



Indian Institute of Technology Kanpur

Kalyanpur

Kanpur -208 016

# **Interim Project Report on Machine Learning And Data Science**

**Topic:**

**CNNs for Face Detection and Recognition**

**Submitted By:**

Name: V. Sai Nikhil (16MIS0257)

College: Vellore Institute of Technology – Vellore

E-mail: [v.sainikhil2016@vitstudent.ac.in](mailto:v.sainikhil2016@vitstudent.ac.in)

## **Abstract of Project:**

We have many face detection methods for detecting the face of human. It is used in many industries, classrooms, offices and other places for security reasons. Usually this surveillance will detect human faces. In our project we do both detection of human face and also recognize the persons face.

This face detection project tackle the main problem of working in real world environment since it has high rate of accuracy and run in real time system. We store the persons image in the dataset and use the system to train these images and detect and recognize the human face. Not only this detect the single face in the frame, but also multiple faces in the image frame.

## **Methodology:**

- First, we need to locate the face and detect it from the image.
- Second, we need to classify the face that is detected, identify number of people in the image.
- We need to check the accuracy and time taken for each face to detect and classify, since it is implemented in real time.

## **Related Works:**

- Sliding window: It detect the present and absent of human face in the image-sliding window. This was CNN based object detection algorithm. However, the drawback of this sliding window is that we need to repeat the process for period of time, which resulting in more computing power.
- Regional proposal method: They take the potential regions for high possible object detection. This number of regions is reduced than the sliding window. This method is also called as R-CNN. However, the time taken to train the model is slow and also consume much memory.
- YOLO and SSD: This method does not use the regions to scan for objects. The input image directly goes through the convolution neural network, and they are divided into multiple grids and perform classification score on each grid. They reduce the training and testing time, but the accuracy is not achieved completely.
- All these above methods have their benefits and flaws. In out method of detection and recognition of human face, we use a new method which takes less time and indeed achieves better accuracy.

## Proposed Method:

- **Two Stream CNN:** In the first part, we detect a single human face from the rest of the image. The output will have the human face coordinate and size of the bounding box as well as the class of it. This module has 6 primary modules, each has one convolution layer, one max-pooling and one leaky ReLU layer. The last layer is connected to fully connected layer one for predicting the location and size.
- **Cascade CNN:** In the above model, we cannot detect multiple faces. We perform sliding window across whole image. Each window is fed into convolution layer for binary classification. second stage, non-maximum suppression(NMS) is used to eliminate highly overlapped detection region. This help to tackle multiple face detection and classification.

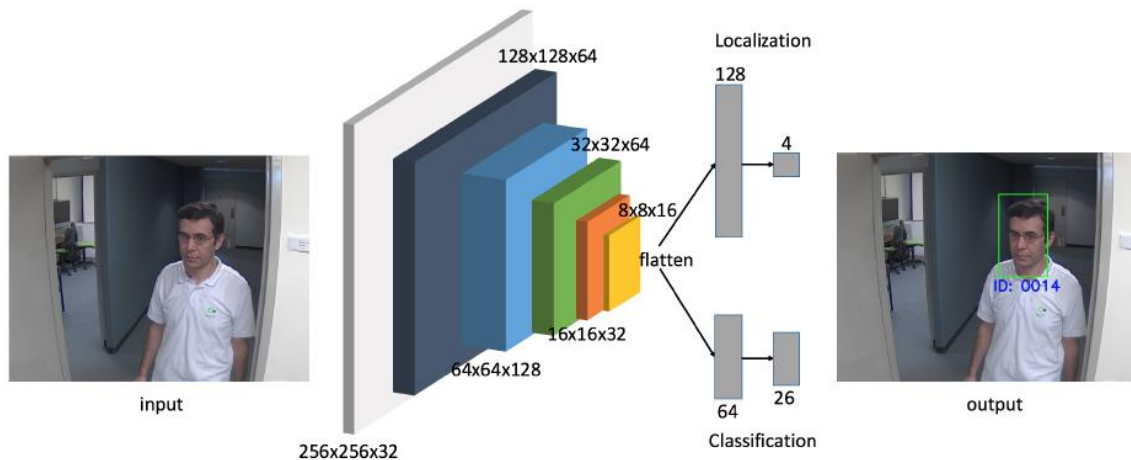


Figure 3. The simplified architecture and framework of Two Stream CNN

## Loss Function:

The loss function is of two parts. The coordinate loss and the size loss of the bounding box as well as the classification loss. In the last layer we use softmax cross entropy loss.