Detection and Correspondence Matching of Corneal Reflections for Eye Tracking Using Deep Learning

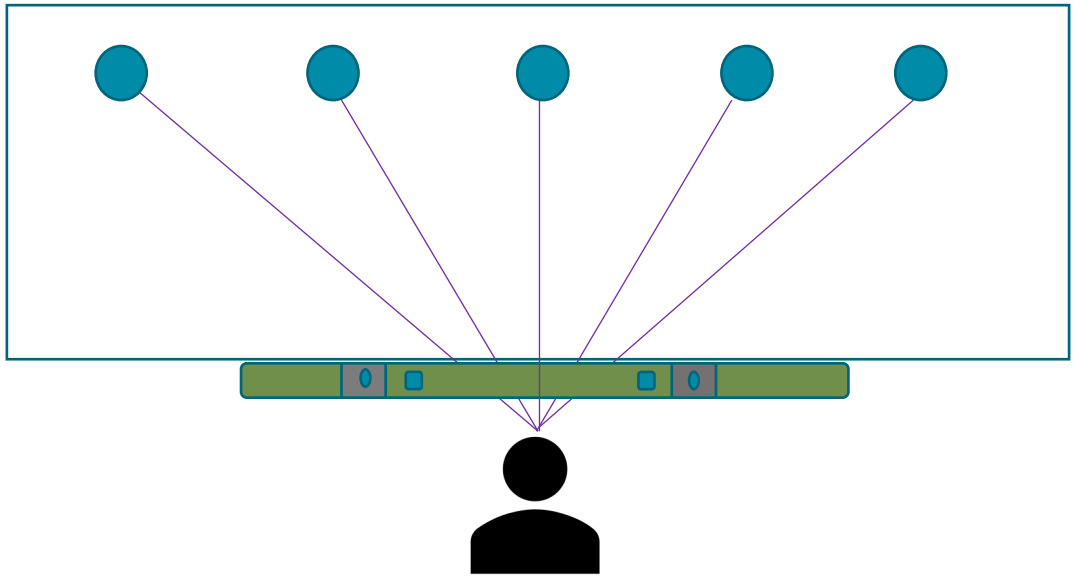
Soumil Chugh, Braiden Brousseou, Jonathan Rose and Moshe Eizenman

Electrical and Computer Engineering Department, University of Toronto,

Toronto, Canada

EYE TRACKING

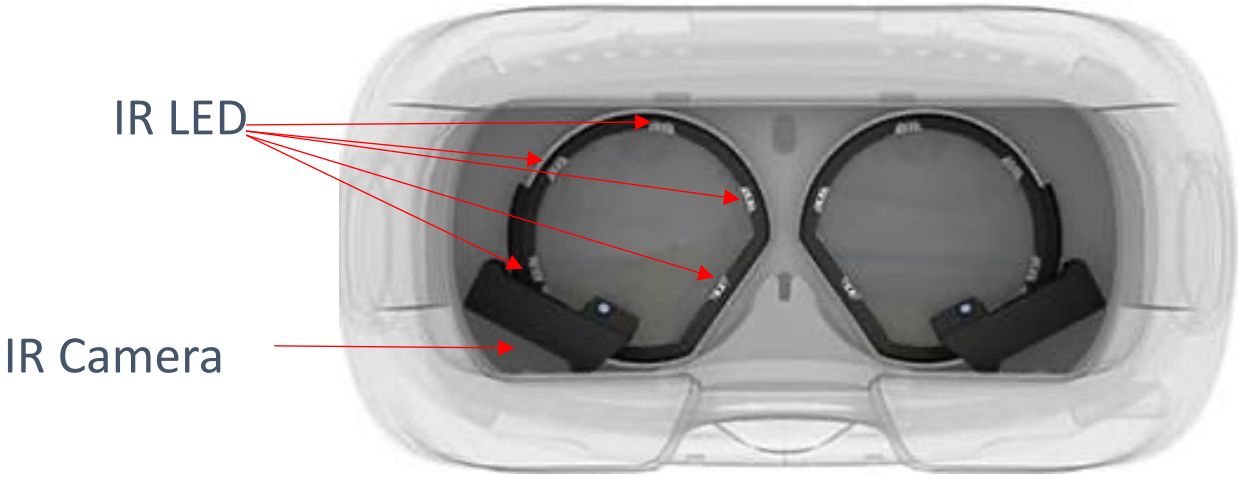
- Ability of computer to determine where person looks on
  - Display Screen
  - Real World



UNIVERSITY OF TORONTO

EYE TRACKING SYSTEM

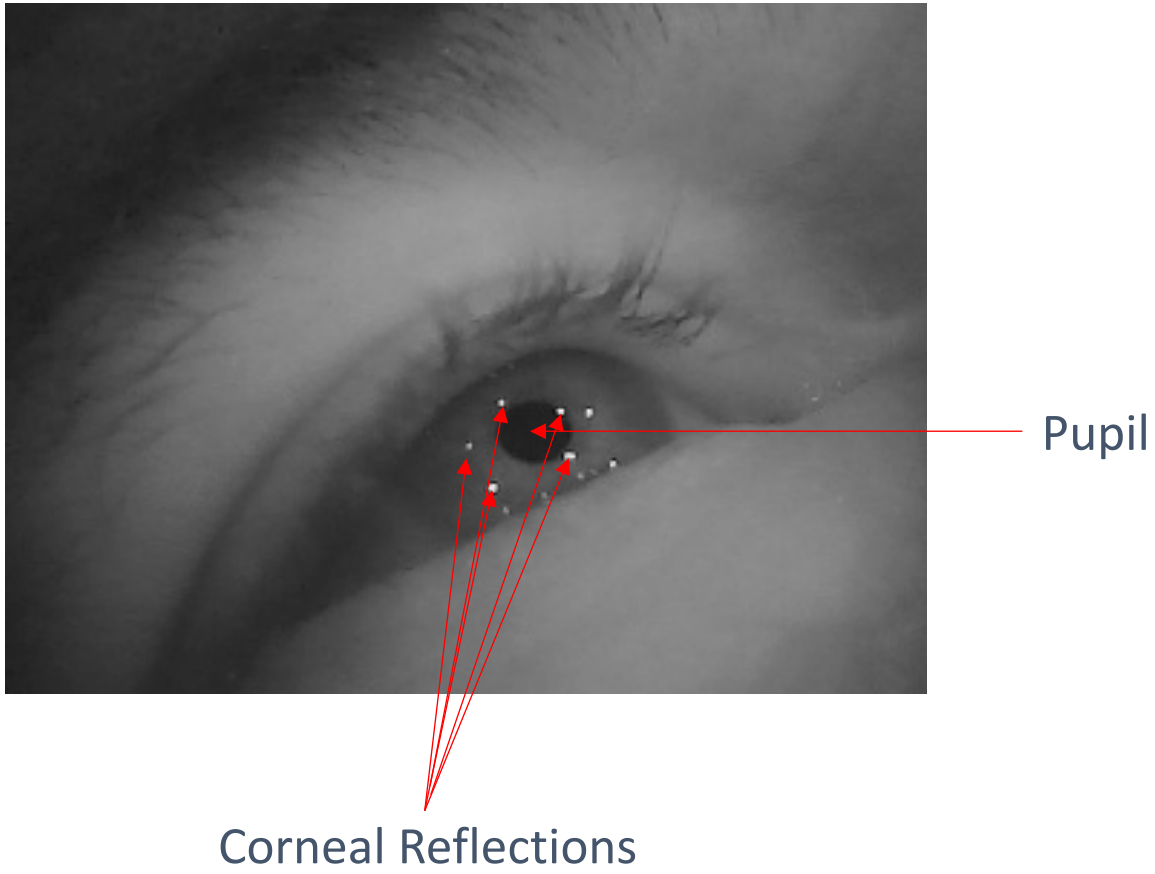
- 5 IR Light sources and an IR camera per eye
  - Manufactured by Pupil Labs [1]



HTC VR Headset

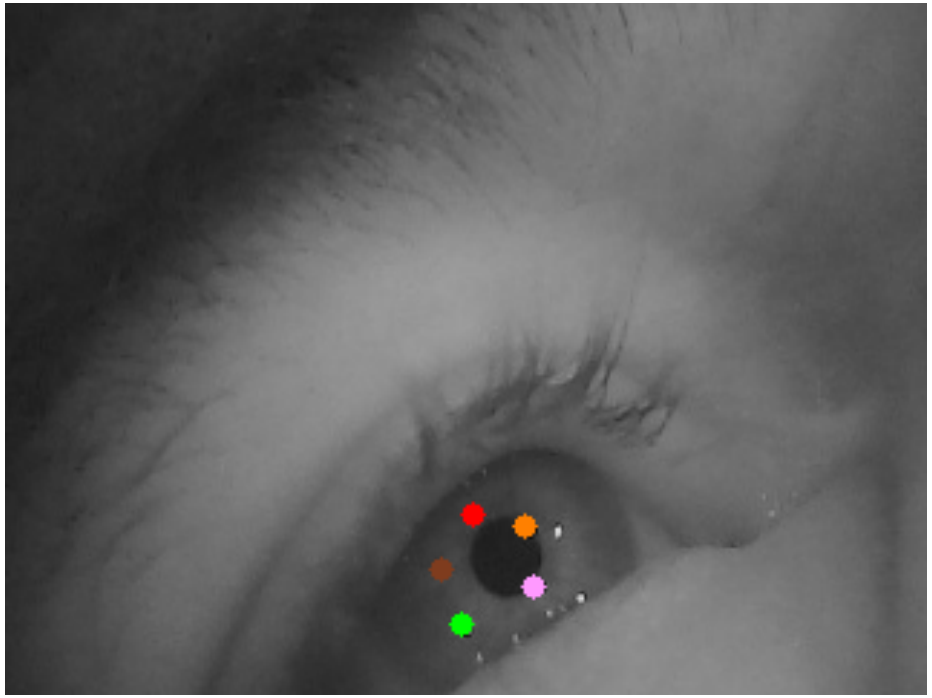
[1] M. Kassner, W. Patera and A. Bulling, “Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction,” In Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing: Adjunct publication, Seattle, Washington, USA, pp. 1151-1160, 2014.

OUTPUT OF EYE TRACKING HARDWARE MODULE



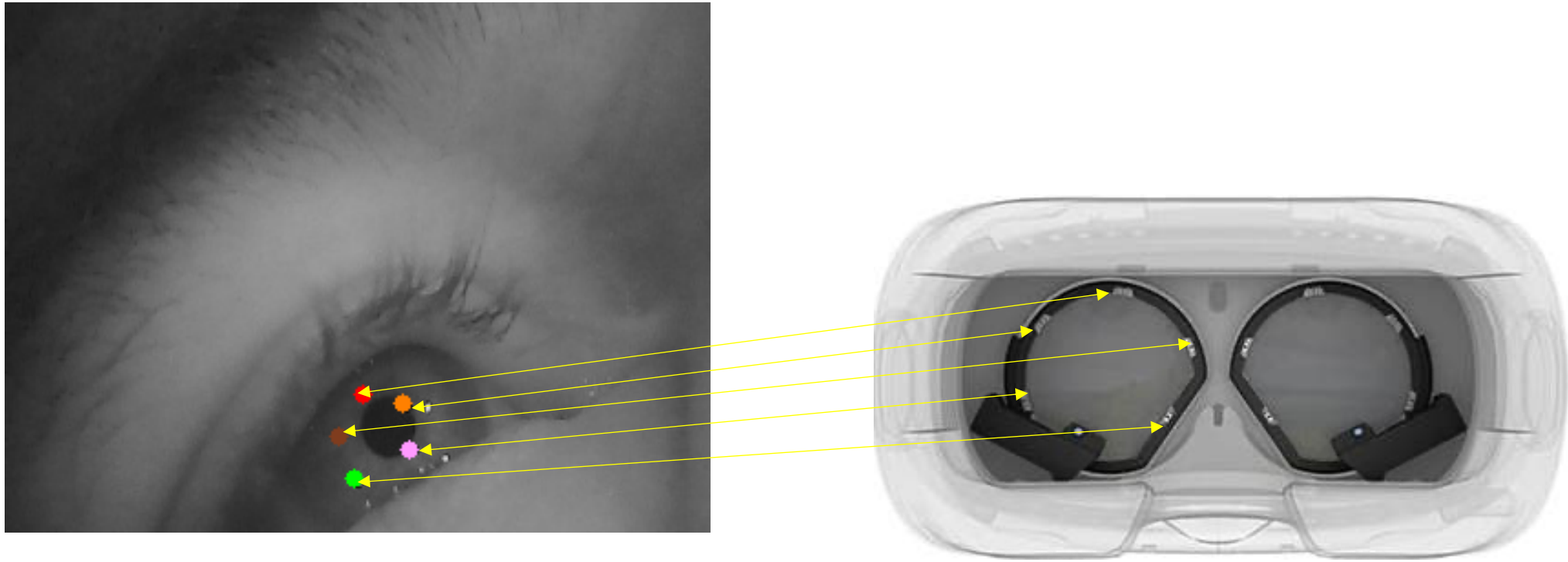
UNIVERSITY OF TORONTO

REQUIREMENT 1 – CORNEAL REFLECTION LOCATIONS



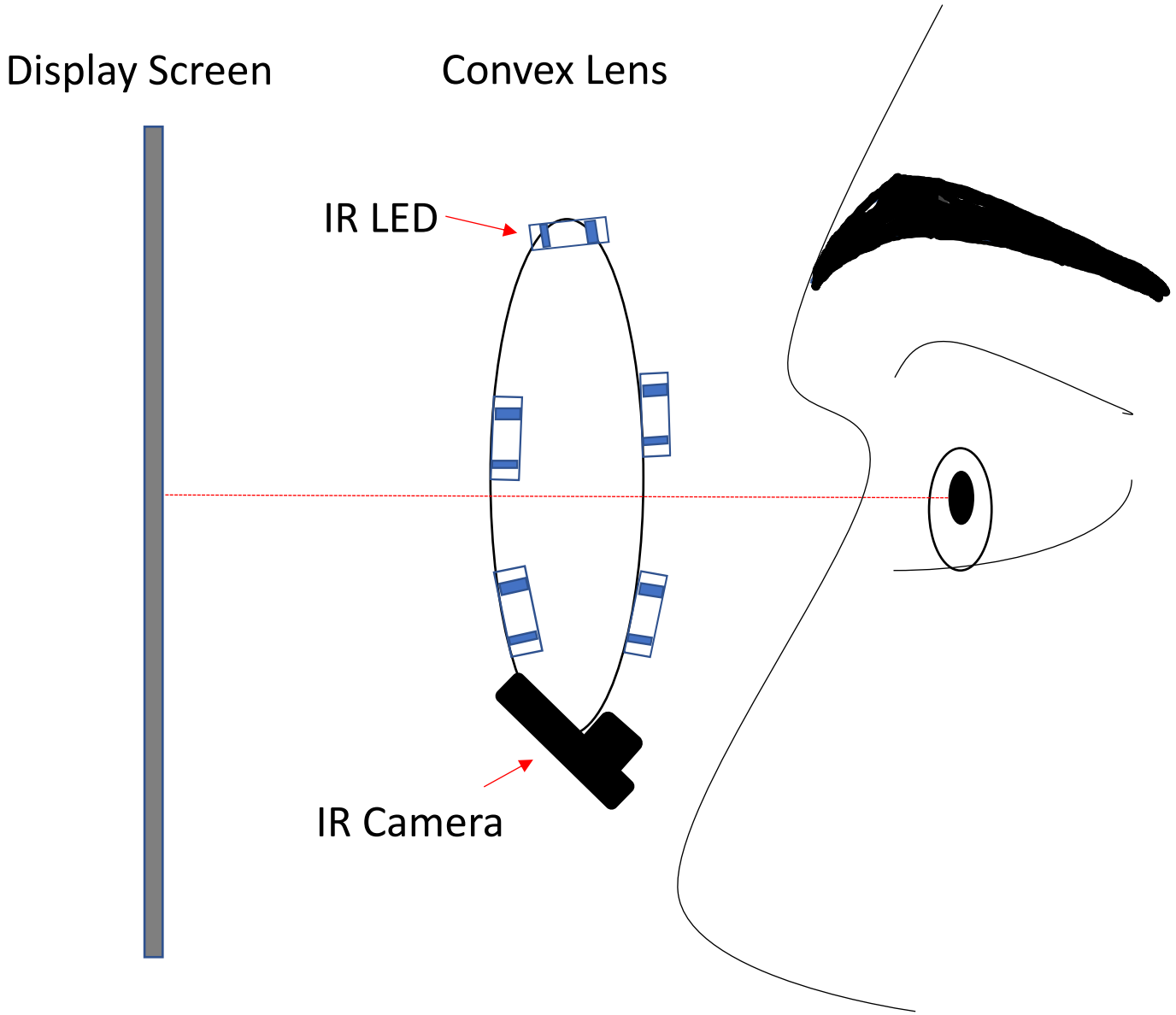
UNIVERSITY OF TORONTO

REQUIREMENT 2 – CORNEAL REFLECTION LED CORRESPONDENCE



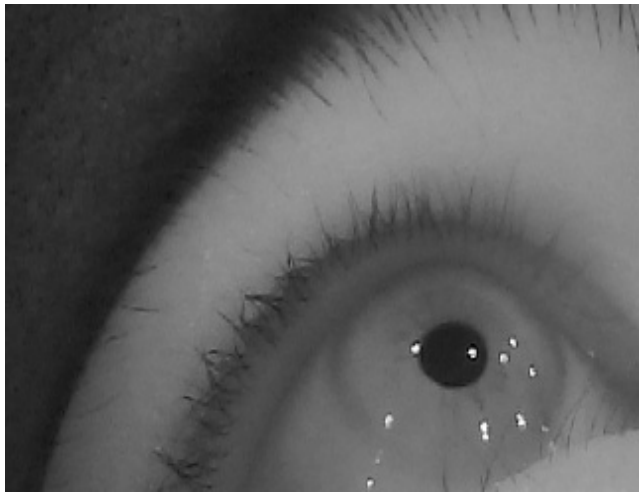
UNIVERSITY OF TORONTO

CHALLENGES OF VR BASED EYE TRACKING

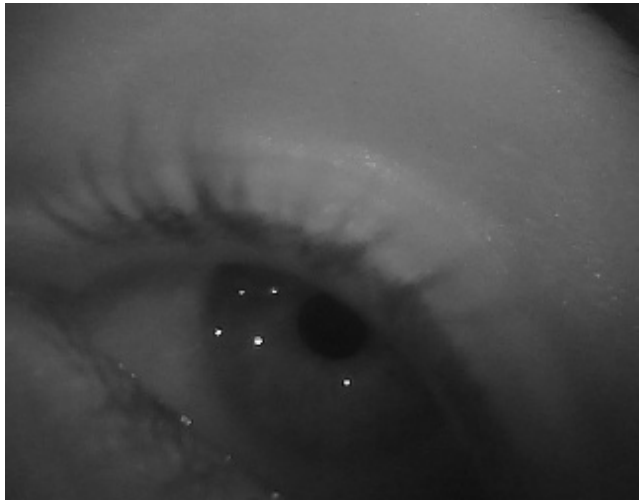


UNIVERSITY OF TORONTO

CHALLENGING SCENARIOS FOR CORNEAL REFLECTION DETECTION AND MATCHING



Natural Spurious Reflections



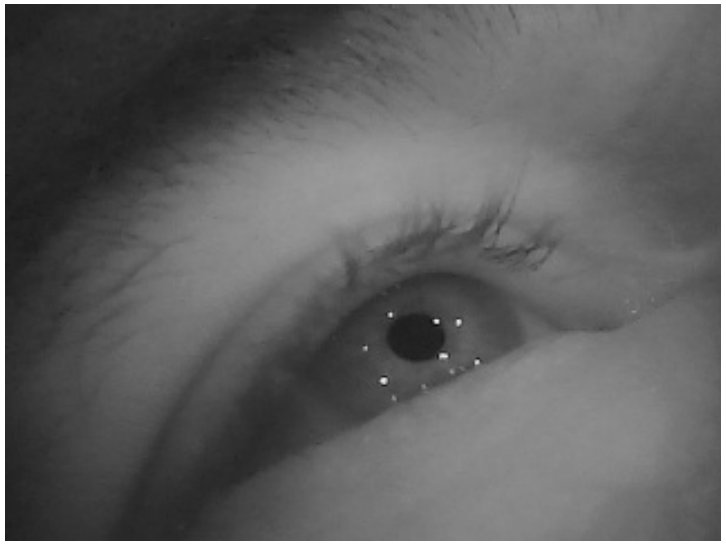
Missing Corneal Reflections due to Eyelid Occlusion

Corneal Reflections Appear/Disappear with Change in Gaze Angle

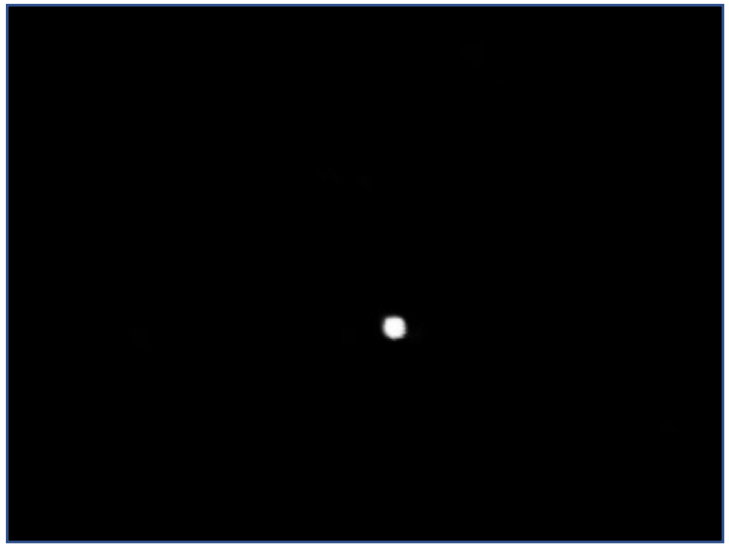
UNIVERSITY OF TORONTO

DEEP LEARNING-BASED SOLUTION

- Goal - Find regions in the image that belong to each of five corneal reflections
- Pixel Wise Classification also known as Semantic Segmentation.
- Input – Image of the eye
- Output – 5 probability maps corresponding to five corneal reflections



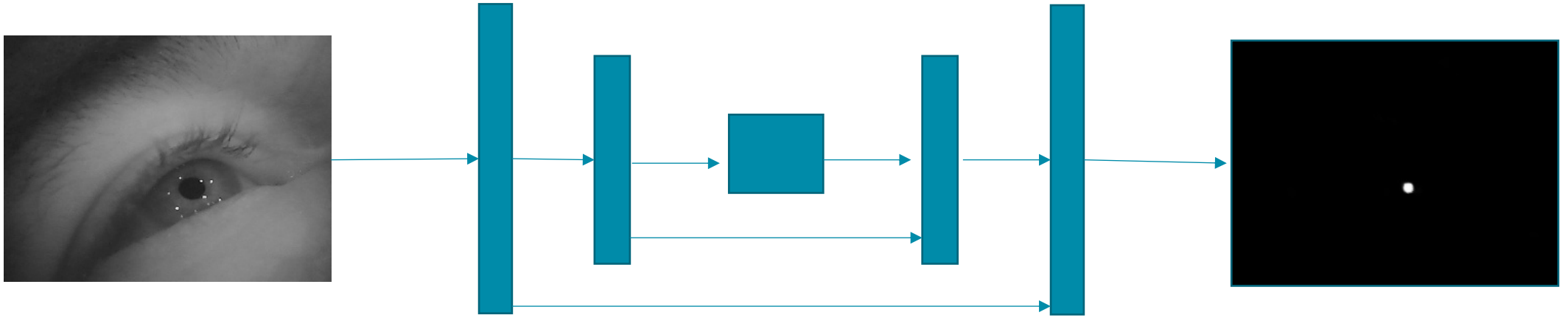
Input



Output

UNET ARCHITECTURE

- Input and output share same size
- Consists of only Convolutional Layers
- Symmetric in nature

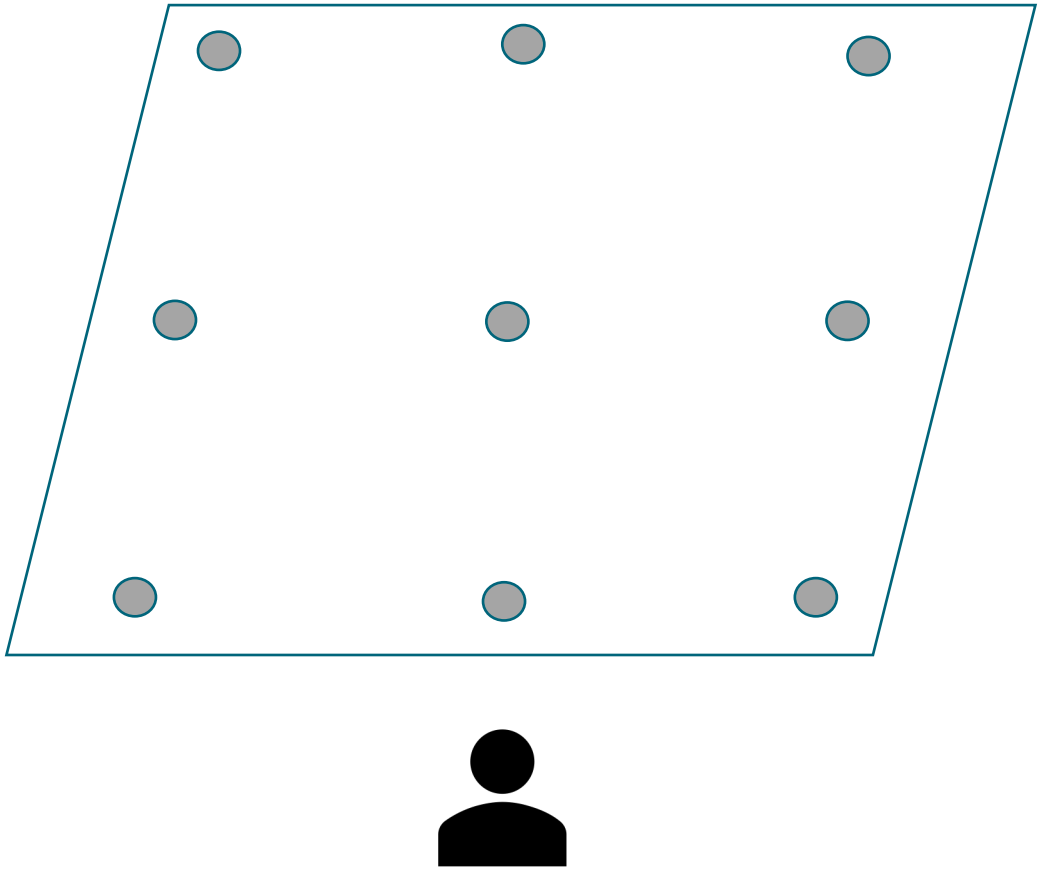


Input

Output

DATA COLLECTION FOR TRAINING

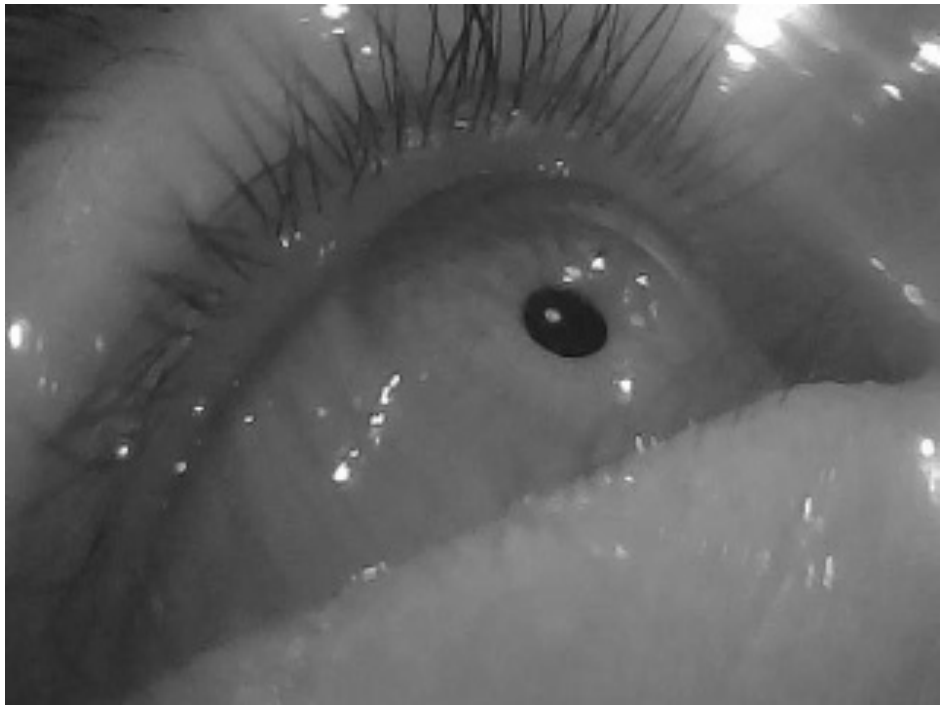
- Left and Right Eye Images from 15 people

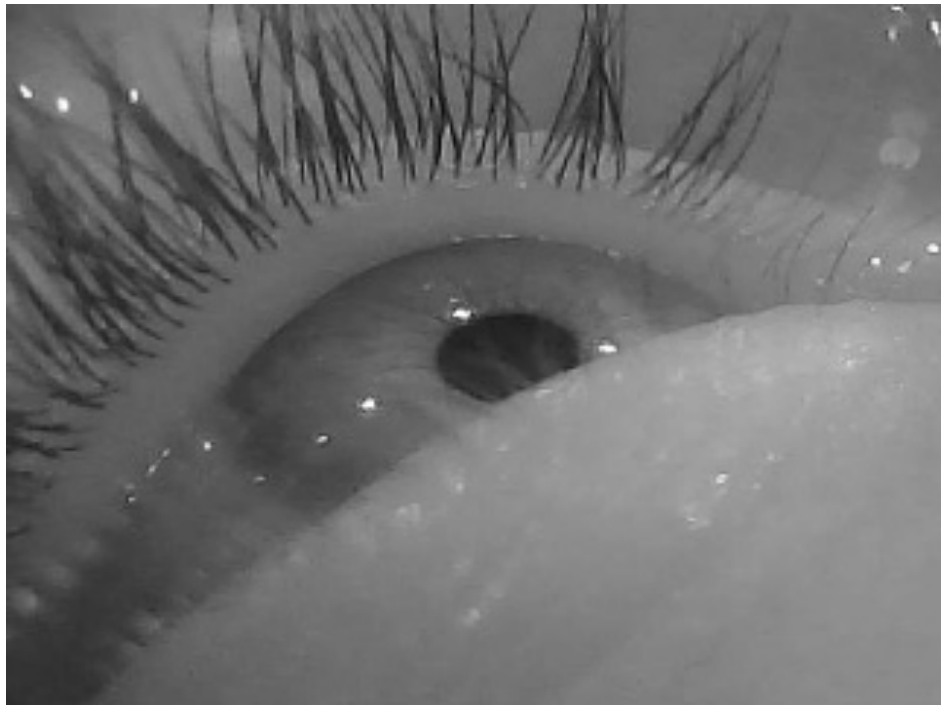


UNIVERSITY OF TORONTO

ANOTHER SOURCE OF DATA

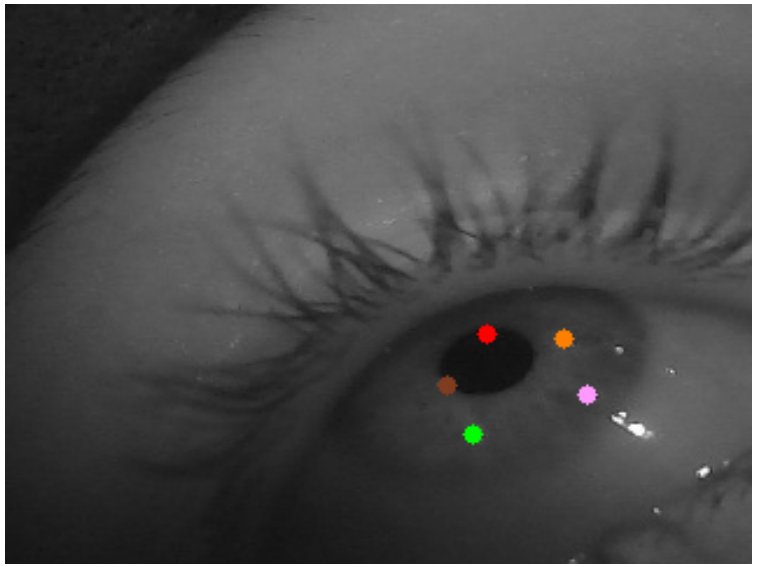
- NVIDIA Dataset
  - Collected using same hardware on 10 people



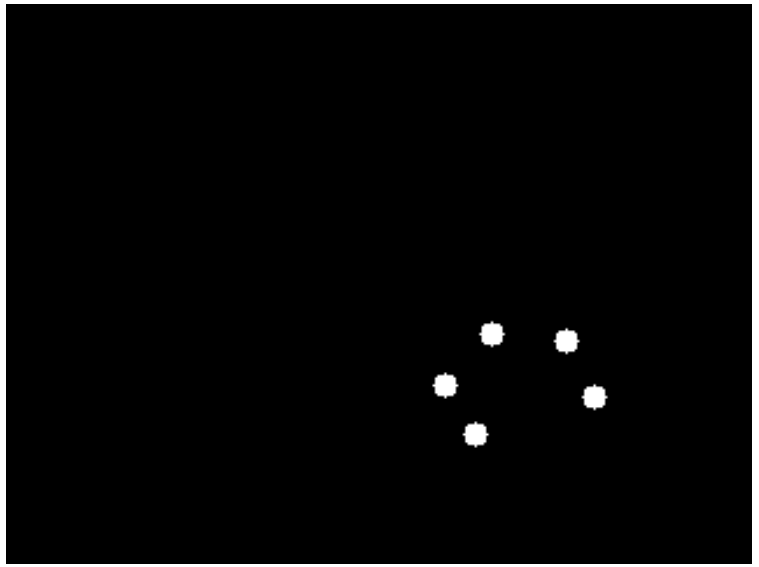


DATA LABELLING

- Manually Labelled 4000 eye images, 25 people
- Binary Mask consist of:
  - Corneal reflections- Circle of radius 5 pixels around the final center



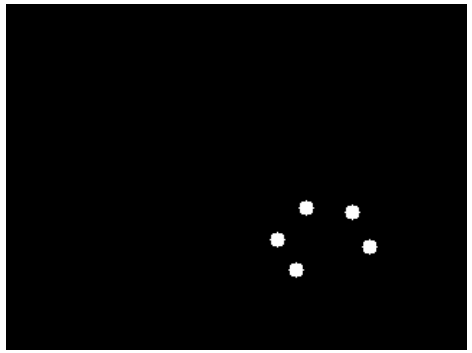
Labelled Input Eye Image



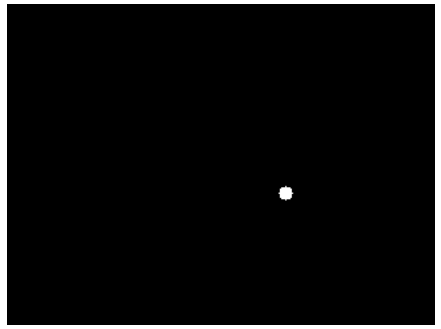
Binary Mask

UNIVERSITY OF TORONTO

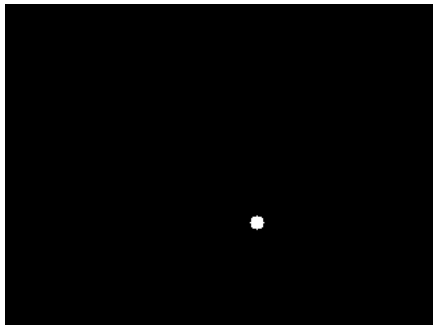
LABELS FOR TRAINING



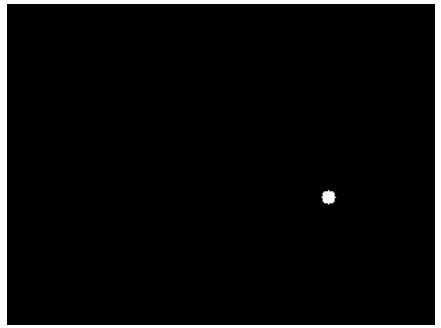
Binary Mask



Light Source 1

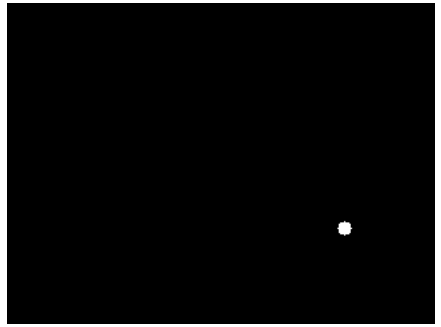


Light Source 5



Light Source 2

Light Source 4

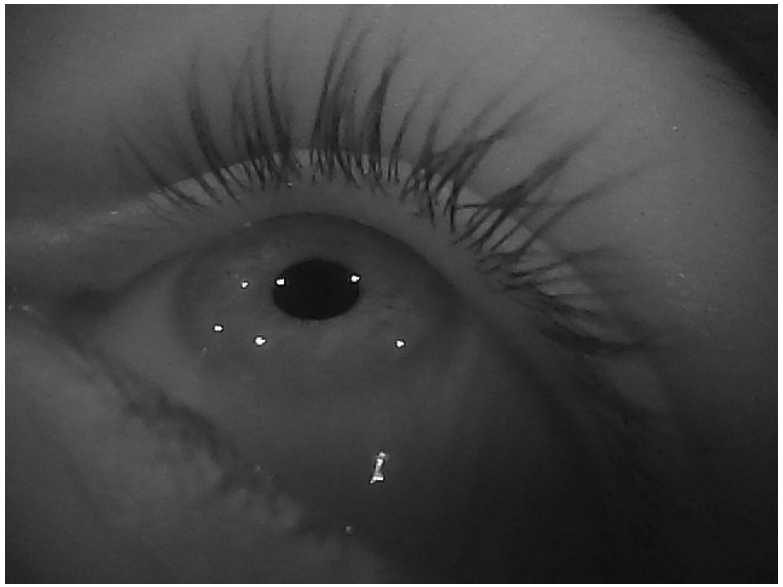


Light Source 5

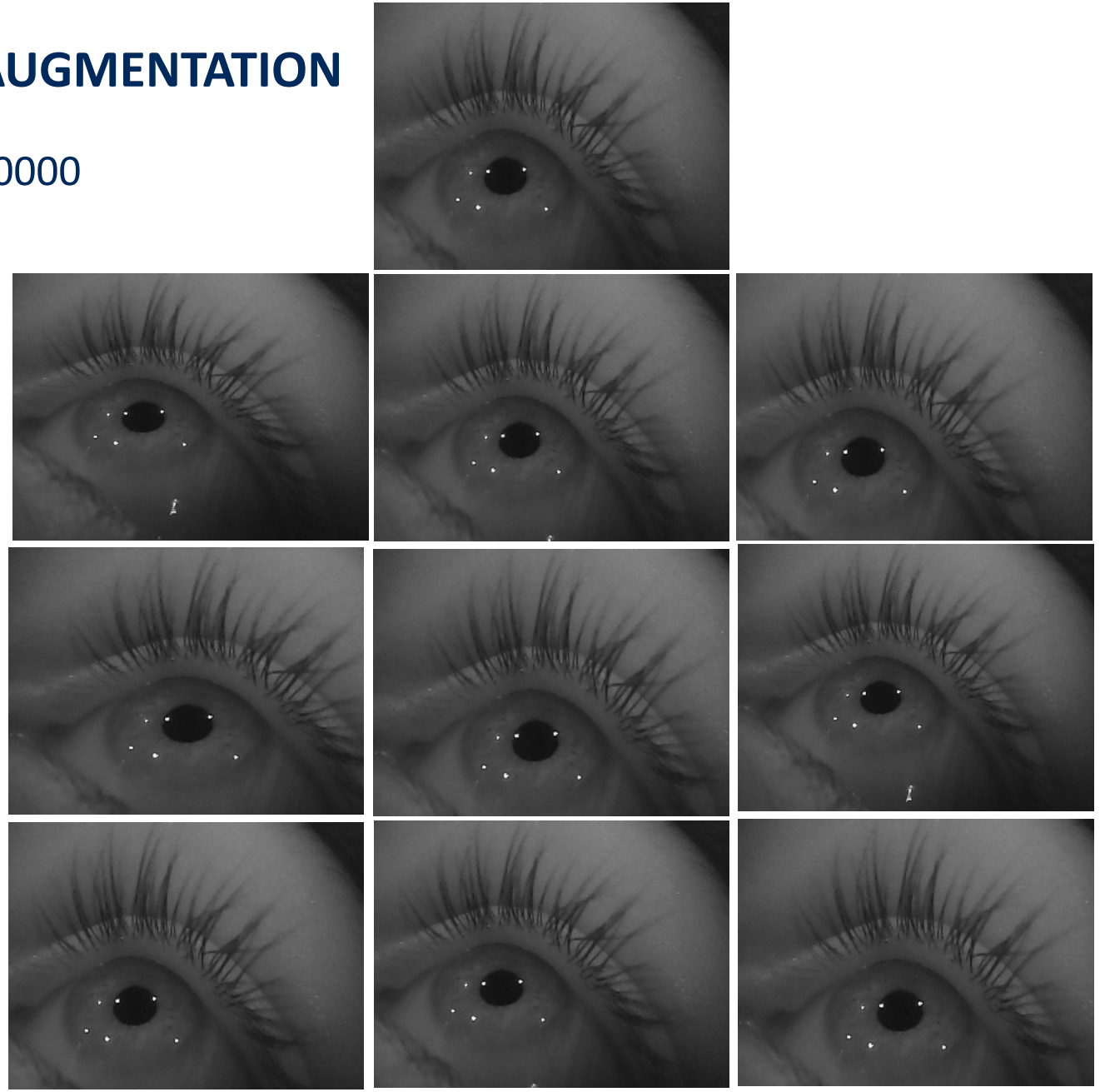
UNIVERSITY OF TORONTO

STATIC DATASET AUGMENTATION

4000 ---> 40000

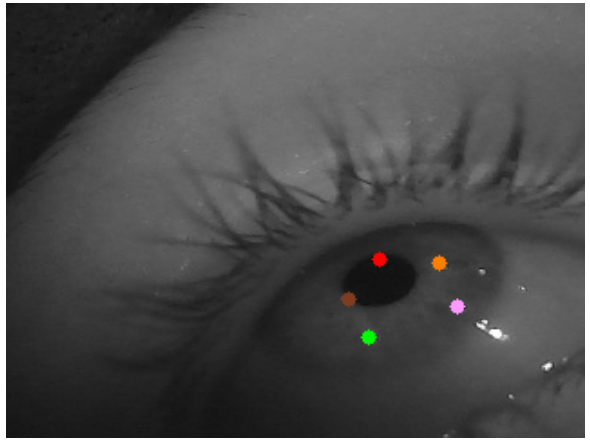


640x480

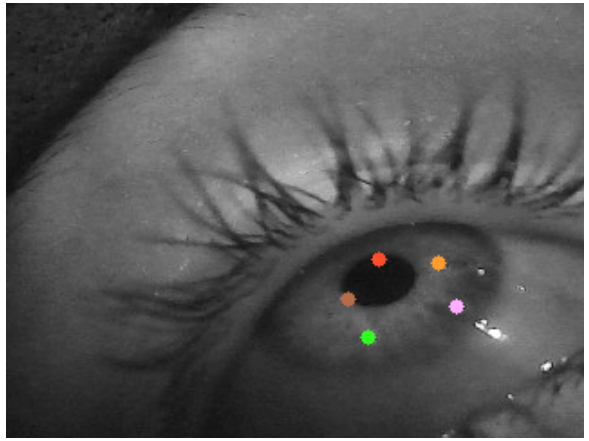


UNIVERSITY OF TORONTO

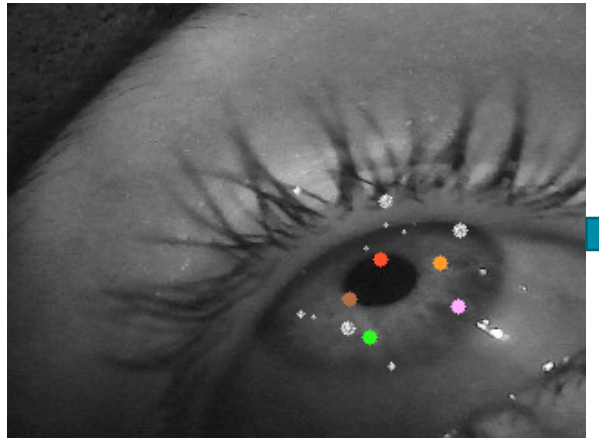
DYNAMIC DATA AUGMENTATION



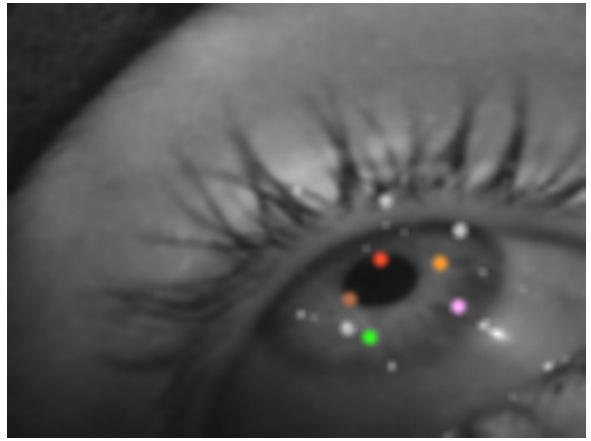
Original Image



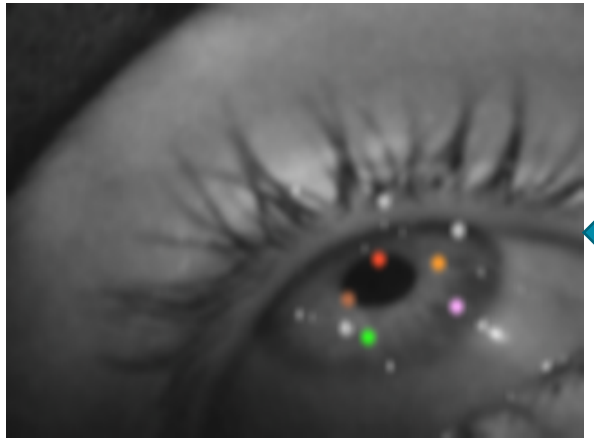
Adjust Contrast and Brightness



Add Spurious Reflections



Gaussian Blurring



Motion Blurring

PERFORMANCE METRICS

Accuracy

Fraction of true corneal reflections detected and matched

Inference Time

Forward pass through neural network on RTX2060 GPU

UNIVERSITY OF TORONTO

SINGLE CORNEAL REFLECTION PERFORMANCE

Method	Accuracy (%)	Inference Time (ms)
Our System	91	5.7
EyeNet [1]	96	60.0

[1] Z.Wu, S.Rajendran, T.Van As, V.Badrinarayanan and A.Rabinovich, “EyeNet: A Multi-Task Deep Network for Off-Axis Eye Gaze Estimation,” In 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), pp. 3683-3687, October 2019.

EYE TRACKING PERFORMANCE

Metrics	NvGaze [1]	EyeNet [2]	Our System
Mean Accuracy (°)	**2.1**	3.0	**1.1**

[1] J. Kim, M. Stengel, A. Majercik, S. De Mello, D. Dunn, S.Laine, M. McGuire, and D. Luebke, "Nvgaze: An anatomically-informed dataset for low-latency, near-eye gaze estimation," InProceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Scotland, pp. 1-12, May 2019.

[2] Z.Wu, S.Rajendran, T.Van As, V.Badrinarayanan and A.Rabinovich, "EyeNet: A Multi-Task Deep Network for Off-Axis Eye Gaze Estimation," In 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), pp. 3683-3687, October 2019.

CONCLUSION

- Accurate corneal reflection detection and matching is challenging in XR systems.
- We propose to solve this problem using a deep learning-based solution.
- Our algorithm reports accuracy of over 90% which is comparable with the only prior deep learning-based technique.
- However, our model runs 10x faster and occupies 33x less space in memory.
- When integrated with the eye tracking algorithm in a VR headset, our system exhibits accuracy of 1°.
- This performance 100% better than current XR based eye tracking systems.