CMPE 258, Deep Learning

Face recognition

April 10, 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

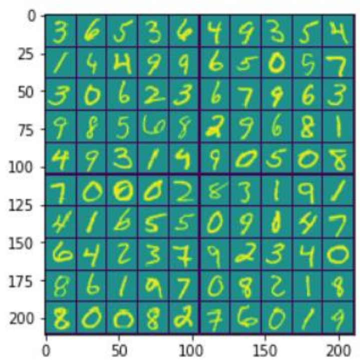
- Justification for the project
- Background: any relevant previous work
- How to collect data set
- Which algorithms / platform will be used
- What is the role for each team member

Last lesson

Object detection

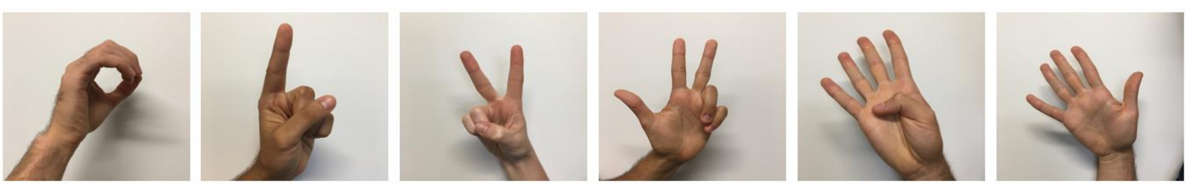
- Sliding windows
- 1 x 1 convolution
- Bounding box
- Intersection over union
- Non-max suppression

Image classification



Images for Hand written digits

Signs images



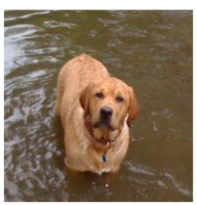
y = 0 y = 1 y = 2 y = 3 y = 4 y = 5

Coursera (Deep Learning specialization)

Image classification

Input image

Output



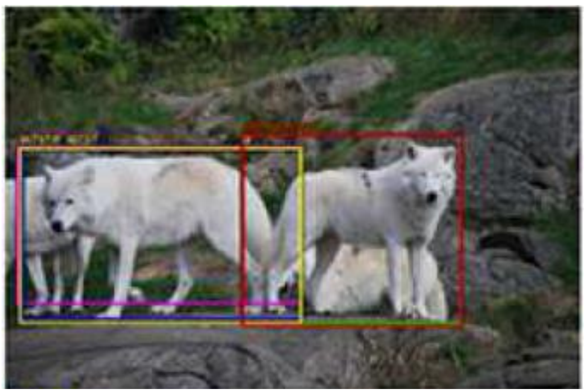
Deep Neural Network

Convolution Neural Network

softmax

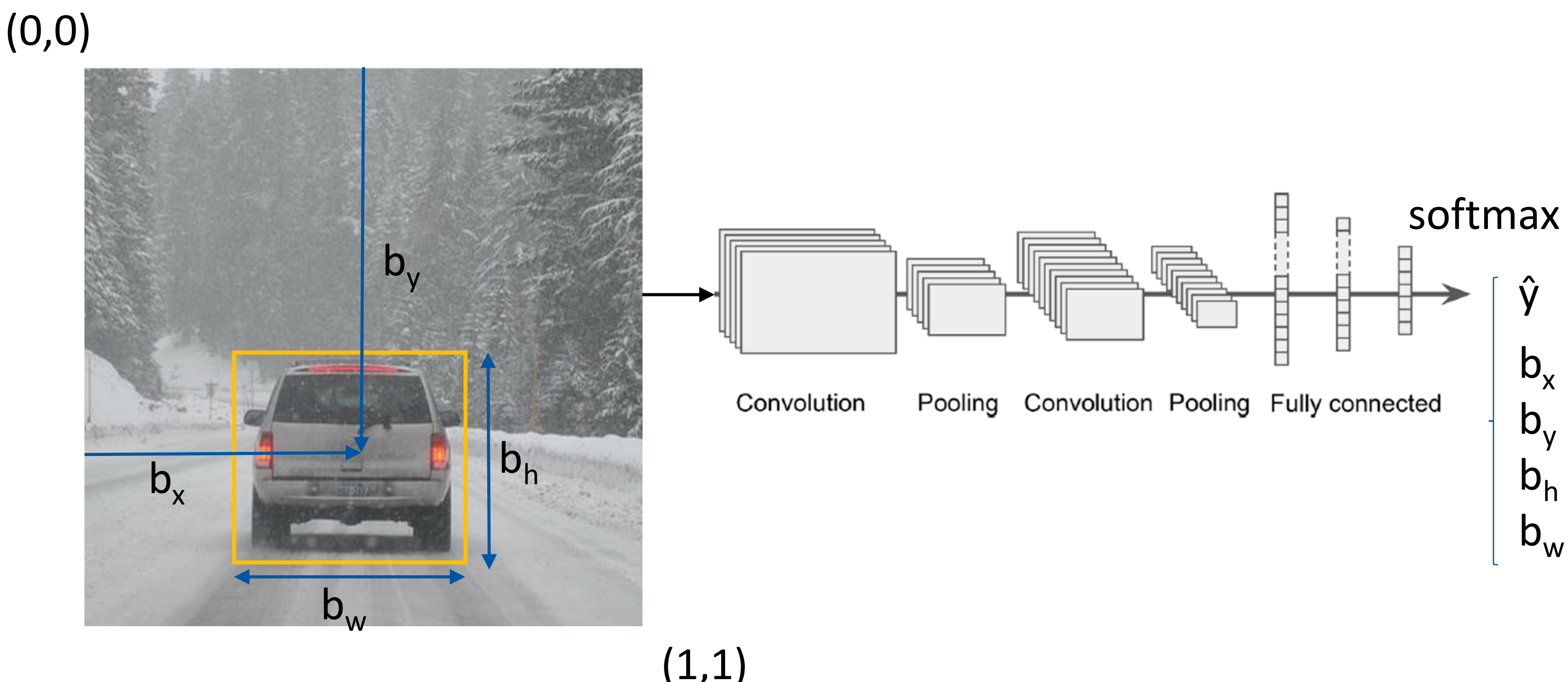
Prediction:
Cat or Dog?

Localization



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

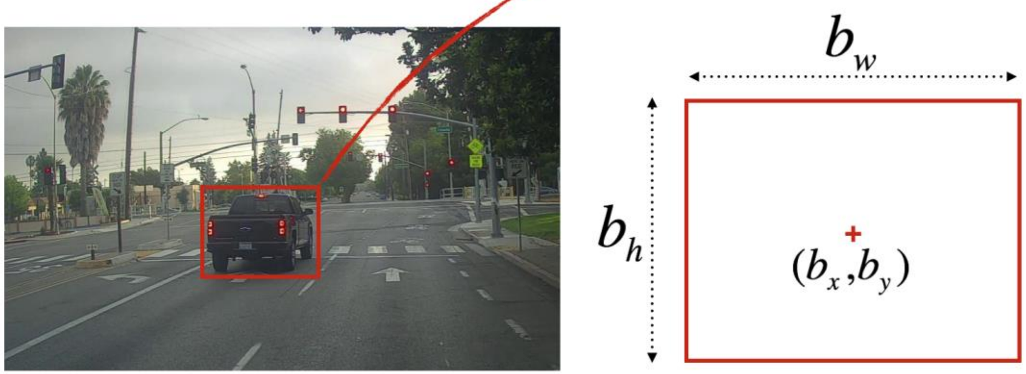
Classification with localization



<deep learning, Andrew Ng>

Example of bounding box

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



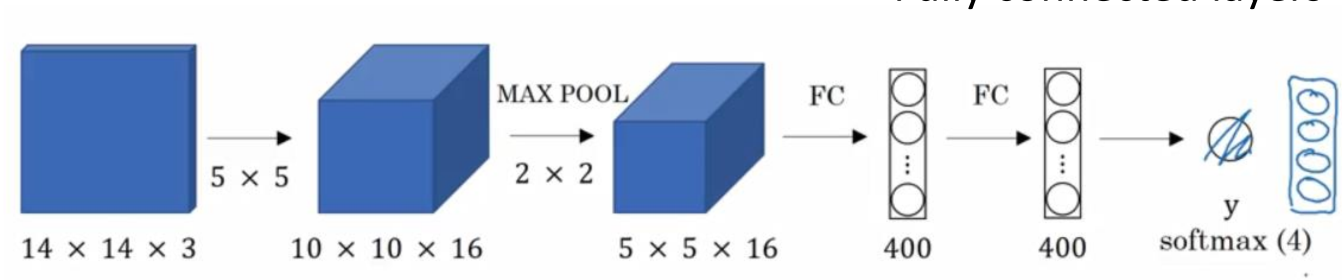
$p_c = 1$: confidence of an object being present in the bounding box

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

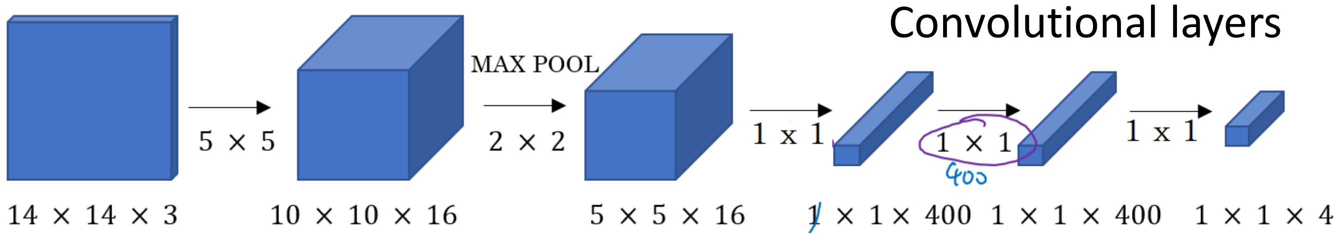
<deep learning, Andrew Ng>

Turning Fully connected layer into convolutional layers

Fully connected layers

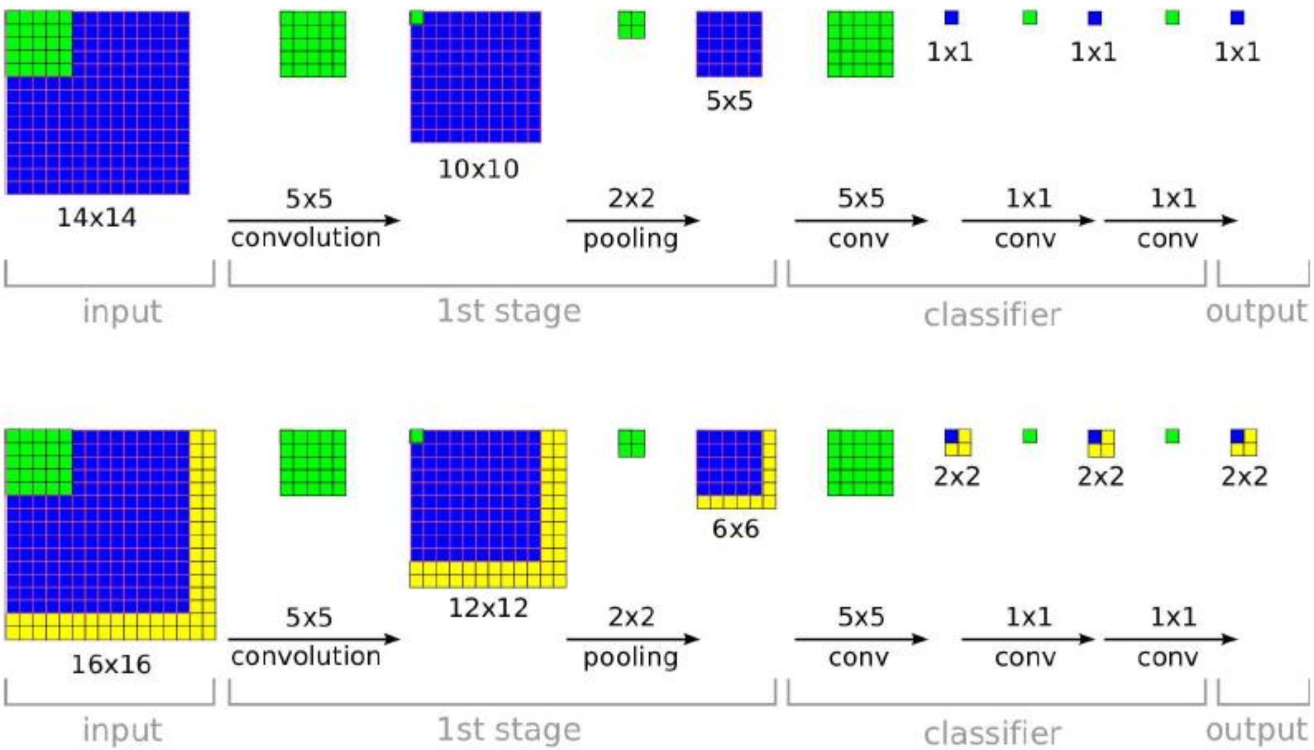


Convolutional layers



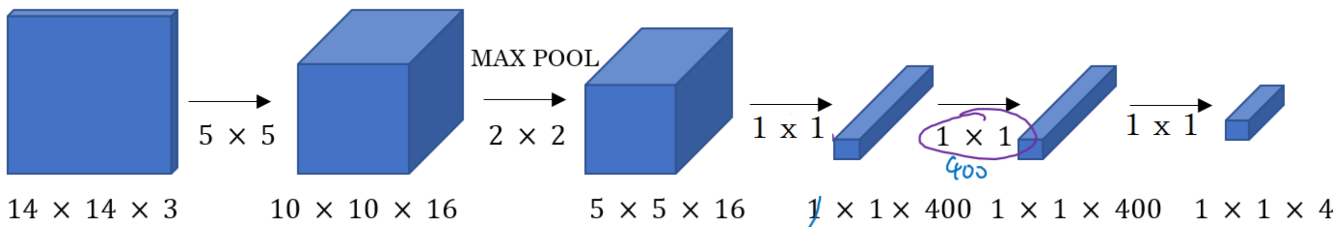
<Deep Learning, Andrew Ng>

Efficiency of ConvNets for detection

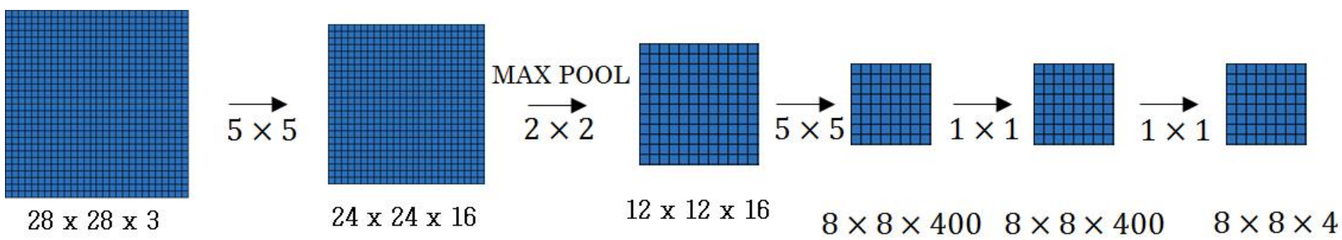


Sermanet et al., 2014, OverFeat: Integrated recognition, localization and detection using convolutional networks

Convolution implementation of sliding windows



Sliding windows : 14 x 14, stride: 2

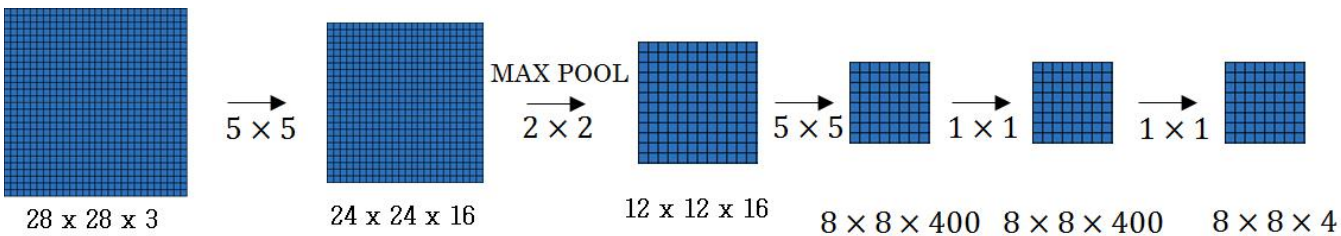


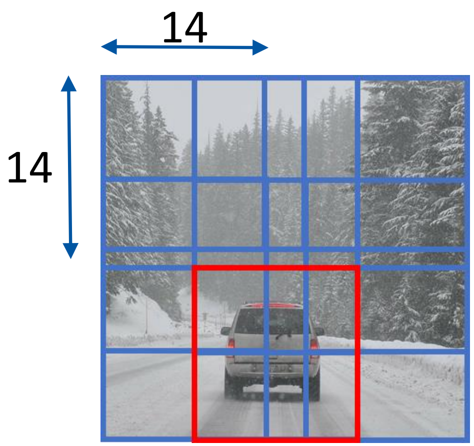
<Deep Learning, Andrew Ng>

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

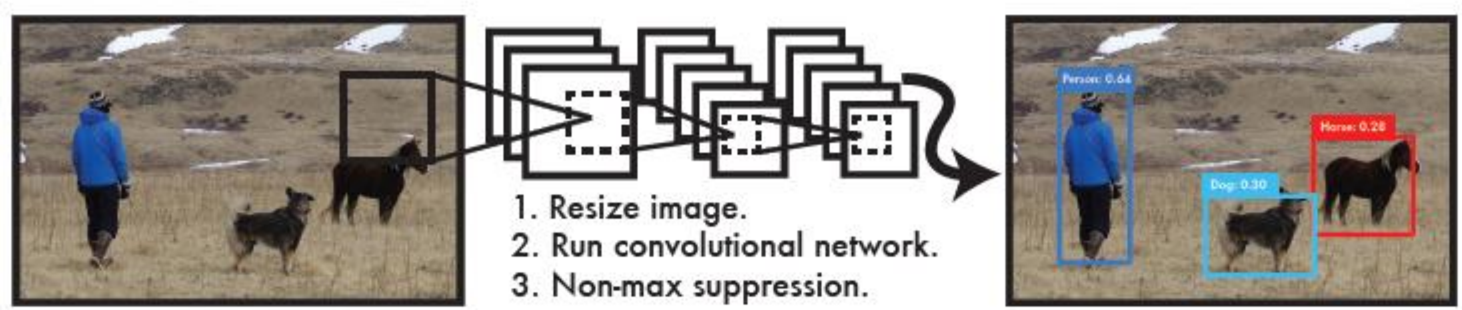




<Deep Learning, Andrew Ng>

YOLO detection system

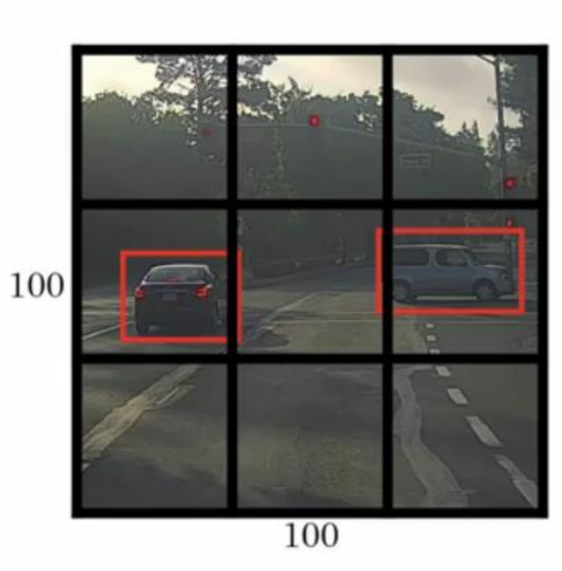
you only look once



(1) Resizes the input image
(2) runs a single convolutional network on the image
(3) thresholds the resulting detections by the model's confidence.

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

Divide image into S x S grid



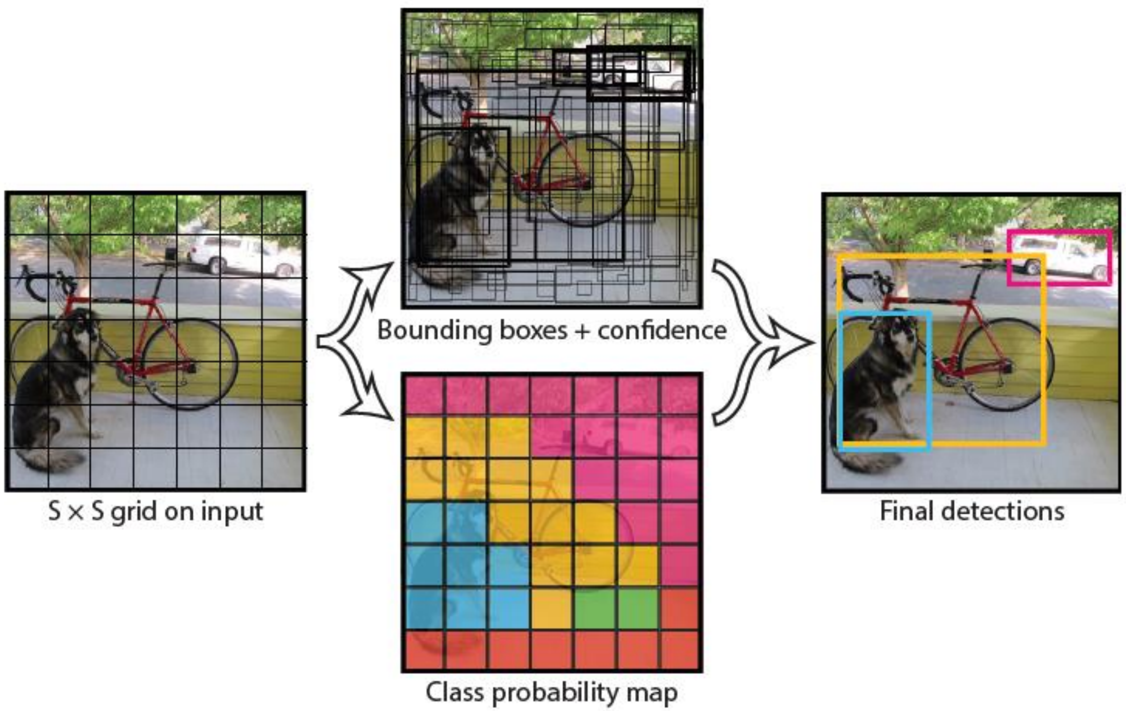
For each grid cell:

- Bounding box
- Confidence for those boxes,
- C class probabilities

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

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

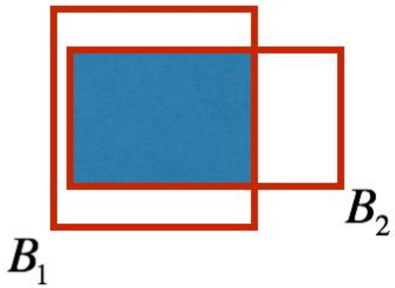
YOLO model



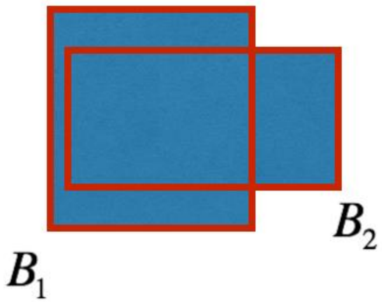
Redmon et al., 2015, You Only Look Once: Unified real-time object detection

Intersection over Union (IoU)

Intersection

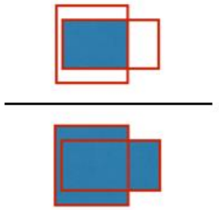


Union



Intersection over Union

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



$= P_c$

"Correct" if IoU ≥ 0.5

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

<Deep Learning, Andrew Ng>

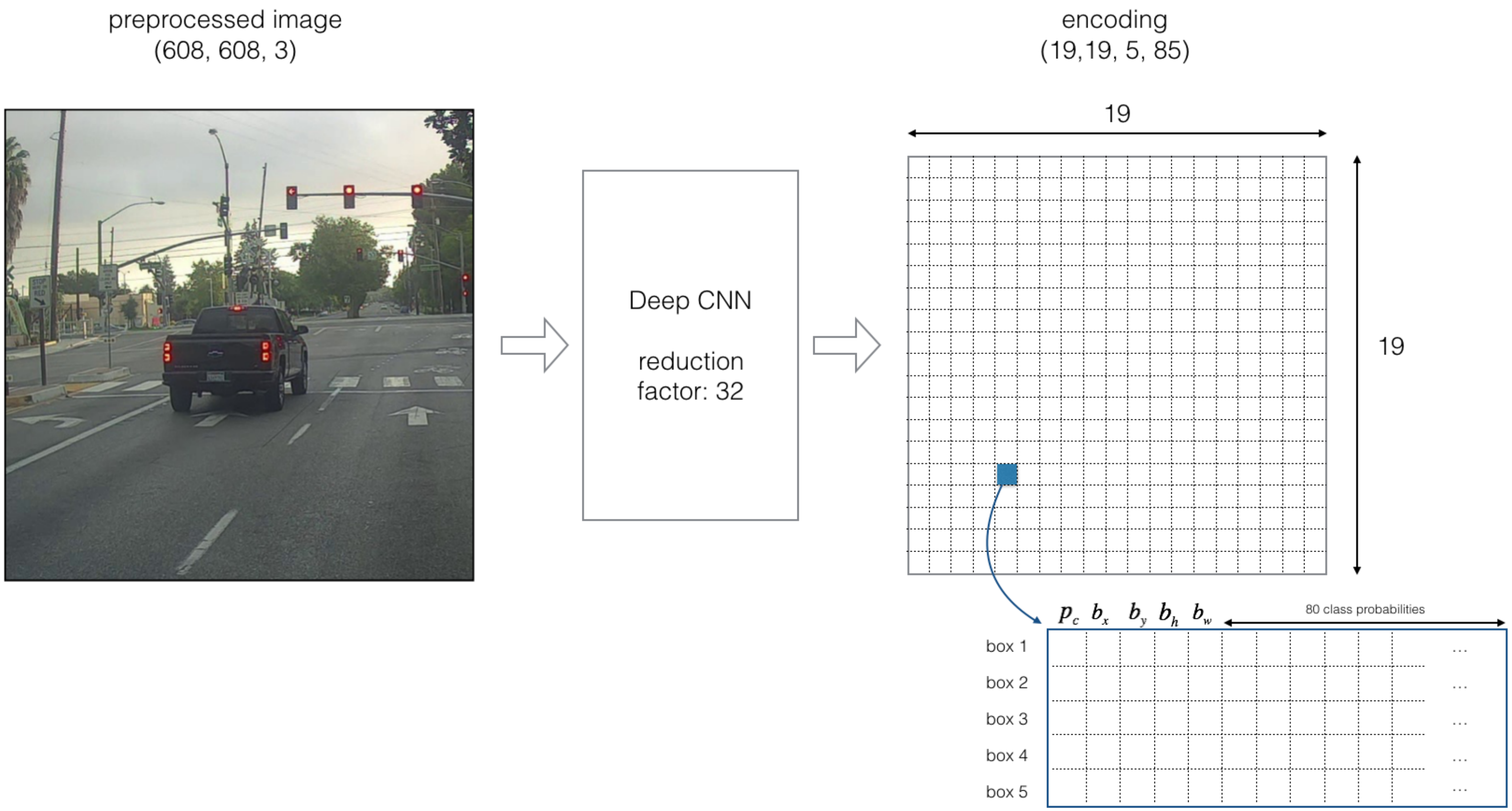
Non-max suppression

Each output prediction is:

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

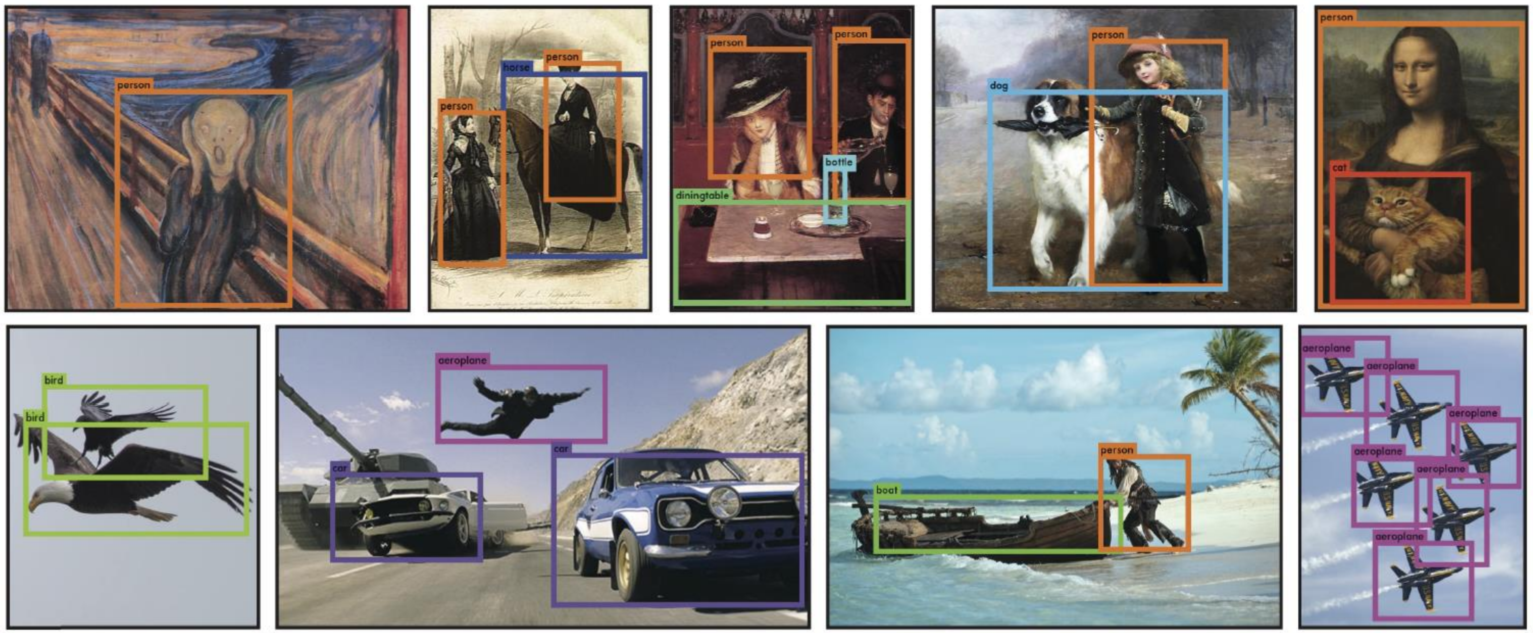
Discard all boxes with $p_c \leq 0.6$
Among remaining boxes,
Pick the box with the largest p_c.
Output that as a prediction.

Encoding architecture for YOLO



<Deep Learning, Andrew Ng>

Results for YOLO



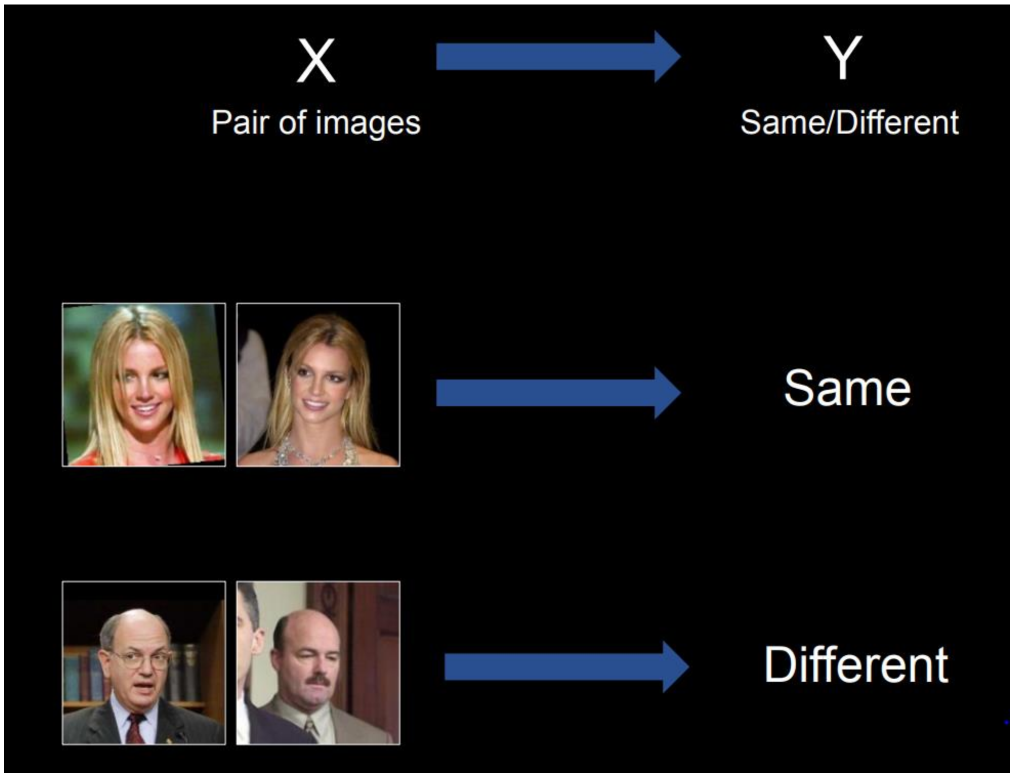
Redmon et al., 2015, You Only Look Once: Unified real-time object detection

Learning to recognize faces

Are these same person?

GTC 2015, Andrew Ng

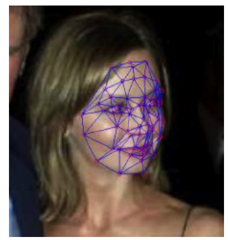
Learning to recognize faces



GTC 2015, Andrew Ng

Face recognition

Input image :



- Has a database of K persons
- Get an input image
- Output identity (name or id) if the image is any of the K persons (or “not recognized”)
- OneShotLearning

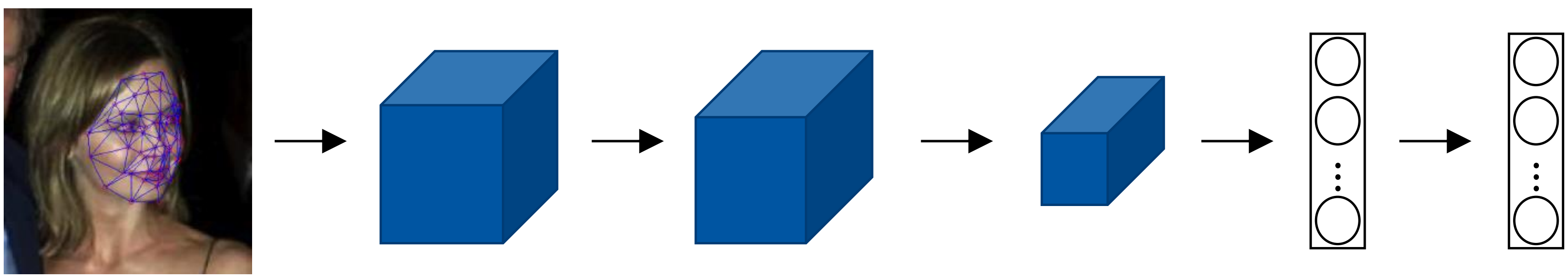
<Deep Learning, Andrew Ng>

Distance between images

- d(img1,img2) = degree of difference between images
- Image similarity
  - d(img1,img2) $\leq \tau$ → same
  - d(img1,img2) $> \tau$ → different

<Deep Learning, Andrew Ng>

Deep face feature vector



If $x^{(i)}, x^{(j)}$ are same person, $d\left(f(x^{(i)}) - f(x^{(j)})\right)$ is small.

If $x^{(i)}, x^{(j)}$ are different persons, $d\left(f(x^{(i)}) - f(x^{(j)})\right)$ is large.

Taigman et. al., 2014. DeepFace closing the gap to human level performance

<Deep Learning, Andrew Ng>

Distance between two feature vectors

- χ^2 distance $= \frac{\left(f\left(x^{(i)}\right)-f\left(x^{(j)}\right)\right)^2}{f\left(x^{(i)}\right)+f\left(x^{(j)}\right)}$

- Siamese network: $d\left(x^{(i)} - x^{(j)}\right) = \left\|f\left(x^{(i)}\right) - f\left(x^{(j)}\right)\right\|_2^2$

Same as L2 norm

Taigman et. al., 2014. DeepFace closing the gap to human level performance

Triplet Loss

Anchor

Positive

Anchor

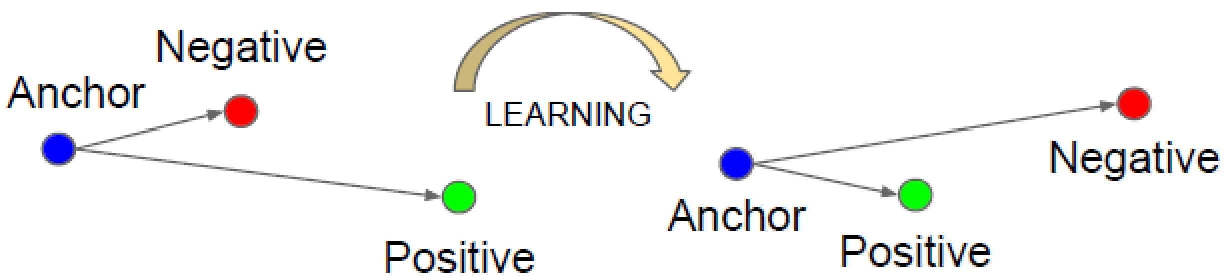
Negative

The triplet loss minimizes the distance between an anchor and a positive, and maximizes the distance between the anchor and a negative.

Schroff et al.,2015, FaceNet: A unified embedding for face recognition and clustering

<Deep Learning, Andrew Ng>

Triplet loss



The triplet loss minimizes the distance between an anchor and a positive, and maximizes the distance between the anchor and a negative.

Schroff et al.,2015, FaceNet: A unified embedding for face recognition and clustering

Triplet Loss

Anchor

Positive

Anchor

Negative

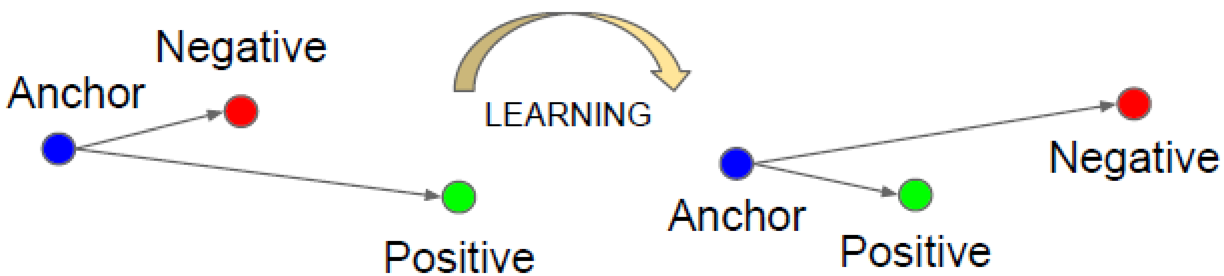
$$\boxed{\|f(x_i^a) - f(x_i^p)\|_2^2 + \alpha < \|f(x_i^a) - f(x_i^n)\|_2^2}$$

α: margin

Schroff et al.,2015, FaceNet: A unified embedding for face recognition and clustering

<Deep Learning, Andrew Ng>

Triplet Loss



$$\boxed{\|f(x_i^a) - f(x_i^p)\|_2^2 + \alpha < \|f(x_i^a) - f(x_i^n)\|_2^2}$$

Schroff et al.,2015, FaceNet: A unified embedding for face recognition and clustering

Triplet Loss

Loss is minimized:

$$\boldsymbol{L} = \sum_{i}^{N} \left[\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right]$$

Schroff et al.,2015, FaceNet: A unified embedding for face recognition and clustering

Triplet Selection

For fast convergence,

Given x_i^a,

Select hard positive x_i^p, to make $\left\|f(x_i^a) - f\left(x_i^p\right)\right\|_2^2$ maximum.

Select hard negative x_i^n, to make $\|f(x_i^a) - f(x_i^n)\|_2^2$ minimum.

$$\boxed{\|f(x_i^a) - f(x_i^p)\|_2^2 + \alpha < \|f(x_i^a) - f(x_i^n)\|_2^2}$$

Summary

Face recognition

- Image similarity
- One Shot Learning
- Deep face feature vector
- Distance between two feature vectors
- Triplet Loss
- Triplet selection