Data Format Description

Deep Learning

ITESO, Universidad Jesuita de Guadalajara

Gregorio Alvarez

Introduction

The goal of this project is to build upon a prior initiative from the Predictive Modeling course, which aimed to develop a model for an app designed to distinguish female and non-female users. The model's objective is to foster a secure environment where women can form a community and exchange experiences within the app.

Problem Definition

The project encompasses three core components: **Image Detection, Classification, and Verification**, each tailored for distinct stages of the user identification process. Initially, the image detection algorithm is tasked with locating the subject's face within an image, while ensuring that it neither detects multiple faces nor processes unclear images. Subsequently, the classification algorithm comes into play, designed to discern between male and female users. Finally, the verification algorithm is used for security aspect, permitting entry to the app only after confirming the individual's identity as a recognized member. The images obtained from the regisration process are used during the verification process.

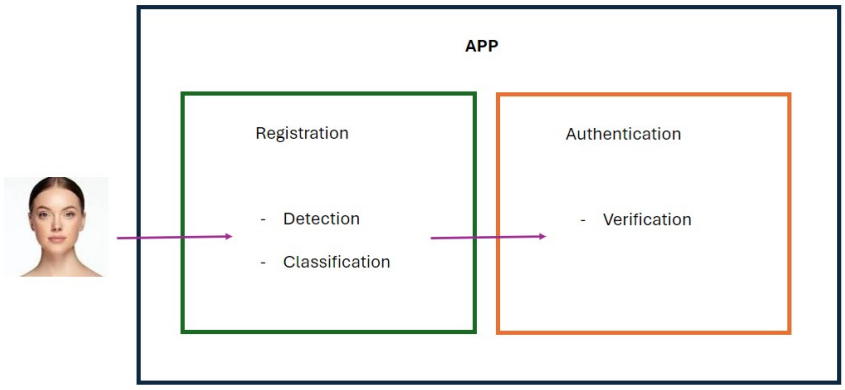

Figure 1: Diagram illustrating the sequence in which models are integrated within the app for user identification and verification

Data Format

Detection:

The predictors X for the detection task are represented as a tensor of images with dimensions $(m, H, W, 3)$, where m represents the batch size or the number of examples processed by the network simultaneously. The parameters H and W denote the height and width of the images, respectively, and the number 3

corresponds to the three color channels in an RGB image. Aiming for an initial model akin to YOLO, the anticipated output tensor is expected to have the dimensions $(m, G, B * 4 * 1)$, where G signifies the number of grid cells and B denotes the quantity of bouding boxes used assigned to each grid cell.

The parameter m will be determined by the computational resources at our disposal. The dimensions H and W for the input images will be adopted from a preceding project, with a resolution of 225x225 pixels.

Following the methodology suggested by Joseph R. in 2016, we will utilize a grid size of 7x7 (indicating $G \times G$) with $B = 2$ bounding boxes for each grid cell in the initial model iteration. This choice is informed by the relatively straightforward nature of the task at hand and the constraints imposed by the limited computational resources available.

Classification:

For the classification problem, the independent variable remains an $(m, H, W, 3)$ tensor. The dependent variable, however, is an (m) vector that represents the likelihood of each image belonging to the target class (female).

Verification

For the verification problem, the independent variable remains an $(m, H, W, 3)$ tensor. The dependent variables are encapsulated within a vector of size d, which serves to locate the representation of a given face within a d-dimensional Euclidean space.

Taking into account the findings of Florian S. from 2015, we will adopt a value of $d = 128$ for the dimensionality of this space. This choice is based on the balance between having a sufficiently rich representation for accurate facial recognition and the computational efficiency of working with a vector of this size.

References

- [Joseph R. 2016] Joseph, R. (2016). You only look once: Unified, real-time object detection. IEEE, 779-788. https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Redmon_You_Only_Look_CVPR_2016_paper.pdf
- [Florian S. 2015] Florian, S. (2015). FaceNet: A unified embedding for face recognition and clustering. IEEE, 815-823. https://arxiv.org/abs/1503.03832