SINTRA ITEA project 2023-2026

Railway Near-Miss Detection with Monocular Camera

Egor Bondarev
Yilei Wang
Giacomo D'Amicantonio
TU/e

Jan Politiek
ARRIVA



ePicture This Workshop 2025
Eindhoven, the Netherlands
28 October 2025

Organized by projects: 2021004 Imagination (Penta)
2023022 Elevation (Xecs)
2024001 Entertain (Xecs)

Background

Train (Arriva, NS) and rail network (ProRail) operators aim to:

- **Detect all hazardous events daily**
- Analyze why these events happen
- Figure out high-risk areas
- Make railway infrastructure more safe
- Make trains more intelligent and safer



Source: Netherlands News

Problem Analysis

Detect all hazardous events daily, how?

- Installing millions cameras everywhere is too expensive
- Manual monitoring is unfeasible and unreliable
- LiDARs are too expensive
- **Video camera-on-train is the only affordable solution**
  - But let's look at 'what' do we need to detect, actually



Source: Arriva Video

Arriva Near-miss Detection

"**Near-miss**“ is incident of highest priority to be detected

- Near-miss is defined as 70 cm lateral distance from object to rail track.
- Train drivers must take emergency action
- No actual collision, but still a serious safety risk

Why Important?

- Happens more often than real accidents
- Provides key insights for safety measures



Source: Arriva Video

Let’s stop for a second

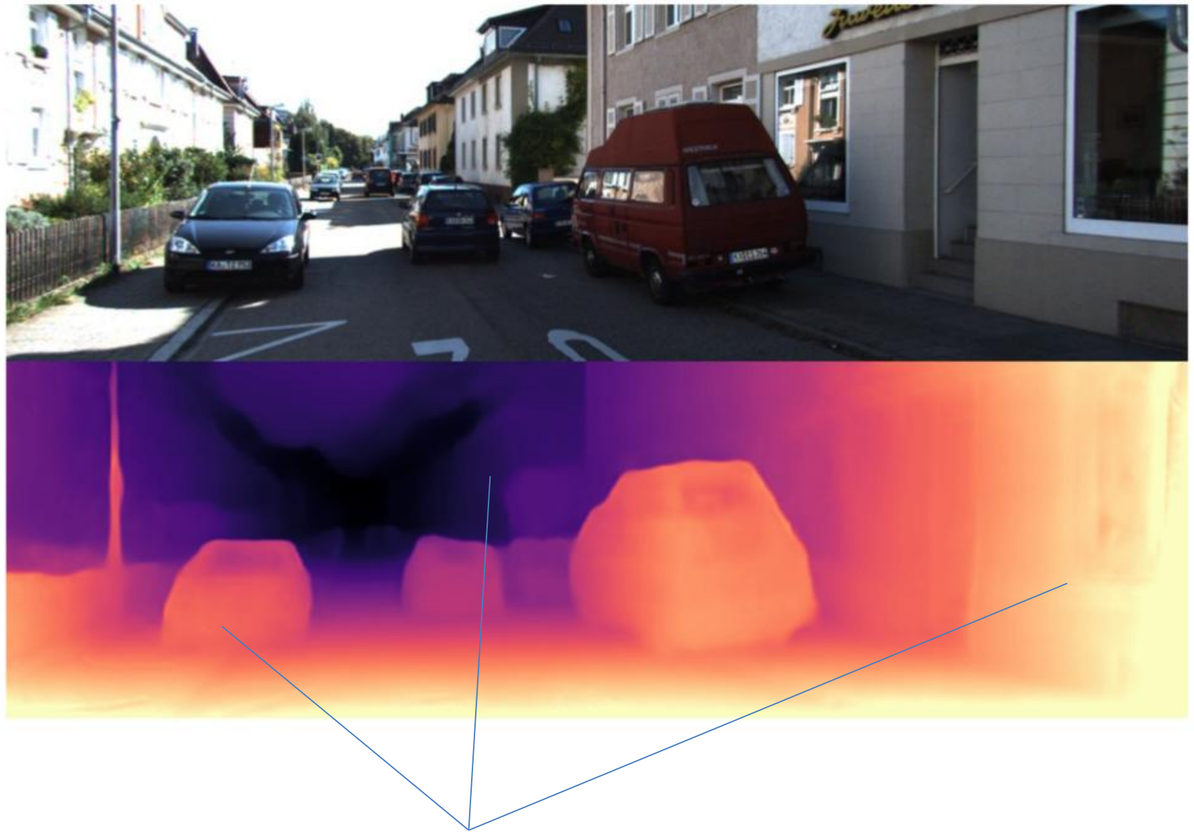
&

look at ‘Near-miss distance estimation’ challenge

Source: Arriva Video

Challenge 1/3

- Current SOTA allows depth estimation of RGB images from monocular camera
- Depth estimation result → abstract depth value (0-255), not the actual distance
- Estimated depth values are very noisy
- Depth-to-distance function is non-linear and unknown
- But there are more fundamental problems…



Depth values in range from 0 to 255

Source: Arriva Video

Challenge 2/3

- We should measure not the ‘camera-to-object’ distance
- But the ‘object-to-object’ distance:
  - Pixels in an image **have no direct correspondence** with a distance value!
- And the worst is coming…



Similar actual distances

Source: Arriva Video

Challenge 3/3

- We need to measure the ‘Lateral Distance’:
  - Definition: side-to-side distance between two points or objects, measured on a horizontal plane
- But people are often not aligned with rails on a horizontal plane
- Are we asked to measure a distance to ‘nothing’?



Source: Arriva Video

Research Question

How can we **estimate accurate lateral distances** between objects and tracks using only a frontal monocular camera?





Part 2

Dataset Pipeline Architecture

Collected Dataset



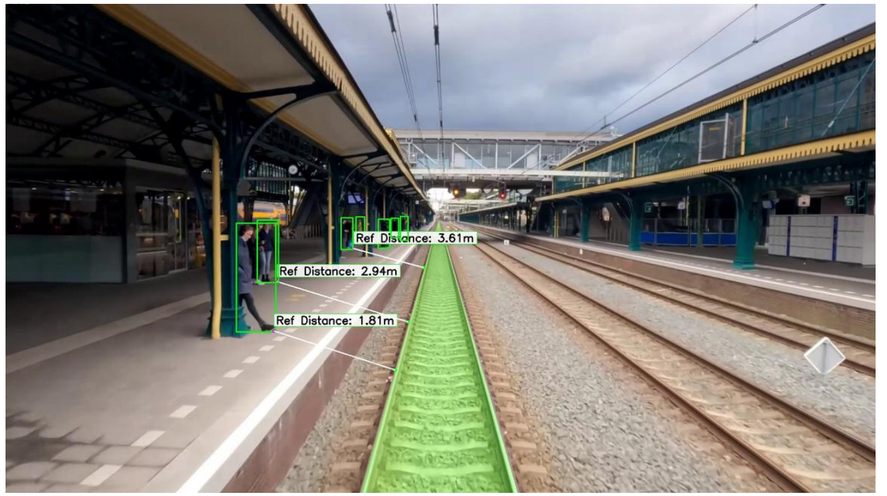
Source: Youtube Video

Dataset Source

Extracted from publicly available train journey on Youtube videos in the Netherlands

Contribution:

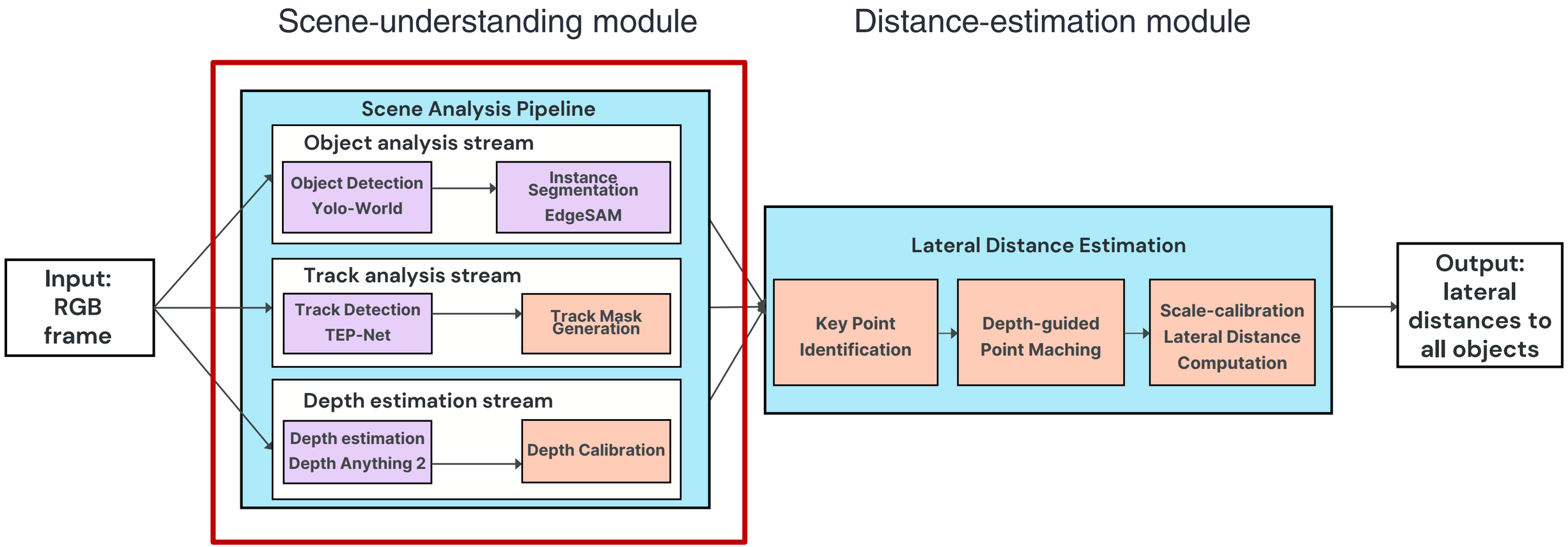
- From Eindhoven ↔ Schiphol Airport – Station Scenes
- 680 high-resolution (1920×1080) images
- 2058 pedestrian annotations
- 20 vehicle annotations
- 680 ego-railway track masks
- 1156 valid lateral distance annotation (<4m)



Distribution of Reference Distance Measurements(<4m)

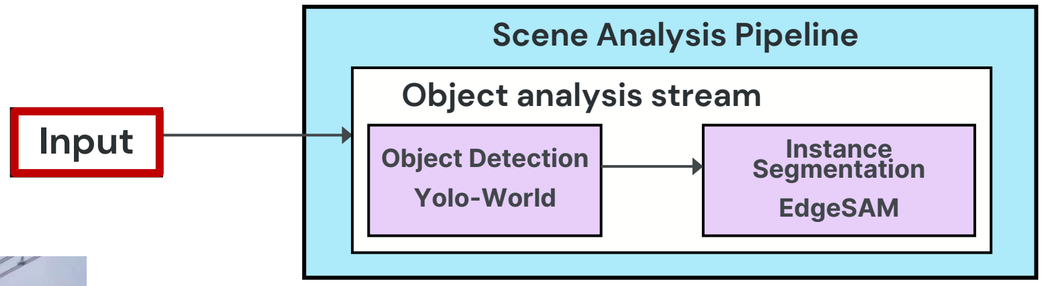
Distance Range	Number of Cases	Percentage (%)
0m - 1m	23	2.0%
1m - 2m	227	19.6%
2m - 3m	774	**67.0%**
3m - 4m	132	11.4%

Pipeline Architecture Overview

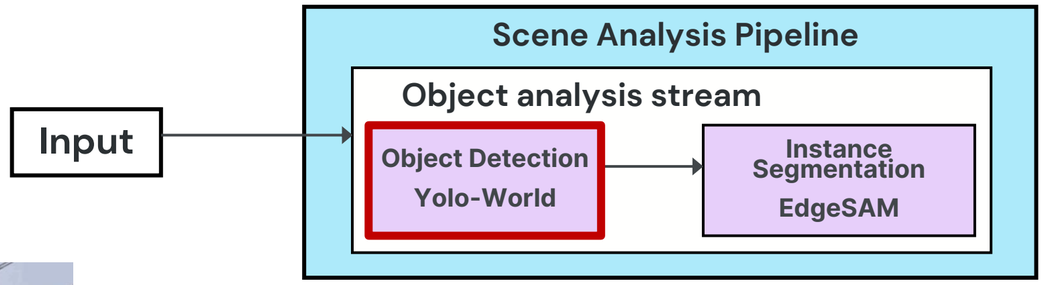


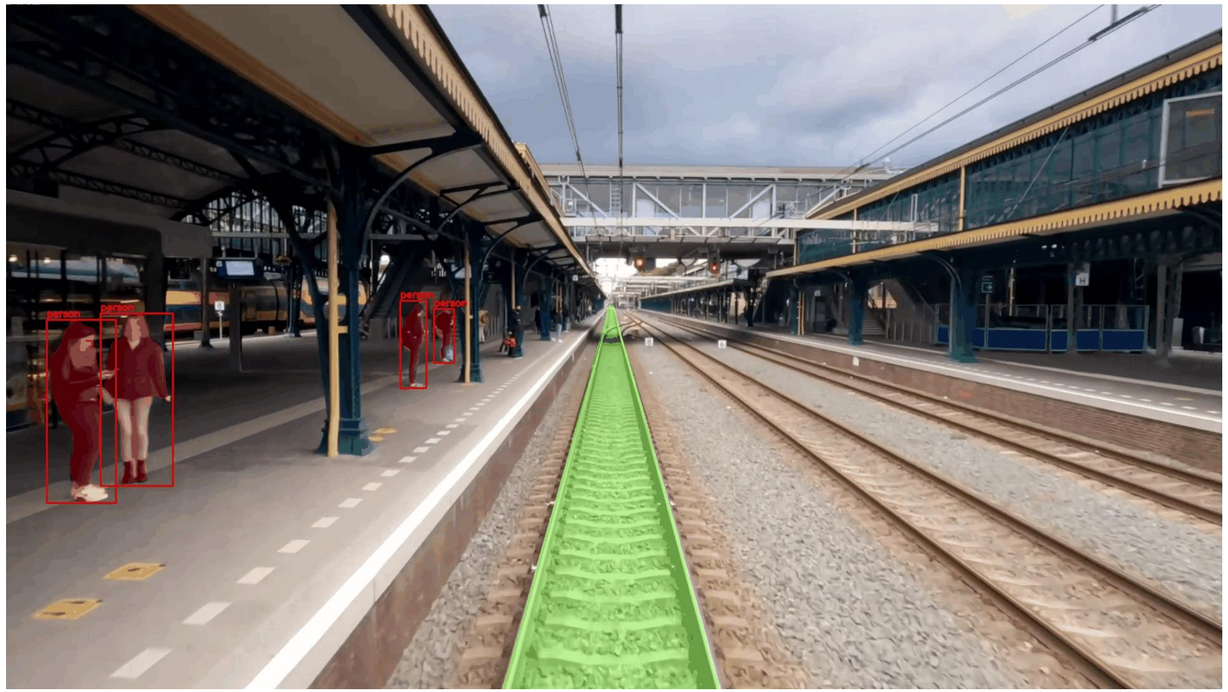
Object analysis stream





Object analysis stream

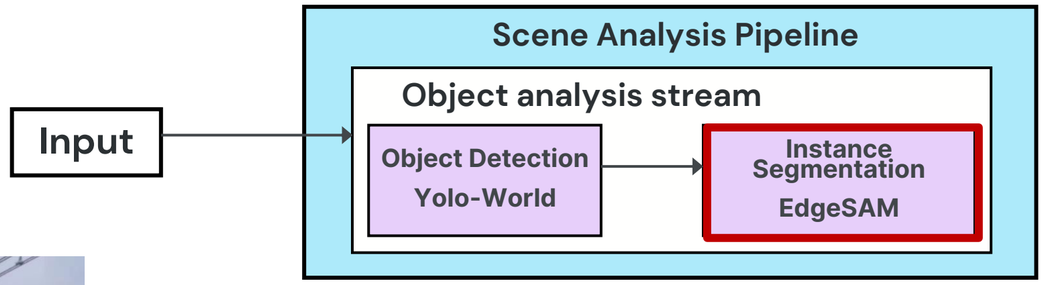




Object detection: Yolo-world

- **Zero-Shot Detection**
  - Recognize unseen objects
- **Enhanced YOLOv8 Framework**
  - Combines a YOLO detector with **CLIP (Transformer-based text encoder)**
- **Detection based on textual descriptions**

Object analysis stream





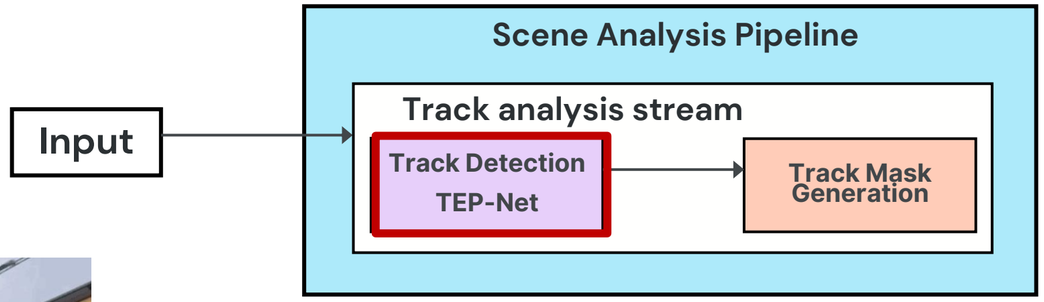
Segmentation: EdgeSAM

- **Fine-grained segmentation**
  - Utilizes YOLO-World's bounding boxes as input

- **SAM with lightweight CNN backbone**
  - RepViT-M1 replaces ViT structure
  - 37x faster while maintaining segmentation accuracy.

Track analysis stream

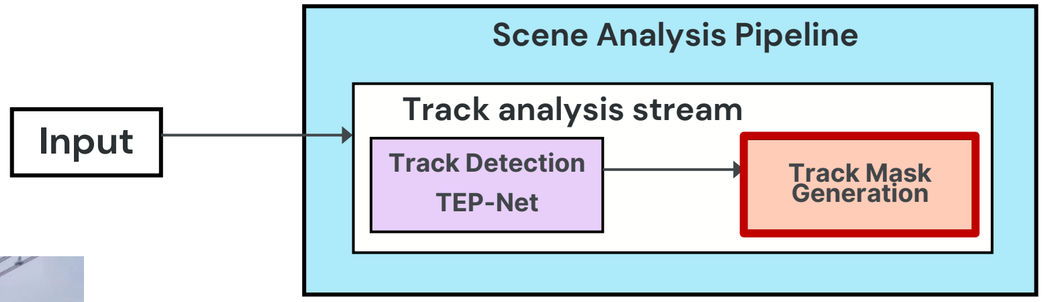




Track detection: TEP-Net

- **Regression-based Method**
  - Directly predict the x- and y-coordinates of the track boundaries.

Track analysis stream

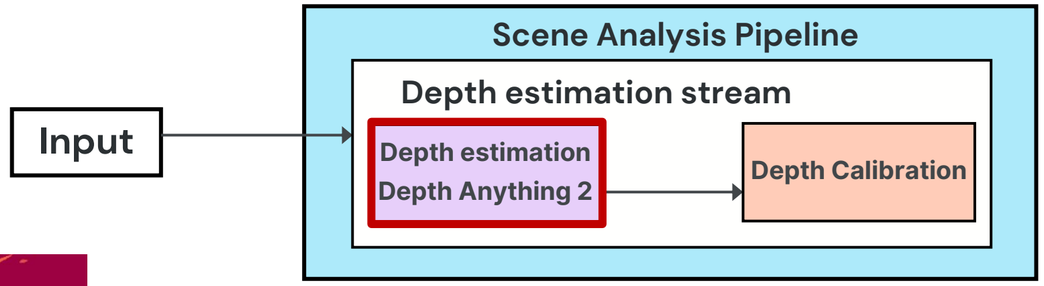


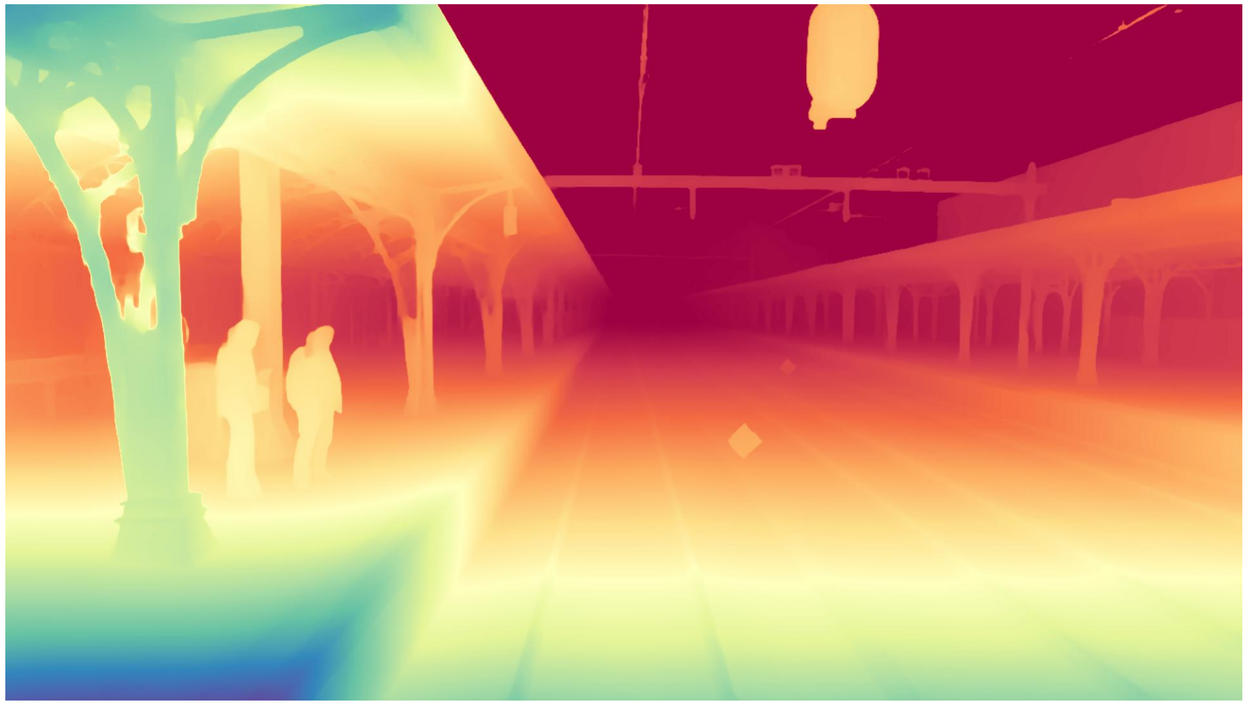


Track mask generation:

- **Coordinate-to-Mask Conversion**
  - Morphological refinement to accurately define track edges

Depth estimation stream

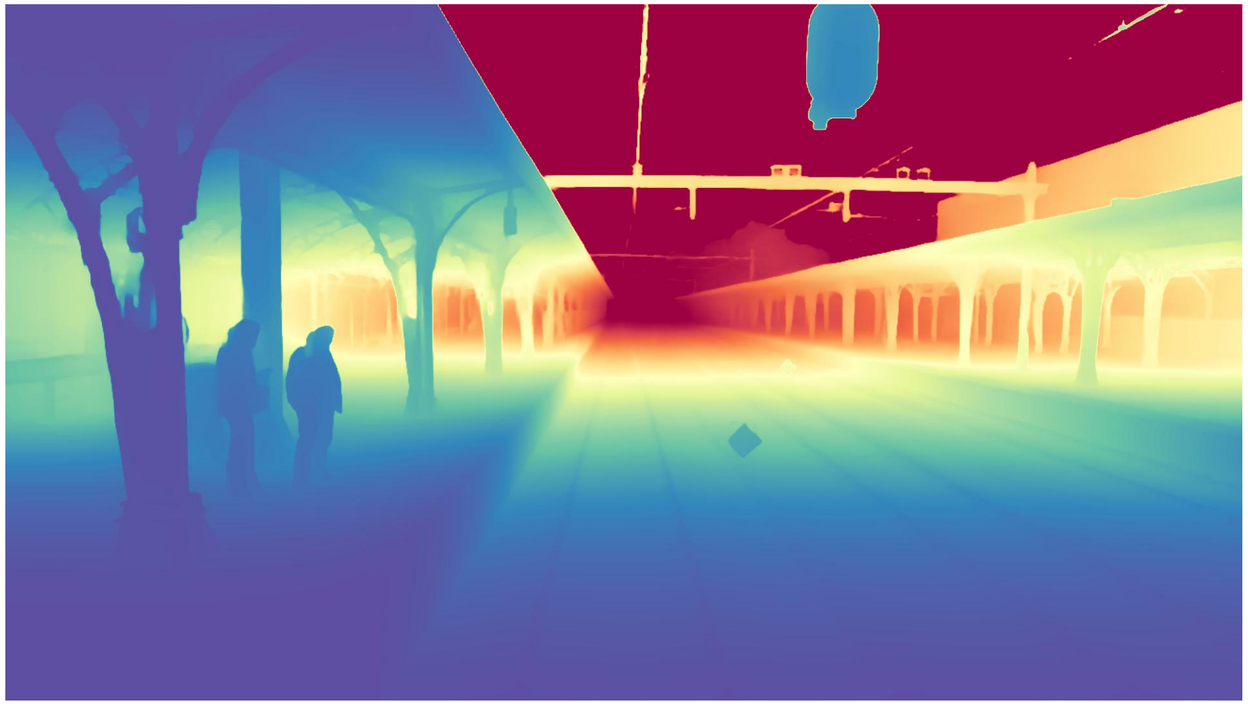


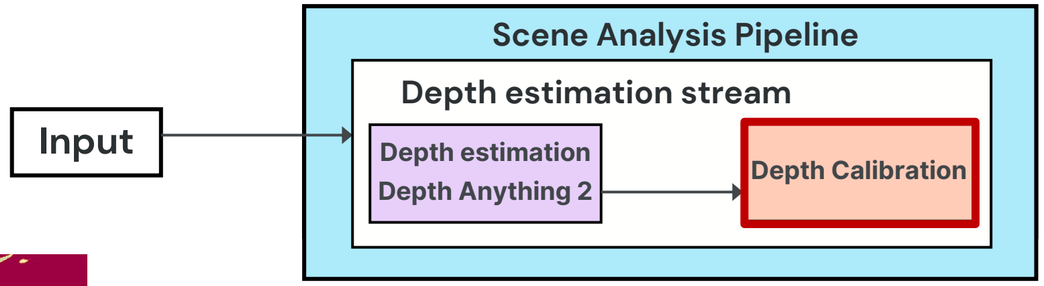


Depth Anything V2:

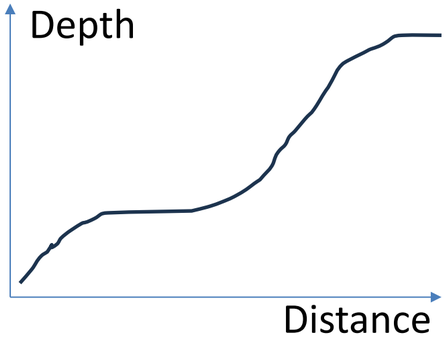
- Dense Prediction Transformers (DPT) - based architecture
- Large-Scale Training
- Strong generalization ability

Depth estimation stream





Depth Calibration:



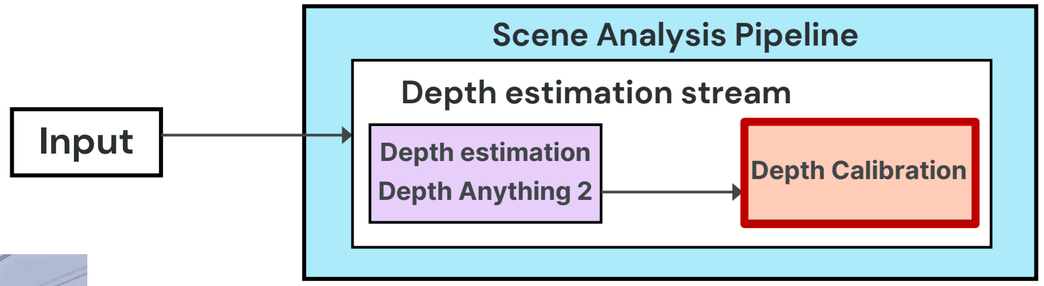
- **Problem to solve**
  - Actual depth-to-distance function is unknown and is non-linear

- **Two-Stage Depth Calibration**
  - **Stage 1**: Use sleeper spacing (0.6m) for initial depth calibration.
  - **Stage 2:** Refines long-distance depth using track geometry constraints.

  - Converts relative depth into real-world depth measurements.

Distance Calibration





Two-Stage Depth Calibration

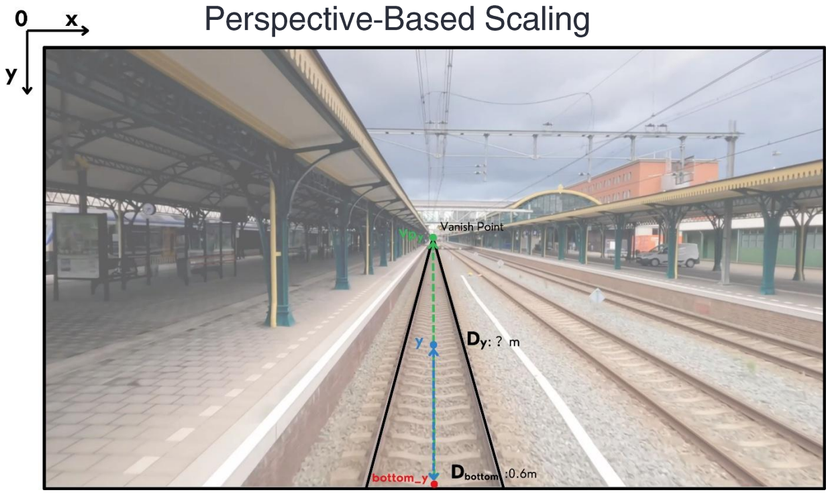
Stage 1: Sleeper-Based Calibration

- Leverages sleeper positions (0.6m apart) to create depth reference points.
- Uses exponential curve fitting to model the non-linear depth relationship.

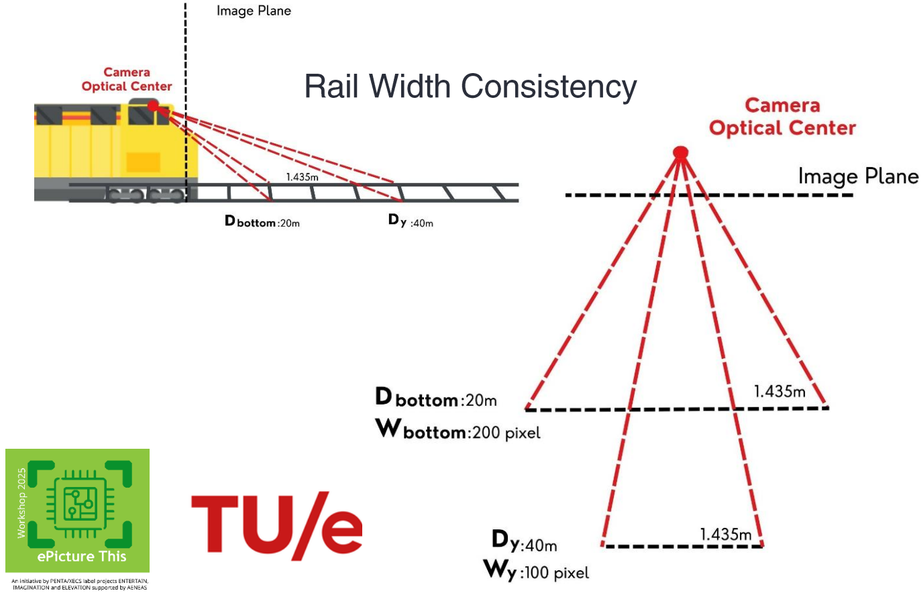
$$D_{calibrated} = a \cdot e^{bD_{raw}} + c$$

D_{raw} is the raw depth estimate from DA V2, and parameters *a, b, c* are calculated through curve fitting using known sleeper-spacing measurements.

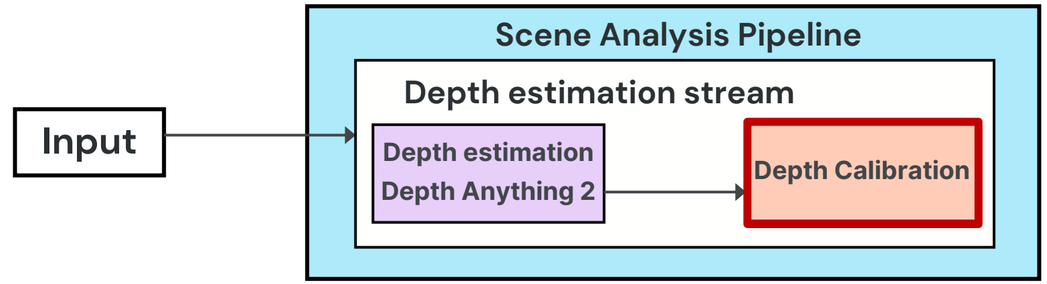
Distance Calibration



$$\frac{D(y)}{D_{bottom}} = \frac{vp_y - y}{vp_y - bottom_y}$$



$$\frac{D(y)}{D_{bottom}} = \frac{W_{bottom}}{W(y)}$$



Two-Stage Depth Calibration

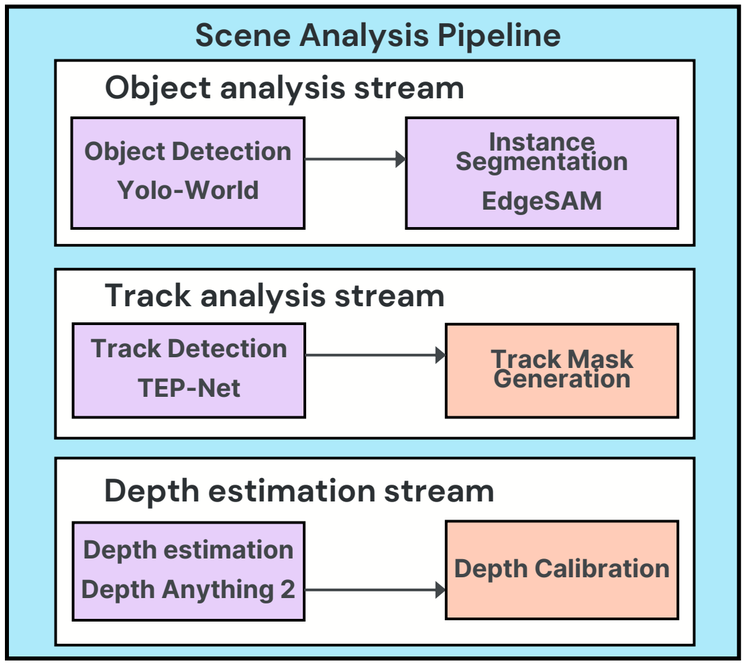
Stage 2: Geometric-Based Calibration

- Improves far-field accuracy using railway track geometry.

Two geometric constraints:

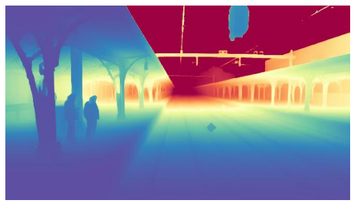
- **Perspective-Based Scaling** - Uses the vanishing point to infer depth relationships.
- **Rail Width Consistency** - Using similar triangles, estimate depth based on this shrinking width in image.

Distance estimation pipeline, so far:

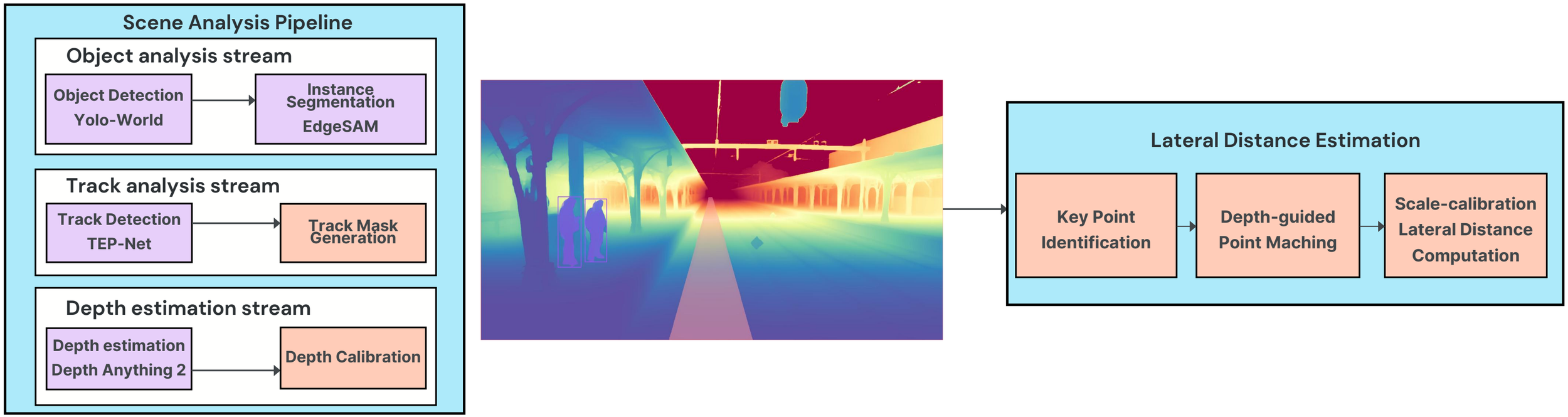




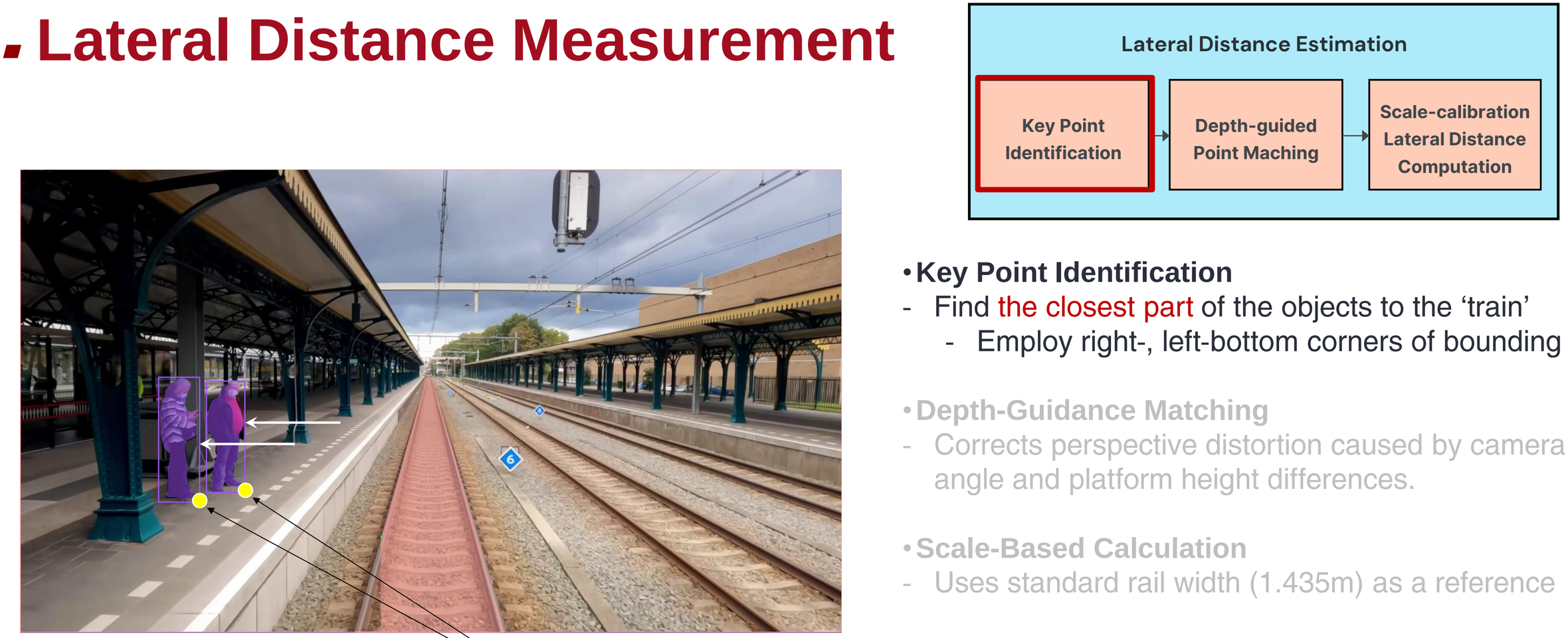




Towards lateral distance estimation



Lateral Distance Measurement



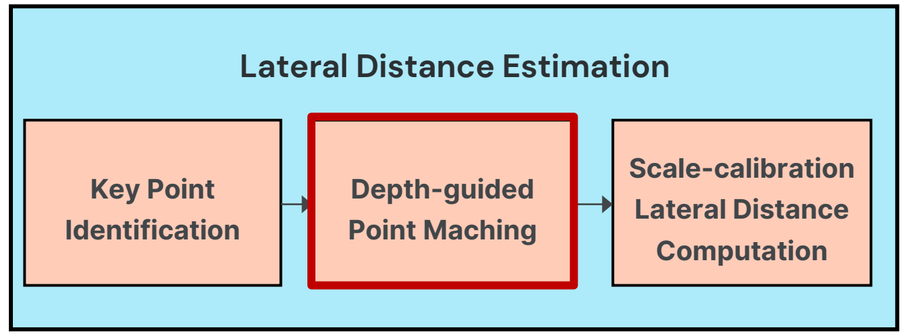
- **Key Point Identification**
  - Find the closest part of the objects to the ‘train’
    - Employ right-, left-bottom corners of bounding box

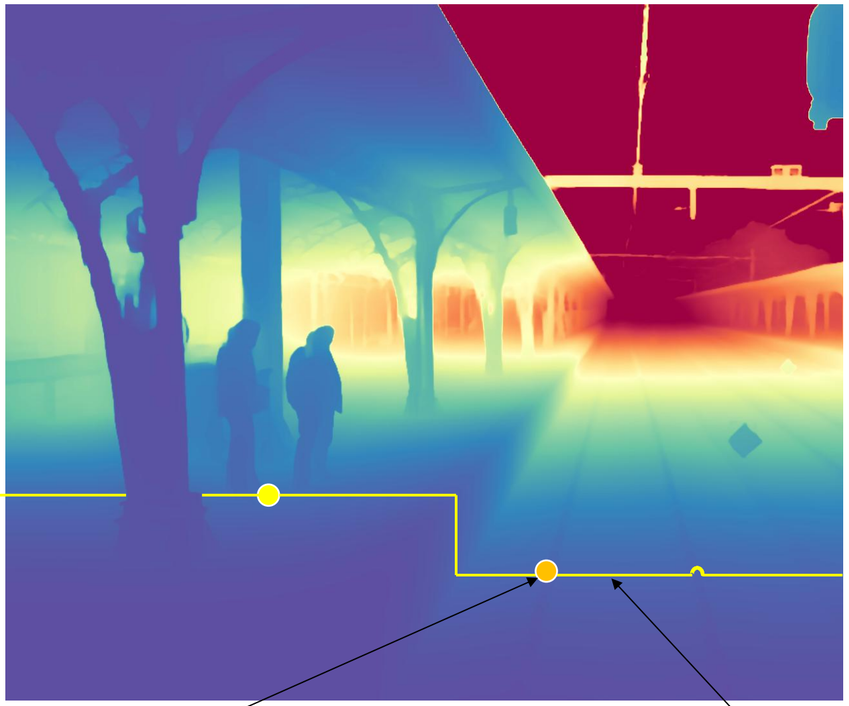
- **Depth-Guidance Matching**
  - Corrects perspective distortion caused by camera angle and platform height differences.

- **Scale-Based Calculation**
  - Uses standard rail width (1.435m) as a reference

Nearest object’s points to a hypothetical train

Lateral Distance Measurement





Match: Nearest rail-track pixel with the depth similar to the object depth

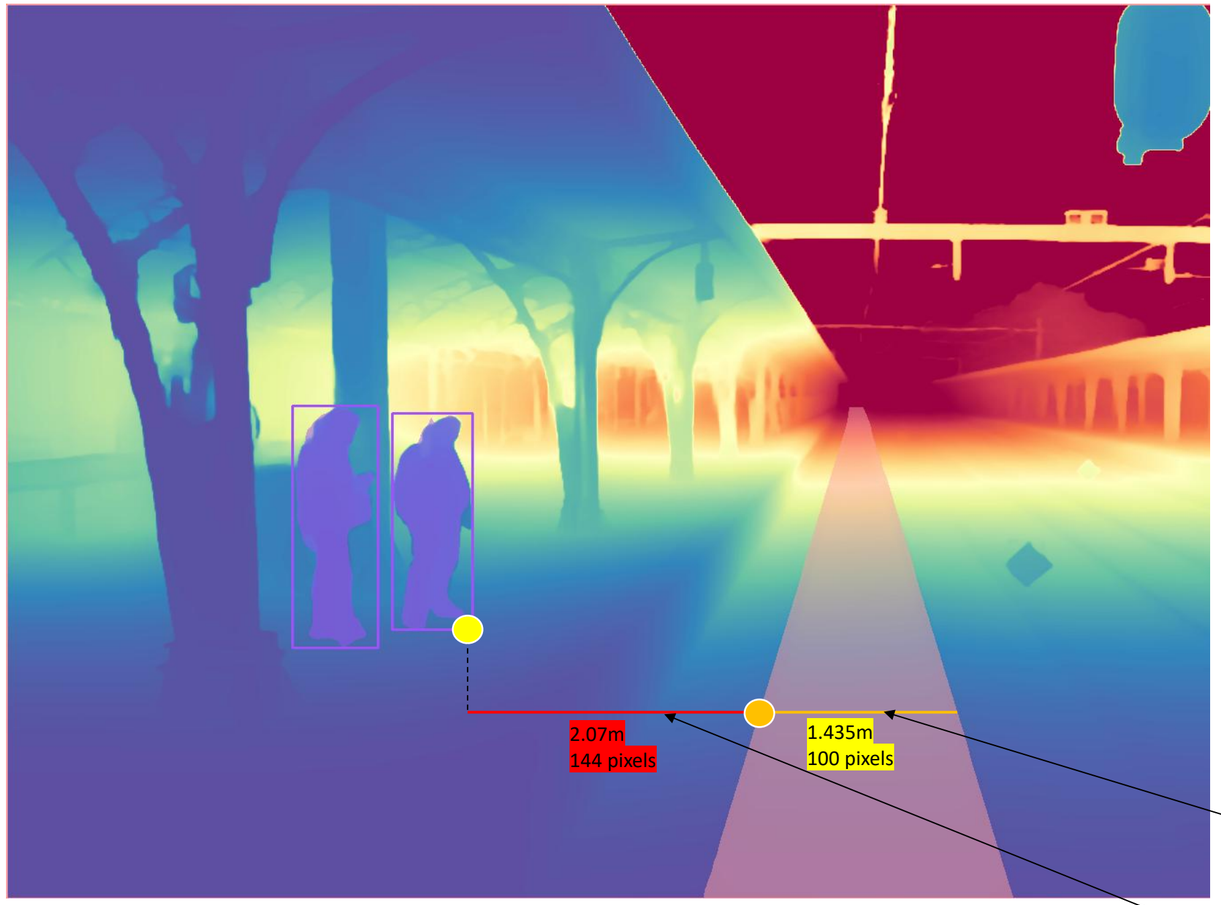
Single depth-value bitmap

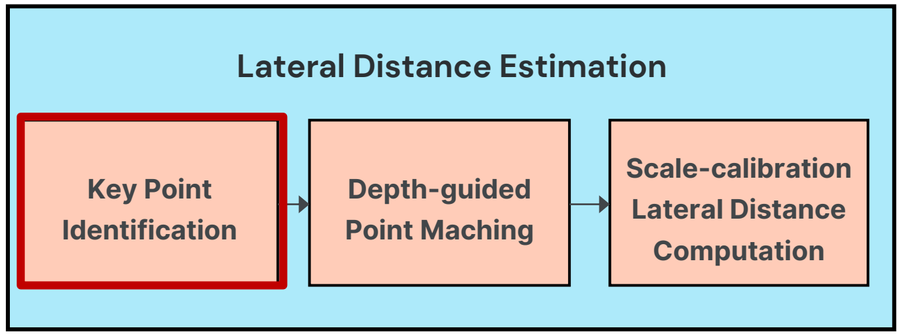
- **Key Point Identification**
  - Locate the closest part of the object to the rail

- **Depth-Guidance Matching**
  - Corrects perspective distortion caused by camera angle and platform height differences.
  - For each object:
    - Create single depth-value bitmap: find all pixels with the depth value of the object's nearest point
    - Find the pixel-pair: select the pixel belonging to the nearest rail-track

- **Scale-Based Calculation**
  - Uses standard rail width (1.435m) as a reference

Lateral Distance Measurement





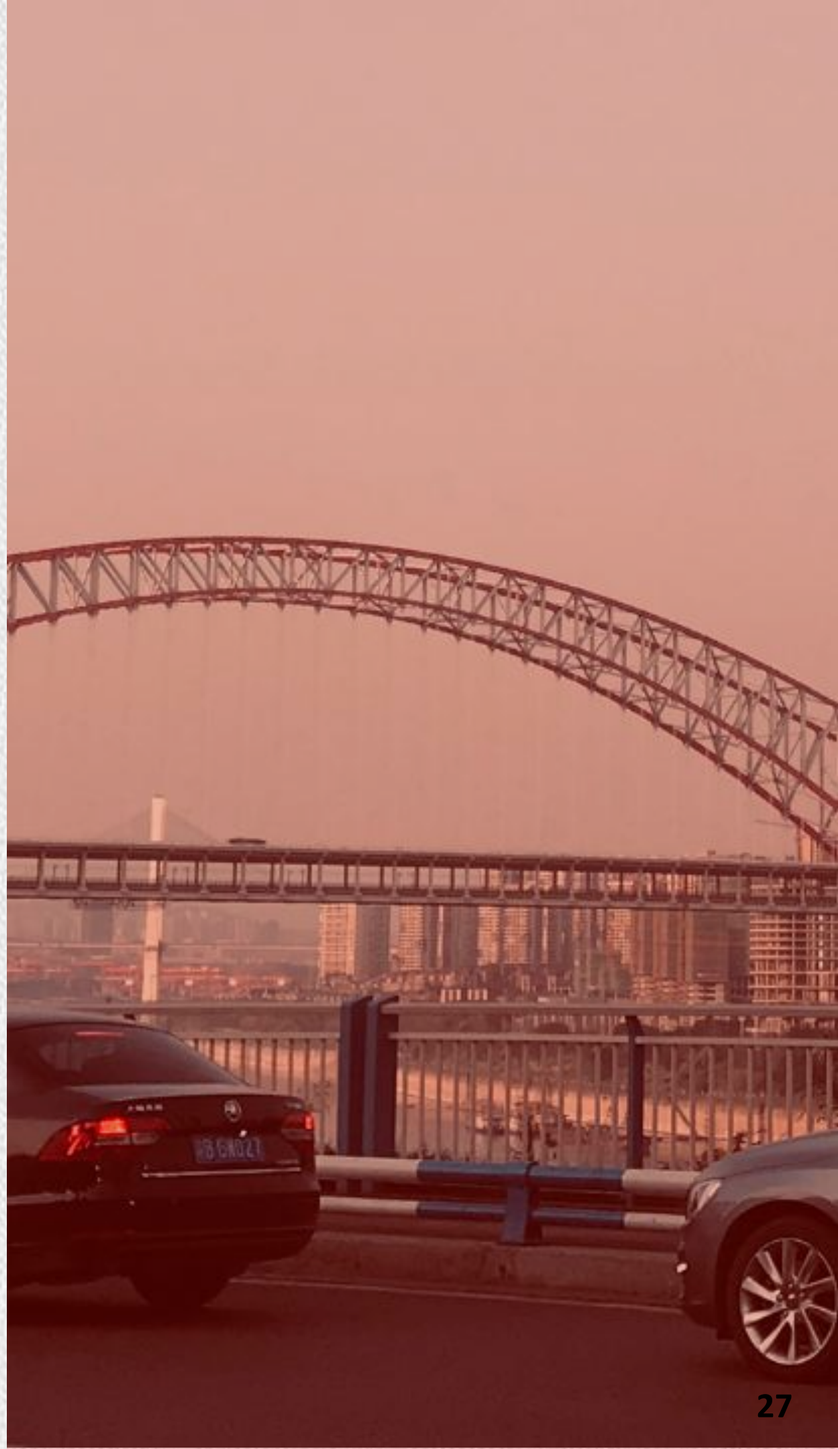
- **Key Point Identification**
  - Locate the closest part of the object to the rail.

- **Depth-Guidance Matching**
  - Corrects perspective distortion caused by camera angle and platform height differences.

- **Scale-Based Calculation of Lateral Distance**
  - Uses standard rail width (1.435m) as a reference.

Part 3

Experimental Results Conclusions



Lateral Distance Measurement Performance

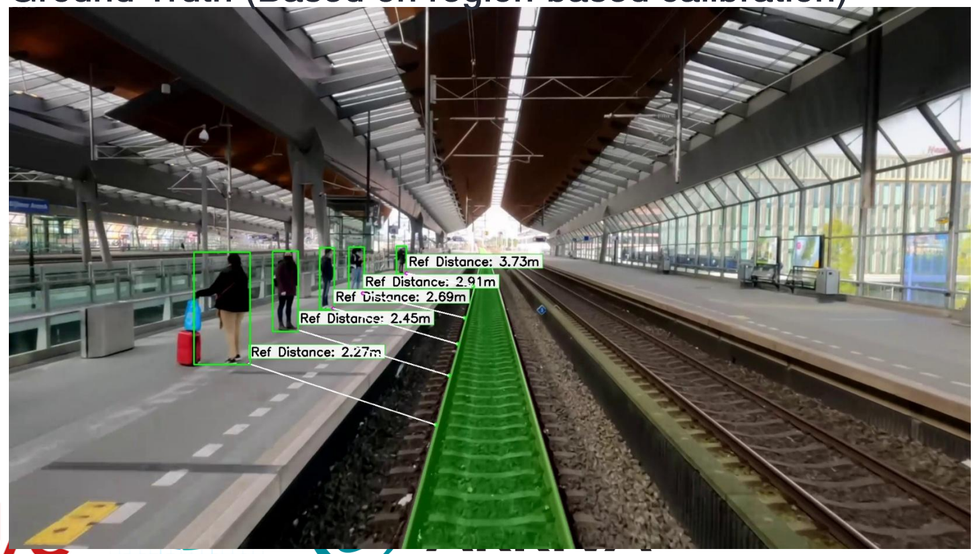
Error Metrics (m)		
Mean	Std	Max
0.290	**0.195**	**2.581**

Distribution of Distance Measurement Errors

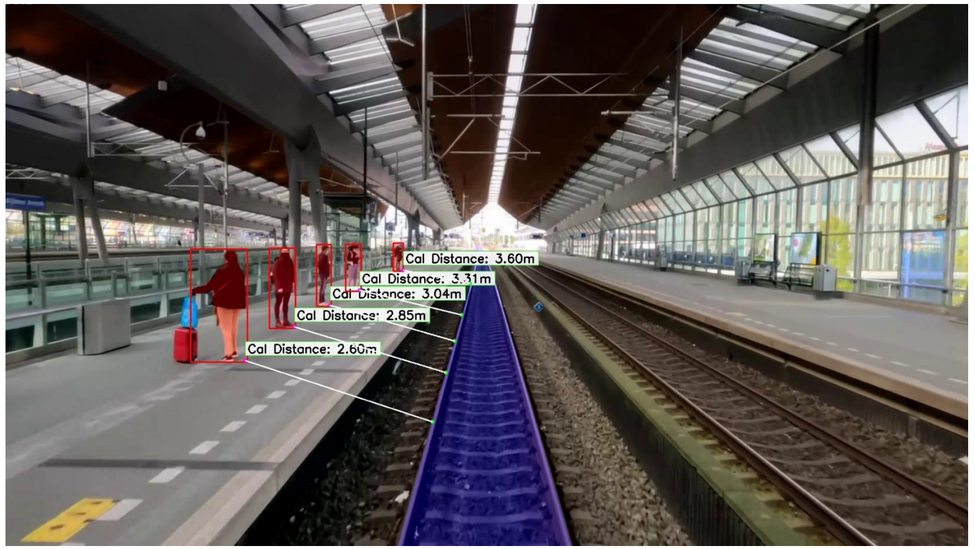
Error Range	Number of Cases	Percentage (%)
$\leq$ 0.1m	95	8.2%
0.1m - 0.3m	592	**51.2%**
0.3m - 0.5m	363	31.4%
$>$ 0.5m	106	9.2%

59.4%

Ground Truth (Based on region-based calibration)



Predict Results



Showcase of project results

Intersecting tracks Scene

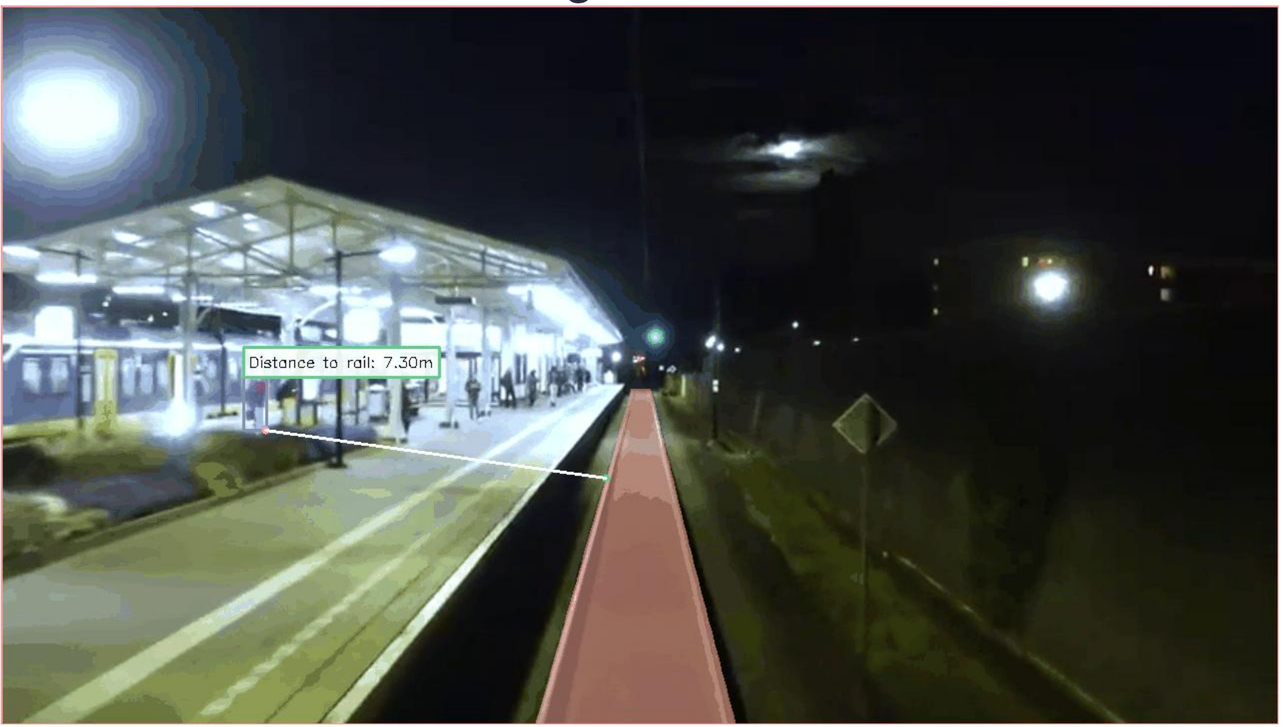


Crowed Scene



Showcase of project results

Night Scene



Field Scene



Metrics and Conclusion



Accuracy

- Mean-accuracy lateral distance estimation: 0.290m

Near real-time in railway monitoring

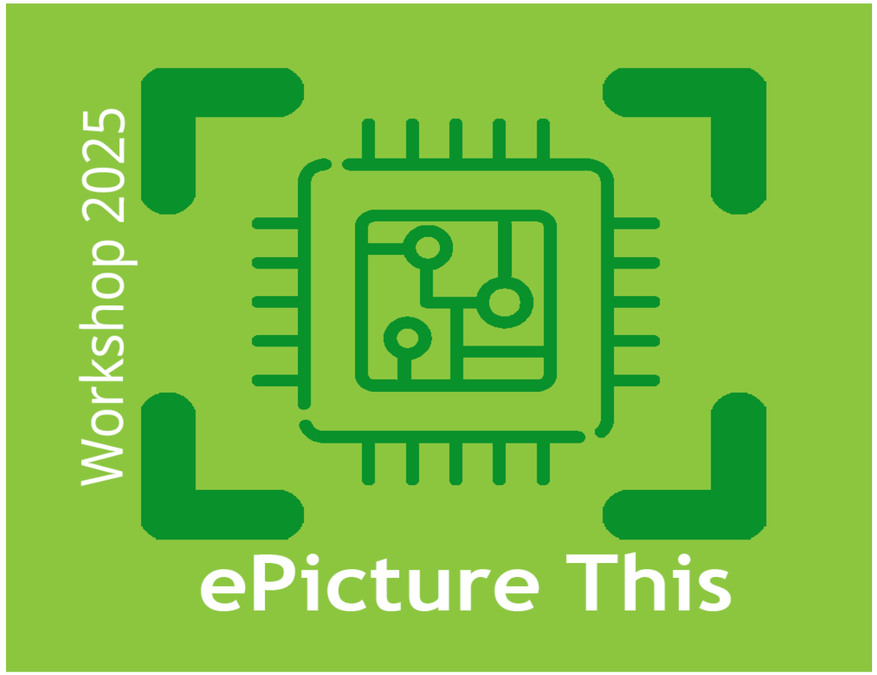
- 2.51 FPS (0.399s per image), fast railway monitoring

System Robustness

- Performs well in low-light conditions, occlusions, and crowded railway platforms.

Simple deployment

- Works with a single monocular camera
- Camera calibration parameters are not necessary



THANK YOU

An initiative by PENTA/XECS label projects ENTERTAIN, IMAGINATION and ELEVATION supported by AENEAS

Penta Xecs EURIPIDES² European Smart Electronic Systems Aeneas

Workshop 2025
ePicture This
An initiative by PENTA/XECS label projects ENTERTAIN, IMAGINATION and ELEVATION supported by AENEAS

TU/e AIMS lab ARRIVA



Thank you for your attention!

Precise Lateral Distance Calibration with Real-World References



Key Reference Measurements

- Track to Platform edge: **0.8m**
- Platform to white brick: **0.9m**
- Bricks to poles : **0.4m**
- Poles to tactile paving: **2.0m**

Total distance (track → tactile paving): **4.1m**



Region-Based Calibration for Accurate Lateral Distance Measurement

- **Region-Based Calibration**
  - Divides image into fixed measurement regions (0.8m, 0.9m, 0.4m, 2.0m).
  - For each region, compute a local homography transformation matrix using the pixel coordinates of region boundary line endpoints and their corresponding real-world distances.
  - Uses homography transformation to correct perspective distortion.
  - Ensures accurate pixel-to-world coordinate mapping for reliable measurement.

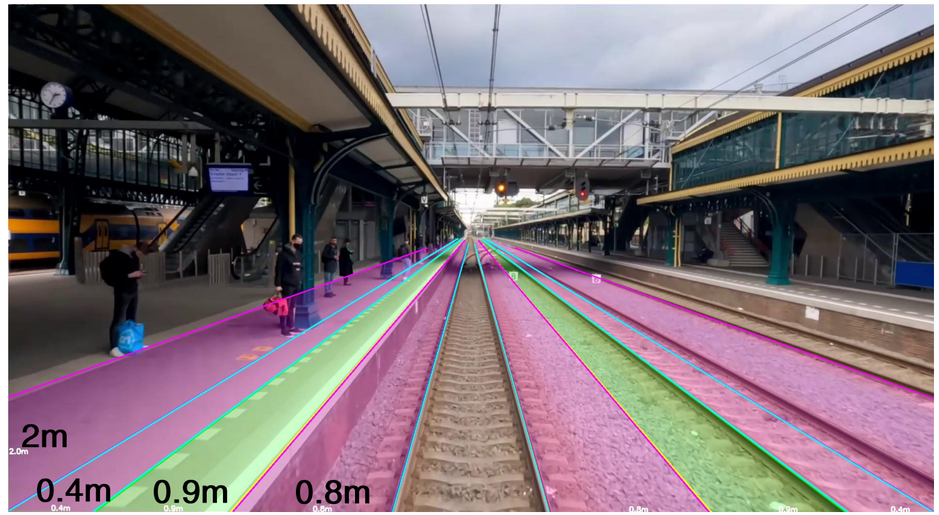
Homography Transformation Formula:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = H_r \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

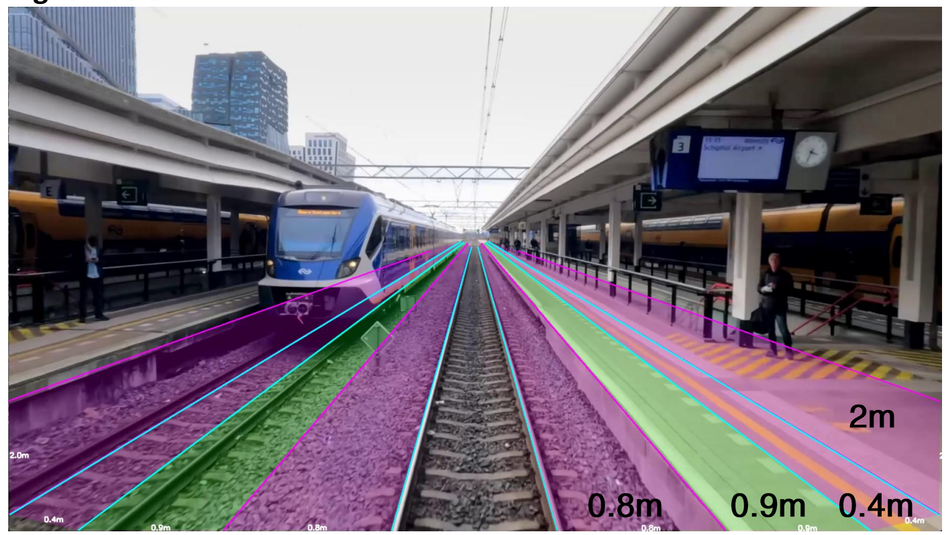
where (x, y) represents the pixel coordinates in the image,
and (x′, y′) represents the corresponding real-world coordinates in meters.

- **Measurement Consistency**
  - Lateral distance is symmetrically mapped using the railway centerline
  - Local homography per region ensures stable multi-zone distance measurements.

Left Platform



Right Platform



Limitations of Existing Datasets & Our Contribution

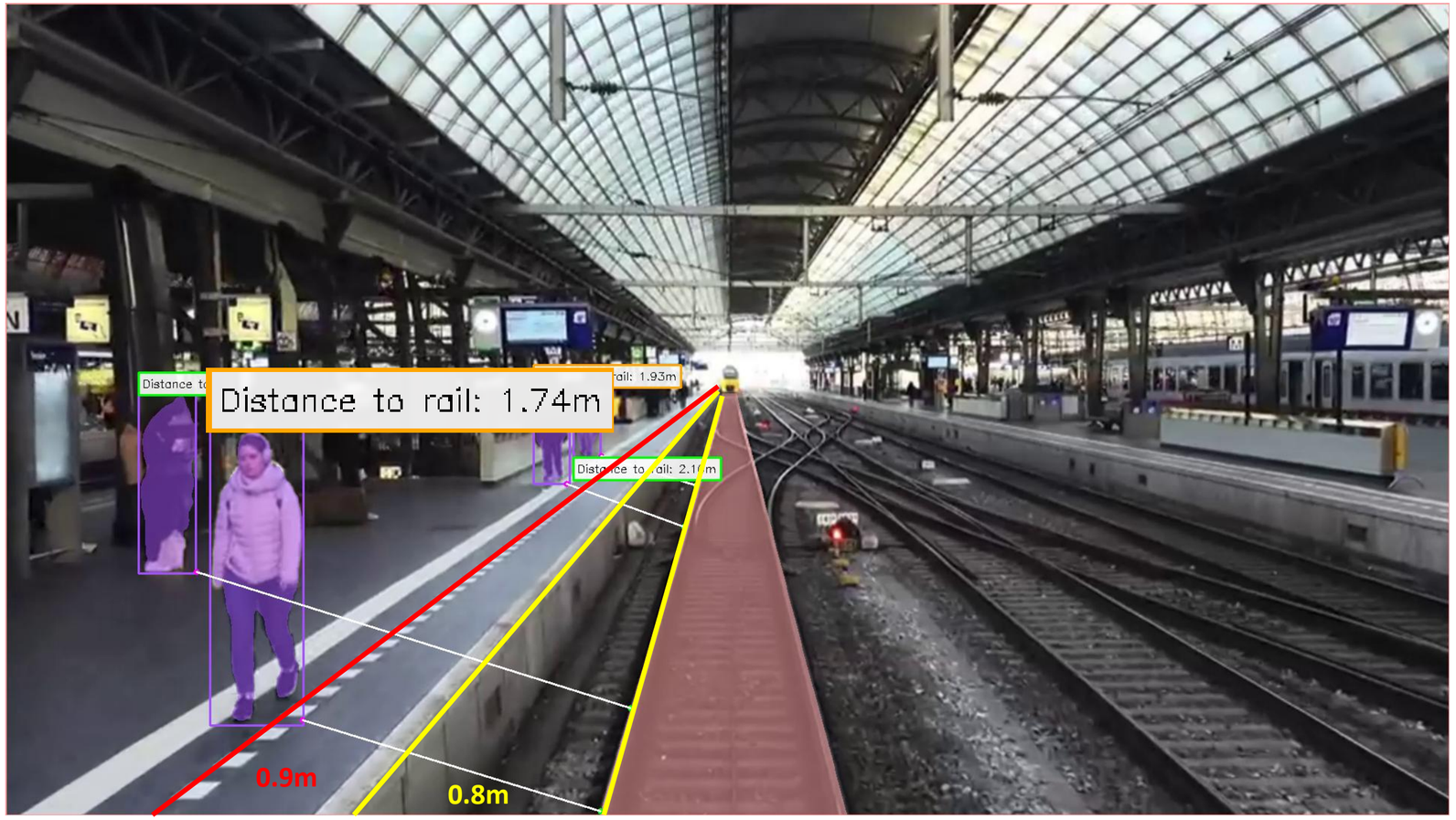
Dataset	Year	Seg.	Det.	Depth	Lateral.	Classes
RailSem19 [39]	2019	✓	✓	✗	✗	19
RAWPED [40]	2023	✗	✓	✗	✗	1+
RailSet [41]	2023	✓	✓	✗	✗	5+
OSDaR23 [42]	2023	✓	✓	✓	✗	20+
Our(Validation)	2024	✓	✓	✗	✓	2

Note: Seg.: Semantic Segmentation; Det.: Object Detection; Lateral.: Lateral Distance

Existing railway datasets (RailSem19, RAWPED, RailSet, OSDaR23) focus on object detection and segmentation but **lack precise lateral distance annotations.**

Qualitative Analysis of Lateral Distance Measurement

This qualitative analysis demonstrates the **effectiveness** of our lateral distance measurement approach.

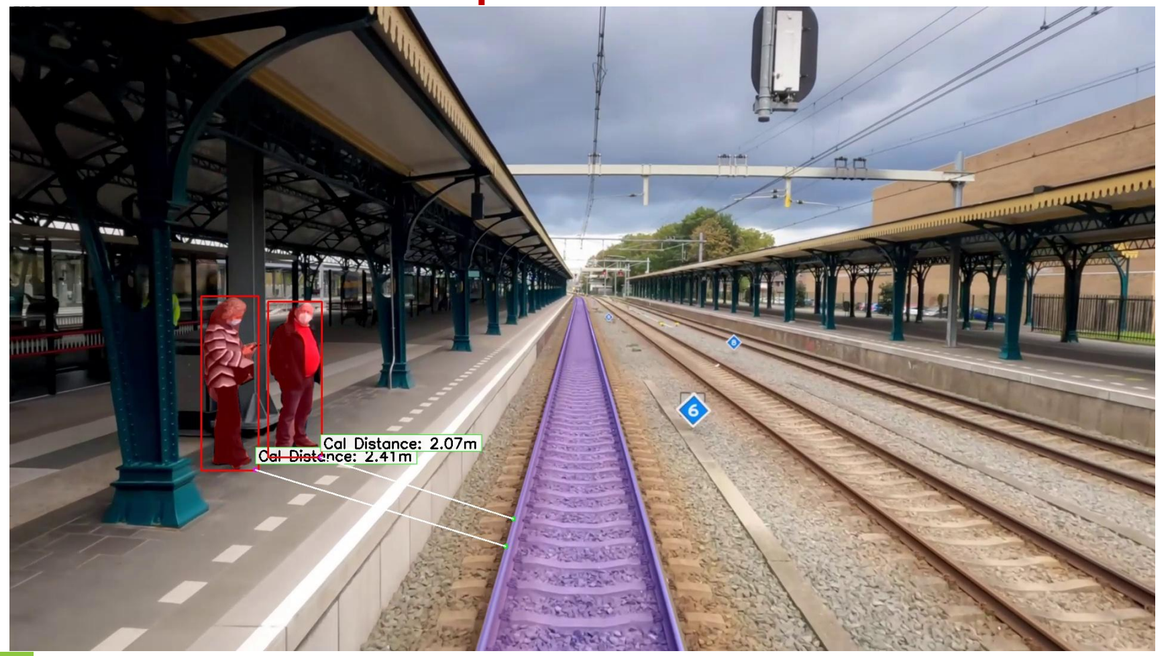


Ablation studies

Impact of Depth-Guided and Calibration

Method	Error Metrics (m)			Error Dist. (%)		
	Mean	Std	Max	<0.3	0.3-0.5	>0.5(m)
Full System	**0.290**	**0.195**	**2.581**	59.4	31.4	9.2
Uncalibrated	0.375	2.154	51.919	59.5	31.5	9.0
No Depth	1.262	3.808	74.650	30.2	15.6	54.2

Depth-Guided



No Depth (simple geometric pixel calculation)

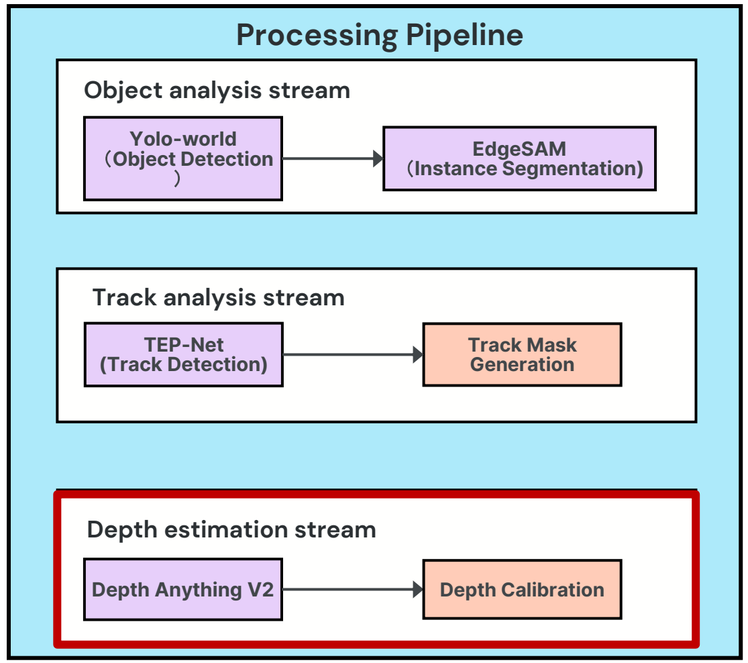


Parallel vs. Sequential Processing: Efficiency Trade-offs

Performance Comparison of Processing Architectures

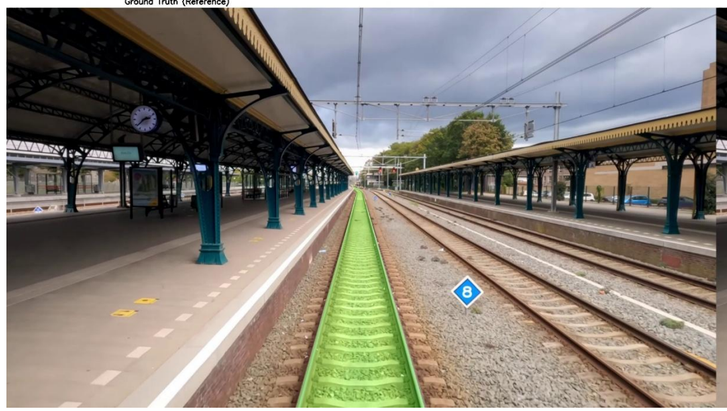
Architecture	Time per Image (s)	Depth part(s)
Parallel	**0.399** 2.51FPS	0.304
Sequential	0.449 2.23FPS	**0.266**

- Parallel processing speeds up inference by 11.2% (2.51FPS vs 2.23FPS)
- Depth estimation slower in parallel mode (0.304s vs. 0.266s).
- Possible causes: GPU contention & thread scheduling overhead.

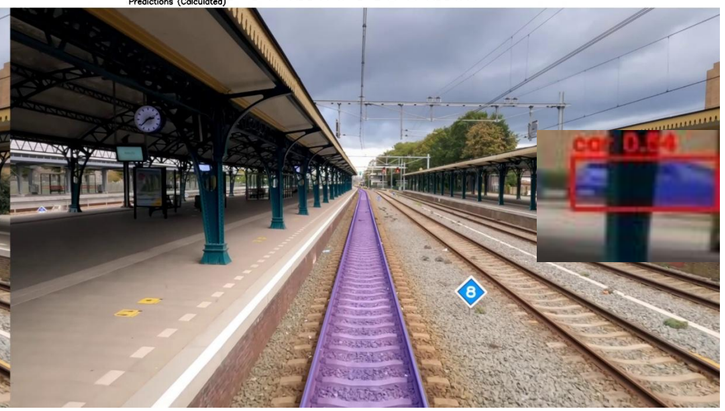


Performance in Vehicle Detection

Ground Truth



Predict Results



High Recall (0.70): Most vehicles are successfully detected.

Low Precision **(0.017):** Many **false positives appear** in **non-railway areas**.

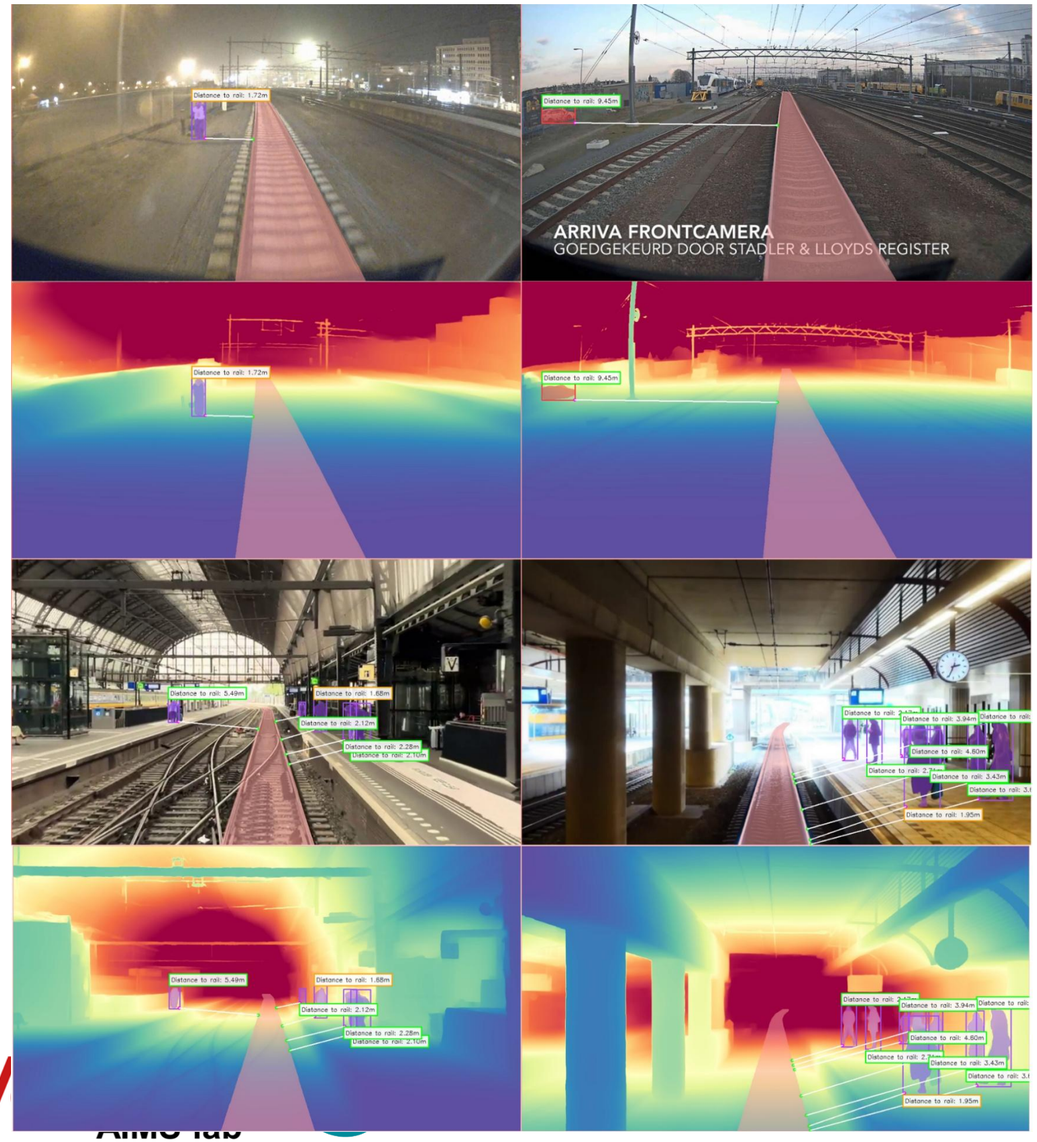
Why:

- Our Dataset annotation **focuses only on safety-critical areas near railway tracks.**
- The model detects a wider range of vehicles, including cars outside the station, leading to over-detections.





System Performance Analysis



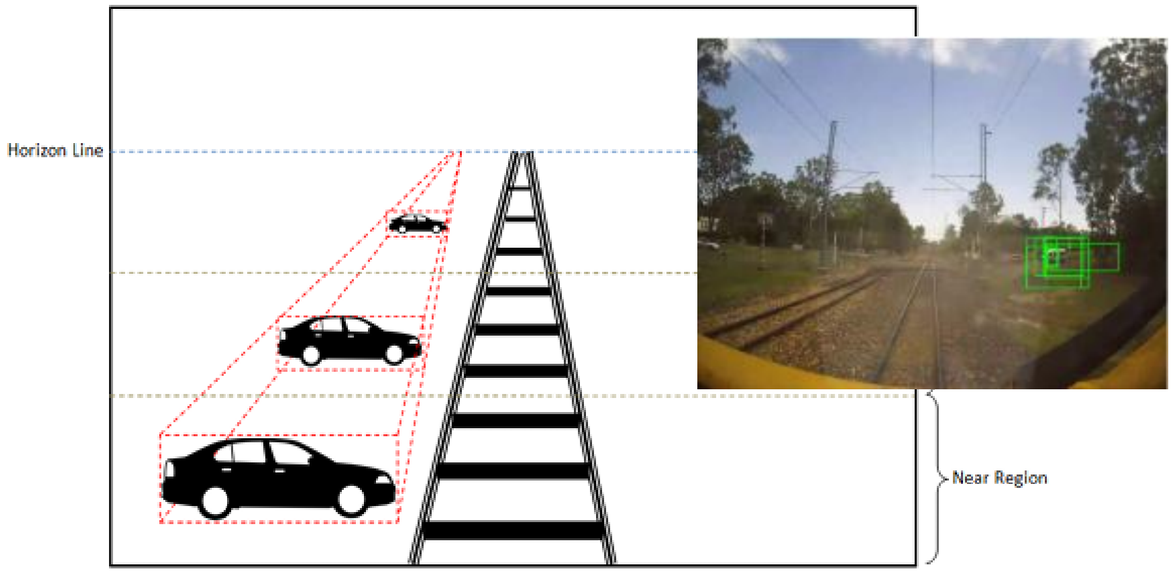
The depth-guided method, combined with the railway gauge (1.435m), reduces lateral distance measurement error from **1.262m to 0.290m (compare with geometric)**.

- Depth calibration method improves stability:
  - Standard deviation: **2.154m → 0.195m.**
  - Maximum error: **51.919m → 2.581m.**

The system is robust across complex scenarios, including crowded environments, low-light conditions, and partial occlusions.

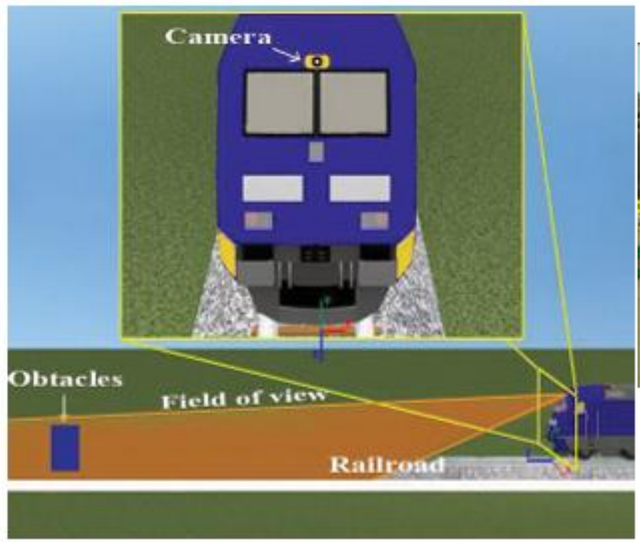
Prior Methods for Near-Miss Detection

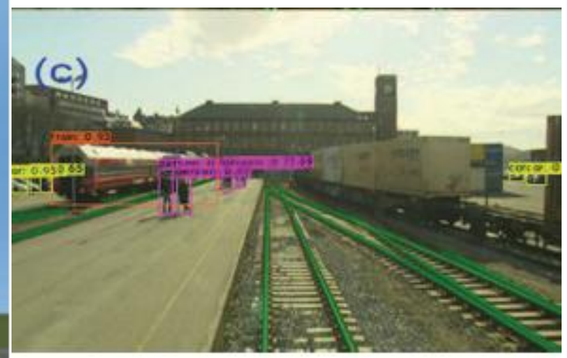
Level Crossing Focus (2015)



- Vehicle-only detection
- GPS and trains operation data integration
- Paper-based reports digitization

Vision-based Approach(2021)





- Track localization & object detection
- Pixel-based risk assessment

Both methods failed to achieve precise lateral distance calculations.

Contribution

Lateral Distance Measurement

Proposed a method that combines depth-guidance matching point with track width calibration for precise measurements.

Depth Calibration

Proposed a novel two-stage calibration to convert relative depth estimates into real-world depth measurement

New Dataset

Created a railway station dataset with accurate lateral distance annotations including annotations for tracks, persons, and vehicles