LATEX Semantic Segmentation task for OpenEDS dataset using Dual Model with Skip Connections

Adish Rao
PES University
Electronic City Campus, Bangalore
contact.adishrao@gmail.com

Aniruddha Mysore
aniruddha.mysore@gmail.com

Poulami Sarkar
poulamisarkar101@gmail.com

Siddhanth Ajri
y2jsiddajri@gmail.com

Abishek Guragol
abhishekguragol@gmail.com

Rahul Suresh
rahs98@gmail.com

Rahul Suresh
rahs98@gmail.com Gowri Srinivasa

gsrinivasa@pes.edu

Abstract

Many AR/VR applications require accurate and precise eye tracking solutions. This solution should work for all persons having different eye colour, and for different orientations of the eye. One means of achieving this is by semantic segmentation of the key regions of the human eye. Our research is based on the OpenEDS: Open Eye Dataset** paper, that presented a pixel-level annotated data set for key eye-regions: iris, pupil and sclera. We present an alternate solution to this problem that beats the baseline model obtained by the OpenEDS team. Our solution received a score of 0.93 on the leaderboard.

1. Introduction

This is our submission for the OpenEDS challenge hosted by Facebook on the CloudCV platform. AR/VR applications require accurate and precise eye tracking solutions. In this paper we explore one method of achieving this, namely semantic segmentation of key regions of the human eye.

2. The OpenEDS dataset

The dataset used in this paper has been created by and released as a workshop challenge at ICCV by Facebook. It has been pre-split into 3 sets, consisting of 8916 train, 2403 validation and 1440 test images respectively. The labels have been included for the train and validation sets.

3. Methodology

3.1. Pipelined encoder-decoder model with skip connections

We have used an encoder-decoder network based on the SegNet architeture to classify different key regions of the eye. The network classifies each pixel of an input image into one of four classes; iris, pupil, sclera and outside regions. Our solution posits a two-fold implementations of encoder-decoder networks to classify each pixel of an input image. The first network is a base model that predicts 3 classes; the sclera, iris and ouside regions, as shown in figure 1 and figure 2. The second model predicts the pixel forming the pupil inside the iris as shown in figure 3 and figure 4. Both models use skip connections that are interlaced between the convolution layers. Results of experimentation with both additive as well as multiplicative skip connections showed that additive skip connections outperformed the multiplicative ones.

The final segmentation output is obtained by classifying each pixel into 3 categories using the base model and identifying the pupils region using the pupil model.

3.2. The architecture

We have implemented an encoder decoder architecture, with 148,739 model parameters for the base model and 32,000 model parameters for the pupil model.

3.3. Post processing

As seen in Figure1. The predicted output of the stitched neural network produces a significant amount of extraneous detection. We have opted to use image processing tech-
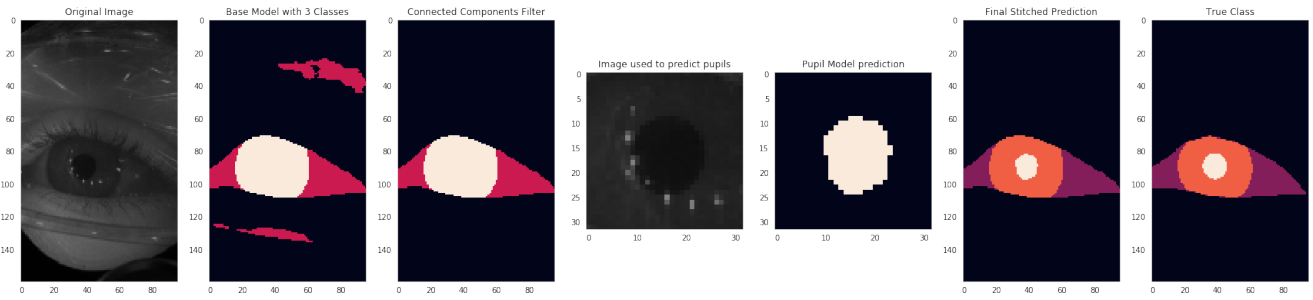
Figure 1: Base-model first predicts the sclera and iris then the pupils-model detects the pupil
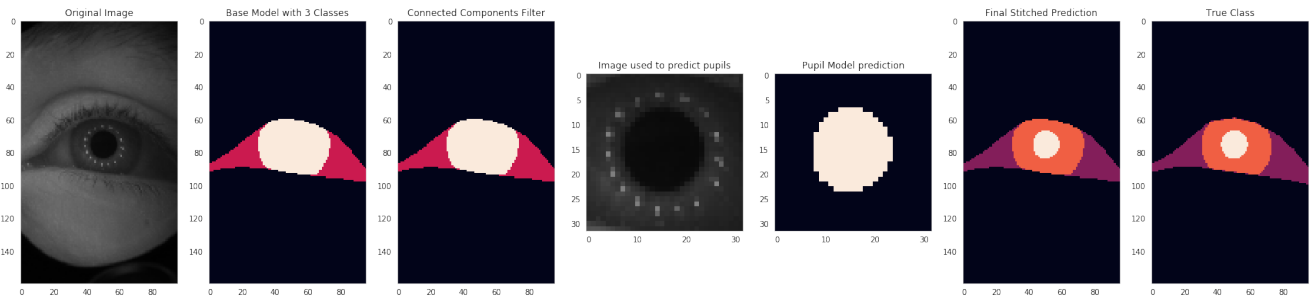

Figure 2: A good prediction

niques in order to detect and remove these erroneous detection. We have used connected components to find extraneous detection in the predicted output of the base model,and eliminate them based on their weighted area. before extracting the smaller image for predicting pupils.
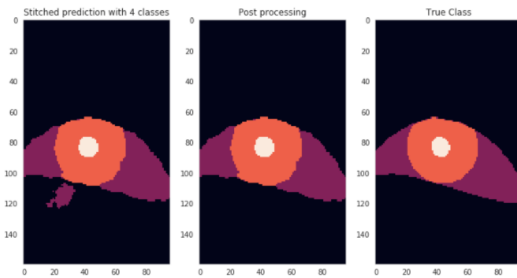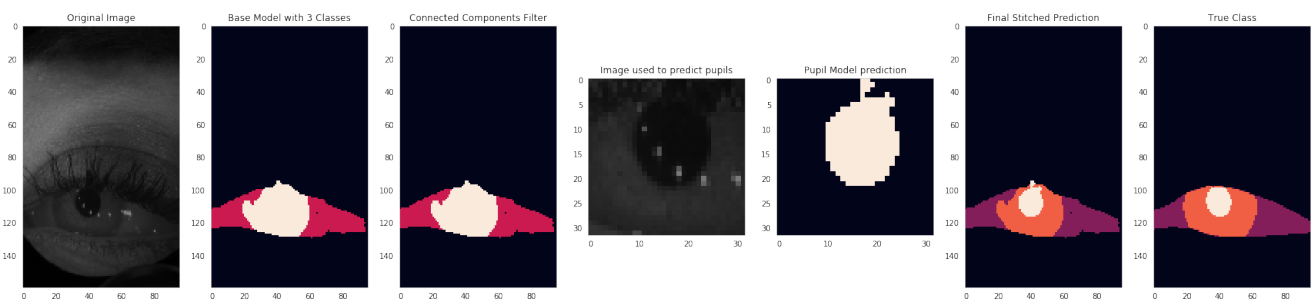

Figure 4: Post processing output

4. Results

5. Conclusion

References

Figure 3: A bad prediction