Team IVA – John Stephen, Jeremiah Lewis, Akbar Khader, Andreas Bruenner

Driver Assistance System using Intelligent Video Analytics on the Edge

Agenda

- Roles & Responsibilities
- Project Background and Context
- Project Goal
- Intelligent Video Analytics Pipeline
- System Platform and Development
- Results
- Outlook

Roles & Responsibilities

	John Stephen	Jeremiah Lewis	Akbar Khader	Andreas Bruenner
Project Management	X			X
Ideation / Concept				X
Technical Writing	X	X	X	X
Data Analysis	X	X	X	X
System Development (HW / SW)				X
Data Acquisition		X		X
Testing				X

Why this project?

Project Background and Context

Global Context

According to the WHO
*(World Health Organization, 2021 and 2022)

- around 1.3 Mio people die and
- around 50 Mio people

are injured every year from traffic accidents.

- 93% of the world's fatalities are happening in low- and middle-income countries.
- Those countries, however, only account for around 60% of the world's vehicles.
- Road traffic crashes are estimated to cost most countries 3% of their GDP.

Distracted Driving

Distracted driving is caused by

- Attending children in the backseat.
- Interacting with the car’s onboard system.
- Cell phone use.
- Etc.

Results often in traffic accidents due to

- Reduced reaction times.
- Overlooking other traffic participants .
- Inability to keep sufficient distance from. proceeding cars.
- Inability to keep the correct lane.

US Context

According to the Department of Transportation
*(Department of Transportation, Washington, DC: NHTSA, 2021)

Distracted driving in 2019 accounted for

- 9% of all fatal and
- 15% of all injury crashes of all police-reported motor vehicle accidents.
- Economic damages of around $129 billion.
- 43,000 people being killed.
- Trend is worsening.

Project Goal

Project Goal

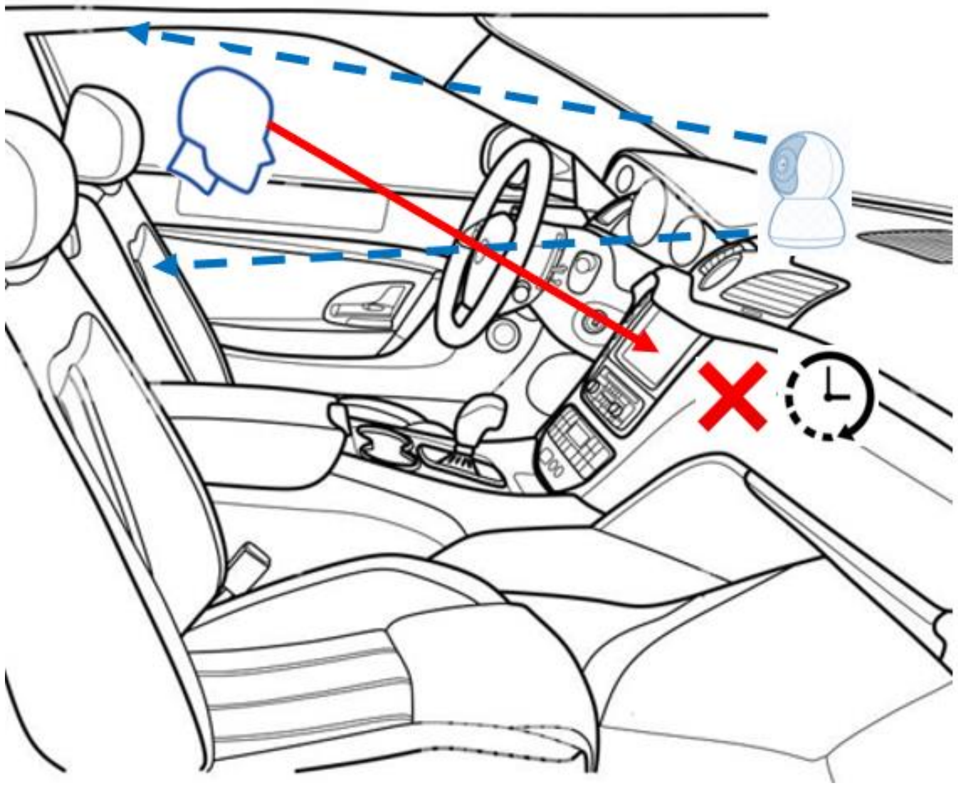
- Contribute to the WHO’s target of halving the global number of deaths and injuries caused by traffic accidents by 2030.

- Provide an **Advanced Driver Assistance System** (ADAS), which is
  - Affordable (to also tackle the problem in low- and middle-income countries)
  - Vehicle vendor- and model-agnostic
  - Supported in **cars**, **vans**, and **light trucks**.

- The system’s purpose in its first version is the real-time detection of prolonged distracted driving and raising an alarm to refocus driver.

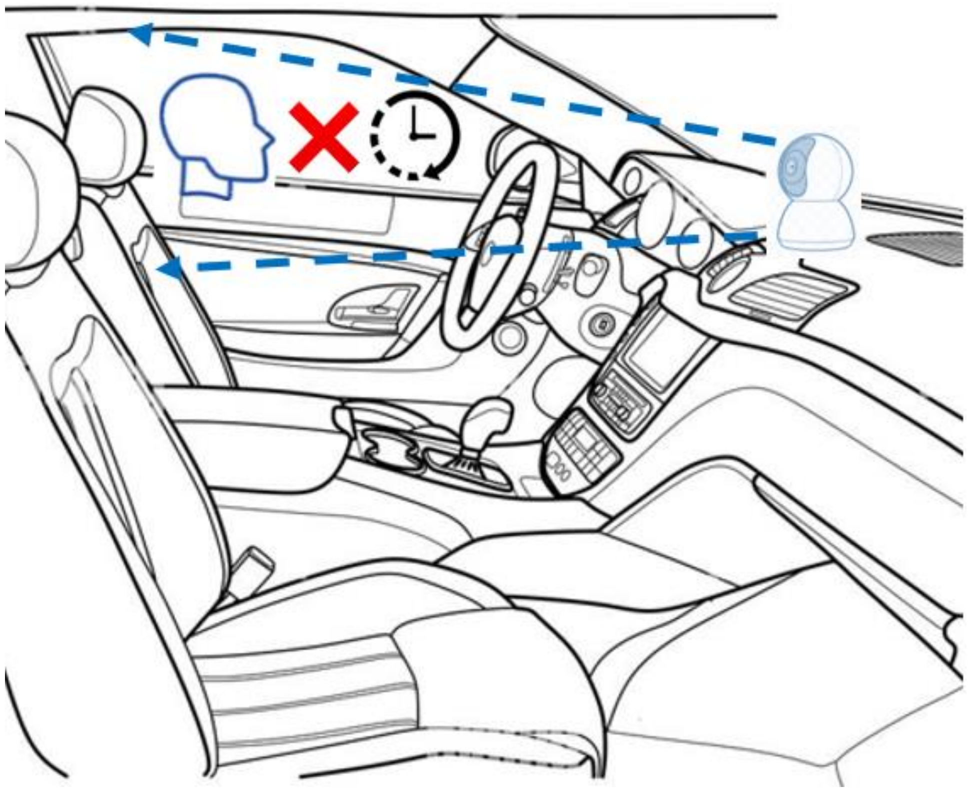
Scenarios to consider (1/3)

The driver looks away from the road (down) for a prolonged period.



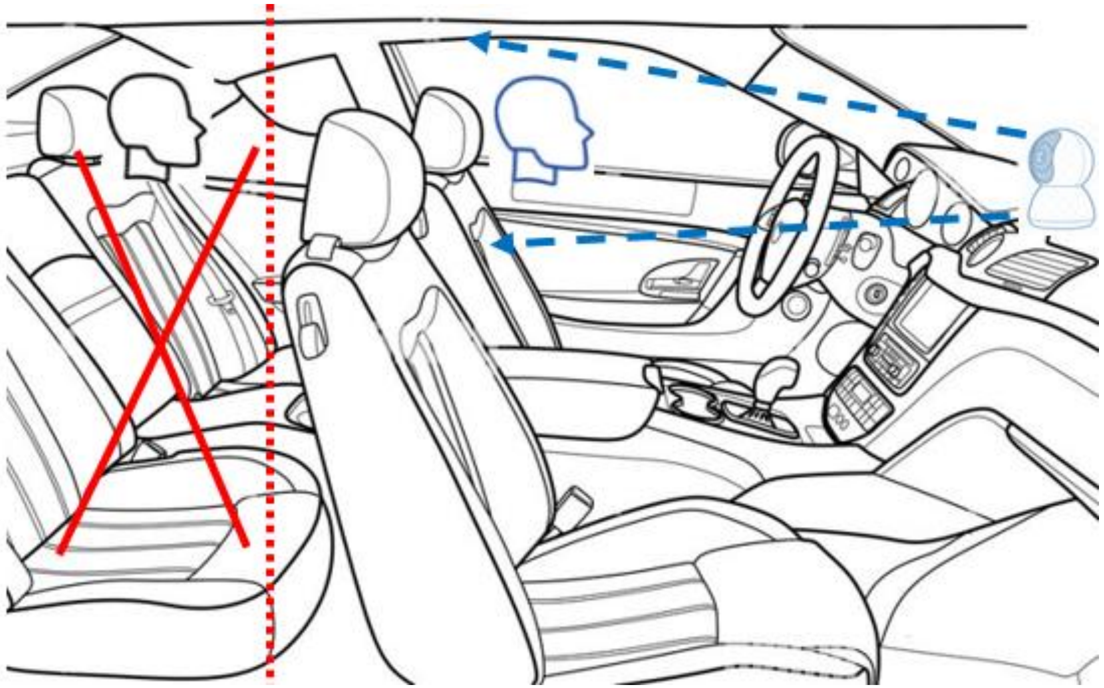
Scenarios to consider (2/3)

The driver has his eyes closed for a prolonged period.



Scenarios to consider (3/3)

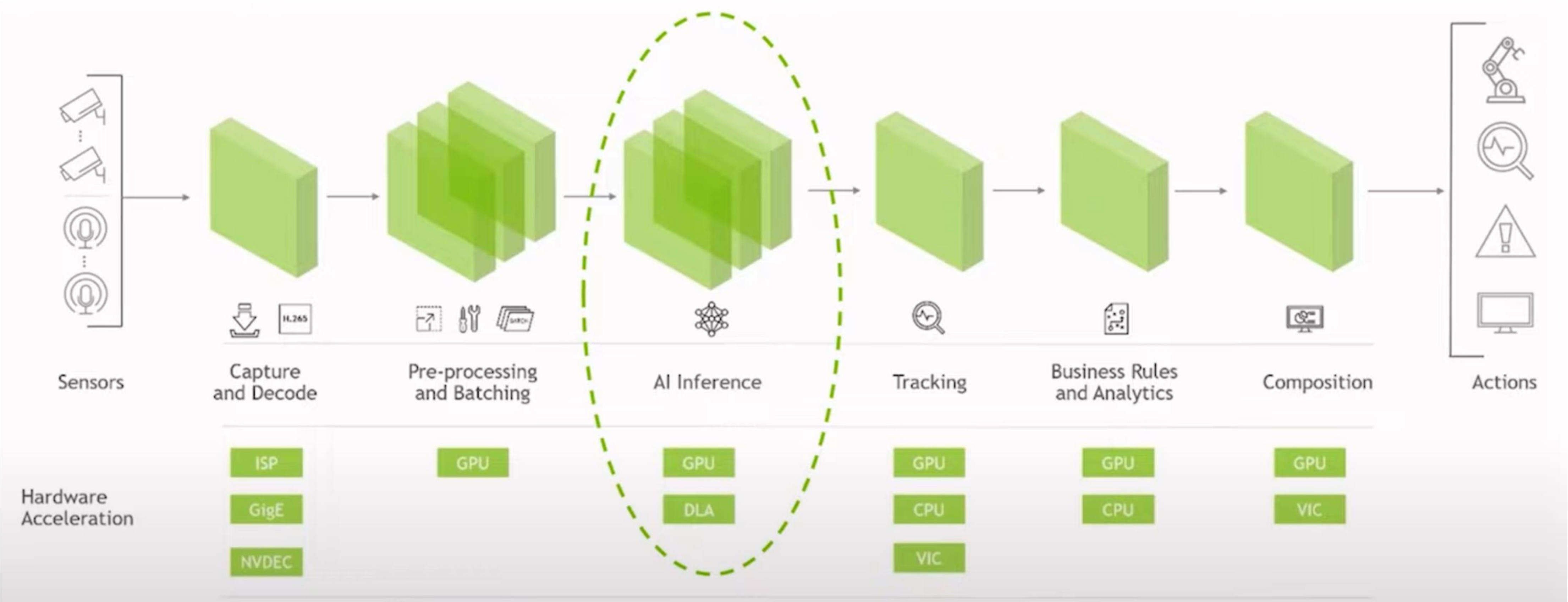
Passengers in the backseat need to be ignored to avoid false detections.



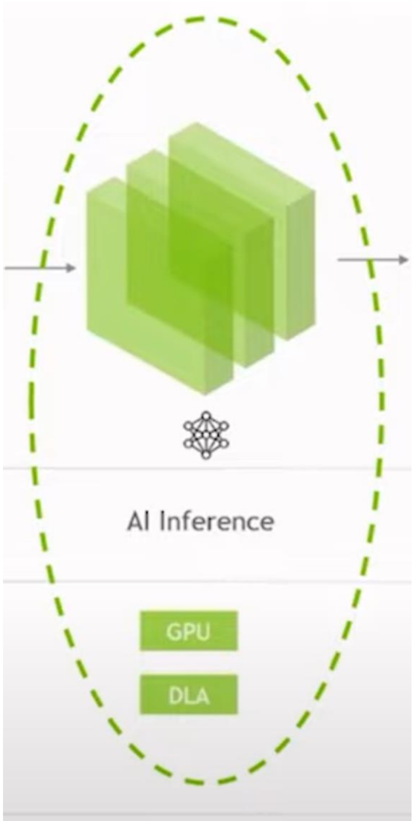
Video Pipeline and Needed Components

Intelligent Video Analytics

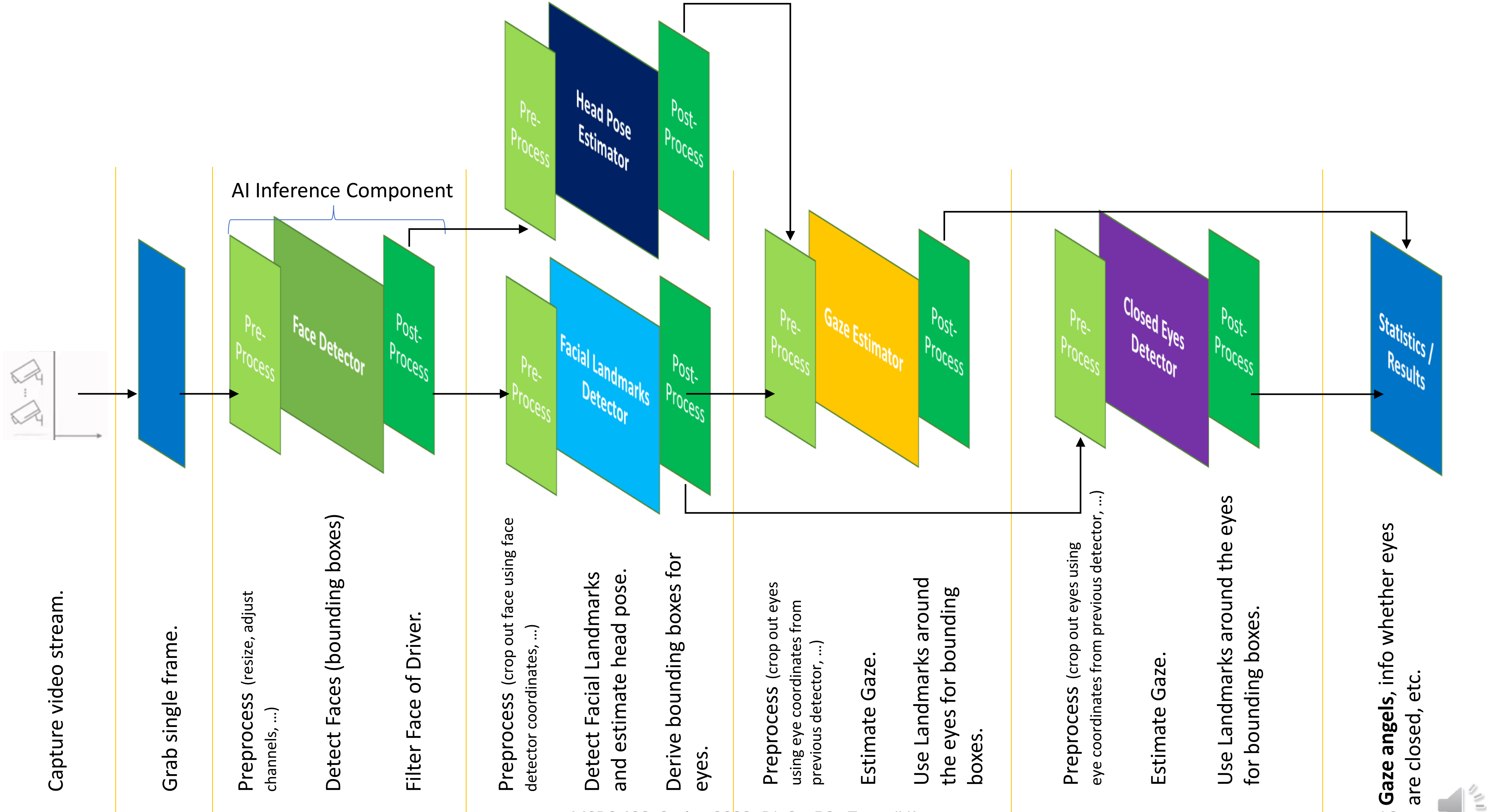
Generic Video Pipeline and Components

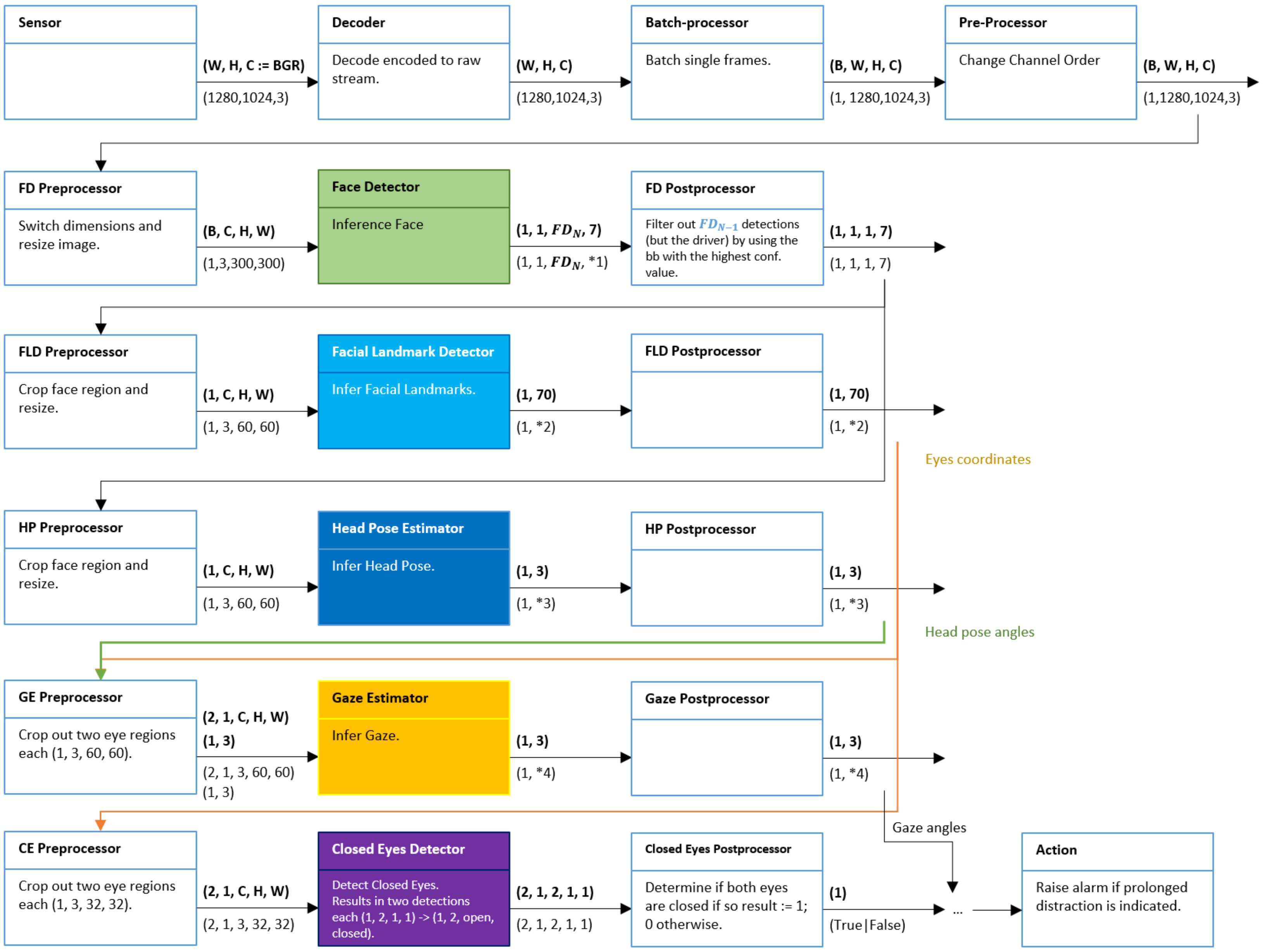


Required AI Inference Components



- The pipeline and especially the AI Inference to track gaze can be configured in many ways.
- We needed at the very least:
  - Face detection
  - Facial landmark detection
  - Left and right eye detection
  - Head pose estimation
  - Gaze estimation


AI Inference Component
Pre-Process
Face Detector
Post-Process
Pre-Process
Head Pose Estimator
Post-Process
Pre-Process
Facial Landmarks Detector
Post-Process
Pre-Process
Gaze Estimator
Post-Process
Pre-Process
Closed Eyes Detector
Post-Process
Statistics / Results
Capture video stream.
Grab single frame.
Preprocess (resize, adjust channels, ...)
Detect Faces (bounding boxes)
Filter Face of Driver.
Preprocess (crop out face using face detector coordinates, ...)
Detect Facial Landmarks and estimate head pose.
Derive bounding boxes for eyes.
Preprocess (crop out eyes using eye coordinates from previous detector, ...)
Estimate Gaze.
Use Landmarks around the eyes for bounding boxes.
Preprocess (crop out eyes using eye coordinates from previous detector, ...)
Estimate Gaze.
Use Landmarks around the eyes for bounding boxes.
Gaze angels, info whether eyes are closed, etc.

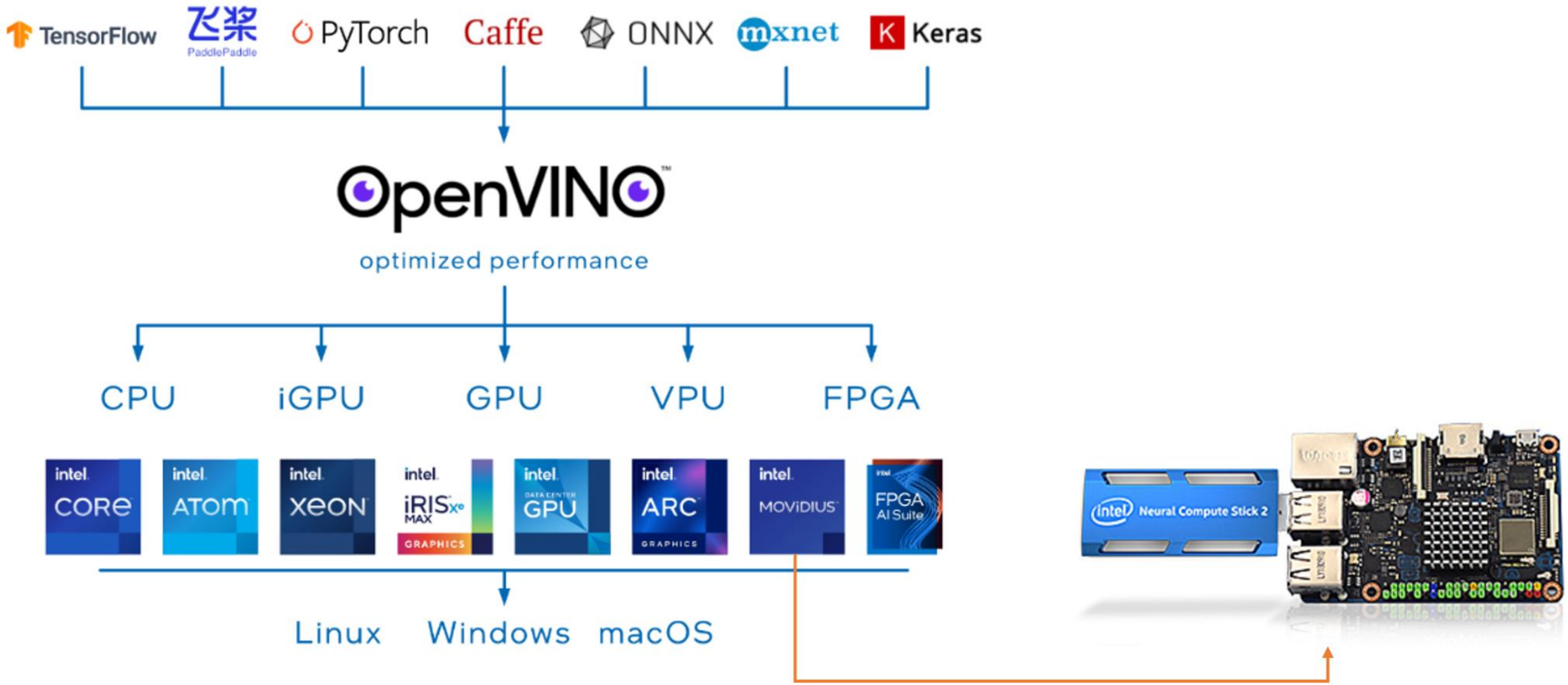


Sensor
(W, H, C := BGR)
(1280,1024,3)
Decoder
Decode encoded to raw stream.
(W, H, C)
(1280,1024,3)
Batch-processor
Batch single frames.
(B, W, H, C)
(1, 1280,1024,3)
Pre-Processor
Change Channel Order
(B, W, H, C)
(1,1280,1024,3)
FD Preprocessor
Switch dimensions and resize image.
(B, C, H, W)
(1,3,300,300)
Face Detector
Inference Face
(1, 1, FD_N, 7)
(1, 1, FD_N, *1)
FD Postprocessor
Filter out FD_{N-1} detections (but the driver) by using the bb with the highest conf. value.
(1, 1, 1, 7)
(1, 1, 1, 7)
FLD Preprocessor
Crop face region and resize.
(1, C, H, W)
(1, 3, 60, 60)
Facial Landmark Detector
Infer Facial Landmarks.
(1, 70)
(1, *2)
FLD Postprocessor
(1, 70)
(1, *2)
Eyes coordinates
HP Preprocessor
Crop face region and resize.
(1, C, H, W)
(1, 3, 60, 60)
Head Pose Estimator
Infer Head Pose.
(1, 3)
(1, *3)
HP Postprocessor
(1, 3)
(1, *3)
Head pose angles
GE Preprocessor
Crop out two eye regions each (1, 3, 60, 60).
(2, 1, C, H, W)
(1, 3)
(2, 1, 3, 60, 60)
(1, 3)
Gaze Estimator
Infer Gaze.
(1, 3)
(1, *4)
Gaze Postprocessor
(1, 3)
(1, *4)
Gaze angles
CE Preprocessor
Crop out two eye regions each (1, 3, 32, 32).
(2, 1, C, H, W)
(2, 1, 3, 32, 32)
Closed Eyes Detector
Detect Closed Eyes. Results in two detections each (1, 2, 1, 1) -> (1, 2, open, closed).
(2, 1, 2, 1, 1)
(2, 1, 2, 1, 1)
Closed Eyes Postprocessor
Determine if both eyes are closed if so result := 1; 0 otherwise.
(1)
(True|False)
...
Action
Raise alarm if prolonged distraction is indicated.

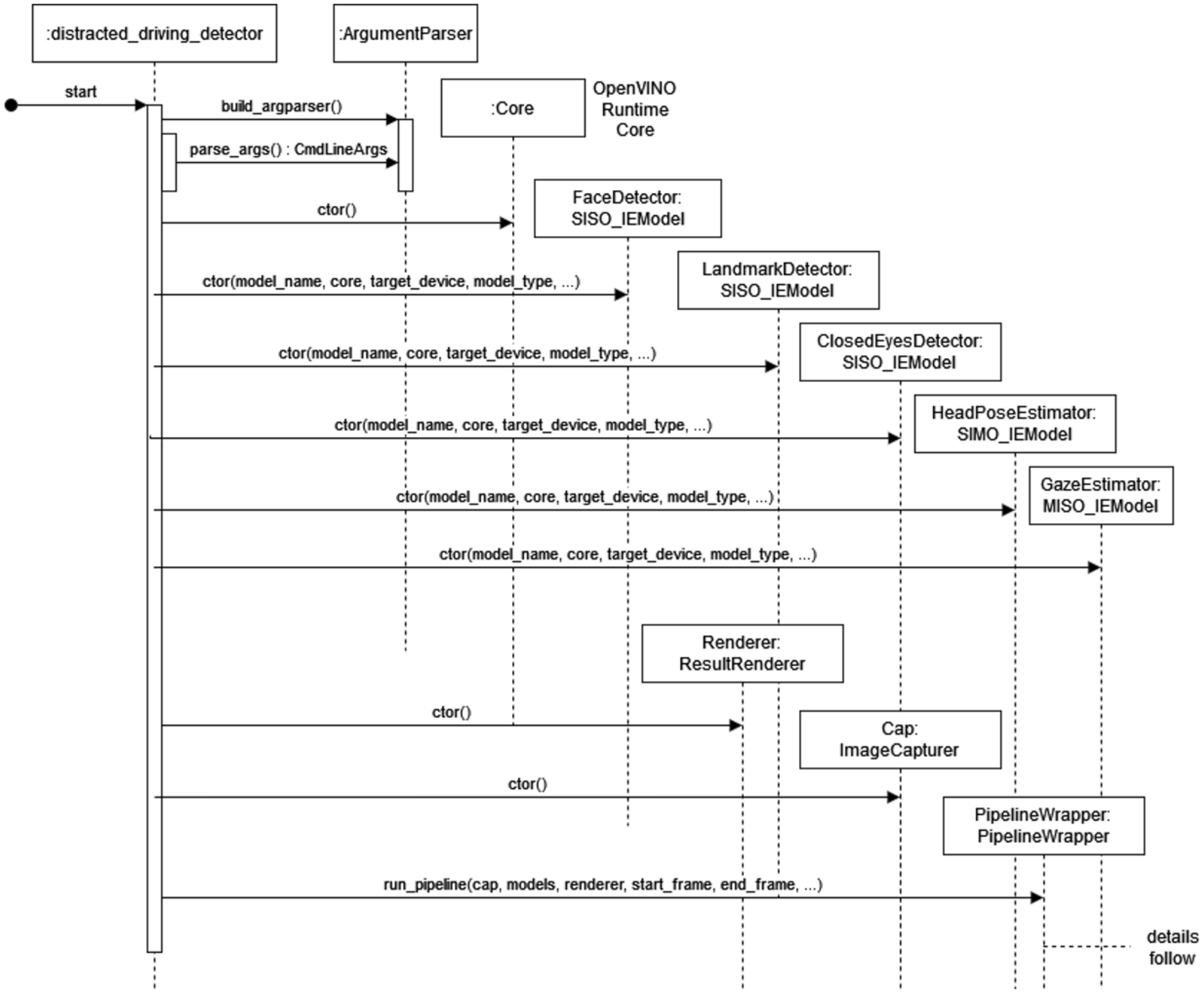
SW and HW Choices

System Platform

OpenVINO and Edge AI Development



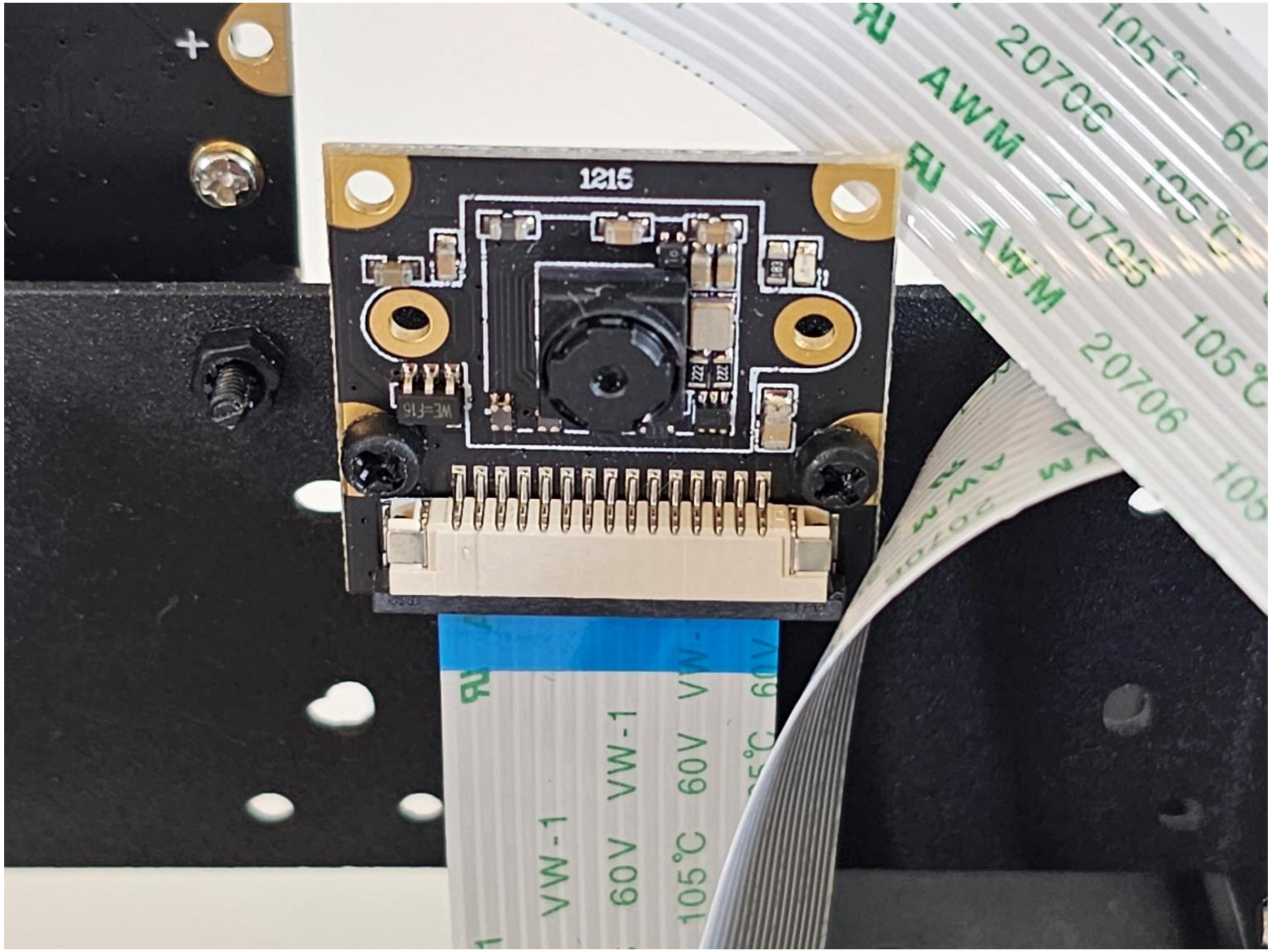
Detector Initialization



NVIDIA Jetson Orin NX Developer Board

IMX219-77 Camera

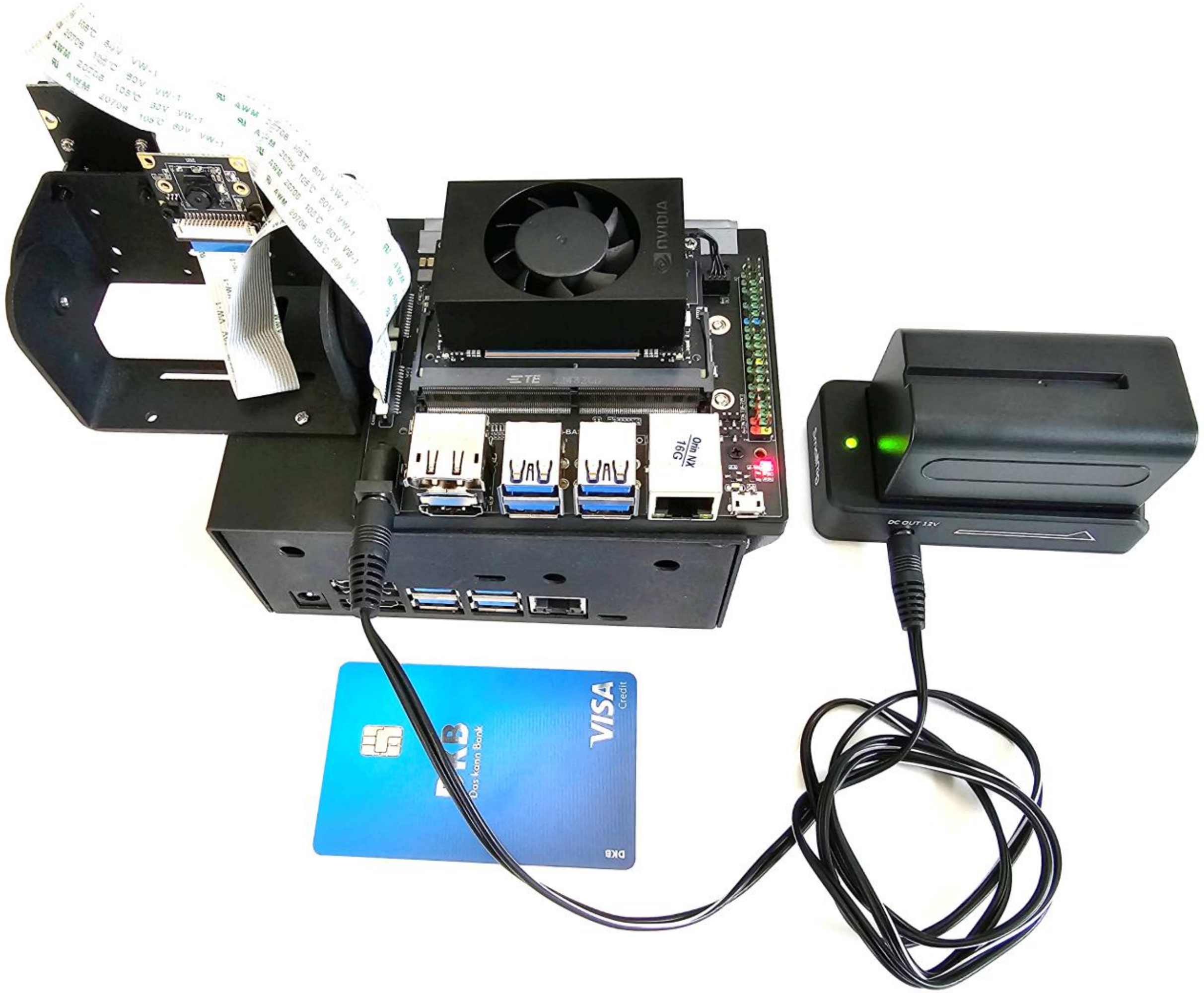
- Sensor:
  - **Sony IMX219**
- Resolution:
  - **8 Megapixels**
- Supported resolutions:
  - **1280 x 720** to **3280 x 2464**
- min **30 FPS**
- Angle of View (diagonal):
  - **~80 degrees**
- Dimension
  - **25mm x 24 mm**



Power Supply

Li-Ion Battery and backplate to provide 12V to the NVIDIA Jetson Orin.

Fully Assembled System



Car installation

Performance on the Edge and Accuracy

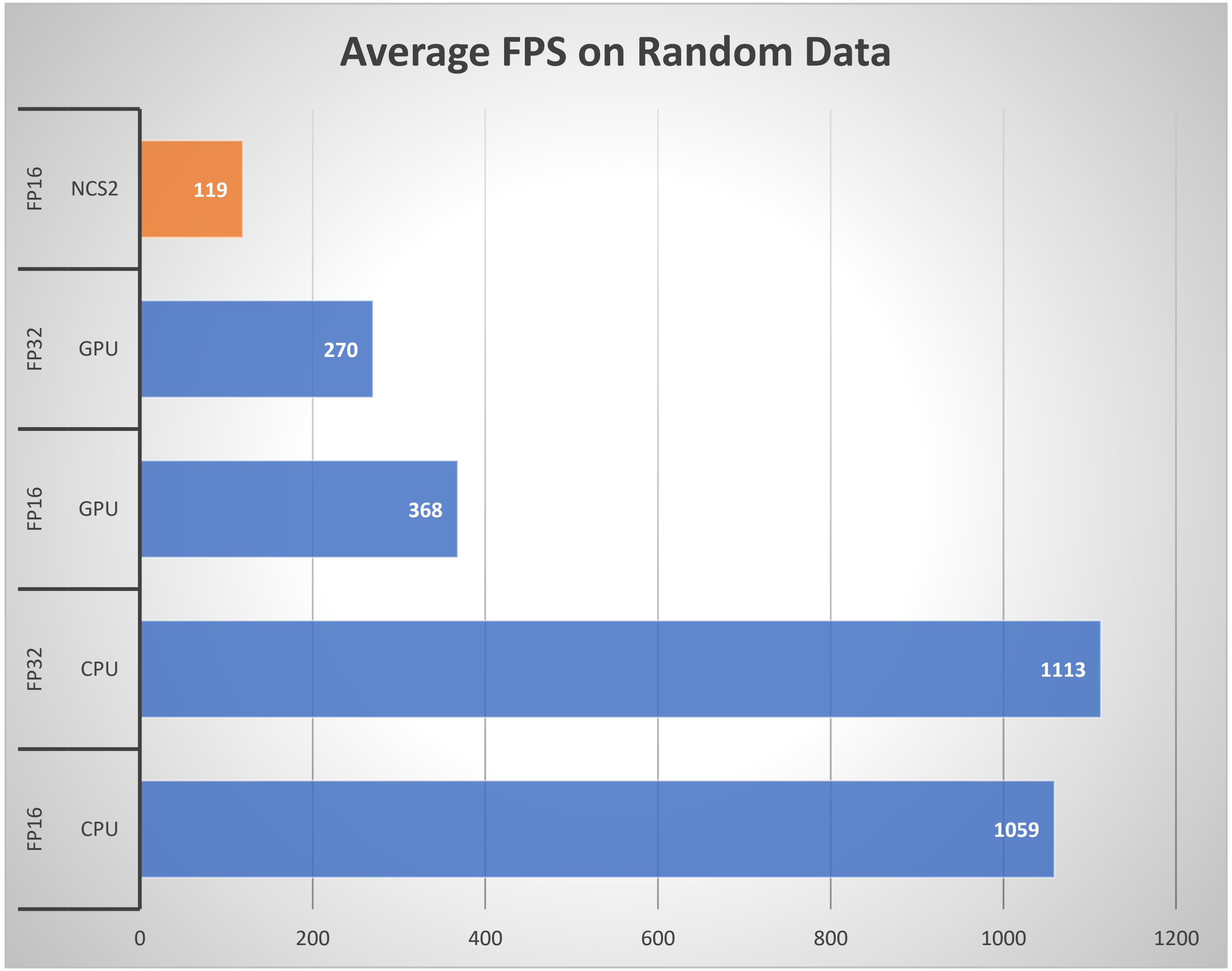
Results

Performance (1/3)

Benchmark results to the right consider only the pure AI component inference performance.

The benchmark used **random data** as input.

Results show the average frame rate in FPS over a duration of 60 seconds.

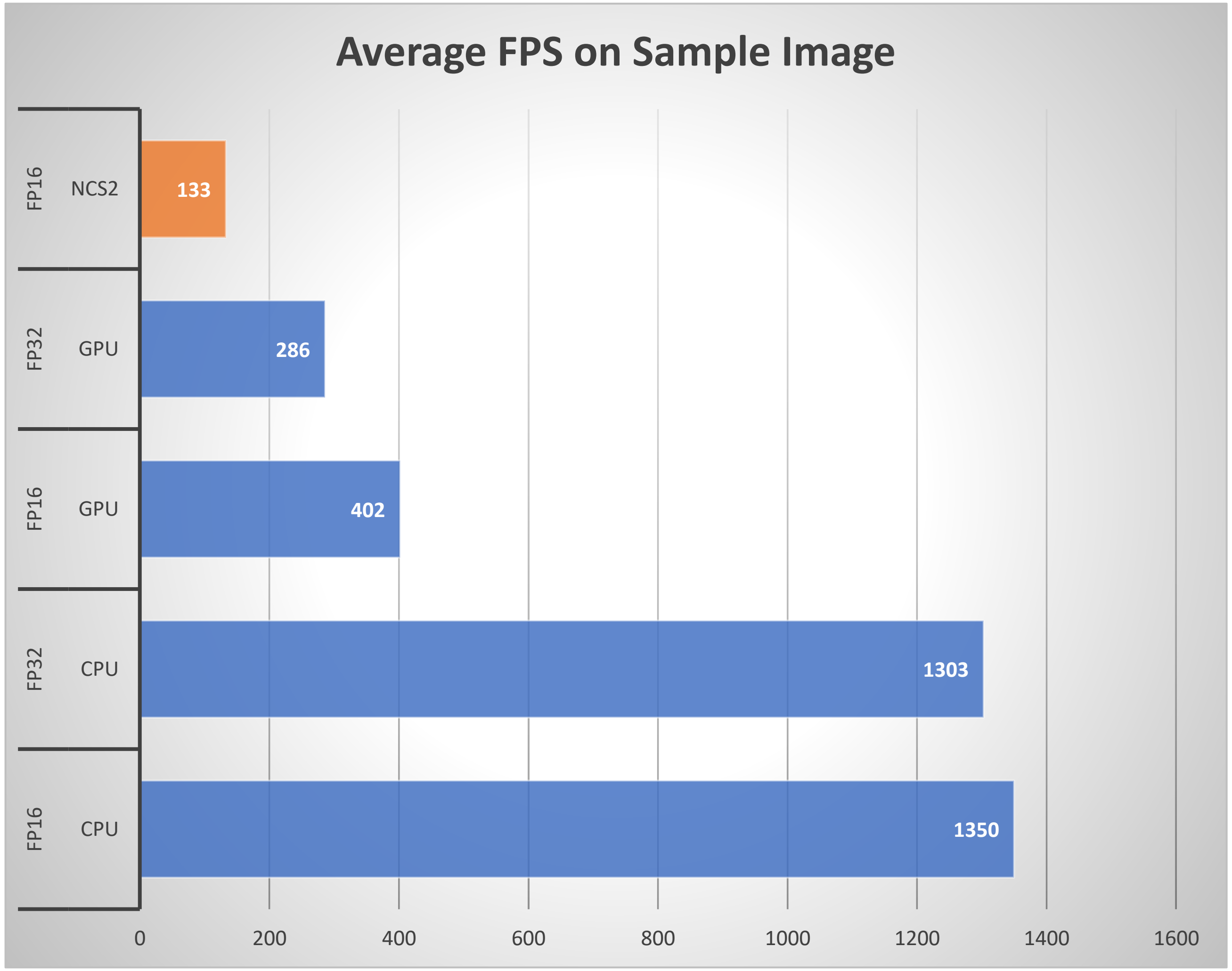


Performance (2/3)

Benchmark results to the right consider only the pure AI component inference performance.

The benchmark used **sample images captured by the ADAS** as input.

Results show the average frame rate in FPS over a duration of 60 seconds.

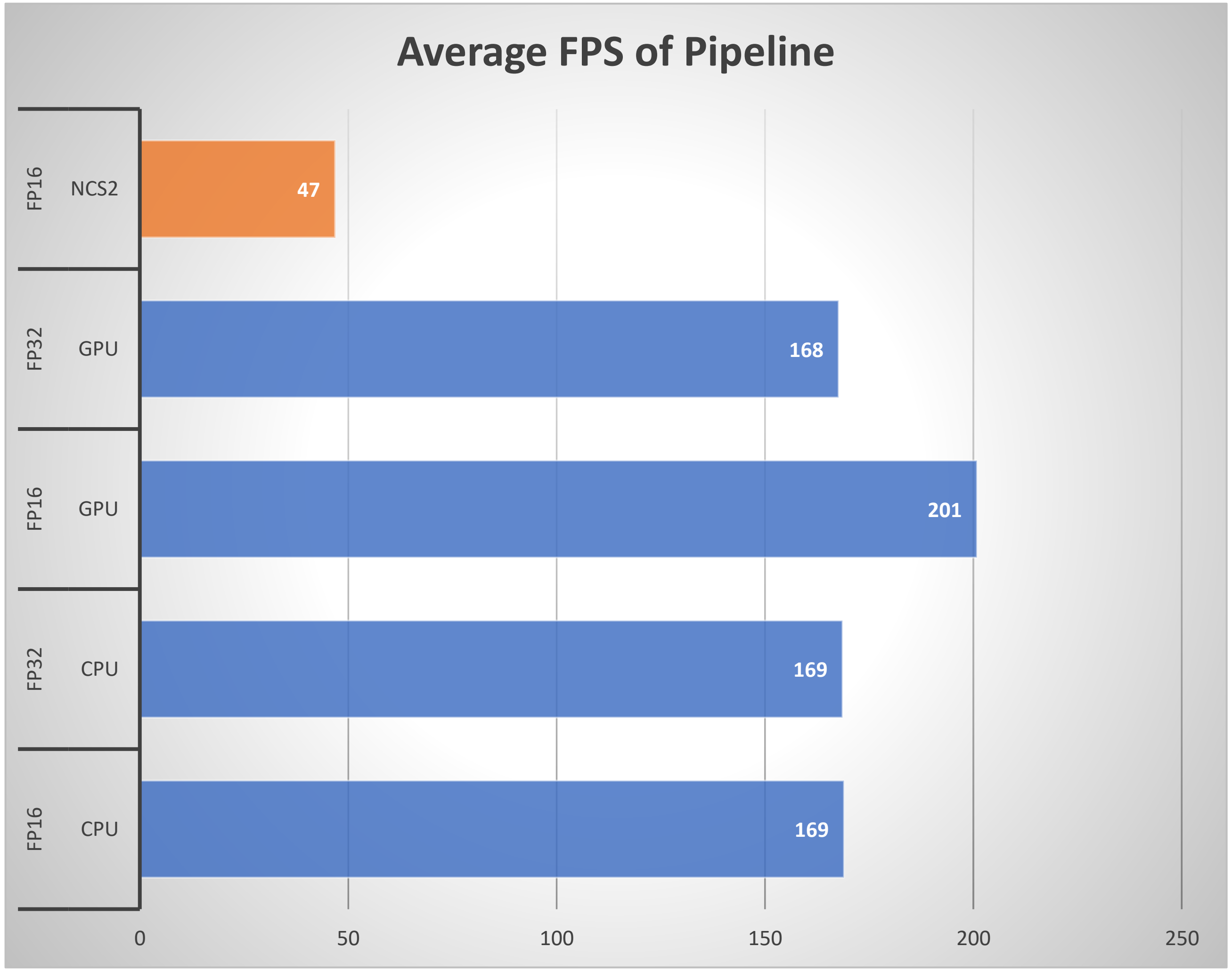


Performance (3/3)

Benchmark of the entire video pipeline, i.e., **all AI and non-AI components**.

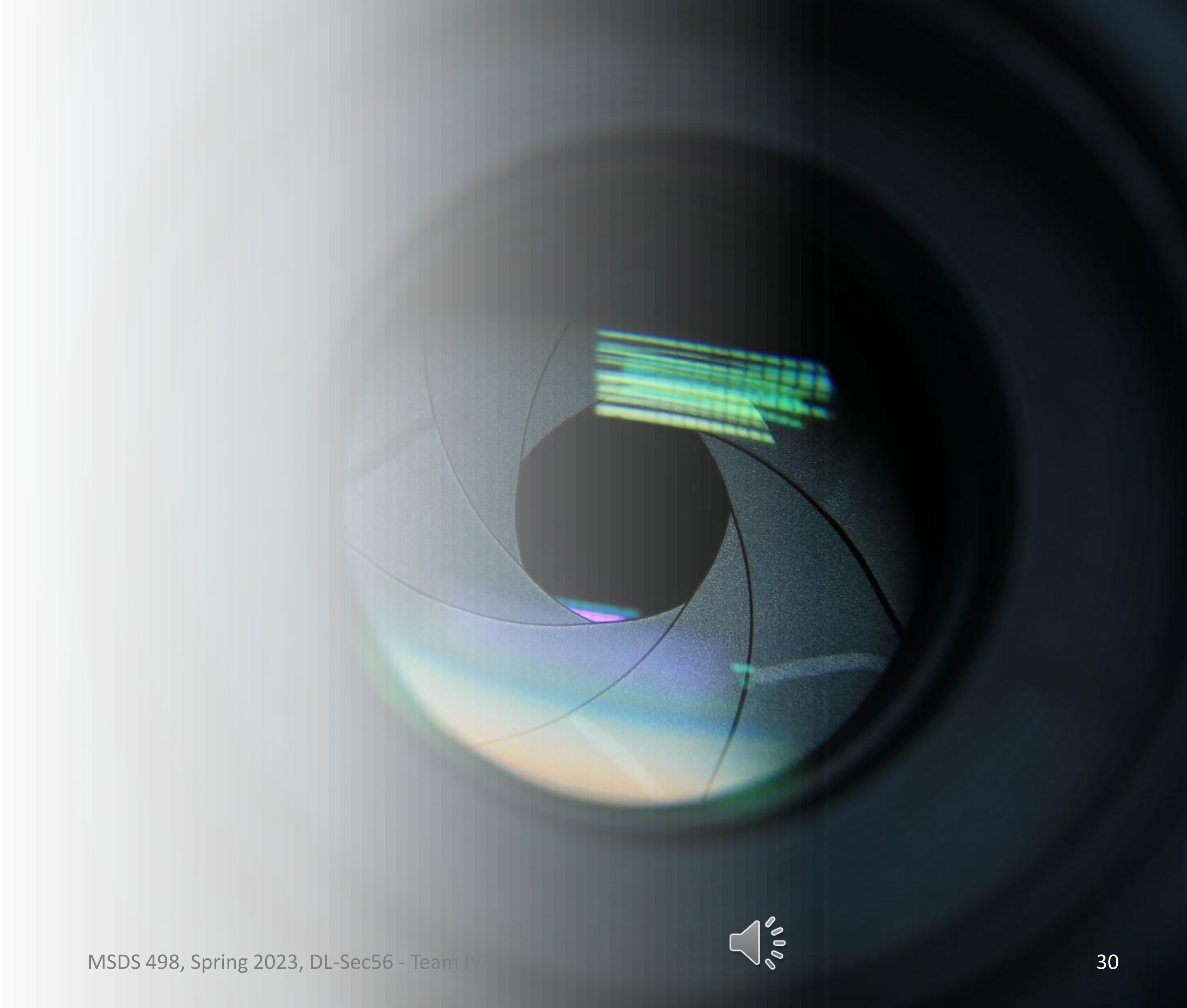
This benchmark shows the actual video pipeline / system performance using a **video stream** as input.

Results to the right show the average frame rate in FPS over 60 seconds.



Accuracy

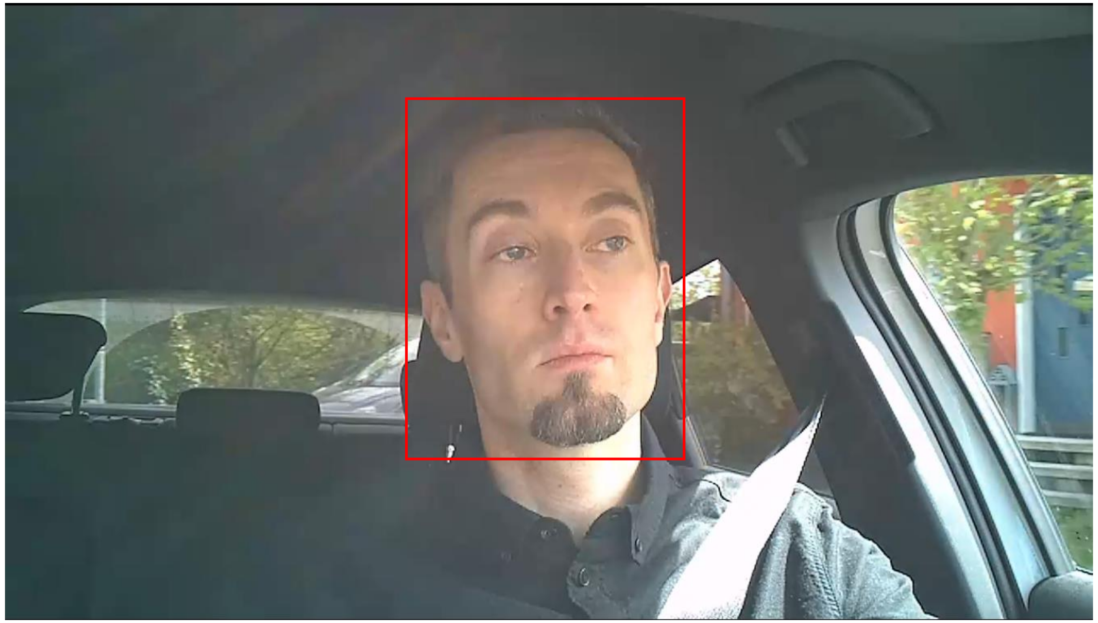
- In addition to an accuracy evaluation common to neural networks we used visual debugging to evaluate accuracy on video streams.
- Those visual annotations are disabled before system roll-out to further increase system performance, since they require non-negligible compute resources.
- The following slides show the used annotations
    - Face bounding box
    - Facial landmarks annotations
    - Eyes bounding boxes
    - Estimated head pose angles
    - Estimated Gaze angles



Face Detector Annotation

Captured Image

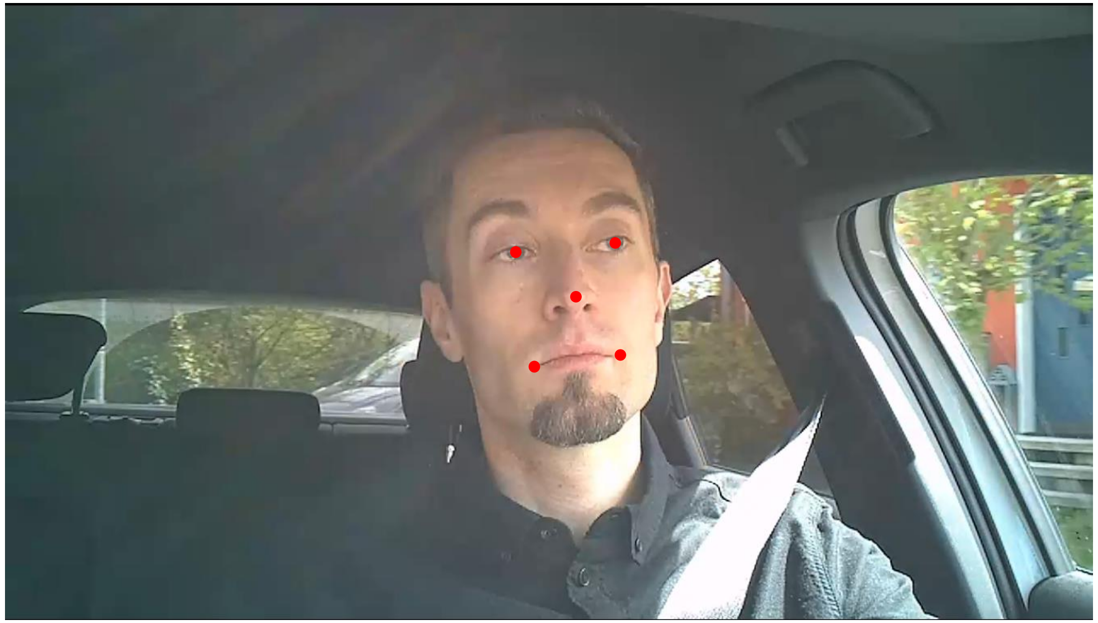
Face Annotated with Bounding Box of Detected Face



Facial Landmarks Detector Annotation (1/2)

Captured Image

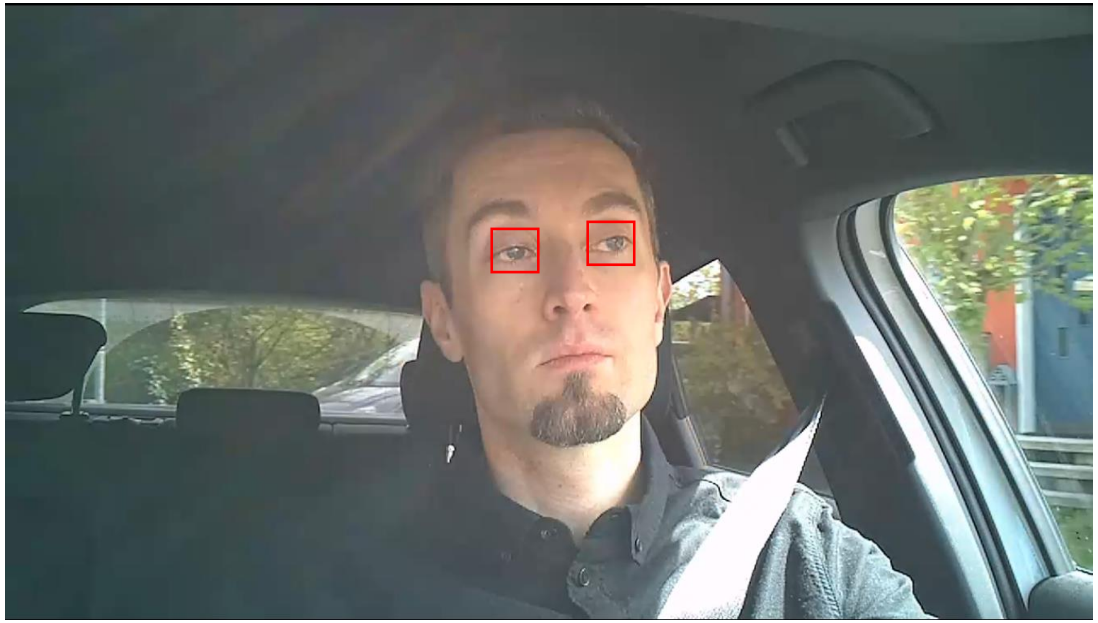
Face Annotated with Detected Landmarks



Facial Landmarks Detector Annotation (2/2)

Captured Image

Eyes Annotated with Bounding Boxes based on Landmarks



Head Pose Estimator Annotation

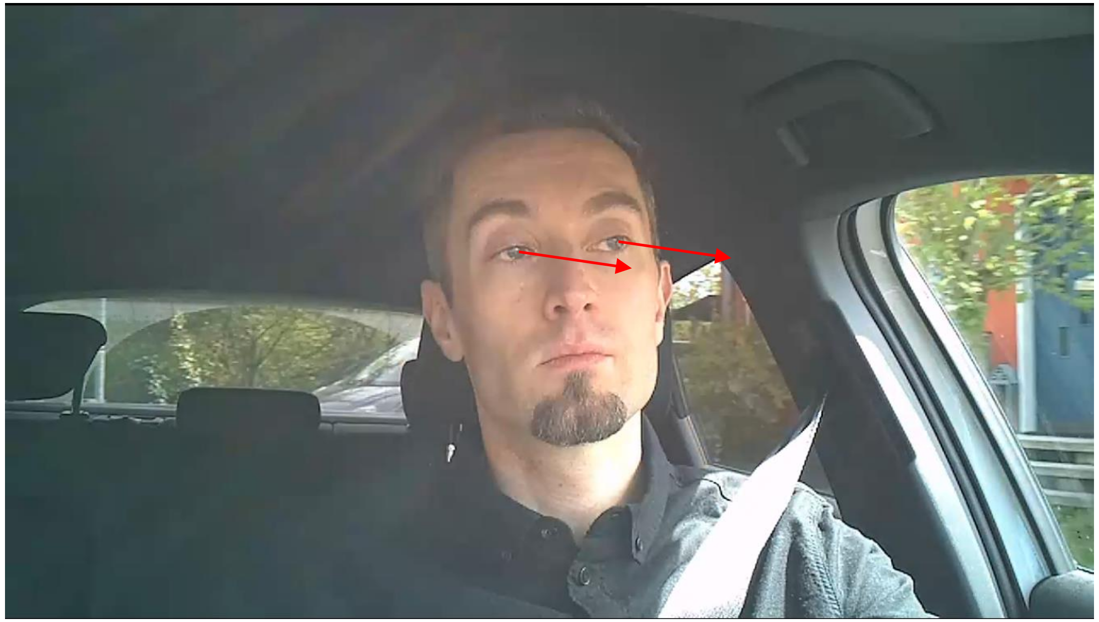
Captured Image

Face Annotated with Estimated Head Pose Angles

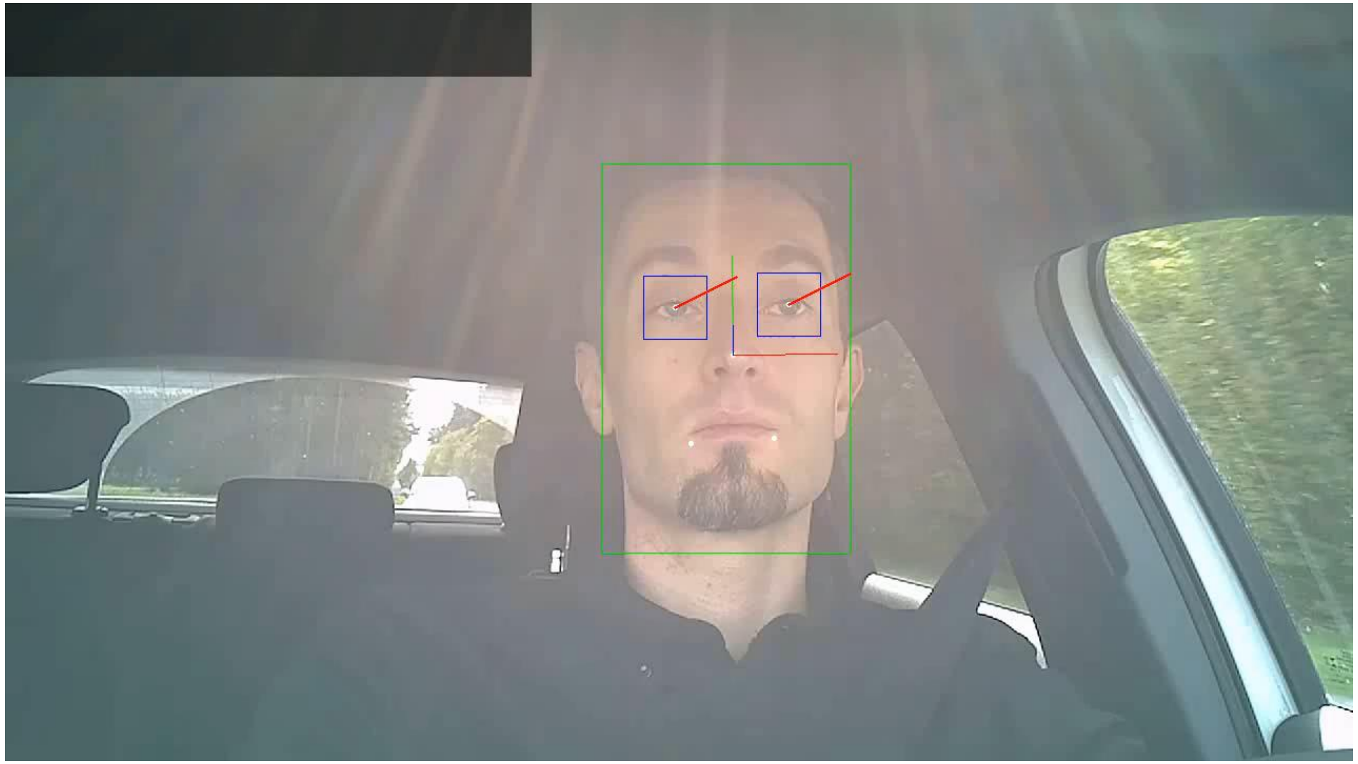
Gaze Estimator Annotation

Captured Image

Face Annotated with Estimated Gaze Angles

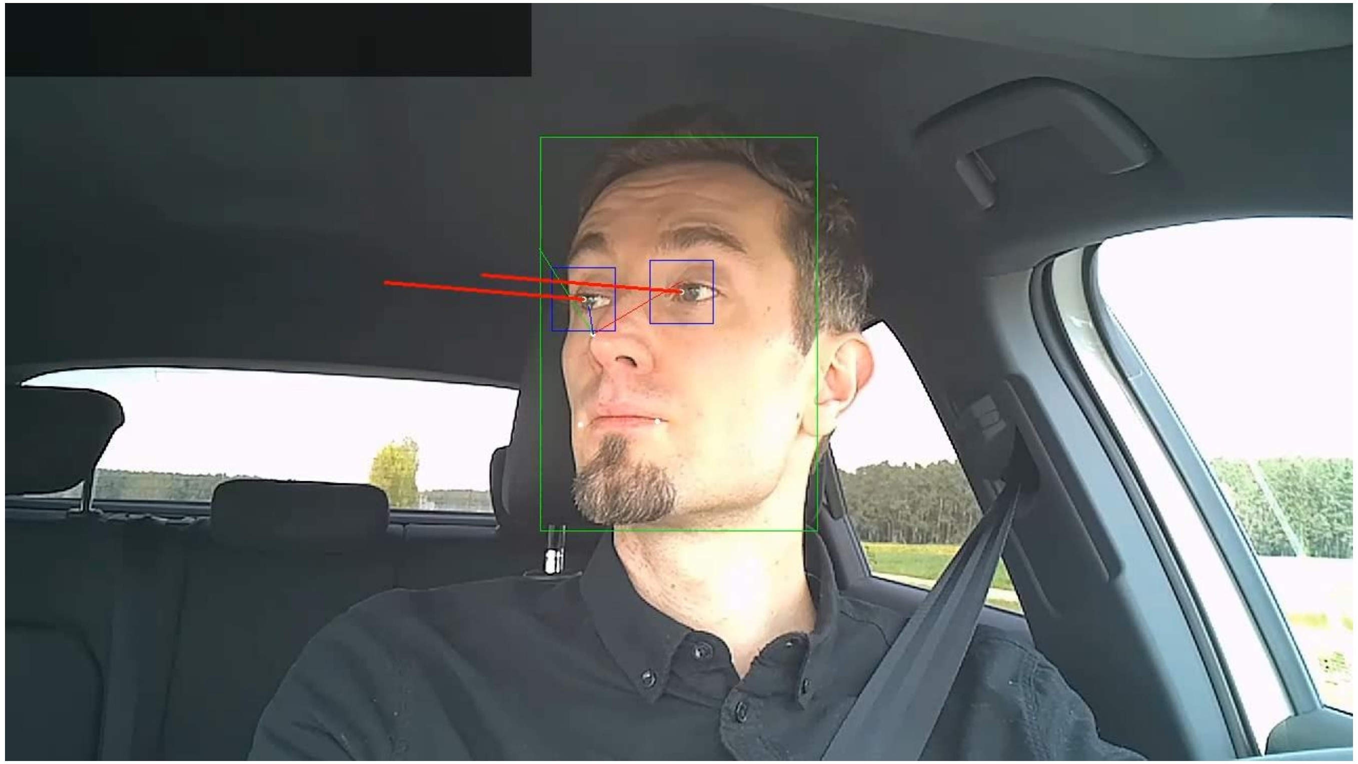


Non-Distracted Driving Scenes



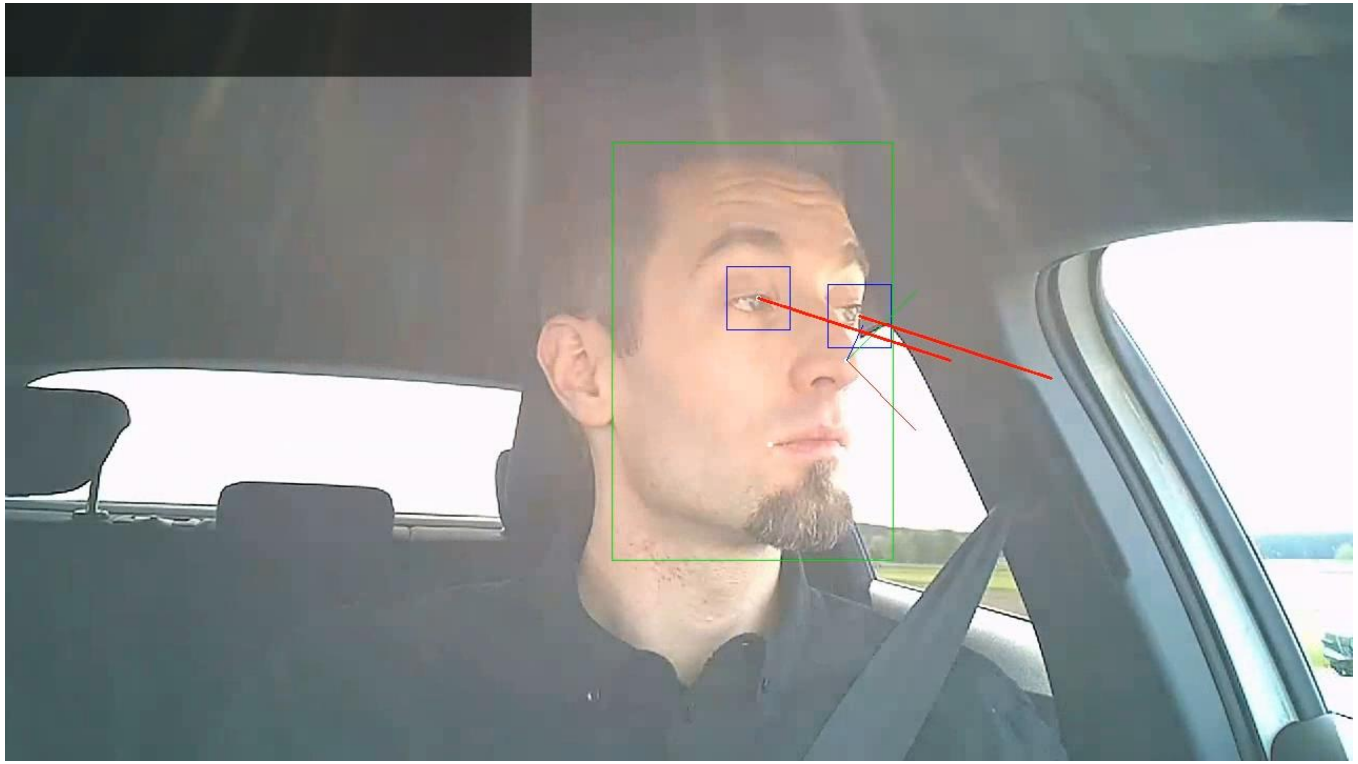
Non-Distracted Driving (1/3)

- Sample shows the system's accurate gaze estimation
- **With changing illuminations** (bright sunlight and shadows)



Non-Distracted Driving (2/3)

- Sample shows the system's accurate gaze estimation
  - While taking a **right turn**
  - Looking to the right
  - Looking down (below distracted driving threshold of 2 seconds)
  - Looking straight ahead

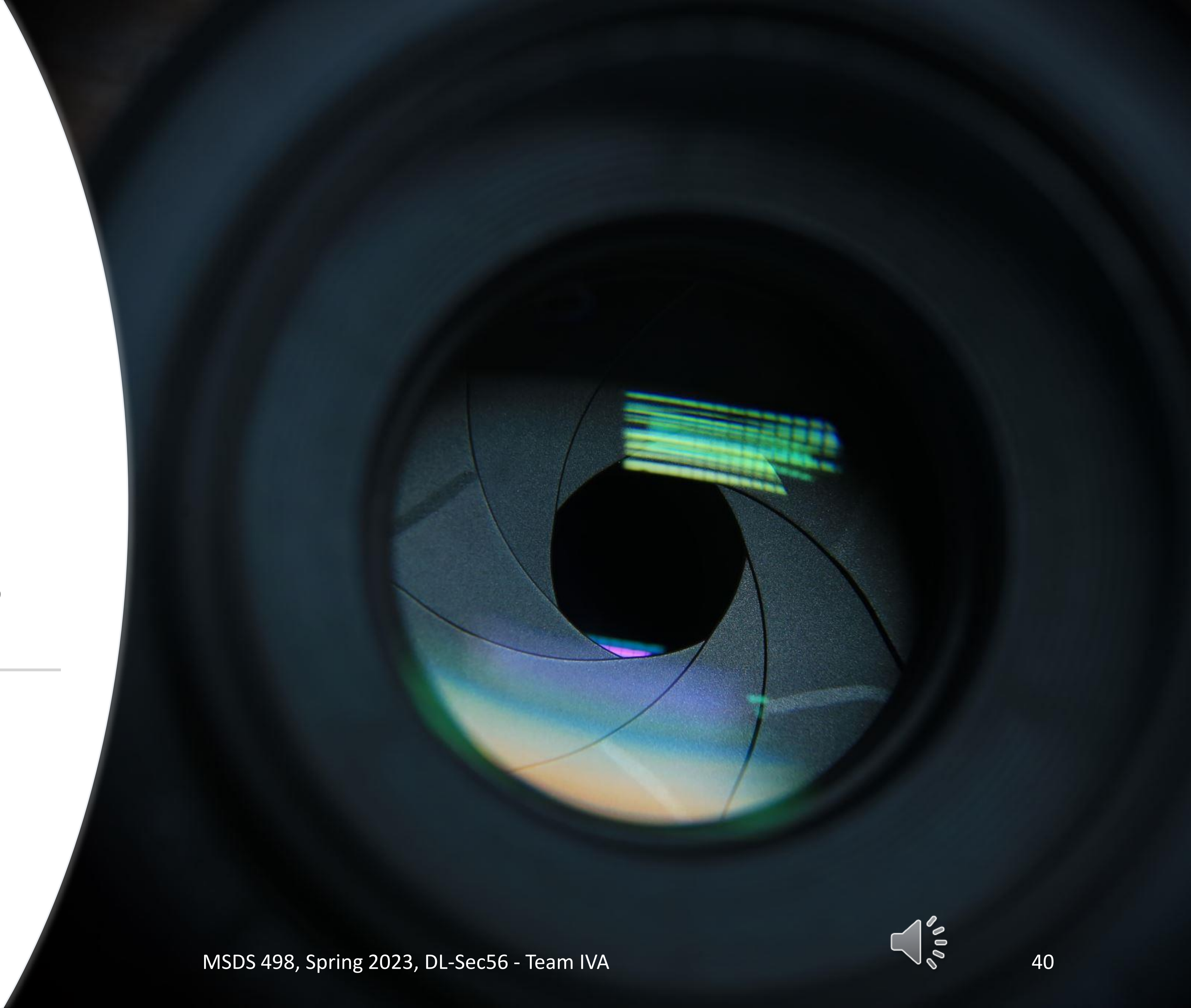


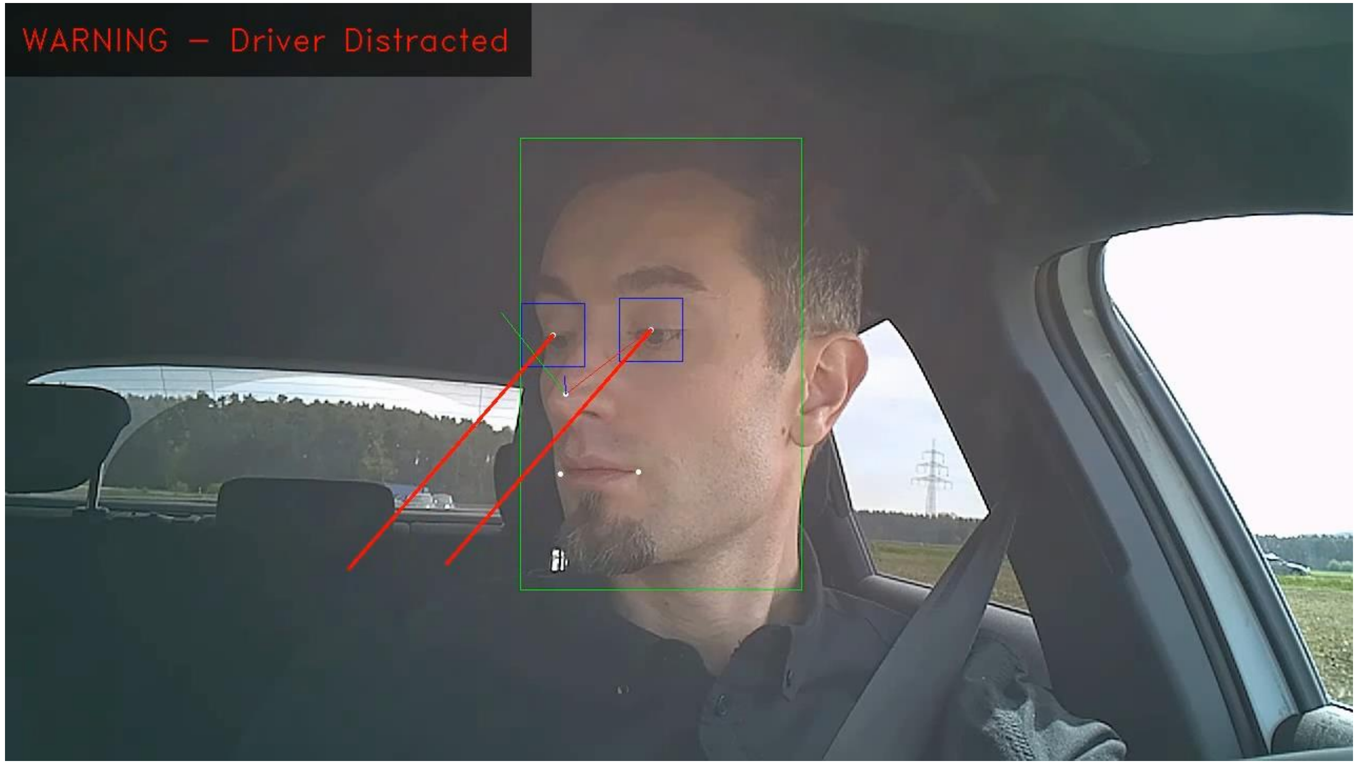
Non-Distracted Driving (3/3)

- Sample shows the system's accurate gaze estimation

  - While taking a **left turn**
  - Looking to the left
  - Looking down (very briefly)
  - Looking straight ahead

Distracted Driving Scenes

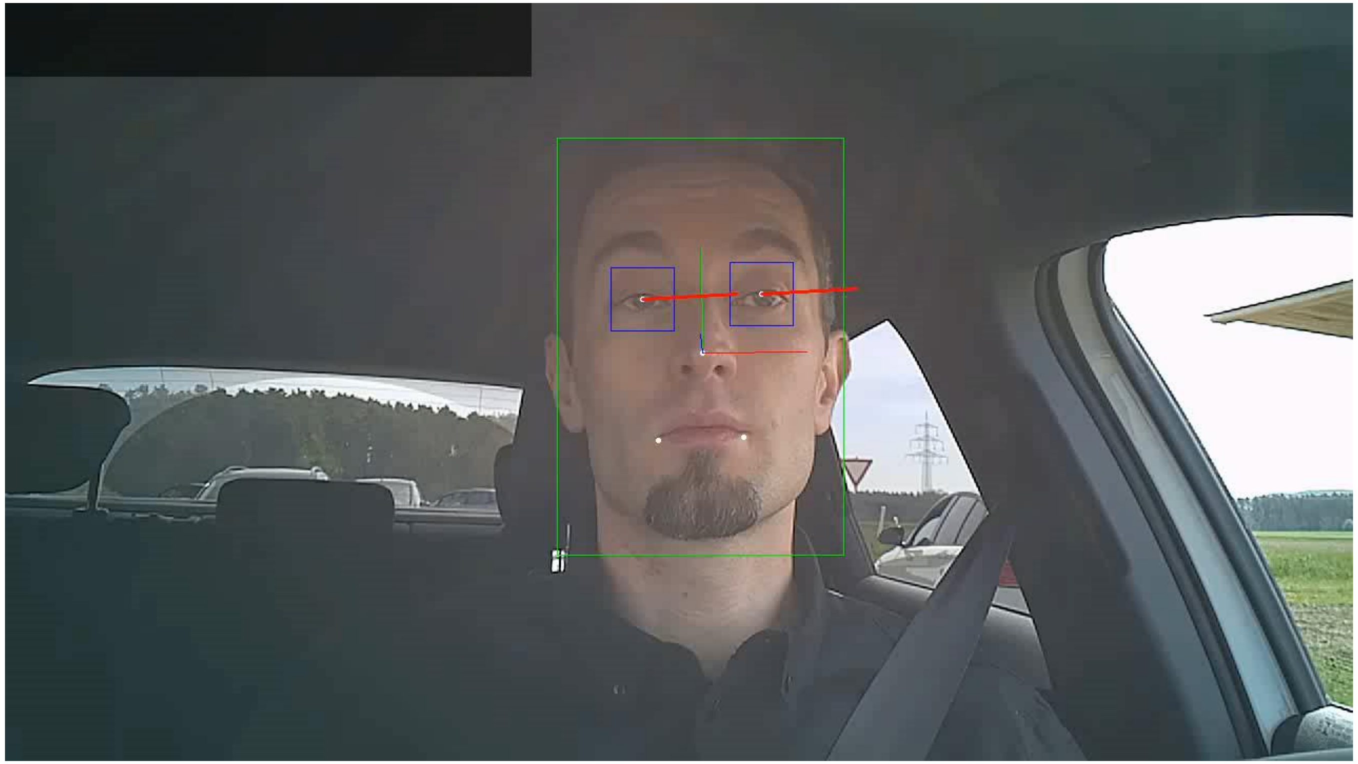




Distracted Driving (1/4)

- Sample shows a distracted driver exceeding the allowed distraction period once with a gaze.

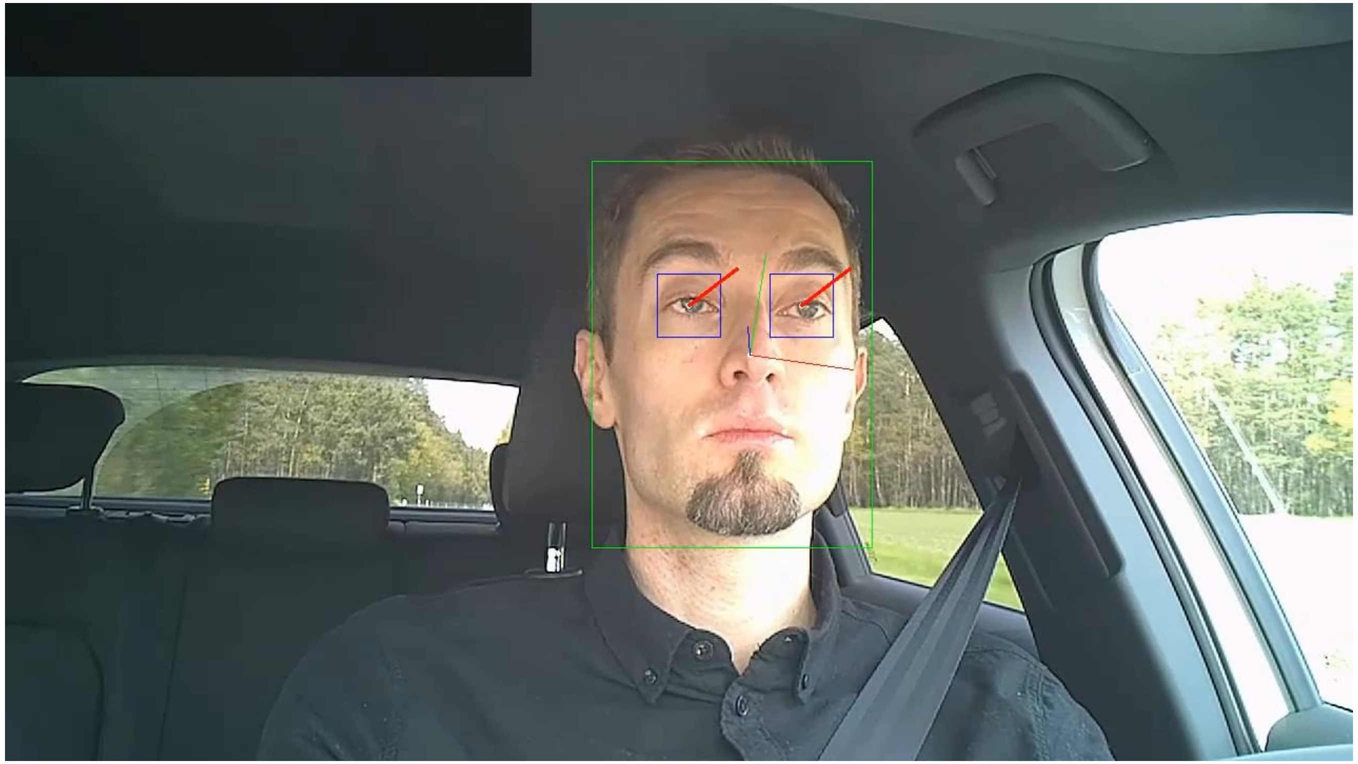
- **To the bottom right**



Distracted Driving (2/4)

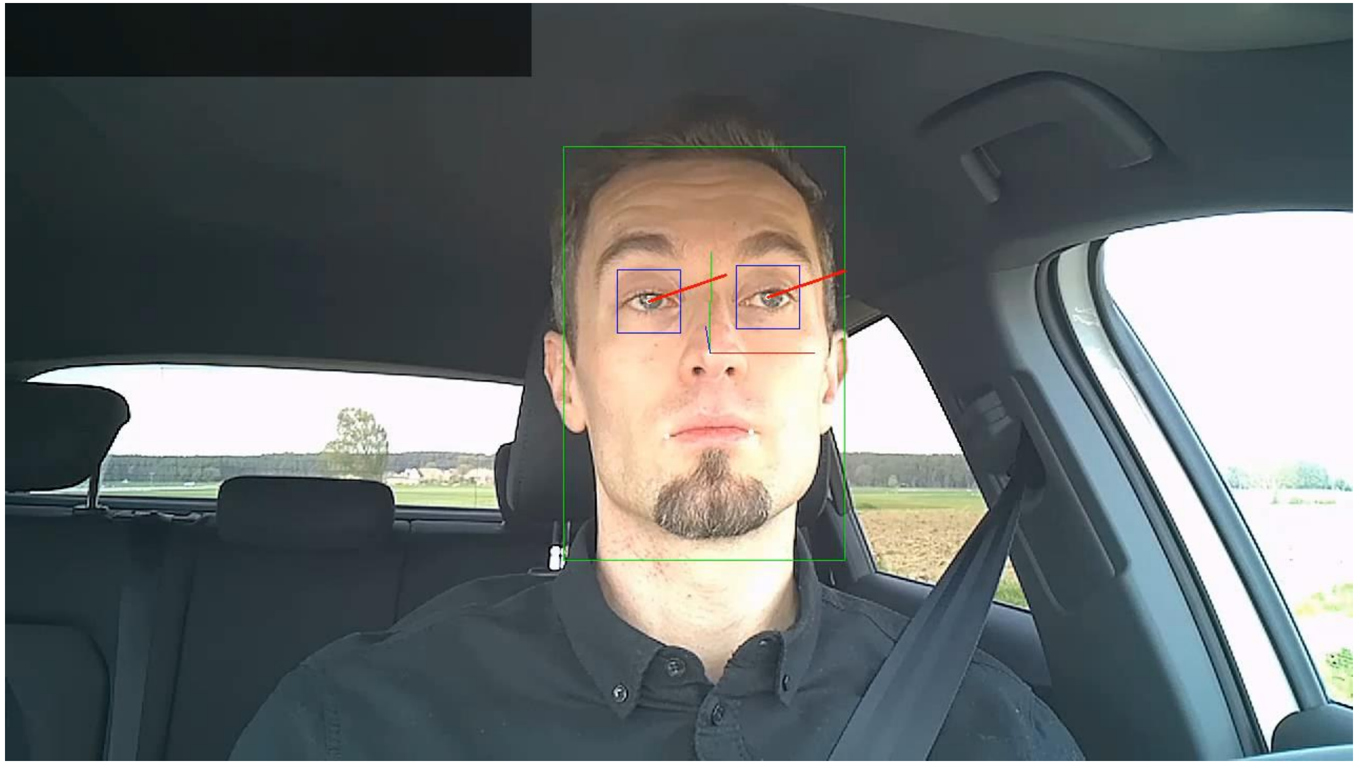
- Sample shows a distracted driver exceeding the allowed distraction period <u>twice</u> with a gaze

- **First down then**
- **To the bottom right**



Distracted Driving (3/4)

- Sample shows a distracted driver exceeding the allowed distraction period slightly <u>twice</u> with a gaze
- **To the left**

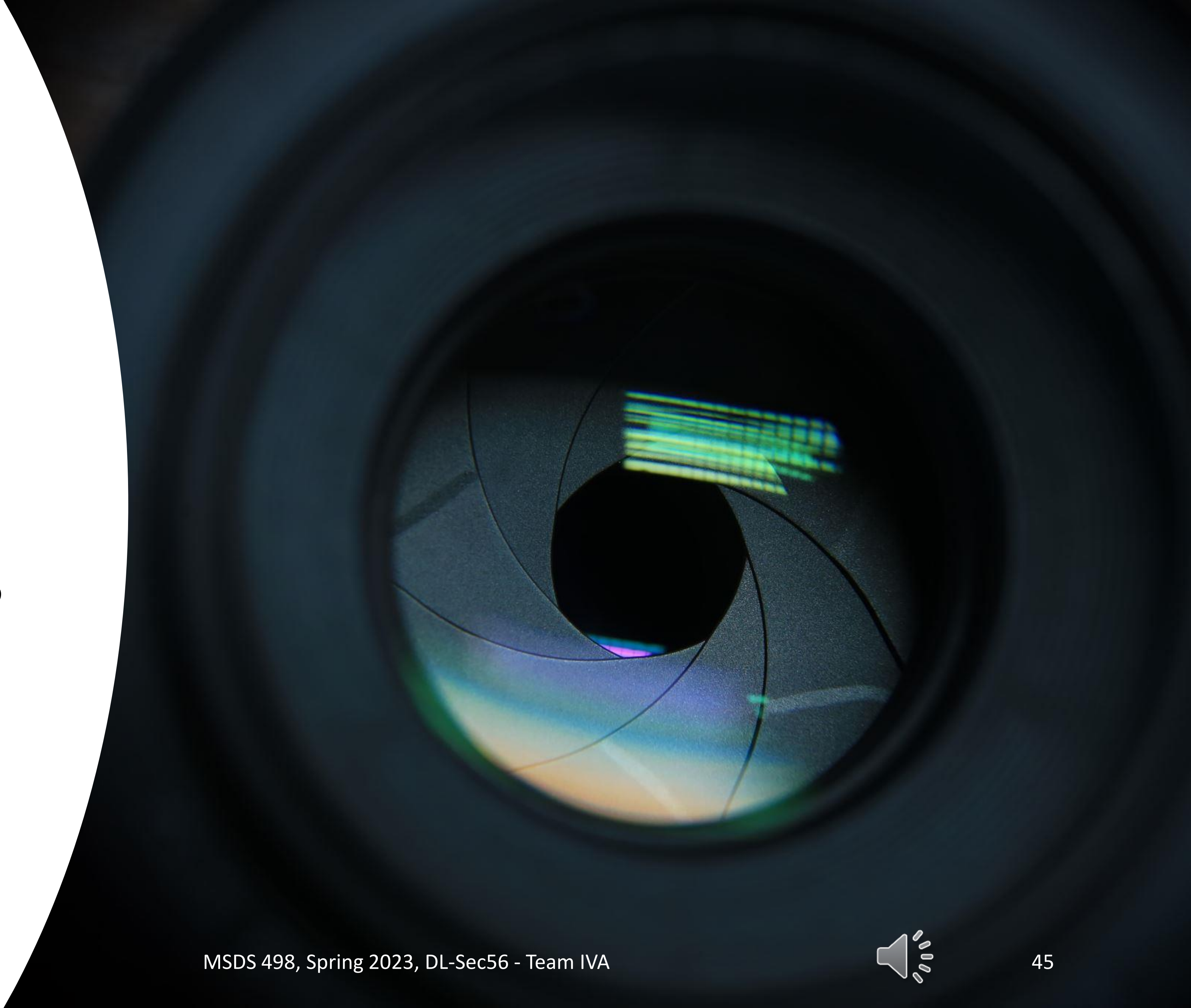


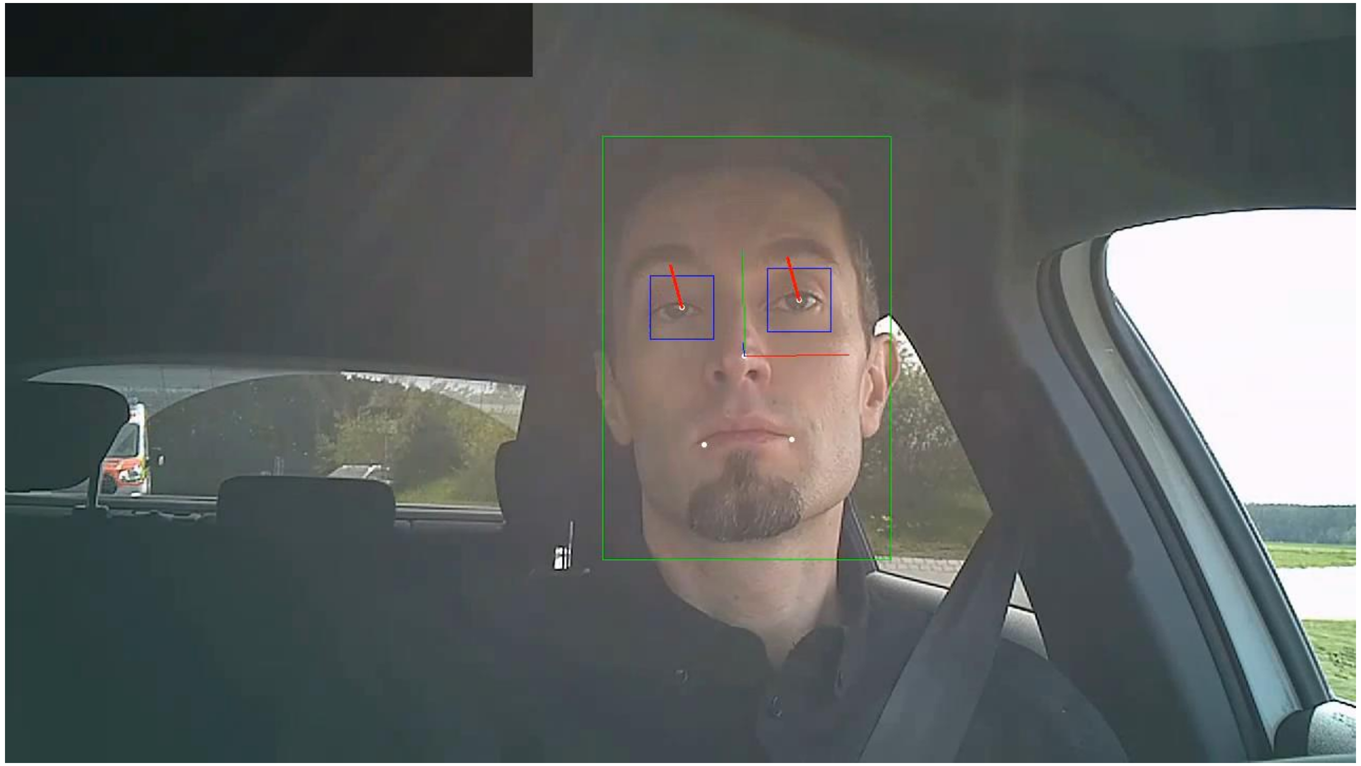
Distracted Driving (4/4)

- Sample shows a distracted driver exceeding the allowed distraction

- **In a different illumination setting** (bright)

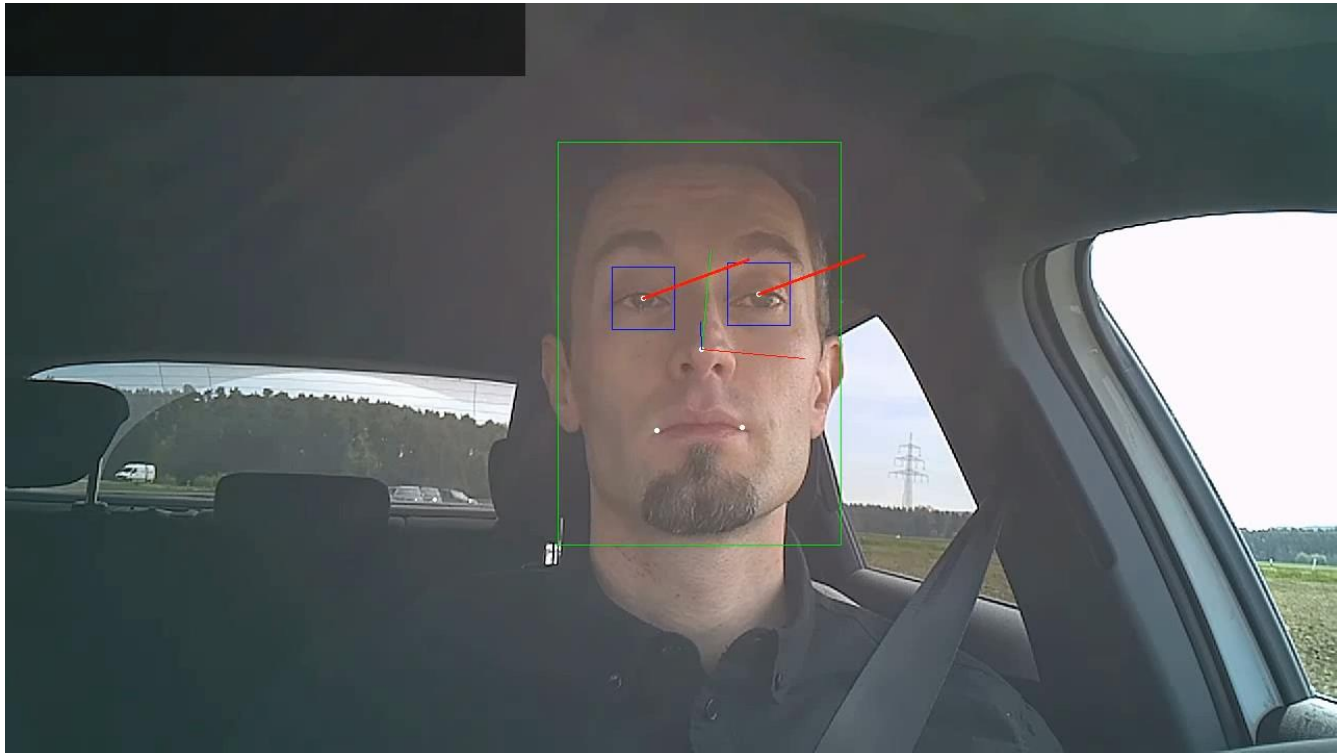
Closed Eyes Driving Scenes





Closed Eyes (1/2)

Sample shows a distracted driver exceeding the allowed distraction period twice.



Closed Eyes (2/2)

Sample shows a distracted driver exceeding the allowed distraction period <u>twice</u>. Once for

- **A prolonged period**
- **A brief period**

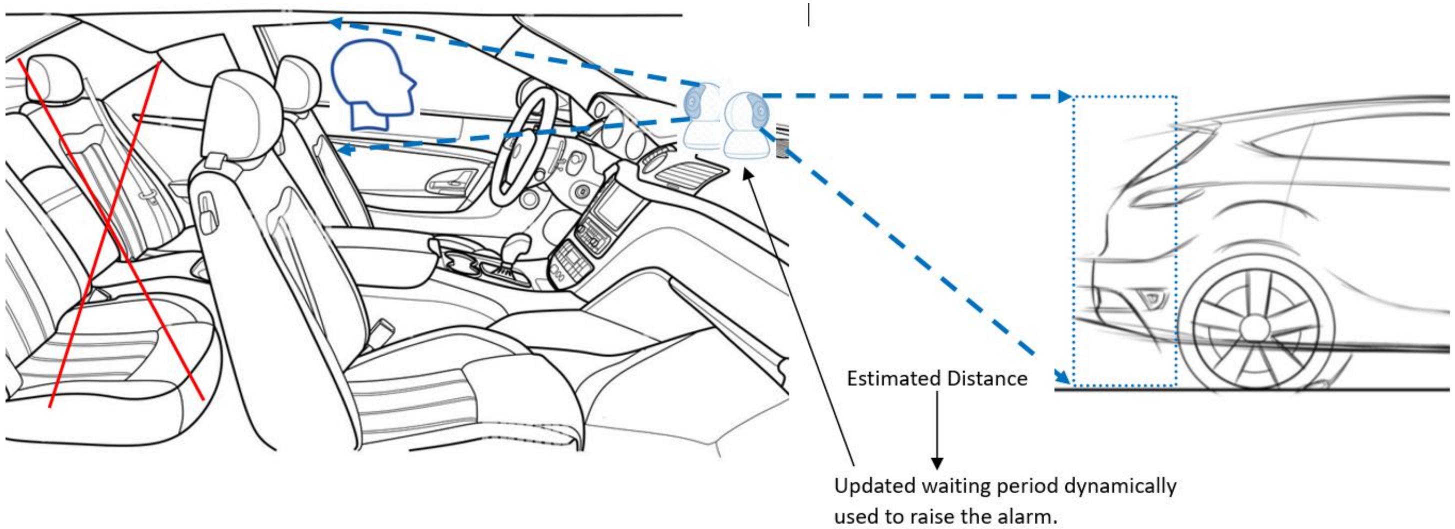
The third times the eyes are closed is slightly below the allowed distracted period.

Possible Extensions to the System

Outlook

Optional Project Extension (1)

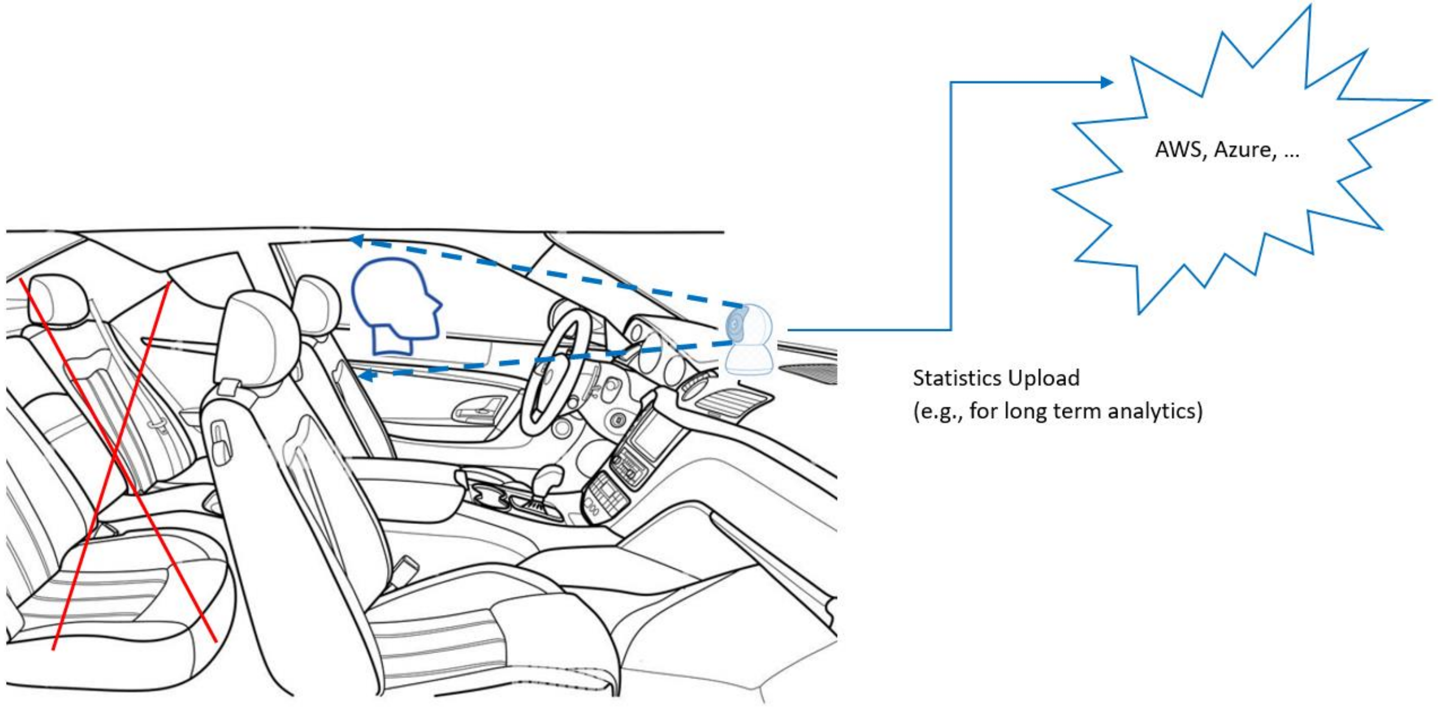
- Integrate a second video feed and a car object tracking model to determine whether a car is in front and if so what the distance to it is.
- Use the estimated distance to dynamically adjust the time of distraction needed to raise the alarm.



Optional Project Extension (2)

Cloud interoperability to

- Upload driver statistics e.g.; to insurance companies or
- Inform police / rescue in case of prolonged closed eyes (possible unconsciousness).



Thank you for your attention