# Gaze Estimation

Gyanig Kumar

CSCI 7000 : Recent Advances in Computer Vision

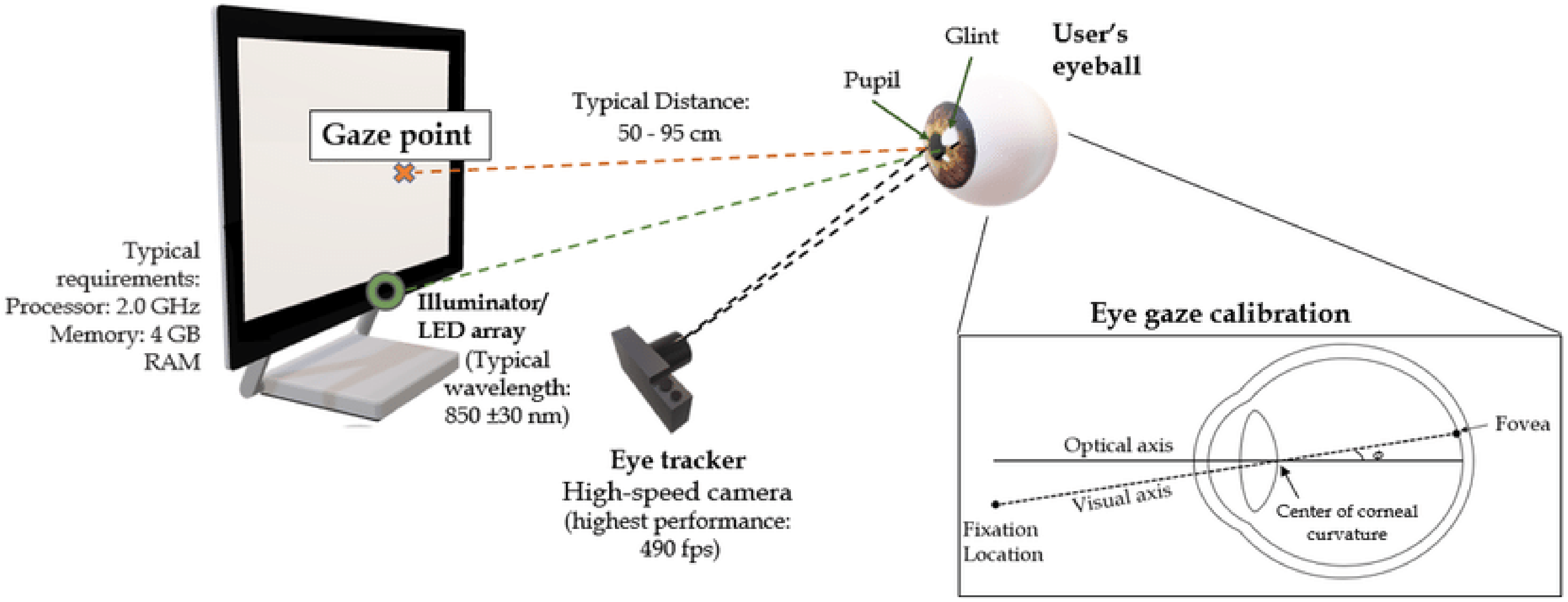
*“Sarvendriyaanaam Nayanam Prad hanam”*

*which loosely translates into,*

*“Of all the sense organs, vision is the most important”.*

[Pupil Invisible eye tracking glasses: Instant insight into human behavior (youtube.com)](youtube.com) 2021

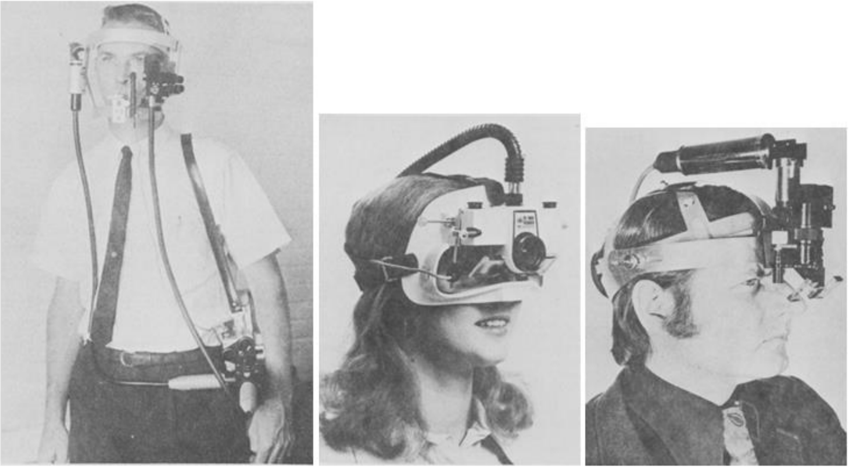
# Introduction to Human Vision




https://www.researchgate.net/figure/Components-of-a-typical-eye-gaze-system-adapted-from-22-38-The-optical-and-the-visual_fig2_332584054

# Historical Perspective

Early approaches to gaze estimation

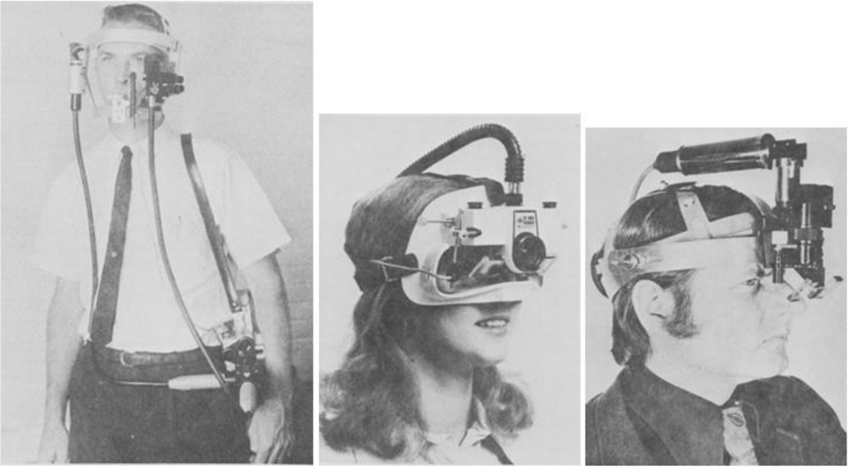


(a) Mobile corneal reflex system (b) Head-mounted system (c) Head-mounted system

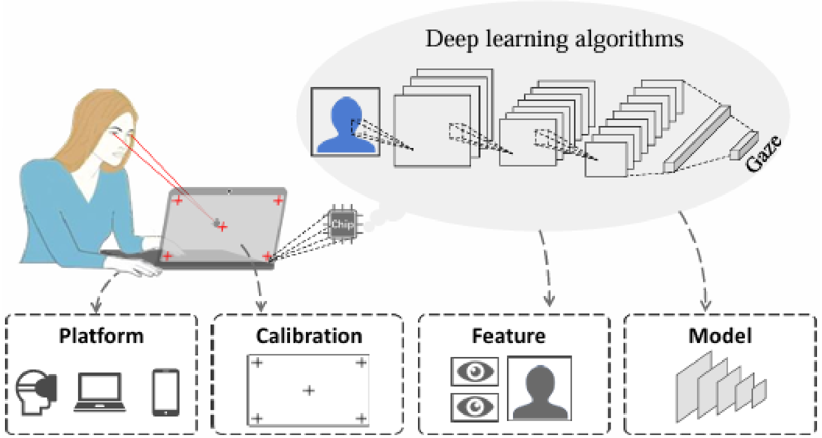
Credits: Appearance-based Gaze Estimation with Deep Learning: A Review and Benchmark, Gaze Direction Estimation · google/mediapipe, Tobii Eyetrackers

# Historical Perspective

Early approaches to gaze estimation



(a) Mobile corneal reflex system (b) Head-mounted system (c) Head-mounted system

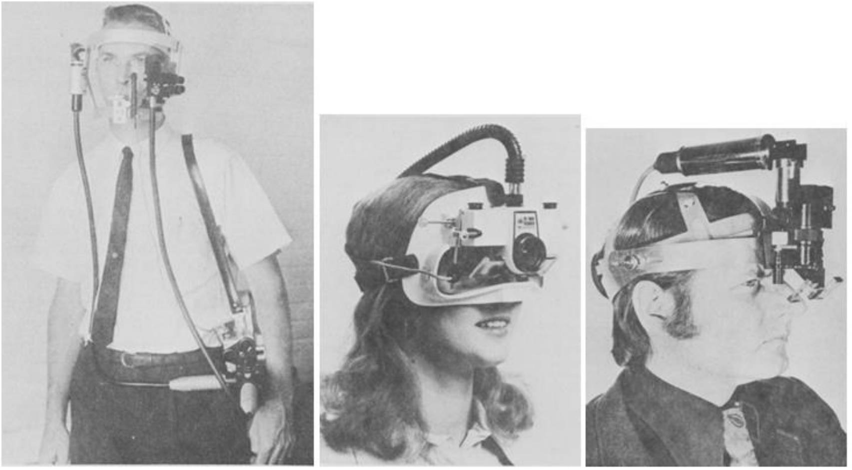



Modern Approach for Gaze Estimation

Credits: Appearance-based Gaze Estimation with Deep Learning: A Review and Benchmark, Gaze Direction Estimation · google/mediapipe, Tobii Eyetrackers

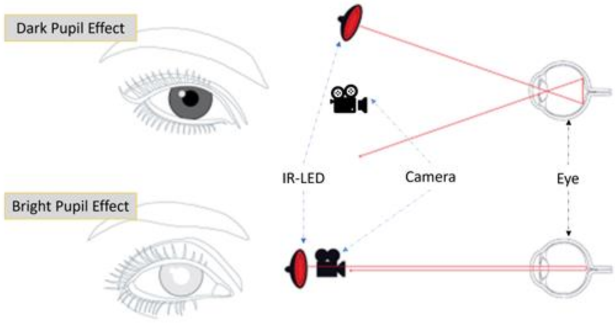
# Historical Perspective

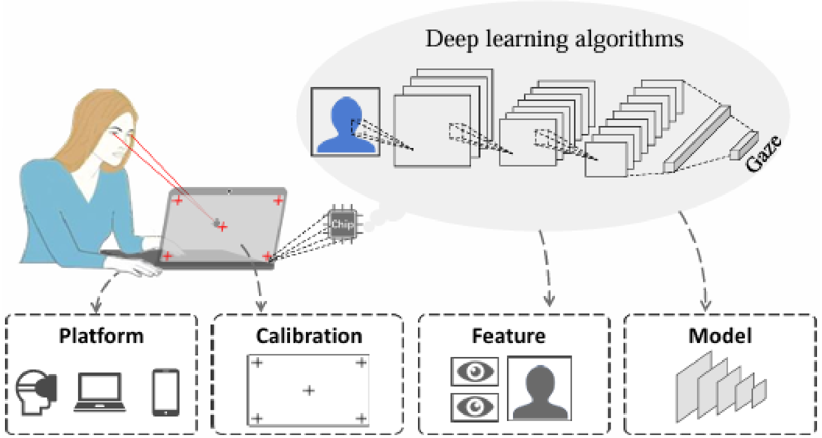
## Early approaches to gaze estimation



(a) Mobile corneal reflex system (b) Head-mounted system (c) Head-mounted system

## Building Wearable EyeTrackers

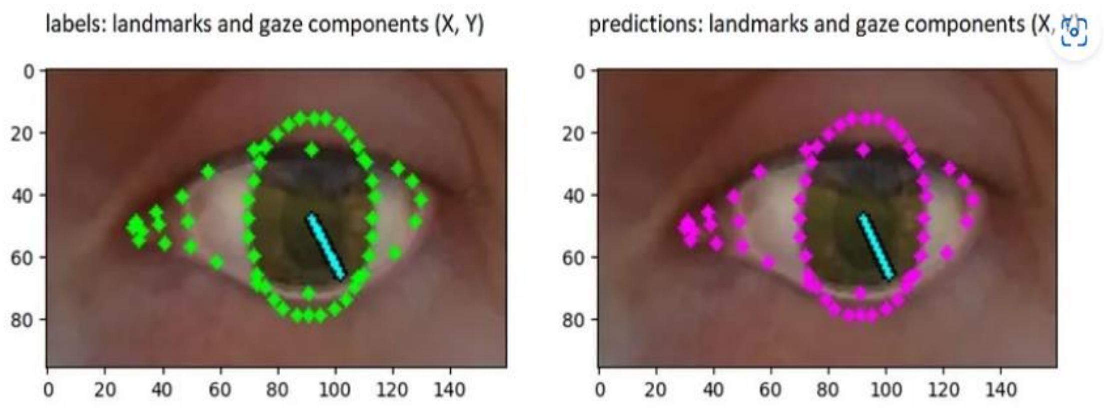


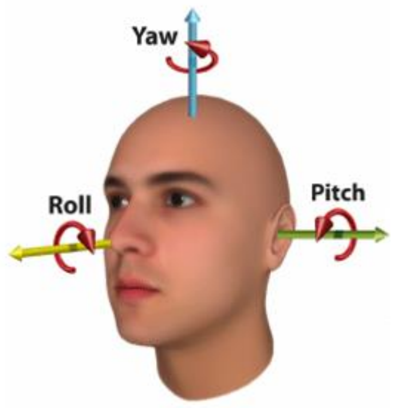

Modern Approach for Gaze Estimation

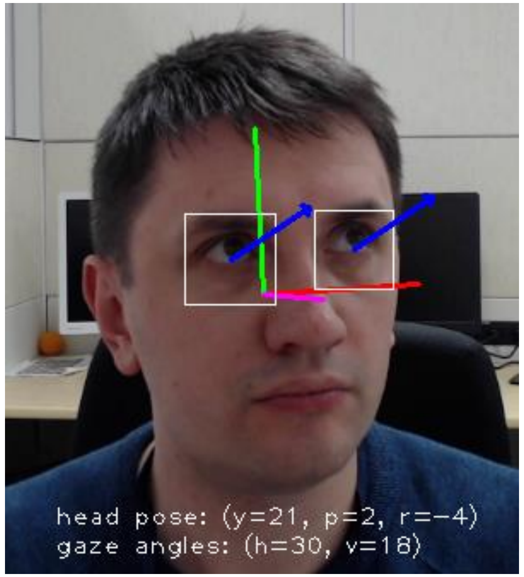
Credits: Appearance-based Gaze Estimation with Deep Learning: A Review and Benchmark, Gaze Direction Estimation · google/mediapipe, Tobii Eyetrackers

# Understanding the Perspective




Pupil Detection




Gaze Estimation

Credits: Gaze Estimation: problem, solutions and practical experiments. | by Olga Mindlina | Medium

Orientation of the head in terms of pitch , roll , and yaw movements... | Download Scientific Diagram (researchgate.net)

# Timeline

**1920-2017** **2015** **2016–2019** **2019** **2021** **2023**

*Introduction*

Discovery of Human Gaze movements basics
Experimental psychology research
Wearable Trackers came out

# Timeline

*Deep learning*

With introduction of Alex Net, first CNN based model for eye tracker introduced

**1920-2017** **2015** **2016–2019** **2019** **2021** **2023**

*Introduction*

Discovery of Human Gaze movements basics
Experimental psychology research
Wearable Trackers came out

# Timeline

**1920-2017** → **2015** → **2016–2019** → **2019** → **2021** → **2023**

*Introduction*

Discovery of Human Gaze movements basics
Experimental psychology research
Wearable Trackers came out

*Deep learning*

With introduction of Alex Net, first CNN based model for eye tracker introduced

*Dataset Work*

Creating a dataset was biggest problem (Face+Eye):
MPIIGaze
GazeCapture
EyeDiap

# Timeline

GAZE2024
6TH INTERNATIONAL WORKSHOP ON GAZE ESTIMATION AND PREDICTION IN THE WILD
Image credit to DALL.E

**1920-2017**

*Introduction*

Discovery of Human Gaze movements basics
Experimental psychology research
Wearable Trackers came out

**2015**

*Deep learning*

With introduction of Alex Net, first CNN based model for eye tracker introduced

**2016–2019**

*Dataset Work*

Creating a dataset was biggest problem (Face+Eye):
MPIIGaze
GazeCapture
EyeDiap

**2019**

*Attention based*

CNN-Models started to adapt the newer attention based models for gaze estimation
ECCV->CVPR workshop

**2021**

**2023**

# Timeline

GAZE2024
6TH INTERNATIONAL WORKSHOP ON GAZE ESTIMATION AND PREDICTION IN THE WILD
Image credit to DALL.E

**1920-2017**

*Introduction*

Discovery of Human Gaze movements basics
Experimental psychology research
Wearable Trackers came out

**2015**

*Deep learning*

With introduction of Alex Net, first CNN based model for eye tracker introduced

**2016–2019**

*Dataset Work*

Creating a dataset was biggest problem (Face+Eye):
MPIIGaze
GazeCapture
EyeDiap

**2019**

*Attention based*

CNN-Models started to adapt the newer attention based models for gaze estimation
ECCV->CVPR workshop

**2021**

*Transformers*

New approach with Transformers
Large-Scale Dataset introduced:
EVE
ETH-Xgaze
Gaze360

**2023**

# Timeline

GAZE2024
6TH INTERNATIONAL WORKSHOP ON GAZE ESTIMATION AND PREDICTION IN THE WILD
Image credit to DALL·E

**1920-2017**

*Introduction*

Discovery of Human Gaze movements basics
Experimental psychology research
Wearable Trackers came out

**2015**

*Deep learning*

With introduction of Alex Net, first CNN based model for eye tracker introduced

**2016–2019**

*Dataset Work*

Creating a dataset was biggest problem (Face+Eye):
MPIIGaze
GazeCapture
EyeDiap

**2019**

*Attention based*

CNN-Models started to adapt the newer attention based models for gaze estimation
ECCV->CVPR workshop

**2021**

*Transformers*

New approach with Transformers
Large-Scale Dataset introduced:
EVE
ETH-Xgaze
Gaze360

**2023**

*Leaving traditional*

First ever approach to building a end-to-end model without taking features from eyes + face

# What are we going to discuss?

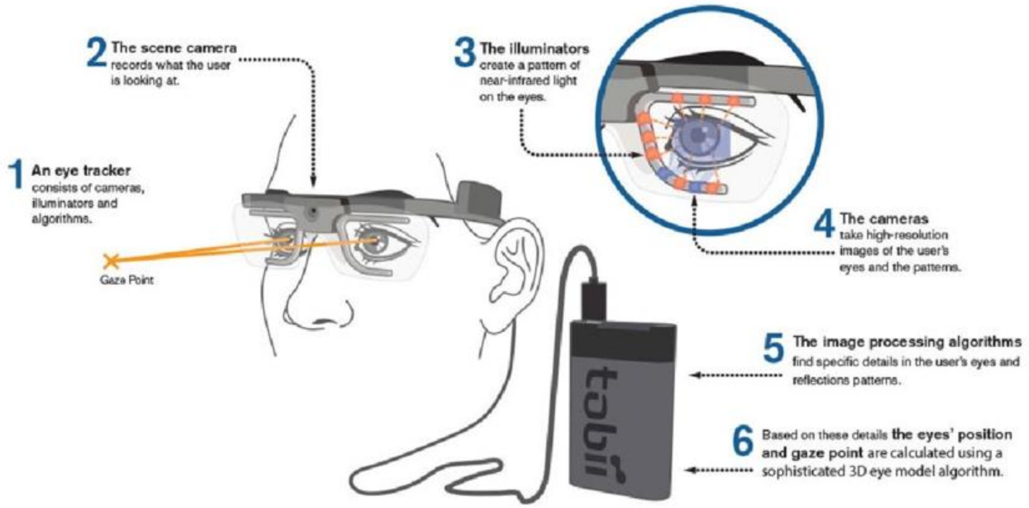
- Types of Gaze Estimation
- Datasets
- Evaluation Metrics of Gaze Estimation
- Appearance Based Gaze Estimation (AGE)
- End-to-End Frame-to-Gaze Estimation (EFE)
- Real World Case Study of Appearance Based Gaze Estimation

# What are we going to discuss?

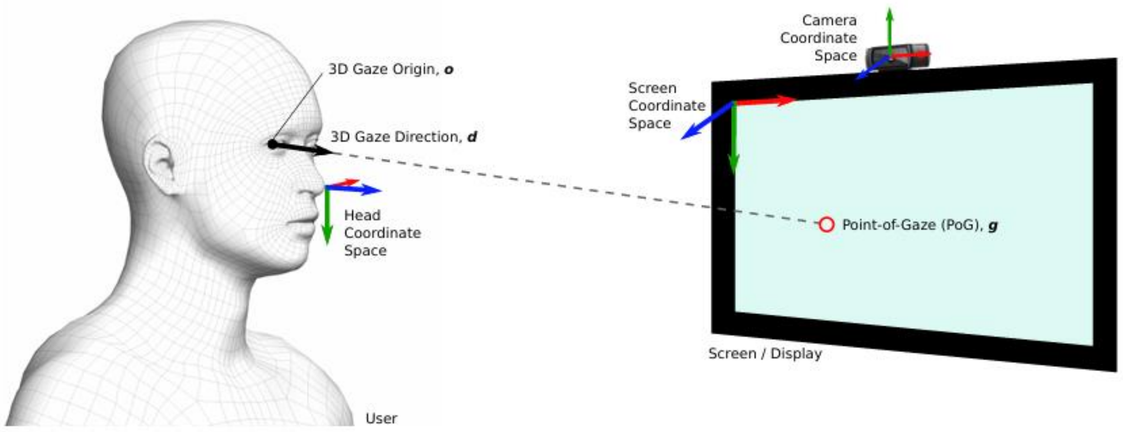
- Types of Gaze Estimation
- Datasets
- Evaluation Metrics of Gaze Estimation
- Appearance Based Gaze Estimation (AGE)
- End-to-End Frame-to-Gaze Estimation (EFE)
- Real World Case Study of Appearance Based Gaze Estimation

# Types of Gaze Estimation

## IR based Tracking




## Webcam based Tracking



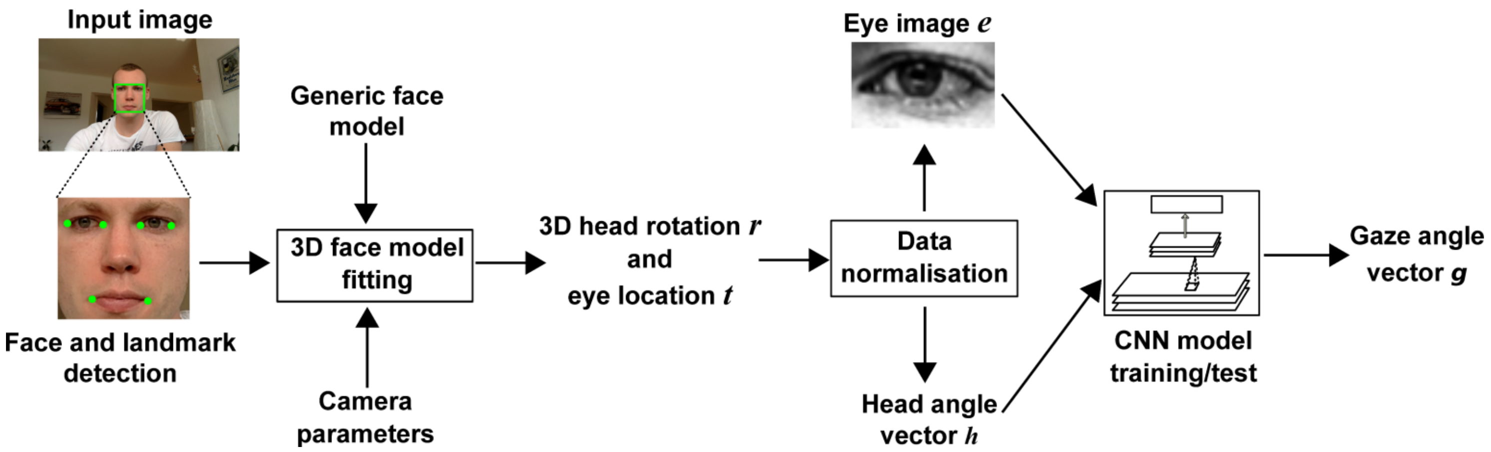

Also called the Appearance Based Gaze Estimation

Wearable-eye-tracker-Tobii-Pro-Glasses-2-Source-64.png (850×420) (researchgate.net)

Eye Gaze Estimation and Its Applications (uni-saarland.de)

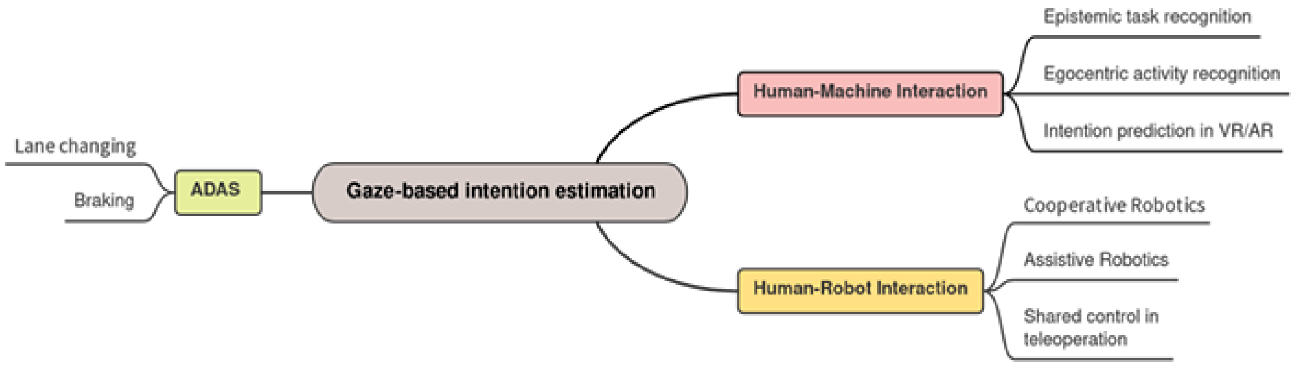
# Different aspects of eye tracking

- Face tracking
- Face landmark
- Data normalization
- Gaze origin
- Gaze direction



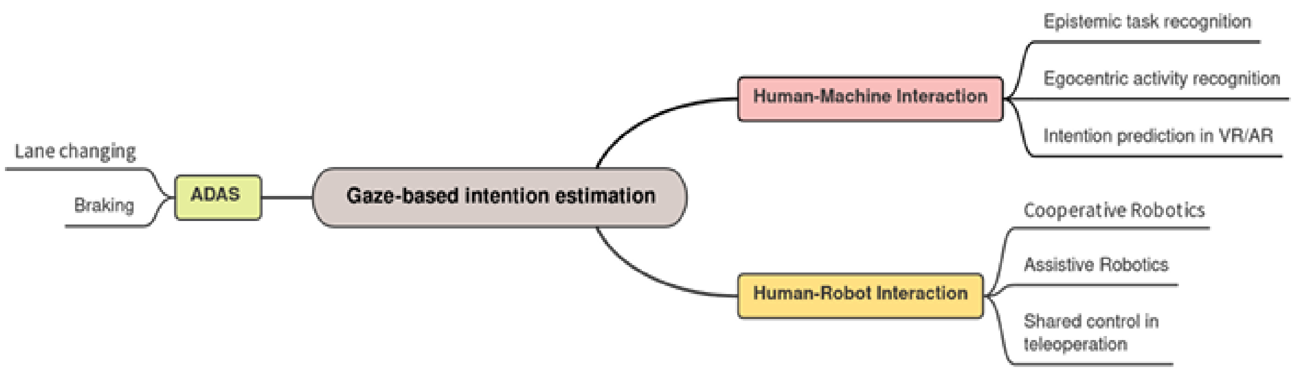

Introduction of MPIIGaze Dataset along with its CNN model

# Application areas



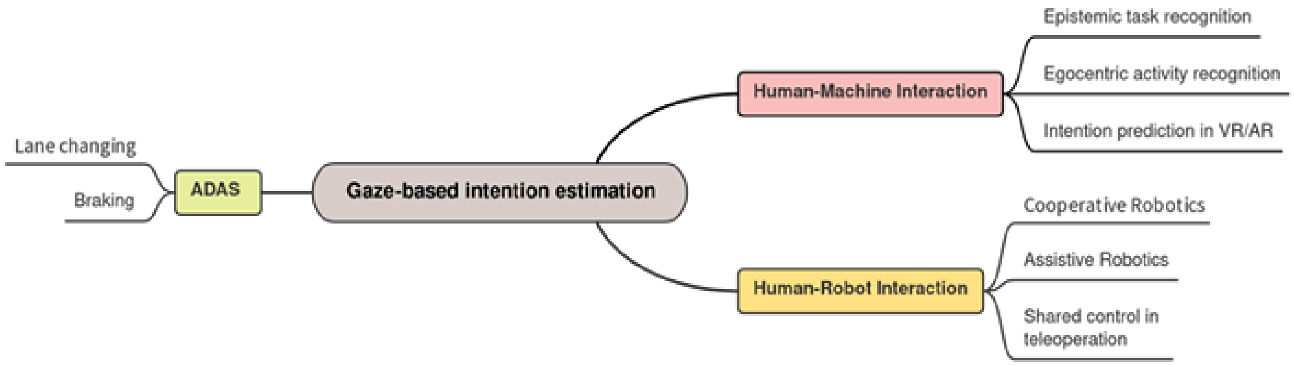

Gaze-based intention estimation: principles, methodologies, and applications in HRI (ANNA BELARDINELLI, Honda Research Institute Europe, Germany)
An End-to-End Review of Gaze Estimation and its Interactive Applications on Handheld Mobile Devices, https://www.mdpi.com/2218-6581/10/2/68/xml
https://flightsafety.org/asw-article/eyeing-the-eyes/ , https://research.adobe.com/publication/vrdoc-gaze-based-interactions-for-vr-reading-experience/

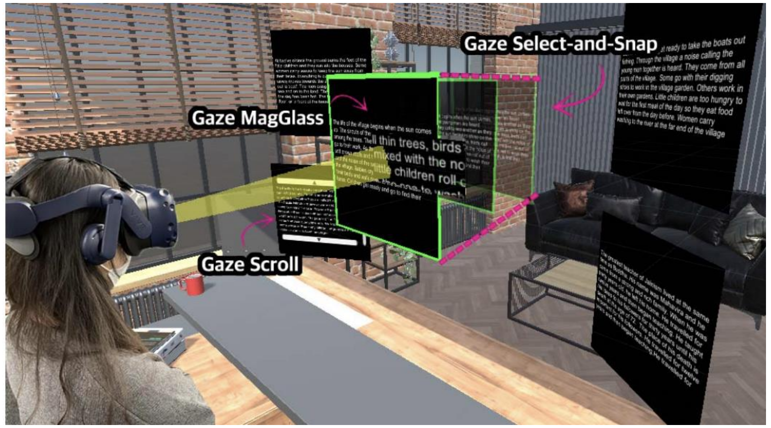
# Application areas




Gaze-based intention estimation: principles, methodologies, and applications in HRI (ANNA BELARDINELLI, Honda Research Institute Europe, Germany)
An End-to-End Review of Gaze Estimation and its Interactive Applications on Handheld Mobile Devices, https://www.mdpi.com/2218-6581/10/2/68/xml
https://flightsafety.org/asw-article/eyeing-the-eyes/ , https://research.adobe.com/publication/vrdoc-gaze-based-interactions-for-vr-reading-experience/

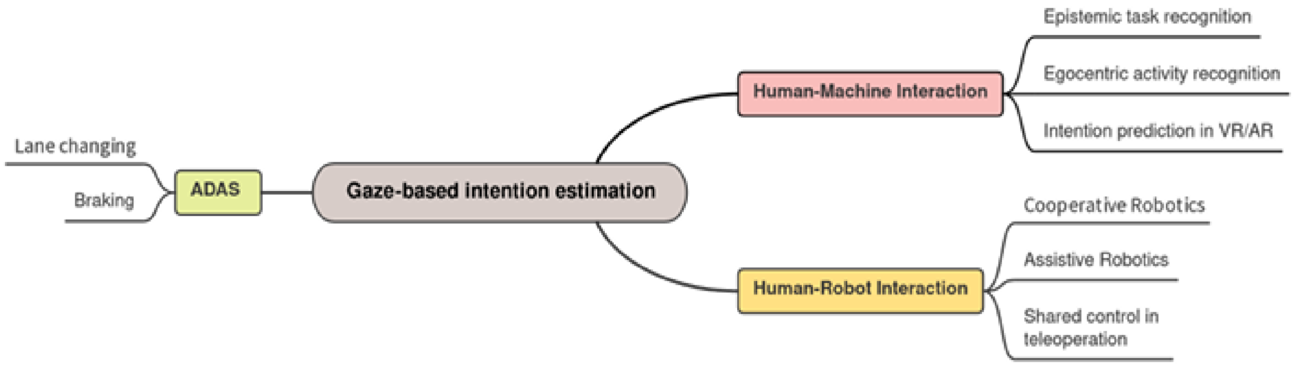
# Application areas

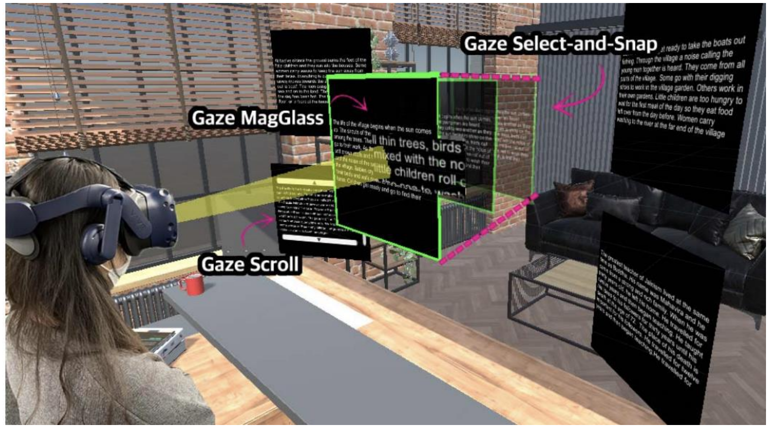


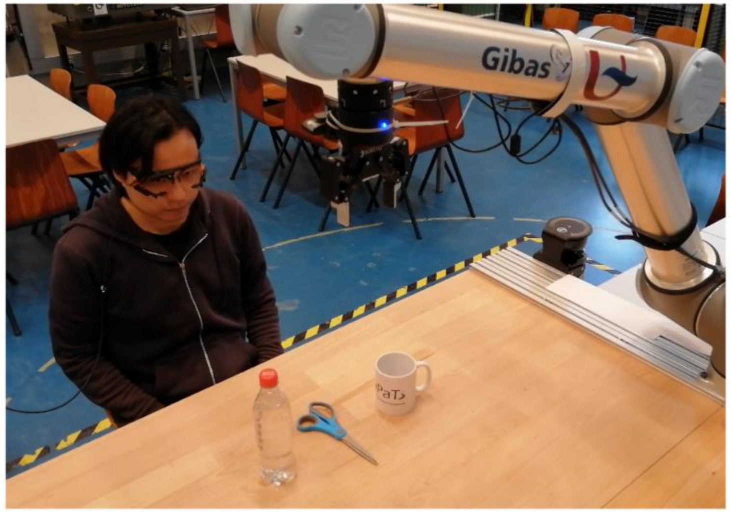




Gaze-based intention estimation: principles, methodologies, and applications in HRI (ANNA BELARDINELLI, Honda Research Institute Europe, Germany)
An End-to-End Review of Gaze Estimation and its Interactive Applications on Handheld Mobile Devices, https://www.mdpi.com/2218-6581/10/2/68/xml
https://flightsafety.org/asw-article/eyeing-the-eyes/ , https://research.adobe.com/publication/vrdoc-gaze-based-interactions-for-vr-reading-experience/

# Application areas

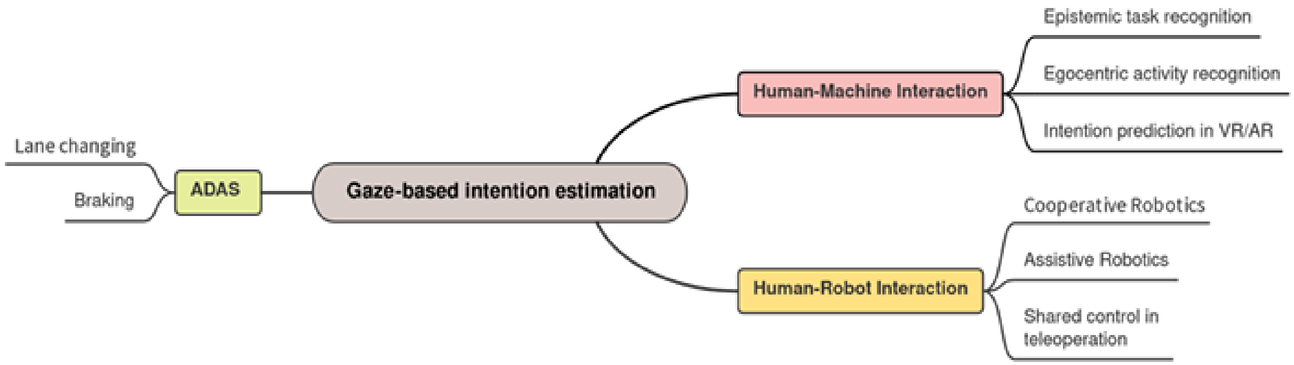


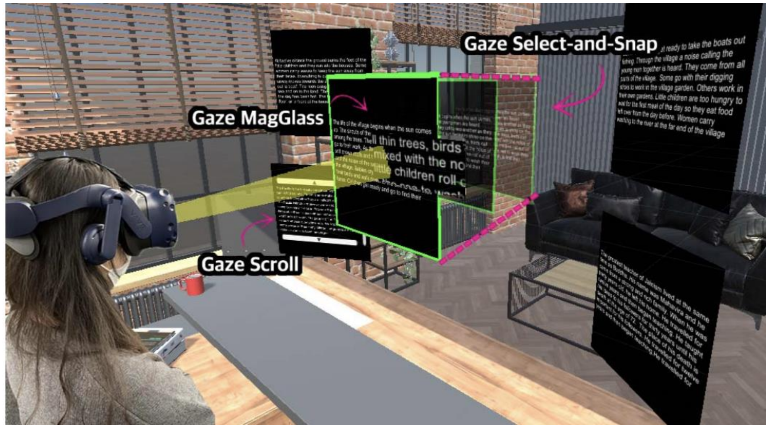



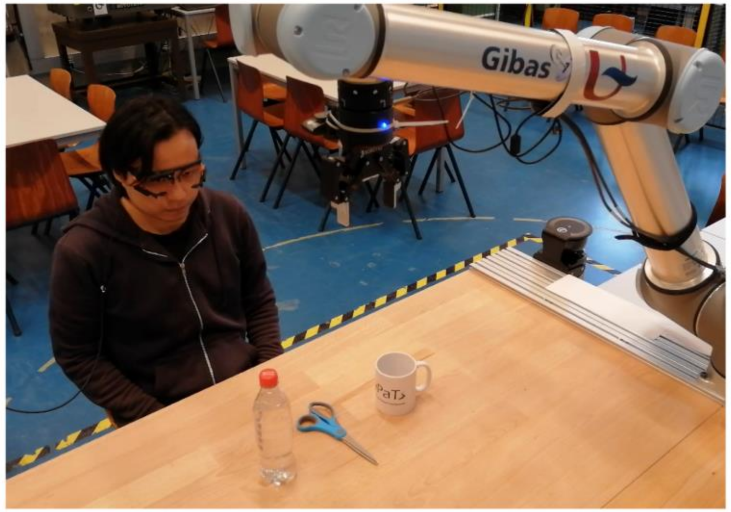
Gaze-based intention estimation: principles, methodologies, and applications in HRI (ANNA BELARDINELLI, Honda Research Institute Europe, Germany)
An End-to-End Review of Gaze Estimation and its Interactive Applications on Handheld Mobile Devices, https://www.mdpi.com/2218-6581/10/2/68/xml
https://flightsafety.org/asw-article/eyeing-the-eyes/ , https://research.adobe.com/publication/vrdoc-gaze-based-interactions-for-vr-reading-experience/

# Application areas







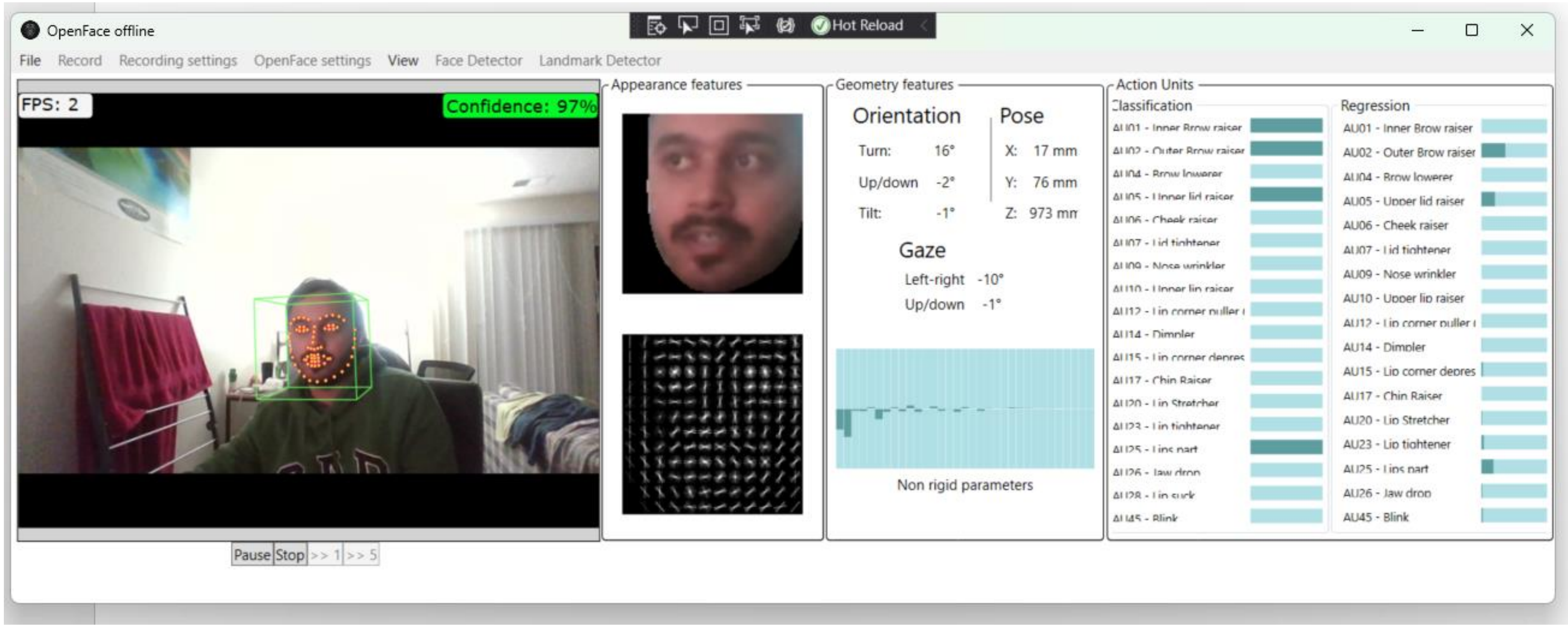


Gaze-based intention estimation: principles, methodologies, and applications in HRI (ANNA BELARDINELLI, Honda Research Institute Europe, Germany)
An End-to-End Review of Gaze Estimation and its Interactive Applications on Handheld Mobile Devices, https://www.mdpi.com/2218-6581/10/2/68/xml
https://flightsafety.org/asw-article/eyeing-the-eyes/ , https://research.adobe.com/publication/vrdoc-gaze-based-interactions-for-vr-reading-experience/

# Other Applications?

# Live Demo of Appearance Based Gaze Estimation

# Live Demo of Appearance Based Gaze Estimation



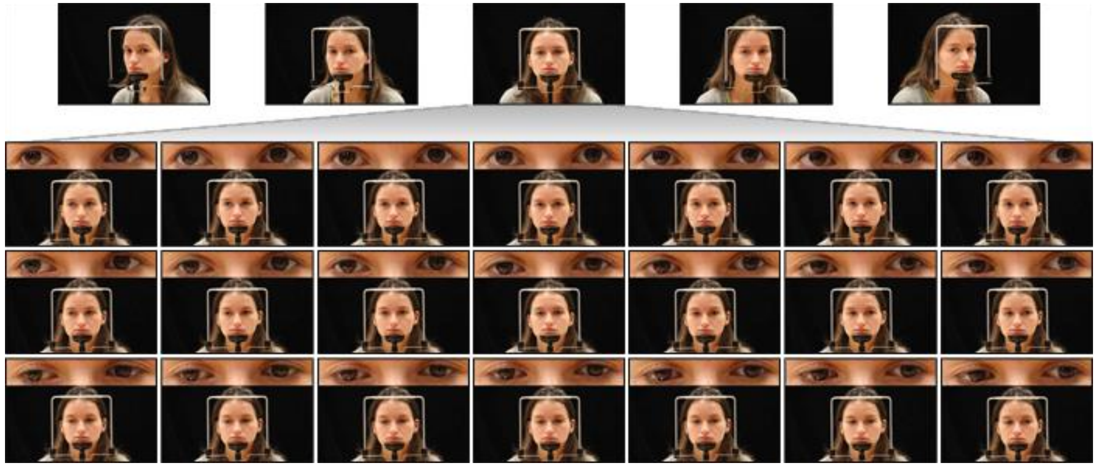

OpenFace Offline Gaze Estimation Demo

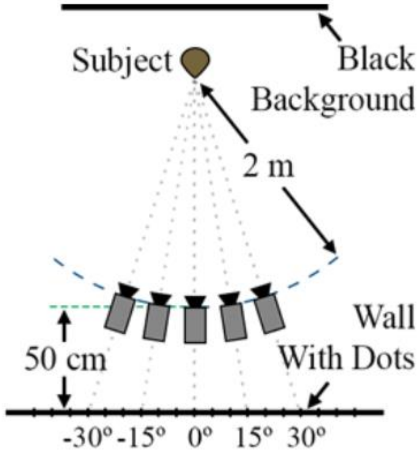
# What are we going to discuss?

- ~~Types of Gaze Estimation~~
- Datasets
- Evaluation Metrics of Gaze Estimation
- Appearance Based Gaze Estimation (AGE)
- End-to-End Frame-to-Gaze Estimation (EFE)
- Real World Case Study of Appearance Based Gaze Estimation

# How to create the Datasets?

Lab setting vs in-the-wild

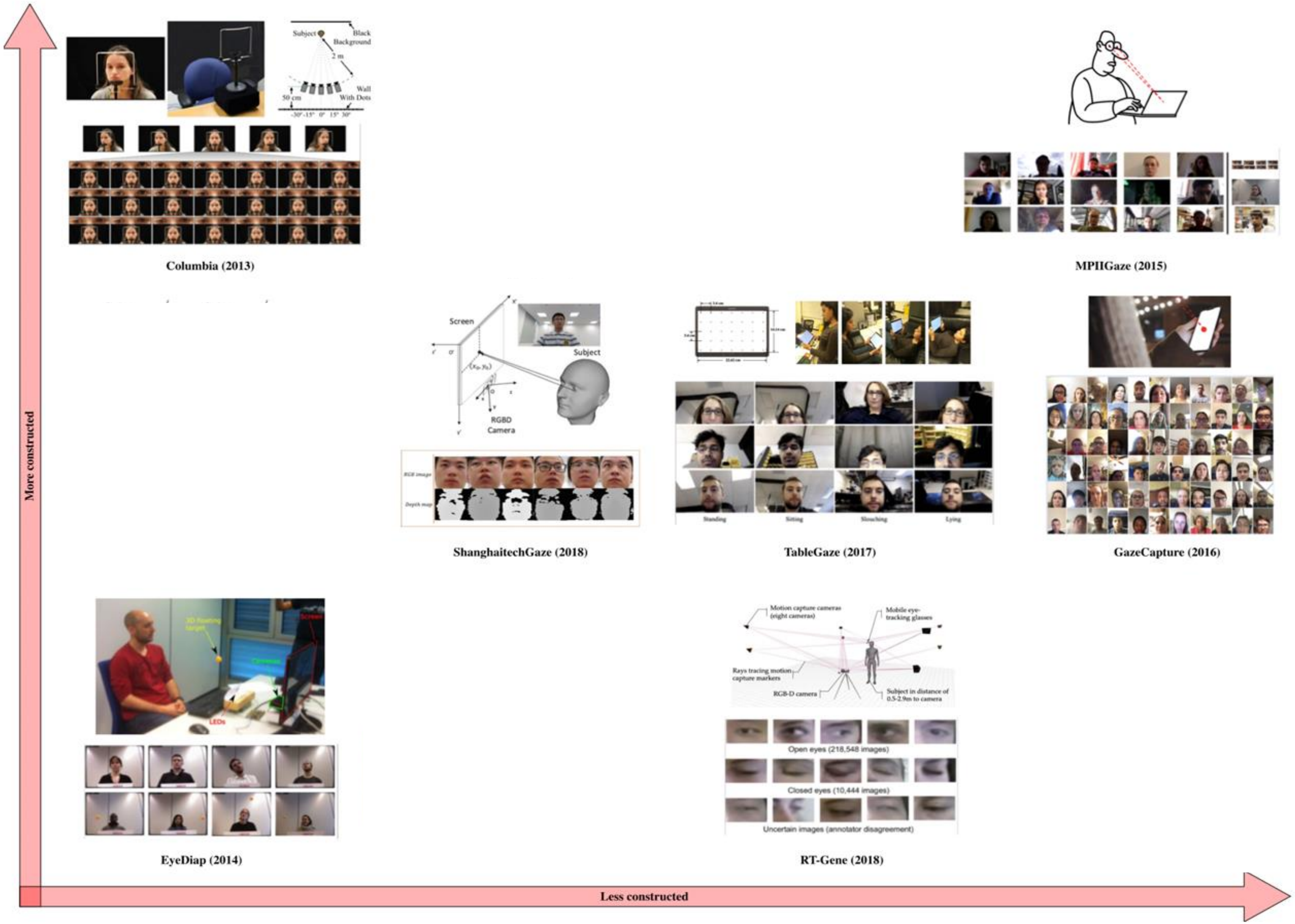





CAVE | Database: Columbia Gaze Data Set

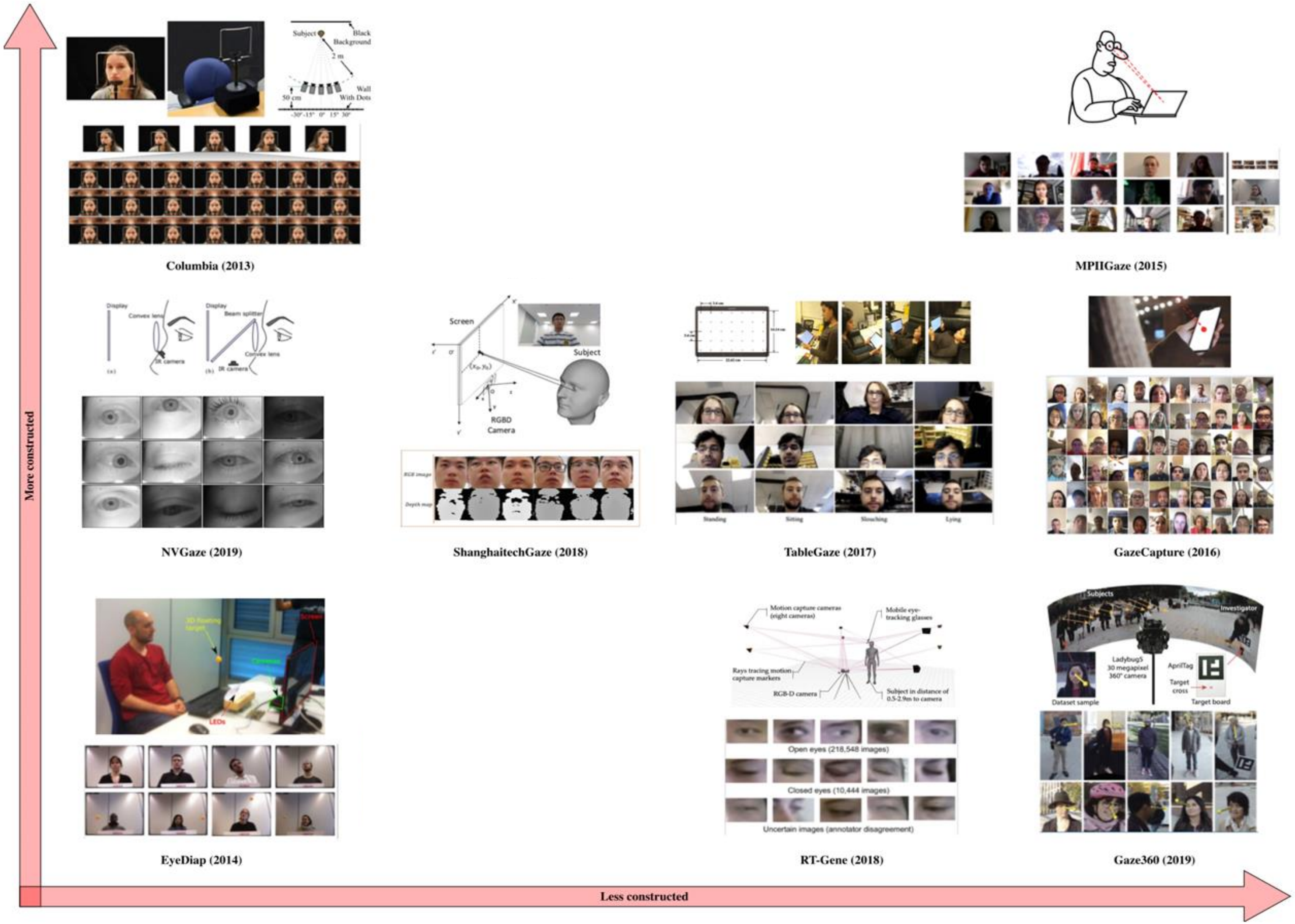
# Datasets

Lab setting vs in-the-wild



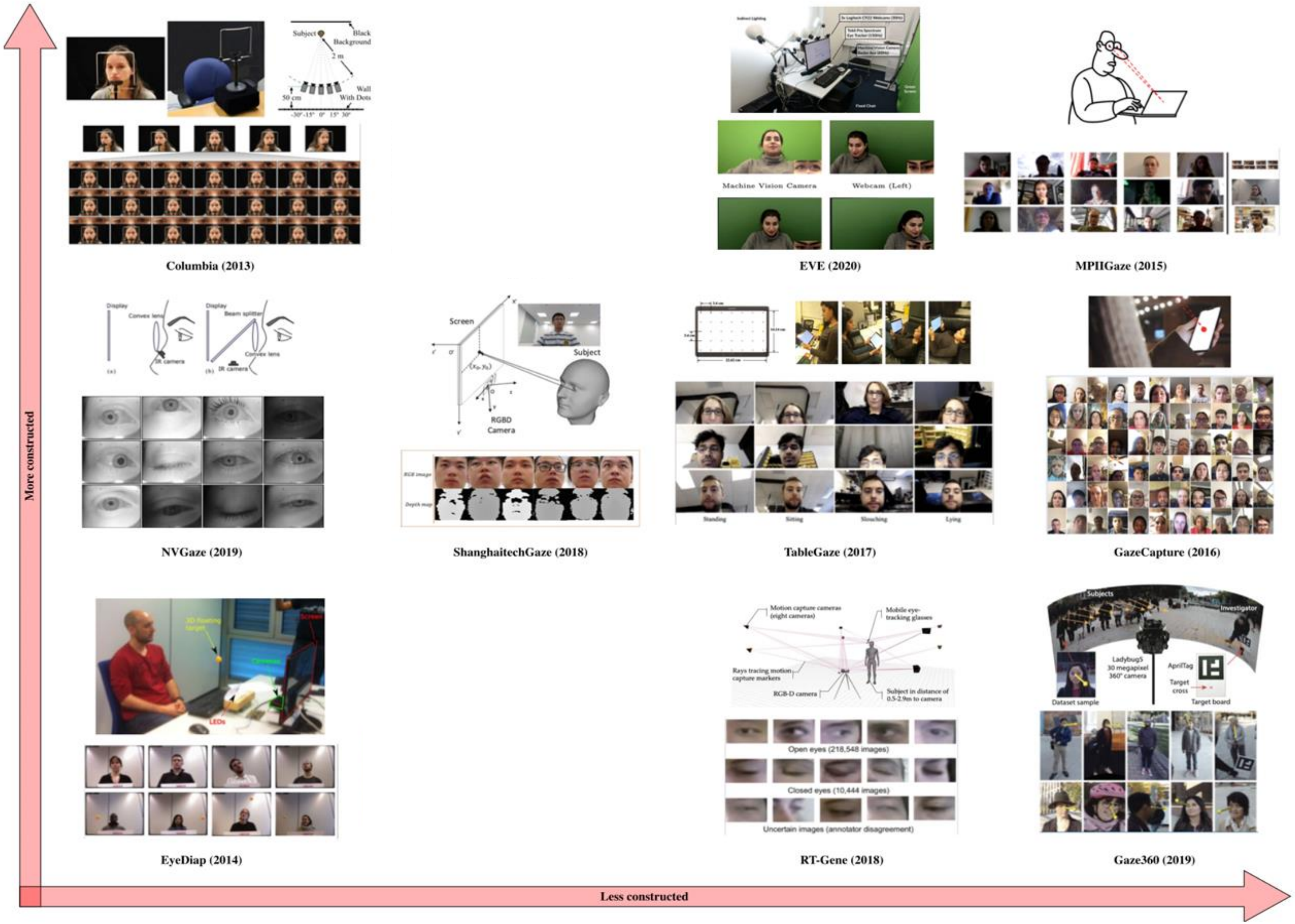
# Datasets

Lab setting vs in-the-wild



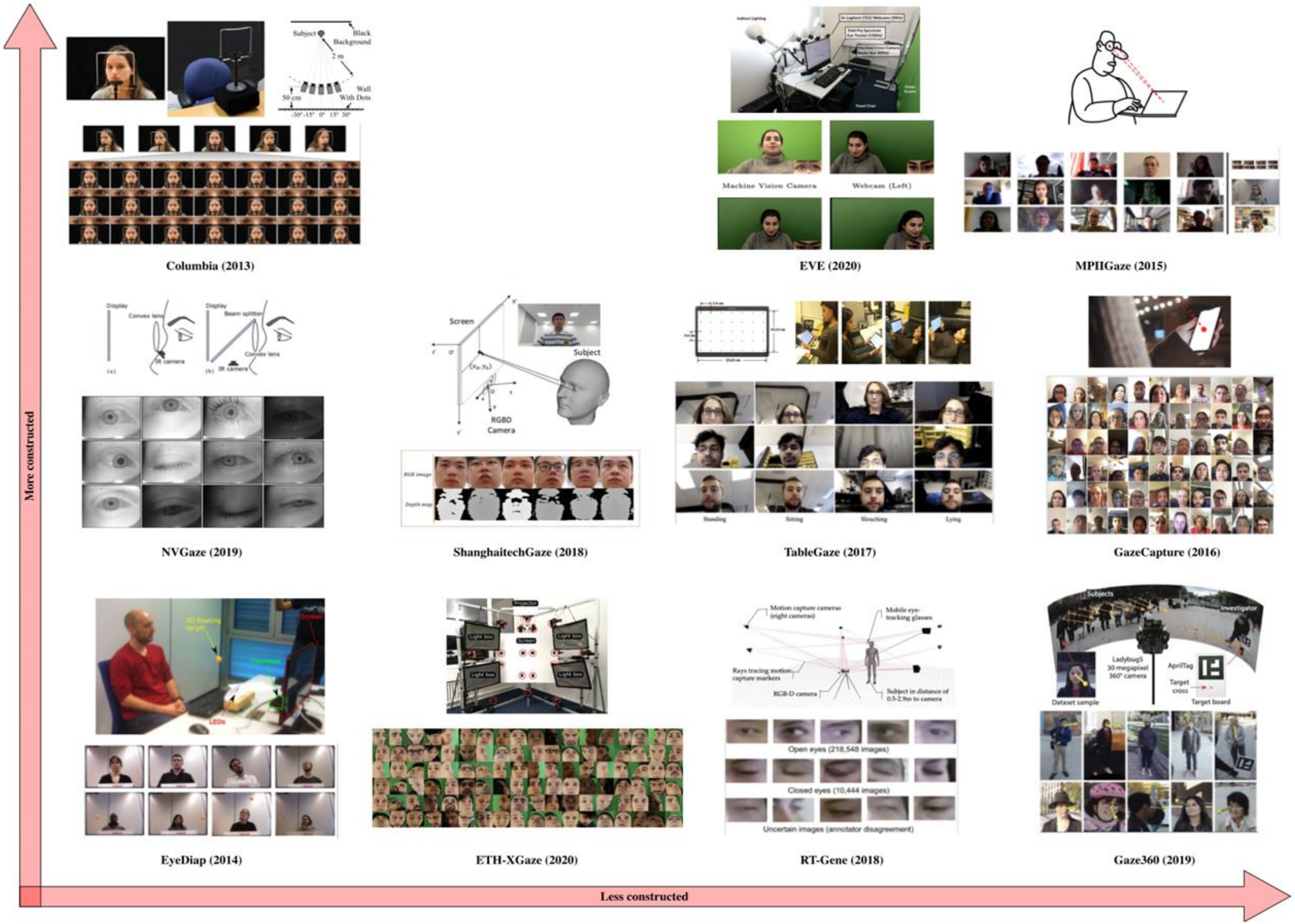
# Datasets

Lab setting vs in-the-wild



# Datasets

Lab setting vs in-the-wild

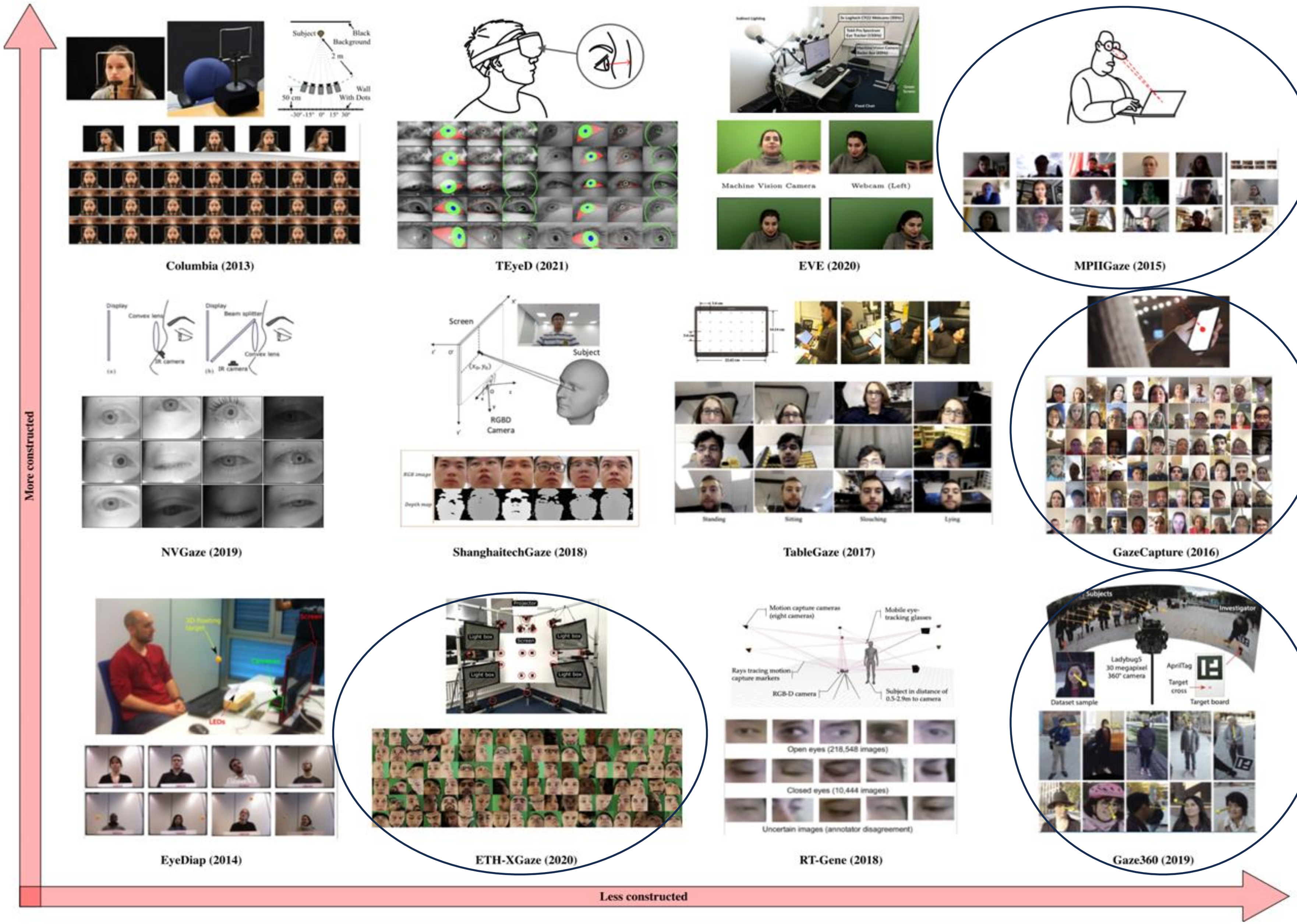


# Datasets

Lab setting vs in-the-wild

# Datasets

Lab setting vs in-the-wild



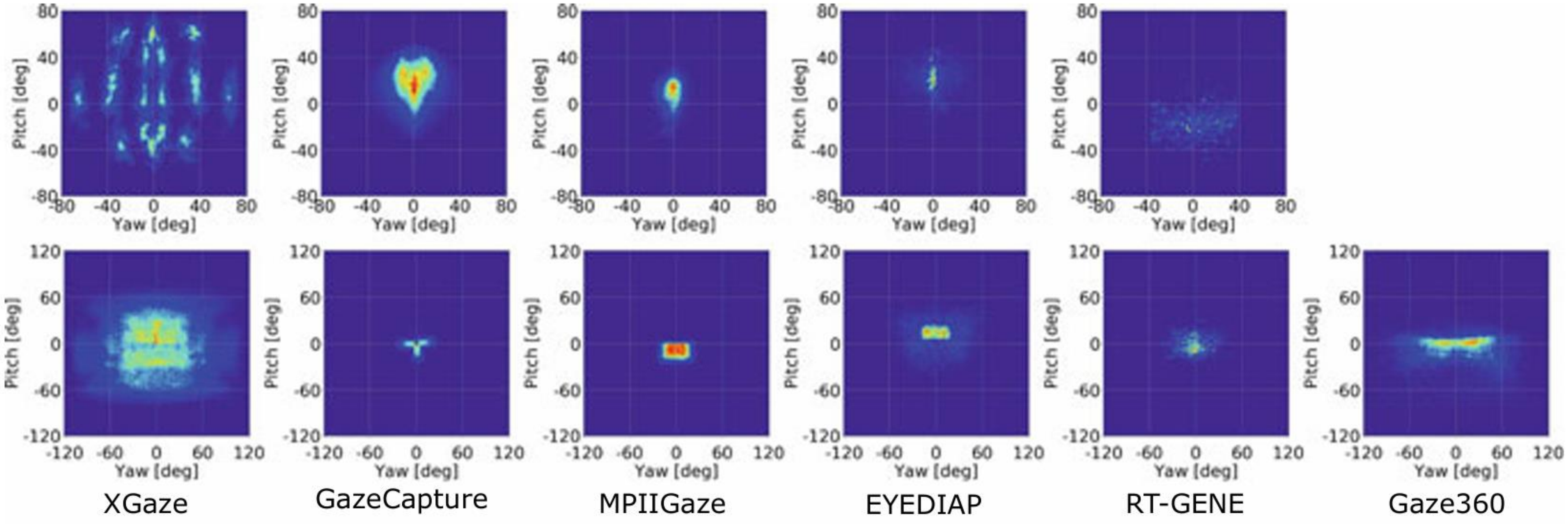
# Gaze Plot of various datasets

How extensively do these gaze angles cover real world in each dataset?

# Gaze Plot of various datasets

Head Pose

Gaze Capture




Area covered by these models are show the area of the gaze points collected via each dataset

Eye Gaze Estimation and Its Applications (uni-saarland.de)

# What are some important aspects of dataset for appearance based gaze estimation?

COMPARISON OF 3-D GAZE ESTIMATION DATA SETS

| Dataset | Channel | Full Face | Camera Number | Total | Subject | Distance | Gaze Pitch/Yaw | Head Pose Pitch/Yaw |
|---|---|---|---|---|---|---|---|---|
| MPIIGaze[15] | RGB | No | 1 | 213,659 | 15 | 40-60 cm | P[-48.94°, 13.05°] Y[-52.38°, 44.07°] | P[-51.39°, 62.14°] Y[-73.14°, 73.14°] |
| MPIIFACEGaze[28] | RGB | Yes | 1 | 213,659 | 15 | 40-60 cm | P[-39.25°, 15.18°] Y[-45.26°, 36.99°] | P[-45.58°, 66.32°] Y[-81.71°, 101.93°] |
| ColumbiaGaze[77] | RGB | Yes | 5 | 5880 | 56 | 200 cm | P[-15.00°, 15.00°] Y[-10.00°, 10.00°] | P[-00.00°, 00.00°] Y[-30.00°, 30.00°] |
| EYEDIAP[78] | RGB-D | Yes | 2 | 94 videos | 16 | 80-120 cm | P[-84.68°, 66.31°] Y[-162.28°, 168.91°] | P[-48.95°, 66.68°] Y[-95.99°, 94.80°] |
| Gaze360°[74] | RGB | Yes | 5 | ≈172K | 238 | ≈ 220 cm | P[-78.22°, 89.26°] Y[-179.88°, 179.88°] | - |
| NISLGaze[41] | RGB | Yes | 1 | 2079 videos | 21 | 90 cm | P[-21.48°, 20.76°] Y[-21.25°, 21.04°] | - |
| RT-GENE[55] | RGB-D | Yes | 1 | 277,286 | 15 | ≈ 182 cm | P[-80.76°, 34.29°] Y[-50.89°, 66.69°] | P[-63.39°, 26.70°] Y[-37.49°, 37.50°] |
| ShanghaiTechGaze[56] | RGB | Yes | 3 | 233,796 | 137 | - | P[0.09 cm, 33.44 cm] Y[2.24 cm, 58.36 cm] | - |
| UnityEye[62] | RGB | No | - | user-defined | - | user-defined | user-defined | user-defined |
| UT-Multiview[5] | RGB | No | 8 | 1,216,000 | 50 | 60 cm | P[-66.48°, 55.43°] Y[-77.92°, 80.80°] | P[-35.51°, 41.93°] Y[-39.90°, 38.78°] |
| XGaze[79] | RGB | Yes | 18 | 1,083,492 | 110 | 100 cm | P[-89.69°, 89.21°] Y[-178.99°, 177.05°] | P[-89.50°, 89.31°] Y[-148.27°, 175.6°] |

Credits: Appearance-based Gaze Estimation with Deep Learning: A Review and Benchmark, Gaze Direction Estimation · google/mediapipe, Tobii Eyetrackers

# What are some important aspects of dataset for appearance based gaze estimation?

COMPARISON OF 3-D GAZE ESTIMATION DATA SETS

| Dataset | Channel | Full Face | Camera Number | Total | Subject | Distance | Gaze Pitch/Yaw | Head Pose Pitch/Yaw |
|---|---|---|---|---|---|---|---|---|
| MPIIGaze[15] | RGB | No | 1 | 213,659 | 15 | 40-60 cm | P[-48.94°, 13.05°] Y[-52.38°, 44.07°] | P[-51.39°, 62.14°] Y[-73.14°, 73.14°] |
| MPIIFACEGaze[28] | RGB | Yes | 1 | 213,659 | 15 | 40-60 cm | P[-39.25°, 15.18°] Y[-45.26°, 36.99°] | P[-45.58°, 66.32°] Y[-81.71°, 101.93°] |
| ColumbiaGaze[77] | RGB | Yes | 5 | 5880 | 56 | 200 cm | P[-15.00°, 15.00°] Y[-10.00°, 10.00°] | P[-00.00°, 00.00°] Y[-30.00°, 30.00°] |
| EYEDIAP[78] | RGB-D | Yes | 2 | 94 videos | 16 | 80-120 cm | P[-84.68°, 66.31°] Y[-162.28°, 168.91°] | P[-48.95°, 66.68°] Y[-95.99°, 94.80°] |
| Gaze360°[74] | RGB | Yes | 5 | ≈172K | 238 | ≈ 220 cm | P[-78.22°, 89.26°] Y[-179.88°, 179.88°] | - |
| NISLGaze[41] | RGB | Yes | 1 | 2079 videos | 21 | 90 cm | P[-21.48°, 20.76°] Y[-21.25°, 21.04°] | - |
| RT-GENE[55] | RGB-D | Yes | 1 | 277,286 | 15 | ≈ 182 cm | P[-80.76°, 34.29°] Y[-50.89°, 66.69°] | P[-63.39°, 26.70°] Y[-37.49°, 37.50°] |
| ShanghaiTechGaze[56] | RGB | Yes | 3 | 233,796 | 137 | - | P[0.09 cm, 33.44 cm] Y[2.24 cm, 58.36 cm] | - |
| UnityEye[62] | RGB | No | - | user-defined | - | user-defined | user-defined | user-defined |
| UT-Multiview[5] | RGB | No | 8 | 1,216,000 | 50 | 60 cm | P[-66.48°, 55.43°] Y[-77.92°, 80.80°] | P[-35.51°, 41.93°] Y[-39.90°, 38.78°] |
| XGaze[79] | RGB | Yes | 18 | 1,083,492 | 110 | 100 cm | P[-89.69°, 89.21°] Y[-178.99°, 177.05°] | P[-89.50°, 89.31°] Y[-148.27°, 175.6°] |

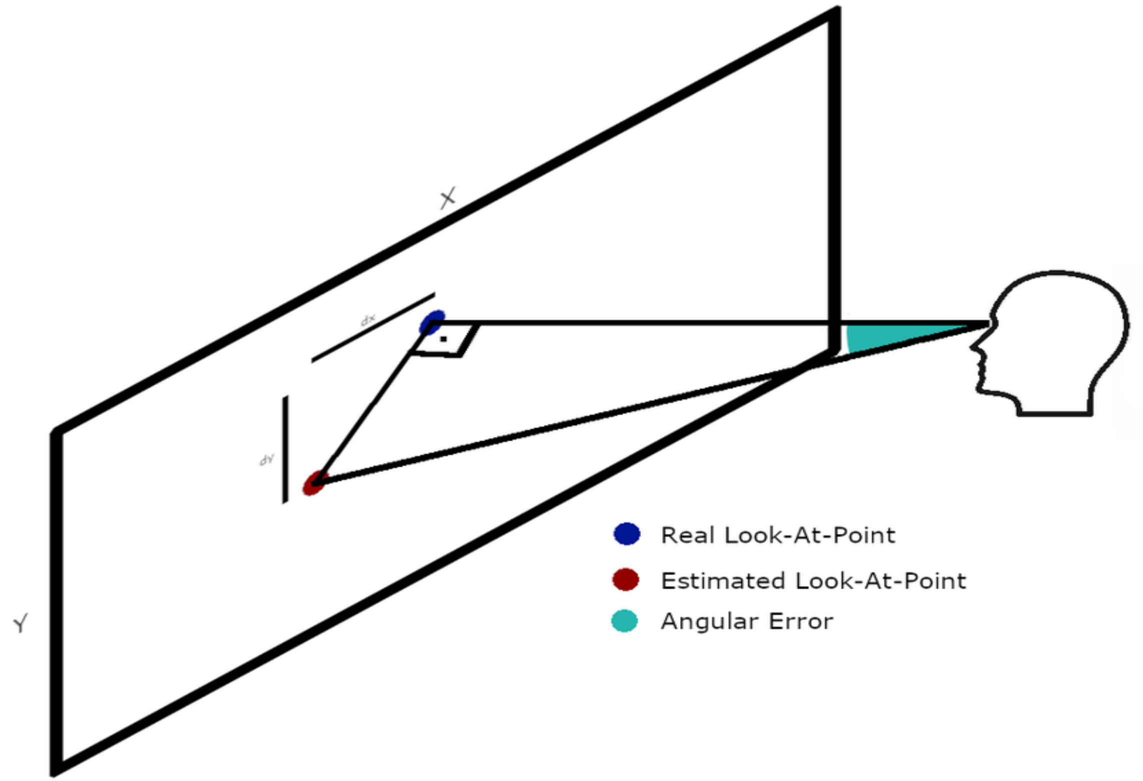
Synthetic Datasets

Credits: Appearance-based Gaze Estimation with Deep Learning: A Review and Benchmark, Gaze Direction Estimation · google/mediapipe, Tobii Eyetrackers

# What are we going to discuss?

- ~~Types of Gaze Estimation~~
- ~~Datasets~~
- Evaluation Metrics of Gaze Estimation
- Appearance Based Gaze Estimation (AGE)
- End-to-End Frame-to-Gaze Estimation (EFE)
- Real World Case Study of Appearance Based Gaze Estimation

# Evaluation Metrics of Gaze Estimation



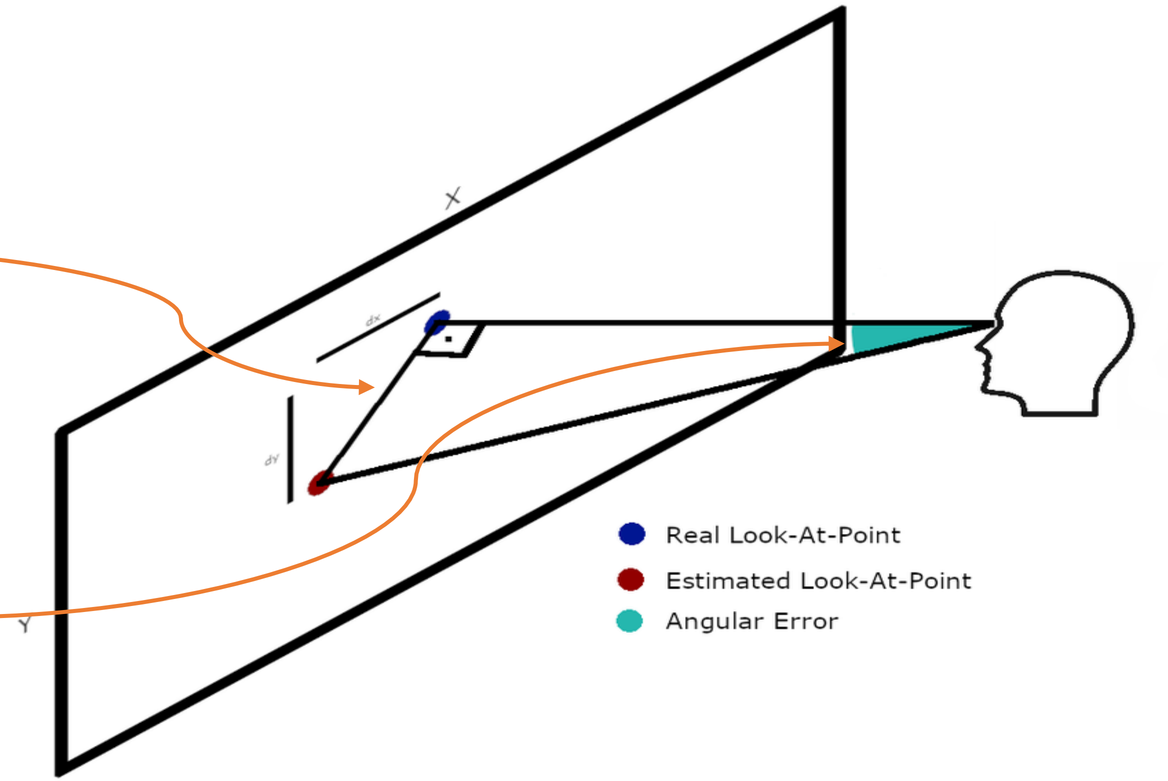
# Evaluation Metrics of Gaze Estimation

## L2 Distance:

The primary metric used to evaluate the accuracy of 2D gaze point estimation

## Angular Error:

The predicted gaze direction vector is produced by connecting the head point to the predicted gaze point



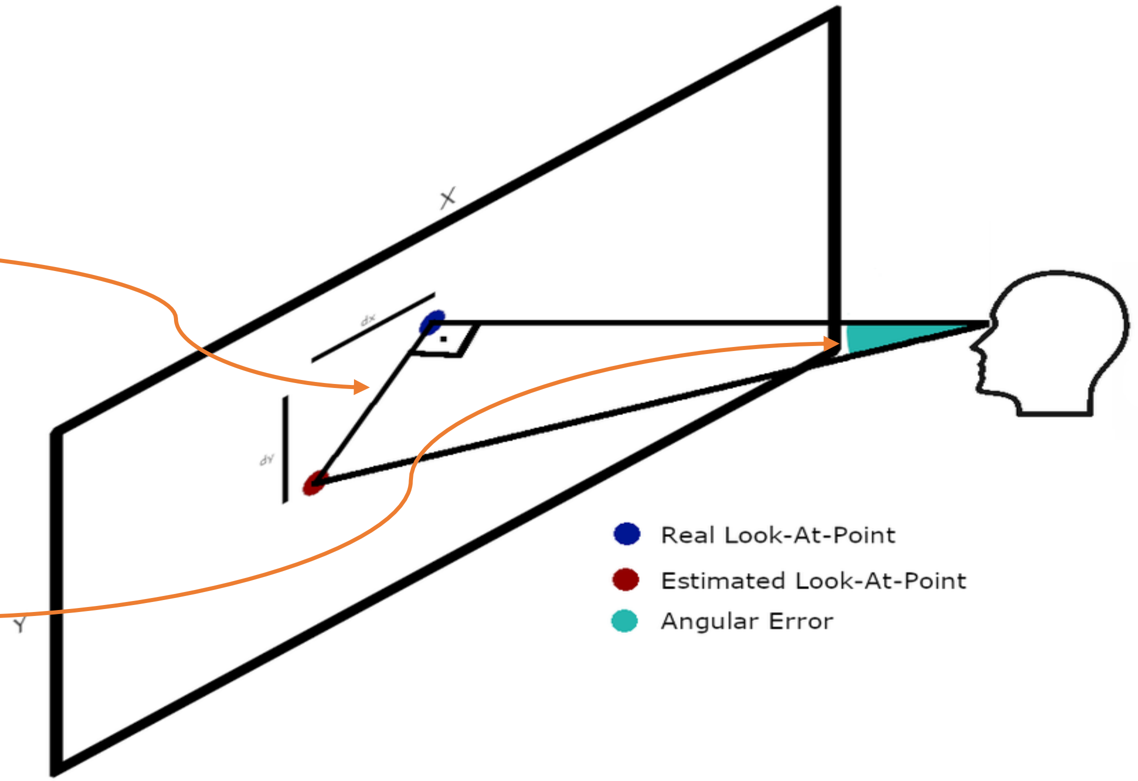
# Evaluation Metrics of Gaze Estimation

## L2 Distance:

The primary metric used to evaluate the accuracy of 2D gaze point estimation

## Angular Error:

The predicted gaze direction vector is produced by connecting the head point to the predicted gaze point




## Other Metrics:

1. Mean Angular Error (MAE)
2. Mean Square Error (MSE)
3. Average Precision (AP)
4. Classification Accuracy

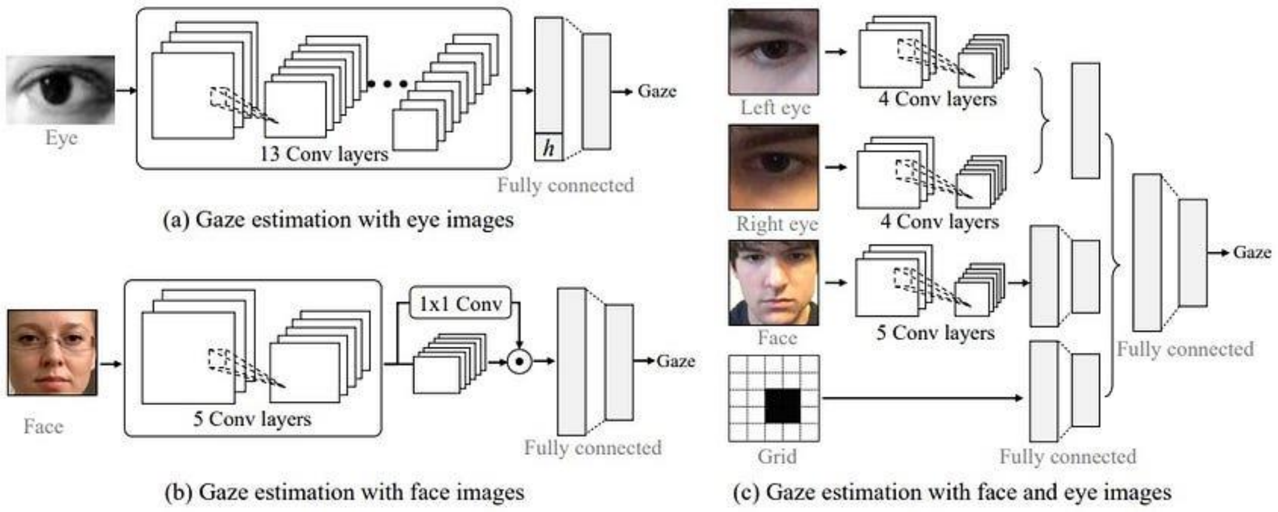
# What are we going to discuss?

- ~~Types of Gaze Estimation~~
- ~~Datasets~~
- ~~Evaluation Metrics of Gaze Estimation~~
- <u>A</u>ppearance Based <u>G</u>aze <u>E</u>stimation (AGE)
- <u>E</u>nd-to-End <u>F</u>rame-to-Gaze <u>E</u>stimation (EFE)
- Real World Case Study of Appearance Based Gaze Estimation

# Appearance Based Gaze Estimation

**What can Model adapt to?**
- Person specific
- Person Independent

**Calibrations Requirement**
1. No calibration
2. One point calibration
3. Multiple point calibration

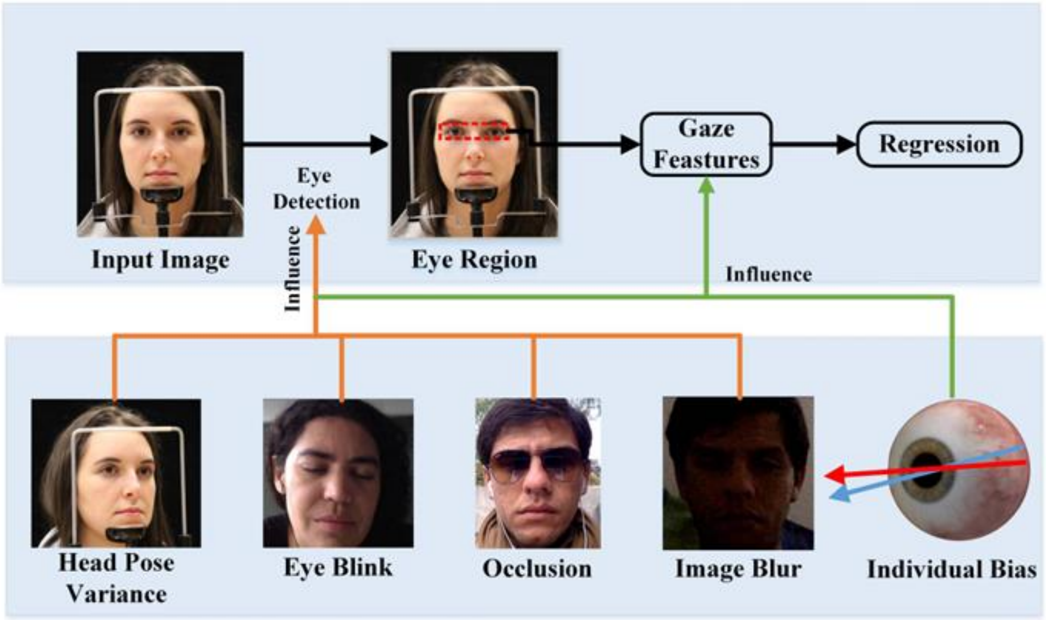



Credits: Appearance-based Gaze Estimation with Deep Learning: A Review and Benchmark, Gaze Direction Estimation · google/mediapipe, Tobii Eyetrackers

# What are some of the problems in Appearance Based Gaze Estimation?

Credits: Appearance-based Gaze Estimation with Deep Learning: A Review and Benchmark, Gaze Direction Estimation · google/mediapipe, Tobii Eyetrackers

# Problems in Appearance based Gaze Estimation




Overview of a simple gaze estimation which consist of eye detection, feature extraction, and gaze regression. Challenges of gaze estimation are head pose variance, eye blink, occlusion, blur images and individual bias.

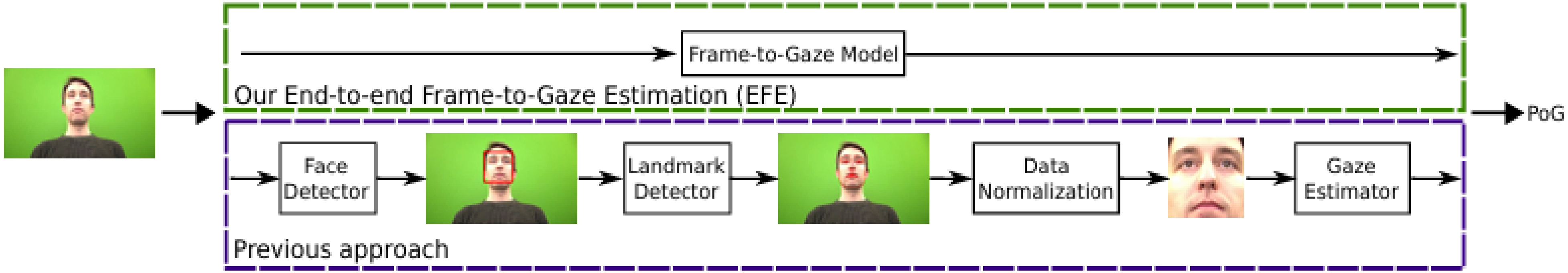
Credits: Appearance-based Gaze Estimation with Deep Learning: A Review and Benchmark, Gaze Direction Estimation · google/mediapipe, Tobii Eyetrackers

# What are we going to discuss?

- ~~Types of Gaze Estimation~~
- ~~Datasets~~
- ~~Evaluation Metrics of Gaze Estimation~~
- ~~Appearance Based Gaze Estimation (AGE)~~
- End-to-End Frame-to-Gaze Estimation (EFE)
- Real World Case Study of Appearance Based Gaze Estimation

# EFE: End to end frame to gaze estimation

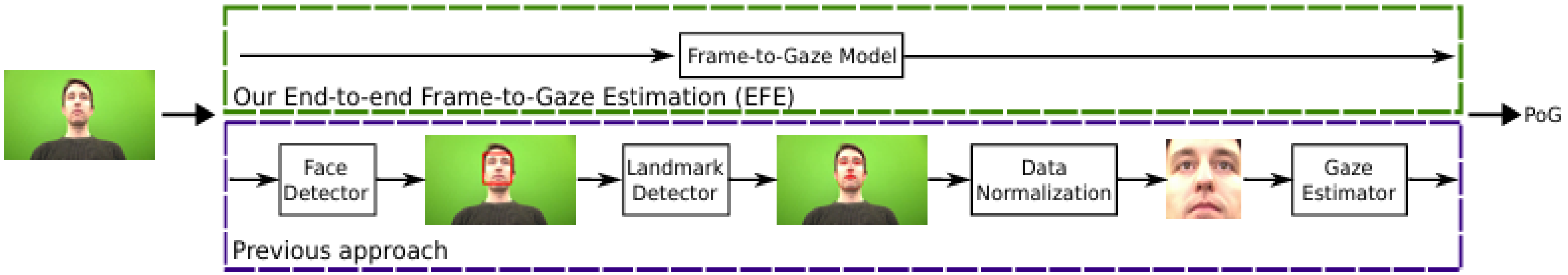
New approach in Gaze estimation = Forget the old approach completely




[2305.05526] EFE: End-to-end Frame-to-Gaze Estimation - arXiv.org

# EFE: End to end frame to gaze estimation

New approach in Gaze estimation = Forget the old approach completely




## Goal of the new approach

directly regresses a 6D gaze ray i.e. (3D origin and 3D direction)
Use frames of the images without preprocessing as the model input

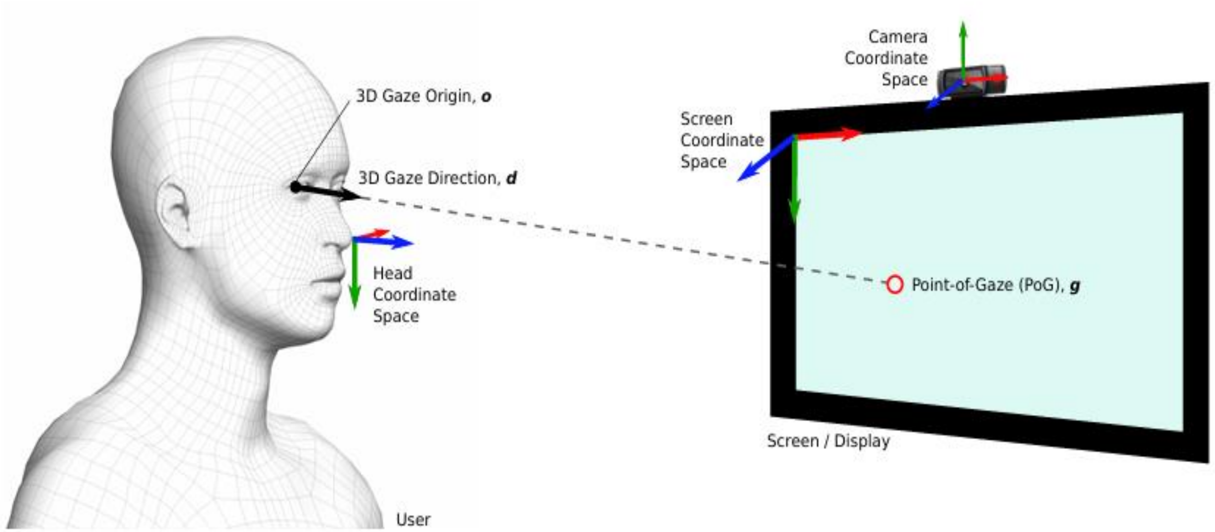
## Main Challenges in new approach

Small Eye Region
Gaze Origin Estimation

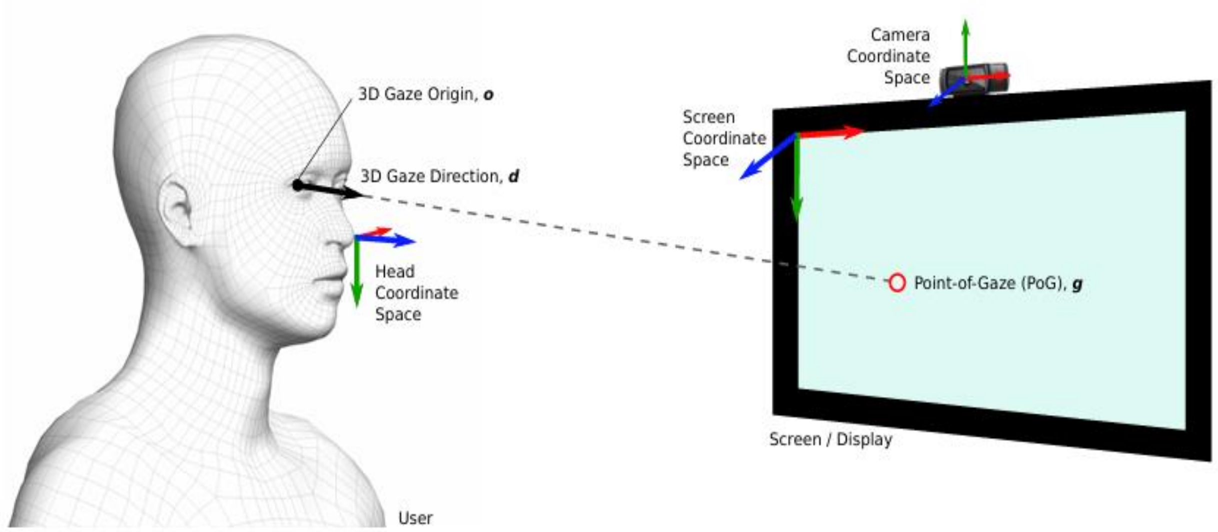
How did we get there?

[2305.05526] EFE: End-to-end Frame-to-Gaze Estimation - arXiv.org

Looking at the traditional approach using the webcam…..



# Looking at the traditional approach using the webcam…..




**Main assumptions of the approach:**

- Person head pose is co-planar to the screen i.e it always faces the screen
- Calculating the 3D head translation estimation from a 2D image without any error
- Diverse head pose with corresponding gaze points calculated
- Ground truth of the datasets are perfect

Before we go into the related works….


Prof. Danna asks :

**How do you find the Gaze Direction for each person in the image?**

[2204.09480] GazeOnce: Real-Time Multi-Person Gaze Estimation


Prof. Danna asks :
**How do you find the Gaze Direction for each person in the image?**

I pointed them out like this.

[2204.09480] GazeOnce: Real-Time Multi-Person Gaze Estimation


Prof. Danna asks :

**How do you find the Gaze Direction for each person in the image?**

I pointed them out like this.

Prof. Danna asks :

**How would you find it doing the current approaches in computer vision that was taught in the class?**

[2204.09480] GazeOnce: Real-Time Multi-Person Gaze Estimation


Prof. Danna asks :
**How do you find the Gaze Direction for each person in the image?**

I pointed them out like this.

Prof. Danna asks :
**How would you find it doing the current approaches in computer vision that was taught in the class?**

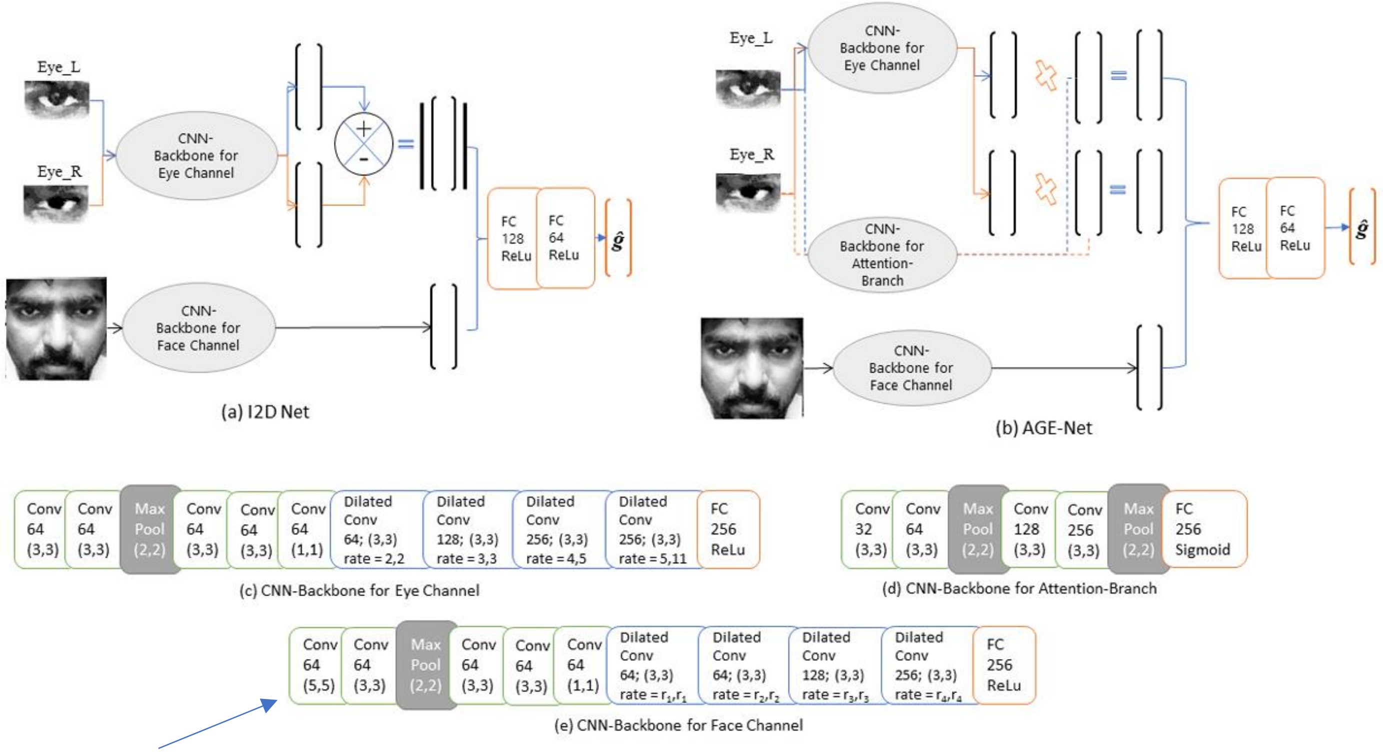
Vision Transformers ? Attention Models ?
Feature pyramids ?  Larger Datasets ?

Utterly Confused?

[2204.09480] GazeOnce: Real-Time Multi-Person Gaze Estimation

# CNN based learning

**How do we extend the CNN's learning capabilities?**




Modified version of AlexNet

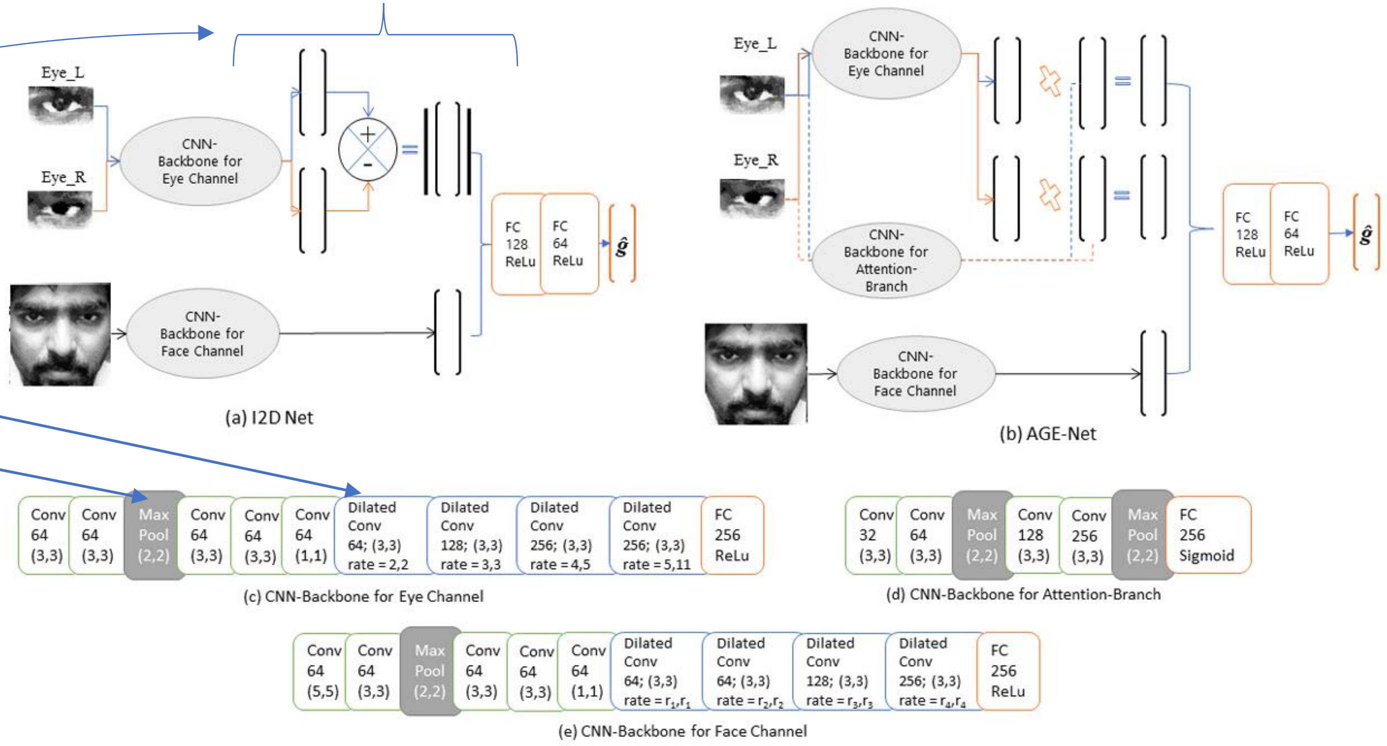
Appearance-based Gaze Estimation using Attention and Difference Mechanism, CVPR 2021

# CNN based learning

**How do we extend the CNN's learning capabilities?**

1. Learning Eye Independent features

2. Modification of model layers
- Diluted Convolution
- Global Max Pooling
- Ensemble model approach



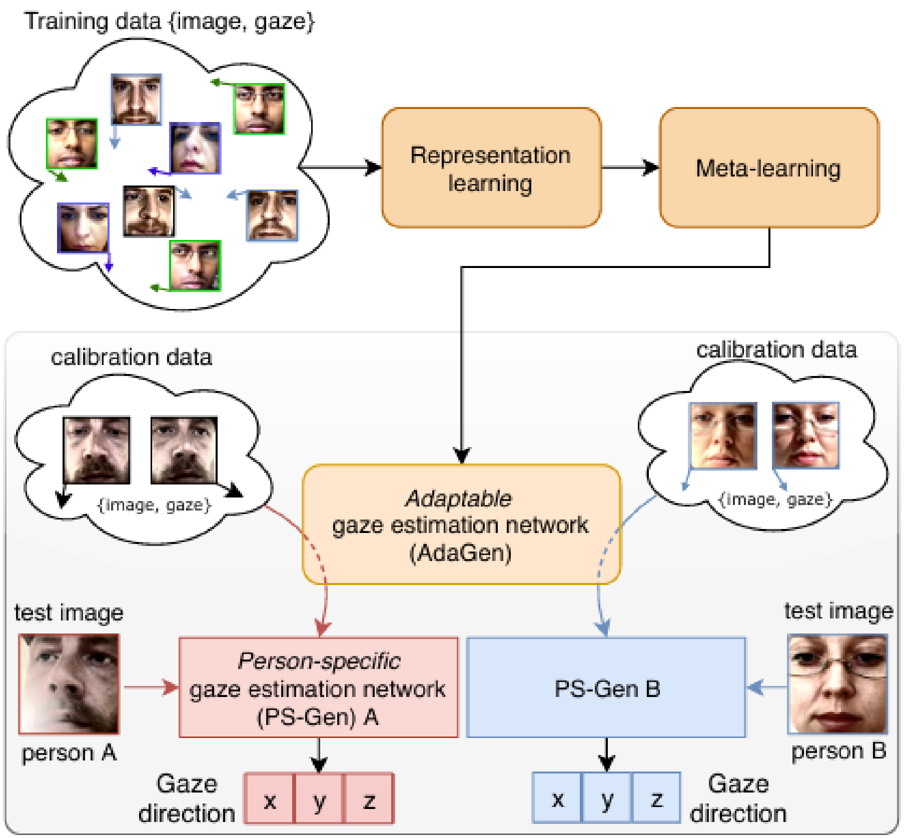

| Model | MPIIGaze | RT-Gene |
|---|---|---|
| iTracker (AlexNet) [23] | 5.6° | - |
| MeNet [35] | 4.9° | - |
| Spatial-Weights CNN [41] | 4.8° | 10.0° |
| Dilated-Net [9] | 4.8° | - |
| RT-GENE (1 model) [16] | 4.8° | - |
| RT-GENE (4 model) [16] | 4.3° | 8.6° |
| FAR* Net [12] | 4.3° | 8.4° |
| Bayesian Approach [34] | 4.3° | - |
| I2D-Net (**Proposed**) | **4.3°** | **8.44°** |
| AGE-Net (**Proposed**) | **4.09°** | **7.44°** |

Lower is better

Modified version of AlexNet

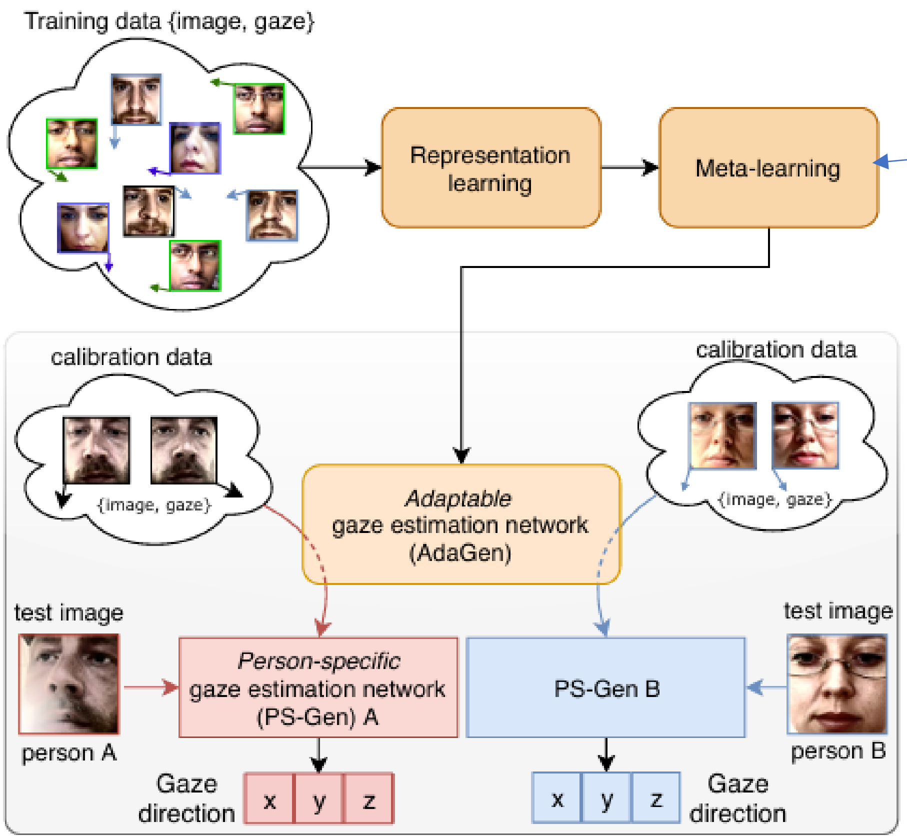
[2104.12668] Appearance-based Gaze Estimation With Deep ...

# Few-Shot Adaptive Gaze Estimation




[1905.01941] Few-Shot Adaptive Gaze Estimation - arXiv.org

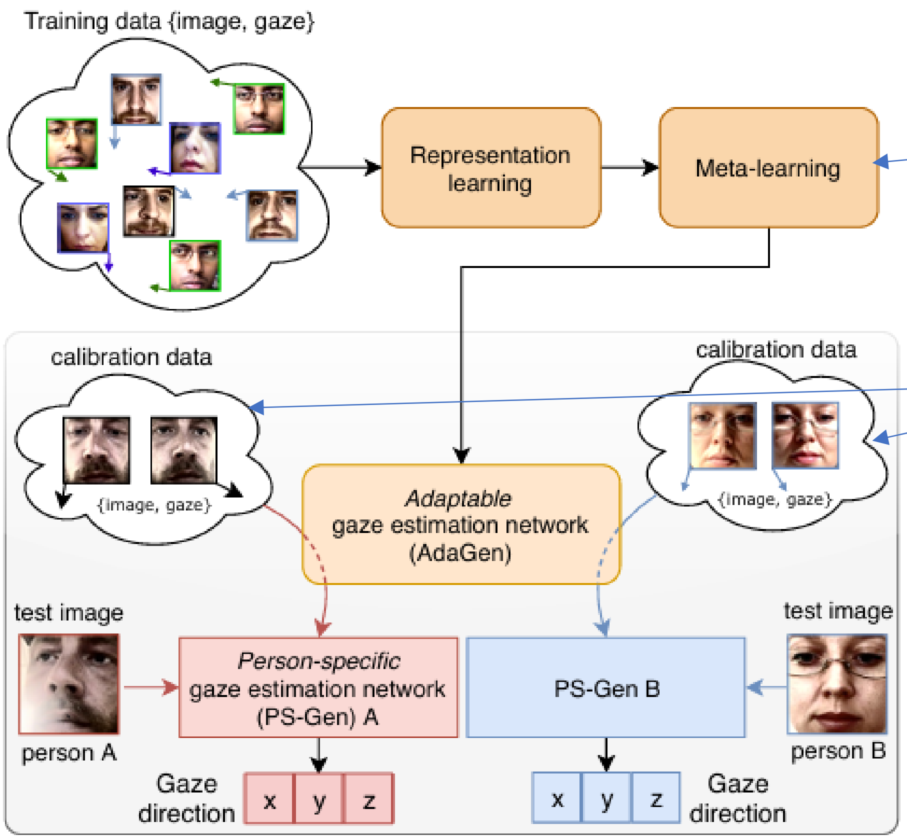
# Few-Shot Adaptive Gaze Estimation




Model-Agnostic Meta-Learning(MAML)

[1905.01941] Few-Shot Adaptive Gaze Estimation - arXiv.org

# Few-Shot Adaptive Gaze Estimation

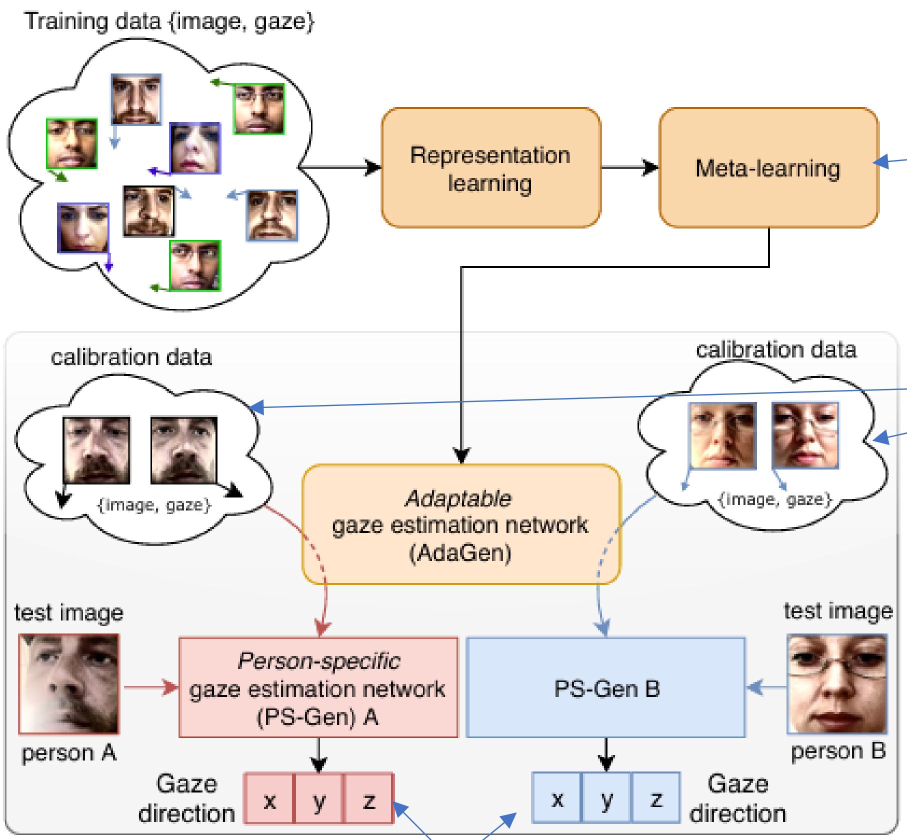



Model-Agnostic Meta-Learning(MAML)

Fine-tuning to person-specific

[1905.01941] Few-Shot Adaptive Gaze Estimation - arXiv.org

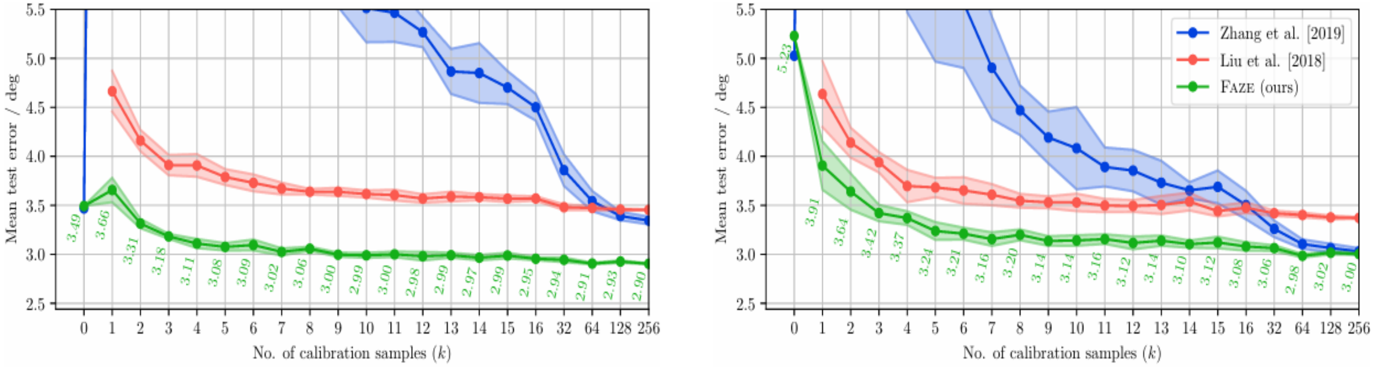
# Few-Shot Adaptive Gaze Estimation




Model-Agnostic Meta-Learning(MAML)

Fine-tuning to person-specific

Focus on Gaze direction

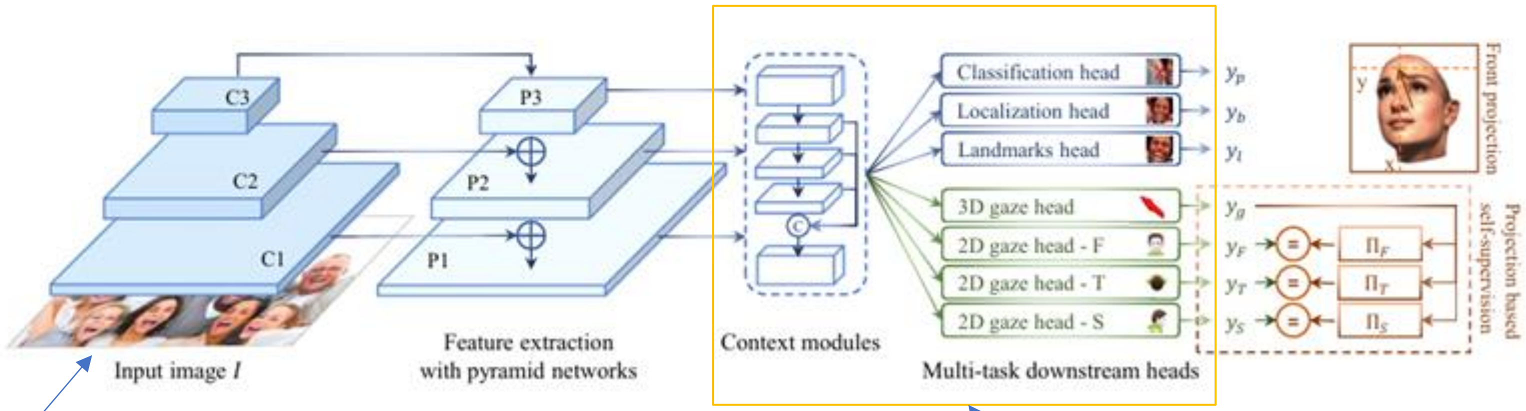



(a) GazeCapture (test)

(b) MPIIGaze

[1905.01941] Few-Shot Adaptive Gaze Estimation - arXiv.org

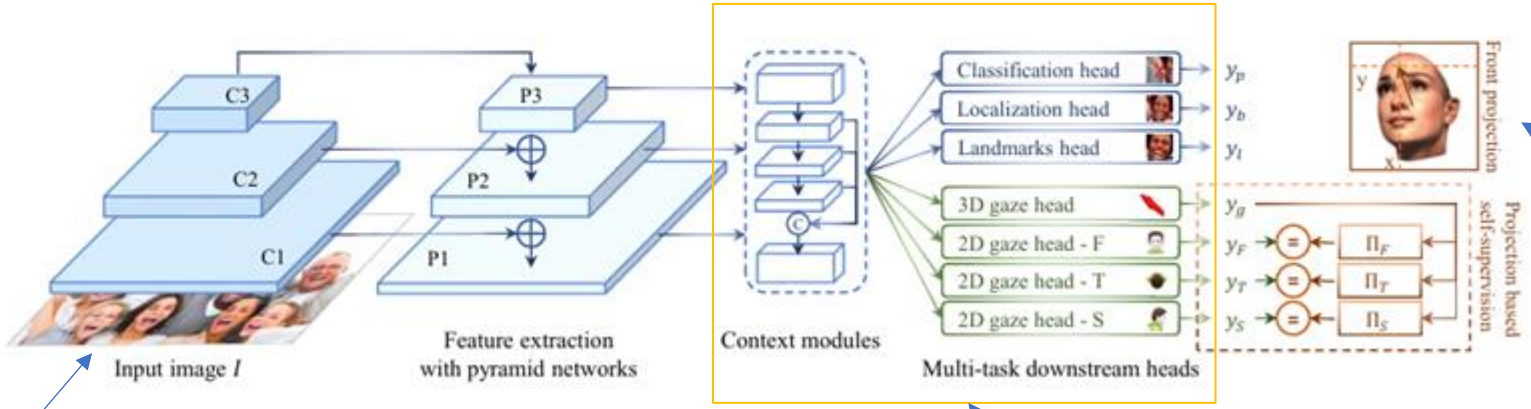
# Gaze Once




Frame as Input

optimizing face localization and gaze estimation

# Gaze Once




Focus on Gaze Direction

Frame as Input

optimizing face localization and gaze estimation

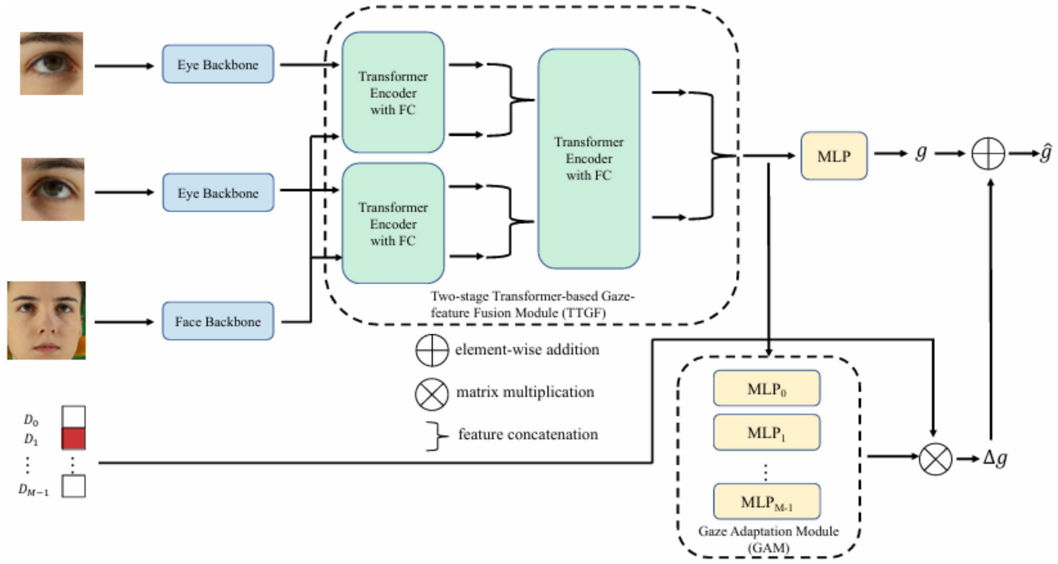
| Method | Backbone | Input | Gaze error (lower is better) w.r.t. **the width of faces** | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | 30-60 | 60-90 | 90-120 | 120-150 | 150-180 | 180-210 | 210-240 | >240 |
| Full-face | AlexNet | 1 normalized face | 24.99 | 20.00 | 17.56 | 17.03 | 16.47 | 14.74 | 13.43 | 12.31 |
| ETH-18 | ResNet18 | 1 normalized face | 28.89 | 21.93 | 16.66 | 14.90 | 14.33 | 12.44 | 11.68 | 10.32 |
| ETH-50 | ResNet50 | 1 normalized face | 29.82 | 21.87 | 16.93 | 14.76 | 13.87 | 11.79 | 11.13 | 9.98 |
| GazeTR | ResNet18 | 1 normalized face | 24.51 | <u>16.84</u> | 14.59 | 13.37 | 13.65 | 11.72 | 10.71 | 9.96 |
| Ours | MobileNet0.25 | 1 full image | <u>22.94</u> | 17.55 | <u>13.69</u> | <u>11.08</u> | <u>11.13</u> | <u>9.41</u> | <u>8.17</u> | <u>7.74</u> |
| Ours | ResNet50 | 1 full image | **21.17** | **13.77** | **10.58** | **7.9** | **8.57** | **6.68** | **6.01** | **5.56** |

[2204.09480] GazeOnce: Real-Time Multi-Person Gaze Estimation


[2204.09480] GazeOnce: Real-Time Multi-Person Gaze Estimation

# Transformers with Gaze Adaptation Module

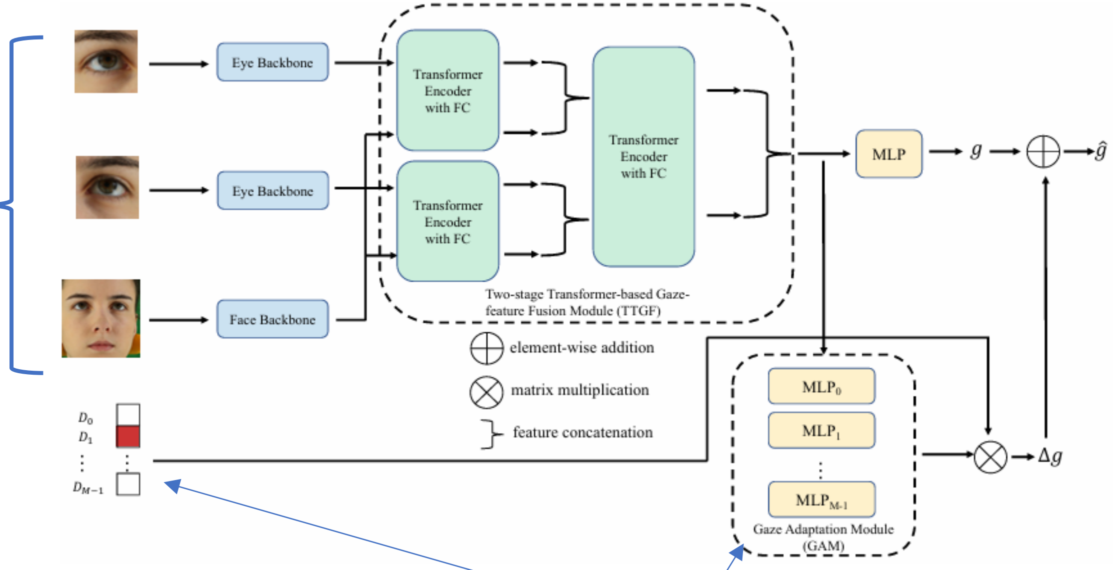
Two-stage Transformer-based Gaze-feature Fusion




Merging Multiple Datasets for Improved Appearance-Based Gaze ...

# Transformers with Gaze Adaptation Module

Two-stage Transformer-based Gaze-feature Fusion




| Model | Transformer | Feature Fusion | MPIIFaceGaze | RT-GENE | EYEDIAP |
|---|---|---|---|---|---|
| FullFace [8] | NO | Face Only | 4.93° | 10.00° | 6.53° |
| RT-GENE [10] | NO | Two Eyes | 4.66° | 8.00° | 6.02° |
| DilatedNet [9] | NO | PAR | 4.42° | 8.38° | 6.19° |
| iTracker [12] | NO | LR-EH | 4.33° | 7.12° | 5.28° |
| iTracker-MHSA [21] | YES | LR-EH | 4.05° | 7.06° | 5.17° |
| GazeTR-Hybrid [19] | YES | Face Only | 4.18° | 7.12° | 5.33° |
| GazeCADSE [42] | YES | Face Only | 4.04° | 7.00° | 5.25° |
| **Proposed** | YES | EH-LR | **3.88°** | **6.46°** | **4.89°** |

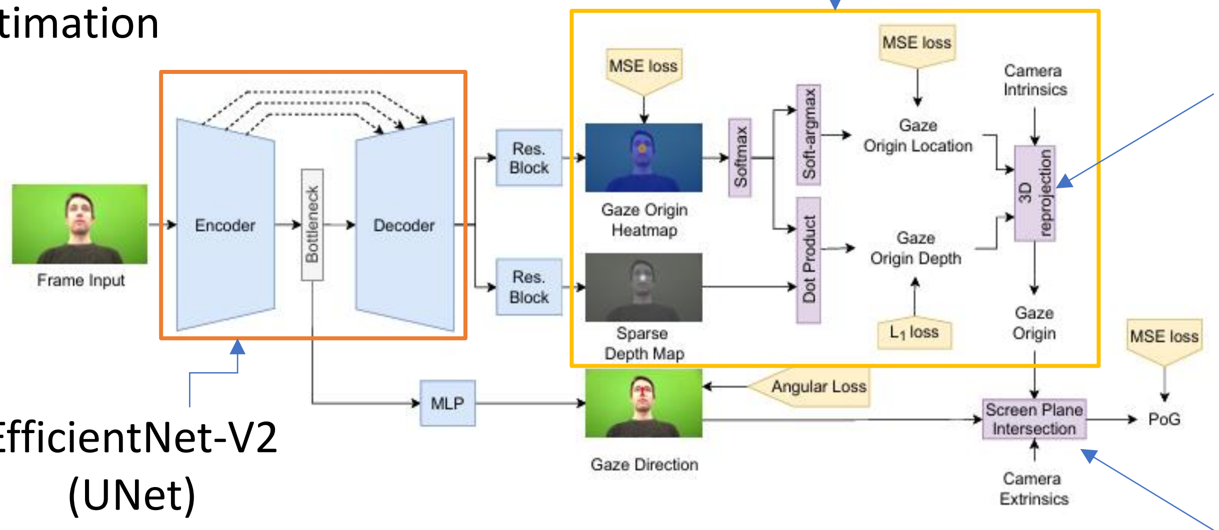
Feature Extraction from multiple datasets

Merging Multiple Datasets for Improved Appearance-Based Gaze ...

# What are we still missing?

# End-to-End Frame-to-Gaze Estimation (EFE)

Learning-based PoG Estimation

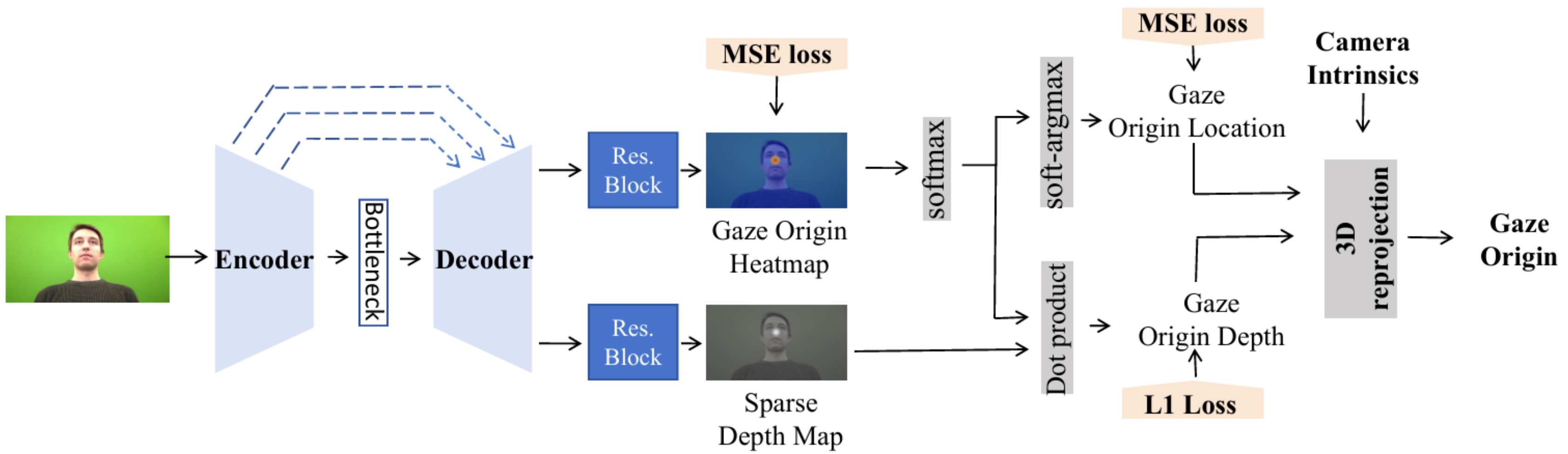



Leveraging camera intrinsic properties directly

EfficientNet-V2 (UNet)

Leveraging camera intrinsic properties directly

# Predicting Gaze Origin




Heatmap Prediction:

$$\mathcal{L}_{\textbf{heatmap}} = \frac{1}{n}\sum_{i=1}^{n}\|\mathbf{h} - \hat{\mathbf{h}}\|_2^2,$$

Sparse depth map loss:

$$\boldsymbol{z} = \boldsymbol{h} \cdot \boldsymbol{d} \qquad \mathcal{L}_d = \|\mathbf{z} - \hat{\mathbf{z}}\|_1,$$

2D location of gaze origin loss:

$$\mathcal{L}_{\mathbf{g}} = \|\mathbf{g} - \hat{\mathbf{g}}\|_2^2,$$

**h** : The 2D gaze origin heatmap
**d** : The Sparse depth map
**g** : 2D position of gaze origin predicted using h
**z** : The gaze depth
$\hat{\mathbf{h}}$ : The ground truth heatmap generated by drawing a 2D Gaussian centered at the gaze origin
$\hat{g}$ : The ground truth gaze location on the cameraframe

# Calculating the Origin of 3D Gaze

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

**1. camera intrinsic matrix K**

**2. The 2D Gazing Point Coordinates g = (x, y)**

**3. The 2D Gazing Point Depth d**

**1. Normalize 2D gaze points**

$$u = \frac{x - c_x}{f_x}$$

$$v = \frac{y - c_y}{f_y}$$

**2. Calculate the origin of 3D gaze**

$$\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = K^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \cdot d$$

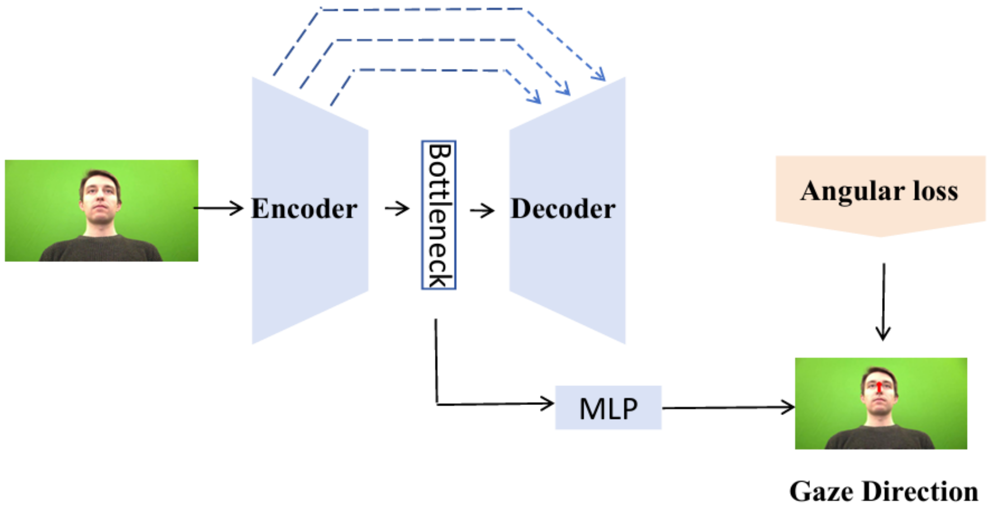
**3. Finally obtaining 3D gaze point o = (x, y, z)**

# Predicting Gaze Direction

Gaze Direction loss:

$$\mathcal{L}_{\mathbf{r}} = \arccos\left(\frac{\hat{\mathbf{r}} \cdot \mathbf{r}}{\|\hat{\mathbf{r}}\|\|\mathbf{r}\|}\right)$$

We can obtain the 3D gaze direction (r).

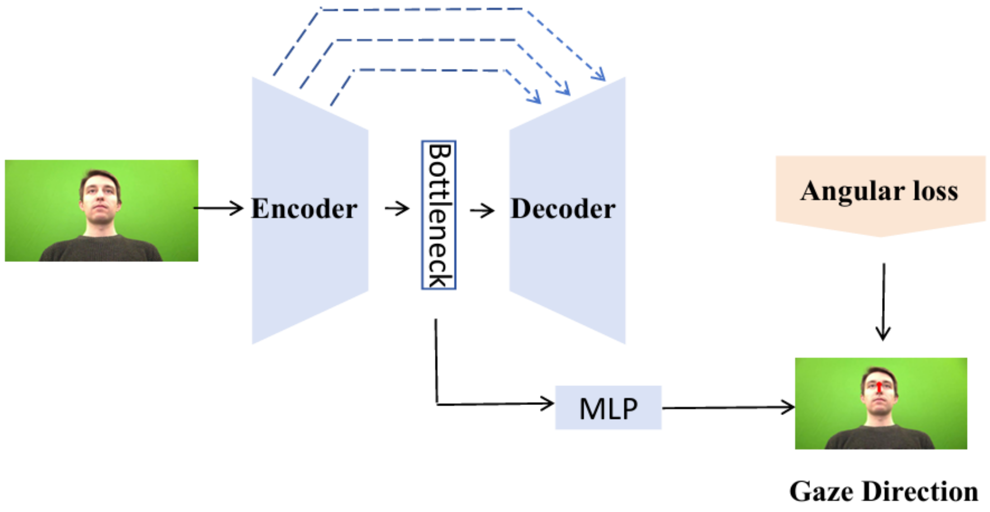


# Predicting Gaze Direction

Gaze Direction loss:

$$\mathcal{L}_{\mathbf{r}} = \arccos\left(\frac{\hat{\mathbf{r}} \cdot \mathbf{r}}{\|\hat{\mathbf{r}}\|\|\mathbf{r}\|}\right)$$

We can obtain the 3D gaze direction (r).

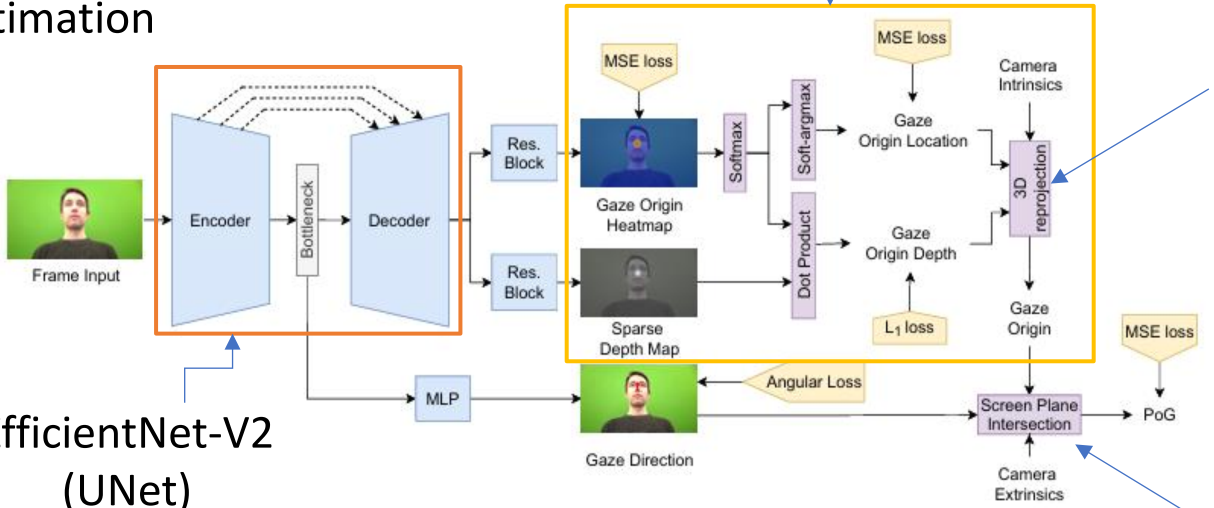



Total Loss

$$\mathcal{L}_{\mathbf{Total}} = \lambda_g \mathcal{L}_{\mathbf{g}} + \lambda_h \mathcal{L}_{\mathbf{heatmap}} + \lambda_d \mathcal{L}_{\mathbf{d}} + \lambda_r \mathcal{L}_{\mathbf{r}} + \lambda_{PoG} \mathcal{L}_{\mathbf{PoG}}$$

# End-to-End Frame-to-Gaze Estimation (EFE)

Learning-based PoG Estimation

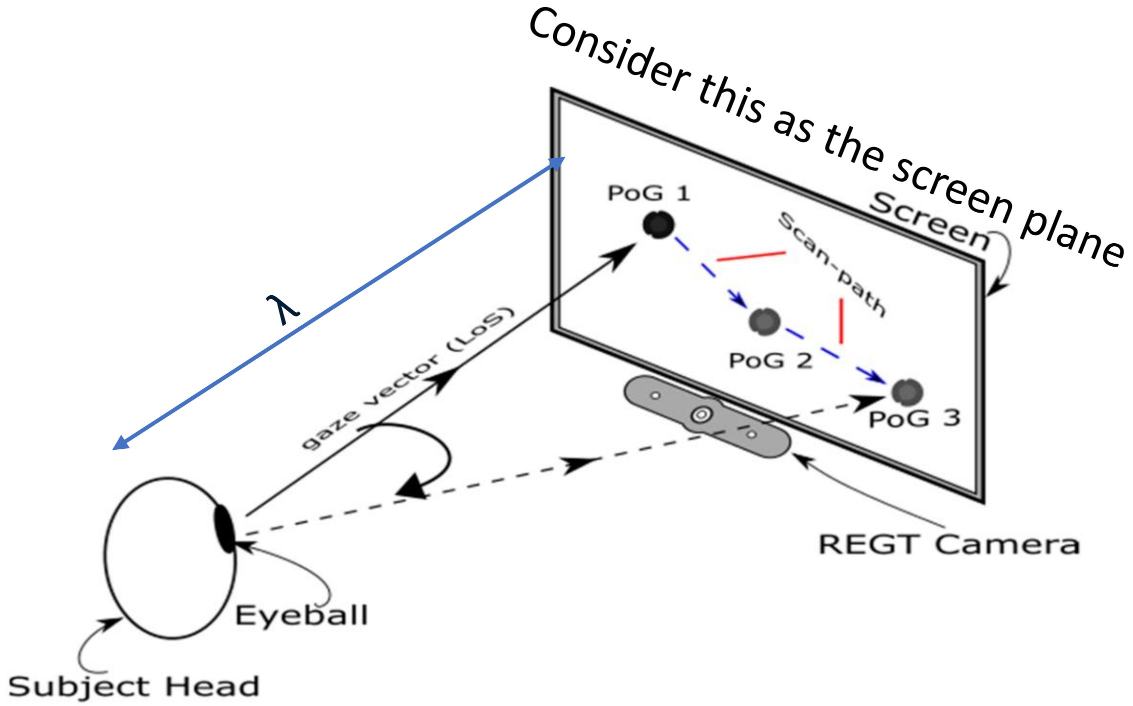



Leveraging camera intrinsic properties directly

EfficientNet-V2 (UNet)

Leveraging camera intrinsic properties directly

# Calculating the Point of Gaze(PoG)

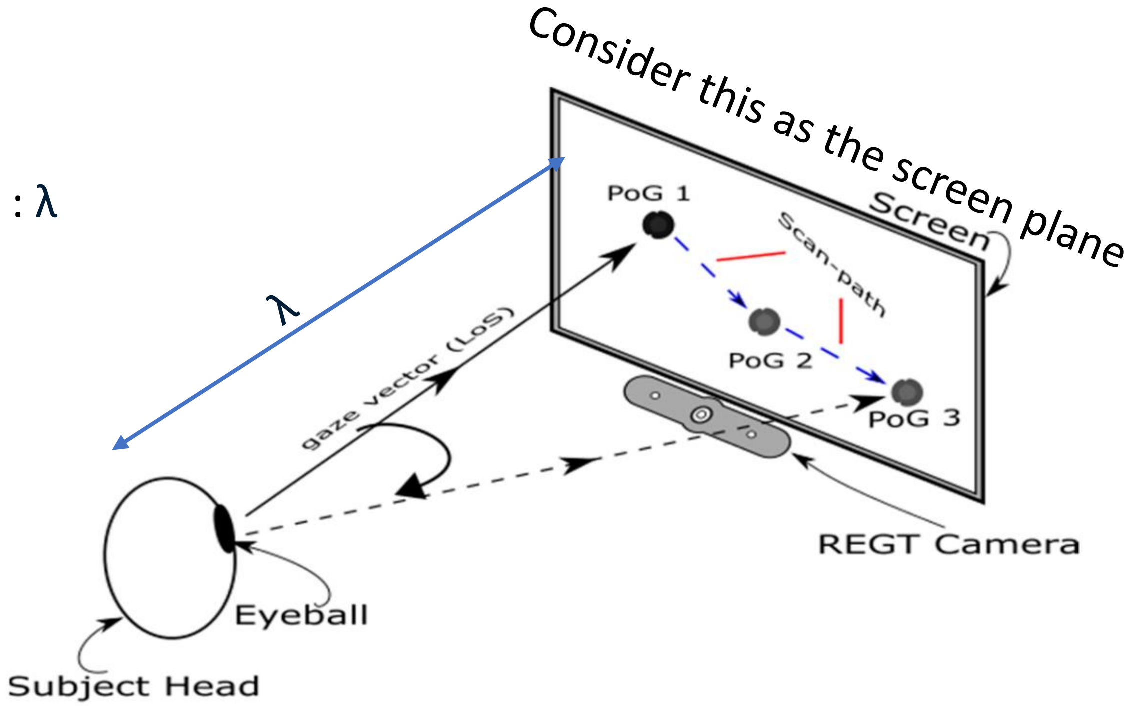



Remote Eye Gaze Tracking Research: A Comparative Evaluation on Past and Recent Progress

# Calculating the Point of Gaze(PoG)

Distance from the screen : λ

$$\lambda = \frac{\boldsymbol{r} \cdot \boldsymbol{n}_s}{(\boldsymbol{a}_s - \boldsymbol{o}) \cdot \boldsymbol{n}_s}$$



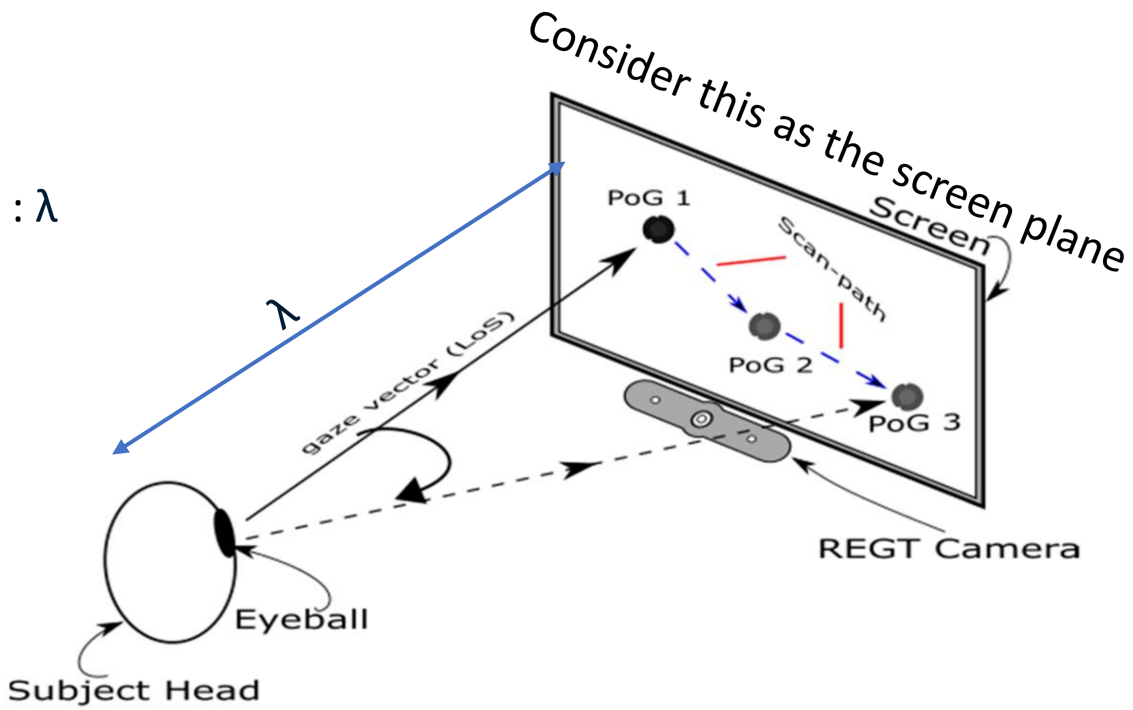

Intersection of the line of sight with the screen plane

$$\boldsymbol{p} = \boldsymbol{o} + \lambda \boldsymbol{r}$$

# Calculating the Point of Gaze(PoG)

Distance from the screen : λ

$$\lambda = \frac{\boldsymbol{r} \cdot \boldsymbol{n_s}}{(\boldsymbol{a_s} - \boldsymbol{o}) \cdot \boldsymbol{n_s}}$$




Intersection of the line of sight with the screen plane

$$\boldsymbol{p} = \boldsymbol{o} + \boldsymbol{\lambda r}$$

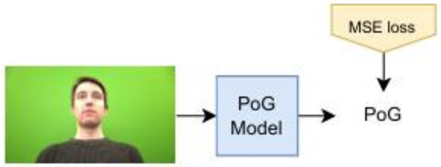
PoG estimation

$$\boldsymbol{L_{PoG}} = ||\boldsymbol{p} - \widehat{\boldsymbol{p}}||_2^2$$

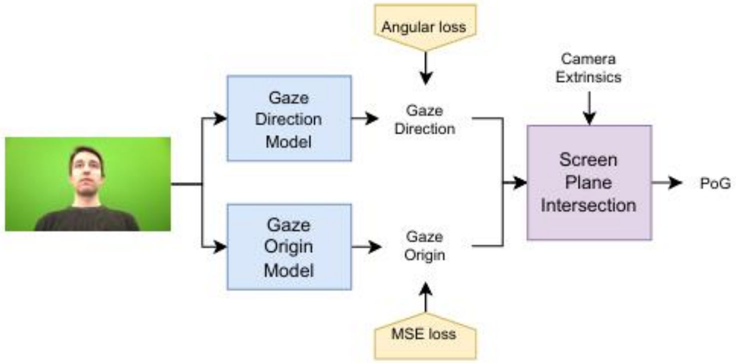
Remote Eye Gaze Tracking Research: A Comparative Evaluation on Past and Recent Progress

# Baseline Models for EFE

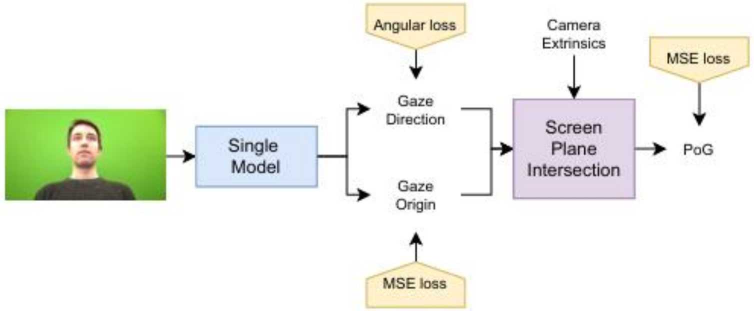
**How to understand the importance of this complex model?**




(a) Direct Regression




(b) Separate Models

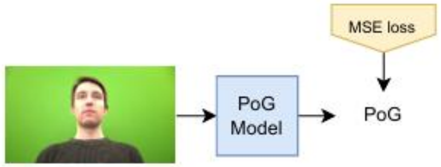



(c) Joint Prediction

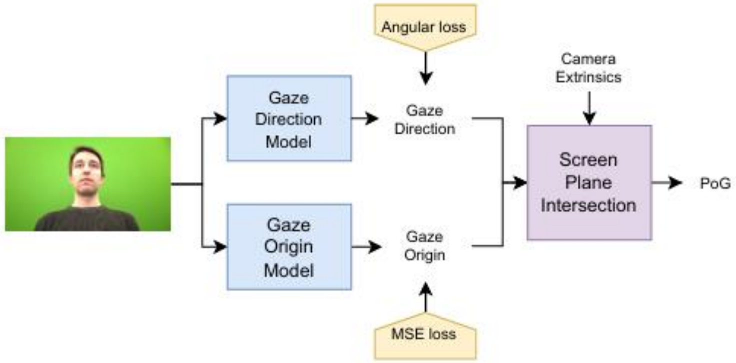
# Baseline Models for EFE

**How to understand the importance of this complex model?**

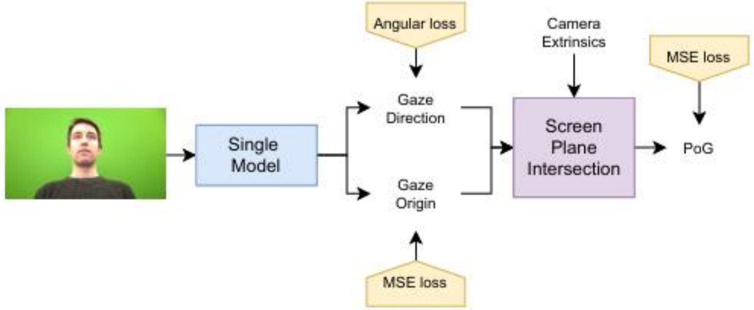
We perform ablation study to understand the model.




(a) Direct Regression




(b) Separate Models




(c) Joint Prediction

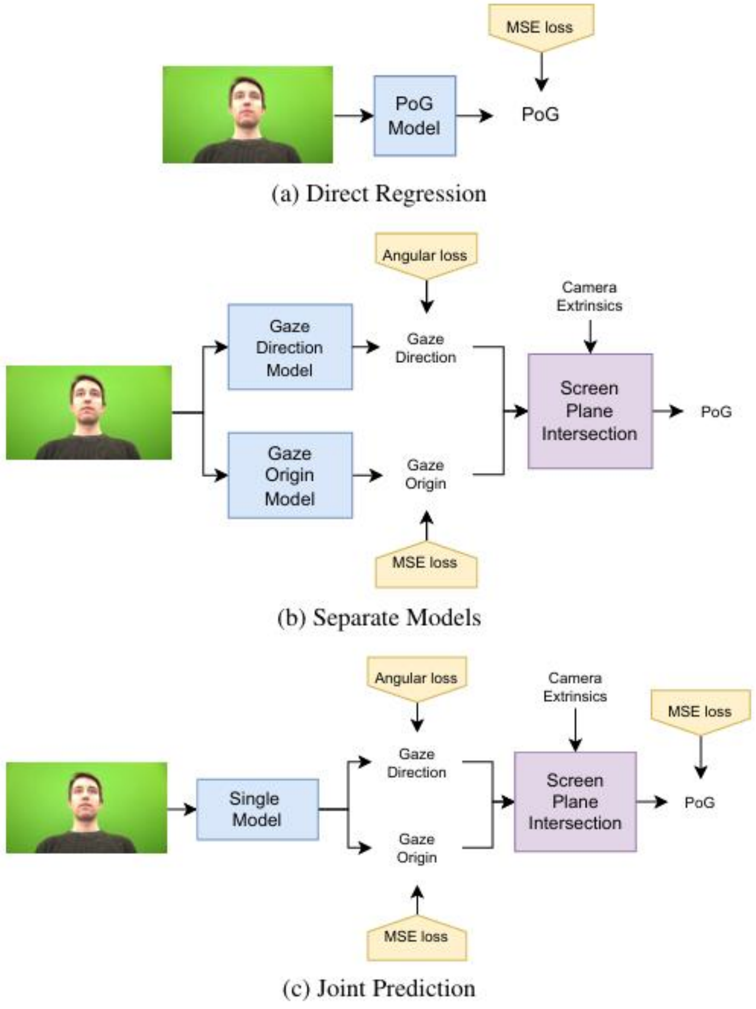
# Baseline Models for EFE

**How to understand the importance of this complex model?**

We perform ablation study to understand the model.

| Model | Heatmap pred. | Depth map pred. | Gaze Origin (mm) | Gaze Dir. (°) | PoG (px) |
|---|---|---|---|---|---|
| Direct Regression | | | - | - | 143.83 |
| Separate Models | | | 16.18 | 3.63 | 141.35 |
| Joint Prediction | | | 20.14 | 3.73 | 143.39 |
| EFE w/o depth map | ✓ | | 18.28 | 3.82 | 146.82 |
| EFE (ours) | ✓ | ✓ | **16.07** | **3.53** | **133.73** |

So, Depth is important feature contributing to PoG estimation



# Importance of Heatmaps


Supervises the distance of the person from screen

# Experimentation

| Model | Inputs | GazeDir.(°) | PoG(px) |
|---|---|---|---|
| EyeNet (static) | Right Eye | 4.75 | 181.0 |
| EyeNet (static) | Left Eye | 4.54 | 172.7 |
| FaceNet | Face | **3.47** | 134.10 |
| EFE(ours) | Frame | 3.53 | **133.73** |

1. Comparison with state-of-the-art methods on the **EVE** dataset.

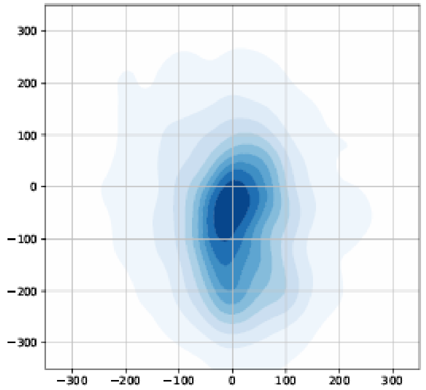
| Model | Inputs | PhonePoG | TabletPoG |
|---|---|---|---|
| iTracker | Face&Eyes | 2.04 | 3.32 |
| iTracker (train aug) | Face&Eyes | 1.86 | 2.81 |
| SAGE | Eyes | 1.78 | 2.72 |
| TAT | Face | 1.77 | 2.66 |
| AFF-Net | Face&Eyes | 1.62 | **2.30** |
| EFE(ours) | Frame | **1.61** | 2.48 |

2. Comparison with state-of-the-art methods on the **GazeCapture** dataset.

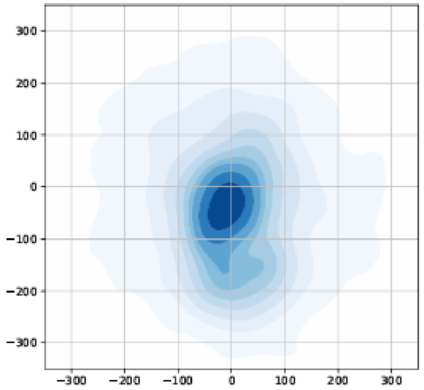
| Model | Input | GazeDir.(°) | PoG(mm) |
|---|---|---|---|
| Full-Face | Face | 4.8 | 42.0 |
| FAR-Net* | Face&Eyes | **4.3** | - |
| AFF-Net | Face&Eyes | 4.4 | 39.0 |
| EFE(ours) | Frame | 4.4 | **38.9** |

3. Comparison with state-of-the-art methods on the **MPIIFaceGaze** dataset.

PoG residual Maps on EVE Dataset




Traditional Approach




New EFE approach

# What are we going to discuss?

- ~~Types of Gaze Estimation~~
- ~~Datasets~~
- ~~Evaluation Metrics of Gaze Estimation~~
- ~~Appearance Based Gaze Estimation (AGE)~~
- ~~End-to-End Frame-to-Gaze Estimation (EFE)~~
- Real World Case Study of Appearance Based Gaze Estimation

Efficient Interaction with Automotive Heads Up Displays using Appearance-based Gaze Tracking, AUTO UI 2021

# User Evaluation

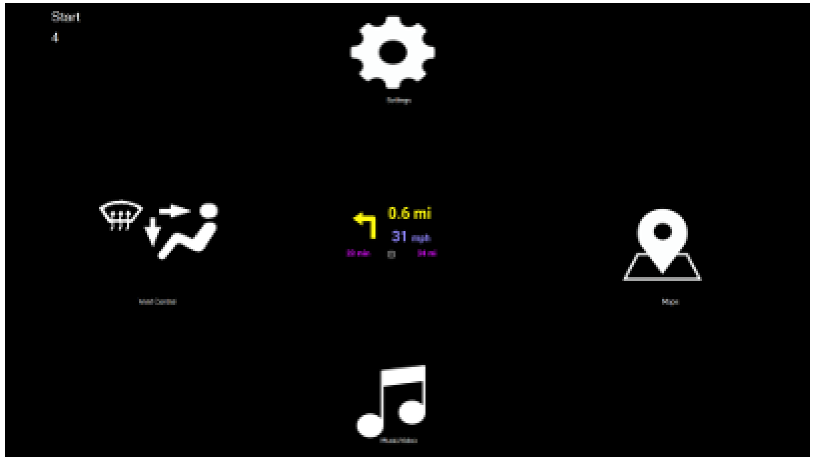



Fig. 3. User Interface Design on Head Up Display

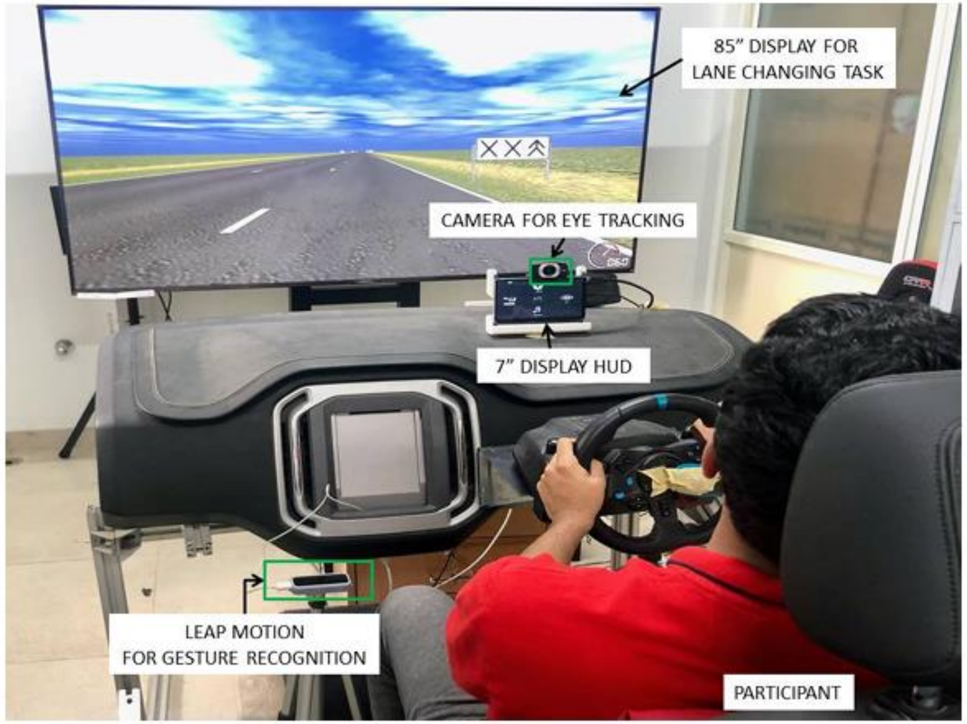
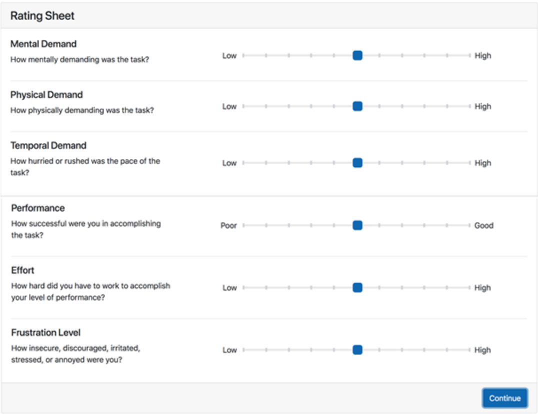



Fig. 2. Illustration of the Experimental Setup

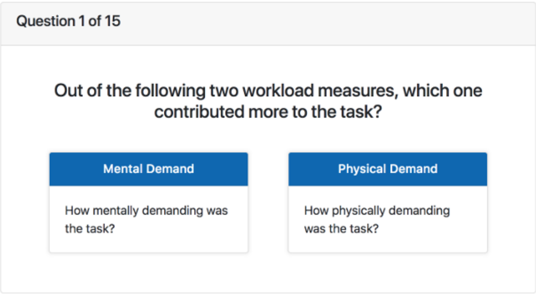
**Study Statistics:**

- 8 (5 Male; 1 left handed, Avg. Age: 24.75yr ) participants (None of the participants had any motor impairment or color blindness)

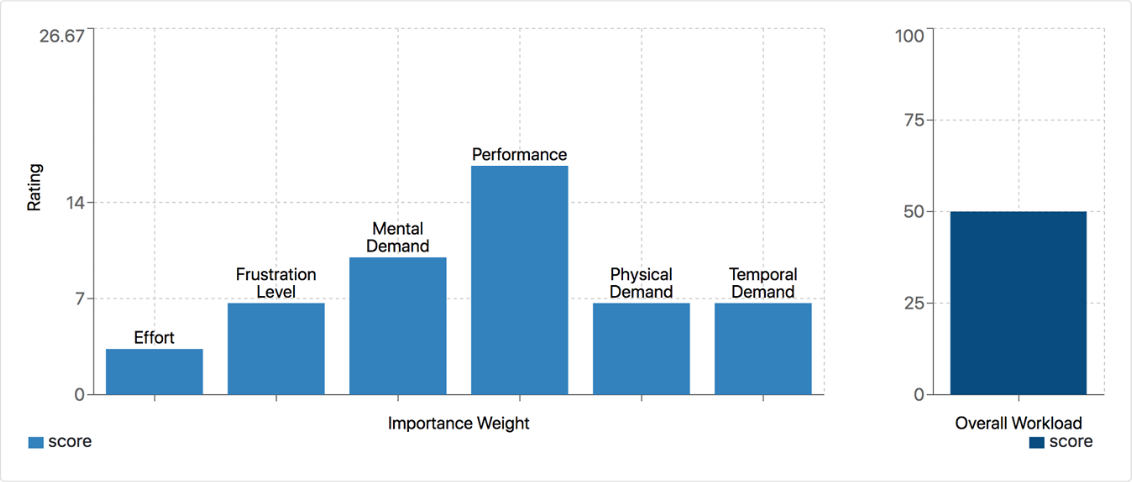
# Understanding Subjective Analysis




(a) Rating sheet for various dimensions of NASA-TLX




(b) Workload dimension comparison card



# Conclusion

NASA TLX (task load index) & SUS(system usability study)

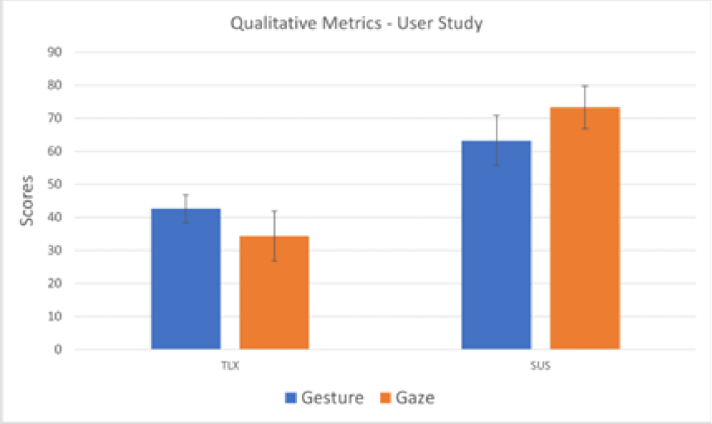



Fig. 5. Qualitative Metrics with Gaze and Gesture Interfaces

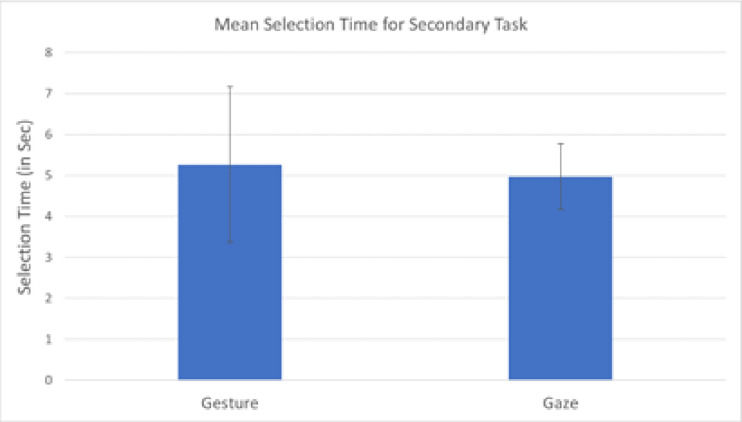



Fig. 6. Mean Selection Time with Gesture & Gaze

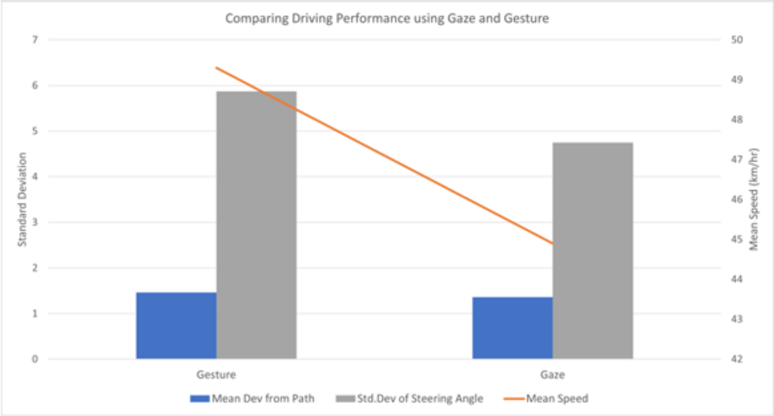



Fig. 4. Comparison of Primary Driving Task Parameters with Gaze and Gesture Interfaces

Applications on Automotive Driving?

Efficient Interaction with Automotive Heads Up Displays using Appearance-based Gaze Tracking, AUTO UI 2021

# Questions?

Thank you