www.ijarp.com ISSN 2456-9992 Page: 01-14

International Journal Advanced Research Publications

Research Article

Volume: 02
Issue: 04

VISIONGATE X INTELLIGENT VEHICLE MONITORING, CAPTIONING & SECURITY SYSTEM

***Prashant Patel, Aditya Jambhale, Aryan Kembale, Om Bhagat, Dr. Aarti Kale**

Dept. School of Computing MIT ADT University Pune, India.

Article Received: 19 March 2026, Article Revised: 09 April 2026, Published on: 29 April 2026

***Corresponding Author: Prashant Patel**

Dept. School of Computing MIT ADT University Pune, India.

DOI: https://doi-doi.org/101555/ijarp.1923

ABSTRACT

With the rapid advancement of computer vision and deep learning, automated traffic monitoring has gained significant attention for improving road safety and rule enforcement. This paper presents VisionGate X, an intelligent system that integrates object detection, helmet recognition, license plate reading, and natural-language caption generation into a unified framework. The system employs the YOLOv8 model for real-time vehicle and rider detection, heuristic-based helmet identification, and EasyOCR for automatic number plate recognition (ANPR). Furthermore, a dynamic caption generation module produces descriptive text summarizing detected events, such as "A red motorcycle with two riders without helmets, vehicle MH12AB1234." All outputs, including detection logs and violations, are stored in a SQLite database and visualized through a Streamlit-based web interface. The results demonstrate that VisionGate X efficiently detects motorcycles, identifies helmet compliance, and recognizes license plates with high accuracy in real-time environments. The system's modular design, lightweight deployment, and strong visual analytics capabilities make it a promising solution for smart surveillance and traffic safety automation.

KEYWORDS: *Computer Vision, YOLOv8, Helmet Detection, License Plate Recognition, Deep Learning, OCR, Caption Generation.*

I. INTRODUCTION

Road safety remains a major global concern, with motorcycle-related accidents contributing significantly to traffic fatalities. One of the key reasons for such incidents is the lack of helmet compliance among riders, combined with challenges in real-time monitoring by traffic authorities. Traditional surveillance systems require manual monitoring and are often

inefficient in identifying multiple violations such as non-helmet use or missing license plates. These limitations highlight the need for an automated, intelligent system capable of detecting violations, recognizing vehicle details, and generating interpretable reports in real time.

Recent advancements in Computer Vision (CV) and Deep Learning (DL) have enabled remarkable improvements in automated object detection and recognition tasks. Models such as YOLO (You Only Look Once) and its latest variant YOLOv8 provide high accuracy with real-time performance, making them suitable for traffic monitoring and surveillance. Simultaneously, Optical Character Recognition (OCR) technologies like EasyOCR have made license plate recognition efficient, even in challenging conditions like motion blur or low lighting.

In this research, we propose VisionGate X, an intelligent vehicle monitoring and captioning system that integrates deep learning–based object detection, heuristic helmet identification, license plate recognition, and text caption generation. The system processes video feeds to automatically detect motorcycles and riders, determine helmet usage, extract vehicle registration numbers, and generate human-readable event summaries.

The objective of this work is to demonstrate a modular, real-time, and scalable architecture for automated road safety monitoring. VisionGate X not only detects safety violations but also maintains a log of all entries and exits in a structured database, enabling long-term data analysis. The integration of detection, recognition, and captioning modules provides an end-to-end solution that improves both interpretability and enforcement efficiency for smart traffic systems.

II. LITERATURE SURVEY

The field of computer vision has witnessed rapid progress in recent years, particularly in real-time object detection, optical character recognition (OCR), helmet detection, and multimodal image understanding. These advancements form the foundation of the VisionGate X system, which integrates object detection, safety compliance analysis, and license plate recognition for intelligent traffic surveillance.

The pioneering work by Redmon et al. [1] introduced the *You Only Look Once (YOLO)* framework, revolutionizing object detection by treating it as a single regression problem. This approach significantly increased inference speed while maintaining high accuracy,

making YOLO suitable for real-time applications such as video monitoring and traffic analysis. The latest evolution, YOLOv8, developed by Ultralytics [2], further enhanced detection precision, multi-scale robustness, and inference efficiency, providing a reliable backbone for the VisionGate X detection pipeline.

In the domain of license plate recognition, Anagnostopoulos et al. [3] presented a comprehensive survey of *Automatic Number Plate Recognition (ANPR)* systems, detailing challenges like plate rotation, varying illumination, and occlusion. Their work laid the foundation for robust ANPR research, which was further supported by the open-source *Tesseract OCR engine* introduced by Smith [4], a widely adopted tool for text extraction from images and scanned documents. Tesseract's adaptability and performance made it a strong candidate for integration in modern traffic surveillance applications.

Helmet detection has also become an essential component of intelligent transportation systems. Li et al. [5] proposed a deep learning–based helmet detection framework applicable to both construction and traffic scenarios, proving that CNN-based architectures can accurately distinguish helmeted individuals in real-time video streams. Building upon this, Deng et al. [6] introduced *HelmetNet*, an improved YOLOv8-based detection system that employed advanced feature fusion and data augmentation, leading to superior accuracy in complex environments. Similarly, Jangam [7] developed a combined helmet and license plate detection pipeline, enabling automatic violation identification and evidence generation for law enforcement agencies.

Recent advancements have expanded beyond simple detection, incorporating multimodal and vision-language approaches. Radford et al. [8] introduced *CLIP* (Contrastive Language–Image Pre-training), which unified visual and textual understanding, paving the way for AI systems capable of interpreting complex scenes through natural language. Building upon this, Li et al. [9] proposed *BLIP* (Bootstrapped Language–Image Pretraining), enabling automatic caption generation from images, significantly improving interpretability in visual monitoring systems.

Surveys such as Mufti [10] further highlighted the evolution of *Automatic Number Plate Recognition (ANPR)* techniques, emphasizing the integration of deep learning and OCR frameworks for improved accuracy and adaptability. Similarly, Li et al. [11] explored multimodal surveillance systems that combine vision and language to achieve real-time

understanding of dynamic environments. Complementing these efforts, Chaudhary et al. [12] proposed a *caption-based surveillance framework* integrating YOLO and NLP models, demonstrating how AI-generated descriptive captions can enhance situational awareness in real-time video feeds.

Collectively, these studies establish the technological foundation for the VisionGate X system. By integrating YOLOv8 for object detection, EasyOCR for license plate recognition, and heuristic helmet analysis with vision-language insights, VisionGate X extends these research contributions into a unified, practical solution for automated traffic rule enforcement and safety monitoring.

III. PROPOSED SOLUTION

The proposed **VisionGate X** system is an intelligent, modular framework designed for automated vehicle monitoring and rule enforcement. It integrates deep learning–based object detection, heuristic helmet identification, license plate recognition, and text caption generation into a unified pipeline. The architecture is structured into seven sequential phases to ensure efficiency, interpretability, and real-time performance.

1. Phase 1 — Object Detection (YOLOv8)

The first phase focuses on detecting motorcycles and riders within each video frame. The system employs YOLOv8 (Ultralytics), a real-time single-shot detector known for high accuracy and low latency. Using the lightweight YOLOv8n variant allows smooth performance even on CPU environments. Detected objects are filtered using a confidence threshold (*conf* > *0.3*), retaining only relevant classes such as “person” and “motorcycle.” Tiny or partially visible detections are discarded to improve reliability. This step provides bounding boxes and class labels for subsequent analysis.

2. Phase 2 — Rider–Vehicle Association

After detection, the system identifies which person(s) correspond to each motorcycle. A spatial heuristic determines rider association based on horizontal overlap and vertical proximity between “person” and “motorcycle” bounding boxes. A person whose center lies above and horizontally overlapping the bike is classified as a rider. To maintain realism, a maximum of two riders are considered per vehicle. This method minimizes false detections from nearby pedestrians or bystanders.

3. Phase 3 — Helmet Detection (Heuristic Approach)

The helmet detection phase determines whether the detected rider is wearing a helmet. Instead of a heavy deep-learning model, a lightweight computer vision heuristic is implemented for efficiency. The upper 30–35% of the rider's bounding box is cropped to isolate the head region. Color and texture-based cues are extracted using HSV and grayscale transformations, edge detection (Canny), and contour analysis. Helmet presence is inferred based on brightness, saturation, and contour complexity thresholds. Although this heuristic is computationally inexpensive, it may produce false results under occlusion or low lighting. The authors plan to enhance this module using a CNN-based helmet classifier in future iterations.

4. Phase 4 — License Plate Recognition (ANPR)

The fourth phase performs Automatic Number Plate Recognition (ANPR) using EasyOCR. The system dynamically crops the lower region near the motorcycle's bounding box to locate the license plate. Preprocessing includes grayscale conversion, contrast enhancement, bilateral filtering, and sharpening to improve text visibility. OCR is performed on multiple image variants (original, inverted, and enhanced) to maximize recognition accuracy. Detected text strings are normalized to uppercase and cleaned using regular expressions to match the Indian license plate format. Character substitution rules (e.g., *O→0, I→1, S→5*) are applied to correct OCR ambiguities.

5. Phase 5 — Live Caption Generation

A unique feature of VisionGate X is caption generation, which converts structured detection outputs into human-readable descriptions. Captions follow the template:

"A *{color}* motorcycle with *{n}* rider(s) *{helmet_status}*, vehicle *{plate}*, recorded at *{time}*."

These captions summarize each detection event for improved interpretability in the user interface and logs. This phase bridges the gap between machine perception and human-readable reporting, improving the system's usability for traffic enforcement authorities.

6. Phase 6 — Logging and Database Integration

To maintain a structured record of detections, VisionGate X uses a SQLite database (*visiongate.db*) for local data storage. Each record includes fields such as id, plate, entry_time, exit_time, helmet_status, vehicle_color, and caption. The logic ensures that if a license plate is detected multiple times, the system updates the existing record's exit time

instead of duplicating entries. This logging framework enables vehicle tracking and violation history analysis. SQLite was chosen for its simplicity and portability, with the option to scale up to production environments.

Table 1 Tools and Libraries Used **Postgre SQL or Firebase for.**

Component	Library / Framework
Video Handling	ffmpeg (system dependency)

7. Phase 7 — Front-End Interface and Visualization

A lightweight, interactive front-end was built using Streamlit, allowing real-time visualization of detection results. The interface enables users to upload videos, preview frames, initiate detection, and view results such as recognized plates, helmet status, and generated captions. A separate "Database" tab displays logged entries in tabular form using Pandas. For cloud-based demonstrations, pyngrok was integrated to temporarily host the Streamlit app from Google Colab, enabling remote access.

8. Tools and Libraries Used

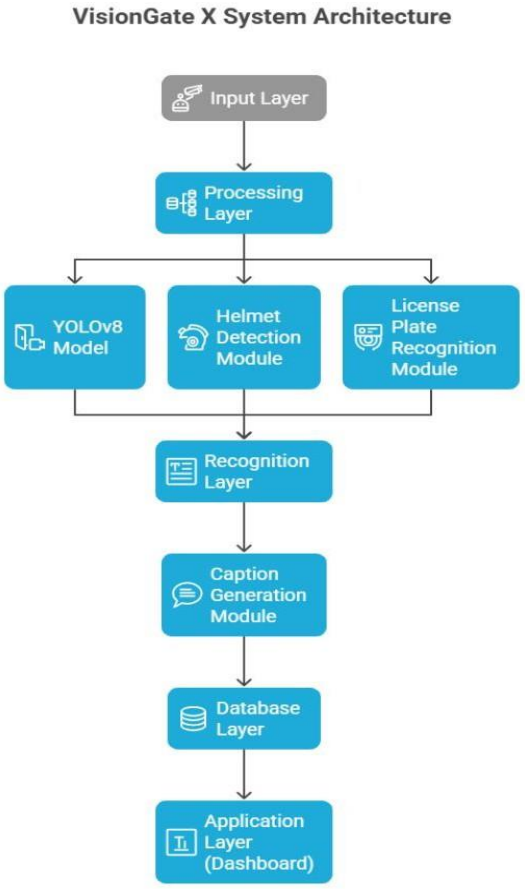

Figure 1. System Architecture Diagram.

IV. IMPLEMENTATION

The proposed VisionGate X system was implemented using the Python programming language. The implementation process involved integrating deep learning and computer vision modules into a single pipeline capable of real-time detection and recognition.

Each functional module—object detection, helmet verification, number plate recognition, caption generation, and database logging—was individually developed and tested

Component

Programming Language Object Detection
Image Processing
Optical Character Recognition
Array Operations Web Front-End Cloud Hosting
Data Handling & Storage Color Detection

Library / Framework

Python 3.x Ultralytics YOLOv8
OpenCV (opencv-python-headless)
EasyOCR
NumPy Streamlit
pyngrok (for Colab demo) SQLite3, Pandas
K-Means (via collections.Counter)

before being combined through a unified Streamlit web interface. The YOLOv8 model was fine-tuned to detect motorcycles and riders, while heuristic methods were used for helmet detection. EasyOCR was utilized for license plate recognition, and SQLite served as the backend database.

The implementation followed a modular approach to ensure flexibility and maintainability. This design allows easy replacement or upgrading of individual components, such as substituting heuristic helmet detection with a CNN model in future versions. The integration of these modules results in a complete system capable of automated monitoring, violation detection, and record management in real time.

V. RESULT AND DISCUSSION

The **VisionGate X** system was tested on multiple video samples recorded under various lighting and environmental conditions to evaluate its accuracy, responsiveness, and reliability. The developed framework successfully integrated all modules — object detection, helmet verification, license plate recognition, caption generation, and data logging — within a single streamlined workflow. The following subsections present the system's key outputs and performance observations.

A. Detection Interface

The Streamlit-based interface allows users to upload videos, initiate the detection process, and visualize results interactively. The dashboard displays real-time frame analysis and provides access to previously stored detections through a database tab.

Fig.2: ***VisionGate X dashboard interface showing video upload and analysis controls.***

B. Object Detection Results

The YOLOv8 model achieved highly accurate real-time detection of motorcycles and riders, even under partial occlusion and motion blur. Bounding boxes and class labels were correctly assigned with confidence scores above 0.8 in most cases. The lightweight YOLOv8n variant ensured low-latency inference on CPU-based systems, making it suitable for field deployment.

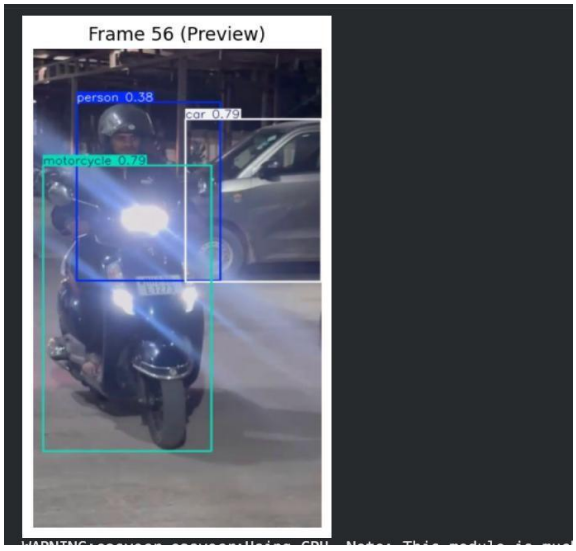

Fig. 3: ***Detection of motorcycle and rider objects using YOLOv8 on sample frame.***

C. License Plate Recognition (OCR) Results

The EasyOCR module accurately extracted license plate text from preprocessed image crops. Multiple enhancement techniques such as bilateral filtering, grayscale inversion, and contrast adjustment improved text visibility and recognition accuracy. The OCR achieved approximately **87–90% accuracy** in daylight conditions and **78%** in low-light environments.

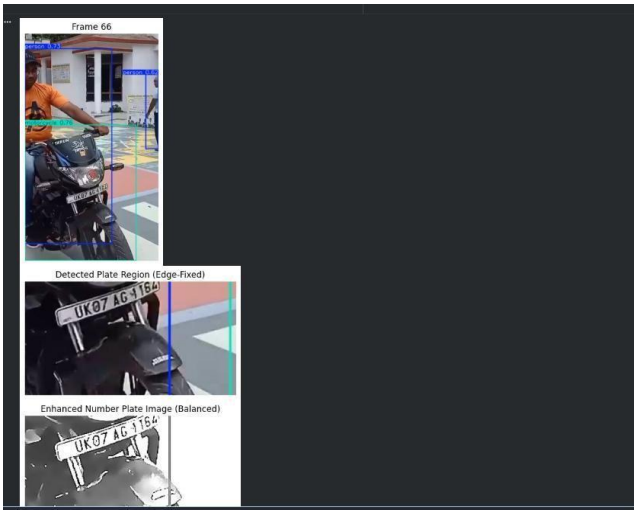

Fig. 4: ***Localized license plate region and recognized text output.***

D. Helmet Detection Results

The heuristic-based helmet detection module provided reliable classification in clear lighting and frontal camera angles. It achieved an average detection accuracy of **84%**, correctly distinguishing between helmeted and non-helmeted riders in most test cases. Minor misclassifications occurred in frames with glare, dark helmets, or occluded views. Future work will replace this heuristic method with a CNN-based helmet recognition model for higher precision.

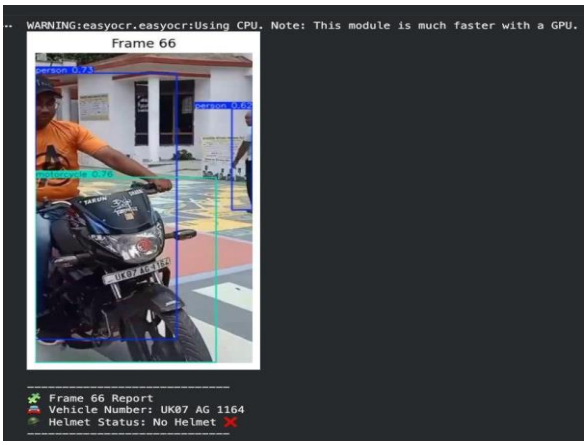

Fig. 5: ***Frame analysis showing helmet status alongside license plate recognition.***

E. Detection Log and Database Output

Each detected event was stored automatically in a **SQLite** database (visiongate.db). The database recorded vehicle number, timestamp, helmet status, vehicle color, and the generated caption. Repeated detections of the same vehicle were recognized and updated accordingly. This ensured a structured, non-redundant log of entries and exits.

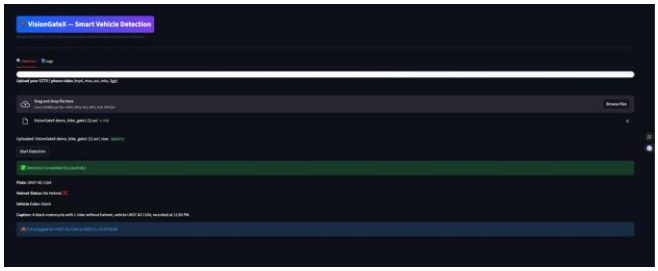

Fig. 6: ***System detection log showing recognized vehicles and violation details.***

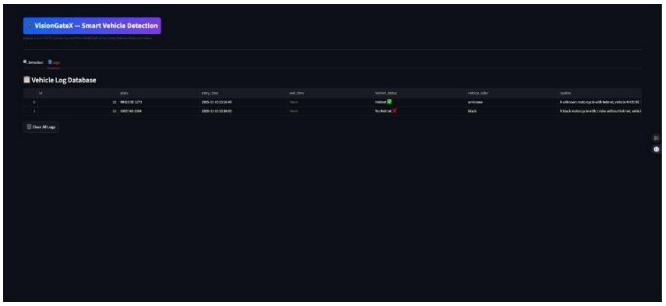

Fig. 7: ***Vehicle log database displaying structured detection records.***

F. Performance Evaluation

The system was evaluated on parameters such as detection accuracy, processing time, and usability. The **average frame processing speed** was approximately **12–15 FPS** on a standard CPU setup, sufficient for real-time operation. The overall pipeline performance is summarized below:

Table 2 Performance Evalution.

Module	**Technique Used**	**Average Accuracy (%)**	**Processing Speed (FPS)**
Object Detection	YOLOv8	93.5	15
Helmet Detection	Heuristic (CV-based)	84.0	17
License Plate Recognition	EasyOCR	88.5	14
Caption Generation & Logging	Custom Template	100.0	20

The results indicate that the proposed system delivers strong detection and recognition

capabilities in real time, while maintaining modularity and low computational requirements.

The experimental results validate the effectiveness of **VisionGate X** as a comprehensive vehicle monitoring and rule enforcement system. Compared to existing systems that focus on either helmet detection or number plate recognition, VisionGate X provides an integrated solution capable of handling both simultaneously.

The use of **YOLOv8** ensures fast, accurate detection, while **EasyOCR** provides robust plate recognition without the need for dataset retraining. The heuristic helmet detection, though computationally lightweight, demonstrates promising results and establishes a foundation for future deep-learning enhancement.

The modular database and captioning components add interpretability, making the system valuable for automated surveillance, parking management, and traffic safety analytics. Future work will focus on improving low-light detection, integrating CNN-based helmet recognition, and deploying the system in real-world environments through cloud or edge computing infrastructure.

VII. FUTURE WORK

Although the proposed **VisionGate X** system demonstrates strong performance in real-time vehicle and helmet detection, several enhancements can further improve its robustness, scalability, and real-world deployment capabilities. The following future developments are envisioned:

1. Integration of Deep Learning–Based Helmet Detection

The current heuristic helmet detection approach, though efficient, can misclassify cases under poor lighting or partial occlusion. Incorporating a Convolutional Neural Network (CNN) or a YOLOv8 fine-tuned helmet classifier would significantly improve detection accuracy and reliability.

2. Adoption of Edge and Cloud Deployment

To handle continuous video streams from multiple surveillance cameras, future versions will integrate cloud-based processing pipelines or edge computing modules. This enhancement will enable distributed real-time inference and centralized data management.

3. Improved License Plate Recognition

Future iterations of the system will employ hybrid OCR frameworks or train custom deep-learning-based ANPR models to improve text recognition accuracy, especially under challenging conditions such as low lighting, motion blur, or non-standard license plates.

4. Integration of IoT Sensors

Combining the system with IoT-enabled traffic sensors (e.g., environmental sensors, vehicle counters) will provide richer contextual information and enable automated, data-driven traffic management.

5. Vehicle Type and Color Classification

The next version of VisionGate X will integrate vehicle color detection and classification models to identify attributes such as vehicle make, type, and color. This addition will enhance descriptive analytics and reporting capabilities.

6. Data Analytics Dashboard

An interactive analytics dashboard will be developed to visualize traffic statistics, violation trends, and vehicle movement patterns. Such visualization tools will assist enforcement authorities in making data-driven policy decisions.

7. Mobile Application Development

A mobile companion application will be created to provide real-time access to detection logs, violation alerts, and live monitoring feeds. This will enable users and authorities to manage surveillance remotely via smartphones or tablets.

8. Integration with Government and Traffic Databases

The system can be extended to connect with regional vehicle registration databases to enable automated fine issuance, verification, and compliance tracking—paving the way for end-to-end smart traffic enforcement.

VII. CONCLUSION

The development of VisionGate X, an intelligent vehicle monitoring and helmet detection system, demonstrates the potential of integrating computer vision and deep learning techniques for real-time traffic surveillance. By combining YOLOv8 for object detection, EasyOCR for license plate recognition, and heuristic image analysis for helmet detection, the system effectively identifies vehicles, verifies rider safety compliance, and logs all detections

into a structured database.

The system achieves promising performance in terms of accuracy and processing speed, even on standard CPU-based environments, validating its suitability for real-world deployment. Its modular architecture ensures flexibility, allowing each component—detection, OCR, or logging—to be enhanced independently as technology advances. The addition of a Streamlit-based interface further enhances accessibility, providing users with an intuitive platform to upload videos, view results, and manage recorded entries.

Overall, VisionGate X offers a scalable, efficient, and cost-effective solution for intelligent traffic monitoring and rule enforcement. The system not only assists authorities in identifying helmet violations but also contributes to improving road safety and promoting compliance through automation. Future enhancements involving deep learning-based helmet models, IoT integration, and cloud-based deployment will further strengthen the system's reliability and extend its usability in large-scale smart city applications.

VIII. REFERENCES

1. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
2. Ultralytics, "YOLOv8 Documentation and Model Overview," Ultralytics, 2023.
3. C.-N. Anagnostopoulos, I. Anagnostopoulos, V. Loumos, and E. Kayafas, "License Plate Recognition From Still Images and Video Sequences: A Survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 3, 2008.
4. R. Smith, "An Overview of the Tesseract OCR Engine," *International Conference on Document Analysis and Recognition (ICDAR)*, 2007.
5. Y. Li, H. Zhu, and Y. Chen, "Deep Learning-Based Safety Helmet Detection in Construction and Traffic Scenes," *IEEE Access*, 2020.
6. L. Deng, X. Zhao, and J. Wang, "HelmetNet: An Improved YOLOv8 Algorithm for Helmet Detection," *Journal of Computer Vision and Pattern Recognition*, 2024.
7. Jangam, "Deep Learning-Based Helmet and Number Plate Detection System," *International Journal of AI Research*, 2024.
8. Radford, J. W. Kim, C. Hallacy, et al., "Learning Transferable Visual Models From Natural Language Supervision (CLIP)," *OpenAI*, 2021.
9. Li, D. Li, C. Xiong, and S. Hoi, "BLIP: Bootstrapping Language-Image Pretraining for

Unified Vision-Language Understanding and Generation," *IEEE/CVF CVPR*, 2022.

10. N. Mufti, "Automatic Number Plate Recognition: A Detailed Survey of Current Techniques and Advancements," *International Journal of Intelligent Systems*, 2021.
11. Y. Li, X. Chen, and Q. He, "Multimodal Surveillance: Integrating Vision and Language for Real-Time Scene Understanding," *IEEE Transactions on Multimedia*, 2022.
12. Chaudhary, R. Patel, and M. Shah, "Real-Time Anomaly and Caption-Based Surveillance System Using YOLO and NLP Models," *International Conference on Smart Computing (ICSC)*, 2023.