

Face detection neural network

Neural network is a network or circuit of neurons, or in a modern sense, an artificial neural network, composed of artificial neurons or nodes. Information flows through a neural network in two ways. When it's learning or operating normally, patterns of information are fed into the network via the input units, which trigger the layers of hidden units, and these in turn arrive at the output units.

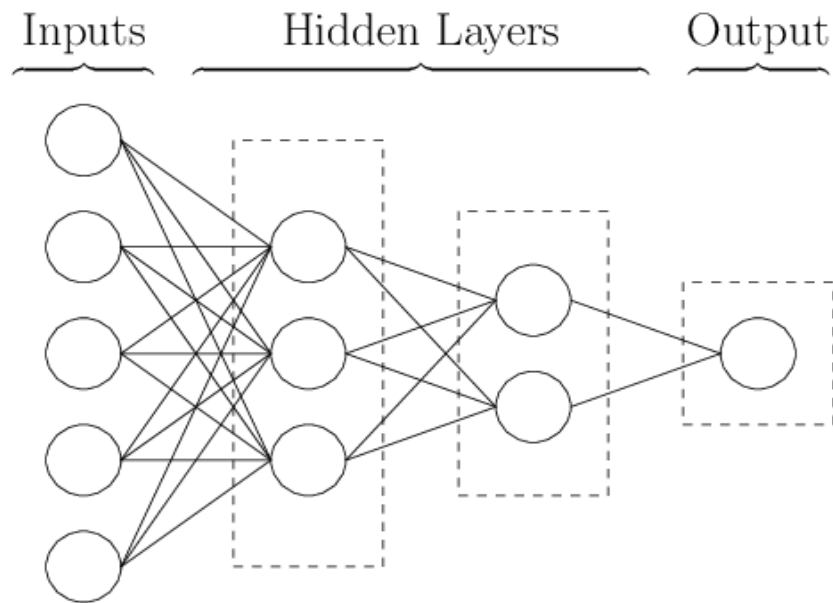


fig: neural network diagram

Steps of Convolutional Layers:

Stage 1:

Feed scaled images of the copies are sent to P-net. After gathering the output bounding boxes with low confidence are deleted but there are still a lot of bounding boxes left, and a lot of them overlap. Non-Maximum Suppression (NMS) is a method that reduces the number of bounding boxes. After converting bounding box coordinates to un-scaled image coordinates, these have to be reshaped to square.

Stage 2:

After padding out of bound boxes, feed scaled images are sent to R-Net. Next outputs are gathered. Bounding boxes with low confidence are then removed. NMS for all boxes convert bounding box coordinates to un-scaled image coordinates. Finally reshaping of bounding boxes to square are required.

Stage 3:

Again, after padding out of bound boxes, feed scaled images are sent to O-Net. After gathering O-Net output, bounding boxes with low confidence are removed. Convert bounding box and facial landmark coordinates to un-scaled image coordinates. NMS for all boxes.

Finally package of all coordinates and confidence levels are stored into a dictionary and return it.

MTCNN (Multi-task Cascaded Convolutional Neural Networks) is an algorithm consisting of 3 stages. P-net, R-net and O-net. For every image we pass in, the network creates an image pyramid. Which is basically multiple copies of that image in different sizes.

Summary of the whole MTCNN process is-

✧ **P-Net:**

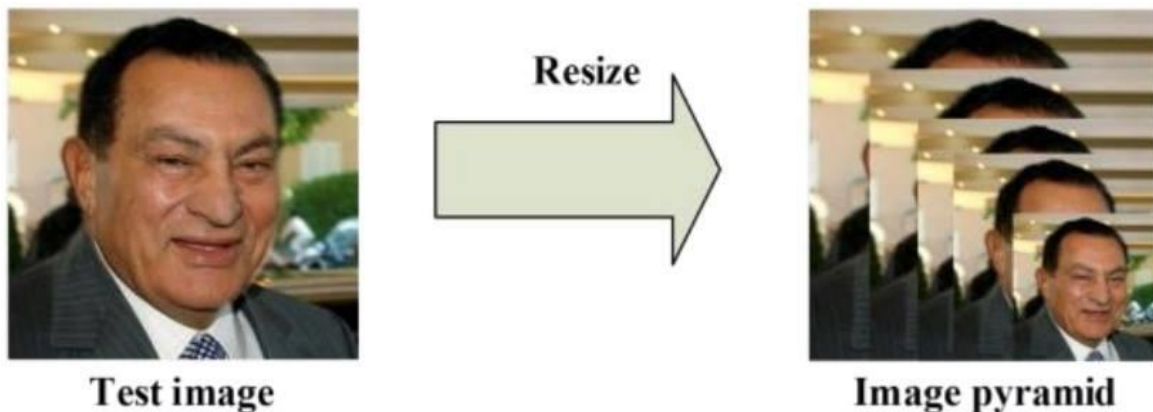


Image 1: Image Pyramid

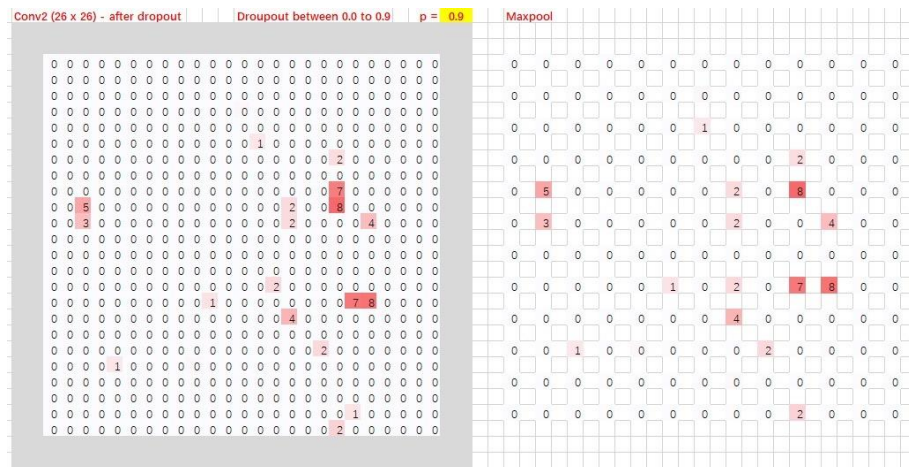


Image 2: Max-Pool



Image 3: P-Net

✧ R-Net:

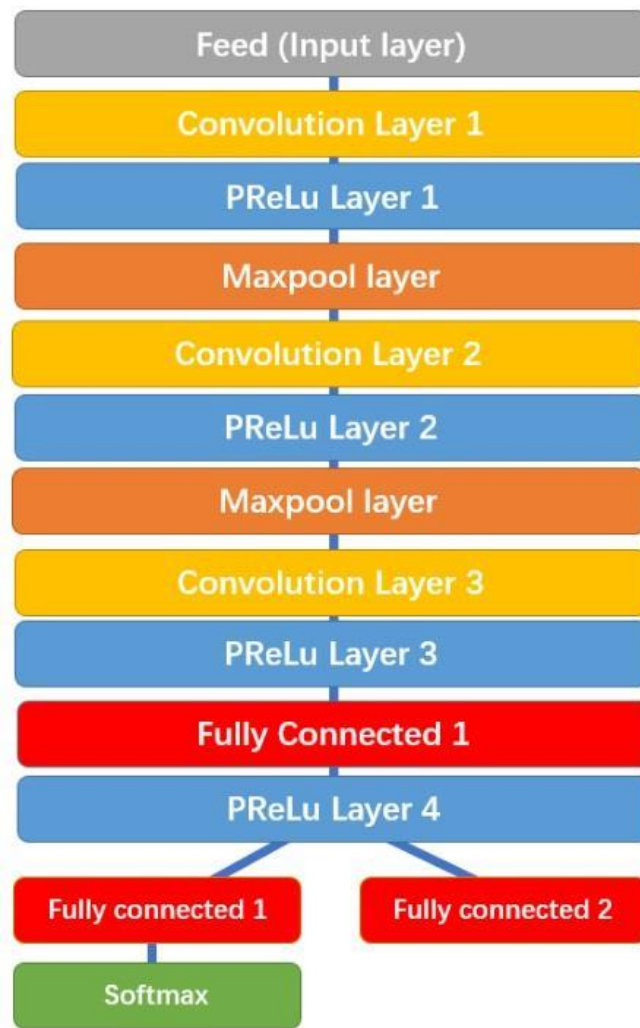


Image 4: R-Net

✧ O-Net

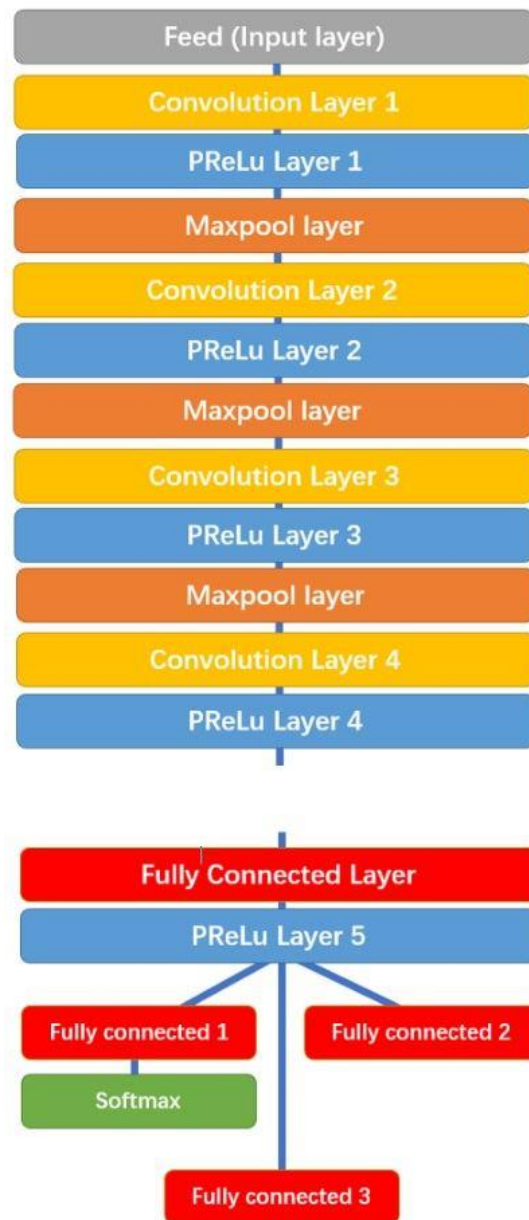


Image 5: O-Net

```
[{'box': [277, 90, 48, 63], 'confidence': 0.9985162615776062, 'keypoints': {'left_eye': (291, 117), 'right_eye': (314, 114), 'nose': (303, 131), 'mouth_left': (296, 143), 'mouth_right': (313, 141)}}]
```

References-

1. <https://towardsdatascience.com/how-does-a-face-detection-program-work-using-neural-networks-17896df8e6ff>
2. <https://towardsdatascience.com/face-detection-neural-network-structure-257b8f6f85d1>
3. <https://towardsdatascience.com/what-is-a-neural-network-6010edabde2b?fbclid=IwAR3Bnl8KEVaB4FTp9xWkikYMHdGH8fm6DQ69i9uvLjkk4HVObkf7sA2MVC8>