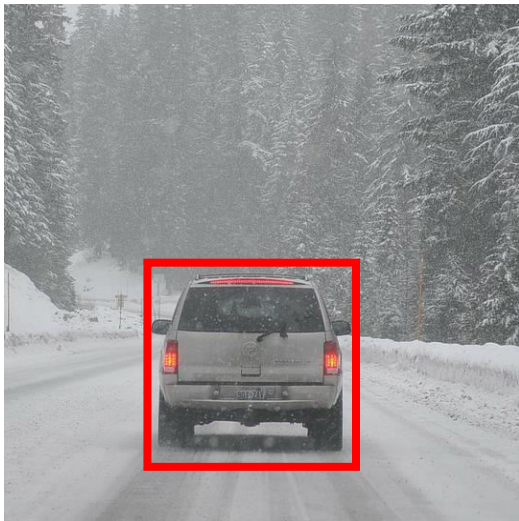deeplearning.ai

Object Detection

Landmark detection

Landmark detection

→ General case of b_x, b_y, b_H and b_w
→ output any (x,y) coordinates of Interest

ConvNet → $y = \begin{bmatrix} face? \\ l_1 x \\ l_1 y \\ l_2 x \\ l_2 y \\ \vdots \\ l_{64} x \\ l_{64} y \end{bmatrix}$ ⟹ y is a 129 × 1 dim vector

1 → whether there is a face or NOT

64 × 2 → points of Interest (landmarks)

b_x, b_y, b_h, b_w

(l_x, l_y)

Say you're interested in the position of the corner of the eye in the above Image (l_x, l_y), then instead of (b_x, b_y, b_w, b_H) in the previous example, you output this in the "y" vector

Say, we want the green points in the face above as well (maybe to extract facial expression/position of mouth etc)
then Instead of (l_x, l_y) → we could output a vector of these points $(l_1 x, l_1 y), (l_2 x, l_2 y) \ldots (l_n x, l_n y)$

collection of points of Interest

Pose detection
→ can have $l_1 x, l_1 y$ to $l_n x, l_n y$
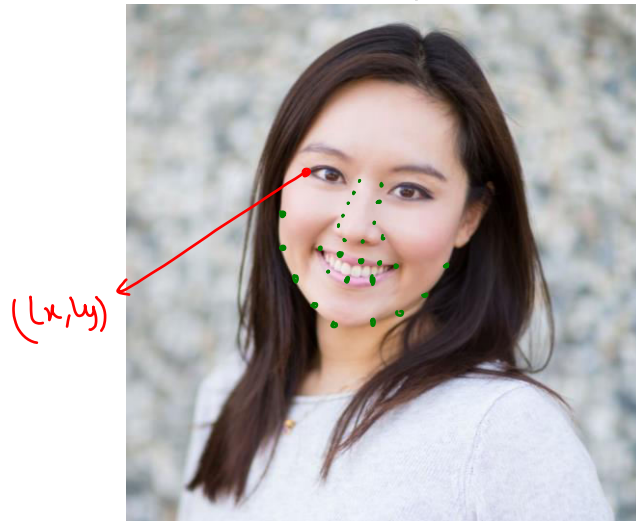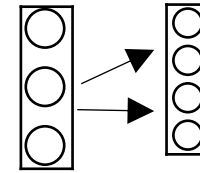
Final note
The labels have to be consistent in landmark detection
eg If $l_1 x, l_1 y$ = coordinates of bottom of chin of a face, it has to be true for all training Images
- $l_1 x, l_1 y$ → Image 1 → chin 1
$(l_1 x, l_1 y)$ → Image 2 → chin 2.

Andrew Ng