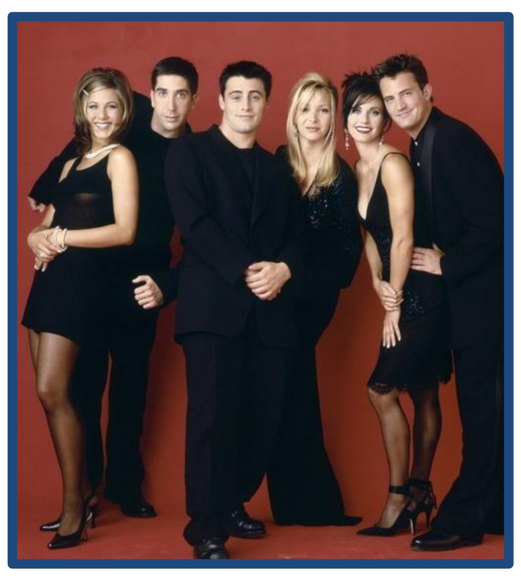
Dynamic Recognition: Extending Faster RCNN Capabilities for Facial Identification in Media Stream

Qian Tang
Department of CS, University of Toronto

Motivation

- I address the growing need for advanced **face identification and tracking**. My solution extends its benefits to individuals with prosopagnosia, aiding them in recognizing actors/actresses. This holds significant value for media companies for cataloging purposes and provides an essential tool for security organizations for surveillance and identification.
- My system leverages **faster Regions-Base Convolutional Neural Network** algorithms to not only detect and identify faces in videos but also to track their movement across different frames. A key feature is its capability to match faces with an existing database for identity verification.
- Ultimately, my project **bridges the gap between technological advancement and enhanced accessibility**, marking a step forward in inclusive media consumption and security measures

Aim to solve cases like below:



Related Work

Faster R-CNN framework: [1]

- **Method:** Applied Faster R-CNN to face detection, marking a shift from traditional, slower R-CNN methods.
- **Limitation:** While effective for detecting faces, Faster R-CNN does not inherently support face identification. It can locate faces in images but cannot distinguish between individual identities.

Face Recognition: [2]

- **Model:** Composed of three convolution layers: two pooling layers, two fully connected layers, and a Softmax regression layer. Employed stochastic gradient descent and the Dropout method to mitigate overfitting.
- **Limitation:** This CNN architecture, though proficient in recognizing faces, may struggle with identification under diverse and complex real-world conditions, such as varying angles, lighting, or facial obstructions.

Face Mesh and 3D reconstruction: [3]

- **Model:** Addressed limitations in face recognition under varied conditions using Face Mesh for 3D reconstruction.
- **Limitation:** The primary focus is on accurate face reconstruction and recognition under diverse conditions. However, the method may face challenges in reliably differentiating between closely resembling faces or in scenarios with extreme variations, which are critical for precise face identification.

References

[1] H. Jiang and E. Learned-Miller, **"Face Detection with the Faster R-CNN,"** 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, USA, 2017, pp. 650-657, doi: 10.1109/FG.2017.82.

[2] K. Yan, S. Huang, Y. Song, W. Liu and N. Fan, **"Face recognition based on convolution neural network,"** 2017 36th Chinese Control Conference (CCC), Dalian, China, 2017, pp. 4077-4081,doi:10.23919/ChiCC.2017.8027997.

[3] Shivalila Hangaragi, Tripty Singh, Neelima N, **"Face Detection and Recognition Using Face Mesh and Deep Neural Network"**, Procedia Computer Science, Volume 218, 2023, Pages 741-749, ISSN 1877-0509

Method

- **Data Processing**

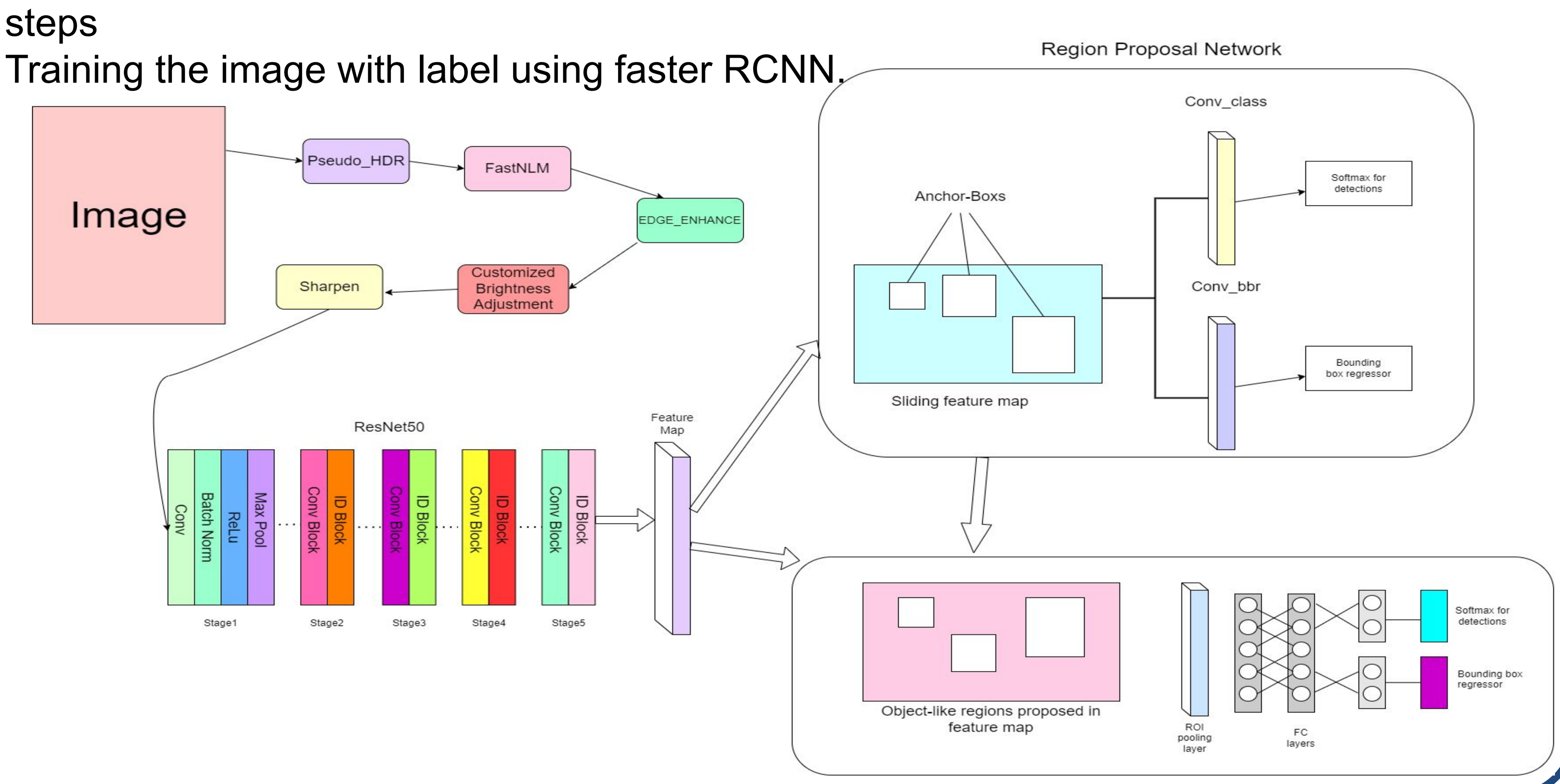
My dataset is meticulously curated from 'Friends' Season 6, employing OpenCV's face recognition capabilities to extract and identify characters from video frames. I leverage a sophisticated algorithm, designated ‘DSFD,' to pinpoint the precise locations of faces within these frames. A bespoke Python script is then utilized to annotate each frame, labeling the actors and actresses for consistent tracking.

To ensure robust training, my dataset is diversified into distinct categories: **90%** with open eyes and **10%** with closed eyes, along with an equal distribution of **50%** frontal and **50%** profile images. This stratified approach is designed to enhance the model's accuracy and performance across a spectrum of visual scenarios.

- **Model Training**

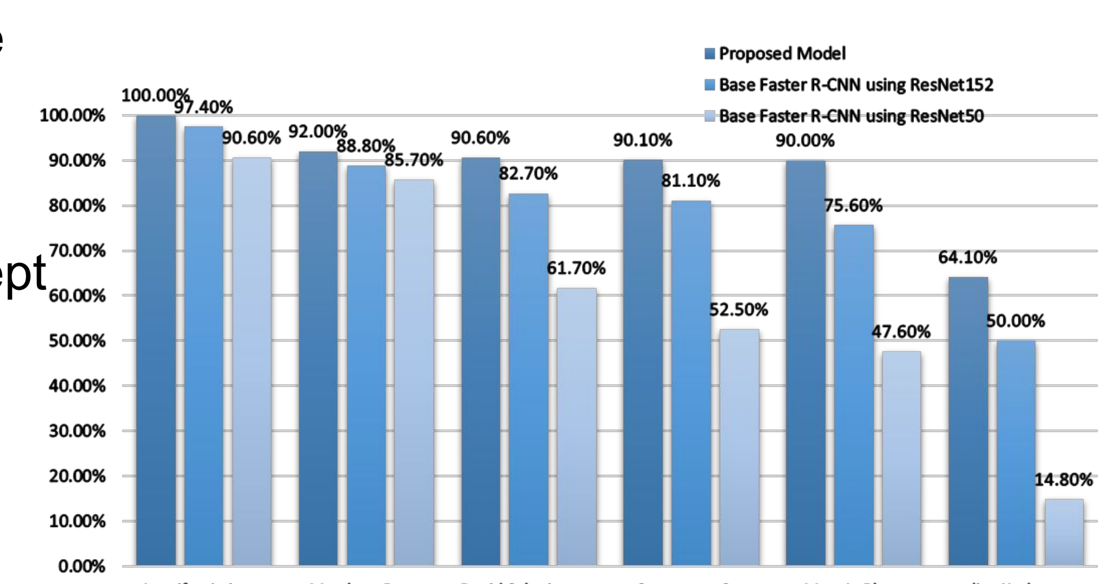
1. Use Pseudo-HDR on images appearance of a wider range between the lightest and darkest areas of the image, providing more detail in highlights and shadows.
2. Use Fast Non-Local Means to do noise reduction.
3. Balance the brightness of each image output.
4. Apply EDGE_ENHANCE in images making the edges within the image more pronounced.
5. Sharpening image softening effects caused by earlier steps
6. Training the image with label using faster RCNN.

comparison(raw data/sample output)

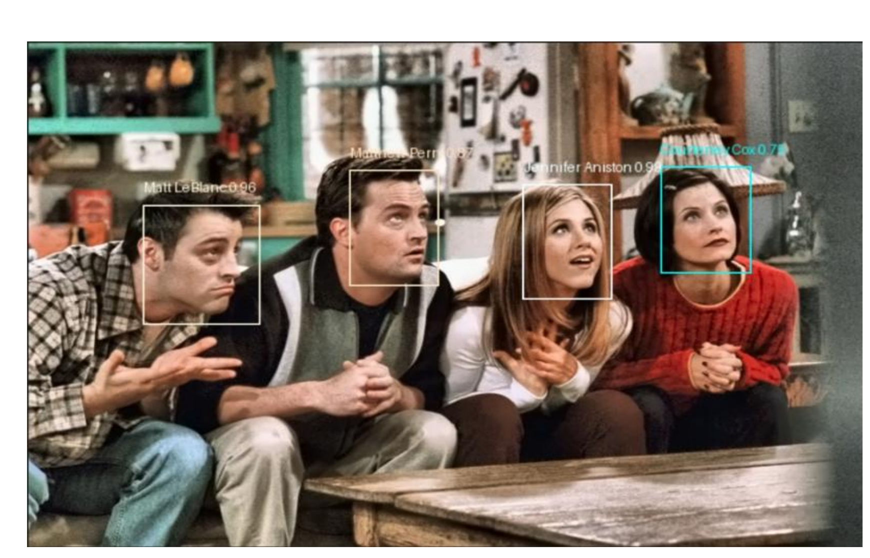


Experimental Results

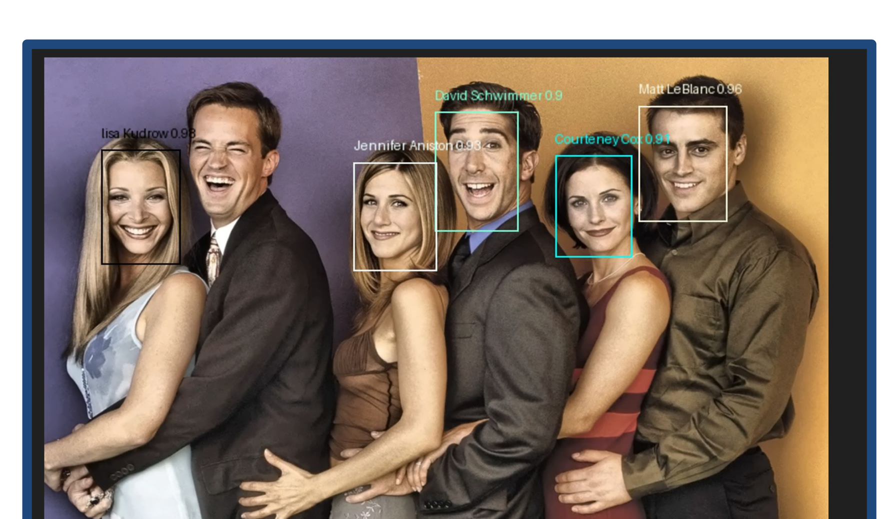
- I have reached an overall mean average precision of **87.81%**, with each actor/actress identification precision shown in the bar chart on the right. Except actress “Lisa Kudrow”, I got other accuracies all above **90%** which is a **significant improvement** compared to Baseline Faster R-CNN models.



- My model still shows good accuracy when encounters **complicated cases** which contains three or fmy actors/actresses in the same frame. Two examples are shown below:



- When an actor has an **overly exaggerated** expression(right side), my model may not be able to identify all people. This is related to **dataset shortage** and I am still working on improving my training set variety.



Computer Science
UNIVERSITY OF TORONTO