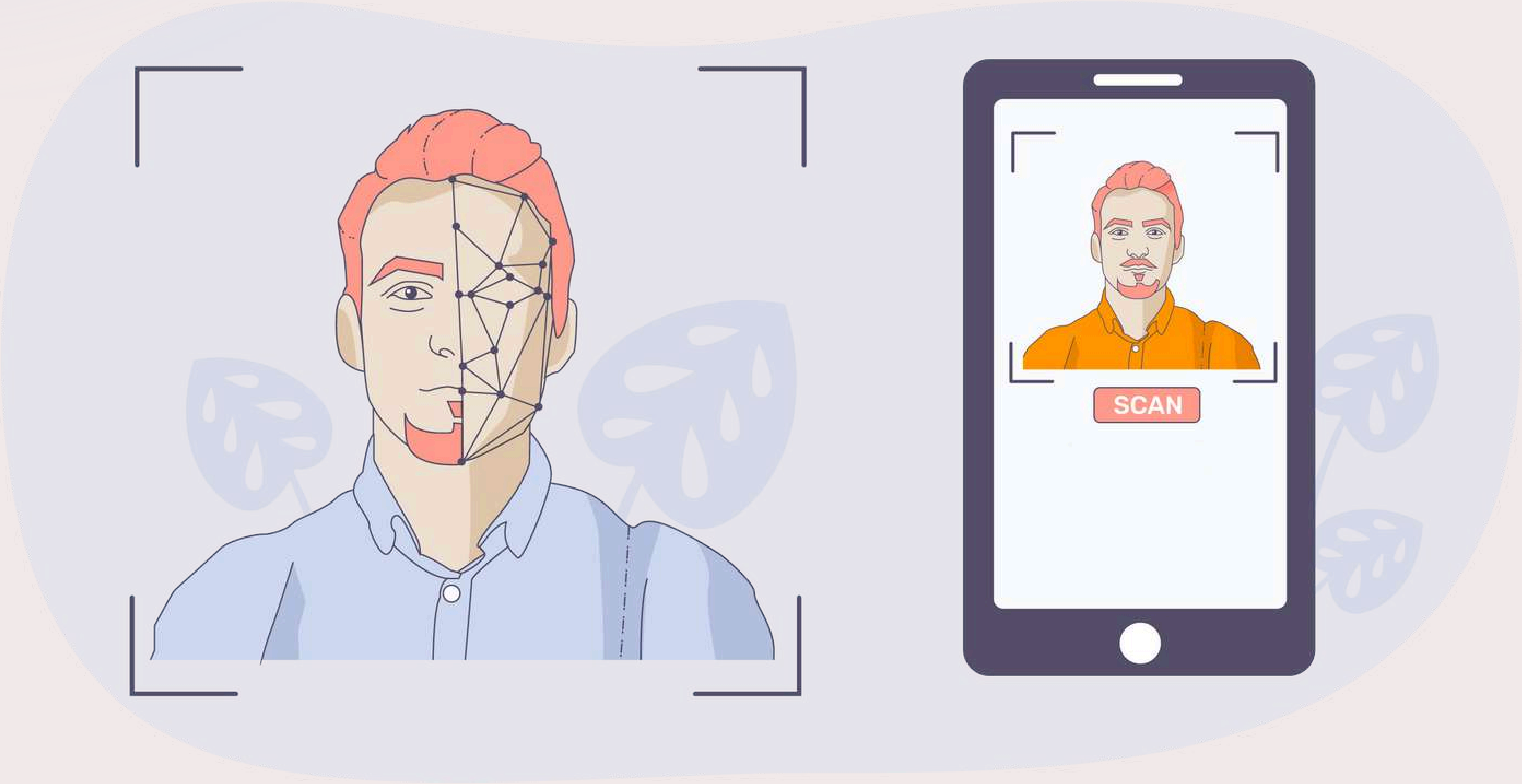
Face Recognition based on Facenet neural network



OUR TEAM

Mohamed Amine Harbaoui

Oussama Bersellou

Ahmed Amine Barkallah

Introduction

Face detection and recognition is one of the biggest areas of research in computer vision.

This is becoming increasingly important with many real-world applications such as unlocking the phone.

01 • Face detection

02 • CNN

03 • MTCNN

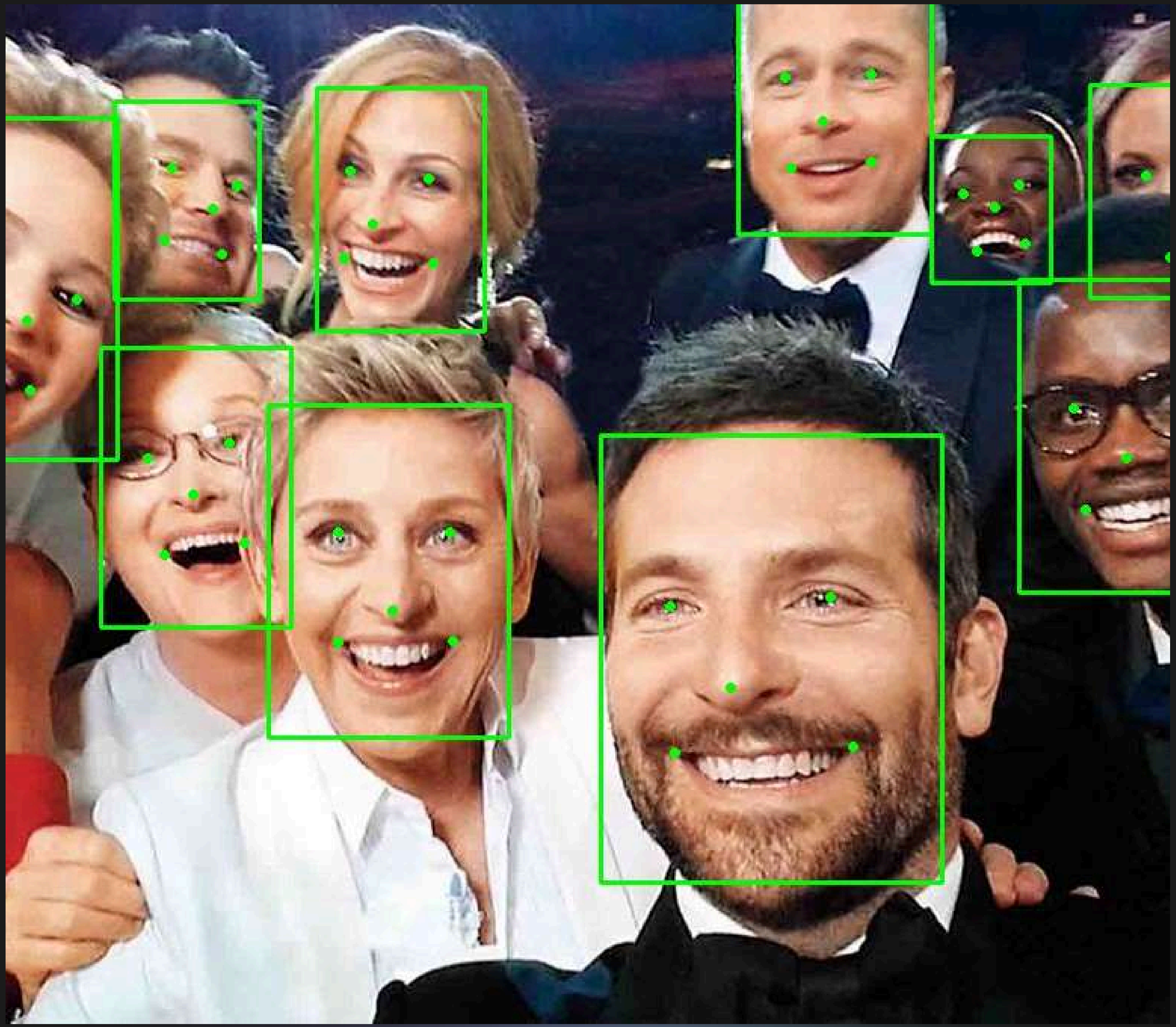
04 • Facial Recognition

05 • Facenet

Face detection

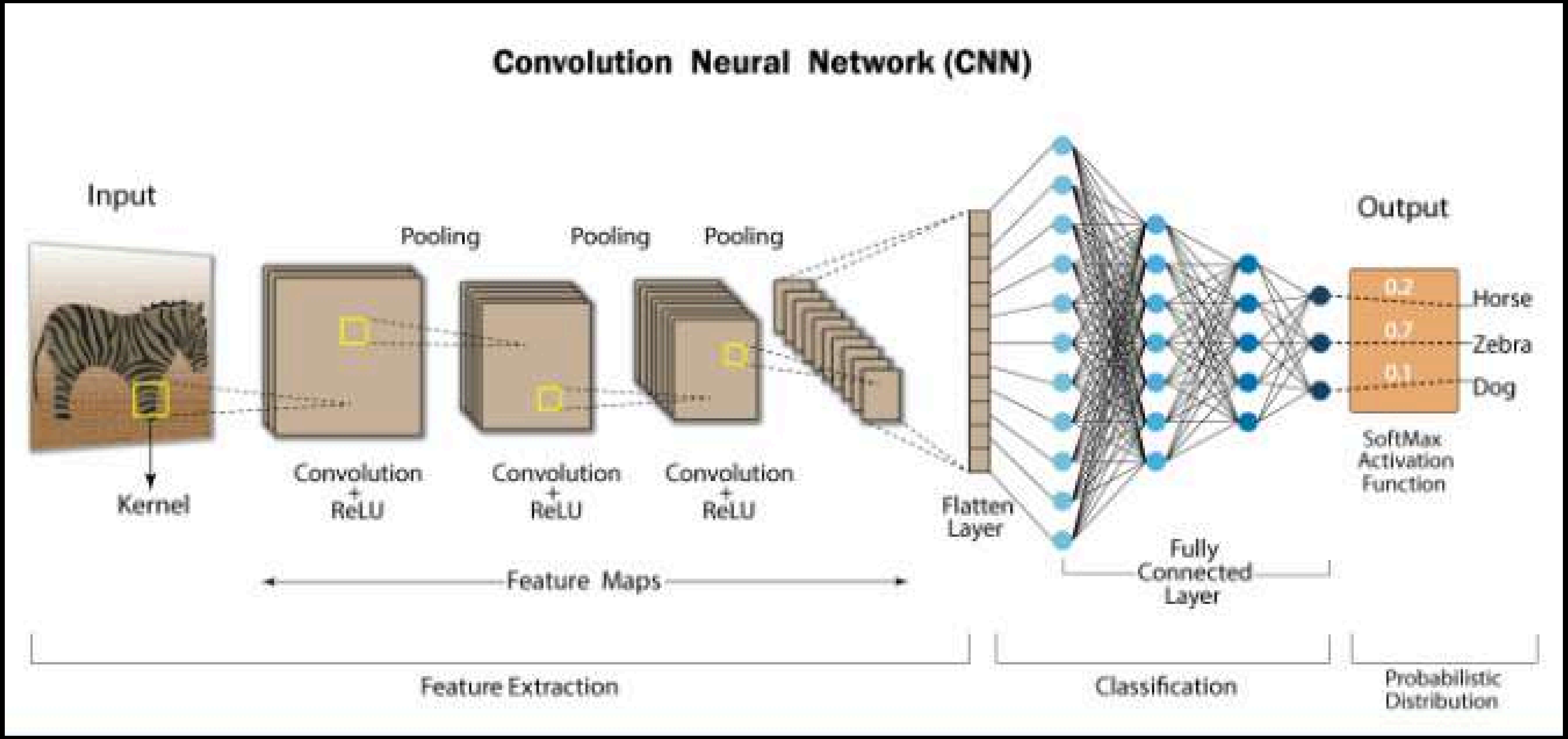
Face detection is the process of finding the face from the image, it consists in identifying all image regions that contain a face regardless of its position, orientation, and three-dimensional lighting states.

Face detection is the treatment performed just before the facial recognition phase, the identification process is not automatic and complete without effective detection phase.



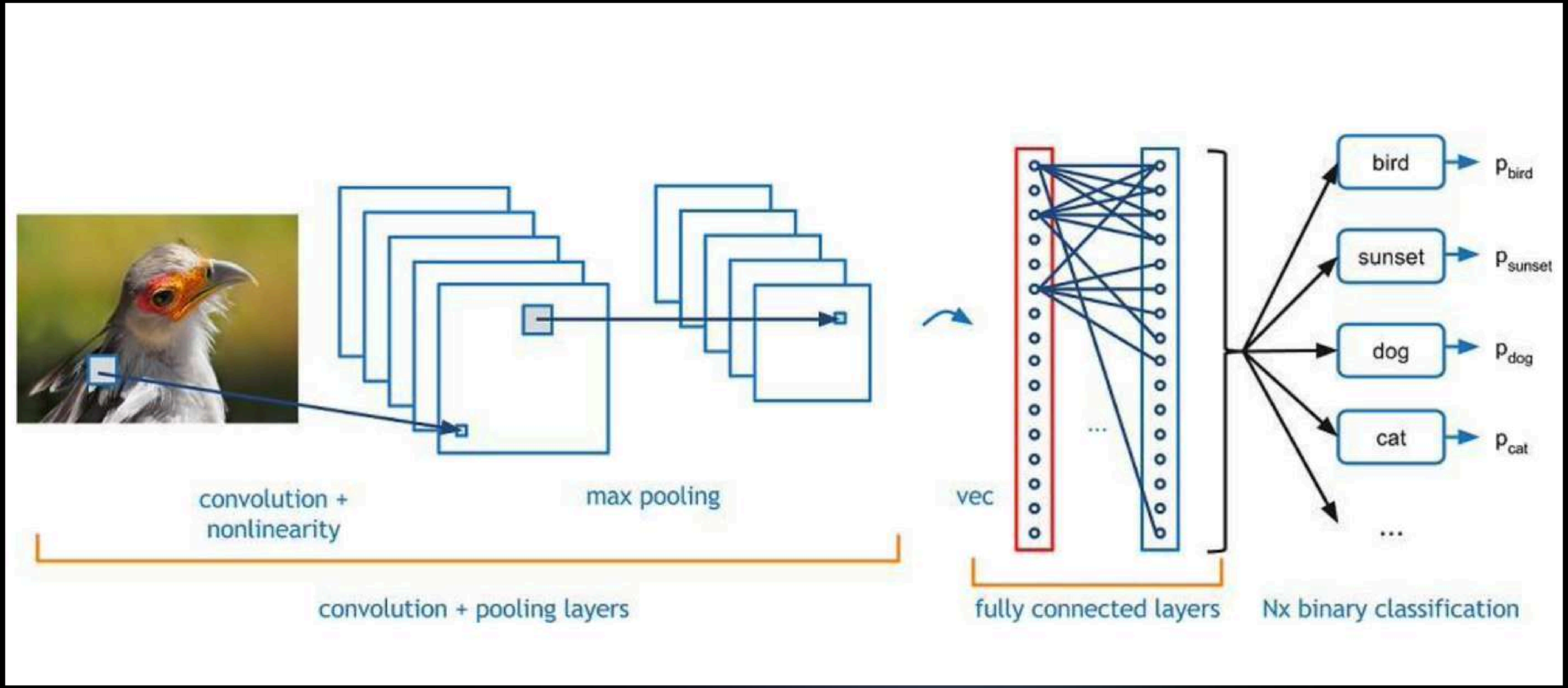
Convolutional Neural Network "CNN"

Convolutional neural networks (CNN) are a category of neural networks that have been shown to be very effective in areas such as image recognition and classification.



CNN Network Layers

- **Convolutional Layer**
- **Pooling layer**
- **ReLu (Rectifier Linear Unit)**
- **Fully Connected Layer**
- **Sofmax**



Convolutional Layer

Original image

1	9	8	4	4	5	7
4	8	6	7	9	1	7
4	0	5	9	3	8	4
7	3	5	9	0	5	4
7	4	1	1	8	1	2
7	6	6	9	8	7	6
3	6	3	5	4	2	7

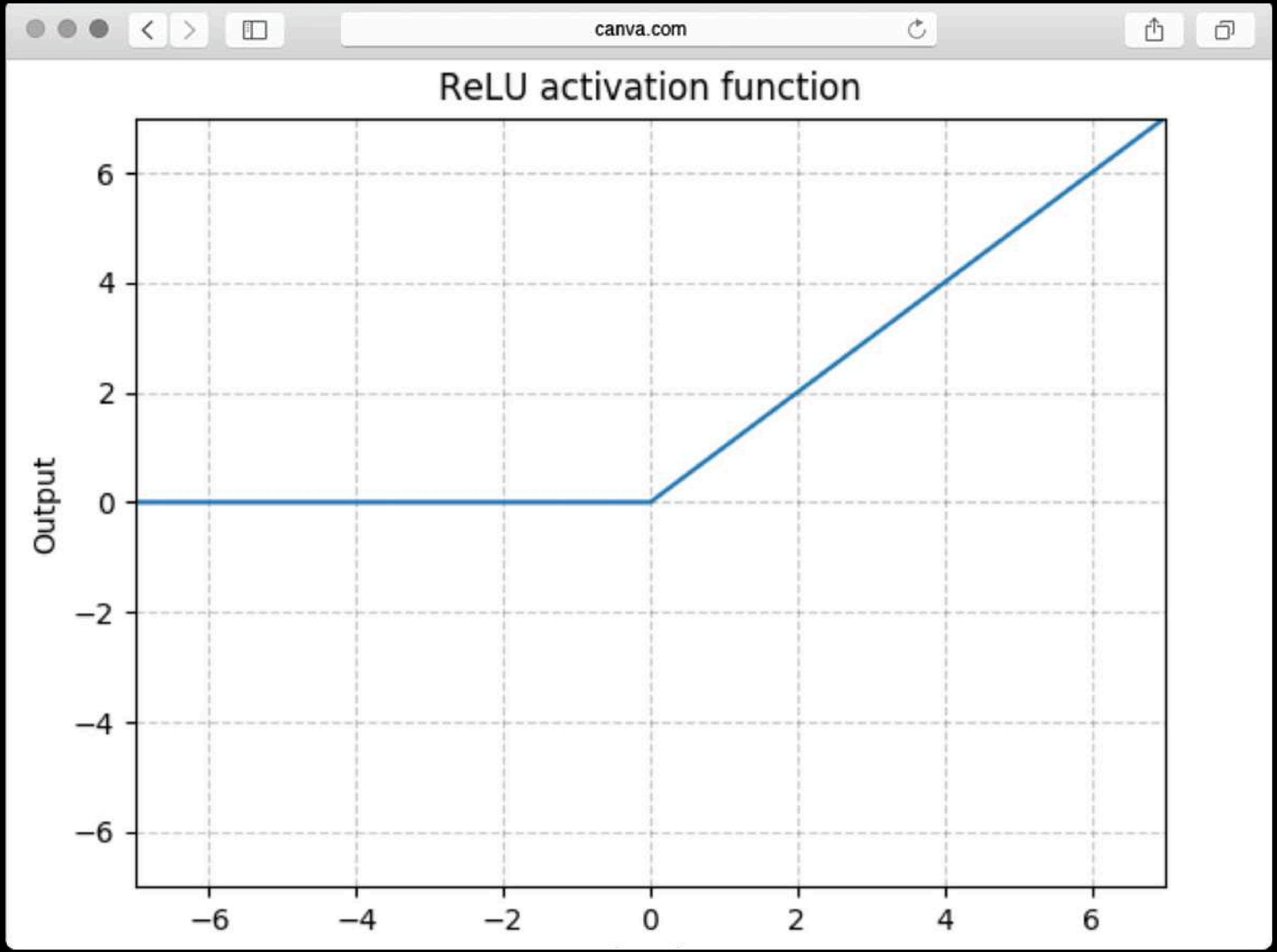
*

Filter 3x3

0	-1	0
-1	5	-1
0	-1	0

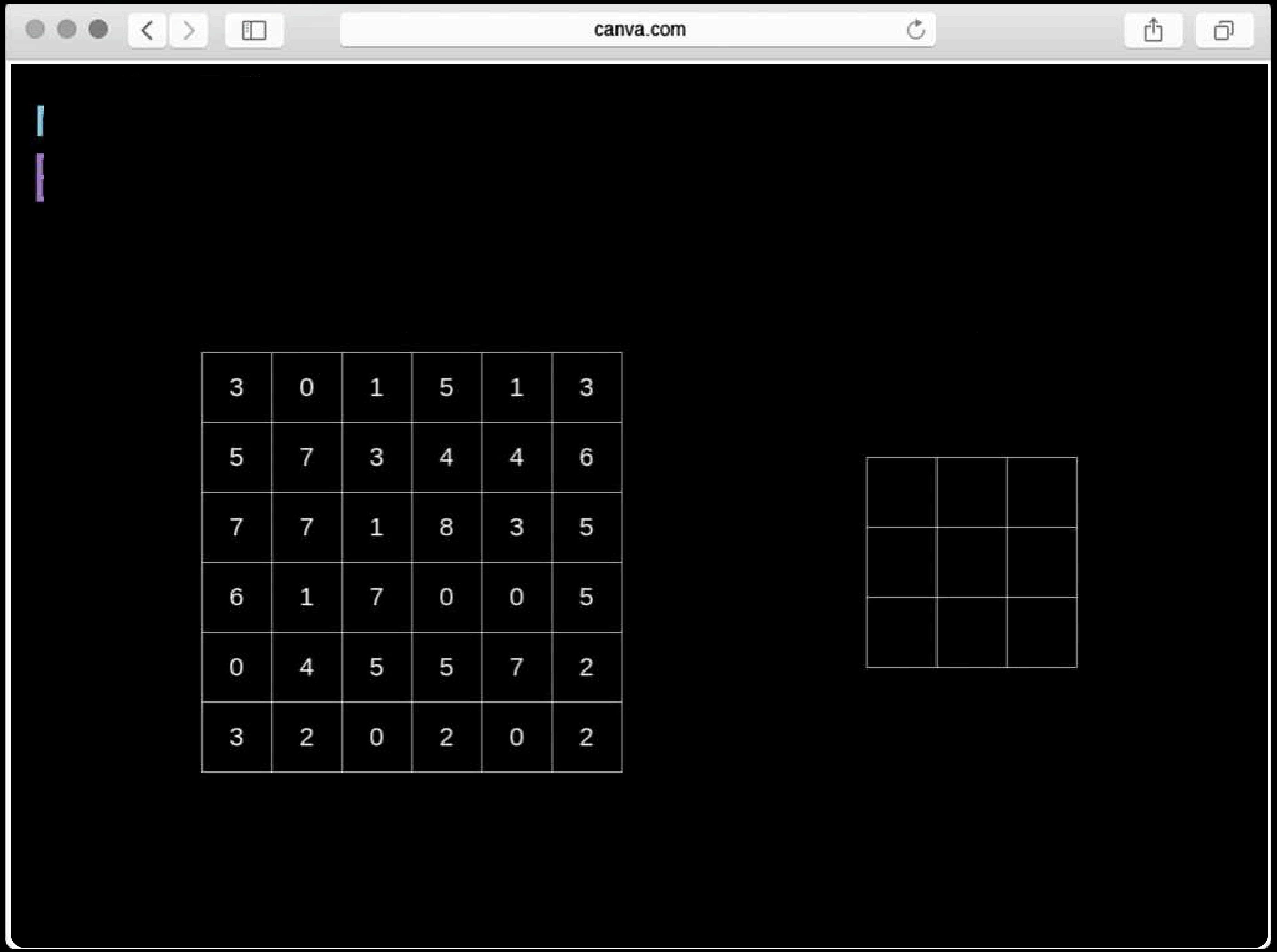
=

Output

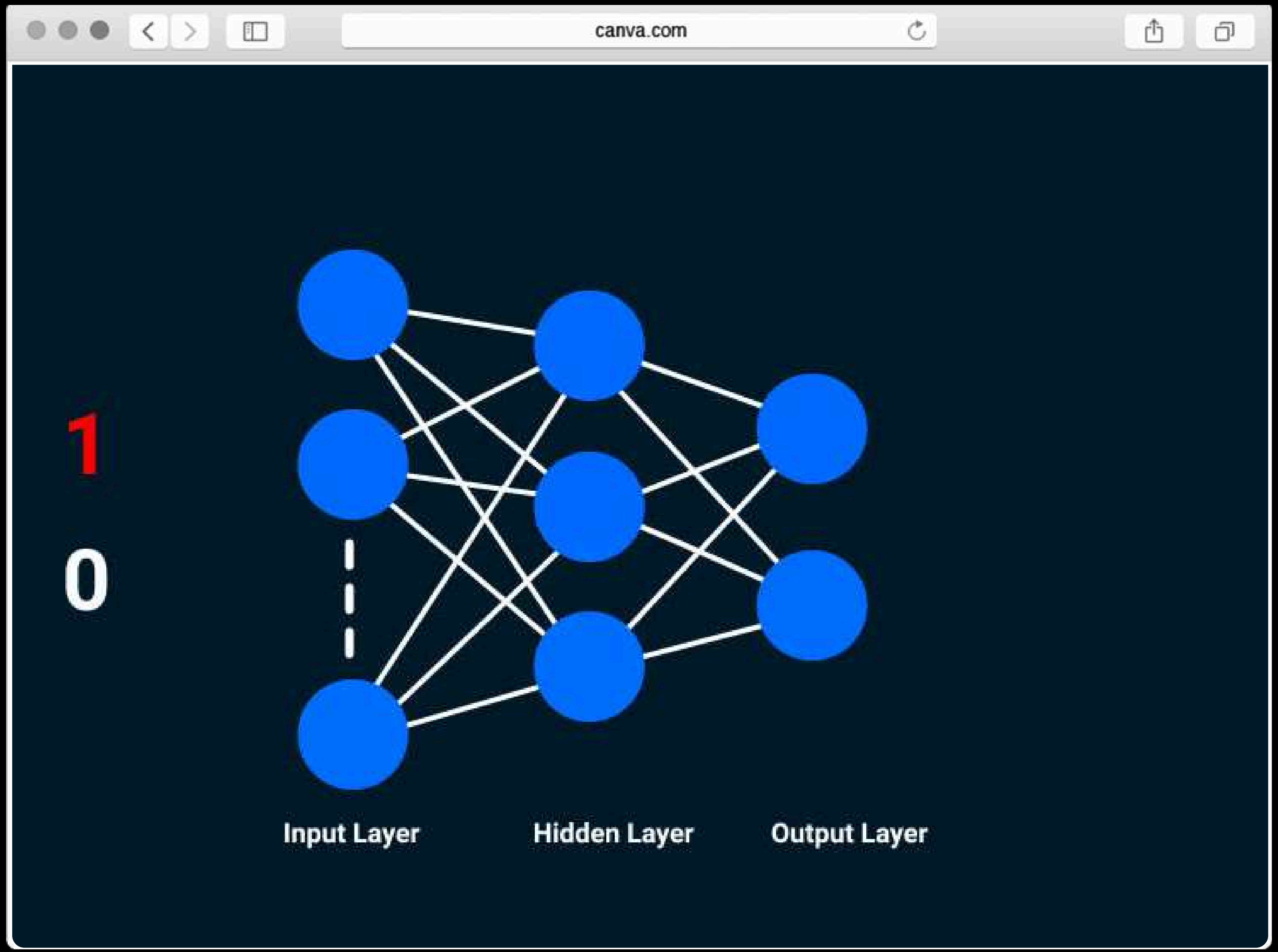


Correction Layer (Relu)

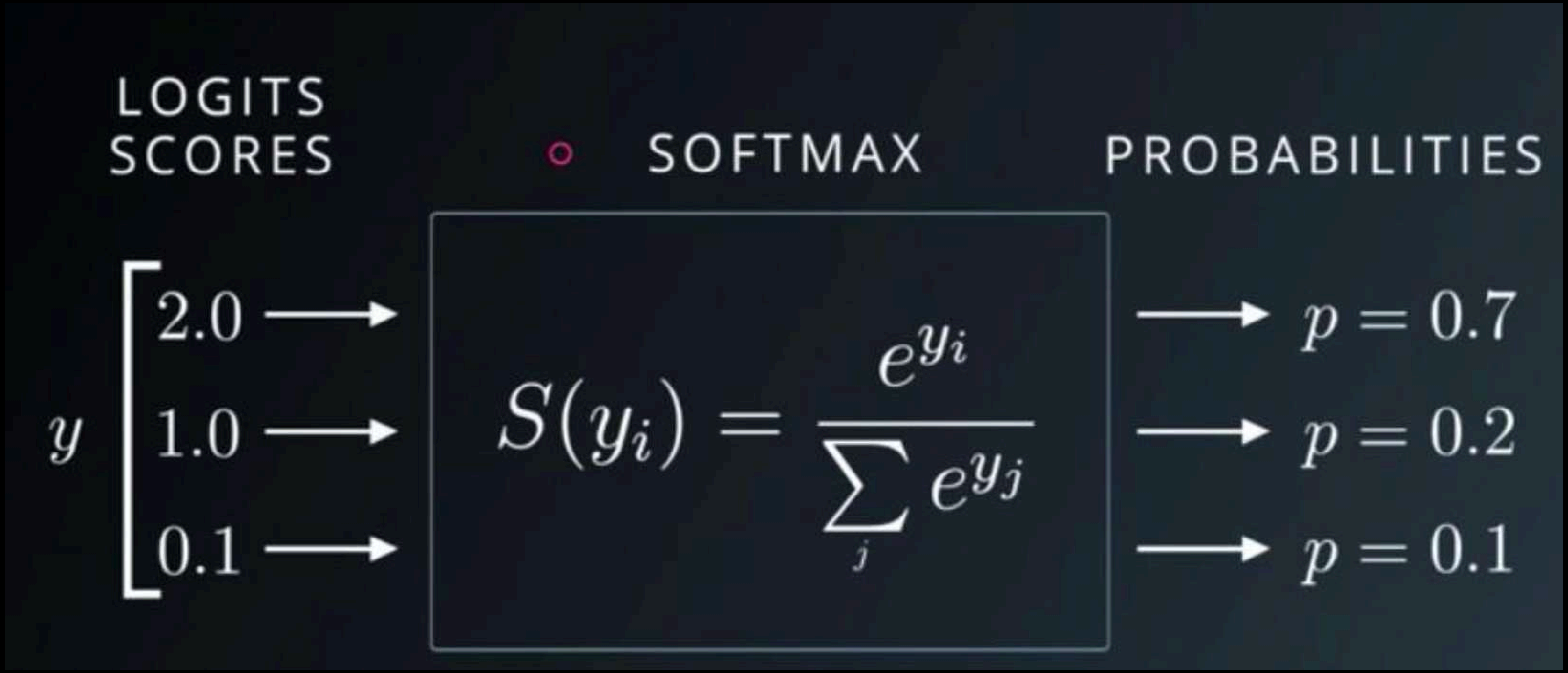
Pooling Layer



3	0	1	5	1	3
5	7	3	4	4	6
7	7	1	8	3	5
6	1	7	0	0	5
0	4	5	5	7	2
3	2	0	2	0	2

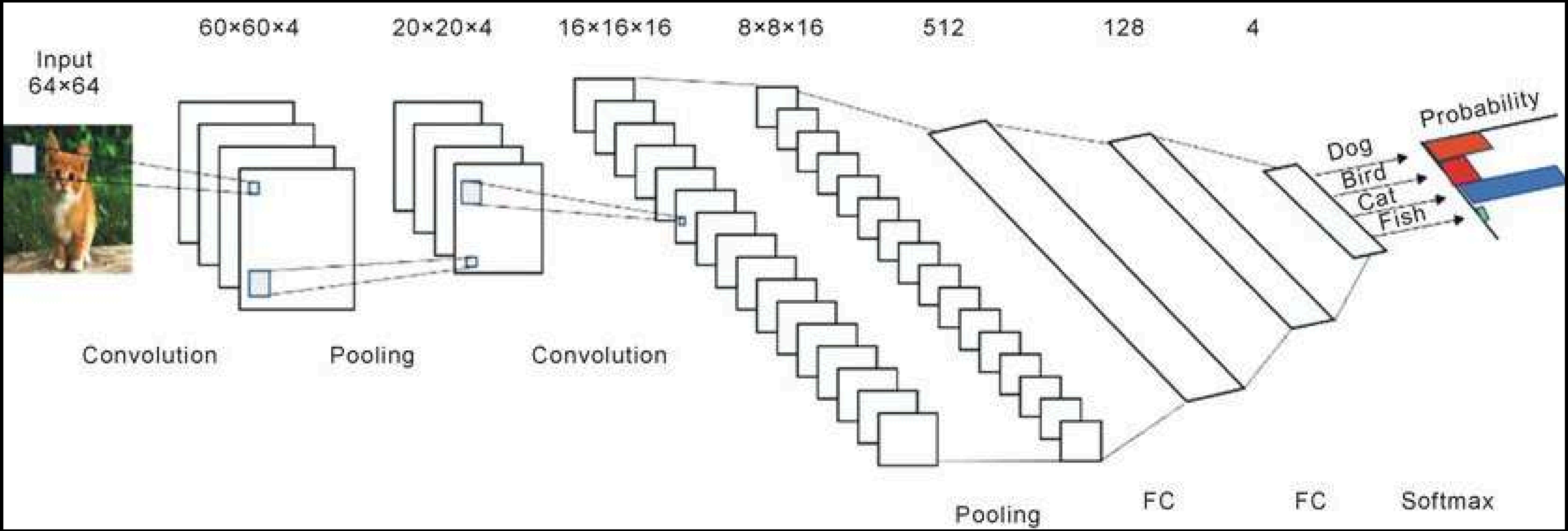


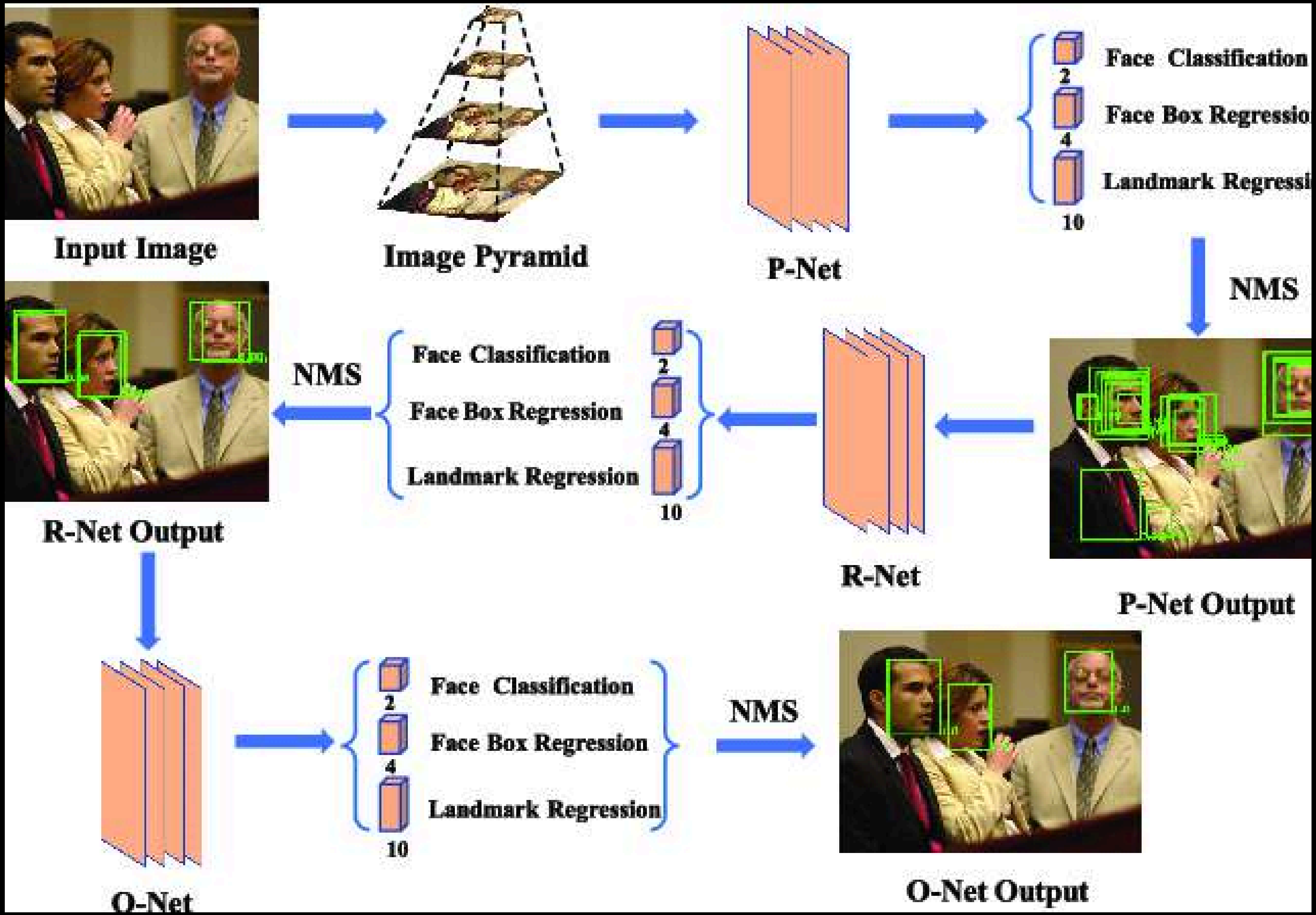
Fully Connected Layer



Classification Output Layer (Softmax)

Convolutional Neural Network "CNN"



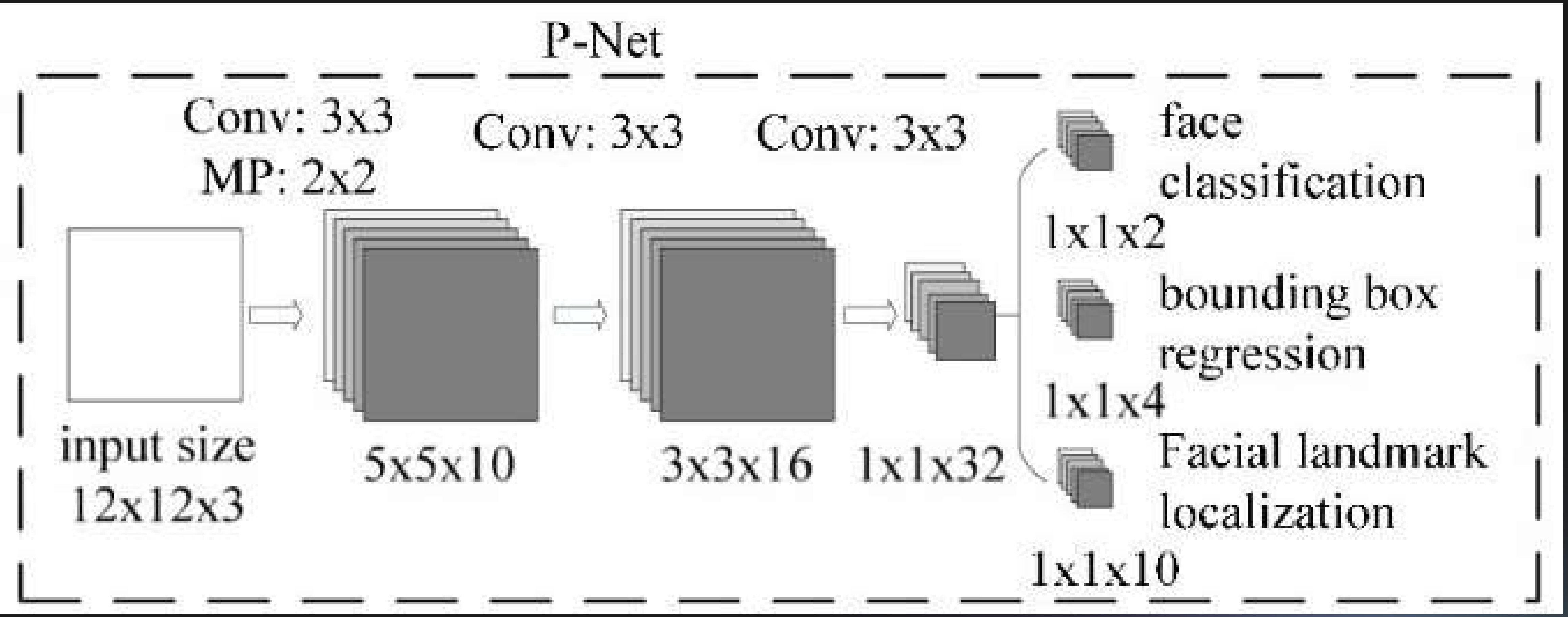


MTCNN

MTCNN is a cascading multi-task convolutional network that uses the inherent correlation between detection and alignment to improve its performance.

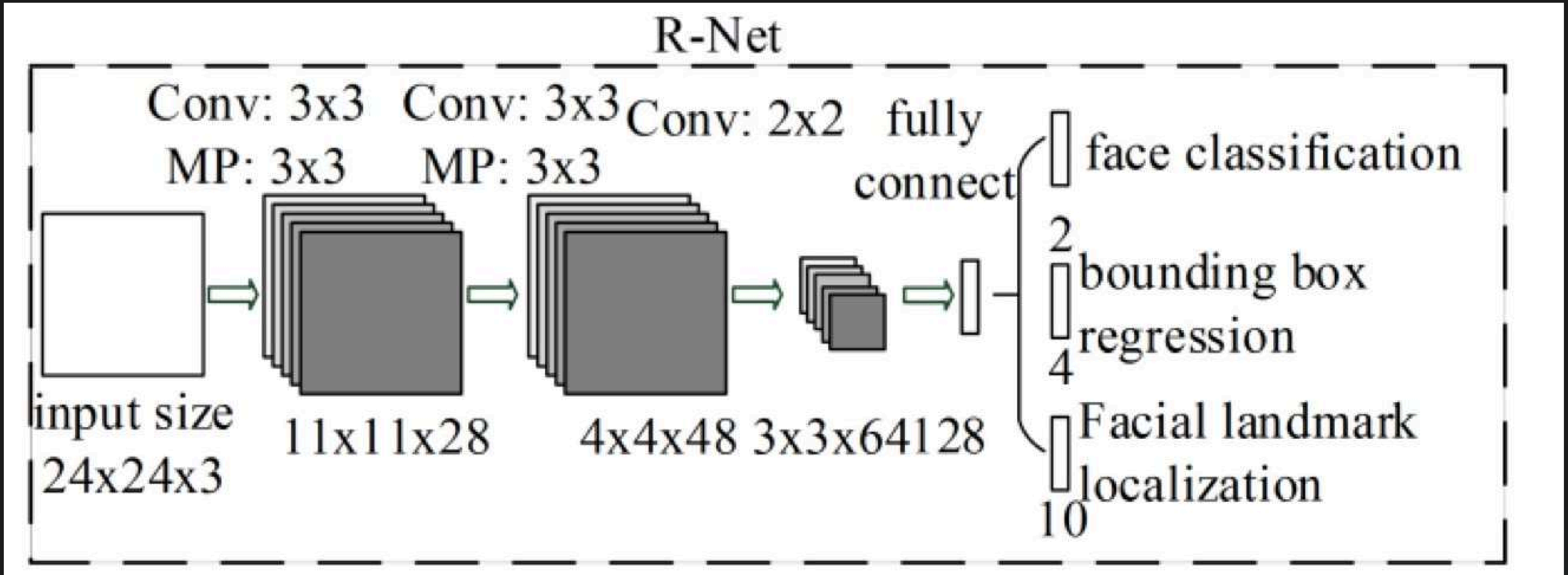
The Proposal Network (P-Net)

This first step is a fully convolutional network (FCN).
This proposal network is used to obtain candidate windows and their bounding box regression vectors.



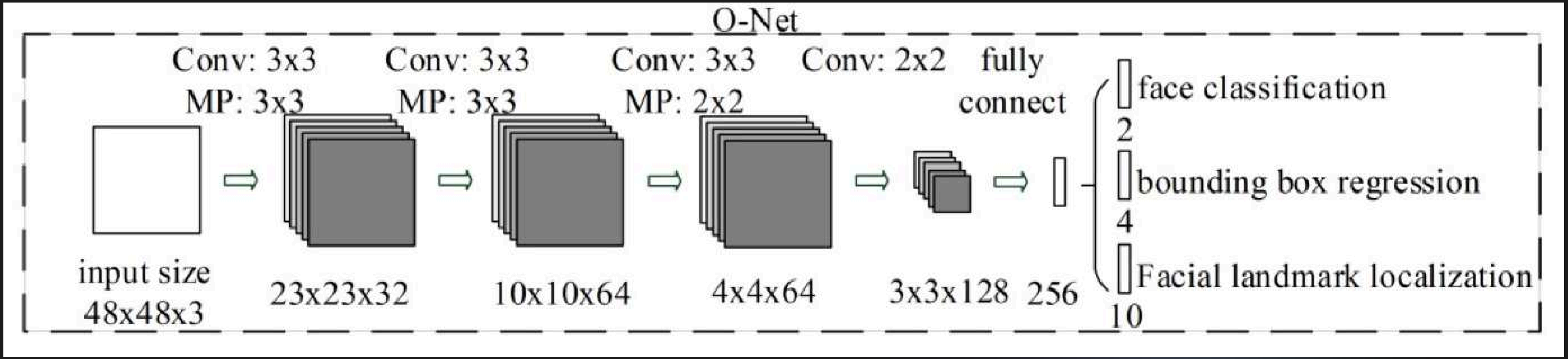
The Refining Network (R-Net)

The R-Net further reduces the number of candidates, performs a calibration with an encompassing box regression, and uses Non-Maximum Suppression (NMS) to merge overlapping candidates.

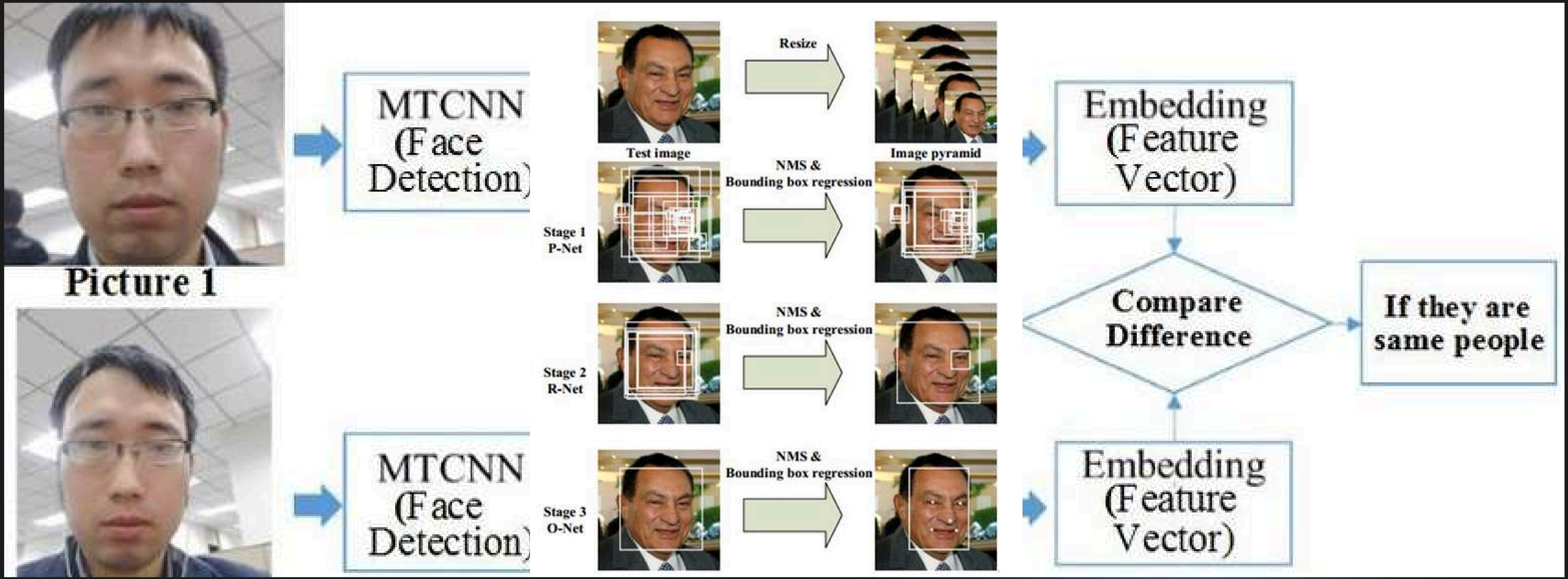


The Output Network (O-Net)

This output network aims to describe the face in more details and provides the positions of the five facial landmarks for the eyes, nose and mouth.



Architecture

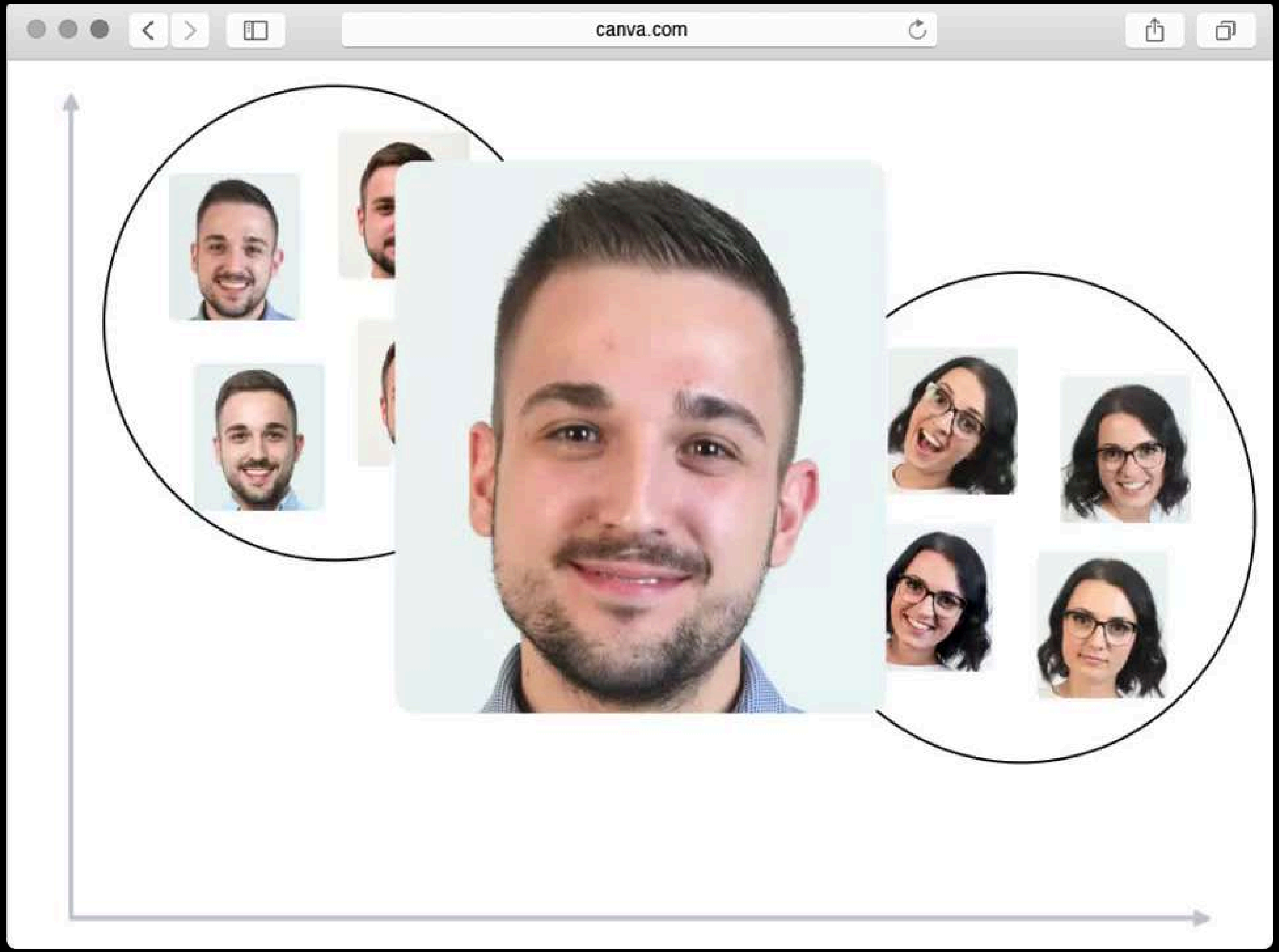


Facial recognition

Face recognition

A biometric process that consists in determining the identity of a person through his face automatically.

Several recognition approaches have been developed, classified into three categories.

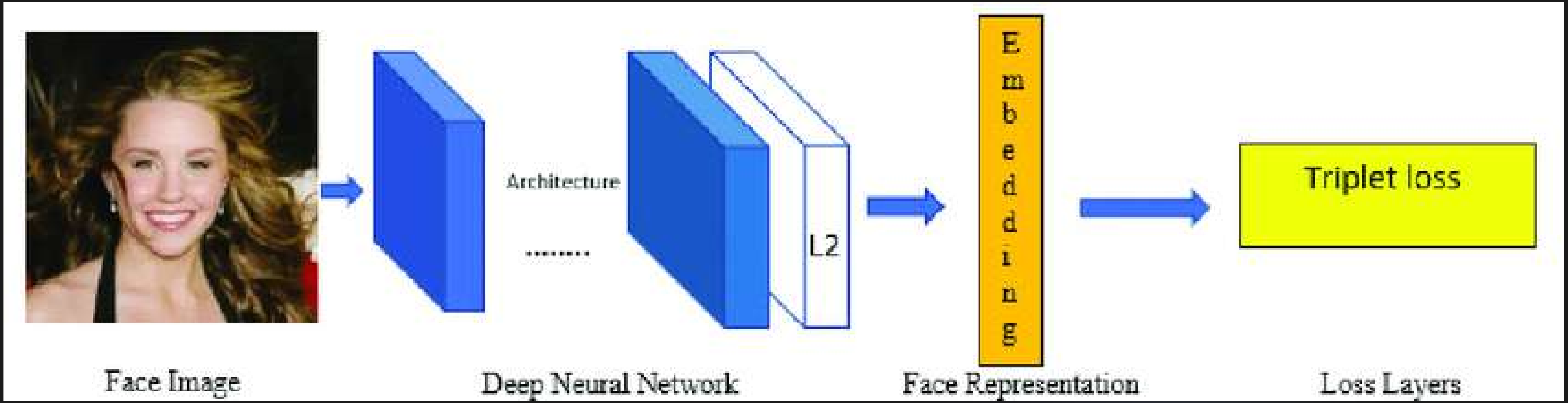


FaceNet

FACENET IS A DEEP LEARNING FRAMEWORK USED IN FACE RECOGNITION.

FACENET USES THE GOOGLENET MODEL, WHICH HAS HIGH ACCURACY IN FACE RECOGNITION.

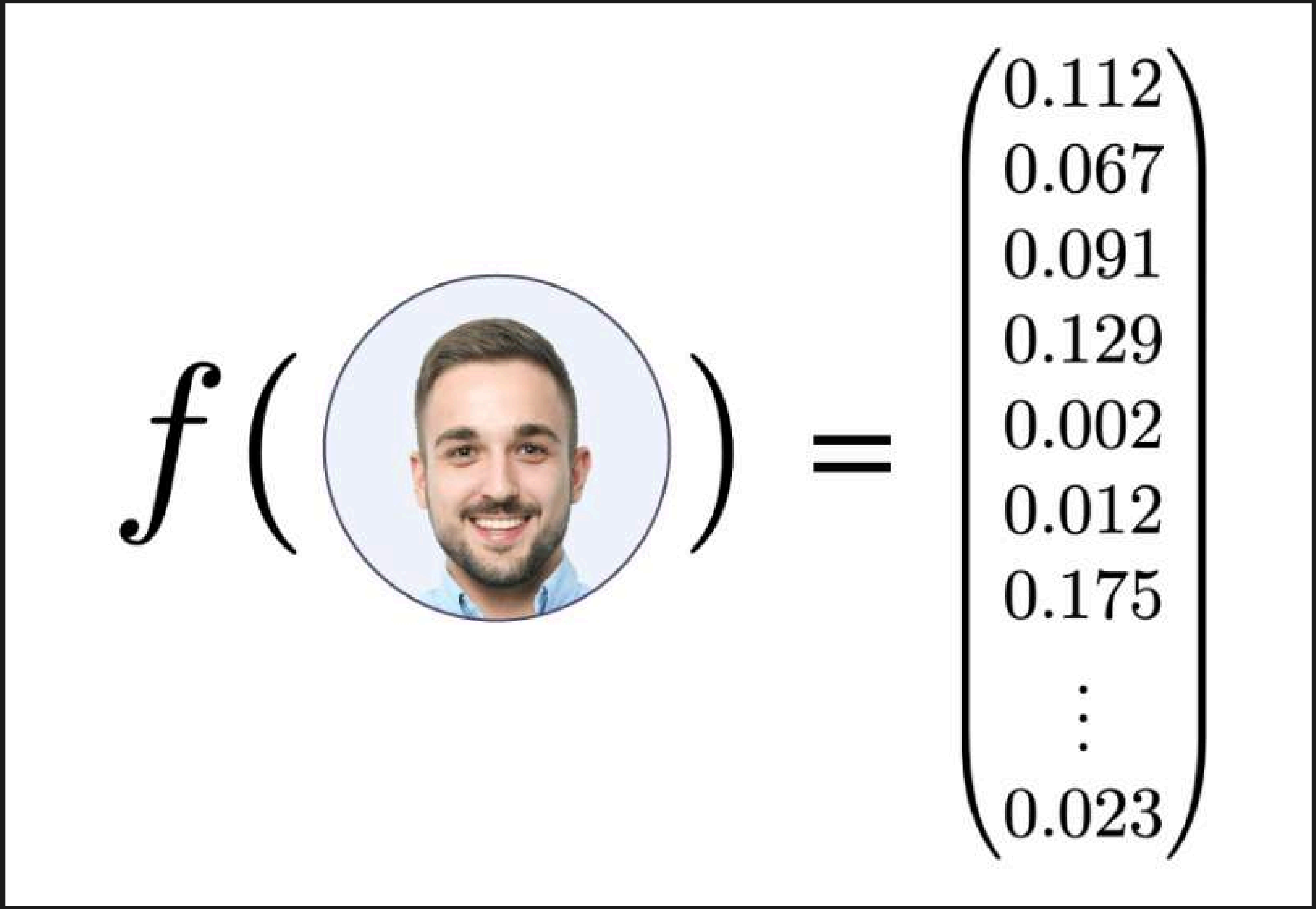
FaceNet



Architecture
....
L2
Embedding
Triplet loss
Face Image
Deep Neural Network
Face Representation
Loss Layers

Embedding

FaceNet takes as input an image of a person’s face and produces a vector of 128 numbers that represent the most important features of a face



Triplet loss

For FaceNet to learn how to generate face embeds, it uses the Triplet loss function:

1. Randomly selects an anchor image.
2. Randomly selects an image of the same person as the cracked image (positive example).
3. Randomly selects an image of a different person from the anchor image (negative example).
4. Adjusts FaceNet network settings so that the positive example is closer to the anchor than the negative example.

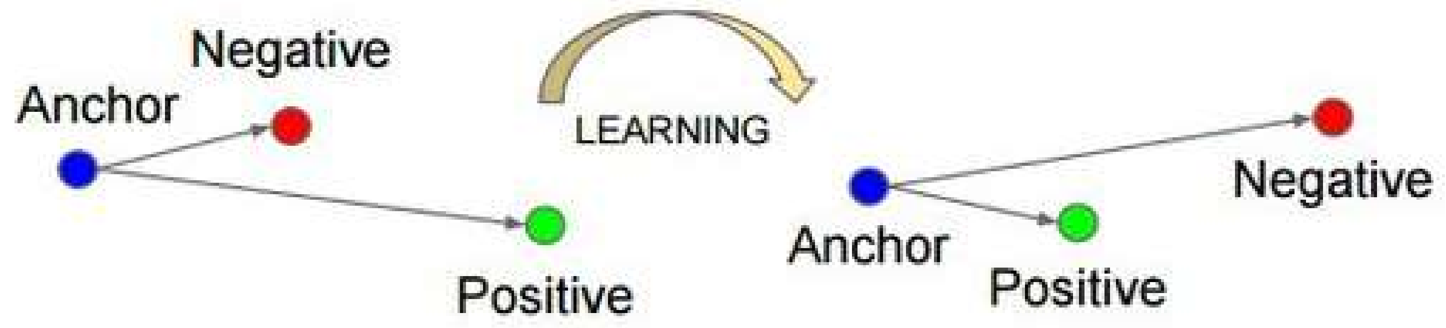
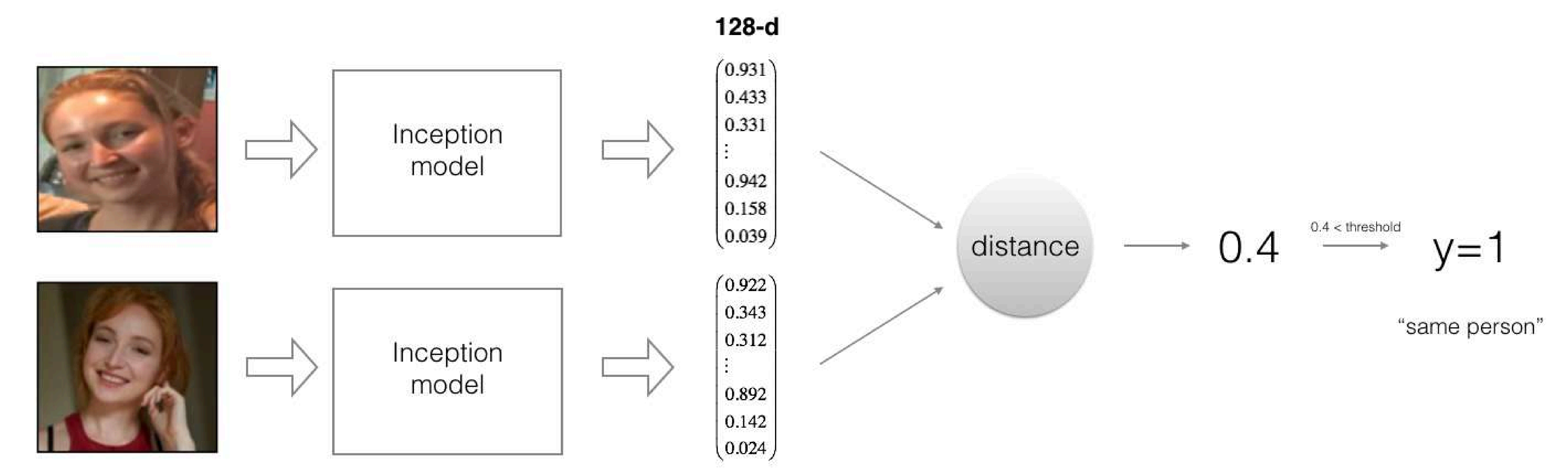
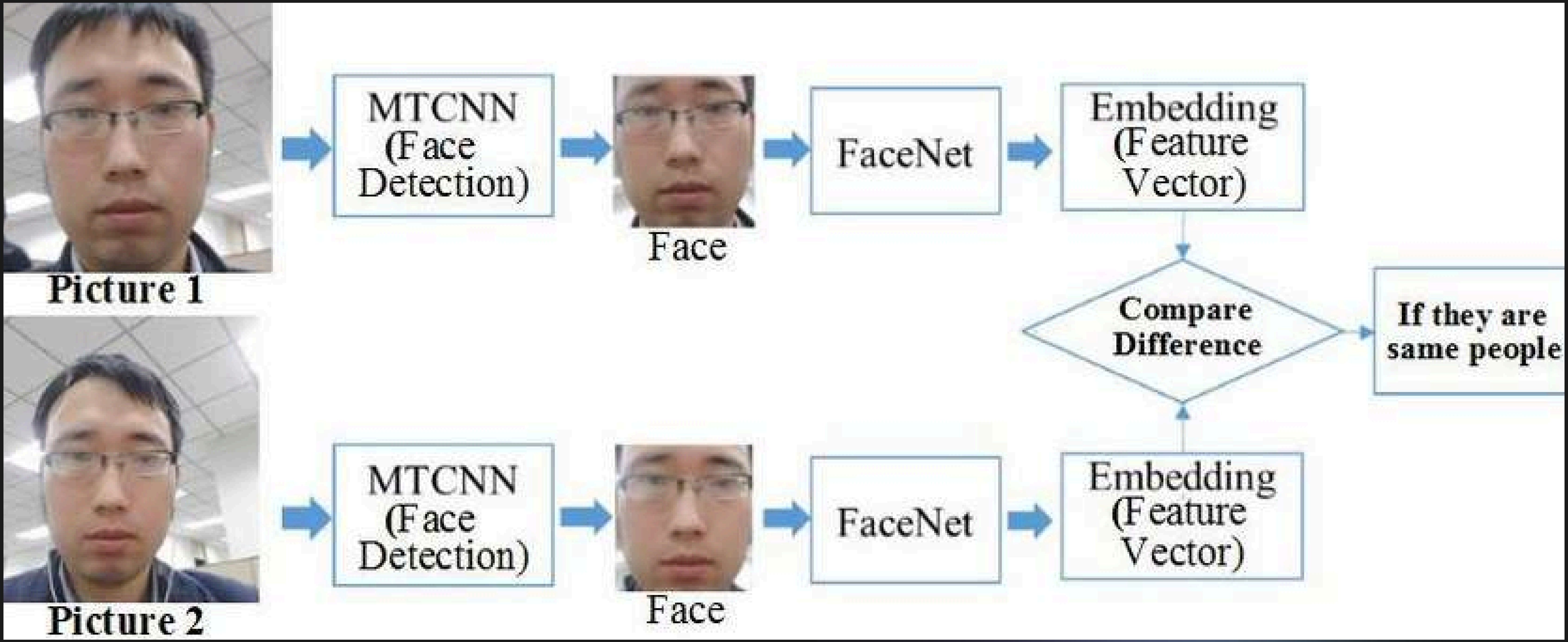


Figure 3. The **Triplet Loss** minimizes the distance between an *anchor* and a *positive*, both of which have the same identity, and maximizes the distance between the *anchor* and a *negative* of a different identity.

Triplet loss



Architecture



Analysis & Performance Evalution of our Application

Analysis

To analyze the performance of our face recognition system with respect to the given **metrics (precision, accuracy, specificity, sensitivity)**, we need to define four key concepts.

01 • TP

02 • TN

03 • FP

04 • FN

Definitions :

- **True Positive (TP)**: The number of instances where the system correctly identifies a face as the person of interest.
- **False Positive (FP)**: The number of instances where the system incorrectly identifies a face as the person of interest when it is not.
- **True Negative (TN)**: The number of instances where the system correctly identifies that a face does not belong to the person of interest.
- **False Negative (FN)**: The number of instances where the system fails to identify a face as the person of interest when it actually is.

Metrics

Let's understand what each term of the given metrics represents and how it applies to our system.

01 • Precision

02 • Accuracy

03 • Sensitivity

04 • Specificity

05 • ROC (Receiver Operating Characteristic)

Accuracy & Precision

- Accuracy: Measures the proportion of correctly identified instances (both positive and negative) out of all instances.

- Precision: Measures the accuracy of positive predictions. It focuses on the proportion of true positive predictions out of all positive predictions made by the system.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

Sensitivity & Specificity

- **Sensitivity (Recall or True Positive Rate):** Measures the ability of the system to correctly identify actual positive instances. It focuses on the proportion of true positives out of all actual positives.

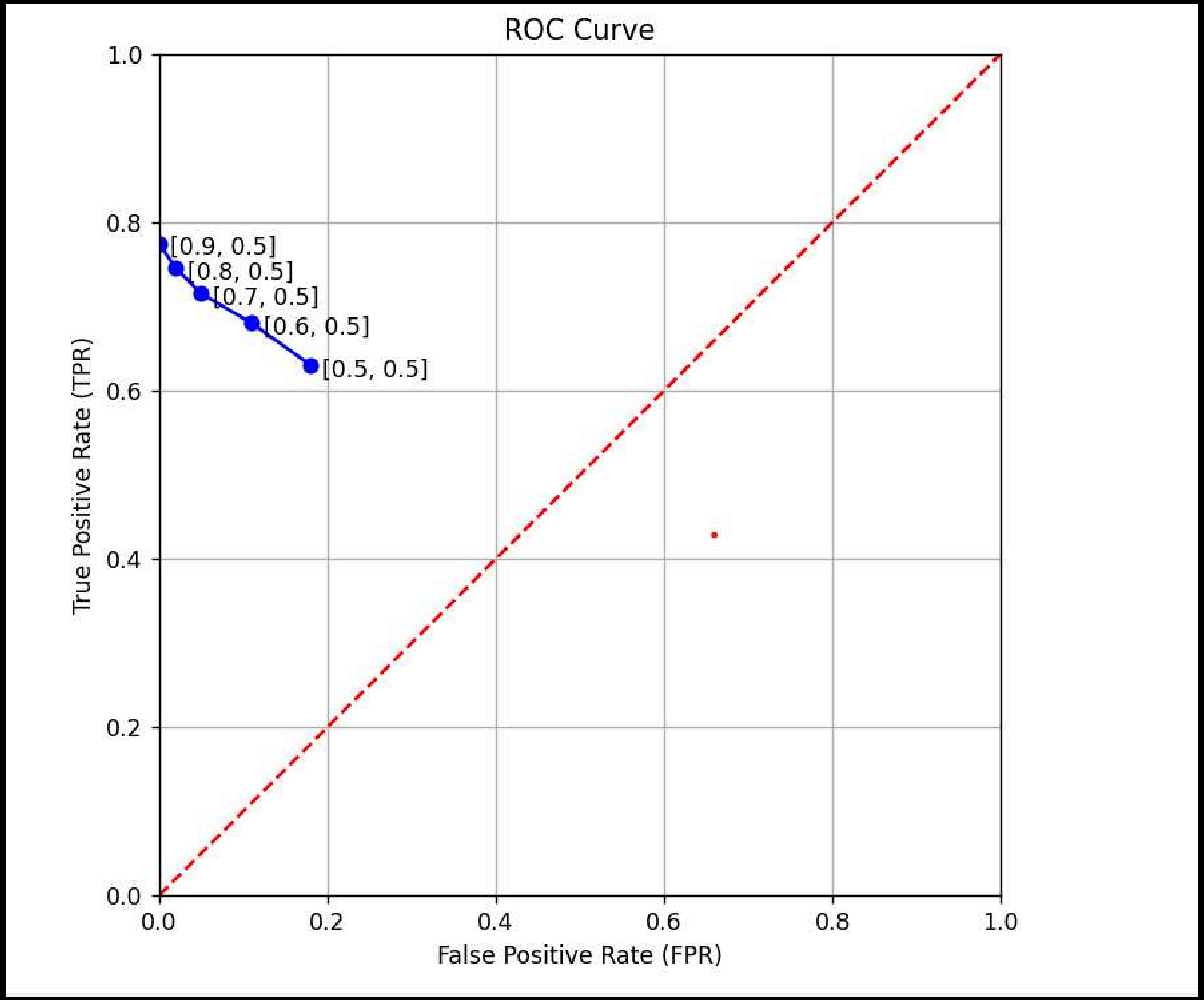
- **Specificity (True Negative Rate)**: Specificity measures the ability of the system to correctly identify actual negative instances. It focuses on the proportion of true negatives out of all actual negatives.

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

ROC

- **ROC (Receiver Operating Characteristic) Curve**: The ROC curve is a graphical representation of the true positive rate (recall) against the false positive rate (FPR) at various threshold settings.



Confidence Threshhold (confidence_t)

High Confidence in Face Detection:

- **0.9 and Above:** A higher confidence threshold ensures that only highly probable faces are considered. This reduces the chances of non-face regions being mistakenly identified as faces.
- **Precision:** High confidence levels increase precision because the face detector is more certain about the detected faces.
- **Reduced Noise:** It filters out low-confidence detections which are often false positives or very uncertain detections.

Avoiding False Positives:

- Lower confidence thresholds might include non-face regions, leading to false positives. By setting a high threshold, the system only processes detections that it is highly confident about.

Recognition Threshold (recognition_t)

Similarity Matching:

- **0.4 and Above:** This ensures that the recognized faces are similar enough to the database encodings. It balances between being too strict and too lenient.
- **Recognition Accuracy:** Setting a threshold above 0.4 ensures that faces are similar enough to the reference images, which helps in accurately recognizing individuals.

Minimizing False Positives:

- if the threshold is too low (e.g., below 0.4), the system might recognize different individuals as the same person due to insufficient discrimination.
- A moderate threshold (e.g., 0.4-0.6) helps in distinguishing between different individuals while still allowing some variability in appearances.

Avoiding Misses:

- If the threshold is too high (e.g., above 0.6), slight variations in facial expressions, lighting, or angle might cause the system to miss recognizing the same person.

To simplify

- **High confidence threshold (confidence_t):**
  - Results in fewer false positives but might miss some true positives.
- **Low recognition threshold (recognition_t):**
  - Increases the chance of matching faces but can introduce false positives.
- **High recognition threshold (recognition_t):**
  - Reduces false positives but might miss true positives.

By testing combinations, we can find a balance that best fits our specific application needs.

The key is to find an acceptable trade-off between precision and recall based on the system's requirements.

Balancing Both Thresholds

High confidence_t with Moderate recognition_t:

- Ensures that the detected faces are indeed faces and not false positives.
- Allows some flexibility in face recognition, accounting for natural variations in the same person's appearance.

In our case we choosed **(confidence_t = 0.9 and recognition_t = above 0.5)**

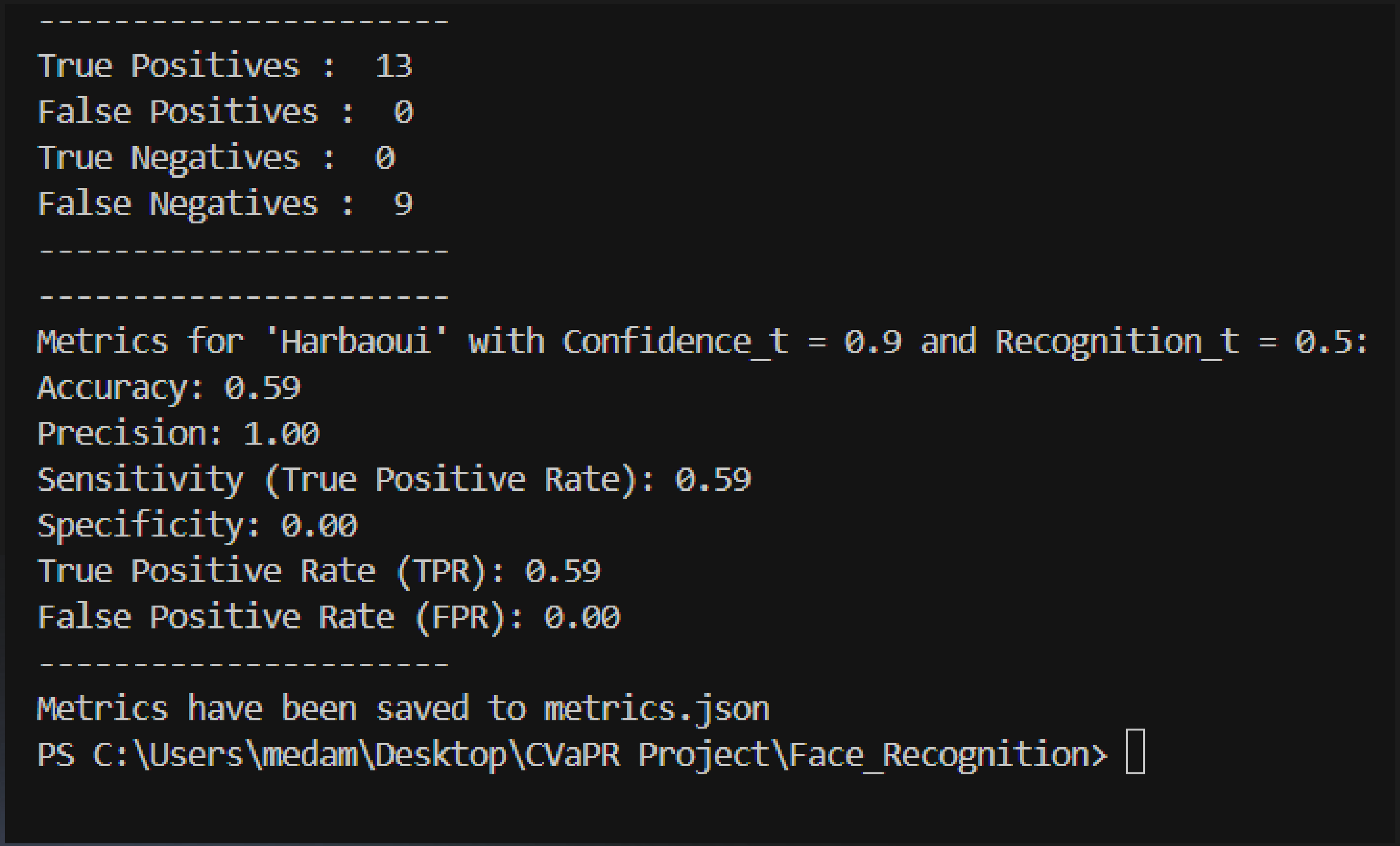
- Choosing confidence_t values above 0.9 and recognition_t values above 0.4 ensures that the face recognition system maintains a good balance between accurately identifying faces and minimizing false positives.

Practical results

TN = 0:

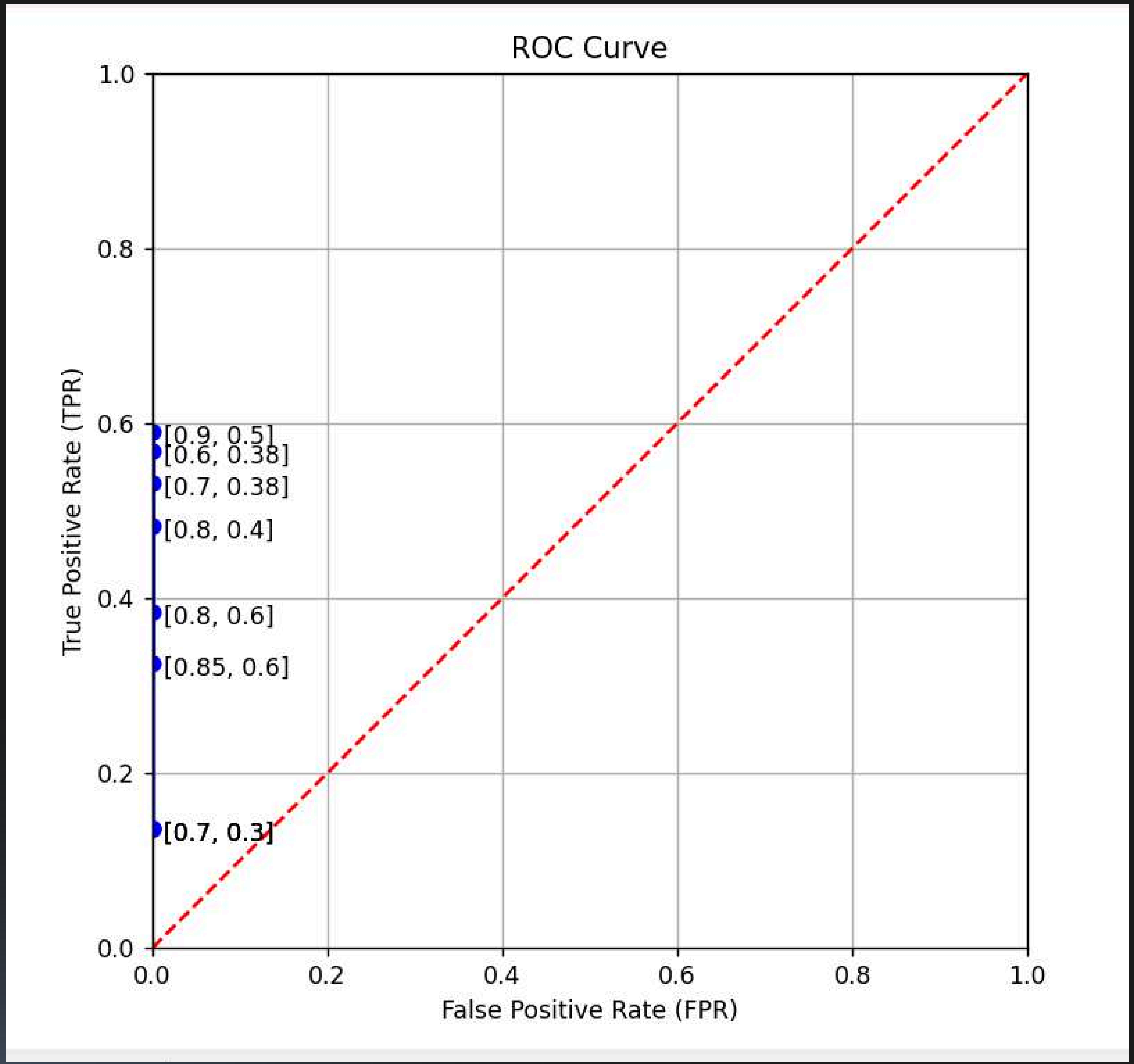
- In face recognition tasks, especially in scenarios where the focus is on **identifying specific individuals** which is our case, TN might not be relevant or calculated because **the model's primary task is to correctly identify the person of interest rather than exclude others.**

- TN becomes relevant in scenarios where the model needs to distinguish between the target class (eg. 'Harbaoui') and others, which is not the case here.



ROC Curve

This figure shows how ROC Curve looks like after testing multiple combinations of Confidence Threshholds (confidence_t) and Recognition Thresholds (recognition_t).



FINISH !