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

Najiya Naj*, Ritik Vaishnav †, Arani Bhattacharya*
{najiyan, arani}@iiitd.ac.in, ritikvaishnav20@gmail.com
* IIIT-Delhi, † MBM University, India

***Abstract*—Autonomous vehicles (AVs) promise to transform mobility, yet their deployment remains constrained by reliability, scalability, and cost barriers. Current AV perception stacks find it challenging to detect distant objects, as computer vision models are powerful enough to detect them. While LIDARs can detect such objects, they are less reliable for distant objects and are expensive. These factors hinder widespread adoption, particularly in consumer-grade vehicles where scalability and robustness are critical.**

In this work, we advocate for *stereo vision* as a scalable and cost-effective alternative for long-range perception and propose a *stereo-guided selective super-resolution (SR) tiling framework* to address the challenge of far-object detection. Our system leverages stereo depth to identify distant regions of interest and selectively applies SR enhancement before detection, improving far-field visibility while minimizing computation and maintaining high accuracy and real-time efficiency. The proposed framework is extensively evaluated on the CARLA simulator and the KITTI benchmark under diverse environmental and traffic conditions. Results demonstrate that our selective SR approach improves far-object detection by 27% in the 50–100 m range and 35%beyond 100 m, while achieves with up to $3\times$ lower SR latency, paving the way for scalable and efficient AV perception.

***Index Terms*—Autonomous Driving, Stereo Vision, Super-Resolution, Dynamic Tiling, Far-Object Detection**

I. Introduction

Autonomous vehicles (AVs) have seen remarkable progress in recent years, particularly in perception, planning, and control. Yet, their widespread adoption remains limited by safety-critical failures encountered in real-world driving conditions. High-profile incidents demonstrate that even AVs equipped with LiDAR can misidentify or fail to respond appropriately to vulnerable road users in complex scenarios [4], [18], [20]. These shortcomings underscore the need for perception systems that are not only accurate but also *scalable, cost-effective, and dependable across diverse environments*.

LiDAR has traditionally been the backbone of AV perception, providing precise long-range depth measurements and object localization [5]. While its accuracy is unmatched, LiDAR comes with significant drawbacks: the cost of a single unit, although reduced from approximately $75,000 in 2009 to $7,500 in 2017, still makes multi-sensor setups expensive, often totaling hundreds of thousands of dollars per vehicle [23]. Additionally, LiDAR consumes substantial power and exhibits performance degradation in adverse environmental conditions such as rain, fog, or snow [13]. Moreover, real-world accidents, including the 2018 Uber crash in Tempe, Arizona, illustrate that LiDAR detection alone cannot guarantee safe response when perception and planning modules fail to interpret the scene in time [4], [25]. These challenges have driven interest in camera-based perception as a viable and scalable alternative.

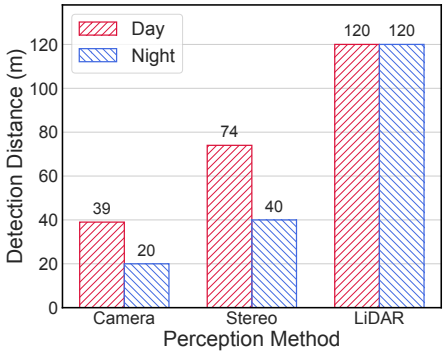


Fig. 1: Comparison of detection distance between camera and LiDAR systems. Stereo vision demonstrates competitive range under daylight conditions, while LiDAR remains superior in full range. The evaluation is performed using CARLA's simulation [6].

Camera systems, particularly stereo vision, have emerged as a compelling alternative for AV perception. Stereo cameras are inexpensive, lightweight, and energy-efficient, while providing dense depth estimation over wide fields of view. Deep-learning-based stereo perception methods such as Pseudo-LiDAR and Stereo R-CNN [15], [27] have significantly narrowed the performance gap with LiDAR, achieving impressive accuracy on benchmarks such as KITTI [8]. Industry prototypes have further demonstrated that stereo-based systems can reliably detect objects up to 130–200 m in favorable lighting [11], [21]. As shown in Fig. 1, stereo cameras achieve competitive performance within critical detection ranges at a fraction of LiDAR's cost, positioning them as an attractive foundation for scalable AV perception.

However, detecting far-away objects (beyond 40–50 m) remains challenging for stereo-based systems. At longer distances, parallax becomes minimal, reducing disparity accuracy and making small or distant objects occupy only a few pixels-too few for reliable classification. This problem is especially critical for safety in highway and urban intersection scenarios, where early recognition of distant

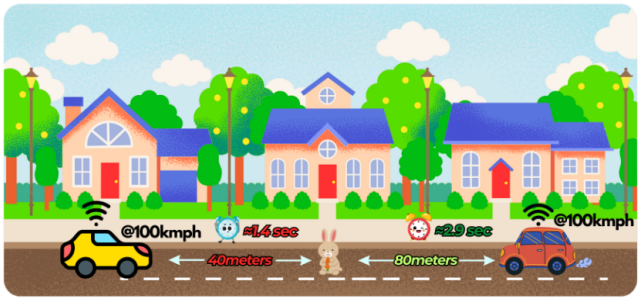


Fig. 2: Extending detection range improves reaction time. At $100\,\mathrm{km/h}$, increasing perception distance from 40 to $80\,\mathrm{m}$ adds $\approx 1.5\,\mathrm{s}$ of reaction time, highlighting the safety advantage of far-object perception.

obstacles directly impacts reaction time and decision-making. As illustrated in Fig. 2, at $100\,\mathrm{km/h}$, detecting an obstacle at $40\,\mathrm{m}$ provides roughly $1.4\,\mathrm{s}$ for reaction, whereas detecting the same obstacle at $80\,\mathrm{m}$ offers about $2.9\,\mathrm{sec}$ an additional $1.5\,\mathrm{sec}$ that can be the difference between collision and avoidance. Thus, extending perception reliability for far objects is a fundamental safety requirement.

A promising way to improve distant object visibility is through *image super-resolution (SR)*, an ML-based interpolation technique that reconstructs high-frequency image details such as edges and textures from low-resolution inputs. SR has shown potential in restoring fine visual cues for small or distant objects [17], [16]. However, applying SR to entire high-resolution frames is computationally expensive, often costlier than object detection itself, and transmitting such frames to the cloud is impractical in bandwidth-limited settings [26].

To overcome these challenges, we develop a *stereo-guided selective SR tiling framework* that selectively applies SR based on scene depth and driving context, balancing far-object accuracy with real-time efficiency. We validate our framework on the CARLA Simulator[6] and KITTI [8] datasets under varied driving and environmental conditions, demonstrating significant improvements in far-object detection accuracy and real-time efficiency compared to baseline detection pipelines.

Our primary contributions are as follows:

- A stereo-guided adaptive decision framework that learns when to activate SR based on curvature, density, and visibility cues.
- A dual-branch local perception pipeline combining YOLOv8n for near/mid-range detection and SR-enhanced YOLOv11x for distant regions.
- A depth-aware selective tiling strategy that applies SR only to far-field regions identified via stereo disparity, improving detection accuracy with minimal overhead.
- Comprehensive evaluation on CARLA and KITTI, demonstrating consistent gains in far-object detection and real-time efficiency.

II. Background and Motivation

In this section, we discuss the role of stereo vision in enabling scalable autonomous vehicle (AV) perception and the challenges it faces in detecting far-away objects. We further explore how image SR can enhance distant object recognition and motivate the need for a stereo-guided, adaptive SR approach that improves perception accuracy and efficiency within an onboard AV pipeline.

Stereo Vision for Scalable Perception. Stereo vision has emerged as a compelling perception modality for autonomous vehicles due to its low cost, compact form factor, and ability to estimate depth from passive sensing. By computing disparity between left–right image pairs, stereo cameras provide dense 3D structure of the scene without expensive LiDAR hardware. While recent learning-based stereo methods [15], [22], [27] have achieved near-LiDAR accuracy in daylight conditions, their performance degrades sharply for distant or small objects due to limited pixel disparity. This limitation becomes critical in scenarios such as high-speed driving or large intersections, where early detection of far-field obstacles is essential for safe decision-making.

Why Far-Object Perception Matters. Far-object detection is a cornerstone of proactive and safe AV behavior. Accurate early perception of distant obstacles allows AVs to perform smoother braking, early lane changes, and proactive trajectory adjustments. Most perception pipelines, however, are tuned for mid-range accuracy, as datasets like KITTI [8] primarily emphasize object distances below $40\,\mathrm{m}$ (Fig. 1). This results in models that perform well in typical urban scenes but degrade significantly in open-road or high-speed highway conditions. To address this limitation, prior studies have explored multi-sensor fusion [3], hybrid stereo–LiDAR systems [28], and monocular depth learning with geometric priors [9]. While these methods improve range perception, they often introduce additional cost, latency, or calibration overhead factors that limit their scalability in real-world deployments.

Super-Resolution for Far-Object Enhancement. A promising alternative to sensor fusion is enhancing perception at the image level using SR. SR is a machine learning-based interpolation technique that reconstructs high-frequency details such as edges and textures from low-resolution images, effectively increasing perceptual clarity and aiding small-object detection. In the context of AV perception, SR can compensate for the loss of spatial detail in distant or low-quality frames. Models such as SRODNet [17], SwinIR [16], and ESRGAN [24] have shown significant gains in image fidelity and downstream detection performance. However, these methods come with substantial computational overhead, making them difficult to integrate into real-time AV pipelines. As shown in Fig. 3(a), the per-frame computation time of transformer-based models like SwinIR or GAN-based models like ESRGAN is prohibitively high, whereas lightweight architectures such as CARN [1] achieve real-time performance with minimal degradation in visual quality.

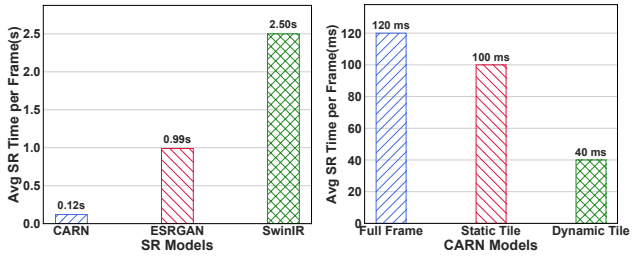


Fig. 3: SR model performance: (a) per-frame computation times for different models, (b) Computation time for CARN-based SR under full-frame, static, and dynamic tiling modes.

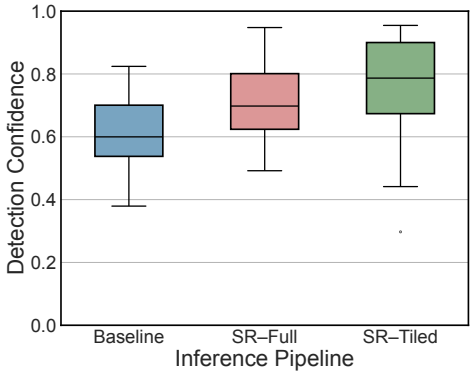


Fig. 4: Average detection confidence for far objects comparing full-frame SR, static tiling, and dynamic tiling. Dynamic SR tiling achieves higher detection confidence while minimizing computation.

Challenges of Full-Frame SR in AV Pipelines. Processing entire frames for SR wastes compute on irrelevant regions (e.g., sky or road texture), while only a small fraction of the frame contains distant, safety-critical objects. Transmitting full-resolution frames for SR enhancement further inflates bandwidth usage [26], [12]. Adaptive or region-based SR strategies have been proposed to focus processing on informative areas [14]. However, most existing methods rely on heuristic region selection or static partitioning, which may miss contextually important regions in dynamic driving scenes. This motivates a stereo-guided, depth-aware SR approach that selectively enhances distant regions in real time.

Our proposed system adopts a dual-branch design: a lightweight *YOLOv8n* detector continuously handles near and mid-range perception, while a parallel *SR-enhanced YOLOv11x* branch processes far-field tiles within the same frame. Computation is dynamically distributed based on stereo-derived depth and scene context, allowing SR to operate only where it yields perceptual gains. We explore both static and dynamic tiling strategies where *static tiling* uses fixed-size patches for predictable computation, and *dynamic tiling* adjusts tile boundaries according to disparity and object density, allocating higher resolution to distant, object-rich regions. As shown in Fig. 3(b), dynamic tiling reduces CARN's SR computation time by roughly $3\times$ compared to full-frame enhancement while improving detection confidence for far objects (Fig. 4). By combining stereo-guided depth cues, selective SR, and adaptive decision control, our framework achieves high far-object detection accuracy with real-time efficiency without any cloud dependence or external offloading.

III. Our Solution: Adaptive Use of Super-resolution

Building on the motivation from Section II, we design a *stereo-guided selective SR tiling framework* that dynamically balances accuracy and efficiency in real time. The system leverages stereo depth to identify distant regions and applies SR selectively before detection, rather than processing the entire frame. A dual-branch design leverages *YOLOv8n* for near and mid-range detection due to its low computational cost, while an *SR-enhanced YOLOv11x* branch focuses on far-field regions to ensure high detection accuracy. An adaptive decision module governs SR activation based on scene-level cues such as road curvature, object density, and visibility, ensuring that SR is used only when beneficial. This fully onboard, dual-branch architecture enhances far-object visibility while maintaining real-time performance and efficient resource utilization.

A. System Overview

The proposed pipeline (shown in Fig. 5) operates onboard and comprises four main components: *(1) stereo-based depth estimation*, *(2) scene characterization*, *(3) adaptive SR decision framework*, and *(4) selective Tiling and Super-Resolution*. Depth and scene metrics are computed in real time to determine when and where SR should be applied. The lightweight YOLOv8n model ensures low-latency near-field perception, while the SR-enhanced YOLOv11x module processes selected far-field tiles in parallel to improve long-range accuracy.

1) Stereo Depth Estimation: Depth is computed from left–right stereo images using a refined Semi-Global Block Matching (StereoSGBM) model [10]. A median filter is applied to suppress low-texture noise. Based on the pinhole camera geometry,

$$D = \frac{f \times b}{\text{disparity}}, \tag{1}$$

where f is the focal length and b is the stereo baseline. The resulting dense depth map allows identification of far-field regions ($D > 40\,\text{m}$), which guide subsequent SR processing.

2) Scene Characterization: Each frame is analyzed to compute contextual metrics such as road curvature [19], object density, and visibility. These parameters reflect the driving context and are used to infer whether SR would contribute meaningful improvement. For example, during high curvature (i.e., on sharp bends or intersections), most objects lie at short or medium range, and the visible far-field area is small or partially occluded. Similarly, in high-density

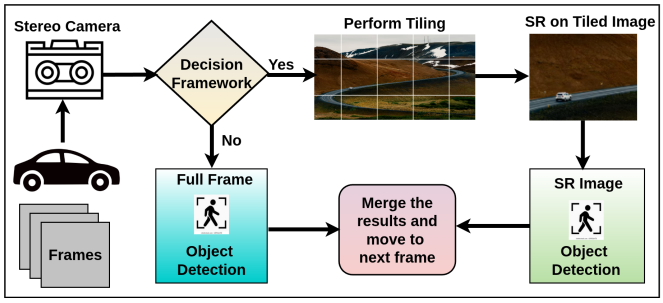


Fig. 5: Proposed stereo-guided pipeline with dynamic tiling and selective super-resolution. Far-field regions identified via stereo depth are enhanced using SR before object detection to improve accuracy.

traffic scenes, vehicles and pedestrians dominate the frame, making SR less useful since distant regions are visually cluttered and contribute little to decision-making. Conversely, under open-road conditions, with low curvature, low density, and good visibility, SR can meaningfully enhance distant texture details, such as lane boundaries, traffic signs, or small, far-away vehicles that are otherwise difficult to detect.

3) Adaptive SR Decision Framework: To activate SR only when beneficial, we introduce an *adaptive scoring-based learner* that quantifies scene complexity per frame. The decision score is defined as:

$$S = 0.35(1 - S_\kappa) + 0.35(1 - S_\rho) + 0.20S_v + 0.10S_h, \quad (2)$$

where S_κ, S_ρ, and S_v denote normalized curvature, density, and visibility scores, respectively, and S_h captures temporal stability to prevent rapid oscillations. Note that curvature and number of objects are negatively correlated with the need for super-resolution, and thus they appear in the above equation as negative terms. SR activation is triggered when S exceeds a learned threshold and distant objects (> 40 m) are detected in the depth map. This ensures SR is used adaptively, conserving compute during complex or occluded scenes.

4) Selective Tiling and Super-Resolution: When SR is activated, the far-field regions are segmented into tiles and enhanced using a lightweight Cascading Residual Network (CARN) [1]. Meanwhile, YOLOv8n continues detection on the native-resolution frame to maintain low-latency perception. This dual-branch setup allocates computation adaptively between regions of interest.

We employ two tiling modes: *Static tiling* divides each 640×360 frame into a uniform 2×2 grid of 320×180 patches for predictable compute cost, whereas *dynamic tiling* adjusts tile boundaries based on disparity and object distribution, typically cropping 50–100 pixel regions around far-field objects. This ensures SR computation remains both efficient and spatially focused

As shown in Fig. 3(b) and Fig. 4, dynamic tiling not only reduces SR computation cost by nearly $3\times$ compared to full-frame enhancement but also improves detection confidence for far objects while preserving visual fidelity.

IV. Implementation

We implement the proposed framework (Fig. 5) as a fully onboard system optimized for real-time operation. Our entire system is evaluated on a system with Nvidia GeForce RTX 3060 GPU and Intel Core i5-12600K with 10 cores. The complete pipeline, including stereo processing, adaptive SR, and detection, is co-located with the CARLA simulator and evaluated on the KITTI dataset. The data from the camera is encoded using H.265 encoder available with ffmpeg's x265 encoder [7].

A. Pipeline Implementation

All modules run on the local GPU without any external offloading. Stereo disparity is computed in real time, followed by lane curvature, traffic density, and visibility estimation. These scene metrics are fed to the adaptive SR decision module, which dynamically determines whether selective SR should be activated for the current frame.

When SR is active, only the identified far-field tiles are enhanced using the CARN model ($4\times$ upscaling) and processed by YOLOv11x, while YOLOv8n continues to handle near and mid-range detection on the original frame. Dual-branch Detections from both models are merged using non-maximum suppression (NMS) and depth-consistency filtering to form a unified high-confidence perception output. This joint design ensures balanced accuracy and efficiency for real-time onboard operation.

B. Adaptive Learning Process

The decision module operates through a three-phase adaptation loop:

1) **Initialization:** Conservative normalized thresholds are set,curvature $(1-S_\kappa = 0.3)$, density $(1-S_\rho = 0.3)$, visibility $(S_v = 0.2)$, and SR distance 40m.

2) **Learning:** Thresholds are refined continuously from percentile-based and variance-aware updates derived from recent driving frames.

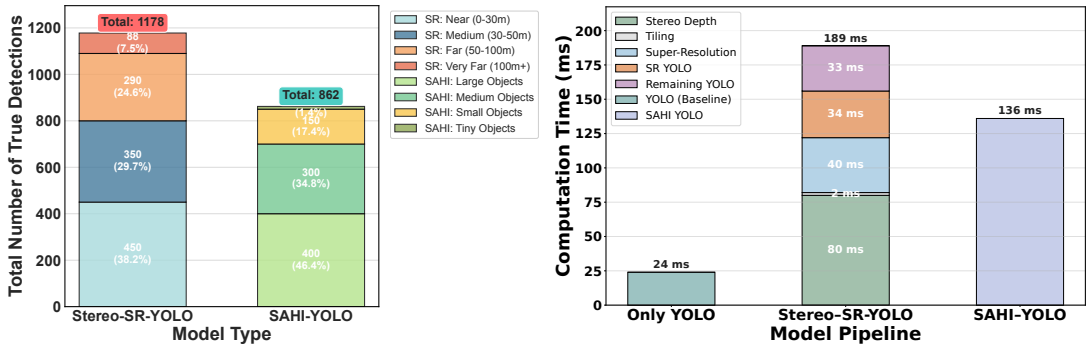
3) **Optimization:** Parameters are periodically tuned to maintain balanced SR activation (40–60%) while minimizing latency and computational cost.

Learned parameters are persistently stored, allowing the system to retain and adapt experience across multiple sessions.

C. Integration with Region-Based Detection

Our framework integrates seamlessly with region-based detectors such as SAHI [2]. SAHI performs *image slicing*, dividing the input into overlapping windows to improve detection of small objects that would otherwise vanish at full-frame resolution. It then applies *slice-level NMS* to merge detections across overlapping patches, enabling high small-object recall independent of input resolution.

Unlike uniform slicing, only SR-enhanced far-field tiles are processed, improving recall for small and distant objects while maintaining real-time performance. This modular integration demonstrates the flexibility of the proposed design and its compatibility with diverse perception backbones.



(a) (b)

Fig. 6: Comparison of Stereo-SR-YOLO and SAHI-YOLO (a) Detection counts by range (0–30 m, 30–50 m, 50–100 m, 100 m+), (b) Computation time per stage, showing how tiling and SR contribute to total latency.

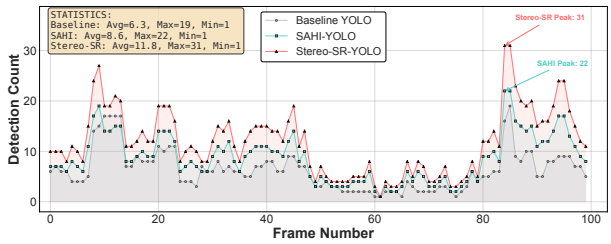


Fig. 7: Per-frame detection comparison: the proposed stereo–SR pipeline consistently achieves higher detection counts and smoother performance.

V. Evaluation

We evaluate the proposed *stereo-guided selective SR framework* on both the CARLA simulator and the KITTI dataset. Our evaluation focuses on three key aspects: (i) improvement in far-object detection accuracy, (ii) computational efficiency of selective SR and tiling, and (iii) stability of detection trends across frames.

We compare our method (Stereo-SR-YOLO) against two baselines: (1) camera-only YOLOv11x (no SR), (2) SAHI-YOLO slicing-based detector.

A. Experimental Setup

All experiments are conducted on an Nvidia GeForce RTX 3060 GPU (CARLA and KITTI) running locally. Each test uses 2,000 CARLA frames and 7,500 KITTI frames under varying visibility and traffic density conditions. The input frame resolution is 640×360, and full frame SR outputs are upscaled to 2560×1440. The stereo-guided decision module activates SR only when the decision score threshold and far-distance condition (> 40m) are satisfied.

B. Far-Object Detection on KITTI

As shown in Fig. 6a, the proposed Stereo-SR-YOLO pipeline achieves consistently higher far-object detection compared to SAHI-YOLO and baseline YOLOv11x. Specifically, it improves true detection counts by 27% in the medium range (50-100 m) and 35% beyond 100 m. This gain comes from selective SR applied to depth-identified distant regions, which enhances texture and edge cues critical for recognition. Overall, Stereo-SR-YOLO demonstrates superior far-object perception across both KITTI and CARLA scenarios.

C. Computation Efficiency

Fig. 6b shows the breakdown of per-frame computation time for different pipelines. The proposed Stereo-SR-YOLO achieves an average total latency of 80 ms per frame, nearly 40% faster than SAHI-YOLO (136 ms). This improvement comes from depth-aware selective tiling, which eliminates unnecessary SR processing. Within each frame, stereo depth estimation, dynamic tiling, and SR modules collectively consume less than 50% of total inference time.

D. Frame-Level Detection Trends

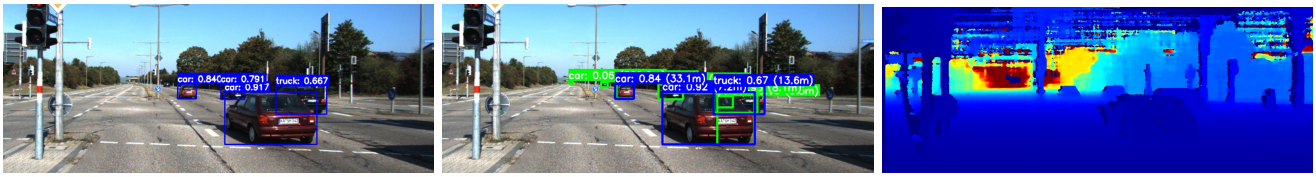
Fig. 7 presents detection trends across 100 KITTI frames. The Stereo-SR-YOLO pipeline maintains stable improvements, consistently detecting 4-6 additional far objects per frame compared to baseline YOLOv11x and SAHI-YOLO. The average detection counts are: YOLOv11x = 6.3, SAHI-YOLO = 8.6, and Stereo-SR-YOLO = 11.8. This stability highlights the reliability of the adaptive SR decision module and its ability to dynamically adjust to changing scene complexity

E. Qualitative Results

Fig. 8 illustrates qualitative improvements. Baseline YOLOv11x Fig 8(a) fails to detect distant obstacles. After stereo-guided SR and tiling Fig 8(b), new detections appear (green boxes). The depth map Fig 8(c) shows the far-field regions (>40 m) that trigger SR enhancement. This highlights that our selective SR approach improves visibility and detection of far-field targets without full-frame processing.

VI. Discussions and Conclusion

In this paper, we show that while detecting distant objects using single-shot computer vision techniques is challenging, these challenges can be mitigated using super-resolution.



(a) (b) (c)

Fig. 8: Qualitative improvements with stereo-guided SR: (a) baseline YOLOv11x detections, (b) detections after SR tiling, (c) stereo depth map showing far-field regions (>40 m).

However, using such super-resolution is compute-intensive in nature, making them challenging to use in autonomous vehicles. We show that by using dynamic tiles, and applying only the tiles that contain distant objects, it is possible to significantly mitigate the problem of high compute latency. We integrate this into our pipeline where the necessary tiles are identified in run-time, integrate it into an autonomous vehicle pipeline and evaluate it on KITTI dataset. We show that accuracy of far-object detection increases by over 35% compared to existing state-of-the-art SAHI-YOLO, while maintaining near real-time operation. Our study shows that use of super-resolution in autonomous vehicles is feasible. We believe that such improved detection of far objects would improve safety and support higher vehicle speeds.

ACKNOWLEDGEMENT

This work was supported by All India Council of Technical Education (AICTE) Doctoral Fellowship, awarded to Najiya Naj (Student ID: S2022405), to do research under the umbrella of smart cities. We also acknowledge the support of DST-PURSE Grant (numbered SR/PURSE/2023/172).

REFERENCES

[1] N. Ahn, B. Kang, and K.-A. Sohn. Fast, accurate, and lightweight super-resolution with cascading residual network. In *Proceedings of ECCV*, pages 252–268, 2018.

[2] F. C. Akyon, S. O. Altinuc, and A. Temizel. Slicing aided hyper inference and fine-tuning for small object detection. In *2022 IEEE international conference on image processing (ICIP)*, pages 966–970. IEEE, 2022.

[3] M. Bijelic, F. Mannan, T. Gruber, W. Ritter, K. Dietmayer, and F. Heide. Seeing through fog without seeing fog: Deep sensor fusion in the absence of labeled training data. *arXiv preprint arXiv:1902.08913*, 1030, 2019.

[4] N. T. S. Board. Collision between vehicle controlled by developmental automated driving system and pedestrian, tempe, arizona, march 18, 2018, 2019. Reissued June 2020.

[5] C. Crash. Enabling pedestrian safety using computer vision techniques: A case study of the 2018 uber inc. self-driving. In *Advances in Information and Communication: Proceedings of the 2019 Future of Information and Communication Conference (FICC)*, volume 1, page 261, 2019.

[6] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun. Carla: An open urban driving simulator. In *Conference on robot learning*, pages 1–16, Mountain View, US, 2017. PMLR.

[7] FFmpeg. https://www.ffmpeg.org/. [online; accessed Sept-2021].

[8] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE CVPR*, pages 3354–3361. IEEE, 2012.

[9] C. Godard, O. Mac Aodha, M. Firman, and G. J. Brostow. Digging into self-supervised monocular depth estimation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3828–3838, 2019.

[10] H. Hirschmuller. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on pattern analysis and machine intelligence*, 30(2):328–341, 2007.

[11] S. R. Institute. Swri develops road autonomous driving tools focused on camera vision. https://www.swri.org/newsroom/press-releases/swri-develops-road-autonomous-driving-tools-focused-camera-vision, 2024. Accessed 2025-09-23.

[12] M. Khani, V. Sivaraman, and M. Alizadeh. Efficient video compression via content-adaptive super-resolution. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4521–4530, 2021.

[13] P. Koopman. Lessons from the cruise robotaxi pedestrian dragging mishap. *IEEE Reliability Magazine*, 1(3):54–61, 2024.

[14] S. Lee, M. Choi, and K. M. Lee. Dynavsr: Dynamic adaptive blind video super-resolution. In *Proceedings of IEEE/CVF WACV*, pages 2093–2102, 2021.

[15] P. Li, X. Chen, and S. Shen. Stereo r-cnn based 3d object detection for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7644–7652, 2019.

[16] J. Liang, J. Cao, G. Sun, K. Zhang, L. V. Gool, and R. Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2021.

[17] Y. R. Musunuri, O.-S. Kwon, and S.-Y. Kung. Srodnet: Object detection network based on super resolution for autonomous vehicles. *Remote Sensing*, 14(24):6270, 2022.

[18] NHTSA/Reuters. U.s. closes probe into gm cruise self-driving vehicles over pedestrian risks. *Reuters*, 2025. NHTSA closes preliminary investigation.

[19] Z. Qin, H. Wang, and X. Li. Ultra fast structure-aware deep lane detection. In *European conference on computer vision*, pages 276–291. Springer, 2020.

[20] Reuters. How gm's cruise robotaxi tech failures led it to drag pedestrian 20 feet. 2024.

[21] B. Rosen. The future of autonomous vehicle sensors: A shift toward advanced stereo vision. https://www.nodarsensor.com/blog, 2024. Accessed 2025-09-23.

[22] J. Sun, Y. Xie, L. Chen, Q. Jiang, X. Zhou, and H. Bao. Disp r-cnn: Stereo 3d object detection via shape prior guided instance disparity estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2020.

[23] Volt Equity. Why LiDAR is doomed. https://www.voltequity.com/article/why-lidar-is-doomed, 2021. Accessed: 2025-09-23.

[24] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pages 63–79, 2018.

[25] Wired. Uber's self-driving car saw the woman it killed, report says. *Wired*, 2018.

[26] R. Xu, S. Razavi, and R. Zheng. Edge video analytics: A survey on applications, systems and enabling techniques. *IEEE Communications Surveys & Tutorials*, 25(4):2951–2982, 2023.

[27] Y. You, Y. Wang, W.-L. Chao, D. Garg, G. Pleiss, B. Hariharan, M. Campbell, and K. Q. Weinberger. Pseudo-lidar++: Accurate depth for 3d object detection in autonomous driving. *arXiv preprint arXiv:1906.06310*, 2019.

[28] J. Zhang, M. S. Ramanagopal, R. Vasudevan, and M. Johnson-Roberson. Listereo: Generate dense depth maps from lidar and stereo imagery. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 7829–7836. IEEE, 2020.