|  |  |
| --- | --- |
| http://sphotos.ak.fbcdn.net/photos-ak-snc1/v2219/119/28/51288393663/n51288393663_1327303_6197.jpg | old_logo_color |

Report on Week7

By

Ahmed Usama Khalifa

Submitted to

Dr. Omar Nasr

FACULTY OF ENGINEERING, CAIRO UNIVERSITY

GIZA, EGYPT

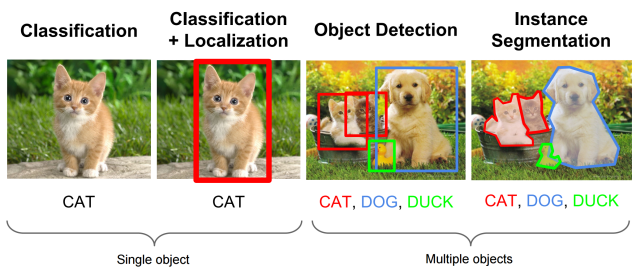
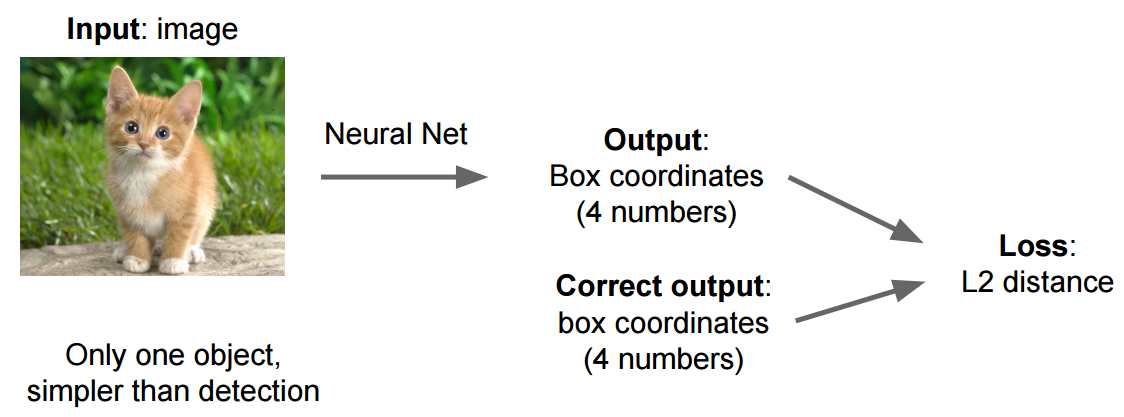
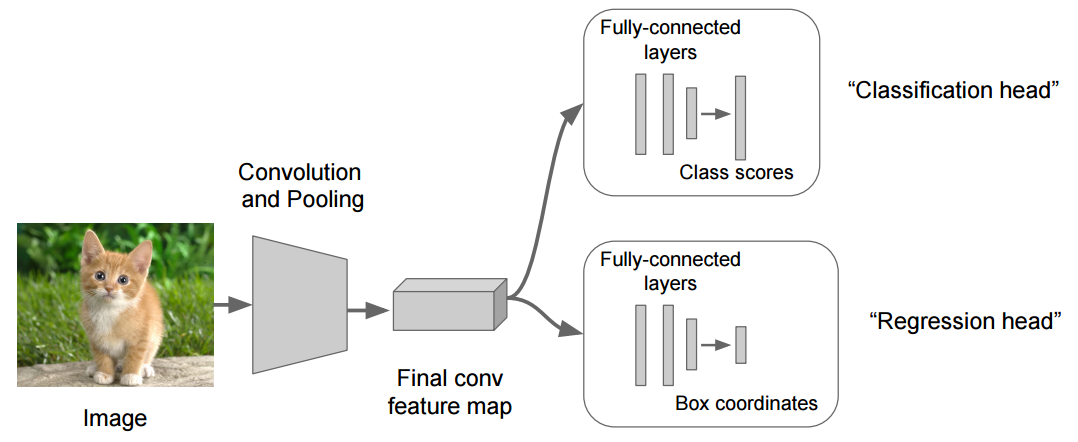
March – 2020

Outlines

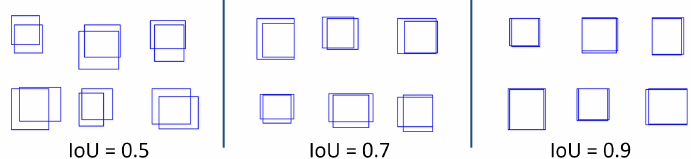
- Detection algorithms

- Art generation with Neural Style Transfer   
 - Face Recognition

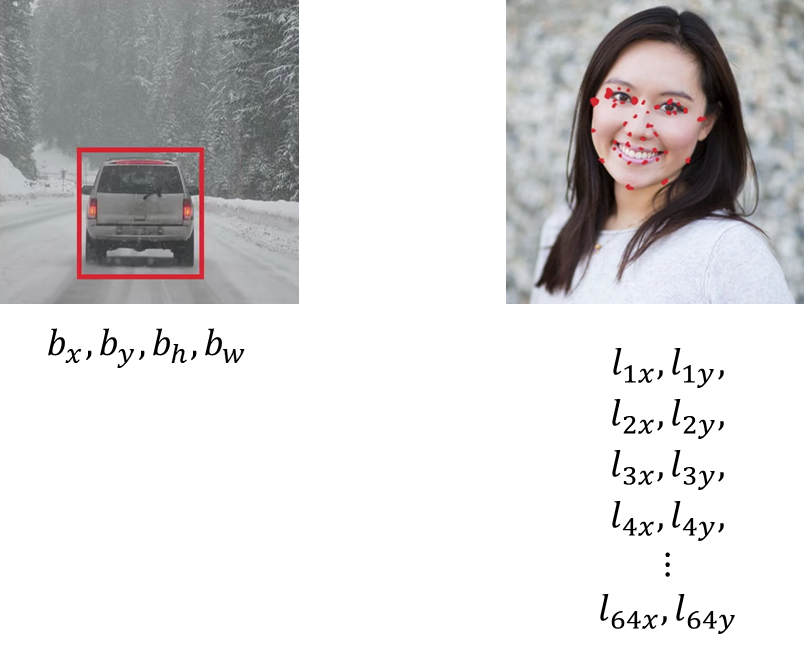
Key Concepts of Detection Algorithms:  
 - Object Localization   
- Landmark Detection   
- Object Detection   
- Convolutional Implementation of Sliding windows  
- Bounding Box Predication  
- Intersection Over Union   
- Non-max Suppression  
- Anchor Boxes

Object Localization  

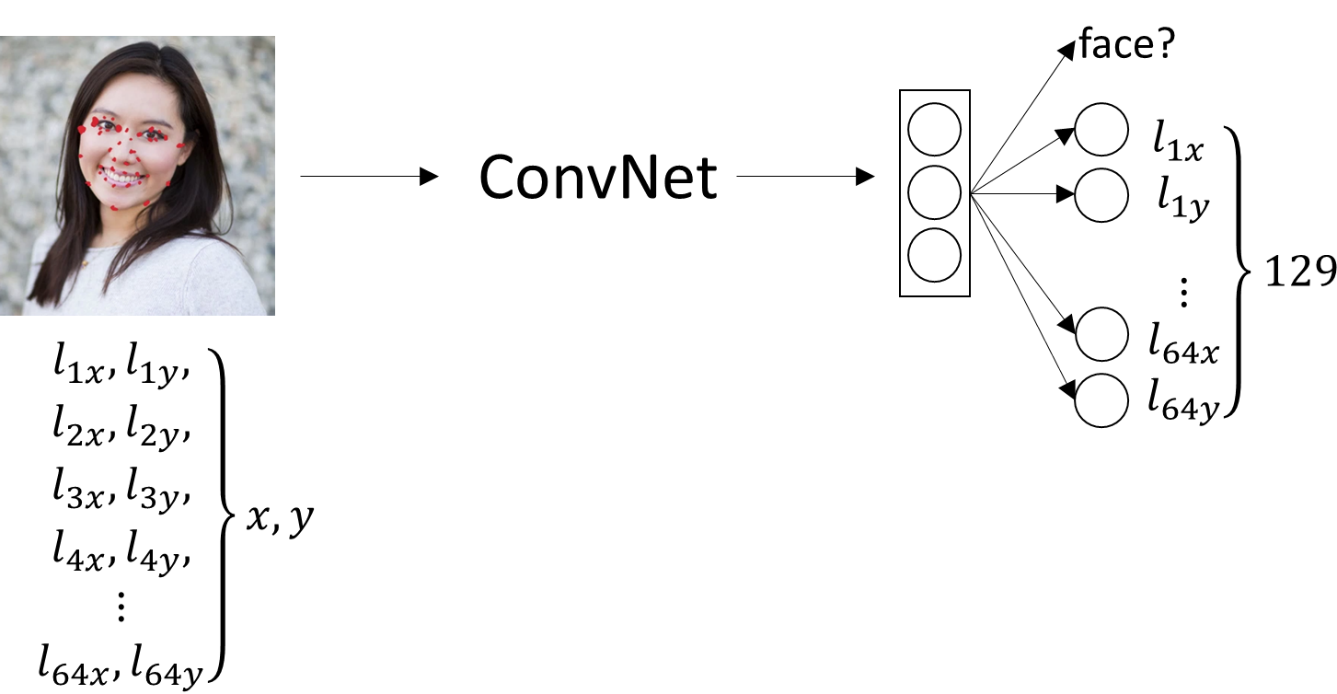
Basically we need to compare if the Intersect Over Union (ioU) between the prediction and the ground truth is bigger than some threshold (ex > 0.5)



Landmark detection



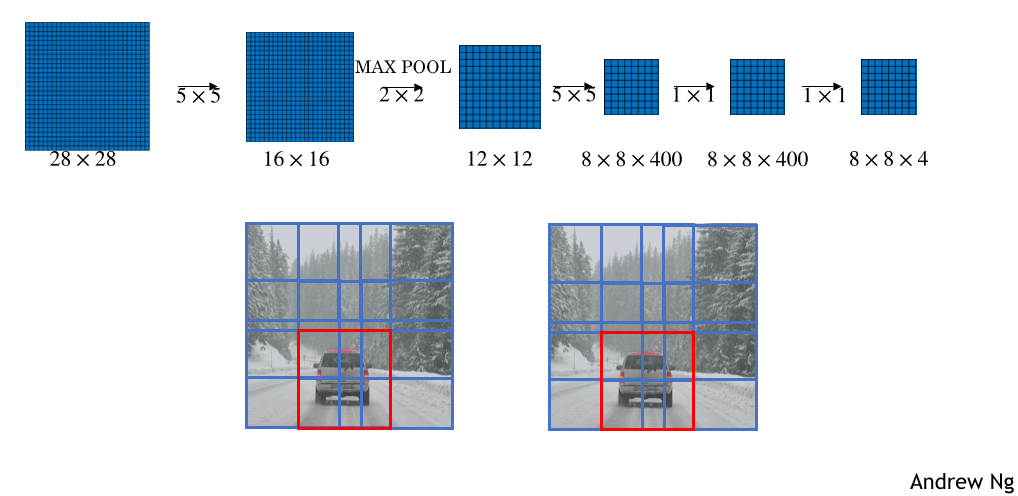
Every point has an x and y coordinate so we can just have a neural network with final layer that outputs two more numbers which we will call lx and ly to specify the coordinates of a point that is for example the person’s eye).

Now, what if we wanted the neural network to tell us all four corners of the eye, or both eyes. If we call the points the first, the second, the third and fourth point, going from left to right, then we can modify the neural network to output l1x, l1y, for the first point, and l2x, l1y for the second point and so on. The neural network can output the estimated position of all those four points of the person’s face. What if we don’t want just those four points? What if we want the output many points? For example what if we want to output different positions in the eye or shape of the mouth to see weather the person is smiling or not. We could define some number, for the sake of argument, let’s say 64 points or 64 landmarks on the face maybe even some points that helps us define the edge of the face, it defines the jawline. By selecting a number of landmarks and generating a label training set that contains all of these landmarks we can then have the neural network which tell us where are all the key positions or the key landmarks on a face.  
  


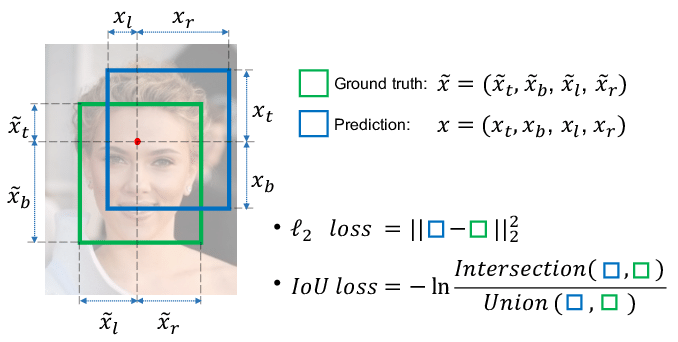
So, what we do is we have this image of person’s face as input, have it go through a convnet and have a convnet then have some set of features maybe have it output 0 or 1, like is there a face in this or not, and then have it also output l1x, l1y and so on down to l64x, l64y. We use l to stand for a landmark.

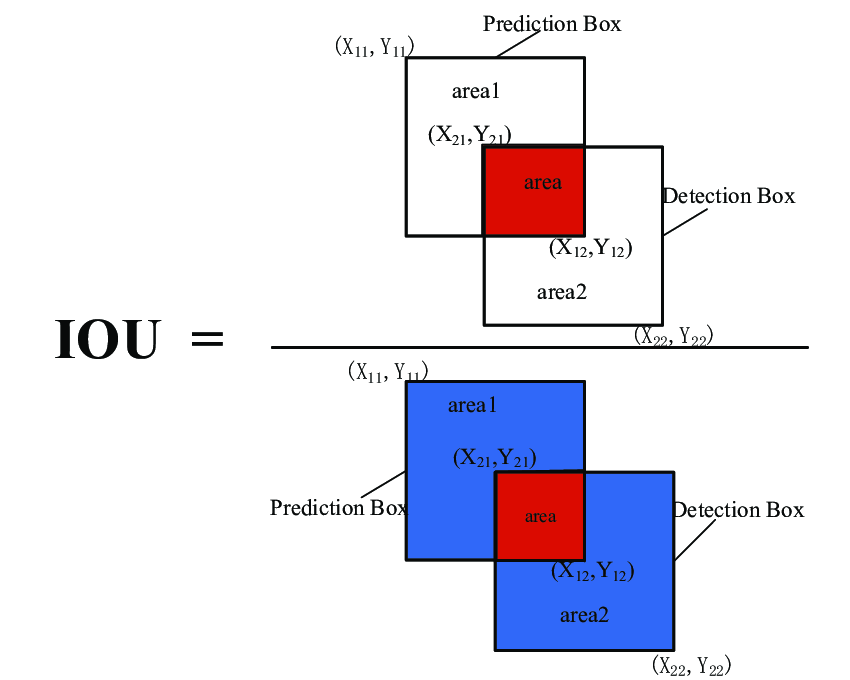
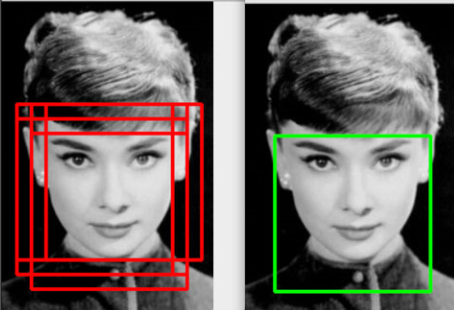
This example would have 129 output units, 1 is for where a face or not, and then if we have 64landmarks that is 64×2 which is equal to 128 plus 1  output units. This can tell us if there’s a face as well as where are all the key landmarks on the face. Of course in order to trade a network like this we will need a label training set. We have a set of images as well as labels Y,  where someone would have had to go through and laboriously annotate all of these landmarks.

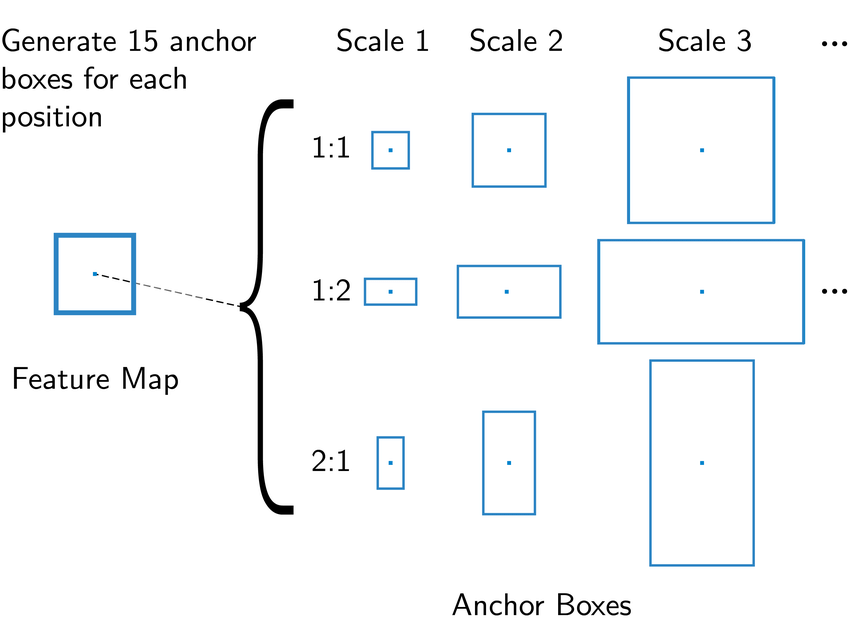
Convolutional Implementation of Sliding windows



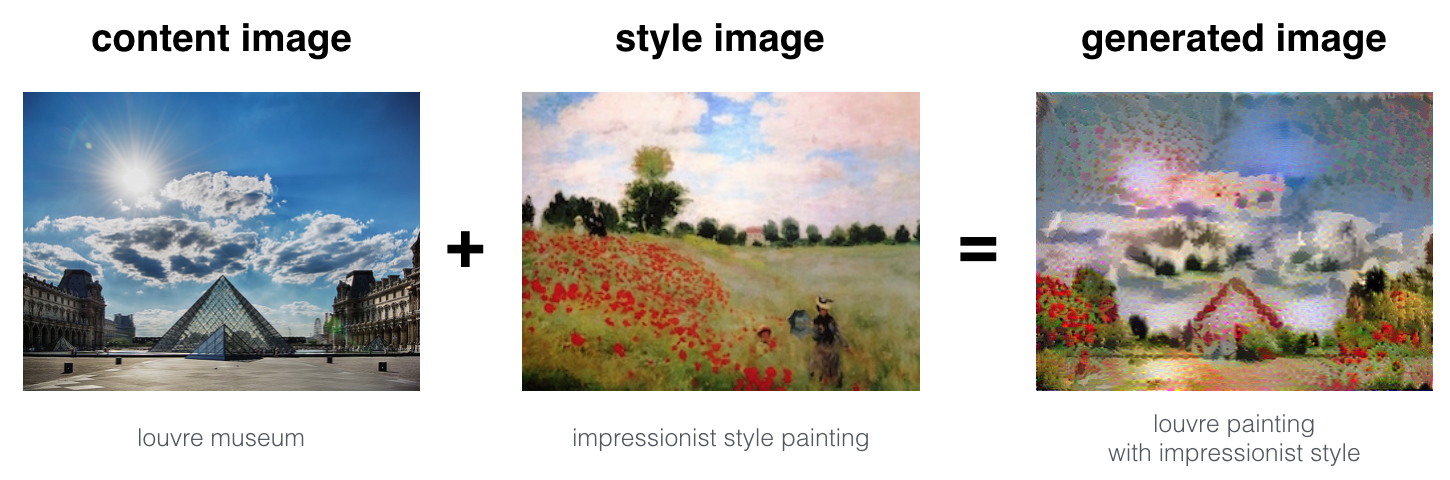
Bounding Box Predication



Intersection Over Union  
  
Non-max Suppression  


Anchor Boxes   


Art generation with Neural Style Transfer



Face Recognition  
