# How Far is too Far? Fixing Vision of Autonomous Vehicles using Selective Super-Resolution

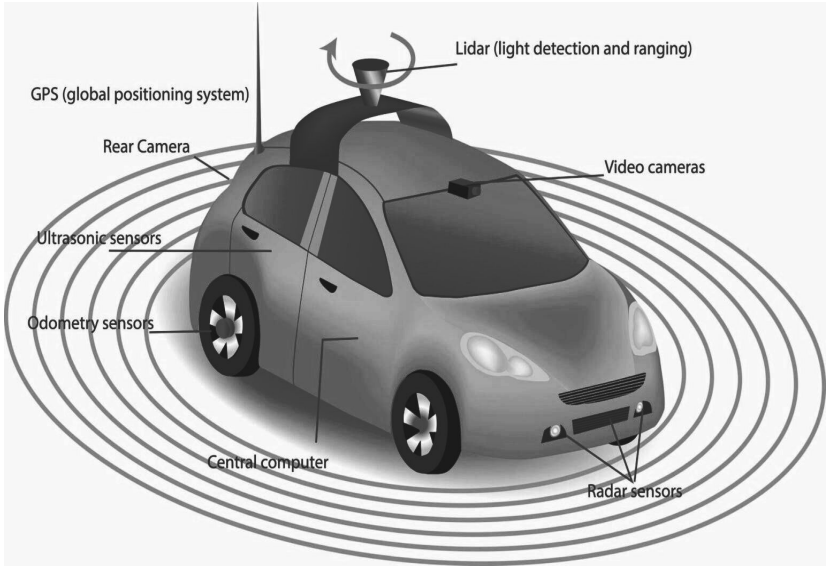
Workshop on Machine Intelligence in Networked Data and Systems (MINDS)
COMSNETS 2026

**Najiya Naj**, Ritik Vaishnav, Arani Bhattacharya


INDRAPRASTHA INSTITUTE of
INFORMATION TECHNOLOGY DELHI

# Autonomous Vehicles (AVs)

A vehicle that can drive itself without human intervention.



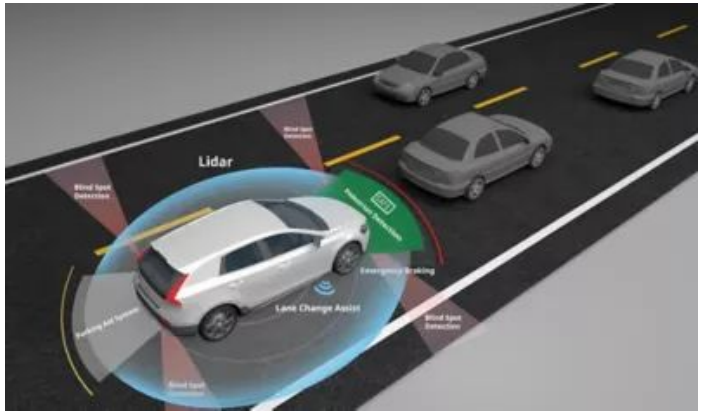

Autonomous Driving (AV) improves safety and efficiency

Supports better utilization of road infrastructure.

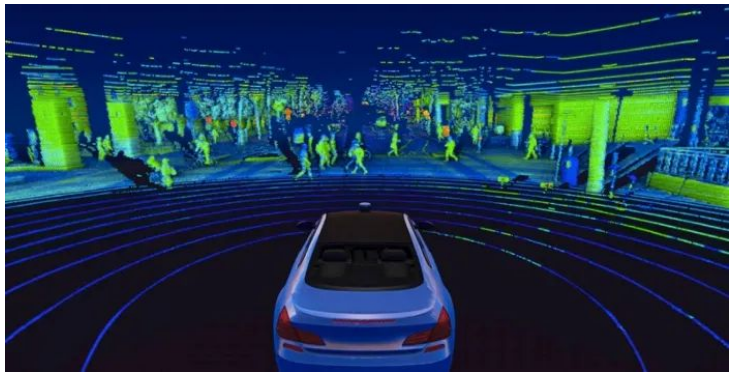
https://fifthlevelconsulting.com/key-components-of-autonomous-vehicles/

# Autonomous Vehicle with LiDAR Sensors

Lidar sensor captures 3D point cloud data




Point cloud data provide information, which can be used for 3D object detection. But these 3D models are computationally heavy.



# The Core Perception Challenge

Reliable perception is critical for safe autonomous driving

– Vehicles must detect obstacles early

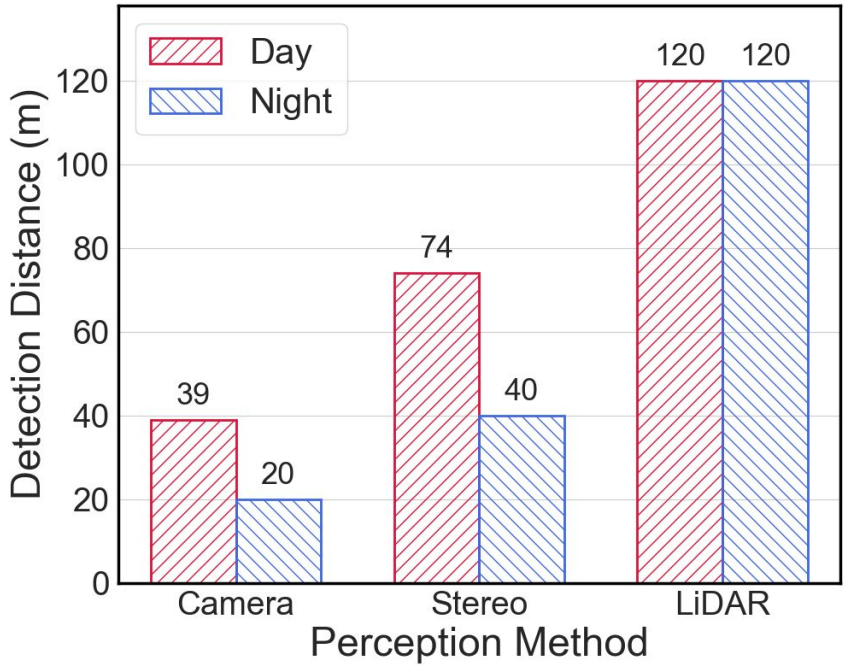
– Detection range directly impacts reaction time

– Errors at long distances are safety-critical [1,2]

[1] Caesar, Holger, et al. Proceedings of the IEEE/CVF CVPR 2020.
[2] Janai et al., 2020, Foundations and trends® in computer graphics and vision.

# Why Far-Object Perception is a Safety Bottleneck

– AVs struggle with far-object detection (>40 m)
 – Critical in high-speed roads and large intersections
 – Early detection enables safer braking and maneuvering



# Limitations of LiDAR-Based Perception

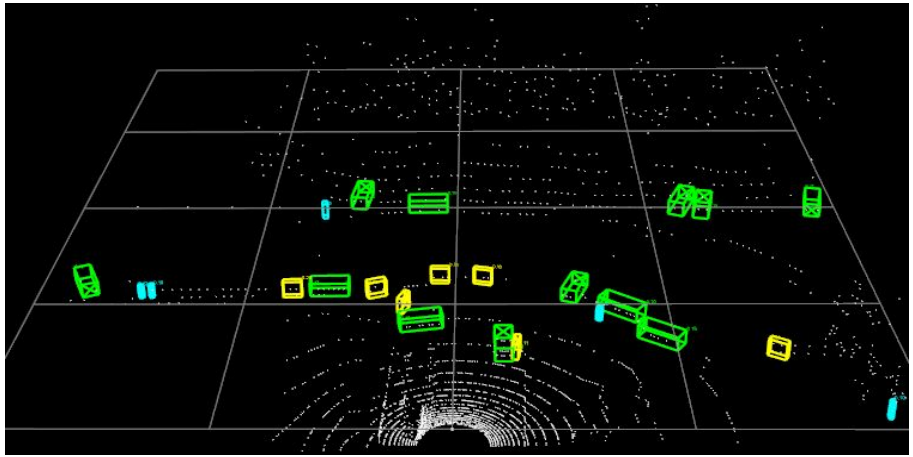
– High sensor and maintenance cost

– Sparse point clouds at long range reduce object detail

– Difficult to infer object shape and semantics, complicating downstream prediction and planning

– Performance degrades in rain, fog, and adverse weather

– Limits large-scale deployment



# Why Stereo Vision? A Cost-Effective Alternative

– Uses a pair of cameras to estimate depth via image disparity

– Passive sensing without active emission

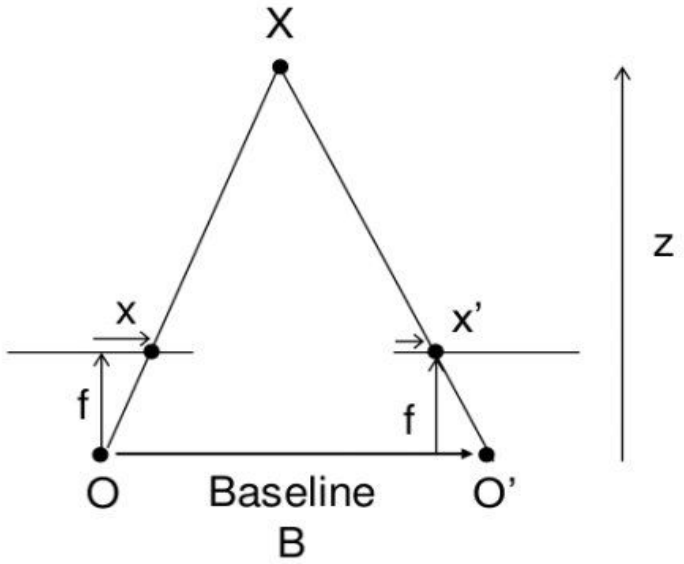
– Cheaper and more scalable than LiDAR

– Achieves near-LiDAR accuracy in favorable conditions

# Stereo Depth Estimation

Depth from Stereo Geometry:

$$z = (f \times B) / \text{disparity},$$




Where X is object and,

– z: object distance

– f: camera focal length

– B: stereo baseline

– disparity: pixel shift between left and right images (disparity =x-x')

# Limitation of Stereo for Far Objects

– Far objects occupy very few pixels

– Small disparity reduces depth precision

– Detection accuracy drops after long range

– Limits reliability at high speed and open intersections

# Super-Resolution (SR) for Far-Object Detection

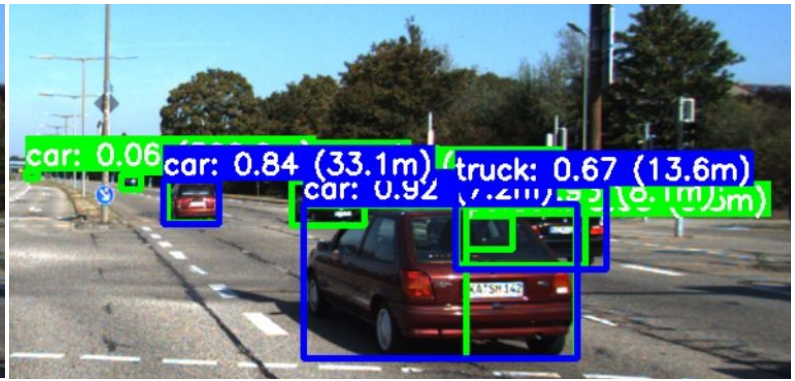
– SR enhances fine details and textures in low-resolution regions

– Improves far-object detection by up to 30–35% beyond 100 m

– Boosts detection confidence for small and distant objects

– Effectively addresses stereo's limited pixel disparity at long range


(a)




(b)

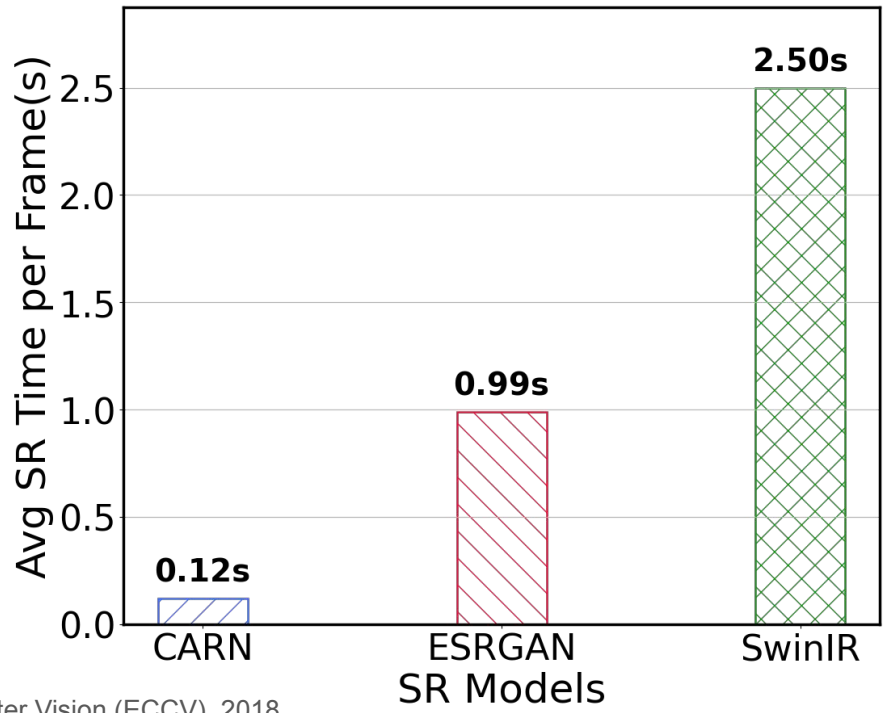
# Challenge of Full-Frame Super-Resolution

– Full-frame SR is computationally expensive

– SR latency can exceed object detection time

– Wastes computation on irrelevant regions (sky, road, nearby objects)

– Impractical for real-time AV pipelines



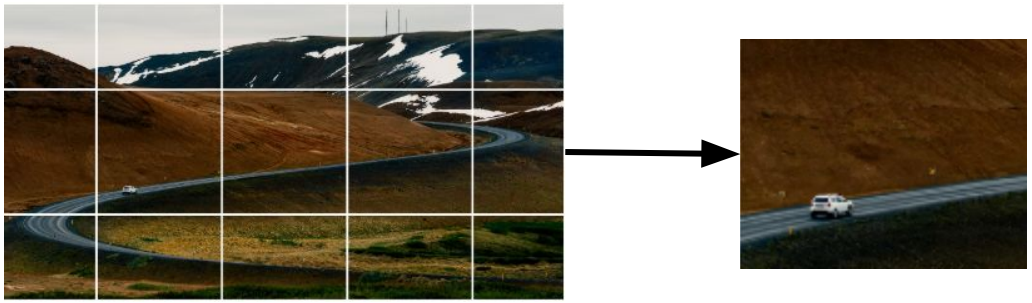

[3] N. Ahn, et al. CARN, In Proceedings of the European Conference on Computer Vision (ECCV), 2018.
[4] X. Wang, et al. ESRGAN, In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, 2018.
[5] J. Liang, et al. SwinIR, In Proceedings of the IEEE/CVF ICCVW, 2021

# Tiling-Based Super-Resolution

– Apply SR only on selected image regions (tiles)

– Two approaches:

- Static Tiling: fixed grid



**Static Tiling**

# Tiling-Based Super-Resolution

– Apply SR only on selected image regions (tiles)

– Two approaches:

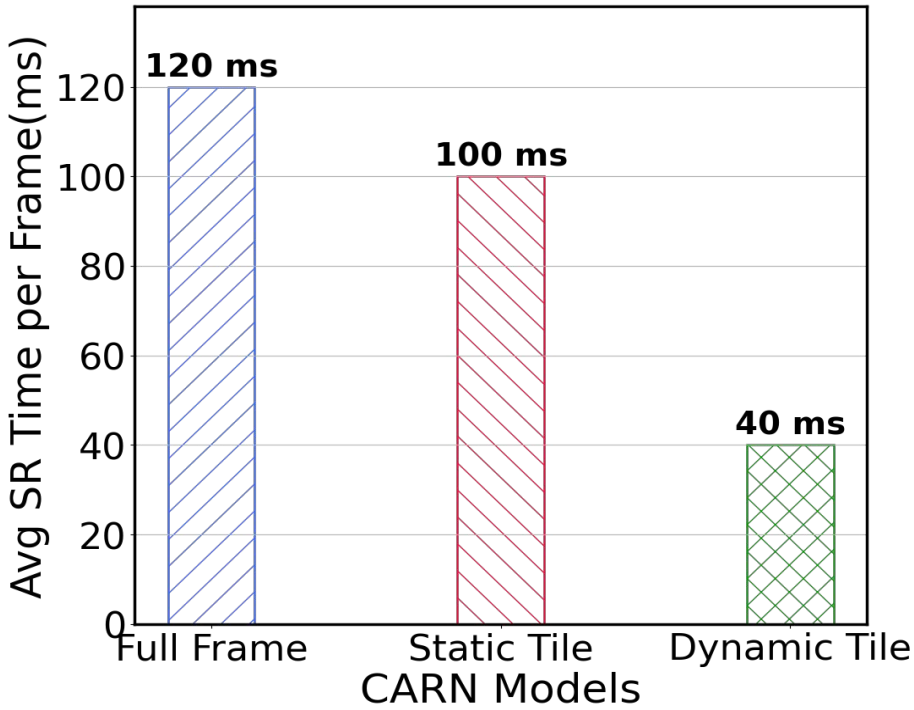
- Static Tiling: fixed grid
- Dynamic Tiling: scene aware adaptive tiles

**Dynamic Tiling**

# Why Dynamic Tiling?

**Dynamic Tiling for Efficient SR**

– Focuses SR on far-field, object-rich regions
– Avoids processing empty or near-field areas
– Reduces SR computation significantly
– Enables real-time far-object enhancement



# When Is Super-Resolution Worth It?

**SR is useful only for specific regions and scenes**

– Far-field regions benefit most from SR
– Dense or highly curved scenes reduce SR usefulness
– SR must be applied *selectively and adaptively*

# Our Solution: Stereo-Guided Selective SR

– Use stereo depth to identify far-field regions

– Apply SR selectively, instead of full-frame processing

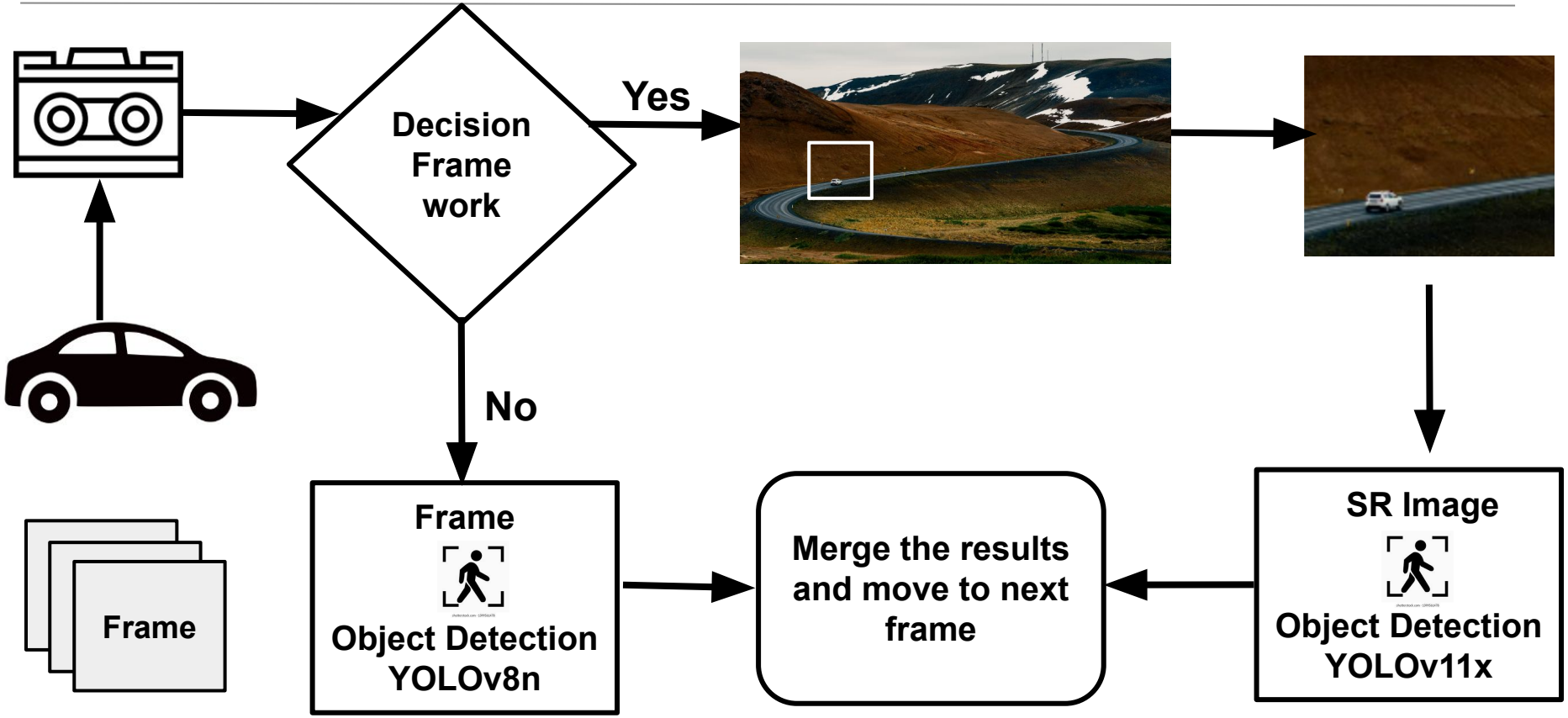
– Dual-branch detection architecture

  - YOLOv8n for near & mid-range (low latency)
  - SR-enhanced YOLOv11x for far-field regions

– Adaptive SR activation based on scene context

– SR is useful only for specific far-field regions and specific scenes

# Architecture: Stereo-Guided Selective SR



# Our Solution: Stereo-Guided Selective SR

– **Adaptive SR decision framework** determines when SR is beneficial using:

- Road curvature
- Object density
- Visibility & temporal stability

– SR activated only when the **decision score (S)** exceeds a threshold

# Our Solution: Stereo-Guided Selective SR

$$S=0.35(1-S_k)+0.35(1-S_\rho)+0.20S_v+0.10S_h$$

where $S_\kappa$, $S_\rho$, and $S_v$ denote normalized curvature, density, and visibility scores, respectively, and $S_h$ captures temporal stability to prevent rapid oscillations.

# Implementation

– Fully onboard real-time system

– NVIDIA RTX 3060 + Intel i5-12600K

– Integrated with CARLA simulator

– Evaluated on KITTI dataset

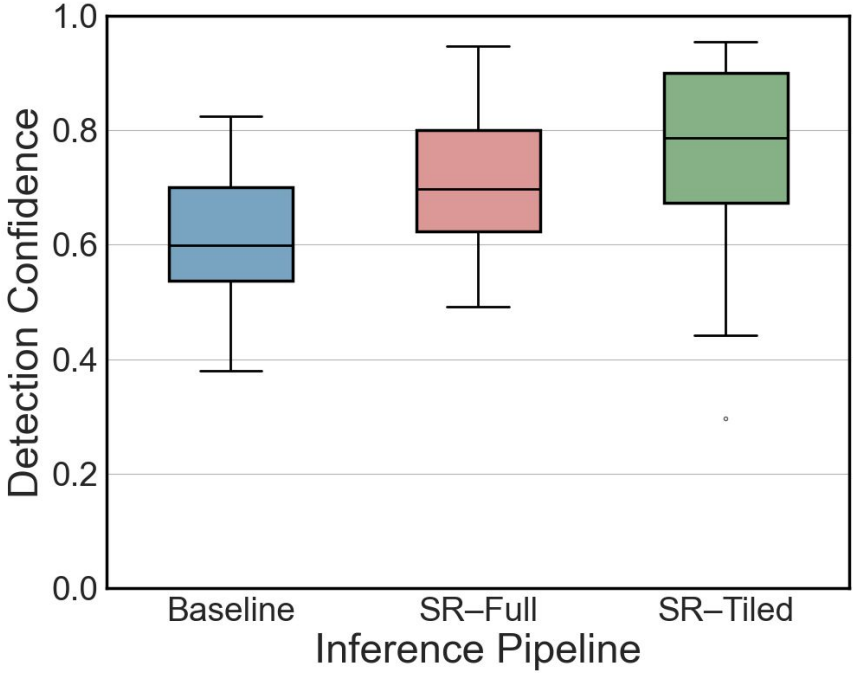
# Evaluation Setup

– Datasets: CARLA simulator and KITTI

– Compared against YOLOv8n, YOLOv11x and SAHI-YOLO

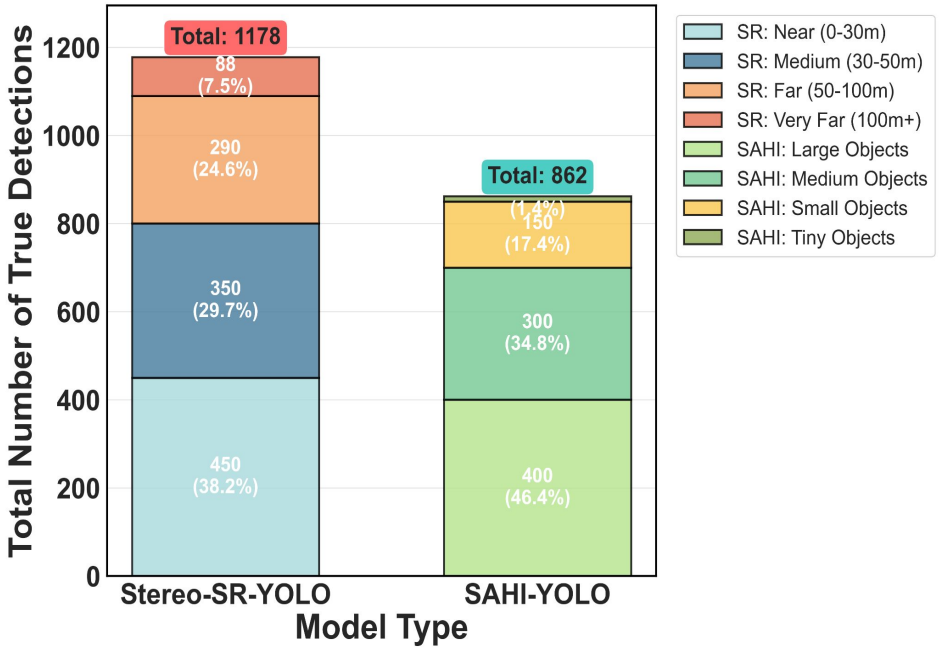
– Metrics:  Far-object detection confidence score, true detection count, latency

# Results: Confidence Improvement



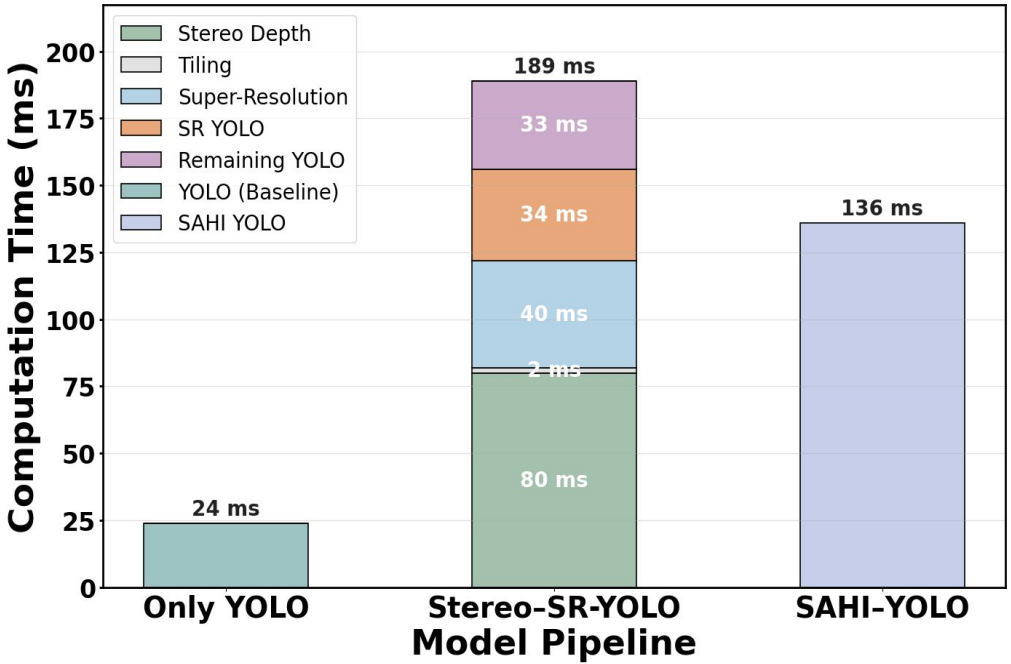

**SR-enhanced stereo increases detection confidence for small and distant objects.**

# Results: Far-Object Detection



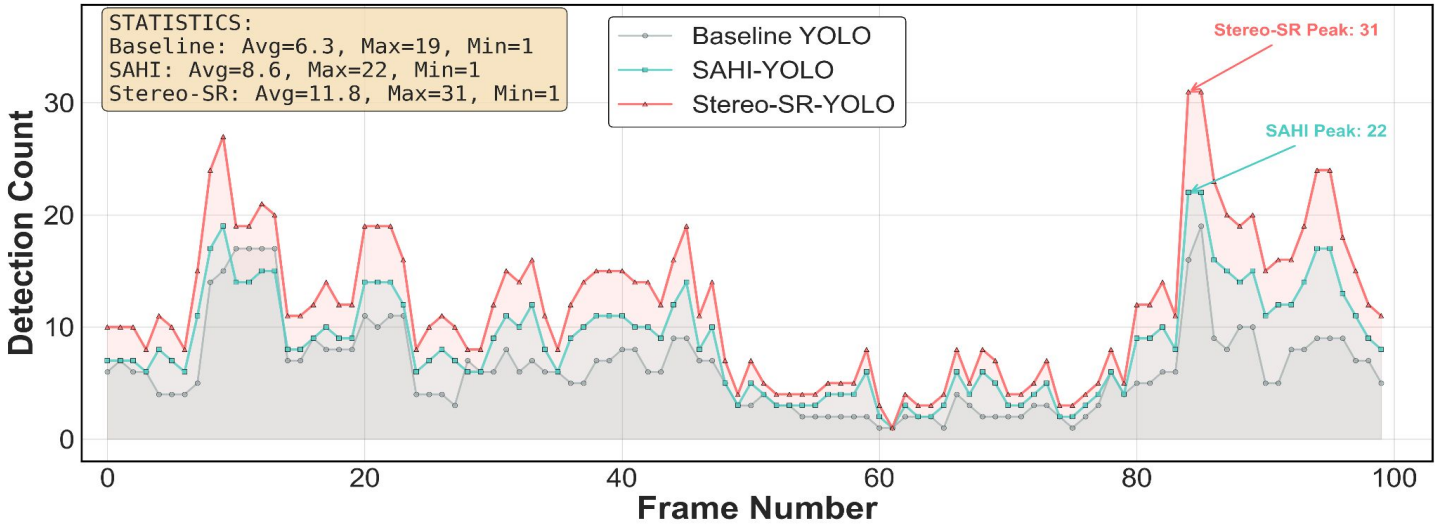

**Stereo-SR-YOLO improves far-object detection in both medium (50–100 m) and long-range (>100 m) scenarios compared to SAHI-YOLO.**

# Results: Computational Efficiency




**Stereo-SR-YOLO achieves significantly lower end-to-end latency avoiding unnecessary SR through depth-aware tiling.**

# Results: Frame-Level Detection




**Stereo-SR-YOLO consistently detects more far-field objects per frame than YOLOv11x and SAHI-YOLO, demonstrating stable frame-level performance.**

# Key Takeaways

– Far-object perception is a critical AV challenge
– Selective SR improves far-object detection by >35%
– Stereo-guided dynamic tiling enables real-time deployment
– Improved far-object perception enables safer driving at higher speeds