

An Integrated Vision-Based Intelligent Driver Assistance System for Driver Monitoring and Collision Avoidance

Bhumika Yadav
Northeastern University

Ishan Chaudhary
Northeastern University

Abstract—Road traffic accidents remain one of the leading causes of deaths worldwide, with driver fatigue, distraction, and delayed reaction times being the main contributing factors. Although modern Advanced Driver Assistance Systems (ADAS) provide partial solutions, many existing systems focus on isolated tasks such as lane keeping or collision warning, limiting their overall effectiveness. This paper presents an integrated vision-based Intelligent Driver Assistance System (IDAS) that simultaneously monitors driver behavior and the external driving environment using monocular cameras. The proposed system combines real-time driver drowsiness detection, attention monitoring, and phone usage detection with lane detection, object detection, distance estimation, and collision risk prediction. Facial landmarks extracted using MediaPipe are utilized to compute the eye aspect ratio, the mouth aspect ratio, and the head pose to assess driver alertness. The understanding of the external scene is achieved using YOLOv8-based object detection, monocular distance estimation, object tracking, and time-to-collision analysis. The system operates in real time on consumer-grade hardware and provides interpretable alerts to the driver. Experimental results demonstrate the effectiveness of the integrated approach in detecting unsafe driver states and potential collision scenarios, highlighting its suitability for cost-effective ADAS applications.

Index Terms—component, formatting, style, styling, insert

I. INTRODUCTION

Despite advancements in automotive safety technologies, road accidents continue to pose a significant global challenge. According to transportation safety studies, a large proportion of accidents are attributed not to vehicle malfunction but to **human factors**, including driver drowsiness, distraction, and reduced situational awareness. Long driving hours, mobile phone usage, and monotonous road conditions further exacerbate these risks.

Conventional ADAS solutions typically address individual problems such as lane departure warning or forward collision warning in isolation. However, safe driving requires a **holistic understanding of both the driver's internal state and the surrounding environment**. A distracted or drowsy driver may

fail to respond appropriately even when external hazards are detected.

This project proposes an **integrated vision-based Intelligent Driver Assistance System (IDAS)** that unifies internal driver monitoring with external environment perception. Using only monocular RGB cameras, the system continuously evaluates driver alertness while simultaneously detecting lanes, road users, and collision risks. The primary goal is to provide early warnings that allow the driver to take corrective action before dangerous situations escalate.

II. RELATED WORK

Driver Drowsiness Detection: Soukupová and Čech [1] introduced the Eye Aspect Ratio (EAR) metric for real-time eye blink detection using facial landmarks. Their method demonstrated robustness and efficiency, forming the foundation for many modern drowsiness detection systems. Our work extends this approach by incorporating mouth aspect ratio (MAR) and temporal smoothing to improve reliability.

Head Pose and Attention Monitoring: Murphy-Chutorian and Trivedi [2] presented a detailed survey on head pose estimation for driver assistance, highlighting its importance in assessing driver attention. Inspired by their findings, our system employs head pose estimation using 3D facial landmarks to detect prolonged off-road gaze.

Vision-Based Collision Avoidance: Xiaozhi Chen [3] demonstrated monocular vision-based object detection and scene understanding for autonomous driving. While their work focused on autonomous vehicles, our approach adapts similar concepts—object detection, tracking, and distance estimation—for real-time driver assistance using lightweight models

III. METHODS

The proposed Vision-Based Safety Intelligence System integrates three core perception and monitoring modules: (i) road and lane understanding, (ii) object detection with collision risk estimation, and (iii) driver monitoring. Each module operates independently on shared video input and contributes to a

centralized alert fusion mechanism. This section describes the methodology and algorithms used in each component in sufficient detail to allow reproducibility.

A. System Input and Video Acquisition

The system operates on monocular RGB video streams obtained either from live cameras or pre-recorded driving videos. Two video streams may be used simultaneously:

- 1) Road-facing camera for lane and object perception
- 2) Driver-facing camera for monitoring driver behavior

If a separate driver camera is unavailable, the road-facing camera is reused. Frames are processed sequentially at a target frame rate of 20–30 FPS. Each frame is copied into parallel processing pipelines to avoid interference between modules.

B. Road Segmentation and Virtual Lane Detection

- Unlike traditional lane detection methods that rely on painted lane markings, this system identifies drivable road regions using semantic segmentation. A YOLOv8 segmentation model is employed to classify each pixel into semantic categories such as *road*, *pavement*, and *terrain*. The segmentation output is a binary mask representing the drivable surface. This approach enables lane inference even in scenarios where lane markings are absent or degraded. To obtain stable lane boundaries, the following steps are applied:
 - 1) Mask Extraction: Pixels belonging to road-related classes are retained.
 - 2) Morphological Filtering: Noise is reduced using erosion and dilation.
 - 3) Lane Boundary Estimation: The left and right boundaries of the road mask are extracted row-wise.
 - 4) Polynomial Fitting: A second-order polynomial is fitted to each boundary using least squares regression.
 - 5) Temporal Smoothing: A sliding window of previous frames is used to smooth lane estimates and reduce jitter.

The resulting virtual lane region represents the ego vehicle's driving corridor and is used to spatially filter collision candidates.

C. Object Detection and Tracking

Object detection is performed using YOLOv8, a real-time single-stage detector. The model is trained to detect traffic-relevant classes including vehicles (cars, trucks, buses, motorcycles), cyclists, and pedestrians. For each frame, the detector outputs bounding boxes, class labels, and confidence scores. Non-maximum suppression (NMS) is applied to eliminate duplicate detections. To maintain object identity across frames, a lightweight centroid-based tracking algorithm is employed. Each detected object is assigned a unique ID based on spatial proximity and confidence matching. A track buffer maintains object history over multiple frames, allowing motion estimation. Objects outside the estimated lane region are optionally deprioritized to reduce false collision alerts from adjacent lanes.

D. Monocular Distance Estimation and Time-to-Collision (TTC)

Collision risk estimation is performed using monocular vision and geometric assumptions. The distance to an object is estimated using the apparent height of the detected bounding box and a known reference height for each object class.

Let H_r denote the real-world height of an object class and h_p the pixel height of its bounding box. The estimated distance D is computed as $D = \frac{H_r \cdot f}{h_p}$, where f is the camera focal length in pixel units. Relative velocity is estimated by tracking distance changes over consecutive frames. Time-to-collision (TTC) is then calculated as $TTC = \frac{D}{\Delta D / \Delta t}$.

Thresholds are applied to categorize collision risk into *safe*, *monitor*, *warning*, and *danger* levels. These thresholds are empirically tuned to balance sensitivity and stability.

E. Driver Face Detection and Landmark Extraction

Driver monitoring is implemented using MediaPipe Face Mesh, which provides 468 facial landmarks in real time. The face mesh model is accessed through its Python API, despite being implemented internally in C++, ensuring efficient execution. Facial landmarks are used to compute geometric ratios and head pose angles without requiring subject-specific calibration. A temporal buffer is maintained to detect sustained patterns rather than transient events.

Eye Aspect Ratio (EAR) and Drowsiness Detection Drowsiness detection is based on the Eye Aspect Ratio (EAR), computed from six eye landmarks:

$$EAR = \frac{\|p_1 - p_4\| + \|p_2 - p_6\| + \|p_3 - p_5\|}{2 \|p_1 - p_4\|}$$

If EAR falls below a predefined threshold for a consecutive number of frames, the driver is classified as drowsy. This temporal constraint reduces false positives caused by normal blinking.

Mouth Aspect Ratio (MAR) and Yawn Detection Yawning is detected using the Mouth Aspect Ratio (MAR), calculated from vertical and horizontal mouth landmarks. Sustained MAR values above a threshold indicate yawning behavior, which contributes to the overall drowsiness score.

Head Pose Estimation and Distraction Detection Driver head pose is estimated using a Perspective-n-Point (PnP) formulation. Selected facial landmarks are mapped to a 3D face model, and rotation vectors are computed using the solvePnP algorithm. Yaw and pitch angles are monitored to detect sustained deviations from the forward-looking direction. If the driver's gaze remains off-road beyond a configurable frame threshold, the driver is classified as distracted.

Phone Usage Detection Phone usage is detected using YOLOv8 object detection. A spatial constraint is applied to ensure that detected phones are located near the driver's head or hand region. This significantly reduces false detections from dashboard-mounted devices or passengers.

Attention Scoring and Alert Fusion Each driver monitoring signal (EAR, MAR, head pose, phone usage) contributes to a weighted attention score. Alerts are fused using a hierarchical decision strategy:

Collision danger alerts

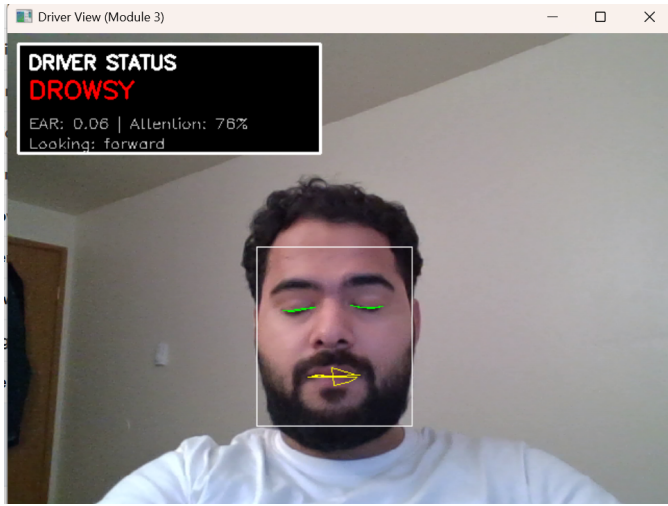


Fig. 1. Driver monitoring output showing driver status.

Severe driver impairment alerts

Moderate attention warnings

Informational alerts

This prioritization ensures that critical warnings override less urgent notifications.

F. Experiments and Results

Experimental Setup The proposed system was evaluated using recorded driving videos. The experiments were conducted on a consumer-grade laptop using a monocular RGB camera operating at a resolution of 1280×720 pixels. No specialized sensors or depth information were used.

The evaluation focused on three key capabilities of the system:

Driver monitoring and attention analysis

Surrounding object detection and tracking

Integrated alert visualization in real time

All modules were executed simultaneously to assess real-time feasibility and system-level behavior.

Driver Monitoring Results Figure 1 illustrates the output of the driver monitoring module. The system successfully detects the driver’s face and extracts facial landmarks in real time. Based on Eye Aspect Ratio (EAR), head pose angles, and temporal analysis, the driver’s attention state is classified and displayed through a dedicated status panel.

The attention score and driver state (e.g., attentive, distracted, or drowsy) are updated continuously. During testing, the system was able to correctly identify prolonged eye closure and head deviation events, triggering appropriate alerts while remaining robust to brief blinks and natural head movements.

Surrounding Object Detection and Collision Awareness Figures 1 and 2 demonstrate the system’s ability to detect and track surrounding objects in the driving environment. Vehicles, pedestrians, and cyclists are correctly identified using the YOLOv8 detector, with bounding boxes and confidence scores overlaid on the video stream.

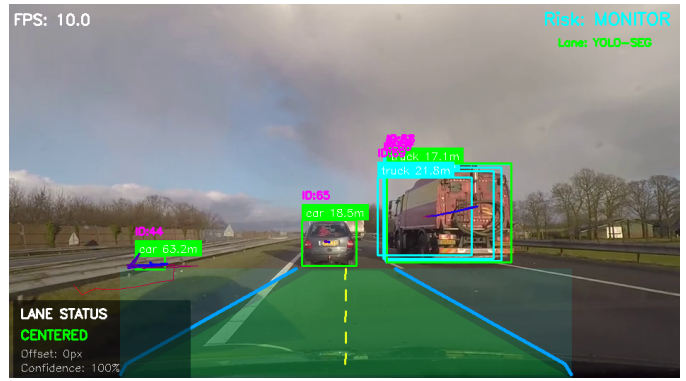


Fig. 2. Detection and tracking of surrounding vehicles in the driving scene.

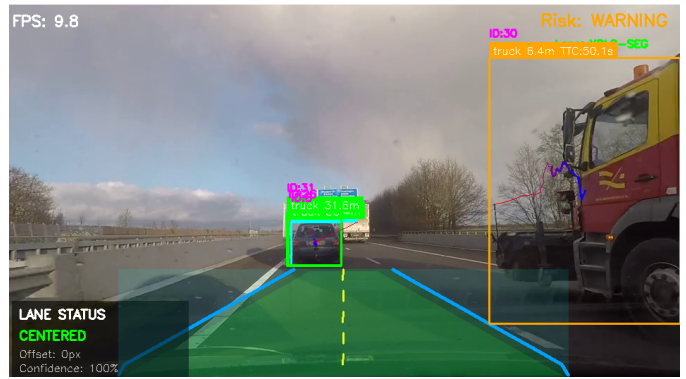


Fig. 3. Collision risk visualization with prioritized objects inside the lane region.

Tracked objects maintain consistent IDs across frames, enabling motion analysis and time-to-collision (TTC) estimation. Objects within the ego vehicle’s lane region are prioritized for collision risk assessment, reducing false warnings from adjacent lanes.

In scenarios where leading vehicles reduced distance rapidly, the system generated warning and danger alerts in a timely manner, visually highlighted using color-coded indicators.

Real-Time Performance The complete system operated at approximately 20–30 frames per second, depending on scene complexity and the number of detected objects. Temporal smoothing and confidence thresholds contributed to stable outputs and reduced alert flickering.

The modular design allowed all three components—lane perception, collision awareness, and driver monitoring—to run concurrently without noticeable lag, demonstrating the feasibility of real-time deployment on non-specialized hardware.

Qualitative Evaluation Qualitative results indicate that the system provides clear, interpretable visual feedback to the driver. The combination of on-screen alerts, lane visualization, object highlighting, and driver status information enables intuitive understanding of both internal and external risk factors.

Overall, the experiments validate the effectiveness of the proposed approach in delivering integrated vision-based safety

assistance under realistic conditions.

G. Failure and Future Scope

The proposed system relies primarily on **monocular RGB vision**, which introduces inherent limitations. Distance and time-to-collision (TTC) estimation are approximate and may be affected by camera pose variations, partial occlusions, and inaccuracies in assumed object dimensions.

Road segmentation-based lane detection, while robust to missing lane markings, can degrade under **poor lighting, adverse weather, shadows, or visually ambiguous road surfaces** such as construction zones or unpaved roads.

The driver monitoring module depends on reliable facial landmark detection and may fail under **low illumination, sunglasses, or extreme head rotations**. Phone usage detection may occasionally produce false positives or negatives due to occlusions or limited field of view.

Finally, the system uses **fixed, empirically tuned thresholds**, which may not generalize optimally across different drivers, vehicles, and camera configurations.

Future work will focus on improving robustness and adaptability. Integrating **multi-sensor fusion** (e.g., depth or radar) can enhance distance estimation and collision prediction accuracy. Learning-based and **personalized driver attention models** may reduce false alerts and adapt to individual driving behavior. The system can be extended using **temporal deep learning models** to better capture long-term fatigue patterns. Enhancements to lane perception through richer semantic understanding and evaluation on **embedded automotive platforms** will further support real-world deployment.

H. Conclusion

This project presented a **Vision-Based Safety Intelligence System** that integrates lane understanding, object detection with collision risk estimation, and driver monitoring into a unified real-time framework. By leveraging lightweight deep learning models and geometric computer vision techniques, the system demonstrates that effective driver assistance can be achieved using monocular RGB video without specialized sensors.

The proposed approach combines road segmentation-based virtual lane detection, object tracking with time-to-collision analysis, and facial landmark-based driver attention monitoring. Experimental evaluation shows that the system operates in real time and provides timely, interpretable alerts for both environmental hazards and driver impairment.

Overall, this work highlights the feasibility of modular, vision-only ADAS solutions and establishes a strong foundation for future enhancements involving sensor fusion, personalized driver models, and deployment on embedded automotive platforms.

I. ACKNOWLEDGMENT

I am deeply grateful to **Professor Bruce Maxwell** for his guidance, encouragement, and the academic freedom provided to explore and design this project. His lectures, feedback, and

course material were instrumental in shaping both the technical direction and implementation of the Vision-Based Safety Intelligence System. I would also like to acknowledge the open-source research community, particularly the developers and contributors behind **MediaPipe**, **YOLO**, and **OpenCV**, whose tools and documentation greatly facilitated the development of this system. Several online tutorials, research papers, and technical blogs were referenced during implementation and provided valuable insights into practical computer vision techniques.

J. References

- [1] T. Soukupová and J. Čech, “Real-Time Eye Blink Detection Using Facial Landmarks,” *Computer Vision Winter Workshop*, 2016.
- [2] E. Murphy-Chutorian and M. M. Trivedi, “Head Pose Estimation in Computer Vision: A Survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009.
- [3] A. Geiger, P. Lenz, and R. Urtasun, “Are We Ready for Autonomous Driving? The KITTI Vision Benchmark Suite,” *CVPR*, 2012.