EXPLORING THE INFLUENCE OF FACIAL FEATURES IN DEEP LEARNING BASED GAZE ESTIMATION

Authors
TAN M. NGUYEN
nguyenmanhtan@student.tudelft.nl
+(31) 613917175

Affiliations
DELFT UNIVERSITY OF TECHNOLOGY

01 Introduction

Gaze estimation holds significant importance in various applications. Pioneering research [1] has demonstrated state-of-the-art performance in gaze estimation models by utilizing deep Convolutional Neural Networks (CNNs) and incorporating full facial images as input, instead of or in addition to solely using one or both eye images. Facial images encode crucial cues that can enhance the accuracy of gaze regression models. However, it remains unclear which specific facial features contribute and to what extent they contribute to the overall estimation accuracy. In this research, we aim to shed light on identifying the influential facial regions and quantifying their contributions to gaze estimation accuracy.

02 Objective

The objective of this study aims to answer the question of which and to what extent different facial landmarks contribute to the performance of gaze estimation by conducting an investigation into the significant regions of the face and quantifying their contribution toward their impact on the accuracy of gaze estimation

03 Methodology

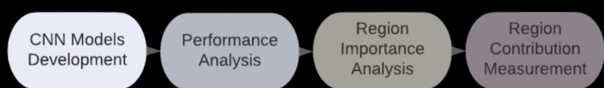

Figure 1: The research methodology employed to investigate the influence of different facial regions on eye gaze estimation.

04 Model Architectures

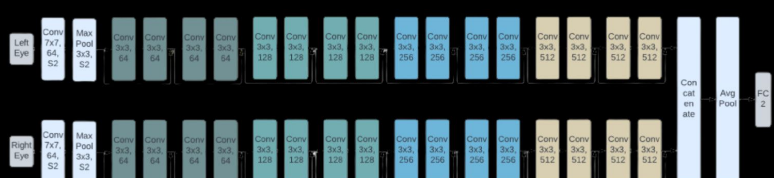

Figure 2: CNN Architecture for Eye baseline model based on ResNet18 Convolutional backbone.

Figure 3: CNN Architecture for full face input model based on ResNet18 Convolutional backbone

05 Performance Analysis Results

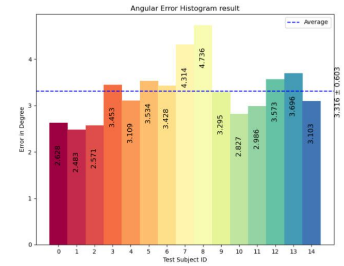

Figure 4: Cross-Validation results for the baseline eye-only model over 15 MPIIFaceGaze test subjects.

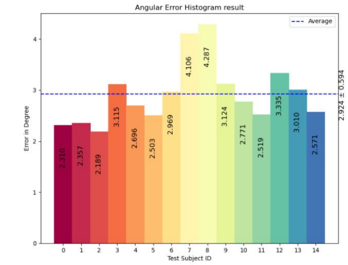

Figure 5: Cross-Validation results for the full face model over 15 MPIIFaceGaze test subjects

06 Facial Contribution Results

Figure 6: Region Importance Analysis sliding windows applied to the first image of test subject 14. Shown from left to right: (16,32), (192,144), (208,48), (208,208), and (64,16) positions with the box filter implemented [2].

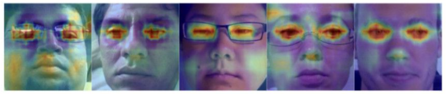

Figure 7: Regional importance heat maps for test subject ids 5, 6, 7, 8, and 9 respectively from left to right.

07 Region Contribution Results

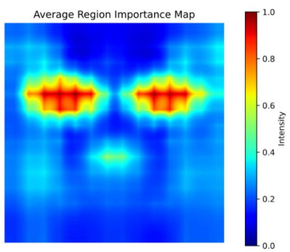

Figure 8: Unweighted Averaged Region Importance Heatmap computed with all 15 test subjects' heatmaps

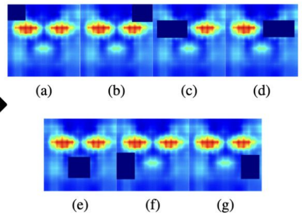

Figure 9: The identified regions contribute the most to the overall accuracy of gaze estimation. The regions are highlighted by a dark rectangular shape

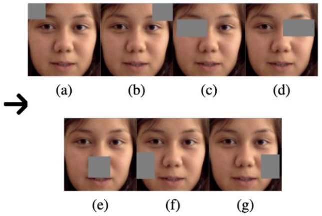

Figure 10: From left to right, the figure shows examples of occluded facial regions: right eyebrow, left eye-brow, right eye, left eye, nose, right cheek, and left cheek for the first image of test subject 14.

Occluded Region	Averaged Angular Error (Deg)
Right Eyebrow	3.007 ± 0.604
Left Eyebrow	2.973 ± 0.649
Right Eye	3.502 ± 0.754
Left Eye	3.518 ± 0.720
Nose	3.033 ± 0.626
Right Cheek	3.000 ± 0.628
Left Cheek	2.968 ± 0.614

Table 1: Averaged angular error evaluated for each occluded facial region in degree

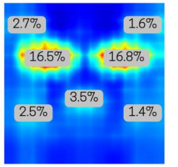

Figure 10: Relative Error Contribution for different facial landmarks. Shown from left to right and top to bottom: right eyebrow, left eyebrow, right eye, left eye, nose, right cheek, and left cheek

08 Conclusion

The studied experiments revealed improved performance when utilizing full facial input, up to 11% in relative error, in comparison to the baseline method of using eye regions. In addition, this work identified 7 different facial landmarks that contribute to the higher accuracy: right eyebrow (2.7%), left eyebrow (1.6%), right eye (16.5%), left eye (16.8), nose (3.5), right cheek (2.5%), and left cheek (1.4%).

09 Limitations

It is important to acknowledge that while the performed experiments capture the independent important facial regions, they do not show the correlation impacts. In future work, we aim to deploy combination feature analysis to gain insight into how combined facial features impact accuracy.

References

[1] Xucong Zhang, Yusuke Sugano, Mario Fritz, and Andreas Bulling. It's written all over your face: Full-face appearance-based gaze estimation. In Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pages 2299–2308, 2017.

[2] Matthew D. Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In European Conference on Computer Vision (ECCV), pages 818–833, 2014