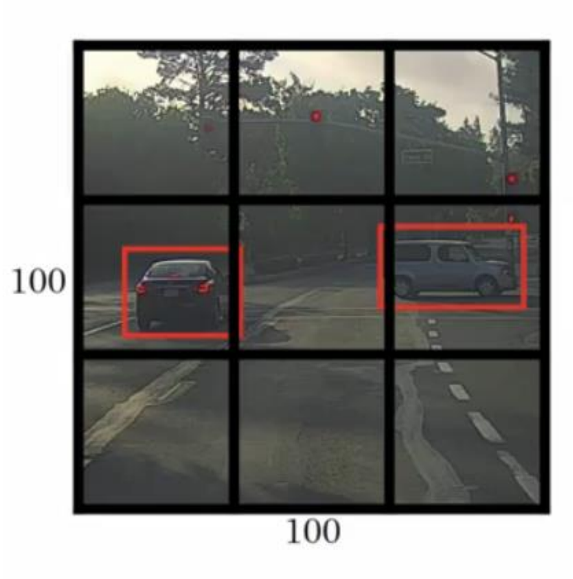
$$\begin{bmatrix} \hat{y} \\ b_x \\ b_y \\ b_h \\ b_w \end{bmatrix}$$

$0 < b_x, b_y < 1$
$0 < b_h, b_w$

<Deep Learning, Andrew Ng>

Redmon et al., 2015, You Only Look Once: Unified real-time object detection

YOLO (you only look once)



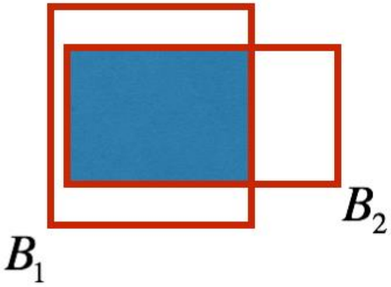
For each grid cell:

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

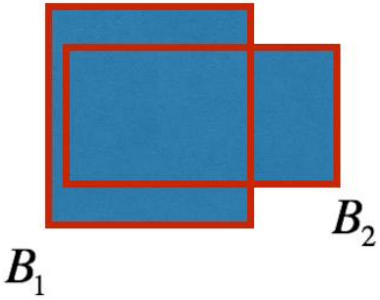
Redmon et al., 2015, You Only Look Once: Unified real-time object detection

Intersection over Union (IoU)

Intersection

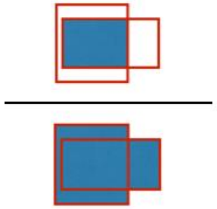


Union



Intersection over Union

$$IoU = \frac{B_1 \cap B_2}{B_1 \cup B_2} =$$



$=P_c$

“Correct” if IoU $\geq$ 0.5

More generally, IoU is a measure of the overlap between two bounding boxes.

<Deep Learning, Andrew Ng>

Non-max suppression

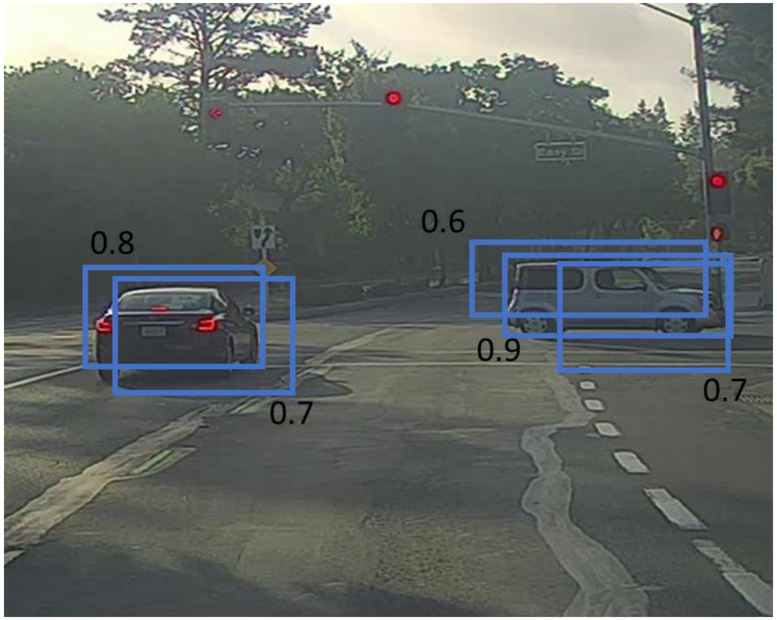
Each output prediction is:

$$\begin{bmatrix} p_c \\ b_x \\ b_y \\ b_h \\ b_w \end{bmatrix}$$

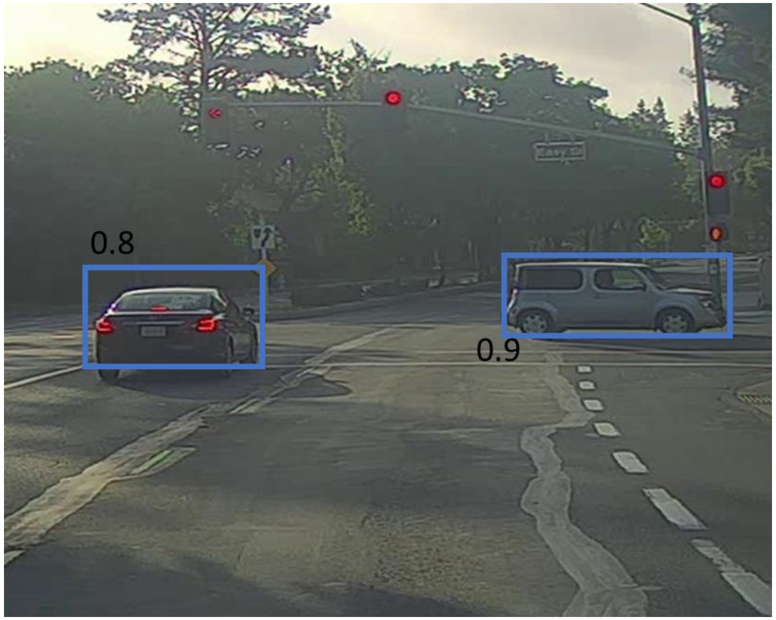
Discard all boxes with $p_c \leq 0.6$
Among remaining boxes,
Pick the box with the largest p_c.
Output that as a prediction.

Non-max suppression

Before non-max suppression

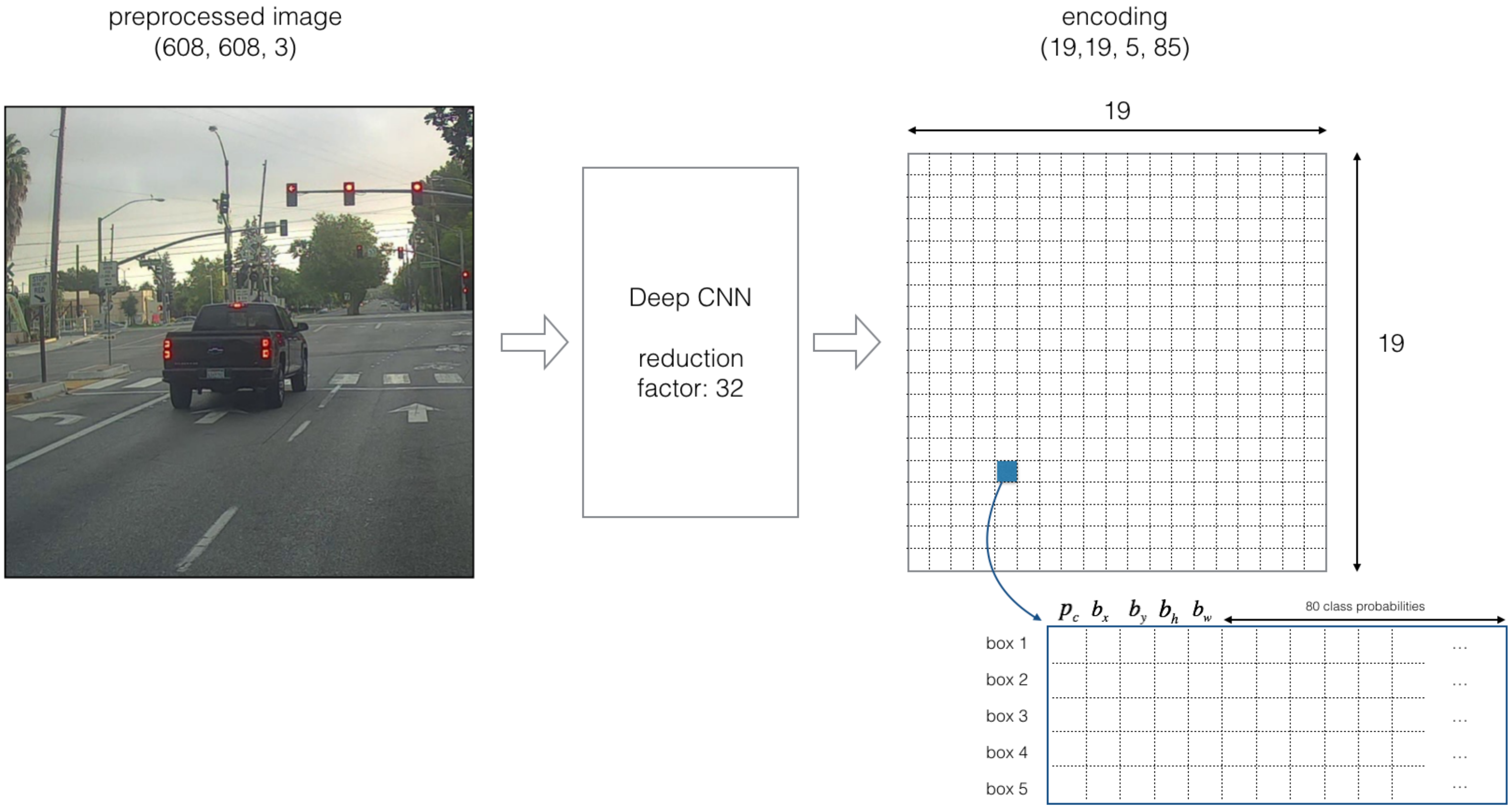


After non-max suppression



<Deep Learning, Andrew Ng>

Encoding architecture for YOLO



preprocessed image
(608, 608, 3)

Deep CNN

reduction
factor: 32

encoding
(19,19, 5, 85)

19

19

p_c b_x b_y b_h b_w

80 class probabilities

box 1 …

box 2 …

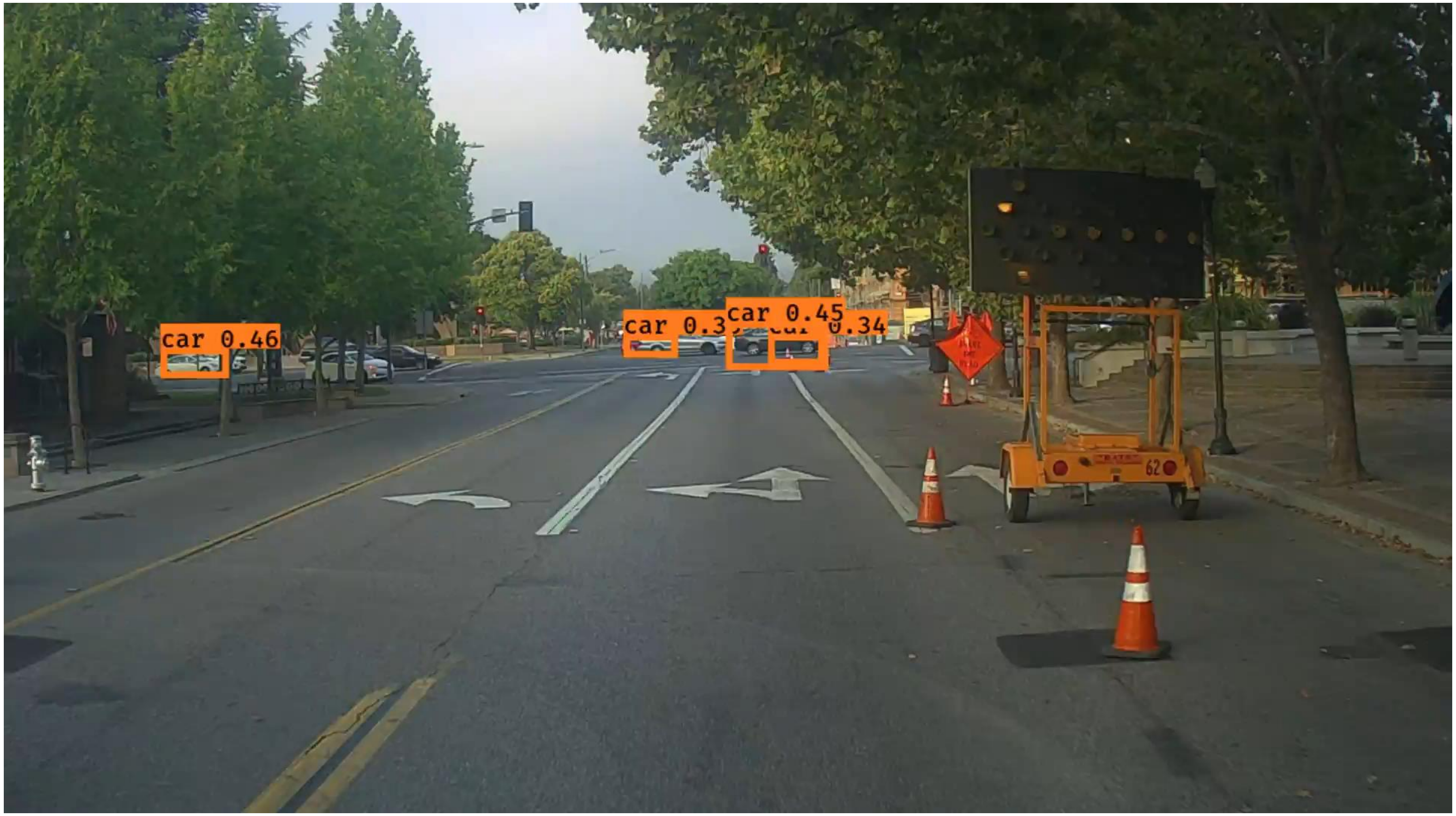
box 3 …

box 4 …

box 5 …

<Deep Learning, Andrew Ng>

Predictions of YOLO model <Deep Learning, Andrew Ng>



Pictures taken from a camera while driving around the Silicon Valley [drive.ai](https://www.drive.ai/)

Summary

Object detection

- Sliding windows
- 1 x 1 convolution
- Bounding box
- Intersection over union
- Non-max suppression