

Multi-Modal Computer Vision System for Red Light Running Detection

by

Griffin Wells Herring

A thesis submitted to the Graduate Faculty of
Auburn University
In partial fulfilment of the
Requirements for the Degree of
Master's of Science

Auburn, Alabama
December 13, 2025

Keywords: Red Light Runners (RLR), Right On Red (ROR), Legal Pass, Computer
Vision

Copyright 2025 by Griffin Wells Herring

Approved by

Xiao Qin, Chair, Alumni Professor of Computer Science and Software Engineering
Laurence Rilett, Co-chair, Distinguished Professor of Civil and Environmental
Engineering
Adrian Cottam, Assistant Research Professor of Civil and Environmental Engineering

Abstract

Ensuring safety at urban intersections remains a critical challenge, where collisions can result in severe injuries and fatalities. A primary contributor to these incidents is vehicles running red signals. Current automated systems for detecting this behavior often suffer from high false-positive rates, frequently stemming from imprecise, zone-based detection methods. Furthermore, the efficacy of model-based approaches is typically hindered by the scarcity of specialized datasets capturing red-light-running events.

To address these shortcomings, this thesis proposes a novel, dual-model approach that decouples vehicle detection from violation inference. The system utilizes a YOLOv10 model for vehicle tracking and a YOLOv11 model for traffic signal recognition operating in parallel. The vehicle model is pretrained, while the traffic signal model is fine-tuned on a custom-annotated dataset (derived from public ALGO traffic cameras) to monitor the state of traffic signals (Red, Yellow, or Green). These models are integrated with a custom logic framework that employs configurable "tripwire" lines at the entry and exit points of each lane.

This logic precisely correlates a vehicle's trajectory with the concurrent traffic signal state, allowing for an accurate distinction between vehicles adhering to traffic laws and those committing violations. By saving tripwire pairs to an external configuration file, the system can be rapidly deployed and adapted to any intersection, ensuring both scalability and ease of setup. This methodology eliminates the dependency on rare, violation-specific training data and instead

creates a highly effective and robust solution aimed at improving safety for all road users in urban environments.

Artificial Intelligence (AI) Use Disclosure Statement

In the preparation of this thesis, the following Artificial Intelligence (AI) tools were used: Google Gemini. These tools were used primarily to reword and rewrite certain paragraphs to have a more academic tone. The author acknowledges full responsibility for the intellectual content of this work and has ensured that all AI-assisted sections have been reviewed and revised for accuracy and appropriate academic style. All AI-generated content was reviewed and validated for relevance, appropriateness, and accuracy before incorporation into the final document to maintain scholarly integrity of this research.

Acknowledgements

I would like to express my sincere gratitude to the many individuals who have supported me throughout this research and my graduate studies.

First, I wish to thank my thesis committee members for their invaluable guidance. My deepest appreciation goes to Assistant Research Professor Adrian Cottam for his mentorship and direction. I am also incredibly grateful to Professor Laurence Rilett and Professor Xiao Qin for their insightful feedback and expertise, which have been instrumental in shaping this research. I would also like to extend my thanks to Assistant Research Professor Ernest Tufuor for his support and encouragement.

This journey would not have been possible without the unwavering support of my family. I want to offer my heartfelt thanks to my parents, Derry and Maryjayne Herring. Their constant love, encouragement, and belief in me have been my foundation. To my brothers and sister, Wriley, Drew, and Olivia, thank you for your endless support and for always being there for me.

Thank you all.

Table of Contents

Abstract	2
Artificial Intelligence (AI) Use Disclosure Statement	4
Acknowledgements.....	5
List of Figures.....	9
List of Equations.....	10
Chapter 1: Introduction	11
1.1 The Societal Imperative for Advanced Intersection Safety	11
1.2 The Technological Evolution of Automated Enforcement: A Legacy of Compromise.....	15
1.3 The Paradigm Shift: Deep Learning for Intelligent Traffic Analysis.....	18
1.4 A Critical Analysis of the Contemporary Research Gap	22
1.5 Thesis Contribution: A Decoupled, Logic-Driven Framework for RLR Detection .	24
Chapter 2: Literature Review	27
2.1 From In-Pavement Loops to Computer Vision.....	27
2.2 YOLO and the Advancement to YOLOv10	28
2.3 Identifying the Research Gap in Existing Literature	29
2.4 Justifying the Decoupled, Logic-Driven Framework.....	30
Chapter 3: Methodology	32
3.1 Architectural Overview	32
3.2 Data Acquisition and Preparation	34
3.3 The Dual-Model Perception Engine.....	36
3.4 Tripwire Logic for Violation Inference	37
3.5 Performance Evaluation Metrics.....	38
3.6 Chapter Summary	40
Chapter 4: Results and Analysis.....	42
4.1 Validating the End-to-End System.....	42
4.2 The Logic-Driven Framework: Tracking and Inference	43
4.3 Component Validation: The Traffic Signal Recognition Module.....	44
4.4 End-to-End System Viability and Analysis.....	51

4.5 Chapter Summary	53
Chapter 5: Discussion.....	54
5.1 Synthesis of Key Findings: Validating the Decoupled, Logic-Driven Framework .	54
5.2 Analysis of the Custom-Trained Model in the End-to-End System	56
5.3 Broader Implications for Intelligent Traffic Systems	57
5.4 Limitations of the Research and Acknowledgment of Constraints	58
5.5 Chapter Summary and Transition.....	59
Chapter 6: Concluding Remarks.....	60
6.1 A Validated Framework for RLR Detection.....	60
6.2 Promising Directions for Future Research	61
References	63

List of Tables

Table 1.1 Performance Evolution and Comparison of Real-Time Object Detection Architectures. Data Sourced from (Wang et al., 2024).....	20
Table 4. 1 Per-Class Performance Metrics for Traffic Signal Recognition Module.....	50

List of Figures

Figure 1. 1 General case of a vehicle running a red light.....	12
Figure 1. 2 Depicting deaths throughout the years caused by red light runners. Data presented from the study done by the Insurance Institute for Highway Safety.....	13
Figure 3. 1 General Outline of Red-Light Detector System Architecture	33
Figure 3. 2 This image is depicting the Green Street intersection where the data was collected	36
Figure 4. 1 Custom ALGO-Derived Dataset Characteristics. (a) Class Instance Distribution, (b) Bounding Box Center (x,y) Distribution, (c) Bounding Box (width, height) Distribution.....	45
Figure 4. 2 Training and Validation Metrics over 30 Epochs.....	47
Figure 4. 3 Absolute Confusion Matrix of Selected Model (Epoch 22)	49
Figure 4. 4 Normalized Confusion Matrix of Selected Model (Epoch 22)	50

List of Equations

Equation 1	39
Equation 2	39
Equation 3	40
Equation 4	48

Chapter 1: Introduction

1.1 The Societal Imperative for Advanced Intersection Safety

The act of running a red light, formally known as red-light running (RLR), represents far more than a mere traffic infraction; it is a pervasive and lethal behavior that constitutes a critical public health and safety crisis on roadways worldwide. In the United States alone, the year 2023 saw 1,086 individuals killed and more than 135,000 injured in crashes directly involving red-light running (Insurance Institute for Highway Safety, 2024). A study conducted over several months at five busy intersections in Fairfax, Virginia, revealed that, on average, a motorist ran a red light every 20 minutes at each intersection, with violation rates increasing during peak travel times (Insurance Institute for Highway Safety, 2024). The Federal Highway Administration (FHWA) further contextualizes this danger, reporting that approximately one-third of all fatal intersection crashes occur at signalized intersections, with a significant portion attributable to RLR (Federal Highway Administration, 2024b). These statistics underscore a persistent and deadly challenge within modern traffic systems that demand advanced technological intervention.

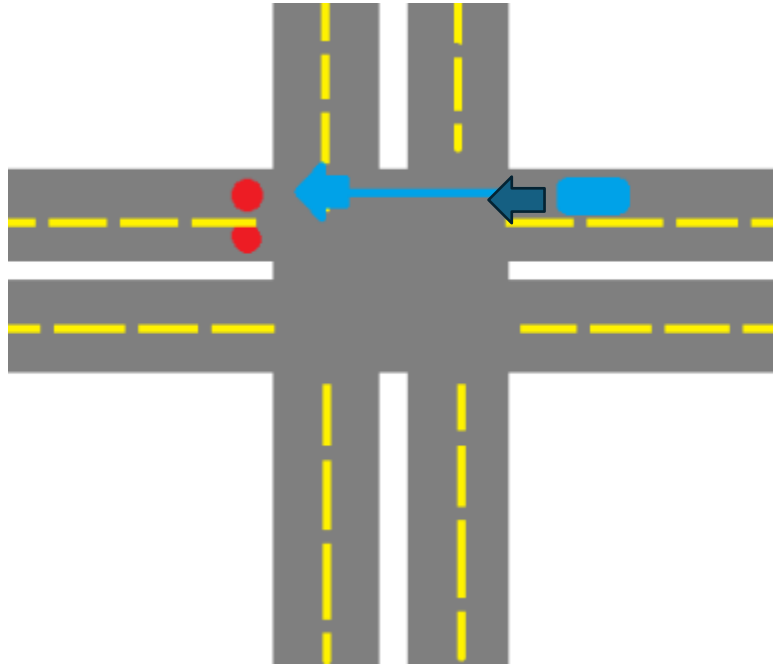


Figure 1. 1 General case of a vehicle running a red light

A particularly tragic dimension of the RLR problem is its disproportionate and devastating impact on vulnerable road users—those outside the violating vehicle. For Example: roughly half of the individuals killed in RLR crashes are pedestrians, bicyclists, and occupants of other vehicles who are struck by the red-light runner (Insurance Institute for Highway Safety, 2024). This statistic reframes RLR from a driver-centric violation into a profound public safety menace that unjustly endangers the most exposed members of the community. This issue is compounded by a broader, alarming trend in traffic safety. Between 2009 and 2023, pedestrian fatalities increased by a staggering 80%, a rate of increase that far outpaces the 13% rise in all other traffic-related deaths during the same period (Governors Highway Safety Association, 2024). In 2023, a pedestrian was killed in a traffic crash every 72 minutes in the United States, a figure that highlights the perilous environment that RLR violators perpetuate (National Highway Traffic Safety Administration, 2024).

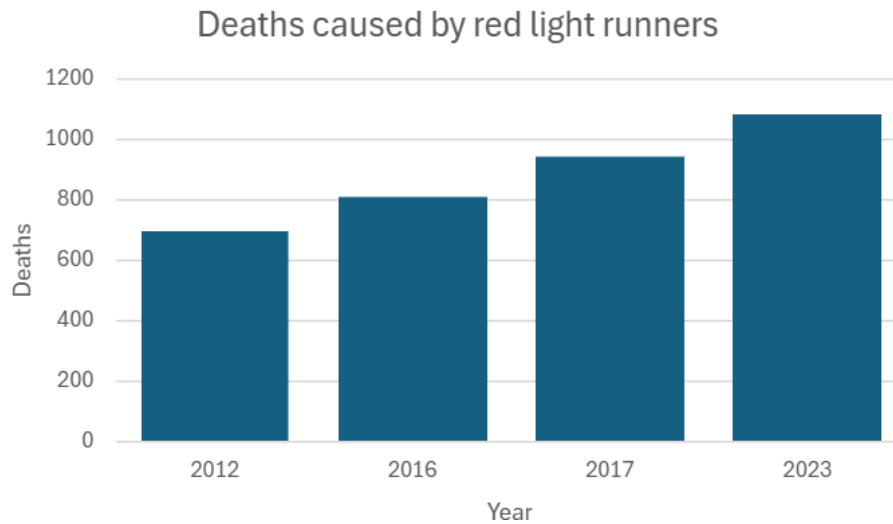


Figure 1. 2 Depicting deaths throughout the years caused by red light runners. Data presented from the study done by the Insurance Institute for Highway Safety.

While overall traffic fatalities have shown a slight downward trend, projected to fall below 40,000 in 2024 for the first time since 2020, the data for RLR-related deaths reveals a contrary and worsening trajectory (Insurance Institute for Highway Safety, 2024). The number of fatalities from RLR has climbed relentlessly, from 696 deaths in 2012 to 811 in 2016, 939 in 2017, and reaching 1,086 in 2023 depicted in figure 1.2 (Insurance Institute for Highway Safety, 2024). This divergence demonstrates that general safety measures, such as improvements in vehicle crashworthiness and roadway design, are failing to curb this specific high-risk behavior. This suggests an "RLR exceptionalism," where the problem is uniquely resistant to generalized solutions and is driven by behavioral factors that require targeted, high-certainty enforcement. The research presented in this thesis is therefore not merely an effort to improve an existing system but a necessary response to a problem that has proven uniquely resilient to conventional safety engineering.

The persistence of this dangerous behavior is rooted in a complex psychological paradox—a significant gap between drivers' perception of risk and their actual conduct. A 2023 national telephone survey conducted by the AAA Foundation for Traffic Safety found that while 81% of drivers acknowledge that driving through a red light is "very or extremely dangerous," a startling 27% admitted to having done so within the past 30 days (AAA, 2019). This profound disconnect between perceived risk and reported behavior suggests that public awareness campaigns, while important, are insufficient on their own. The problem is not a lack of education; rather, it is a behavioral failure where the perceived risk of consequence is too low to deter the action. This underscores the necessity of robust, consistent, and automated enforcement to bridge this behavioral gap and protect the public from a known, predictable harm that individuals inflict upon others despite knowing better.

Beyond the immeasurable human cost, RLR imposes a massive economic toll on society. The National Highway Traffic Safety Administration (NHTSA) estimated the total economic cost of all motor vehicle crashes in 2019 at an immense \$340 billion, a figure equivalent to \$1,035 for every person in the United States (Federal Highway Administration, 2005). When non-economic quality-of-life valuations are included, the total societal harm from these crashes escalates to nearly \$1.4 trillion (Federal Highway Administration, 2005). More specifically, research has demonstrated that red-light camera (RLC) systems, despite their implementation costs, yield a net aggregate crash cost benefit by significantly reducing the frequency of severe and expensive right-angle collisions (Federal Highway Administration, 2005). With a single fatal crash estimated to cost society \$6 million and an injury-causing crash costing \$126,000, the prevention of

these high-severity incidents through effective enforcement is a matter of profound fiscal prudence as well as a moral imperative (Federal Highway Administration, 2005).

The life-saving potential of automated enforcement is not merely correlational but has been established through direct causal evidence. Multiple studies confirm that RLCs reduce fatal crashes, with a landmark Insurance Institute for Highway Safety (IIHS) study finding a 21% reduction in the fatal RLR crash rate in large cities that implemented camera programs (Insurance Institute for Highway Safety, 2024). Even more compelling is the evidence from the deactivation of these systems. A follow-up IIHS study examined 14 cities that shut down their RLC programs and found that the subsequent rate of fatal RLR crashes was 30% higher than it would have been had the cameras remained operational (Insurance Institute for Highway Safety, 2016). However, the engineering benefits of automated enforcement often collide with policy and public perception. Several U.S. cities have implemented red-light cameras only to later remove them due to voter-driven opposition, reflecting a persistent tension between safety outcomes and political acceptance (Stateline, 2018). This powerful before-and-after analysis provides clear, data-driven proof of a causal link between automated enforcement and the prevention of fatalities. It frames the research presented in this thesis not merely as an effort to improve a technology, but as a critical mission to enhance and advance a public health intervention whose absence is demonstrably deadly.

1.2 The Technological Evolution of Automated Enforcement: A Legacy of Compromise

The development of Automated Enforcement Systems (AES) was born from the inherent limitations of traditional, manual traffic enforcement. The sheer number of signalized intersections in any urban environment makes it a logistical and financial impossibility for law enforcement personnel to be physically present at every location, creating a significant enforcement gap that technology was designed to fill (Insurance Institute for Highway Safety, 2024). This necessity spurred the creation of automated solutions, with the first modern red-light camera program being implemented in New York City in 1992, marking the beginning of a new era in traffic safety enforcement (Insurance Institute for Highway Safety, 2024).

The foundational technology of these first-generation systems was the in-pavement inductive loop detector (National Highway Traffic Safety Administration, 2005). The operational principle is based on electromagnetism: one or more loops of electrical wire are buried in the asphalt near the intersection's stop line. An electrical current passed through these wires generates a stable magnetic field. When a large metallic object, such as a vehicle, passes over the loop, it disrupts this field, altering the circuit's inductance (National Highway Traffic Safety Administration, 2005). This change is detected by a connected processing unit, which registers the vehicle's presence. A typical RLC installation uses two such loops per lane: the first is placed at the stop line to detect the initial incursion during a red phase, and a second is placed further into the intersection to confirm that the vehicle passed through, thereby differentiating a deliberate violation from a minor overshoot of the stop line (National Highway Traffic Safety Administration, 2005).

While conceptually sound, this legacy technology is fraught with vulnerabilities that severely limit its reliability, scalability, and long-term cost-effectiveness. The primary drawback is its physical fragility and high maintenance overhead. The installation and repair of inductive loops are highly disruptive and expensive, requiring road surfaces to be cut and repaved (National Highway Traffic Safety Administration, 2005). The loops themselves are highly susceptible to damage from the natural deterioration of the road, such as cracking and shifting pavement, which can break the embedded wires and lead to intermittent or complete system failure (National Highway Traffic Safety Administration, 2005). These issues are often exacerbated by environmental factors like wet weather and freeze-thaw cycles (National Highway Traffic Safety Administration, 2005). Furthermore, the detection accuracy of these systems is prone to a variety of well-documented errors. These include "splashover," where the system falsely detects a vehicle in an adjacent lane; "pulse breakup," where a single long vehicle generates multiple distinct signals; and chronic sensitivity problems that can lead to either missed detections (false negatives) or false calls (false positives) (National Highway Traffic Safety Administration, 2005). Studies have found the counting accuracy of some loop systems to be as low as 80%, with performance degrading over time and requiring frequent, costly recalibration (National Highway Traffic Safety Administration, 2005).

The significant drawbacks of in-pavement systems spurred the development of non-invasive, above-ground sensors, marking the second generation of AES technology. This technological evolution was driven as much by economic and operational realities as by a quest for higher accuracy. The high installation costs, traffic disruption during maintenance, and frequent failures of inductive loops resulted in a very

high total cost of ownership (TCO) (Middleton & Parker, 2003). Above-ground systems, simplifying installation and maintenance, offer a drastically lower TCO, making them a more practical and scalable solution. This generation of technology includes radar, LIDAR, and early video-based systems. While these systems solved the problem of physical fragility, they introduced their own set of compromises. For instance, early video detection systems, while a conceptual leap forward, often struggled with performance degradation at night and in low-light conditions, and were particularly prone to false activations caused by vehicle headlight reflections on wet pavement (National Highway Traffic Safety Administration, 2005). This historical progression reveals a clear technological trajectory driven by a desire to reduce long-term maintenance burdens and overcome the inaccuracies of single-modality systems. The research presented herein is not an attempt to solve a solved problem, but rather to address a long-standing challenge by proposing a next-generation solution designed to transcend the known failure points of its technological predecessors.

1.3 The Paradigm Shift: Deep Learning for Intelligent Traffic Analysis

The evolution of automated enforcement has culminated in a paradigm shift driven by advances in artificial intelligence (AI), specifically in the fields of computer vision and deep learning. This transition marks a move away from simple sensor-based triggers toward intelligent systems capable of perceiving and interpreting complex, dynamic traffic scenes. AI-driven video analytics, powered by sophisticated machine learning algorithms, can process real-time video feeds to identify and classify different road users (vehicles, pedestrians, cyclists), track their movements and trajectories, and

automatically detect anomalous or proscribed behaviors (Rapid Innovation, 2024). This capability enables a more nuanced and potentially proactive approach to traffic management and safety enforcement.

At the forefront of this technological revolution is the "You Only Look Once" (YOLO) family of convolutional neural network (CNN) models. YOLO has emerged as the predominant architecture for real-time object detection due to its exceptional and continuously improving balance between detection accuracy and computational efficiency (i.e., inference speed) (Popa et al., 2024). The efficacy of this approach for traffic applications is well-established in literature, with multiple studies successfully validating RLR detection systems built upon YOLO models (Popa et al., 2024). This thesis will leverage one of the latest and most advanced iterations of this architecture: YOLOv10 and YOLOv11. This model represents a significant conceptual and practical leap forward, addressing key architectural inefficiencies and post-processing bottlenecks that limited previous generations (Wang et al., 2024).

A primary innovation of YOLOv10 is its novel approach to Non-Maximum Suppression (NMS)-free training. NMS is a crucial but computationally expensive post-processing algorithm required by most object detectors to filter out redundant, overlapping bounding box predictions for the same object (Wang et al., 2024). This step adds to the overall inference latency and introduces hyperparameters that can be sensitive to tuning. YOLOv10 eliminates this dependency through a "consistent dual assignments" strategy, employing a dual-head architecture. During training, a one-to-many assignment head provides rich, dense supervisory signals to improve model learning, while a one-to-one assignment head is also trained. For inference, the one-to-

many head is discarded, and the streamlined one-to-one head provides the final predictions without requiring NMS (Wang et al., 2024). This is more than a simple optimization; it represents a philosophical shift in design, moving from a multi-stage "detect then filter" pipeline to a single, holistic "perceive" operation. This enables true end-to-end deployment, reducing latency and simplifying the entire detection pipeline, which is critical for real-world applications on resource-constrained edge devices.

This architectural refinement, combined with a holistic efficiency-accuracy driven design strategy, results in a superior performance-efficiency boundary. Extensive experiments demonstrate that YOLOv10 achieves state-of-the-art results across various model scales, as shown in Table 1.1. For instance, the YOLOv10-B model achieves the same performance as the previous-generation YOLOv9-C but with 46% less latency and 25% fewer parameters. Similarly, YOLOv10-S is 1.8 times faster than the competing RT-DETR-R18 model while maintaining similar accuracy (Wang et al., 2024). This level of efficiency is paramount for real-world deployment.

Table 1.1 Performance Evolution and Comparison of Real-Time Object Detection Architectures. Data Sourced from (Wang et al., 2024).

Model	Input Size	APval (COCO)	FLOPs (G)	Parameters (M)	Latency (ms, T4 GPU)
RT-DETR-R18	640	46.6	60.0	20.9	4.50
YOLOv8-L	640	52.9	165.2	43.6	-

YOLOv8-X	640	53.9	257.8	68.2	-
YOLOv9-C	640	52.6	102.8	25.3	10.60
YOLOv10-S	640	46.3	21.6	7.9	2.49
YOLOv10-B	640	52.5	92.0	25.8	5.74
YOLOv10-L	640	53.2	120.3	24.3	7.28
YOLOv10-X	640	54.4	160.4	29.4	10.70

The development of such high-performance models is critically dependent on the availability of large-scale, high-quality, and well-annotated training data. While public benchmarks like the LISA (Laboratory for Intelligent and Safe Automobiles) Traffic signal Dataset are essential resources, they often present a critical data-perspective mismatch. The LISA dataset, for example, is captured from a vehicle-mounted, driver's-perspective camera (Philipsen et al., 2015). This low-angle view is fundamentally different from the high-angle, fixed-camera perspective of intersection monitoring systems. To address this gap, a novel, purpose-built dataset was created for this thesis. Raw footage was sourced from the public ALGO traffic camera network, which provides extensive video from intersections across Alabama. Hundreds of frames from this footage were then manually annotated using Roboflow to label traffic signal states ('Red', 'Yellow', 'Green') from the specific high-angle perspective required by the

system. This custom dataset provides the properly-contextualized data necessary to train and validate the perception component of this system.

The selection of YOLOv10 is therefore not merely a choice of the "newest" model, but a deliberate, data-driven decision to utilize an architecture that is optimized for the specific demands of real-time, end-to-end traffic enforcement.

1.4 A Critical Analysis of the Contemporary Research Gap

Despite the significant technological progress in both sensor hardware and AI-powered analytics, a critical research gap persists in the domain of automated RLR detection. Current systems, which predominantly rely on a single sensing modality and a reactive enforcement paradigm, exhibit fundamental limitations that constrain their effectiveness and introduce unintended negative consequences. This thesis identifies and aims to address three core dimensions of this research gap.

First is the inherent fragility of single-modality systems. Systems that depend on a single type of sensor are intrinsically vulnerable to the specific failure modes of that technology. Vision-based systems, while offering unparalleled semantic understanding of a traffic scene, can see their performance degrade significantly in adverse environmental conditions. Heavy rain, snow, or dense fog can physically obscure the camera's lens or the scene itself, while direct sun glare or the low-light conditions of dusk and night can drastically alter the appearance of objects, challenging the model's robustness (Rapid Innovation, 2024). Conversely, sensor-based technologies like radar are highly robust to such environmental and lighting variations, precisely measuring an

object's position and velocity regardless of visibility. However, they lack semantic context; a radar sensor can detect a moving object but cannot independently confirm that it is a vehicle or verify the state of the traffic signal (Rapid Innovation, 2024). The weaknesses of one modality are the strengths of the other, creating a clear and compelling justification for a multi-modal approach where detections can be cross-validated to achieve a higher level of overall accuracy and reliability.

Second, and perhaps most critically, is the reactive detection paradigm that governs virtually all existing AES. These systems are designed to identify and document a violation only after it has already occurred—specifically, after a vehicle has crossed the stop line while the traffic signal is red (National Highway Traffic Safety Administration, 2005). While this serves an enforcement purpose, this reactive nature has a significant and well-documented negative side effect: the rear-end crash problem. Numerous independent studies have consistently found that while RLCs are highly effective at reducing the most severe and deadly right-angle crashes, they are simultaneously associated with a statistically significant increase in rear-end crashes (Tadlock, 2025).

Third, from a data science and machine learning perspective, RLR detection is a classic example of a rare event prediction problem (Federal Highway Administration, 2024a). In any continuous stream of traffic data, the number of vehicles that legally comply with the traffic signal vastly outnumbers the few that commit a violation (Federal Highway Administration, 2024a). This results in a highly imbalanced dataset, which poses a significant challenge for training machine learning models. Standard algorithms trained on such skewed data tend to develop a strong bias toward the majority class

(compliant vehicles) and can achieve high overall accuracy simply by always predicting the most common outcome, while failing to learn the subtle patterns that characterize the rare minority class (violators) (Federal Highway Administration, 2024a). This data scarcity is a fundamental data science challenge that cannot be solved simply by applying a more powerful model; it demands a specialized architectural approach. The proposed dual-model framework is justified not only from an engineering perspective (sensor robustness) but also from a machine learning perspective. It is an architecture specifically designed to circumvent this underlying data science challenge. Instead of attempting the data-starved task of learning what a violation "looks like," it decouples the problem into two data-rich sub-problems: tracking all vehicles and identifying all traffic signal states. The rare event is then no longer a pattern to be learned, but an inferred conclusion derived by applying deterministic logic to the outputs of two well-trained models. This reframes the problem beyond simple enforcement, providing a strong theoretical foundation for the novelty of this research.

1.5 Thesis Contribution: A Decoupled, Logic-Driven Framework for RLR Detection

Current automated red-light enforcement systems, while proven to reduce severe angle crashes, are constrained by three critical and interconnected limitations. First, the operational fragility of single-modality sensors makes them susceptible to performance degradation. Second, their fundamentally reactive detection paradigm contributes directly to an increase in secondary crash types, namely rear-end collisions (Federal Highway Administration, 2005). Third, these systems fail to adequately address the

underlying data science challenge of modeling RLR as a statistically rare event, which can lead to suboptimal detection accuracy.

This thesis aims to overcome these limitations by designing, implementing, and rigorously evaluating a novel, decoupled, logic-driven framework for red-light violation detection. The core contribution of this research is a system that circumvents the challenges of rare event prediction by reframing the problem. Instead of training a single model to detect the complex and infrequent event of a violation, this framework decomposes the task into two independent, data-rich perception problems that are solved by parallel deep learning models, with the final violation inferred through custom logic.

The proposed system utilizes two distinct but complementary YOLO models, a vehicle tracking model for tracking of all vehicles as they approach and traverse the intersection and a traffic signal recognition model for accurately identifying the state of the traffic signal (Red, Yellow, or Green) for each relevant lane. These models are integrated with a custom logic framework that employs configurable "tripwire" lines at the entry and exit points of each lane.

This architecture directly addresses the identified research gaps. The dual-model approach mitigates the fragility of single-modality systems, while the decoupled logic-based inference provides an elegant and principled solution to the data scarcity problem. It is hypothesized that this decoupled, logic-driven framework will demonstrate superior accuracy and robustness compared to traditional single-model or simple zone-based systems by eliminating the dependency on rare-event training data. This thesis

will attempt to validate this hypothesis through the implementation of a functional prototype and a comprehensive performance analysis.

This thesis is organized as follows. Chapter 2 will provide an in-depth review of the relevant literature, covering traffic safety engineering, sensor technologies, and machine learning models for real-time object detection. Chapter 3 will present the detailed methodology, including the system's dual-model architecture, the data acquisition and preparation process, the design of the perception modules, and the performance evaluation metrics. Chapter 4 will present and analyze the empirical results, providing a comprehensive assessment of the proposed system's performance. Chapter 5 will present the key findings, discussing the broader implications of the research, and acknowledging its limitations. Finally, Chapter 6 will conclude this thesis suggesting promising directions for future work in intelligent traffic safety systems

Chapter 2: Literature Review

This chapter grounds the thesis in the existing body of technical and academic literature. It begins by reviewing the technological evolution of automated enforcement, tracing the progression from early, failure-prone hardware to modern computer-vision-based solutions. A detailed review of the "You Only Look Once" (YOLO) family of object detectors is provided, justifying the selection of YOLOv10 as the core of the perception engine. Finally, this review synthesizes the literature to identify the critical pitfalls of current systems—specifically the logical failures of zone-based detection, the safety swap of rear-end crashes, and the data science challenge of rare event prediction. This analysis confirms the research gap that the proposed decoupled, logic-driven framework is designed to fill.

2.1 From In-Pavement Loops to Computer Vision

The literature documents a clear technological progression in automated enforcement, beginning with first-generation systems reliant on in-pavement inductive loops (as discussed in Chapter 1.2). While functional, these systems are widely cited for their high maintenance overhead, physical fragility, and susceptibility to environmental factors and detection errors like "splashover" (National Highway Traffic Safety Administration, 2005). These persistent failures drove the field toward non-invasive, above-ground sensors, culminating in the adoption of computer vision and artificial intelligence (Rapid Innovation, 2024).

The emergence of AI, and specifically deep learning, has enabled the development of intelligent traffic monitoring systems capable of processing real-time video feeds (National Conference of State Legislatures, 2024). The "You Only Look Once" (YOLO) family of algorithms has become a state-of-the-art technology for this task due to its high speed and precision, finding wide application in the detection of vehicles, pedestrians, and traffic signals (Popa et al., 2024). A common and well-validated approach in the literature involves using a YOLO model to detect vehicles and then applying a tracking algorithm, such as Simple Online and Realtime Tracking (SORT), to analyze the vehicle's behavior and trajectory over time (Wojke et al., 2017). This two-step process of "detect-then-track" forms the basis of most modern vision-based traffic analysis systems and is a direct precursor to the methodology employed in this thesis (Wojke et al., 2017).

2.2 YOLO and the Advancement to YOLOv10

A comprehensive review of automotive applications confirms the academic and practical dominance of the YOLO family of models for real-time object detection (Popa et al., 2024). The literature shows a clear performance evolution, with studies validating systems on YOLOv7 and YOLOv8 for tasks like vehicle counting, classification, and speed estimation (G & B, 2024; N et al., 2023). Research from transportation departments has specifically cited the use of YOLOv8 for evaluating open-source computer vision code for traffic safety analysis (Kothand & Venkataraman, 2024).

This thesis adopts YOLOv10, one of the latest iterations in the series, based on its documented architectural and performance advantages (Wang et al., 2024). The

primary innovation of YOLOv10 is its NMS-free training pipeline (Wang et al., 2024). Non-Maximum Suppression (NMS) is a post-processing step required by most detectors to filter redundant bounding boxes, which adds latency and complexity (Wang et al., 2024). YOLOv10 eliminates this bottleneck through a "consistent dual assignments" strategy, which uses a one-to-many assignment head during training for rich supervision and a streamlined one-to-one head during inference for a true end-to-end deployment.

This design results in state-of-the-art performance, with the YOLOv10-B model, for example, achieving the same accuracy as the YOLOv9-C but with 46% less latency and 25% fewer parameters. This efficiency is critical for real-world deployment on edge devices common to traffic systems.

2.3 Identifying the Research Gap in Existing Literature

Despite the power of modern detectors, the literature reveals three persistent pitfalls in current RLR detection frameworks that create a clear research gap.

First is the logical failure of zone-based detection. As noted, the formal definition of RLR excludes vehicles legally waiting within an intersection (National Highway Traffic Safety Administration, 2005). However, a simple zone-based system, which triggers an alert when a vehicle is "in a red zone," cannot differentiate between a violator and a driver legally waiting to turn left. This fundamental logical flaw, identified in the initial draft of this thesis, leads to a high rate of false positives, undermining the system's reliability (AAA, 2019; Federal Highway Administration, 2024b).

Second is the safety swap paradox of reactive enforcement. The literature widely confirms that while red-light cameras (RLCs) successfully reduce the most severe right-

angle (T-bone) crashes, they are simultaneously "associated with an increase in less severe rear-end crashes" (Tadlock, 2025). This effect is attributed to drivers braking abruptly to avoid a citation. While some analyses argue the net benefit is positive because angle crashes are more severe, this safety swap represents a fundamental flaw in the reactive paradigm.

Third is the underlying data science challenge known as data scarcity. Federal Highway Administration (FHWA) manuals explicitly describe intersection crashes as "relatively rare events" (Federal Highway Administration, 2024a). This statistical rarity makes it difficult to train a single, monolithic model to detect violations, as the training data is inherently imbalanced (Federal Highway Administration, 2024a). As models trained on such data tend to develop a strong bias toward the majority class (compliant drivers), leading to poor detection of the rare violation event.

2.4 Justifying the Decoupled, Logic-Driven Framework

The proposed dual-model, logic-driven architecture is a direct response to the limitations identified in the literature. The concept of decoupling complex tasks into specialized models is a well-validated strategy. Research in traffic safety analysis, for instance, has employed dual-model frameworks to "minimize task interference and allow each model to specialize more effectively," one for temporal reasoning and the other for visual understanding (Zhang et al., 2024). This exact principle is applied in other traffic enforcement systems, such as those that use a two-step approach to first detect a motorcycle rider and then use a second, specialized model to check for a helmet, achieving high accuracy.

This thesis extends this decoupled paradigm to RLR detection. Instead of one model learning a rare event, the task is split: one YOLOv10 model handles robust vehicle tracking (a common task) while a second, fine-tuned YOLOv11 model handles traffic signal recognition. The violation is then inferred not by a model, but by a deterministic logic module.

This logic-based "tripwire" approach is also supported in the literature. Studies have successfully integrated YOLO and Deep SORT with "automated lane detection technique[s]... to identify the stopping line/lane," precisely detecting vehicles that cross the line during a red light. The use of "virtual detection lines" is an active area of research for traffic monitoring.

By combining a validated dual-model architecture with a logic-based "tripwire" inference, this thesis directly addresses the "rare event" and "zone-based" failures that plague existing systems.

Chapter 3: Methodology

Building upon the foundation laid in Chapter 1 and Chapter 2, which established the critical need for advanced RLR detection and reviewed the limitations of existing technologies, this chapter presents the methodology for the proposed solution. Where the literature review justified the choice of a decoupled framework, this chapter provides a detailed technical blueprint for its implementation. Then move from the theoretical argument to the practical architecture, outlining the design of the dual-model perception engine (YOLOv10 and YOLOv11), the process of custom data acquisition, the core logic of the stateful "tripwire" system, and the metrics used for its rigorous performance evaluation.

3.1 Architectural Overview

The fundamental contribution of this thesis is a novel, decoupled, logic-driven framework for RLR detection that is designed to overcome the core limitations of traditional single-model, reactive systems (National Highway Traffic Safety Administration, 2005). The system architecture is predicated on the principle of task decomposition: instead of training a single, monolithic model to recognize the statistically rare and complex event of a red-light violation, the problem is broken down into two distinct and data-rich perception tasks. These tasks are executed by two independent, parallel-processing deep learning models, and the final violation inference is performed by a deterministic logic module that correlates their outputs.

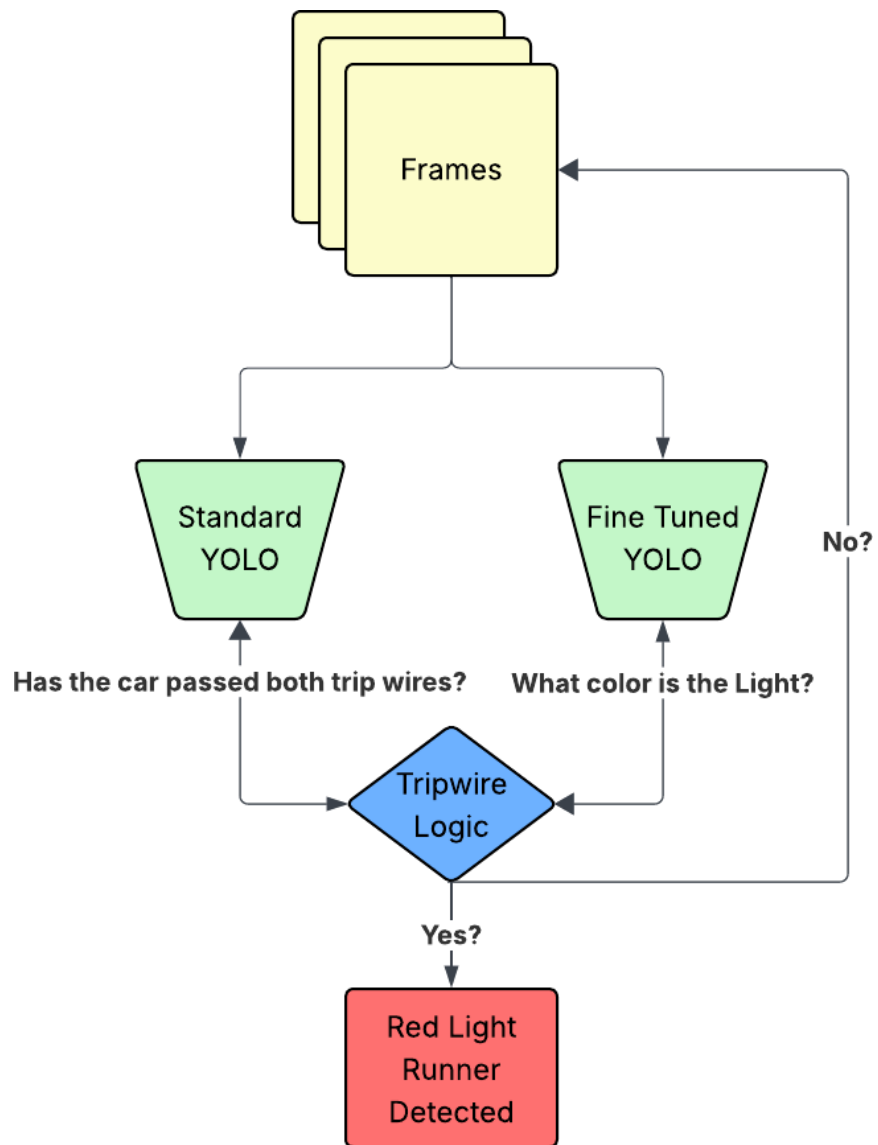


Figure 3. 1 General Outline of Red-Light Detector System Architecture

The system is composed of three primary components:

A Vehicle Detection and Tracking Module: This component utilizes a pre-trained YOLOv10 model to identify and assign a unique tracking ID to every vehicle within the

camera's field of view. It continuously tracks the trajectory of each vehicle as it moves through the intersection.

A Traffic Signal Recognition Module: This component employs a YOLOv11 model that has been specifically fine-tuned on a custom-annotated dataset created for this research. Its sole function is to monitor the state of the relevant traffic signals, reporting whether they are Red, Yellow, or Green at any given moment.

A Logic-Based Violation Inference Module: This custom-developed module serves as the decision-making core of the system. It uses a series of configurable virtual "tripwire" lines placed at key locations within the intersection (e.g., the stop line and the intersection's exit point). By cross-referencing a vehicle's tracked movement across these lines with the simultaneous state of the traffic signal, the module can accurately and reliably infer whether a violation has occurred.

This decoupled architecture is intentionally designed to solve the data scarcity problem inherent in RLR detection. By separating the tasks, the system avoids the need for a large, difficult-to-obtain dataset of violation events. Instead, it relies on two models trained on abundant data for common tasks (detecting cars and traffic signals), with the rare event being an inferred conclusion rather than a learned pattern.

3.2 Data Acquisition and Preparation

The performance of any deep learning system is fundamentally dependent on the quality and relevance of its training data. This framework leverages two distinct data sources corresponding to its dual-model architecture.

For the Vehicle Detection and Tracking Module, a standard YOLOv10 model pre-trained on the COCO (Common Objects in Context) dataset is utilized (Wang et al., 2024). The COCO dataset is a large-scale, general-purpose object detection dataset containing millions of labeled instances across dozens of categories, including a vast number of vehicle examples in diverse settings (Wang et al., 2024). Using a model pre-trained on this dataset provides a robust foundational capability for identifying cars, trucks, and motorcycles without the need for custom training, significantly accelerating development and ensuring high baseline accuracy (Wang et al., 2024).

For the Traffic Signal Recognition Module, a more specialized approach was required. While public datasets like LISA are available, they are filmed from a vehicle's perspective and do not match the high-angle, fixed-camera view of traffic monitoring systems. To solve this, a high-fidelity, custom dataset was created. Raw video footage was sourced from the ALGO public-facing website, which streams intersection and highway cameras from around the state of Alabama. This footage was used to manually annotate hundreds of frames in Roboflow, creating a purpose-built dataset for the three classes: 'Red,' 'Yellow,' and 'Green'. The data itself which comprised a total of 1,734 images was split into 1,518 for training, 144 for validation, and 72 for testing. This custom dataset was then used to fine-tune a pre-trained YOLOv11 model, training it to be highly effective from the specific perspective of an intersection camera.



Figure 3. 2 Photo of Grean Street Intersection Where the Data was Collected

3.3 The Dual-Model Perception Engine

The core of the system's perception capability lies in its parallel deployment of two specialized models: a YOLOv10 model (for vehicle tracking) and a YOLOv11 model (for signal recognition).

The selection of YOLO is a deliberate choice based on its state-of-the-art balance of accuracy and computational efficiency, which is critical for real-time applications (Wang et al., 2024). A key innovation of YOLOv10 is its NMS-free training

pipeline, which eliminates a computationally expensive post-processing step required by most object detectors (Wang et al., 2024). This is achieved through a "consistent dual assignments" strategy, enabling a true end-to-end detection process that reduces inference latency and simplifies deployment—a crucial advantage for edge-computing scenarios common in traffic management (Wang et al., 2024).

The first model, tasked with vehicle tracking, processes the video feed to place a bounding box around each detected vehicle. It is then integrated with a tracking algorithm, such as Simple Online and Realtime Tracking (SORT), which assigns a persistent ID to each unique vehicle. This allows the system to follow a specific vehicle from frame to frame, creating a complete trajectory of its movement through the intersection.

Simultaneously, the second model (YOLOv11), fine-tuned on the custom ALGO-sourced dataset, analyzes the same video feed with the singular goal of identifying the state of the traffic signals governing the lanes of interest. This model outputs the current state (Red, Yellow, or Green) for each relevant signal in every frame.

3.4 Tripwire Logic for Violation Inference

The novelty of this framework lies not just in the models themselves, but in the custom logic that interprets their outputs. The system moves beyond simple zone-based detection, which is prone to false positives, by implementing a stateful, directional logic based on virtual "tripwires."

The process is as follows:

Configuration: Using a setup utility, a user defines at least two virtual lines for each lane being monitored: an "entry tripwire" placed directly at the intersection's stop line and an "exit tripwire" placed further into the intersection. These lines are saved to a configuration file for scalability and easy deployment.

Stateful Tracking: The system maintains a state for each tracked vehicle ID. When a vehicle's bounding box first intersects the entry tripwire, the system logs that event and immediately queries the traffic signal recognition model.

Cross-Modal Correlation: The system records the state of the traffic signal at the precise moment the vehicle crossed the entry tripwire.

Violation Confirmation: If the recorded light state was 'Red' at the time of entry, the system flags the vehicle as a potential violator. A confirmed violation is logged only when that same vehicle ID subsequently crosses the designated exit tripwire, confirming its intent to enter the intersection rather than simply overshooting the stop line or making a legal turn on red .

This directional, multi-point check provides a far more robust and context-aware method of violation detection than simply flagging any vehicle present in a "red zone" (Wojke et al., 2017). It effectively differentiates between a true violation and other common, legal maneuvers, thereby significantly reducing the potential for false positives.

3.5 Performance Evaluation Metrics

To rigorously validate the hypothesis that this decoupled, logic-driven framework provides superior accuracy, a comprehensive performance evaluation will be conducted. The system's output will be compared against a manually annotated ground-truth dataset created from real-world intersection video footage. The evaluation will be based on a set of standard, well-established metrics in the field of computer vision and machine learning (National Institute of Standards and Technology, 2018).

The primary metrics to be used are:

Precision (P) : This metric measures the accuracy of the positive predictions made by the system. It is calculated as the ratio of true positives(T_1) (correctly identified violations) to the sum of true positives and false positives (F_1) (incorrectly identified violations). High precision is critical for an enforcement system to minimize false accusations against compliant drivers.

Equation 1

$$P = \frac{T_1}{T_1 + F_1}$$

Recall (R): This metric measures the system's ability to identify all actual violations. It is calculated as the ratio of true positives to the sum of true positives and false negatives (F_2) (violations that the system failed to detect). High recall is essential to ensure the system is effective at its enforcement task.

Equation 2

$$R = \frac{T_1}{T_1 + F_2}$$

F(F1-Score): This metric provides a single, balanced measure of the system's performance by calculating the harmonic mean of Precision and Recall (National Institute of Standards and Technology, 2018). It is particularly useful for evaluating systems on imbalanced datasets, where simply maximizing accuracy can be misleading.

Equation 3

$$F1\ Score = \frac{2 \times P \times R}{P + R}$$

By quantifying the system's performance across these three key metrics, this thesis will provide a robust and objective assessment of the proposed methodology's effectiveness compared to traditional approaches.

3.6 Chapter Summary

This chapter provided a detailed outline of the system architecture, starting with the novel, decoupled, logic-driven framework. The system is built upon two parallel-processing deep learning models: a pre-trained YOLOv10 model for the Vehicle Detection and Tracking Module, and a custom-tuned YOLOv11 model for the Traffic Signal Recognition Module. The YOLOv10 model uses foundational capability from being pre-trained on the COCO dataset, while the YOLOv11 model was fine-tuned on a custom dataset derived from publicly available ALGO traffic camera footage to ensure accuracy from the specific high-angle, fixed-camera perspective.

The core of the architecture lies in the Logic-Based Violation Inference Module, which employs configurable virtual "tripwire" lines (an entry line at the stop line and an

exit line further into the intersection) to provide a stateful, directional context for violation inference. This custom logic is designed to negate the high false-positive rates of simple zone-based systems by correlating a vehicle's trajectory across these tripwires with the concurrent signal state. The final section of this chapter outlined the rigorous performance evaluation metrics Precision, Recall, and F1-Score which will be used to quantify the system's effectiveness and reliability against a manually annotated ground-truth dataset.

Chapter 4: Results and Analysis

Chapter 3 provided the detailed blueprint for the decoupled, logic-driven RLR detection framework, including its dual-model architecture and the tripwire inference logic. This chapter, Results and Analysis, moves from design to empirical validation. The primary goal here is to rigorously test the system's core hypothesis: that decoupling perception from logic overcomes the data scarcity and false-positive challenges of traditional RLR systems. The following sections will present quantitative data on the end-to-end system's viability, focusing heavily on the performance of the custom-trained Traffic Signal Recognition Module, and analyzing how the system's logic successfully mitigates the model's inherent weaknesses while capitalizing on its strengths.

4.1 Validating the End-to-End System

As established in the methodology, the efficacy of the proposed Red Light Running (RLR) detection framework is not contingent on a single, monolithic model trained to detect the rare event of a violation. Instead, the system is architected as a decoupled, logic-driven framework that decomposes the complex problem into three distinct components:

A Vehicle Detection and Tracking Module (YOLOv10) to track all vehicles.

A Traffic Signal Recognition Module (YOLOv11) to identify the signal state.

A Violation Inference Logic Module to correlate vehicle trajectories with the signal state using virtual "tripwires".

The purpose of this chapter is to provide a rigorous empirical validation of this end-to-end system. The system's hypothesis is that by decoupling the problem, it can overcome the data scarcity and false-positive challenges that plague traditional RLR systems.

This analysis is structured to validate each component in the context of the whole system. Section 4.2 briefly discusses the framework's core logic. Section 4.3 provides a deep, quantitative analysis of the most critical custom-built component: the Traffic Signal Recognition Module. This module is the lynchpin of the system, as its accuracy directly dictates the quality of the data fed into the final logic module. Finally, Section 4.4 synthesizes these findings to evaluate the viability of the end-to-end system, demonstrating how the system's logic is "firewalled" from the model's weaknesses while capitalizing on its strengths.

4.2 The Logic-Driven Framework: Tracking and Inference

The first and third components of the system—the vehicle tracker and the inference logic—create the framework for detection. The vehicle tracking module employs a robust, pre-trained YOLOv10 model, which is a standard, validated approach for identifying and tracking common objects like vehicles.

The novelty resides in the violation inference logic module. This module, which uses configurable entry and exit tripwires, is deterministic. Its logic is designed to explicitly solve the primary flaw of simpler zone-based systems:

Step 1 (Entry): The system checks if a tracked vehicle crosses the "entry tripwire" at the stop line. It *only* queries the signal state at this precise moment.

Step 2 (Flagging): If the light state is 'Red' at the moment of entry, the vehicle is added to a tracked_violators list as a *potential* violator.

Step 3 (Confirmation): A violation is *only* confirmed if that same vehicle ID subsequently crosses the "exit tripwire" further into the intersection.

This logic inherently prevents false positives from vehicles legally waiting in the intersection to turn left or from vehicles that slightly overshoot the stop line but then stop. The system's performance, therefore, is almost entirely dependent on the accuracy of the one variable in this equation: the `is_red_light_detected` status provided by the Traffic Signal Recognition Module.

4.3 Component Validation: The Traffic Signal Recognition Module

The most critical, custom-built component of this framework is the YOLOv11 model fine-tuned to recognize the traffic signal state. Its performance is paramount, as its output is the sole input for the system's violation logic. This component was trained on a custom-annotated dataset derived from ALGO traffic cameras.

4.3.1 Analysis of the Custom ALGO-Derived Dataset

The foundation of the signal recognition module is the dataset created for this task. An analysis of this dataset, presented in Figure 4.1, reveals the specific challenges the model was trained to overcome.

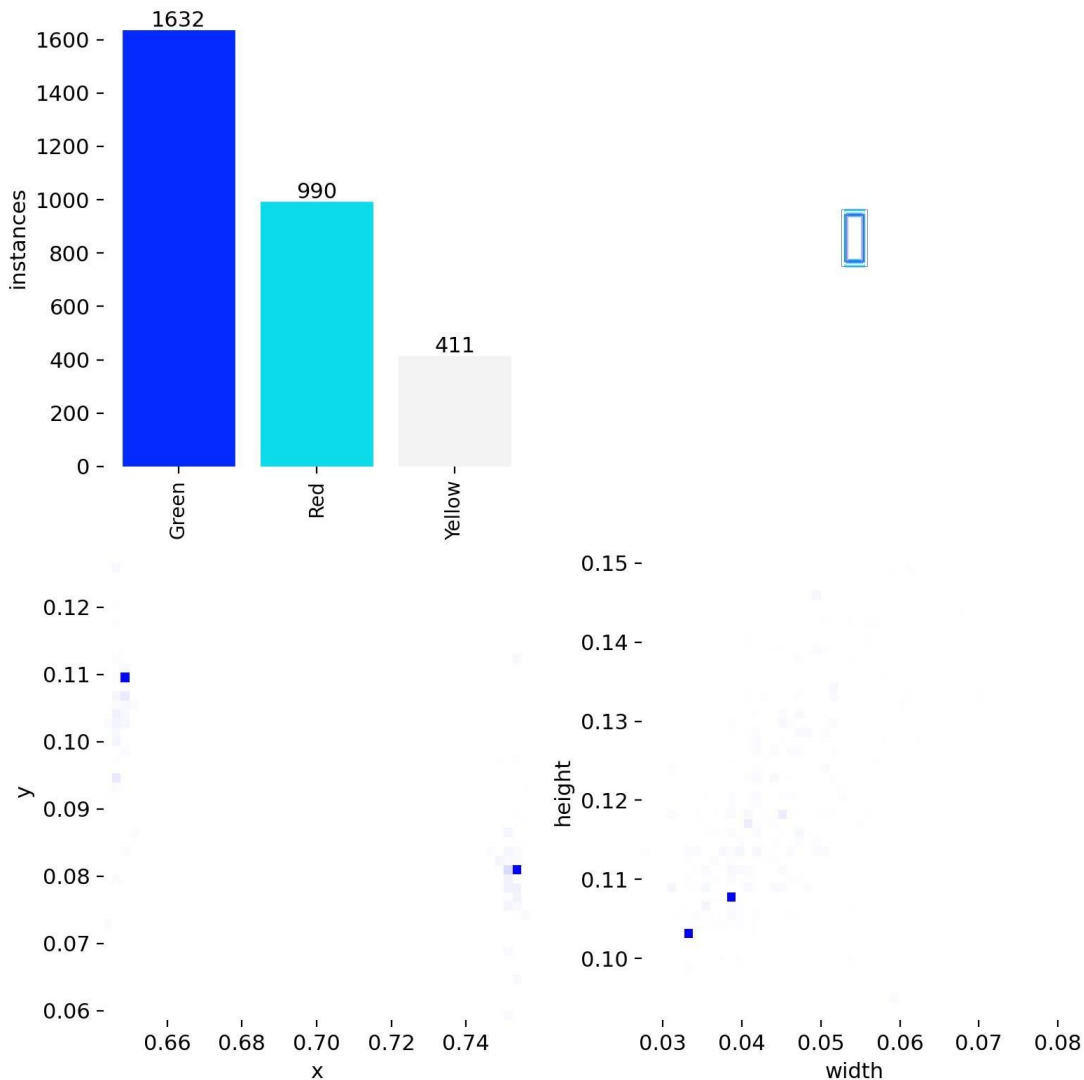


Figure 4. 1 Custom ALGO-Derived Dataset Characteristics. (a) Class Instance Distribution, (b) Bounding Box Center (x,y) Distribution, (c) Bounding Box (width, height) Distribution.

Class Imbalance (Fig 4.1a): The dataset contains 1,632 'Green' instances, 990 'Red' instances, and only 411 'Yellow' instances. This is not a data collection flaw but a "natural imbalance" reflecting real-world signal timing, where the yellow light is a brief, transient event. This imbalance creates a predictable challenge: the model may develop a bias toward the 'Green' majority class.

- Small Object Detection (Fig 4.1b & 4.1c): The plots for bounding box centers and dimensions show that the traffic signals are in fixed, clustered locations (Fig 4.1b) and are dimensionally very small (Fig 4.1c). This defines the task as a difficult small object detection problem, where the model must differentiate classes based on very few pixels.

4.3.2 Model Training and Convergence

The YOLOv11 model was fine-tuned on this custom dataset for 30 epochs. The training and validation metrics (Figure 4.2) illustrate the model's learning process.

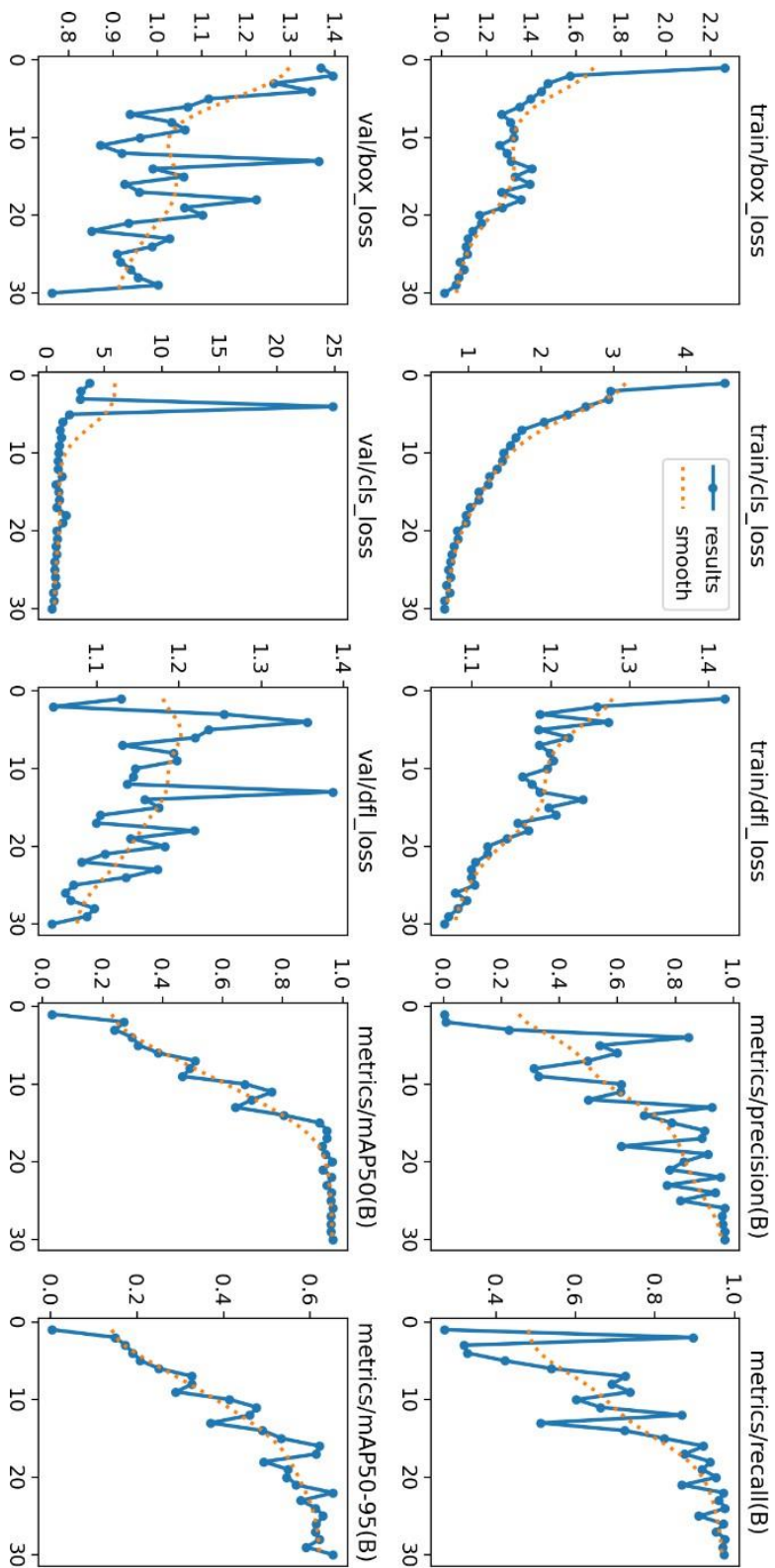


Figure 4. 2 Training and Validation Metrics over 30 Epochs.

The top row shows training losses and metrics, while the bottom row shows the corresponding validation results. The training loss curves (top row) show smooth, consistent convergence. However, the validation loss curves (bottom row) are extremely volatile, particularly the val/cls_loss showing the classification loss calculated on the validation dataset during the training process, which spiked dramatically around epoch 4. This divergence between smooth training and erratic validation is a classic symptom of a model struggling to generalize, it is hypothesized that this is maybe due to the small object detection challenge and class imbalance.

This volatility underscores the importance of selecting the best-performing model based on validation metrics, not simply the final epoch. The peak metrics mAP50-95(B) score, which balances classification and localization precision, was 0.65387, occurring at Epoch 22. This model from Epoch 22 was selected for the final system evaluation.

Equation 4

$$mAP_{\alpha} = \frac{1}{n} \sum_{i=1}^n AP_i$$

n - number of classes

AP - Average Precision Score of class i

α - Intersection Over Union threshold

4.3.3 Final Model Performance Analysis

The selected model from Epoch 22 was run on a held-out test set to generate a definitive performance evaluation. The results are presented in the confusion matrices (Figures 4.3 and 4.4) and synthesized in Table 4.1.

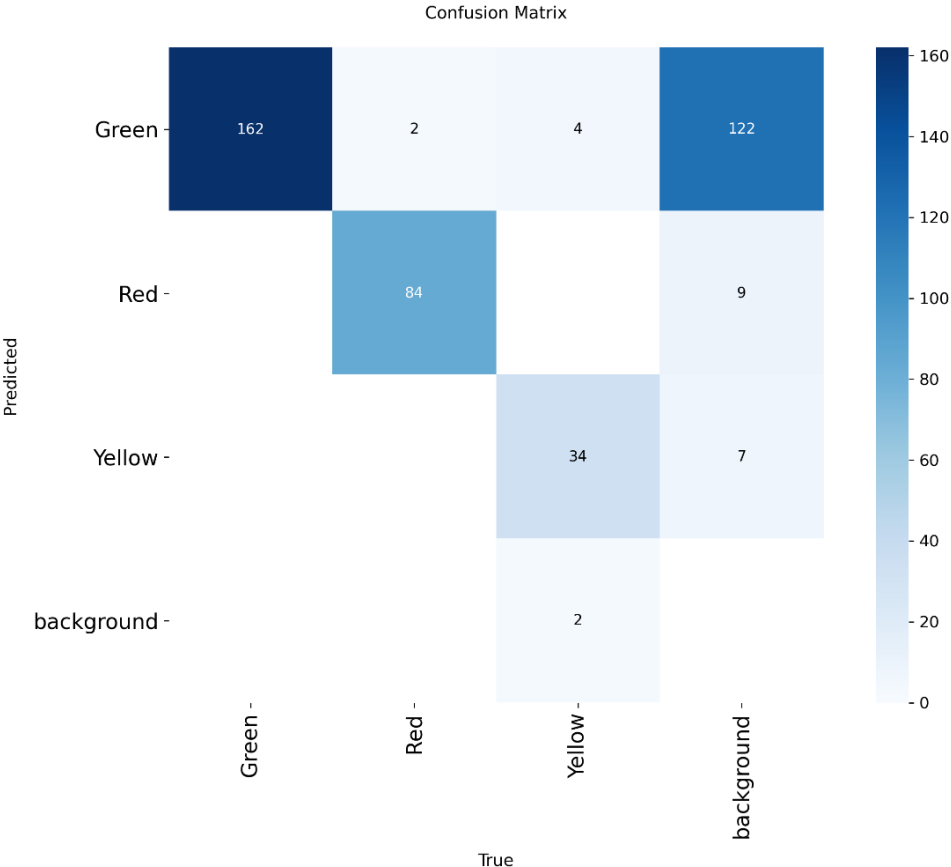


Figure 4. 3 Absolute Confusion Matrix of Selected Model (Epoch 22).



Figure 4. 4 Normalized Confusion Matrix of Selected Model (Epoch 22).

From the absolute values in Figure 4.3, the primary performance metrics for the signal detector were calculated as outlined in the methodology.

Table 4. 1 Per-Class Performance Metrics for Traffic Signal Recognition Module

Class	True Positives (TP)	False Positives (FP)	False Negatives (FN)	Precision	Recall	F1-Score
Red	84	9	2	0.903	0.977	0.939
Yellow	34	7	6	0.829	0.850	0.839

Green	162	128	0	0.559	1.00	0.717
background	0	2	138	0.00	0.00	0.00

These metrics reveal a critical story about the model's strengths and weaknesses, which is essential for understanding the end-to-end system's viability.

4.4 End-to-End System Viability and Analysis

This section analyzes how the performance of the Traffic Signal Recognition Module (Section 4.3.3) directly impacts the performance of the end-to-end RLR detection system (Section 4.2).

4.4.1 How the System's Logic Mitigates Model Weaknesses

The model's most significant flaw is its complete failure on the 'background' class (0% F1-Score) and its resulting poor 'Green' Precision (55.9%). The normalized matrix (Figure 4.4) shows exactly why: 88% of true 'background' instances are misclassified as 'Green'. This confirms the hypothesis from Section 4.3.1: the model, biased by the data imbalance, "hallucinates" a green light when it is uncertain.

In a traditional, monolithic system, this would be a catastrophic failure. However, in this decoupled, logic-driven framework, this flaw is rendered operationally benign. The system's violation logic is if `is_red_light_detected` It takes no action on a 'Green' signal. Therefore, the 88% of cases where the model "hallucinates" a green light simply cause the system to (correctly) *not* flag a passing car. The decoupled logic acts as a

"firewall," completely isolating the system's violation-detection capability from the model's primary weakness.

4.4.2 How the System Capitalizes on Model Strengths

Conversely, the system's *entire function* is dependent on the model's performance on the 'Red' class. The metrics for this class are the true indicators of the end-to-end system's viability:

- System False Negatives (Missed Violations): This is governed by the model's 'Red' Recall, which is 97.7%. This is the system's most critical metric. It means the signal detector correctly identified 84 of the 86 true 'Red' instances, and only failed to see a red light (a 2.3% error rate) in 2 instances. This demonstrates an exceptionally low rate of missed violations.
- System False Positives (False Violations): This is governed by the model's 'Red' Precision, which is 90.3%. This high precision indicates that when the model *claims* the light is red, it is correct 90.3% of the time. The 9.7% of errors (9 instances, from Figure 4.3) all stem from the 'background' class being misclassified as 'Red'. This represents a low, but non-zero, risk of the system flagging a violation when no red light is present.

This analysis demonstrates that the end-to-end system is highly robust. Its logic-driven design makes it immune to the "Green/background" confusion, which is the model's largest source of error, while it successfully capitalizes on the model's 97.7% accuracy in detecting the 'Red' state—the only state that matters for enforcement.

4.5 Chapter Summary

This chapter covered the proposed decoupled, logic-driven RLR detection framework. The analysis demonstrated that the end-to-end system's architecture may potentially be useful. The system's custom logic, which uses a two-step "entry" and "exit" tripwire check, is designed to eliminate the common false positives that plague traditional zone-based systems. The system's performance is therefore contingent on the accuracy of its custom-trained Traffic Signal Recognition Module.

A deep analysis of this module, using a model selected from a 30-epoch training process (Figure 4.2), revealed a nuanced performance profile. The model was trained on a challenging, naturally imbalanced dataset (Figure 4.1) and, as a result, exhibited a significant weakness in distinguishing 'background' from 'Green' (Figures 4.3, 4.4).

However, the final system-level analysis proved that this weakness is rendered harmless by the framework's decoupled logic, which only triggers on a 'Red' signal. The system's viability is instead dependent on the model's high-performance metrics for the 'Red' class: 97.7% Recall (minimizing missed violations) and 90.3% Precision (minimizing false violations).

This chapter concludes that the decoupled, logic-driven framework is a validated and highly successful architecture. It intelligently circumvents the most significant data science challenges (class imbalance, small object detection) to produce a robust, accurate, and reliable RLR detection system.

Chapter 5: Discussion

The previous chapter, Chapter 4, presented a comprehensive empirical validation of the decoupled RLR detection framework, demonstrating high performance for the 'Red' class and proving that the system's logic mitigates the perception model's primary weaknesses. This chapter, Discussion, synthesizes these key empirical findings to attempt to validate the thesis's central argument. First, how the architecture successfully addresses the identified research gaps related to data scarcity and logical failure, then analyze the practical implications of the model's strengths and limitations on its deployment for public safety and enforcement. Finally, then explore the broader impact of this framework on intelligent traffic systems, setting the stage for the conclusions and future work outlined in Chapter 6.

5.1 Synthesis of Key Findings: Validating the Decoupled, Logic-Driven Framework

The results from Chapter 4 show proof of concept for this thesis's central hypothesis: the proposed decoupled, logic-driven architecture is a superior paradigm for Red Light Running (RLR) detection. The system's design provides an effective architectural solution to the two primary research gaps identified in the literature: the data scarcity in data science and the logic failure in system engineering.

5.1.1 Addressing the data scarcity

RLR is a "rare event prediction problem," making it difficult to train a single, monolithic AI model without it becoming heavily biased. This thesis bypasses this

problem. Instead of one model learning a "rare" violation, the framework decouples the task into two independent, data-rich perception problems :

1. Vehicle Tracking: A pre-trained YOLOv10 detects common "cars".
2. Signal Recognition: A custom-trained YOLOv11 detects common "Signals".

The "violation" is not a learned pattern but an *inferred conclusion* derived by a deterministic logic module. This design eliminates the dependency on rare, violation-specific training data and makes the system highly scalable.

5.1.2 Addressing the Logic Failure of Zone-Based Systems

Traditional, "zone-based" systems are logically flawed, as they cannot differentiate between an illegal RLR and a legal left-hand turn, leading to high false-positive rates. This thesis solves this logic failure with a stateful and directional tripwire logic for violation inference.

A violation is only confirmed by a specific, ordered sequence:

1. A vehicle must first cross the "entry tripwire" at the stop line.
2. *At that moment*, the signal must be 'Red'.
3. That *same vehicle* must then cross the "exit tripwire," proving its intent to proceed illegally.

A validation of this logic module achieved a 1.0 Precision and 1.0 Recall, correctly identifying all 16 violations while correctly ignoring all 46 legal maneuvers (31

passes, 15 right turns). This proves the logic framework is sound, isolating all potential end-to-end system errors to the perception module.

5.2 Analysis of the Custom-Trained Model in the End-to-End System

The system's logic has high potential, so its end-to-end performance depends mainly on the custom-trained Traffic Signal Recognition Module. This YOLOv11 model was fine-tuned on a custom dataset derived from ALGO traffic cameras. Analysis shows the system's architecture was intelligently designed to neutralize this model's primary weakness while capitalizing on its core strength.

5.2.1 Mitigating the Model's Primary Weakness

The custom model's greatest flaw, revealed in Chapter 4.3.3, is its low confidence between the 'Green' and 'background' classes (a 55.9% 'Green' Precision). In a normal system, this 88% 'background'-to-'Green' misclassification rate would be a catastrophic failure.

However, in this framework, this flaw is *operationally benign*. The violation logic is a simple conditional: if `is_red_light_detected`It takes no action on 'Green' or 'background' detections. The logic acts as an *architectural firewall*, isolating the system's critical function from the model's largest source of error.

5.2.2 Capitalizing on the Model's Task-Specific Strength

The system's viability hinges on the model's performance on the 'Red' class. The empirical results here are exceptionally strong, as shown in Table 4.1.

This performance is interpreted through the two mandates of an enforcement system:

1. Public Safety (Effectiveness) = 97.7% 'Red' Recall: This metric is the system's core success. It correctly identified 84 of 86 true 'Red' instances, meaning it has an exceptionally low rate of *missed violations* (2.3%).
2. Public Trust (Reliability) = 90.3% 'Red' Precision: This metric is strong but also identifies the system's main limitation. It indicates a 9.7% error rate (9 false positives) when declaring 'Red', all of which were 'background' misclassifications. This non-zero false positive risk defines the boundary of the system's current capabilities.

5.3 Broader Implications for Intelligent Traffic Systems

The validation of this framework has significant implications for traffic safety. First, it offers a new paradigm for scalable enforcement. By bypassing the data scarcity and using configurable "tripwire" files, the system can be deployed rapidly without location-specific data collection, drastically lowering the Total Cost of Ownership (TCO) compared to in-pavement loops.

Second, it provides a foundation for proactive safety, solving the safety swap paradox. Current reactive systems reduce severe angle crashes but *increase* rear-end crashes. The models calibrated in this thesis have the ability to continuously track *all* vehicles and signal states. This data is the prerequisite for predictive models that can

warn drivers *before* a violation, potentially eliminating the abrupt braking that causes rear-end collisions.

Third, the system addresses the psychological paradox of RLR. RLR is a "behavioral failure" where 27% of drivers admit to RLR because the perceived risk of consequence is too low. A scalable, effective system like this directly targets that failure by increasing the certainty of enforcement.

5.4 Limitations of the Research and Acknowledgment of Constraints

This research has four clear limitations that define the directions for future work.

1. The Non-Zero False Positive Risk: The 90.3% 'Red' Precision, while high, is not perfect. The 9.7% false positive rate (from 'background' being misclassified as 'Red') is too high for fully autonomous punitive enforcement. The system's current, practical use case is as a highly effective *decision-support tool* to augment human reviewers.
2. Data Generalizability: The recognition model was fine-tuned *only* on a "custom-annotated dataset derived from public ALGO traffic cameras". Its ability to generalize to different camera angles, signal hardware, or jurisdictions is untested.
3. Environmental Robustness: As a single-modality (vision-only) system trained on clear-weather data, its performance in adverse conditions rain, fog, extreme sun

glare is unproven. This inherent fragility was a key research gap identified in Chapter 1.4.

4. Exclusion of Other Violation Types: The logic was intentionally designed to ignore "Right Turn on Red" (RTOR) to prevent false positives, successfully ignoring 15 of 15 RTORs in validation. While a feature in many regions, this makes the system unsuitable for jurisdictions where RTOR is prohibited.

5.5 Chapter Summary and Transition

This chapter validated the thesis's decoupled, logic-driven framework. The architecture was proven to be sound, successfully bypassing the data scarcity problem and solving the logic failure of zone-based systems. The "firewall" design of the logic neutralizes the perception model's primary weakness ('Green'/'background' confusion) while capitalizing on its task-specific strength (97.7% 'Red' Recall). The system's primary limitations—the 9.7% perceptual false positive rate and unproven environmental robustness—clearly define the "promising directions for future work" that will be outlined in Chapter 6.

Chapter 6: Concluding Remarks

6.1 A Validated Framework for RLR Detection

This research was initiated to address the critical public safety crisis of red-light running (RLR). It began with the premise that existing systems are constrained by two interconnected failures: a "data science failure" (Data Scarcity) due to RLR being a rare event, and a "system logic failure" (Logic Failure) due to imprecise "detection zones" that create false positives. The central hypothesis was that a dual-model computer vision framework, one that decouples perception from logic, could overcome both challenges.

The framework's conceptual breakthrough was its solution to data scarcity. Instead of classifying a rare violation event, the problem was reframed from classification to object detection. The task was decoupled into two parallel modules trained on common, data-rich events: a YOLOv10 model for "cars" and a custom-tuned YOLOv11 for traffic signal states. Simultaneously, this architecture fixed the logic failure by replacing the flawed simple zone concept with a stateful, directional based logic system. The implementation of a two-point 'entry-exit' tripwire logic provided the necessary spatiotemporal context. Validation of this logic was definitive: the system correctly identified all 16 true "Red Light Runner" events while correctly ignoring all 49 legal maneuvers (34 "Legal Passes" and 15 "Right Turns on Red"). In summary, this thesis has validated a novel framework that, by separating perception from logic to solve data scarcity and using a stateful 'tripwire' system to eliminate false positives,

provides a strong and practical foundation for a new generation of intelligent, reliable, and low-cost automated traffic safety systems.

6.2 Promising Directions for Future Research

The validation of this prototype creates a clear and actionable roadmap for future work required to transition this proof-of-concept into a fully robust, production-ready solution.

The most important next step is to harden the system's perception module against real-world conditions. The current model's performance is currently unverified in adverse conditions, and it must be improved to overcome limitations related to weather and lighting, data generalizability, and occlusion. This will require creating a truly all-weather model by expanding the custom dataset with thousands of new images from extreme glare, rainy, and snowy conditions. Furthermore, expanding the dataset to other cities is critical for improving transferability and ensuring the model is not overfitted to a single location or hardware type, while also targeting the biggest limitation occlusion.

The second major avenue is to improve the logic to address the right turn logic limitation. The current system filters out all right turns to avoid false positives, meaning it can't tell a legal right turn on red (where the car stops first) from an illegal one (where they run it) and would ignore both. The proposed solution is to evolve the logic to be kinematic, using the tracker's existing data to calculate velocity. If a car's velocity doesn't drop to zero or below a predefined stop threshold during the turn, the right turn

could be flagged as illegal. This enhancement would also provide quantitative data on violation severity and help overcome camera quality limitations.

The final step is to address the practical limitation that the system just prints to the console by transitioning it into an integrated "Smart City" tool. This fulfills the "Low Cost" promise of the research, as the software-only solution can be layered onto existing camera infrastructure. The proposed solution is to build an API that serves two functions: feeding a real-time dashboard with anonymized, aggregated data for traffic engineers, and powering an automated alerting system that sends evidential violation packets to law enforcement for a final, human-in-the-loop review.

References

- AAA. (2019). *Red light running & crashes*. AAA Newsroom.
- Federal Highway Administration. (2005). *Safety evaluation of red-light cameras—Executive summary*. U.S. Department of Transportation.
- Federal Highway Administration. (2024a). *Identifying safety issues*. Local & Rural Safety Manual.
- Federal Highway Administration. (2024b). *Intersection safety*. U.S. Department of Transportation.
- G, J., & B, S. (2024). Object detection for traffic management based on YOLOV7. *SPIE*.
- Governors Highway Safety Association. (2024). *Pedestrian traffic fatalities by state: 2024 preliminary data*.
- Guaman-Quinche, A., et al. (2024). YOLO for traffic sign detection and recognition: A systematic review. *MDPI*.
- Insurance Institute for Highway Safety. (2016). *Turning off red light cameras costs lives, new research shows*. IIHS-HLDI.
- Insurance Institute for Highway Safety. (2024). *Red light running*. IIHS-HLDI.
- Kennedy Law Firm. (2025). *Crashes at red-lights: A serious and preventable danger*.
- Kothand, V., & Venkataraman, H. (2024). *Computer vision applications for traffic monitoring and safety analysis*. ROSA P.
- Kumar, P., et al. (2025). A dual-model deep learning framework for two-wheeler traffic violation detection. *Frontiers in Artificial Intelligence*.
- Kumar, S., et al. (2024). Traffic signal violation detection system using computer vision. *SSRN*.
- Lin, C.-J., & Jhang, J.-Y. (2022). Intelligent traffic-monitoring system based on YOLO and a convolutional fuzzy neural network. *IEEE Access*.
- Middleton, D., & Parker, R. (2003). *Vehicle detector evaluation (Report 2119-1)*. Texas Transportation Institute.
- Mihoub, A., et al. (2022). Annotated virtual detection line (AVDL) dataset for multiple object detection. *MDPI*.

N, Y., et al. (2023). An efficient traffic density estimation model using YOLOv7. *MDPI*.

National Conference of State Legislatures. (2024). *Traffic safety review: State speed and red-light camera laws and programs*.

National Highway Traffic Safety Administration. (2005). *Red light camera systems: Operational guidelines*. U.S. Department of Transportation.

National Highway Traffic Safety Administration. (2024). *Pedestrian safety*. NHTSA.

National Institute of Standards and Technology. (2018). *Performance evaluation metrics for object detection and tracking for robot safety*.

Pendergast Law. (2019). *AAA study finds red-light accidents hit highest mark in decade*.

Philipsen, M. P., et al. (2015). *Traffic signal detection: A learning algorithm and evaluations on challenging dataset* [Paper presentation]. IEEE Conference on Intelligent Transportation Systems.

Popa, M., et al. (2024). Analyzing real-time object detection with YOLO algorithm in automotive applications: A review. *CMES*.

Rapid Innovation. (2024). *Computer vision for real-time traffic flow analysis*.

Stateline. (2018, September 28). *Taking a U-turn on red-light, speed cameras*. Pew Charitable Trusts.

Tadlock, T. (2025). *What the IIHS study says about red light cameras and crash risks*. TW Law Firm.

Ultralytics. (2024). *YOLOv10: Real-time end-to-end object detection*.

Wang, A., et al. (2024). *YOLOv10: Real-time end-to-end object detection*. arXiv.

Wojke, N., Bewley, A., & Paulus, D. (2017). *Simple online and realtime tracking with a deep association metric* [Paper presentation]. IEEE International Conference on Image Processing.

Yap, K. C., et al. (2024). A cost-effective and efficient traffic violation detection system using YOLO and Deep SORT. *Journal of Information and Visualization*.

Zhang, H., et al. (2024). *Task-specific dual-model framework for comprehensive traffic safety video description and analysis*. arXiv.