# ISMAR 2021: Edge-Guided Near-Eye Image Analysis for Head Mounted Displays

Zhimin Wang[1], Yuxin Zhao[1], Yunfei Liu[1], Feng Lu[1,2]

1 State Key Laboratory of VR Technology and Systems, School of CSE, Beihang University

2 Peng Cheng Laboratory, Shenzhen, China

ISMAR 2021
OCTOBER 4-8 BARI - ITALY

IEEE COMPUTER SOCIETY
IEEE Advancing Technology for Humanity

# Authors




Zhimin Wang

Yuxin Zhao

Yunfei Liu

Feng Lu

# Outline

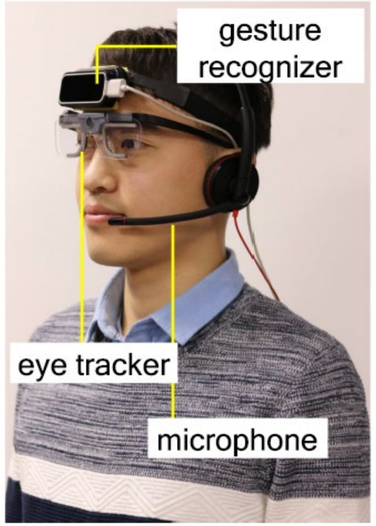
- Background
- Related Work
- Our Method
- Experiments
- Future Work

# Background

Eye tracking has been used in many applications.

Foveated rendering
[Meng et al. 2020]


Redirected Walking
[Langbehn et al. 2018]




Gaze-based Interaction
[Wang et al. 2020]

Behavior Analysis
[Lang et al. 2018]

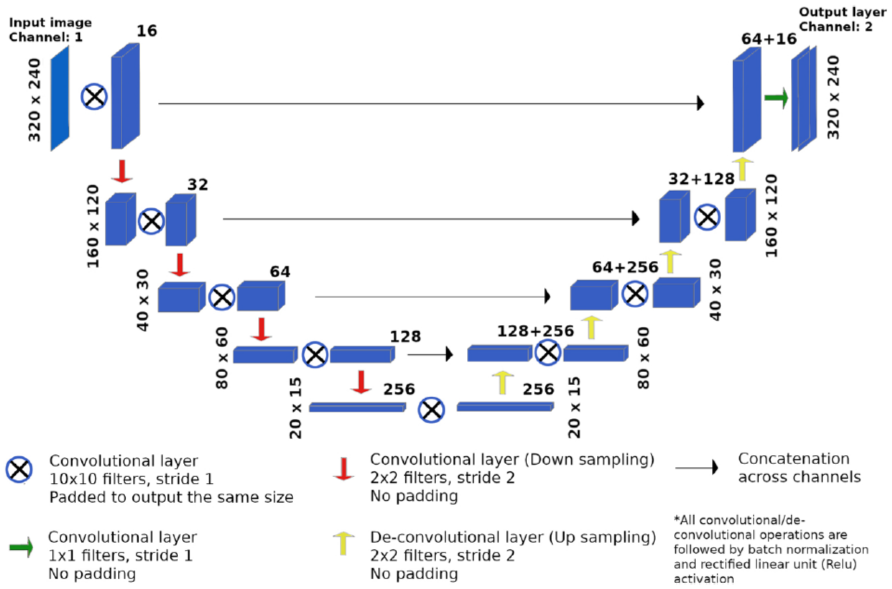
# Related Work

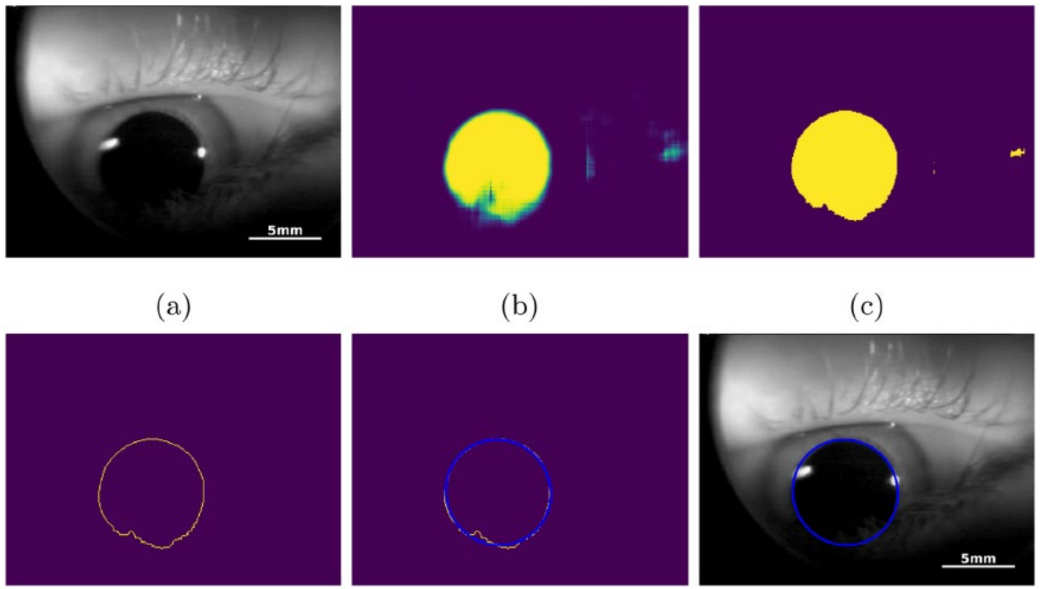
Eye tracking methods need to compute gaze-relative features, from infrared (IR) eye images.

| Gaze-relative Features | Methods |
|---|---|
| Pupil Center | Guestrin et al. 2014, NVGaze, Lu et al. 2020 |
| Pupil Ellipse | Else, Pure, DeepVOG, RITNet, EllSeg |
| Iris Ellipse | Hansen et al. 2005, RITNet, EllSeg |

# Related Work

- DeepVOG: uses the U-Net to segment out the pupil area.
- Fit an ellipse on the segmentation map.

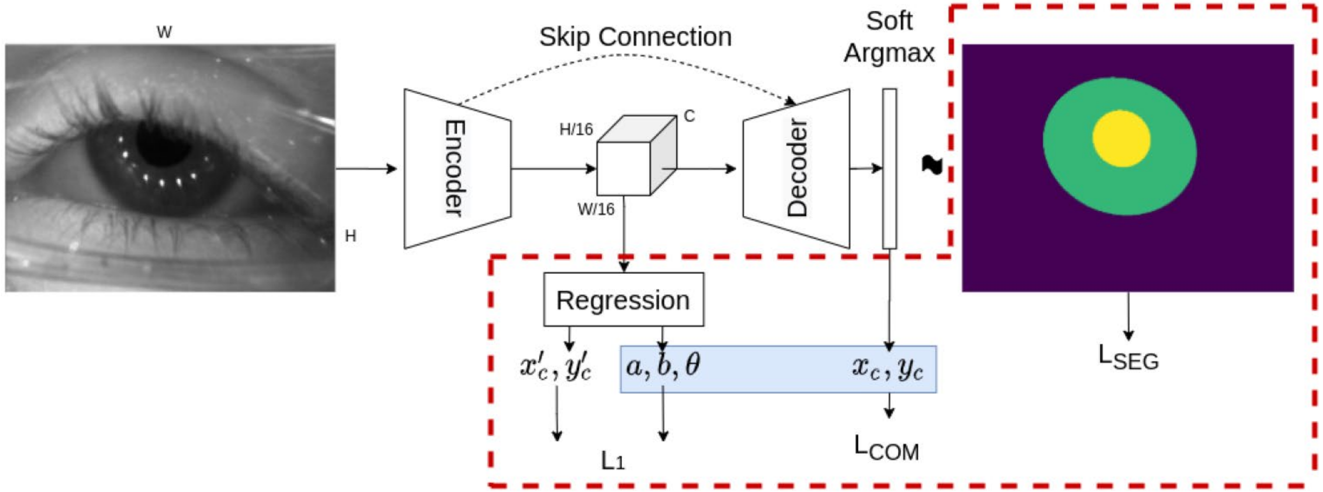



Y.-H. Yiu, M. Aboulatta, T. Raiser, L. Ophey, V. L. Flanagin, P. Zu Eulenburg, and S.-A. Ahmadi. DeepVOG: Open-source pupil segmentation and gaze estimation in neuroscience using deep learning. Journal of neuroscience methods, 324:108307, 2019.

# Related Work

- EllSeg: segments out complete pupil and iris structures
- Robust to occlusions
- Improved ellipse estimates as compared to segmenting eye parts




R. S. Kothari, A. K. Chaudhary, R. J. Bailey, J. B. Pelz, and G. J. Diaz. Ellseg: An ellipse segmentation framework for robust gaze tracking. IEEE TVCG, 27(5):2757–2767, 2021.

# Our Method

## Motivation

- Most discriminate information in the eye image is encoded in certain edge areas, including two eyelids, pupil contour and iris contour. We call it task-related edges.
- The intersection of these contours exactly corresponds to the ellipse of the pupil and iris.

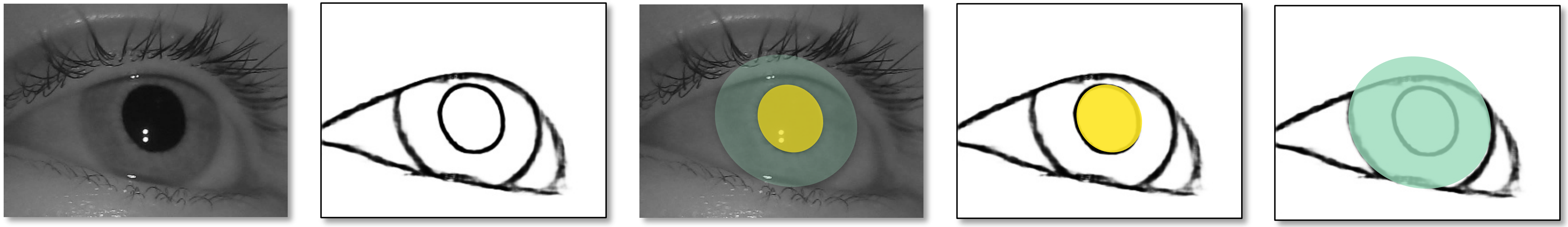
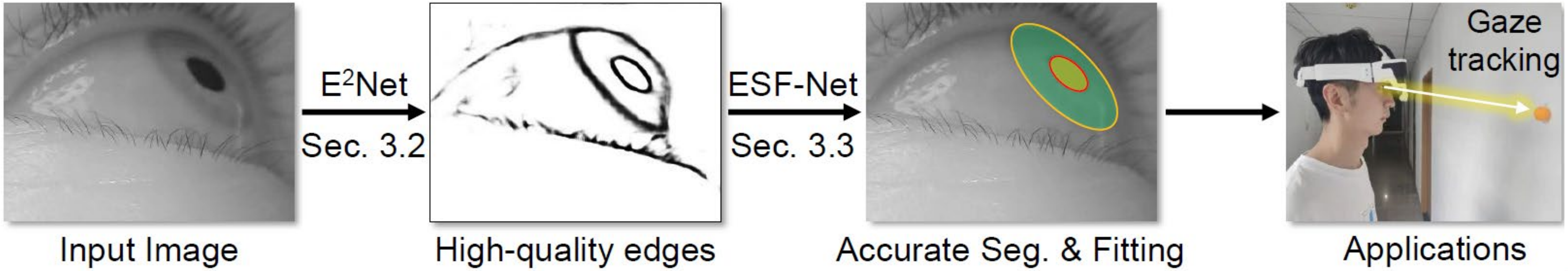


Image | Task-related Edges | Ground Truth | Pupil | Iris

# Our Method

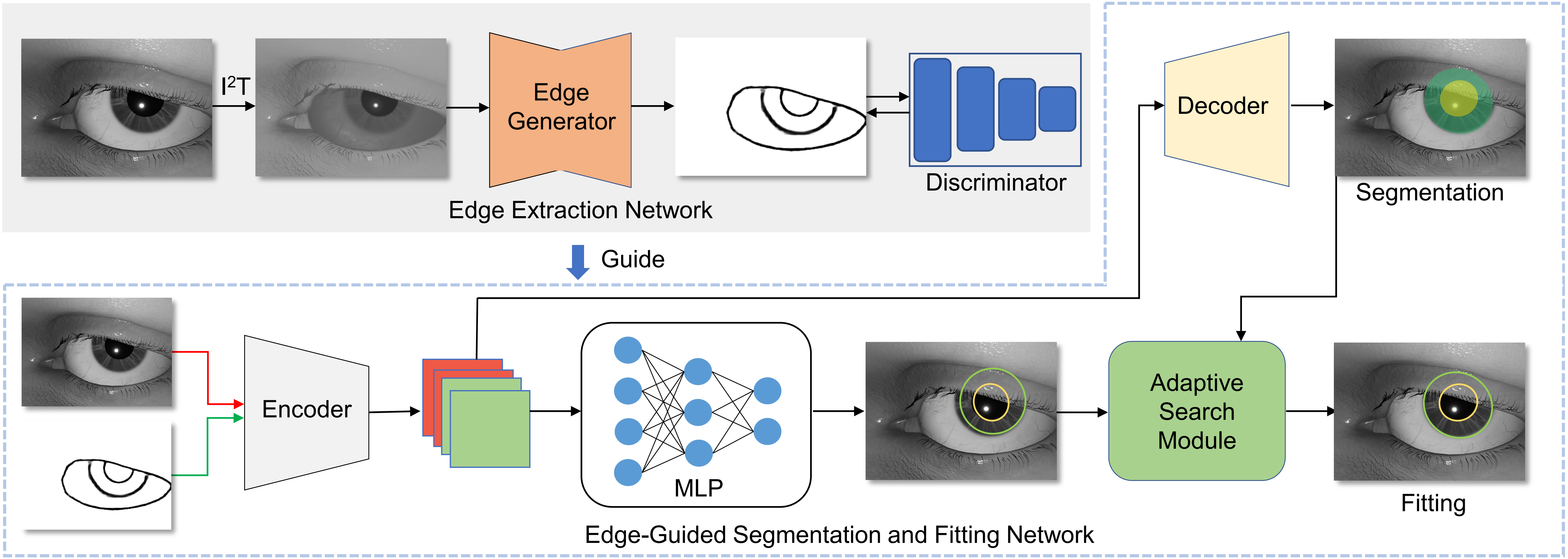
## Overview

- We propose a novel near-eye image analysis method with edge maps as guidance.
- We first utilize an Edge Extraction Network (E2-Net) to predict high-quality edge maps, which only contain eyelids and iris/pupil contours without other undesired edges.
- Then we feed the edge maps into an Edge-Guided Segmentation and Fitting Network (ESF-Net) for accurate segmentation and ellipse fitting.



# Our Method

## Pipeline

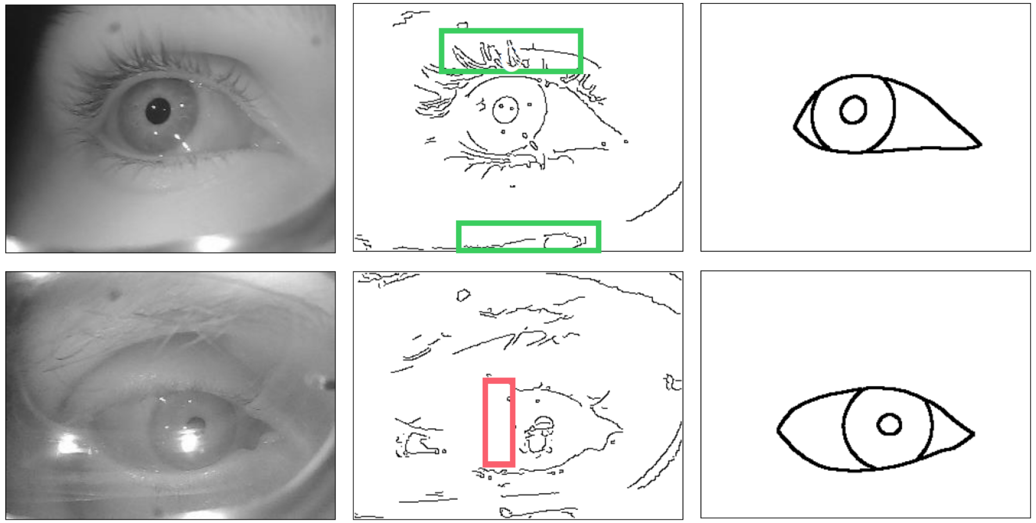


# Our Method

## Edge Extraction Network($E^2$-Net)

Challenge:

- To eliminate task-unrelated edges;
- To complete task-related edges.



Eye images | Results of Canny detection | Task-related Edges

- The edges in green boxes represent task-unrelated edges.
- We try to eliminate task-unrelated edges and restore the lost edges in red box.

# Our Method

## Edge Extraction Network(E$^2$-Net)

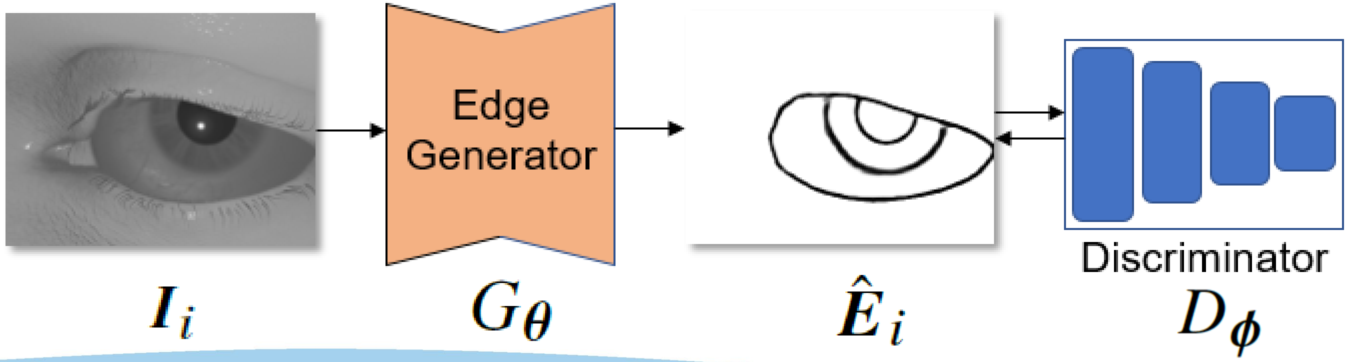
E$^2$-Net contains an edge generator G and a discriminator D

Loss functions:

$$\mathcal{L}_G(\theta) = \alpha \sum \ell_{\text{bce}}\left(\hat{\boldsymbol{E}}_i, \boldsymbol{E}_i\right) + \beta \sum \ell_{\text{pure}}(\boldsymbol{\theta}; \boldsymbol{I}_i), \quad \ell_{\text{pure}}(\boldsymbol{\theta}; \boldsymbol{I}_i) = (D_{\boldsymbol{\phi}}(R_{\boldsymbol{\theta}}(\boldsymbol{I}_i)) - 1)^2 = (D_{\boldsymbol{\phi}}(\hat{\boldsymbol{E}}_i) - 1)^2.$$

$$\mathcal{L}_D(\boldsymbol{\phi}) = \sum (D_{\boldsymbol{\phi}}(\boldsymbol{E}_i) - 1)^2 + \sum (D_{\boldsymbol{\phi}}(\hat{\boldsymbol{E}}_i))^2.$$

$\ell_{\text{bce}}$ is the binary cross-entropy loss, $\alpha$ and $\beta$ are fixed weight parameters.

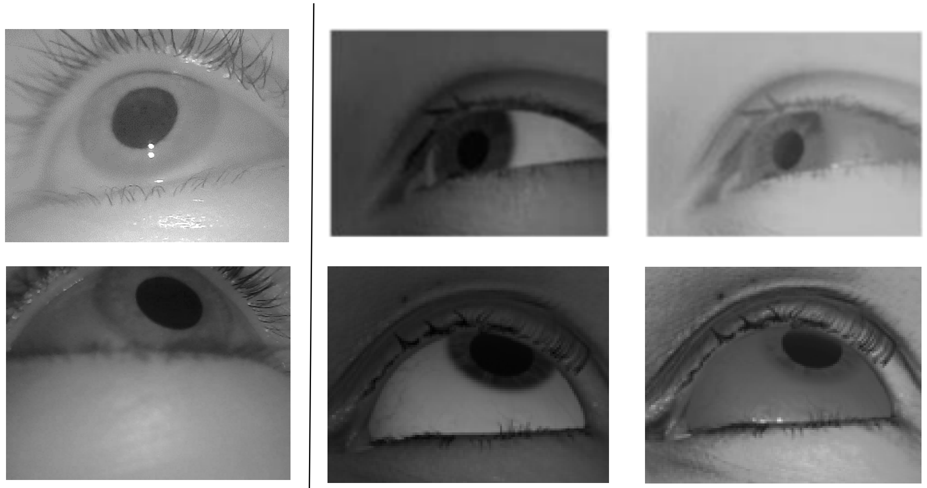



Please consult our paper for more details.

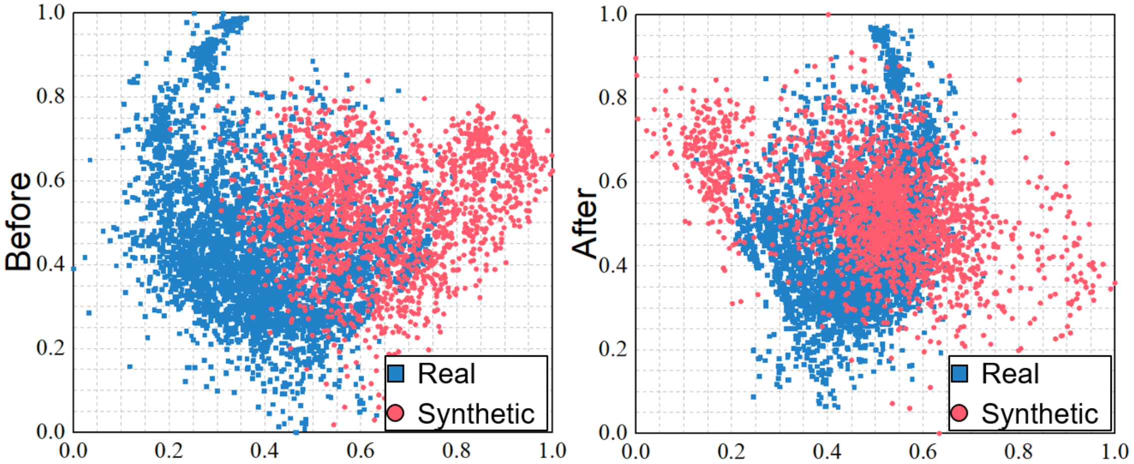
# Our Method

## Image Intensity Transfer ($I^2T$)

- Synthetic datasets have abundant eye images and accurate labels.
- There is large difference between the real and synthetic images.
- We propose the $I^2T$ method for producing realistic images from synthetic images.



Real Images | Synthetic Images | Transferred Images

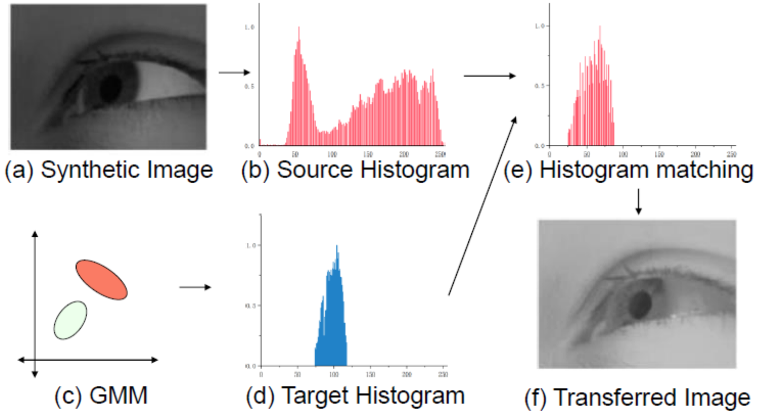



Visualization of real and synthetic images in 2-D space vis t-SNE

# Our Method

## Image Intensity Transfer ($I^2T$)

- ***Sample data:*** Sample 10000 images from four real near-eye datasets.
- ***Compute histogram:*** Divide each image into three subregions which respectively contain 1) iris and sclera, 2) skin, and 3) pupil. Then calculate the intensity histogram of each subregion.
- ***Fit mixture gaussian distributions:*** Fit three mixture gaussian distributions (GMM) which corresponds to the distributions of three subregion histograms.
- ***Align image intensity:*** Employ the histogram matching algorithm and perform a mapping that transforms intensities of the source image towards the target.

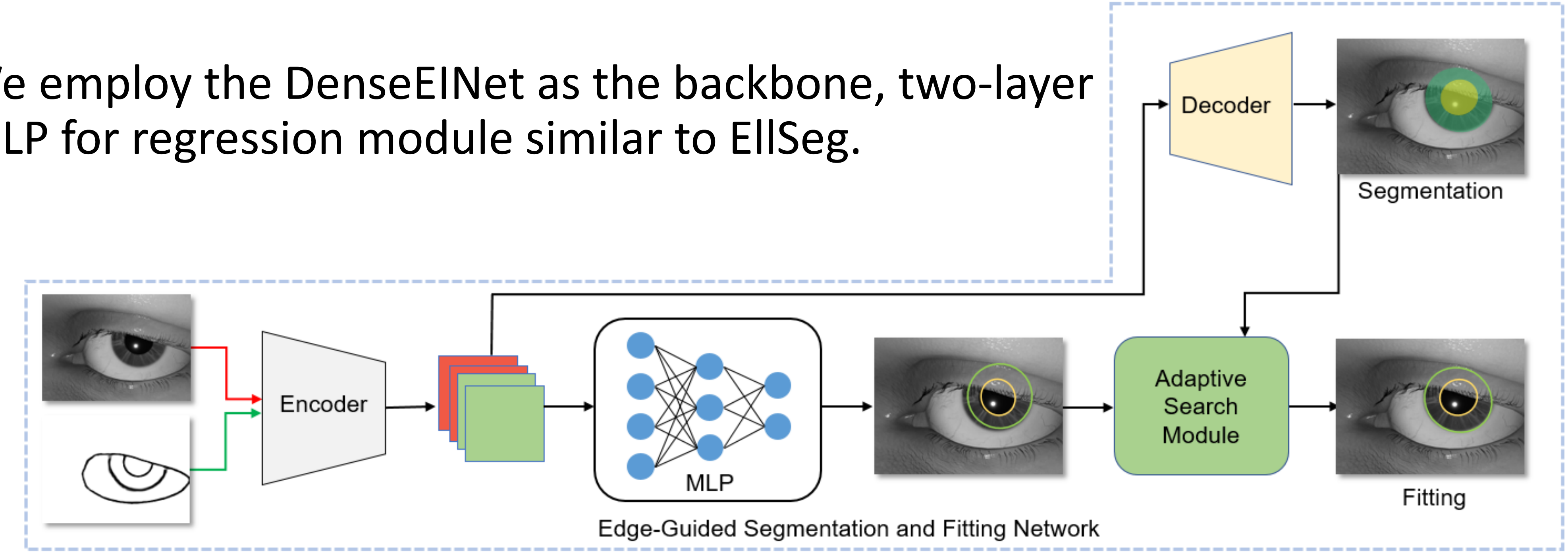



The procedure of ***Align image intensity***

# Our Method

## Edge-Guided Segmentation and Fitting Network (ESF-Net)

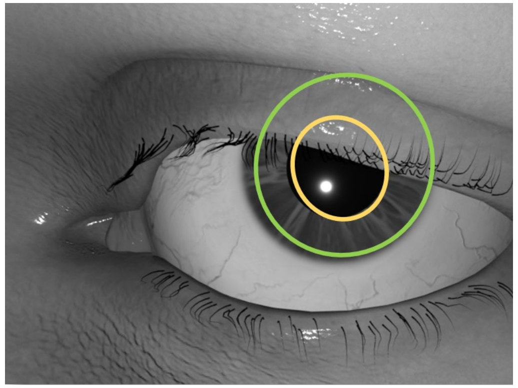
We employ the DenseElNet as the backbone, two-layer MLP for regression module similar to EllSeg.



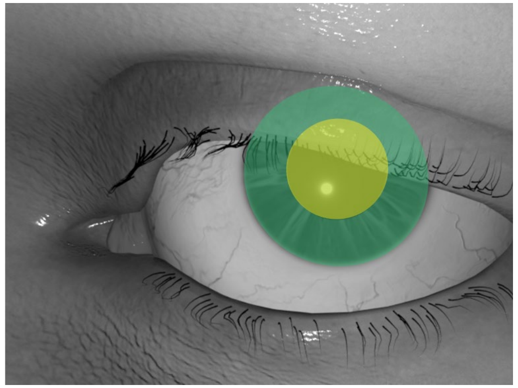
# Our Method

## Adaptive Search Module(ASM)

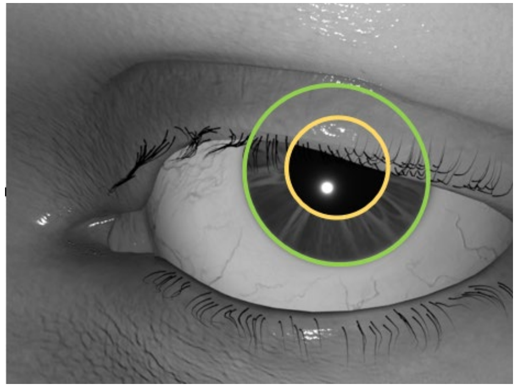
- Motivation: The regressed parameters usually are inaccurate.
- ASM searches the optimal ellipse on the guidance of segmentation maps
- Goal : Maximize the value of IoU between segmentation map and the ellipse form by parameters.



Origin Fitting



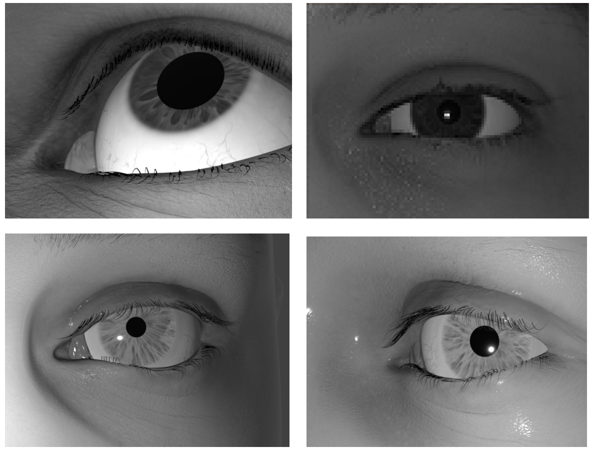
Segmentation



Optimal Fitting

# Experiments

- Train networks on synthetic dataset
- Evaluation on four publicly real datasets and customized AR HMD
- Compared methods: DeepVOG, RITNet, EllSeg



Sample images from RIT-Eyes dataset

| Dataset | Source | Purpose | Image Count | Sample Count |
|---|---|---|---|---|
| RITEyes-General | Synthetic | Train | 45516 | 45516 |
| NVGaze-AR* | AR HMD | Test | 2265127 | 11051 |
| OpenEDS | VR HMD | Test | 11202 | 11200 |
| LPW* | Head-mounted Eye tracker | Test | 130856 | 10865 |
| Fuhl* | Head-mounted Eye tracker | Test | 5665053 | 11197 |

Summary of train and test datasets

# Experiments

- Quantitative comparison between DeepVOG, RITNet, EllSeg and our methods (along rows) in four public datasets.

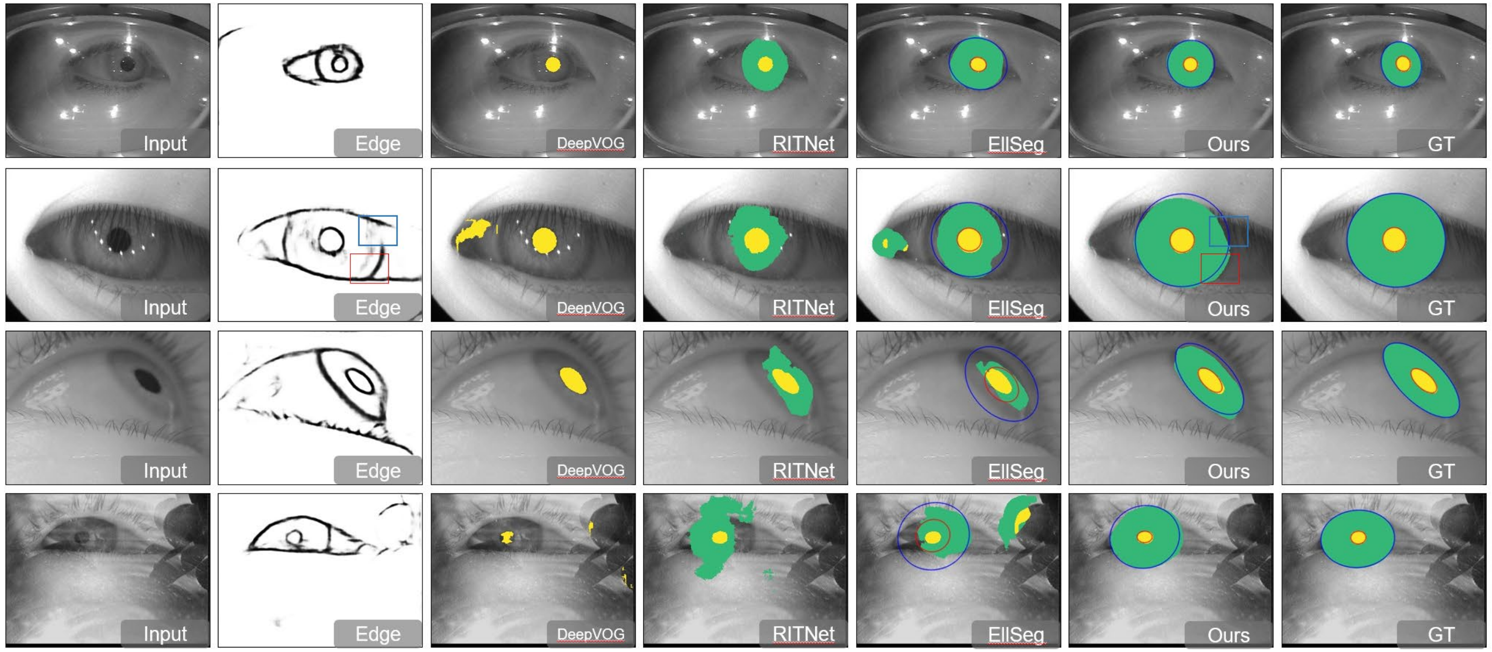
| Model \ Metric | $IoU_{pupil}\uparrow$ | $IoU_{iris}\uparrow$ | $PE_{pupil}\downarrow$ | $PE_{iris}\downarrow$ | $BIoU_{pupil}\uparrow$ | $BIoU_{iris}\uparrow$ | $IoU_{pupil}\uparrow$ | $IoU_{iris}\uparrow$ | $PE_{pupil}\downarrow$ | $PE_{iris}\downarrow$ | $BIoU_{pupil}\uparrow$ | $BIoU_{iris}\uparrow$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Benchmark** | **LPW** | | | | | | **NVGaze** | | | | | |
| DeepVOG | 0.833 | - | 4.66 | - | - | - | 0.867 | - | 1.23 | - | - | - |
| RITNet | 0.822 | 0.509 | 7.47 | 12.28 | - | - | 0.881 | 0.773 | 1.85 | 3.66 | - | - |
| EllSeg | 0.876 | 0.527 | 5.17 | 12.29 | 0.679 | 0.622 | 0.878 | 0.758 | 1.25 | 3.21 | 0.758 | 0.718 |
| Ours (E) | 0.885 | **0.689** | 4.41 | 11.84 | **0.780** | 0.700 | **0.890** | **0.814** | 1.24 | **2.90** | **0.765** | **0.747** |
| Ours (E+I) | **0.896** | 0.688 | **3.68** | **10.27** | 0.762 | **0.745** | 0.884 | 0.812 | **1.20** | 3.24 | 0.763 | 0.745 |
| **Benchmark** | **OpenEDS** | | | | | | **Fuhl** | | | | | |
| DeepVOG | 0.890 | - | 1.41 | - | - | - | 0.856 | - | 4.38 | - | - | - |
| RITNet | 0.889 | 0.611 | 2.41 | 6.24 | - | - | 0.862 | 0.718 | 5.09 | 7.91 | - | - |
| EllSeg | 0.921 | 0.740 | 1.39 | **5.32** | 0.799 | 0.727 | 0.893 | 0.737 | 3.70 | 8.02 | 0.778 | 0.739 |
| Ours (E) | 0.915 | **0.850** | 1.44 | 6.06 | 0.810 | 0.778 | 0.893 | 0.795 | 2.84 | 7.63 | 0.790 | 0.750 |
| Ours (E+I) | **0.925** | 0.821 | **1.27** | 5.40 | **0.813** | **0.780** | **0.904** | **0.813** | **2.80** | **7.25** | **0.800** | **0.780** |

Ours (E) means the result of ESF-Net that only inputs edge maps.

Ours (E+I) means the result of ESF-Net that combines images with edge maps.

# Experiments

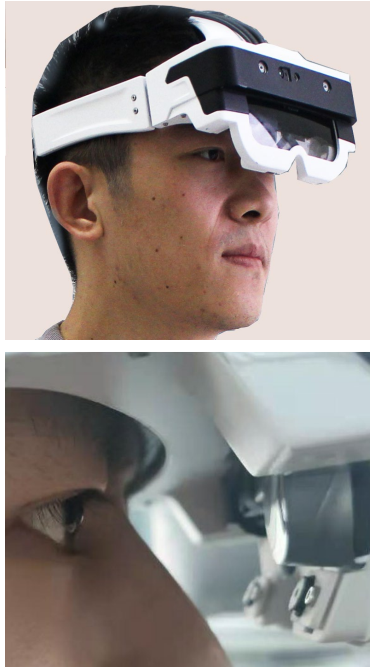
- Visual comparisons of segmentation and fitting results.
- Under the guidance of the task-related edge, our model 1) smooths the contour of segmentation map 2) produces more accurate ellipse shape 3) is robust with disturbances.



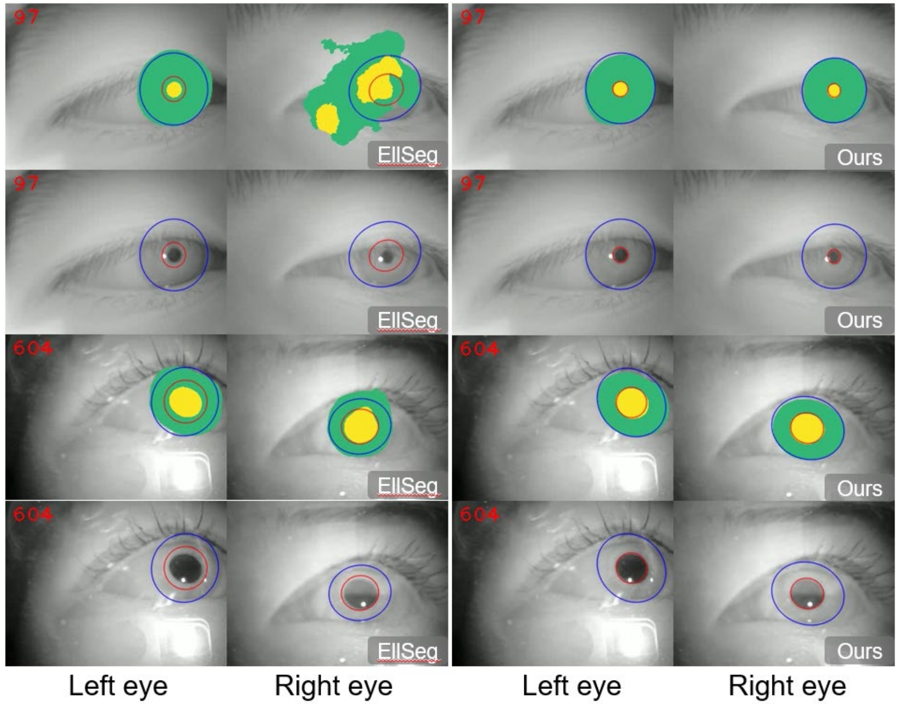

Results in four public datasets

# Experiments

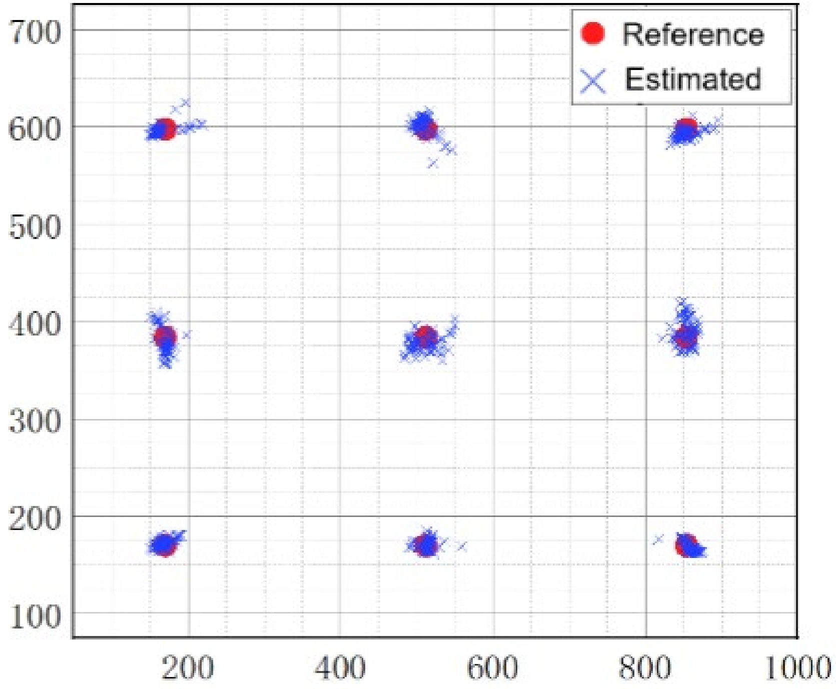
- Assessed our model in customized AR device.
- Implemented the 2D gaze estimation task in the AR device, achieving 0.38° accuracy.



Our AR Device



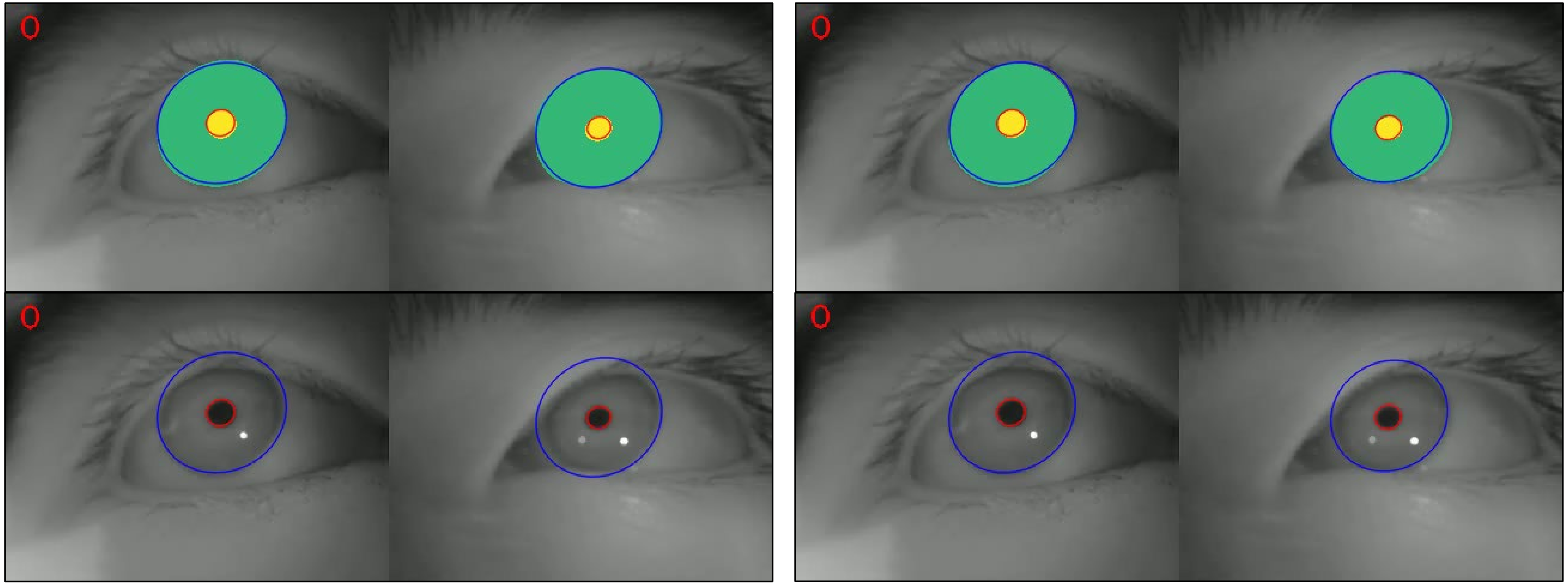

Segmentation Result




Gaze Estimation Result

# Experiments

- Segmentation and fitting results on our customized AR device. (Binocular)




Left eye Right eye

EllSeg (IEEE TVCG' 21)

Left eye Right eye

Ours

* Captured in outdoor environment.

# Future Work

- Conduct more online tests with the AR device in real time.
- Optimize our network by knowledge distillation.
- Fit 3D eye model based segmentation and fitting results by our method, which can mitigate the effect of device slippage and improve the stability of gaze estimation.

Our work: https://github.com/zhaoyuhsin/Edge-Guided-Near-Eye-Image-Analysis

ISMAR 2021

OCTOBER 4-8 BARI - ITALY

IEEE COMPUTER SOCIETY

IEEE

Advancing Technology for Humanity

Thank you for your time and interest in our work!