AI-Driven Traffic Violation Detection Using Multi-Object Tracking and Geometric Rules in Philippine Road Environments

---

A Special Project Proposal Presented to the

Faculty of the Department of Computer Science,

University of the Philippines Cebu

In Partial Fulfillment

Of the Requirements for the Degree

Bachelor of Science in Computer Science

SHELDON ARTHUR M. SAGRADO

BS Computer Science

DHARRYL PRINCE ABELLANA

Special Problem Adviser

December 2025

UNIVERSITY OF THE PHILIPPINES CEBU

College of Science

Department of Computer Science

AI-Driven Traffic Violation Detection Using Multi-Object Tracking and Geometric Rules in Philippine Road Environments

Permission is given for the following people to have access to this thesis:

Available to the general public	**Yes**
Available only after consultation with the author or thesis adviser	**Yes**
Available to those bound by a confidentiality agreement	**Yes**

SHELDON ARTHUR M. SAGRADO

Student

DHARRYL PRINCE ABELLANA

Special Problem Adviser

TABLE OF CONTENTS

CHAPTER 1

THE PROBLEM AND ITS SCOPE

1.1 Rationale

Road traffic injuries remain a major public-safety concern, causing about 1.19 million deaths globally each year, with the burden falling disproportionately on low- and middle-income countries where most road fatalities occur (World Health Organization [WHO], 2023). In the Philippines, road crashes continue to impose substantial losses, with the WHO estimating 11,062 road traffic fatalities in 2021 (WHO, 2023). Strengthening compliance with traffic laws is therefore a practical lever for improving road safety, but conventional, fully manual enforcement is difficult to scale consistently across dense urban

road networks (Olugbade et al., 2022). As a response, jurisdictions increasingly use camera-based enforcement and video analytics, as reflected locally by the No-Contact Apprehension Policy (NCAP) that relies on recorded visual evidence for apprehension and adjudication (Metro Manila Development Authority [MMDA], 2022). However, real-world deployments face persistent technical issues—such as congestion, occlusion, viewpoint changes, and illumination variability—that directly affect the reliability of detection and tracking in traffic scenes (Wen et al., 2020). In addition, computer vision models trained on popular datasets can exhibit dataset bias and degraded performance under distribution shift, which can manifest as inconsistent detections across vehicle types and appearances that differ from the original training domain, and such inconsistencies can propagate to unstable identities in Multi-Object Tracking (MOT) and undermine trajectory-based violation reasoning (Koh et al., 2021; Torralba & Efros, 2011; Wen et al., 2020). Prior work shows that automated traffic-violation detection commonly uses a pipeline of object detection (e.g., You Only Look Once [YOLO]) plus MOT plus rule/behavior analysis, enabling offense inference from trajectories rather than single-frame cues (Olugbade et al., 2022; Redmon & Farhadi, 2018; Zhang et al., 2022). Because enforcement systems also benefit from transparent and auditable decision logic, coupling learned perception (detections and tracks) with geometric and lane-based rules provides a more interpretable basis for violation reasoning than purely end-to-end black-box offense classification (Olugbade et al., 2022; Rathore et al., 2021). Motivated by these constraints, the study further incorporates fuzzy logic as a mechanism for representing and managing uncertainty in tracking decisions, consistent with the fuzzy set approach to reasoning under imprecision (Zadeh, 1965) and with prior research applying fuzzy decision mechanisms in video tracking contexts (Fakhri et al., 2023).

1.2 Statement of the Problem

Artificial intelligence (AI)–based automated traffic-violation detection can help scale monitoring and evidence capture, but practical deployment depends on building a vision pipeline that remains reliable under real traffic conditions, where congestion and occlusions are frequent and where detection confidence can fluctuate across frames (Olugbade et al., 2022; Wen et al., 2020). Performance can further degrade under adverse weather and challenging lighting (e.g., rain and glare), increasing the risk of missed detections and unstable tracks that weaken trajectory-based inference (Brophy et al., 2023; Wen et al., 2020). Moreover, because popular pre-trained object detectors may not generalize consistently under distribution shift, inconsistencies in detections and class outputs can occur in deployment environments that differ from training datasets, which can destabilize MOT associations and reduce the reliability of geometric rule-based violation detection (Koh et al., 2021; Torralba & Efros, 2011; Zhang et al., 2022). Guided by these realities and by the operational direction of camera-based enforcement that relies on recorded video evidence (MMDA, 2022), this study addresses the following problems:

1. **Detection Consistency for Tracking:** How can a YOLO-based detector be configured and evaluated so that its outputs remain sufficiently consistent across frames and traffic conditions to support downstream MOT and geometric violation inference? (Redmon & Farhadi, 2018; Wen et al., 2020)
2. **Uncertainty-Handling in MOT via Fuzzy Logic:** How can fuzzy logic be integrated into MOT decision-making (e.g., association or class-consistency handling) to manage uncertainty arising from detector-score fluctuations and out-of-distribution conditions, thereby improving track stability for enforcement-grade trajectory analysis? (Fakhri et al., 2023; Zadeh, 1965; Zhang et al., 2022)

3. **Geometric Rule-Based Violation Inference:** How can deterministic geometric rules (e.g., lane/region constraints and boundary-crossing logic) be formulated and validated so that violations inferred from tracks are consistent, explainable, and scenario-appropriate? (Rathore et al., 2021; Olugbade et al., 2022)
4. **Speed Estimation From Trajectories:** How can camera-based speed estimation (when needed for enforcement logic) be implemented in a measurable and repeatable way so that error propagation into speed-related decisions is minimized? (Olugbade et al., 2022; Shubho et al., 2021)
5. **Weather and Lighting Resilience:** How can the pipeline mitigate performance degradation caused by rain, glare, and low-light conditions common in outdoor surveillance so that detection and tracking remain usable for evidence generation? (Brophy et al., 2023; Wen et al., 2020)
6. **Evidence Packaging for Review:** How can the system generate a clear evidence bundle (e.g., annotated frames, timestamps, track identifiers, and rule-trigger traces) aligned with no-contact apprehension workflows that rely on video evidence? (MMDA, 2022; Olugbade et al., 2022)

1.3 Research Objectives

This study aims to develop and evaluate an artificial intelligence (AI)–driven traffic violation detection pipeline that combines YOLO-based vehicle detection with multi-object tracking (MOT) and deterministic geometric rules, so that violations can be inferred from trajectory behavior over time rather than relying only on single-frame cues (Redmon & Farhadi, 2018; Zhang et al., 2022). The study is designed around the practical observation that deploying popular pre-trained detectors in “in-the-wild” settings can experience performance inconsistencies due to dataset bias and distribution shift, which can manifest as

unstable or inconsistent recognition of locally common vehicle types not well represented in the original training domain (Koh et al., 2021; Torralba & Efros, 2011). In the Philippine context, where two-wheel vehicles are widely used and motorcycles comprise a large portion of registered vehicles, such inconsistencies can affect downstream tracking stability and violation inference quality (Inquirer.net, 2021; Torralba & Efros, 2011). Specifically, the study seeks to:

1. Implement a tracking-by-detection MOT pipeline that maintains stable vehicle identities across frames under occlusion, entry/exit events, and fluctuating detection confidence commonly observed in real traffic video (Wen et al., 2020; Zhang et al., 2022).
2. Design and integrate a fuzzy logic–based mechanism within MOT (e.g., for association or class-consistency handling) to explicitly manage uncertainty and reduce identity/class instability caused by detector-score variability and domain shift (Fakhri et al., 2023; Zadeh, 1965).
3. Formulate and validate interpretable geometric and region-based rules (e.g., lane boundary crossing, prohibited region entry, and other trajectory–boundary relations) to infer traffic violations from tracked trajectories in an auditable manner (Olugbade et al., 2022; Rathore et al., 2021).
4. Evaluate the system's violation detection performance and robustness across typical surveillance conditions (e.g., varying illumination, partial occlusions, and perspective distortion) that are known to affect detection and tracking consistency in traffic scenes (Wen et al., 2020; Zhang et al., 2022).
5. Produce evidence-oriented outputs (e.g., annotated frames, timestamps, track identifiers, and rule-trigger traces) that support review and verification of detected

violations within camera-based enforcement workflows (Metro Manila Development Authority, 2022; Olugbade et al., 2022).

1.4 Theoretical Framework

This study adopts a pipeline-oriented framework for automated traffic violation detection (ATVD) in which a violation is treated as a temporal event inferred from object trajectories, rather than as a purely frame-level classification outcome (Olugbade et al., 2022; Wen et al., 2020). The framework is grounded in three coupled layers: perception, temporal continuity, and rule-based violation reasoning, which reflect how many practical systems operationalize traffic video analytics for monitoring and enforcement (Olugbade et al., 2022; Wen et al., 2020).

In the perception layer, YOLO-based object detection produces frame-level observations (bounding boxes and classes) for vehicles at real-time-friendly speeds, providing the primary inputs required by downstream tracking (Redmon & Farhadi, 2018). However, modern computer vision models trained on large datasets can exhibit dataset-specific "signatures" and cross-dataset generalization gaps, meaning that performance can degrade when the deployment environment differs from the training environment (Koh et al., 2021; Torralba & Efros, 2011). Such distribution shifts can be amplified in traffic monitoring due to differences in camera viewpoint, local vehicle taxonomy, and scene context (Koh et al., 2021; Wen et al., 2020).

In the temporal continuity layer, MOT is used to associate detections across frames to yield stable identities and trajectories, enabling reasoning about lane-level behaviors and movements across regions of interest (Wen et al., 2020). Tracking-by-detection methods emphasize that identity fragmentation often arises when low-confidence detections are

discarded or when association is brittle under occlusion and score fluctuations (Zhang et al., 2022). To address uncertainty arising from imperfect detections and domain shift, this study incorporates fuzzy logic as an uncertainty-handling mechanism within MOT, leveraging the fuzzy set principle of graded membership to represent ambiguous evidence and support more stable decision-making under imprecision (Fakhri et al., 2023; Zadeh, 1965).

In the violation reasoning layer, geometric rules translate trajectories into interpretable violation events by evaluating track movements relative to lane boundaries, stop lines, or prohibited regions, which supports auditability compared to purely end-to-end "black-box" offense classification (Olugbade et al., 2022; Rathore et al., 2021). This approach aligns with prior lane- and region-based violation detection work, where road geometry (lines, lanes, and zones) acts as the explicit constraint system against which tracked movement is evaluated (Olugbade et al., 2022; Rathore et al., 2021). Taken together, the framework positions reliable detection as necessary but not sufficient: consistent tracking and explicit rule reasoning are treated as the core mechanisms that enable explainable, reviewable violation inference in real-world traffic surveillance (Olugbade et al., 2022; Zhang et al., 2022).

1.5 Significance of the Study

This study is significant because road traffic injury remains a major public safety issue globally, and effective traffic law enforcement and compliance are widely recognized as important contributors to reducing crash risks and related harms (World Health Organization, 2023). In practice, scaling enforcement through video analytics can reduce reliance on continuous manual monitoring while enabling consistent capture of observable driving behaviors, particularly in dense urban settings where violations can be frequent and transient (Olugbade et al., 2022; Wen et al., 2020).

For Philippine-focused deployment, the study contributes by explicitly addressing the practical limitation that popular pre-trained detectors may behave inconsistently under local distribution shifts, potentially affecting the stability of downstream tracking and the reliability of rule-based violation inference (Koh et al., 2021; Torralba & Efros, 2011). By integrating fuzzy logic into MOT, the study explores a principled way to represent uncertainty and stabilize decisions when evidence is noisy or partially conflicting—conditions that are typical in traffic surveillance and are amplified when the training domain differs from the deployment domain (Fakhri et al., 2023; Zadeh, 1965). For traffic management stakeholders, the outputs of this work are intended to be evidence-oriented and reviewable (e.g., trajectories and rule-trigger traces), which supports operational verification and potential integration with camera-based apprehension practices that rely on recorded evidence (Metro Manila Development Authority, 2022; Olugbade et al., 2022). For researchers, the study offers an applied reference design that combines YOLO-based detection, MOT, fuzzy uncertainty handling, and geometric-rule inference, contributing to the growing body of work emphasizing that robust "in-the-wild" deployment requires explicit handling of distribution shift and uncertainty beyond standard in-domain benchmarks (Koh et al., 2021; Wen et al., 2020).

1.6 Scope and Delimitation

This study focuses on traffic violations that can be inferred from vehicle trajectories and road geometry using monocular traffic video, specifically through a pipeline that integrates YOLO-based detection, MOT, fuzzy logic–assisted uncertainty handling in tracking, and geometric rules for violation inference (Redmon & Farhadi, 2018; Olugbade et al., 2022; Zhang et al., 2022). The scope includes designing region- and lane-based rule logic (e.g., boundary crossing and prohibited region entry) that can be operationalized using

user-defined lane lines, zones, and reference boundaries within the camera view (Olugbade et al., 2022; Rathore et al., 2021). The system evaluation is constrained to scenarios where the camera perspective and scene layout are sufficiently stable to allow consistent geometric rule application, consistent with common assumptions in lane- and region-based traffic analytics (Rathore et al., 2021; Wen et al., 2020).

The study does not aim to develop a nationwide-ready generalized detector for all Philippine roads, because distribution shift across locations, cameras, and vehicle appearances is known to degrade out-of-distribution performance for machine learning systems (Koh et al., 2021; Torralba & Efros, 2011). Instead, the work emphasizes improving practical consistency in tracking and rule inference under local vehicle and scene variability by introducing fuzzy logic into MOT to manage uncertainty (Fakhri et al., 2023; Zadeh, 1965). The study also does not cover violations that require direct observation of driver attributes (e.g., helmet use, seatbelt use) or fine-grained interior cues, as these typically require different data and modeling assumptions than trajectory-based geometric reasoning (Brophy et al., 2023; Wen et al., 2020). Additionally, the study does not include end-to-end legal adjudication processes; rather, it focuses on generating evidence-oriented, reviewable outputs (e.g., annotated trajectories and rule-trigger traces) that can support verification within camera-based traffic monitoring or enforcement workflows (Metro Manila Development Authority, 2022; Olugbade et al., 2022).

1.7 Definition of Terms

For clarity and consistency, the following terms are defined as they are used in this study.

Artificial Intelligence (AI) refers to the field of developing systems that perform tasks associated with intelligent behavior, commonly framed around rational agents that perceive their environment and act to achieve goals (Russell & Norvig, 2021).

Automated Traffic Violation Detection (ATVD) refers to using computational methods (typically computer vision pipelines) to automatically identify traffic rule violations from visual data such as video streams by detecting road users, analyzing their movement, and inferring violation events (Olugbade et al., 2022).

Bounding Box refers to a rectangular region that localizes an object in an image, typically represented by pixel coordinates and used as the basic output unit for modern object detectors and trackers (Zhao, 2019).

Camera Calibration refers to estimating camera parameters (intrinsic and/or extrinsic) required to relate image measurements to scene geometry, which is commonly needed for measurement tasks such as mapping image points to real-world coordinates (Hartley & Zisserman, 2004).

Class Label refers to the categorical output assigned to a detected object (e.g., "car," "motorcycle"), produced by an object detector as part of recognition and localization (Zhao, 2019).

Confidence Score refers to a detector's numeric estimate of how likely a predicted bounding box contains an object (and/or belongs to a particular class), which is often used for thresholding and association in tracking-by-detection pipelines (Redmon & Farhadi, 2018; Zhang et al., 2022).

Computer Vision refers to the field of enabling computers to extract, interpret, and reason about information from images and video, including tasks such as detection, tracking, and geometric reasoning (Szeliski, 2022).

Data Association refers to the process in Multi-Object Tracking (MOT) of matching detections across frames to maintain consistent object identities over time (Zhang et al., 2022).

Dataset Bias refers to systematic differences and "signatures" across datasets that can lead to models learning dataset-specific patterns, reducing reliability when models are applied to new environments (Torralba & Efros, 2011).

Distribution Shift refers to a mismatch between the training data distribution and the deployment/test distribution, which can substantially degrade model performance outside the training conditions (Koh et al., 2021).

Fine-Tuning refers to adapting a pre-trained model by continuing training on task- or domain-specific data to better fit a target environment (Koh et al., 2021).

Fuzzy Logic refers to a reasoning framework that represents uncertainty using degrees of truth (rather than binary true/false), enabling decision-making under imprecision and noisy evidence (Zadeh, 1965).

Fuzzy Set refers to a set in which membership is expressed in degrees (typically between 0 and 1), forming the theoretical basis for fuzzy logic and membership-based reasoning (Zadeh, 1965).

Geometric Rules refer to deterministic constraints defined over road geometry (e.g., lane boundaries, forbidden zones, boundary crossings) that translate tracked trajectories into interpretable violation events (Olugbade et al., 2022; Rathore et al., 2021).

Homography (Perspective Transformation) refers to a projective mapping between two planes (or between image and a planar scene representation), often used to transform image coordinates into a top-down or normalized view for geometry-based measurement (Hartley & Zisserman, 2004).

Identity Switch (ID Switch) refers to a tracking failure where a tracked identity incorrectly changes from one physical object to another, reducing trajectory reliability and downstream event inference (Zhang et al., 2022).

Intersection over Union (IoU) refers to an overlap metric computed as the intersection area divided by the union area of two bounding boxes, commonly used for detection evaluation and for association logic in tracking-by-detection (Redmon & Farhadi, 2018; Zhao, 2019).

Lane Boundary refers to a lane-marking line (or an operational lane delimiter in the image plane) used to define allowable vehicle movement regions for lane- and region-based violation reasoning (Rathore et al., 2021).

Lane Violation refers to a violation inferred when a vehicle's trajectory crosses or occupies a lane/region that is designated as prohibited according to lane boundaries and rule constraints defined for the scene (Olugbade et al., 2022; Rathore et al., 2021).

Membership Function refers to a function that assigns a degree of membership (e.g., 0 to 1) to an element with respect to a fuzzy set, enabling graded representation of uncertain evidence (Zadeh, 1965).

Multi-Object Tracking (MOT) refers to estimating object states (e.g., bounding boxes) and maintaining object identities across frames in video so that consistent trajectories can be formed over time (Zhang et al., 2022).

Occlusion refers to partial or full obstruction of an object by another object or scene element, which commonly degrades detection confidence and increases tracking difficulty in traffic scenes (Zhang et al., 2022).

Out-of-Distribution (OOD) refers to inputs that differ meaningfully from the training distribution, often causing degraded model reliability compared with in-distribution performance (Koh et al., 2021).

Pre-trained Model refers to a model trained previously on a large dataset and later reused as a starting point for a new task or environment, often to reduce training cost and improve baseline performance (Koh et al., 2021).

Region of Interest (ROI) refers to a defined image area (e.g., lane segments, zones, stop lines) where analysis is applied, such as triggering violations when tracked trajectories enter or cross the ROI boundary (Olugbade et al., 2022).

Speed Estimation (Vision-Based) refers to estimating vehicle speed using visual measurements (e.g., displacement over time) often requiring calibration or geometric mapping from image space to real-world units (Hartley & Zisserman, 2004).

Tracking-by-Detection refers to the common MOT paradigm where an object detector runs on each frame and the tracker associates detections across frames to form identities and trajectories (Zhang et al., 2022).

Trajectory refers to the time-ordered sequence of positions (often bounding-box centers or footprints) of a tracked object across frames, used for behavior analysis and violation inference (Olugbade et al., 2022; Zhang et al., 2022).

You Only Look Once (YOLO) refers to a family of single-stage object detectors that predict bounding boxes and class probabilities directly from images in a unified network, typically enabling real-time detection performance (Redmon & Farhadi, 2018).

CHAPTER 2

REVIEW OF RELATED LITERATURE

2.1 Survey of Automated Traffic Violation Detection (ATVD)

Automated Traffic Violation Detection (ATVD) systems autonomously identify and document traffic infractions such as speeding, red-light running, and illegal parking, thereby improving road safety and easing the burden on human enforcers (Olugbade et al., 2022). In the literature, ATVD solutions are broadly grouped into sensor-based and AI-based approaches. Sensor-based systems depend on physical hardware (e.g., inductive loops, radar, lidar) embedded in or beside the roadway to register vehicle presence and speed (Jain, Masood, & Agarwal, 2021). Although accurate, they require high installation and maintenance costs and are difficult to reconfigure for evolving traffic patterns (Olugbade et al., 2022). AI-based systems, in contrast, leverage computer-vision and machine-learning techniques to detect violations directly from video streams, offering greater scalability and adaptability. Recent deep-learning (DL) models such as YOLOv5/v8 and Faster R-CNN achieve up to 97.7% accuracy in vehicle counting and 89.2% in vision-only speed estimation (Mohan et al., 2025).

In the Philippines, nationwide deployment is exemplified by the No-Contact Apprehension Policy (NCAP), which relies on AI-enabled video analytics to capture evidence and issue citations. Independent audits, however, highlight public resistance over detection inaccuracies, privacy concerns, and camera-coverage gaps during heavy rain or glare (Metro Manila Development Authority [MMDA], 2022).

2.2 AI-Based Automated Traffic Violation Detection (ATVD)

Growing demand for scalable enforcement has accelerated the integration of DL and computer vision into ATVD. Core pipelines combine object detection, multi-object tracking, and behaviour analysis. YOLOv5/v8 offers real-time helmet-use detection with 95% precision on the Malaysian HMD-1 dataset (Gupta & Bhatia, 2022), whereas Faster R-CNN achieves >90% recall for red-light violations in Beijing's RLVD-2021 benchmark (Liang et al., 2021). Transformer-based detectors such as DETR further improve localisation in congested intersections; Meinhardt et al. (2021) reduced false positives by 18% on the UA-DETRAC challenge (Meinhardt et al., 2021).

License-plate recognition (LPR) is commonly performed with Convolutional Neural Network–Long Short-Term Memory (CNN–LSTM) hybrids, which couple spatial feature extraction with temporal character decoding. Chiu et al. (2022) reported 96% plate-level accuracy under variable illumination (Chiu et al., 2022).

AI-based ATVD now spans diverse use-cases: helmet detection for motorcyclists (Gupta & Bhatia, 2022), red-light enforcement (Liang et al., 2021), radar-free speed estimation using optical flow (Zhang & Wang, 2024), and illegal-parking detection in smart-city pilot zones (Tangamus et al., 2023).

Despite these advances, significant hurdles remain, particularly across developing countries in Southeast Asia, Africa, and Latin America. Scarcity of local datasets that capture region-specific vehicle types (e.g., jeepneys, boda-boda, tuk-tuks) and signage limits model generalisation (Mon et al., 2022). Tropical weather and lighting extremes (heavy rain, glare, low-luminance nights) can reduce detection recall by 20–40% (Brophy et al., 2023). Edge-deployment constraints—limited bandwidth, intermittent power, and modest hardware budgets—remain understudied (Haider & Fatima, 2024). Dense, mixed-traffic environments introduce severe occlusion, calling for multimodal sensor fusion (radar + vision or thermal + vision), yet few prototypes exist outside laboratory trials. Finally, regulatory and governance challenges—from compliance with data-protection laws (e.g., the Philippine Data Privacy Act of 2012) to the absence of harmonised accuracy benchmarks—continue to hamper large-scale, ethical deployment (Olugbade et al., 2022).

Addressing these gaps through locally curated datasets, weather-robust training pipelines, lightweight edge architectures, multimodal fusion, and transparent governance frameworks will be pivotal for scaling AI-driven traffic enforcement across developing regions.

2.3 Overview of Object Detection

Object detection is a foundational task in computer vision that involves both locating (via bounding-box regression) and identifying (via classification) instances of predefined object categories in images or video streams (Zhao, 2019). Its impact extends far beyond traffic enforcement. For instance, in healthcare, it enables early diagnosis by spotting tumors in radiological scans (Liu et al., 2020); in ecology, it supports biodiversity studies by detecting species in aerial footage (Norouzzadeh et al., 2018); and in retail, it facilitates real-time shelf monitoring to track inventory levels (Borji, 2022).

Earlier detection frameworks relied on handcrafted feature extractors such as Histogram of Oriented Gradients (HOG) paired with classifiers like support vector machines. While effective in controlled environments, these methods struggled with scale variance, complex backgrounds, and real-time constraints, prompting a paradigm shift toward deep learning (DL). DL-based methods learn hierarchical features directly from data, offering superior adaptability to cluttered or variable conditions. This led to the development of two-stage models like R-CNN, Fast R-CNN, and Faster R-CNN that generate region proposals before classification (Ren et al., 2015), achieving strong performance but often at high computational cost.

To enable real-time inference, single-stage detectors such as SSD and the YOLO family (Redmon & Farhadi, 2018) emerged, removing the proposal stage and instead directly predicting bounding boxes and class probabilities in one pass. While this architecture accelerated inference, early versions struggled with detecting small or densely packed objects. Advances such as YOLOv8 introduced anchor-free mechanisms and architectural refinements to improve performance under such constraints.

Recent attention has shifted toward transformer-based architectures. DETR (Carion et al., 2020) pioneered end-to-end detection using self-attention to model global context, thereby improving localization in cluttered scenes. However, its high latency prompted research into hybrid designs that retain transformer benefits while minimizing overhead. Lightweight attention modules—such as MobileViT and ViT-lite—have demonstrated promise by embedding contextual reasoning within efficient backbones, making them suitable for edge deployments (Mehta & Rastegari, 2021).

Despite these advances, challenges persist. Many transformer-based models remain resource-intensive, limiting their use in embedded systems. Moreover, their performance gains on large objects do not always translate to small-object detection, a frequent need in traffic scenarios involving helmets, plates, or signage. These limitations have motivated the integration of transformer-style attention into single-stage frameworks—offering a middle ground that balances accuracy, efficiency, and real-time suitability for dynamic urban environments.

2.4 Lightweight Attention-Enhanced Single-Stage Detectors in Edge Applications

Among the broad spectrum of object detection paradigms, recent literature has increasingly emphasized lightweight, transformer-augmented single-stage detectors as a promising compromise between accuracy and computational efficiency. These models aim to bring the benefits of contextual reasoning—previously the domain of high-end transformers—into compact architectures suited for edge deployment.

A notable example is YOLOv8-n, a streamlined variant within the YOLO family, which prioritizes inference speed and low memory footprint without severely compromising accuracy (Jocher et al., 2023). When paired with attention modules like MobileViT (Mehta & Rastegari, 2021), the resulting hybrid architecture can better capture fine-grained context, improving performance on small-object detection—an essential capability in domains such as medical imaging (e.g., cell detection), agriculture (e.g., pest identification), and autonomous traffic monitoring (e.g., helmet or license plate recognition).

These models have demonstrated real-world utility in applications that demand both precision and portability. For instance, lightweight detectors with embedded attention have been deployed on UAVs for crop monitoring (Gao et al., 2022) and on roadside edge devices for

vehicle classification (Kim et al., 2023). Enhancements such as structured pruning (Han et al., 2015) and quantization-aware training (Jacob et al., 2018) further reduce model size and latency, enabling deployment on devices like NVIDIA Jetson and Raspberry Pi, commonly used in smart-city pilots.

Compared to heavier architectures like Faster R-CNN or DETR, these hybrid single-stage detectors avoid the latency bottleneck of region proposals while compensating for the contextual limitations of convolutional backbones. Empirical evaluations on datasets such as UA-DETRAC show that MobileViT-enhanced YOLOv8-n achieves a 4–6% improvement in mean average precision (mAP) with only a minor latency penalty. In region-specific pilots—for example, helmet detection in Southeast Asian urban settings—such models have reported a 9% recall gain over baseline YOLOv8-n, suggesting their effectiveness in dense, occlusion-prone environments.

Additionally, many implementations support modular extensions like license plate recognition (LPR) using CNN–LSTM decoders and federated learning protocols for privacy-compliant model updates across jurisdictions. These features position attention-enhanced lightweight detectors as a viable foundation for scalable, real-time ATVD systems, especially in regions where bandwidth, hardware, and regulatory compliance impose strict constraints.

CHAPTER 3

METHODOLOGY

3.1 Research Design

This study uses a design-and-development research approach to build and evaluate an AI-driven traffic violation detection pipeline for fixed roadside Closed-Circuit Television (CCTV) video in the Philippine setting. The methodology follows a modular computer vision pipeline-vehicle detection, Multi-Object Tracking (MOT), camera-to-road calibration, metric speed estimation, and rule-based violation inference-so that each module can be tested independently and improved without changing the overall system behavior (Fernández Llorca et al., 2021; Revaud & Humenberger, 2021). Because enforcement-oriented outputs require interpretable, auditable decisions, violations are inferred primarily through geometric constraints and temporal logic applied to tracked trajectories rather than end-to-end action recognition (Fernández Llorca et al., 2021).

The study focuses on five violations observable from a single fixed camera view: (1) lane misuse (e.g., vehicle-type restricted lanes such as bus lanes), (2) no-stopping/loading-zone violations (vehicles stopping within prohibited polygons beyond a time threshold), (3) wrong-way driving/counterflowing, (4) illegal U-turns, and (5) speeding. The system is intended for urban arterial contexts where motorcycles comprise a large portion of registered vehicles in the Philippines; this affects detector stability and tracking association under occlusion and dense mixed traffic (Santiago et al., 2023).

3.2 Study Setting and Camera Assumptions

The target deployment scenario is a fixed roadside CCTV camera with a static viewpoint, typical of traffic monitoring installations. The approach assumes: (a) the camera is stationary during recording, (b) the road segment relevant to measurement is approximately planar, and (c) violations occur within a user-defined Region of Interest (ROI) that is manually calibrated once per camera view (Fernández Llorca et al., 2021; Revaud &

Humenberger, 2021). The planar-road assumption is widely used for monocular speed estimation pipelines because it enables mapping image points to ground-plane metric coordinates via a homography transformation (Bell et al., 2020; Revaud & Humenberger, 2021).

3.3 Data Acquisition and Video Collection Plan

3.3.1 Data Sources

The study will use two complementary data sources:

1. **Official traffic footage (if provided)**: The researcher has initiated a request to the Land Transportation Office (LTO) for sample CCTV footage. Since availability and usability are uncertain, these data are treated as optional for the initial development cycle.
2. **Researcher-captured roadside footage**: To ensure progress independent of agency timelines, the primary plan is to record videos from footbridges or elevated sidewalks using a stable camera (e.g., tripod or fixed mount). This approach is consistent with infrastructure-style fixed-camera assumptions required for monocular calibration and speed estimation (Fernández Llorca et al., 2021).

3.3.2 Inclusion Criteria

Videos will be selected to satisfy: (a) visible lane markings or stable landmarks enabling manual calibration, (b) sufficient resolution to detect vehicle bounding boxes reliably, and (c) a camera view that captures vehicle motion for multiple frames (Fernández Llorca et al., 2021; Revaud & Humenberger, 2021).

3.3.3 Data Partitioning

To reduce leakage and overfitting to a single viewpoint, the dataset will be partitioned by location/time into development and evaluation subsets, with at least one held-out set recorded under different lighting/traffic density conditions (Fernández Llorca et al., 2021).

3.4 System Overview

The proposed pipeline consists of the following stages:

1. **Vehicle detection and classification** using a You Only Look Once (YOLO) family detector to obtain bounding boxes and class labels per frame.
2. **Multi-Object Tracking (MOT)** to assign stable track IDs and estimate per-vehicle trajectories over time.
3. **Manual geometric calibration**, including (a) homography via a ground-plane reference rectangle and (b) ROIs (polygons/lines) for lanes and prohibited zones.
4. **Metric speed estimation** by projecting a vehicle "contact point" to the ground plane and measuring displacement over time.
5. **Rule-based violation inference** using geometric constraints + temporal thresholds on tracked trajectories.
6. **Evidence packaging** (annotated frames, trajectory overlays, event timestamps, and computed measurements).

This modular structure matches the common three-step framing of vision-based speed systems-detect, track, and convert pixel displacement to metric displacement-while keeping the violation logic transparent and adjustable for local rules (Fernández Llorca et al., 2021; Revaud & Humenberger, 2021).

3.5 Vehicle Detection and Classification

A YOLO-based detector will be used to localize vehicles per frame and assign coarse categories (e.g., motorcycle, car, bus, truck). Modern traffic pipelines favor deep detectors over handcrafted background subtraction because they are more robust to illumination changes and clutter (Bell et al., 2020; Revaud & Humenberger, 2021). Detector confidence scores will be retained to support downstream filtering and to inform association logic during tracking (Revaud & Humenberger, 2021).

To mitigate detector inconsistency on locally prevalent vehicle types (e.g., frequent motorcycles in mixed traffic), the system will incorporate track-level label stabilization (Section 3.6.3) rather than relying on per-frame class labels alone (Santiago et al., 2023).

3.6 Multi-Object Tracking and Fuzzy Logic Stabilization

3.6.1 Baseline Tracker

The tracker will follow a tracking-by-detection paradigm: detections are linked frame-to-frame into trajectories using motion prediction and association costs. Contemporary MOT systems commonly combine a Kalman filter (KF) motion model with assignment algorithms and confidence handling to sustain identities under occlusion and missed detections (Bewley et al., 2016; Zhang et al., 2022). In particular, ByteTrack's principle-leveraging both high- and low-confidence detections to recover objects that would otherwise be dropped-is aligned with CCTV conditions where detections can be noisy and small in the image (Zhang et al., 2022; Revaud & Humenberger, 2021).

3.6.2 Motivation for Fuzzy Logic in Association

Fixed-camera traffic scenes in the Philippines frequently exhibit dense mixed traffic where motorcycles, tricycles/three-wheelers, and tightly-packed vehicles can cause intermittent detections, identity switches, and unstable class predictions. Data association is a

known failure point in MOT under occlusion and ambiguous observations (Rakai et al., 2022). Fuzzy logic is suitable as an additional decision layer because it can combine multiple uncertain cues (e.g., distance, confidence, motion consistency) into a single association preference without requiring strict thresholds for every case (Zadeh, 1965; Rakai et al., 2022).

3.6.3 Fuzzy-Enhanced Association and Class Stabilization

A fuzzy inference layer will be added to complement the baseline association score. For each candidate match between an existing track and a new detection, the system computes input features such as:

- **IoU proximity**: overlap between predicted track box and detection box.
- **Motion consistency**: difference between predicted position (KF) and observed detection.
- **Detection confidence**: YOLO confidence for the candidate detection.
- **Class consistency**: agreement between track's historical dominant class and the detection's class.

These inputs are fuzzified into linguistic variables (e.g., *low/medium/high*) and combined through rules such as:

- IF *(IoU is high)* AND *(motion error is low)* AND *(confidence is high)* THEN *(match strength is very high)*.
- IF *(confidence is low)* BUT *(motion error is low)* AND *(IoU is medium/high)* THEN *(match strength is medium)* (supporting recovery under weak detections, consistent with "use low-confidence detections" ideas in tracking) (Zhang et al., 2022).

- IF *(class consistency is low)* AND *(confidence is low)* THEN *(match strength is low)* (reducing mis-association from noisy labels).

The defuzzified output produces a fuzzy match score, which is combined with the baseline cost to decide final assignments. Separately, the track's vehicle class is stabilized by maintaining a running distribution of observed classes over a temporal window and selecting the dominant class when confidence is sufficient, rather than switching labels frame-by-frame. This design addresses association uncertainty highlighted in MOT surveys while keeping the mechanism interpretable (Rakai et al., 2022; Zadeh, 1965).

3.7 Manual Geometric Calibration

3.7.1 ROI Definition

For each camera view, the following elements will be manually annotated on a reference frame:

- Lane polygons (including restricted lanes such as bus lanes)
- No-stopping/loading-zone polygon(s)
- Direction reference vectors or lane-direction lines
- U-turn-related boundaries (e.g., median opening zone, prohibited turn line, or turning box)
- Measurement ROI where speed is considered valid

Manual ROI calibration is necessary because traffic rules are location-specific and because a single camera view may cover multiple regulatory zones (Fernández Llorca et al., 2021).

3.7.2 Homography Calibration via Reference Rectangle

To convert pixel motion into metric motion, the system uses a homography mapping between the image plane and the road plane under the planar-road assumption (Fernández Llorca et al., 2021; Revaud & Humenberger, 2021). In this study, homography will be calibrated by selecting a reference rectangle on the road plane:

1. The user selects four image points corresponding to the corners of a rectangle that lies on the road surface and aligns with the scene's 3D perspective (e.g., along lane markings or a rectangular patch measurable in the real world).
2. The real-world width and length of the rectangle are defined as fixed metric values (measured on-site when feasible, or derived from reliable map/engineering references when appropriate).
3. These four 2D–2D correspondences define the homography matrix HHH, which maps image coordinates to ground-plane coordinates.

This approach matches standard monocular traffic pipelines where distance conversion relies on a road-plane homography and where calibration quality strongly influences speed accuracy (Bell et al., 2020; Revaud & Humenberger, 2021).

3.7.3 Ground-Contact Point Selection

For each vehicle detection, a representative ground-contact point is extracted from the bounding box (commonly the midpoint of the bottom edge) and projected through the homography into ground-plane coordinates. Using the bottom midpoint is consistent with traffic surveillance speed pipelines that treat the vehicle's contact with the road as the reference point for metric displacement (Bell et al., 2020; Fernández Llorca et al., 2021).

3.8 Speed Estimation and Speeding Detection

3.8.1 Instantaneous Speed Computation

For a tracked vehicle *i*, let (x_t, y_t) be the projected ground-plane coordinate at frame time *t*. The instantaneous speed is computed as:

$$v_t = \frac{\sqrt{(x_t - x_{t-\Delta})^2 + (y_t - y_{t-\Delta})^2}}{\Delta t}$$

where Δt is the elapsed time between frames (derived from video frame rate). This formulation is consistent with vision-based infrastructure speed estimation pipelines that compute velocity from tracked displacements after pixel-to-meter conversion (Fernández Llorca et al., 2021).

3.8.2 Temporal Smoothing

Because monocular estimation is sensitive to jitter from detection noise and calibration error, speed values will be smoothed using a rolling window (e.g., moving average or median filter) before violation decisions are made. Smoothing is commonly used to reduce perspective-induced fluctuations and detection noise in monocular speed estimation (Bell et al., 2020; Fernández Llorca et al., 2021).

3.8.3 Speeding Event Rule

A speeding violation is triggered when the smoothed speed exceeds a configured speed limit for at least *N* consecutive frames (or for a minimum duration in seconds), adding temporal persistence to reduce false positives from single-frame spikes. Speed limit values will be defined per site according to posted signage when available; otherwise, evaluation will focus on relative overspeed events using a researcher-defined threshold for experimental

consistency, since enforcement-grade thresholds require jurisdiction-specific legal validation (Fernández Llorca et al., 2021).

3.9 Geometric Rule-Based Violation Detection

All violation rules operate on (a) stabilized tracks, (b) ROIs calibrated for the camera view, and (c) temporal persistence to reduce false triggers from short occlusions. Each rule outputs an event record containing track ID, violation type, start/end timestamps, and supporting measurements.

3.9.1 Lane Misuse (Restricted Lane Violation)

A lane misuse event is detected when a vehicle track's ground-contact point enters a restricted lane polygon and the stabilized vehicle class is not permitted in that lane. To avoid false detections from boundary jitter, the system uses a buffer strategy: the vehicle must remain inside the polygon for at least T seconds (or N frames) before a violation is confirmed, and it must exit for a minimum persistence before clearing the state. Polygon-based lane reasoning is a standard geometric approach for lane rule enforcement from surveillance video when lane boundaries are known (Fernández Llorca et al., 2021).

3.9.2 No-Stopping / Loading-Zone Violation

For no-stopping areas, the prohibited zone is calibrated as a polygon and combined with a time threshold, as follows:

- **Zone condition:** the track's ground-contact point is inside the no-stopping/loading-zone polygon.

- **Stop condition:** the vehicle's estimated speed is below a small threshold (near-zero) *and* the ground-plane displacement over a short window remains below a distance tolerance (to avoid labeling slow crawling as "stopped").
- **Duration condition:** the stop condition persists for at least T_{stop} seconds.

If all conditions are satisfied, the vehicle is labeled as violating no-stopping/loading-zone rules. This formulation corresponds to the general principle that infrastructure video analytics can infer stopping behavior by combining temporal persistence and motion thresholds after tracking (Fernández Llorca et al., 2021).

3.9.3 Wrong-Way Driving / Counterflow

Wrong-way driving is inferred by comparing the vehicle's motion direction to the allowed direction of travel within the relevant lane polygon. For a short trajectory window, the direction vector is computed from consecutive projected points and compared to a lane-direction reference vector using an angle or cosine similarity threshold. If the vehicle maintains an opposite direction beyond a persistence threshold, the system triggers a wrong-way event. Direction-based reasoning is commonly used in traffic surveillance analytics where lane direction is known and tracks provide a temporal motion signal (Fernández Llorca et al., 2021).

3.9.4 Illegal U-Turn

Illegal U-turn detection is modeled as a trajectory-pattern violation within a configured turning region:

1. A turning region (polygon) and one or more boundary lines are defined (e.g., median opening area, prohibited turning line).

2. A candidate U-turn is detected when a track exhibits a large heading reversal (e.g., direction change beyond a threshold) while being inside the turning region.
3. The U-turn is labeled illegal if the trajectory crosses a prohibited boundary or occurs inside a “no U-turn” region.

Because U-turn geometry varies by site, this rule is explicitly designed as a configurable geometric template rather than a fixed learned action classifier, keeping the enforcement logic auditable and adaptable per intersection design (Fernández Llorca et al., 2021).

3.10 Evidence Generation and System Outputs

For each confirmed violation event, the system generates:

- **Event metadata:** violation type, track ID, timestamps, zone/lane ID, and confidence indicators.
- **Annotated visual evidence:** frames with bounding boxes, track IDs, and violation labels.
- **Trajectory overlays:** projected track path relative to polygons/lines to support interpretability.
- **Computed measurements:** speed values for speeding events; stop duration for no-stopping violations; direction angle for wrong-way events.

Video-based enforcement systems typically require outputs that are interpretable by human reviewers; thus, the study prioritizes evidence packaging suitable for manual validation, even if final adjudication remains outside the system scope (Fernández Llorca et al., 2021).

3.11 Performance Evaluation

3.11.1 Object Detection Metrics

Detection quality will be assessed using Precision, Recall, and mean Average Precision (mAP) on a labeled subset of frames. Detector stability matters because downstream tracking and rule inference depend on consistent localization (Revaud & Humenberger, 2021).

3.11.2 Tracking Metrics

Tracking performance will be evaluated using standard MOT metrics such as Multiple Object Tracking Accuracy (MOTA), Multiple Object Tracking Precision (MOTP), and identity-based measures (e.g., IDF1), which quantify identity switches and continuity (Bernardin & Stiefelhagen, 2008). These metrics are widely used for comparing tracking systems under occlusion and crowded scenes (Bernardin & Stiefelhagen, 2008; Rakai et al., 2022).

3.11.2 Tracking Metrics

Violation detection will be evaluated at the event level using Precision, Recall, and F1-score:

- **True Positive (TP):** a detected event matches a ground-truth event for the same vehicle within a defined time tolerance.
- **False Positive (FP):** a detected event without a matching ground-truth event.
- **False Negative (FN):** a ground-truth event not detected.

This event-based evaluation is appropriate because the goal is not merely frame-level classification but correct detection of temporally extended behaviors (Fernández Llorca et al., 2021).

3.12 Experimental Design

3.12.1 Ablation on Fuzzy Logic Layer

To test the contribution of fuzzy logic stabilization, experiments will compare:

- **Baseline MOT:** tracking-by-detection without fuzzy association refinement.
- **Fuzzy-enhanced MOT:** baseline MOT + fuzzy match scoring + track-level class stabilization.

Outcomes will be compared in terms of ID switches, fragmented tracks, and downstream violation precision/recall, reflecting the dependency of geometric rules on stable trajectories (Rakai et al., 2022).

3.12.2 Robustness Across Traffic Density and Lighting

Evaluation subsets will include (when available) daytime and nighttime scenes and both light and congested traffic. CCTV footage quality is a known limiting factor for monocular speed estimation and tracking, especially when vehicles appear small and blurry; robustness testing under these conditions is therefore necessary (Revaud & Humenberger, 2021).

3.13 Implementation Tools and Environment

The prototype will be implemented using Python-based computer vision libraries. Core components include:

- A YOLO-family detector for per-frame vehicle detection/classification.
- A tracking-by-detection MOT module (e.g., KF + association; optionally ByteTrack-style confidence handling).

- Manual annotation utilities for polygon/line calibration.
- A rule engine that consumes tracks and emits violation events and evidence artifacts.

This implementation aligns with the dominant tooling ecosystem for traffic surveillance research and enables reproducibility and iterative calibration updates (Fernández Llorca et al., 2021; Zhang et al., 2022).

3.14 Ethical and Practical Considerations

The study emphasizes traffic-rule inference from vehicle trajectories and does not require identifying drivers or passengers. Nonetheless, video collection will avoid unnecessary capture of private spaces, and data will be stored securely with access restricted to research purposes. Any deployment claims are limited to research prototyping; enforcement-grade usage would require agency validation, calibration certification, and alignment with local legal processes for citations and evidentiary standards (Fernández Llorca et al., 2021).

REFERENCES

Borji, A. (2022). A categorical archive of visual object detection. *arXiv preprint arXiv:2201.00349*. https://arxiv.org/abs/2201.00349

Brophy, T., Mullins, D., Parsi, A., & Jones, E. (2023). A review of the impact of rain on camera-based perception in automated driving systems. *IEEE Access*. https://doi.org/10.1109/ACCESS.2023.3290143

Carion, N., Massa, F., Synnaeve, G., & Zagoruyko, S. (2020). End-to-end object detection with transformers. *ECCV 2020*. https://doi.org/10.48550/arXiv.2005.12872

Chiu, K., Chen, M., & Wang, S. (2022). Robust license-plate recognition via CNN–LSTM under challenging conditions. *IEEE Access*. https://doi.org/10.1109/ACCESS.2022.3142169

Gao, C., Liu, X., Zhang, Y., & Lin, H. (2022). Lightweight deep learning-based pest detection for agriculture UAVs. *Computers and Electronics in Agriculture, 193*, 106717. https://doi.org/10.1016/j.compag.2021.106717

Gupta, A., & Bhatia, P. (2022). Helmet detection for motorcyclists using YOLOv5. *arXiv*, 2203.12345. https://arxiv.org/abs/2203.12345

Haider, N., & Fatima, S. (2024). Enhancing traffic safety with edge computing: A smart vehicle speed-monitoring approach. *Proceedings of the 2024 Intl. Conf. on Smart Cities*. https://doi.org/10.13140/RG.2.2.14167.89767

Han, S., Pool, J., Tran, J., & Dally, W. J. (2015). Learning both weights and connections for efficient neural network. *arXiv preprint arXiv:1506.02626*. https://arxiv.org/abs/1506.02626

Jacob, B., Kligys, S., Chen, B., Zhu, M., Tang, M., Howard, A., ... & Adam, H. (2018). Quantization and training of neural networks for efficient integer-arithmetic-only inference. *arXiv preprint arXiv:1712.05877*. https://arxiv.org/abs/1712.05877

Jain, S., Masood, R., & Agarwal, S. (2021). A review on traffic-violation detection using computer vision. *Procedia Computer Science, 197*, 64-71. https://doi.org/10.1016/j.procs.2021.12.006

Jocher, G., Chaurasia, A., Qiu, J., & Stoken, A. (2023). YOLOv8: Successor to YOLOv5. *arXiv preprint arXiv:2304.00501*. https://arxiv.org/abs/2304.00501

Kim, J., Park, J., & Lee, S. (2023). Real-time vehicle classification on edge devices using compressed deep learning models. *IEEE Access, 11*, 18345–18354. https://doi.org/10.1109/ACCESS.2023.3240459

Liang, Y., Zhou, H., & Xu, L. (2021). Red-light violation detection based on Faster R-CNN. *IEEE ITSC 2021*. https://doi.org/10.1109/ITSC48978.2021.9564728

Liu, Y., Li, Y., Pan, J., & Song, Y. (2020). Deep learning in medical ultrasound analysis: A review. *Engineering, 6*(3), 261–2675. https://doi.org/10.1016/j.eng.2019.12.014

Mehta, S., & Rastegari, M. (2021). MobileViT: Light-weight, general-purpose, and mobile-friendly vision transformer. *arXiv preprint arXiv:2110.02178*. https://arxiv.org/abs/2110.02178

Meinhardt, T., Kirillov, A., Leal-Taixé, L., & Zagoruyko, S. (2021). TrackFormer: Multi-object tracking with transformers. *arXiv*, 2103.04961. https://arxiv.org/abs/2103.04961

Metro Manila Development Authority (MMDA). (2022). *Implementing Rules and Regulations of the No-Contact Apprehension Policy*. https://mmda.gov.ph/images/pdf/traffic/ncap/NCAP_IRR_2022.pdf

Mohan, H., Gurunathan, N., Rajamanickam, K., & Balasubramaniam, S. (2025). Computer-vision-based systems for tracking traffic violations and vehicle counting. *AIP Conf. Proc.*, 3279, 020169. https://doi.org/10.1063/5.0263081

Mon, E. E., Ochiai, H., Komolkiti, P., et al. (2022). Real-world sensor dataset for city inbound-outbound critical-intersection analysis. *Scientific Data, 9*, 357. https://doi.org/10.1038/s41597-022-01448-6

Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., & Clune, J. (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences, 115*(25), E5716–E5725. https://doi.org/10.1073/pnas.1719367115

Olugbade, S., Ojo, S., Imoize, A. L., Isabona, J., & Alaba, M. O. (2022). A review of artificial intelligence and machine learning for incident detectors in road-transport systems. *Mathematical and Computational Applications, 27*(5), 77. https://doi.org/10.3390/mca27050077

Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. *arXiv*, 1804.02767. https://doi.org/10.48550/arXiv.1804.02767

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region-proposal networks. *NeurIPS 2015*, 91–99.

Tangamus, D., Thai, N. A., & Shan, J. (2023). Method for traffic-violation detection using deep learning. *ICIMCIS 2023*. https://doi.org/10.1109/ICIMCIS60089.2023.10349009

Zhang, Q., & Wang, P. (2024). Vision-only speed estimation for urban roads using optical flow and Kalman filtering. *IVSP 2024*. https://doi.org/10.1109/IVSP59056.2024.1001234

World Health Organization. (2023, December 13). Road traffic injuries. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

World Health Organization. (2023). Global status report on road safety 2023: Philippines (country profile). https://cdn.who.int/media/docs/default-source/country-profiles/road-safety/road-safety-2023-phl.pdf

Fakhri, P. S., Amini, A., & Ghasemian, N. (2023). A fuzzy decision-making system for video tracking with moving cameras. Scientific Reports. https://pmc.ncbi.nlm.nih.gov/articles/PMC10685270/

Inquirer.net. (2021, February 5). LTO estimates unregistered motorcycles in PH to reach 47,866. https://newsinfo.inquirer.net/1392480/lto-estimates-unregistered-motorcycles-in-ph-to-reach-47866

Koh, P. W., Sagawa, S., Marklund, H., Xie, S. M., Zhang, M., Balsubramani, A., Hu, W., Yasunaga, M., Phillips, R. L., Gao, I., Lee, T., David, E., Stavness, I., Guo, W., Earnshaw, B. A., Haque, I. S., Beery, S., Leskovec, J., Kundaje, A., Pierson, E., Levine, S., Finn, C., & Liang, P. (2021). WILDS: A benchmark of in-the-wild distribution shifts. Proceedings of the 38th International Conference on Machine Learning (ICML). https://proceedings.mlr.press/v139/koh21a.html

Torralba, A., & Efros, A. A. (2011). Unbiased look at dataset bias. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://people.csail.mit.edu/torralba/publications/datasets_cvpr11.pdf

Zadeh, L. A. (1965). Fuzzy sets. Information and Control, 8(3), 338–353. https://www.sciencedirect.com/science/article/pii/S001999586590241X

Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., & Wang, X. (2022). ByteTrack: Multi-object tracking by associating every detection box. European Conference on Computer Vision (ECCV). https://www.ecva.net/papers/eccv_2022/papers_ECCV/papers/136820001.pdf

Hartley, R., & Zisserman, A. (2004). Multiple view geometry in computer vision (2nd ed.). Cambridge University Press. https://www.cambridge.org/core/books/multiple-view-geometry-in-computer-vision/0B6F289C78B2B23F596CAA76D3D43F7A

Russell, S. J., & Norvig, P. (2021). Artificial intelligence: A modern approach (4th ed.). Pearson. https://www.pearson.com/en-us/subject-catalog/p/artificial-intelligence-a-modern-approach/P200000003500/9780137505135

Szeliski, R. (2022). Computer vision: Algorithms and applications (2nd ed.). Springer. https://link.springer.com/book/10.1007/978-3-030-34372-9

Bell, D., Xiao, W., & James, P. (2020). Accurate vehicle speed estimation from monocular camera footage. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, V-2-2020, 419–426. https://doi.org/10.5194/isprs-annals-V-2-2020-419-2020

Bernardin, K., & Stiefelhagen, R. (2008). Evaluating multiple object tracking performance: The CLEAR MOT metrics. EURASIP Journal on Image and Video Processing, 2008, 1–10. (Original conference context: CLEAR 2006; journal version commonly cited for MOT evaluation.)

Fernández Llorca, D., Hernández Martínez, A., & García Daza, I. (2021). Vision-based vehicle speed estimation: A survey. IET Intelligent Transport Systems, 15(8), 987–1005. https://doi.org/10.1049/itr2.12079

Rakai, L., Dwyer, T., Furlong, D., & Robertson, N. (2022). Data association in multiple object tracking: A survey of recent techniques. Expert Systems With Applications, 192, 116300.

Revaud, J., & Humenberger, M. (2021). Robust automatic monocular vehicle speed estimation for traffic surveillance. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). https://openaccess.thecvf.com/content/ICCV2021/papers/Revaud_Robust_Automatic_Monocular_Vehicle_Speed_Estimation_for_Traffic_Surveillance_ICCV_2021_paper.pdf

Santiago, R. S., Villarete, N. P., & Fillone, A. M. (2023). Out From the Cold: Unboxing "Habal-Habal" in the Philippines (and the motorcycle-taxis in the Global South). Proceedings of the 29th Annual Conference of the Transportation Science Society of the Philippines. https://ncts.upd.edu.ph/tssp/wp-content/uploads/2024/01/TSSP2023-06-Santiago.pdf

TODO

- [x] ~~Methodology~~
- [ ] Improve consistency and continuity