HYBRID 5G–SATELLITE ADAPTIVE VIDEO STREAMING FOR LOW-LATENCY CLOUD-BASED COOPERATIVE PERCEPTION

by

EKHLASS MESLEH ALKHARABSHEH

A dissertation submitted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY IN ELECTRICAL AND COMPUTER ENGINEERING

2026

Oakland University
Rochester, Michigan

Doctoral Advisory Committee:

Shadi Alawneh, Ph.D., Chair
Osamah Rawashdeh, Ph.D.
Mohammad Wardat, Ph.D.
Li Li, Ph.D.

© Copyright by EKHLASS MESLEH ALKHARABSHEH, 2026
All rights reserved

To my parents and my wonderful family, whose unwavering love, encouragement, and sacrifices made this journey possible

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to Prof. Shadi for his invaluable guidance, insight, and continuous encouragement throughout my doctoral research. His expertise in intelligent transportation systems, cloud computing, and autonomous perception has profoundly shaped the direction and depth of this work. I am also sincerely thankful to the members of my dissertation committee—Prof. Osamah Rawashdeh, Prof. Mohammad Wardat, and Prof. Li Li—for their thoughtful feedback, support, and commitment to strengthening the quality of this research.

I am deeply grateful to my family, especially my parents, for their unwavering love, patience, and faith in me. Their belief in hard work and perseverance has been my greatest source of motivation and strength. I am equally thankful to my husband for his steadfast support, understanding, and encouragement, and to my children, whose joy and resilience have been a constant source of inspiration throughout this journey.

EKHLASS MESLEH ALKHARABSHEH

PREFACE

The journey toward this dissertation has been shaped by a deep fascination with the convergence of intelligent transportation, cloud computing, and real-time perception systems. The rapid evolution of connected and automated vehicles has introduced both opportunities and challenges—particularly in achieving the ultra-low-latency, high-reliability communication necessary for safe and scalable autonomous operation. This work was motivated by the pressing need to bridge the gap between laboratory prototypes and deployable, infrastructure-supported perception frameworks capable of operating in diverse, real-world environments.

Throughout this research, the focus remained on developing a practical and adaptive framework for optimizing the transmission and processing of high-volume video data generated by roadside unit (RSU) cameras. The system integrates recent advances in hybrid 5G–Starlink communication, adaptive video compression, and cloud-based GPU inference to achieve robust performance under varying network and environmental conditions. The results presented here demonstrate how intelligent networking and computational adaptation can accelerate the realization of cooperative perception as a cornerstone of future intelligent transportation systems.

This dissertation represents not only a technical investigation but also a personal and professional journey of discovery, resilience, and collaboration. The insights gained extend beyond the engineering domain, underscoring the importance of interdisciplinary thinking, teamwork, and persistence in advancing complex technologies. It is my hope that this work will contribute meaningfully to the development of safer, more efficient, and more connected mobility ecosystems—where vehicles, infrastructure, and cloud intelligence work seamlessly together for the benefit of society.

ABSTRACT

HYBRID 5G–SATELLITE ADAPTIVE VIDEO STREAMING FOR LOW-LATENCY CLOUD-BASED COOPERATIVE PERCEPTION

by

EKHLASS MESLEH ALKHARABSHEH

Adviser: Shadi Alawneh, Ph.D.

The growing complexity of autonomous and connected vehicle systems necessitates reliable, low-latency communication between vehicles, infrastructure, and cloud resources. This dissertation presents a comprehensive framework for optimizing real-time data transmission in cloud-based cooperative perception systems (CPS), focusing on adaptive video compression, hybrid networking, and cloud-based inference. The research aims to reduce end-to-end latency while maintaining high perception accuracy and connectivity reliability across heterogeneous communication environments.

The proposed architecture integrates adaptive H.265 (HEVC) compression, hybrid 5G–Starlink networking, and GPU-accelerated inference to support scalable roadside unit (RSU) camera deployments. An adaptive compression controller dynamically adjusts frame rate, resolution, and bitrate in response to bandwidth fluctuations, while a hybrid link manager employs make-before-break switching and multipath transmission to sustain continuous connectivity. The system was evaluated through a combination of simulation, laboratory bench testing, and real-world field experiments under diverse network conditions and video workloads.

Results demonstrate that the integrated approach achieves end-to-end latency below 100 milliseconds and connectivity uptime exceeding 99 percent, outperforming conventional single-network and fixed-parameter configurations. Statistical analyses confirm significant reductions in latency variance, frame loss, and recovery time following network disruptions. Moreover, the proposed methods maintain detection accuracy (mAP > 0.75) and system stability under high-load scenarios involving up to 100 concurrent RSU streams.

This work contributes (i) a validated, modular framework for real-time cloud-assisted cooperative perception, (ii) empirical benchmarks quantifying compression–network trade-offs, and (iii) deployment guidelines for scalable CPS implementations in intelligent transportation systems. Collectively, these contributions advance the state of knowledge toward deployable, cloud-integrated Level-5 autonomous driving infrastructures, enabling robust perception pipelines with demonstrable reliability, scalability, and cost efficiency.

TABLE OF CONTENTS

TABLE OF CONTENTS—Continued

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADAS	Advanced Driver Assistance Systems
AV	Autonomous Vehicle
CAV	Connected and Automated Vehicle
C-ITS	Cooperative Intelligent Transport Systems
CPM	Collective Perception Message
C-V2X	Cellular Vehicle-to-Everything
DSRC	Dedicated Short-Range Communication
ETSI	European Telecommunications Standards Institute
HEVC	High Efficiency Video Coding (H.265)
IoV	Internet of Vehicles
ITS	Intelligent Transportation Systems
RSU	Roadside Unit
SAE	Society of Automotive Engineers
V2C	Vehicle-to-Cloud
V2I	Vehicle-to-Infrastructure
V2P	Vehicle-to-Pedestrian
V2V	Vehicle-to-Vehicle
V2X	Vehicle-to-Everything

CHAPTER ONE

INTRODUCTION

Autonomous and connected vehicle technologies have rapidly evolved over the past two decades, reshaping the landscape of intelligent transportation systems (ITS). The growing reliance on perception-driven automation has underscored the importance of real-time environmental awareness, communication reliability, and cloud-based computation. Yet, achieving safe and scalable autonomy requires more than advanced onboard sensors—it demands coordinated perception across vehicles and infrastructure, supported by robust data transmission pipelines. This chapter introduces the context and motivation for optimizing real-time data transmission in cloud-based cooperative perception systems. It traces the historical evolution of autonomous driving, outlines the enabling role of advanced driver assistance systems, identifies key challenges in perception and connectivity, and highlights the promise of cooperative perception as a pathway toward full automation. The chapter concludes with the research aim, objectives, methodology, and contributions that frame the remainder of this dissertation.

Historical Foundations of Autonomous Driving

The evolution of autonomous driving represents one of the most transformative developments in modern engineering. What began as small-scale academic experiments has matured into a global research endeavor encompassing robotics, artificial intelligence, communications, and human–machine interaction. Early prototypes in the 1980s demonstrated that vehicles could, in principle, perceive their surroundings and make driving decisions without continuous human control. Projects such as Carnegie

Mellon University's Navlab and the DARPA Autonomous Land Vehicle (ALV) program integrated cameras, sonar sensors, and primitive onboard computing to navigate predefined routes, establishing the conceptual and technical basis for vision-based driving [1], [2].

Figure 1: DARPA Grand Challenge – Stanford University's Autonomous Vehicle "Stanley" [7]

European initiatives soon followed: the EUREKA PROMETHEUS Project and the University of Parma's ARGO vehicle pioneered lane-keeping and highway automation using stereo-vision and rule-based control [3], [4]. Collectively, these efforts forged the earliest perception and control architectures that would later evolve into modern Advanced Driver Assistance Systems (ADAS) and fully autonomous platforms.

A significant leap occurred with DARPA's Grand Challenges (2004–2005) and the Urban Challenge (2007), where teams were tasked with navigating complex off-road and urban environments without human intervention. Stanford's *Stanley* and Carnegie Mellon's *Boss* demonstrated robust multi-sensor fusion, real-time path planning, and decision-making in dynamic environments [5], [6]. These breakthroughs catalyzed industrial efforts by companies such as Waymo and Tesla, setting the foundation for today's autonomous ecosystems. Despite steady progress, most deployments still operate at SAE Level 2–4 automation, as defined in SAE J3016 [7], restricted to narrow operational design domains (ODDs). Achieving Level 5 autonomy, capable of full self-driving in all conditions, remains a core challenge motivating this research.

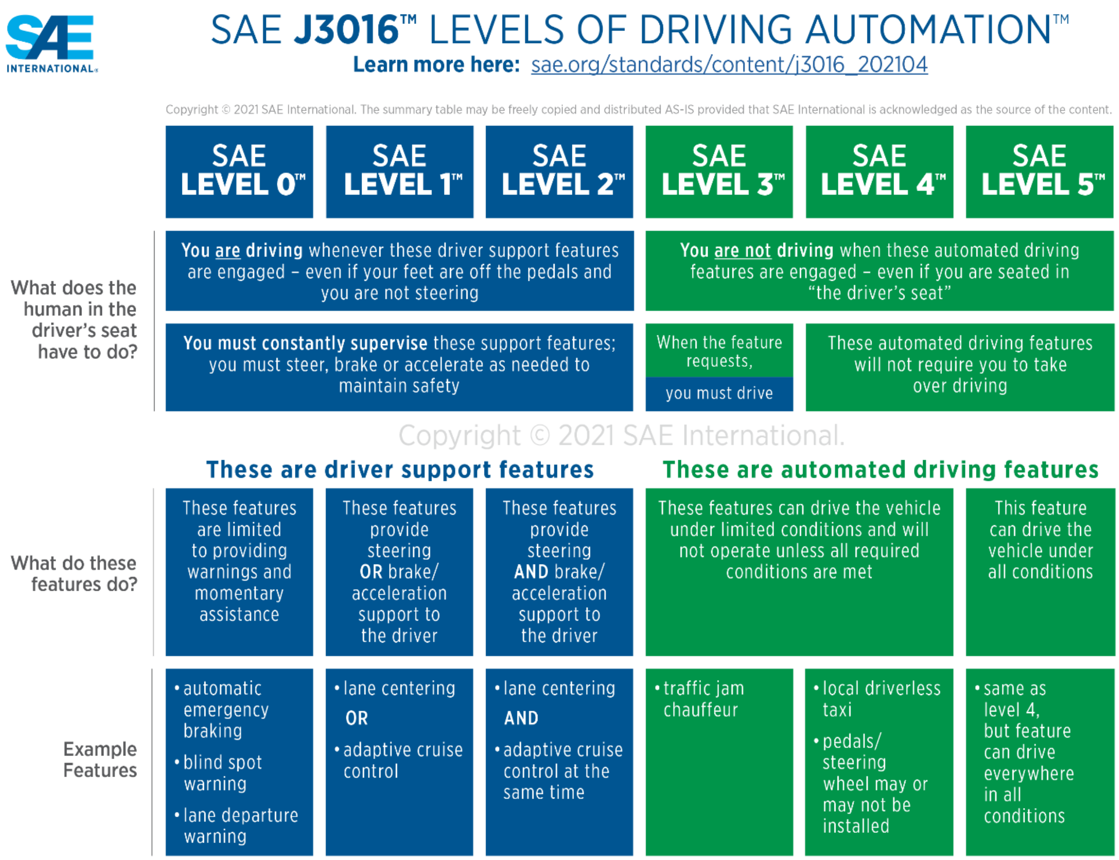

Figure 2: SAE Levels of Autonomy (Level 0–5) [8]

Role of Advanced Driver Assistance Systems (ADAS)

Before vehicles can reach full autonomy, incremental automation through ADAS serves as a practical bridge between human-driven and autonomous vehicles. ADAS features such as adaptive cruise control, lane-keeping assist, blind-spot monitoring, and automatic emergency braking (AEB) have demonstrably improved road safety [8]. Empirical evidence indicates that AEB alone can reduce rear-end collisions by nearly 50 percent in specific driving conditions [9]. However, ADAS primarily rely on localized onboard sensors—cameras, radar, and LiDAR—whose field of view is limited by occlusions, weather, and illumination. For instance, the American Automobile Association (AAA) found that pedestrian detection accuracy degrades sharply at night [10]. These constraints reveal that purely onboard perception is insufficient for comprehensive situational awareness.

Recent developments are transforming ADAS into connected, continuously learning systems. Cloud connectivity enables over-the-air (OTA) updates and large-scale data aggregation, allowing real-time algorithm improvements [11]. This shift from isolated sensing to collaborative perception marks a pivotal step toward fully autonomous and cooperative mobility.

Challenges in Autonomous Driving

Although autonomous perception and control technologies have advanced considerably, widespread deployment still faces formidable technical and societal challenges. These include sensor limitations, computational constraints, network reliability, and system resilience. Perception challenges arise from limited sensor range, occlusion, and environmental degradation. Accidents involving automated vehicles have underscored the risks of overreliance on local sensing alone [12]. Computational limitations are also evident: high-performance onboard GPUs improve inference accuracy but increase energy consumption, cost, and thermal load, reducing electric vehicle range and affordability [13].

Connectivity gaps persist globally; while 5G promises ultra-reliable low-latency communication (URLLC), coverage remains uneven, particularly outside dense urban centers [14]. Satellite-based systems such as Starlink offer complementary low-latency backhaul for underserved regions [15]. Meanwhile, public trust in autonomous systems continues to evolve, influenced by safety incidents and public perception studies [16]. Addressing these challenges requires a holistic approach that unites sensing, communication, and computing—a vision realized through cooperative perception.

Promise and Benefits of Cooperative Perception

Cooperative perception (CP) extends a vehicle's situational awareness by enabling real-time data exchange with other vehicles and roadside infrastructure via Vehicle-to-Everything (V2X) communication. By integrating distributed sensor data in the cloud, CP allows vehicles to detect and respond to hazards that are occluded or beyond their own sensor range [17], [18].

The benefits of CP span several domains:

- Safety: improved detection of hidden objects and reduced collision risk.
- Efficiency: smoother traffic flow and reduced congestion through coordinated driving.
- Cost: lower onboard computational load through cloud offloading.

- Scalability: centralized learning and model updates for fleets operating under varying conditions.

Cloud-based CP also supports collective intelligence by fusing multi-camera RSU data, improving robustness in complex environments such as intersections or urban canyons [19]–[22]. These capabilities position CP as a key enabler for achieving Level-5 autonomy and advancing Cooperative Intelligent Transport Systems (C-ITS).

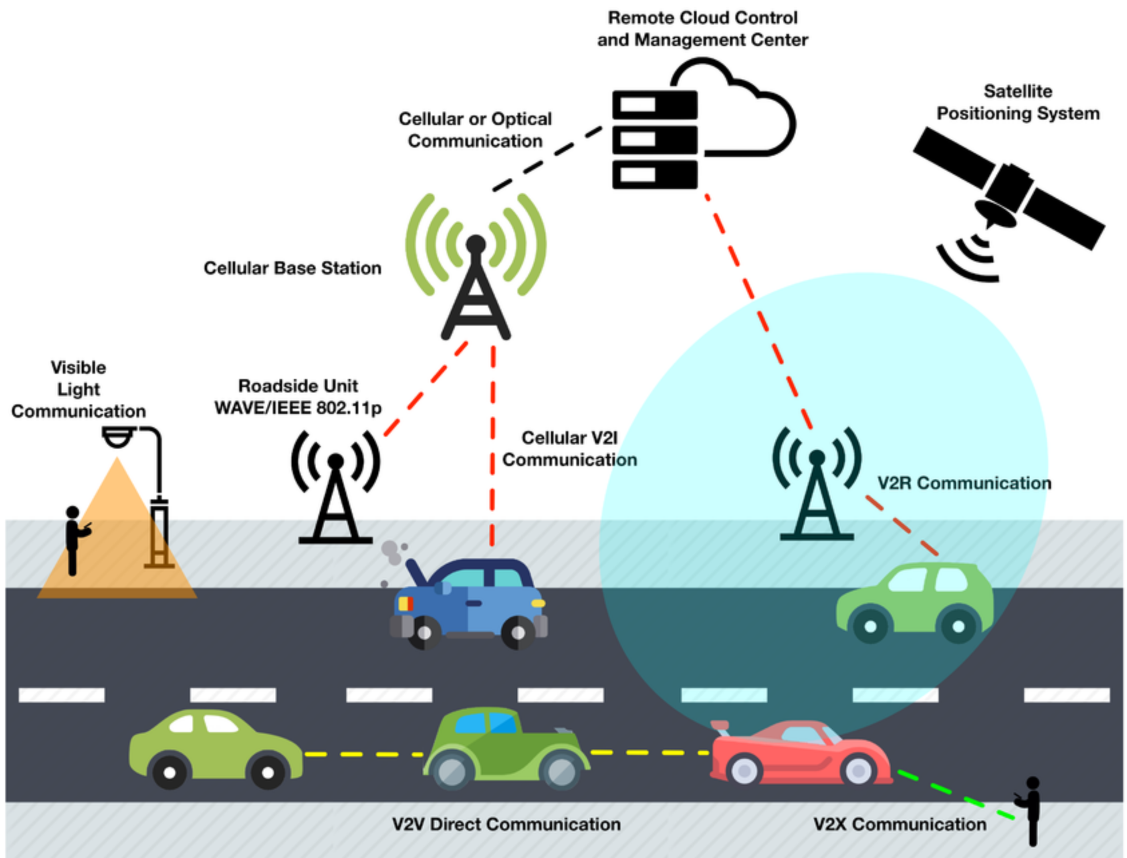

Figure 3: V2V, V2I, and V2X communications [21]

Recent research has focused on enabling technologies for cooperative perception, including wireless communication standards, cloud computing frameworks, and collaborative perception algorithms. Foundational technologies such as IEEE 802.11p/DSRC and Cellular V2X (C-V2X) enable reliable short-range communication for exchanging perception data [23]. Standardization bodies such as ETSI have defined Collective Perception Messages (CPMs) that allow vehicles and RSUs to share detected objects across networks.

Cloud-assisted frameworks like Carcel offload perception computation to scalable infrastructure [24], while cloud-in-the-loop architectures integrate simulation and cloud resources for continuous validation [25]. European initiatives, including the C-Roads platform, have demonstrated the feasibility of RSU–cloud integration for cooperative ITS applications [26].

The evolution of 5G and upcoming 6G technologies promises ultra-reliable, low-latency communication with network slicing and edge computing capabilities [27], [29]. Hybrid approaches combining DSRC, C-V2X, and 5G with intelligent selection algorithms have shown potential for dynamic link optimization [28]. Recent cooperative perception frameworks—Cooper, AutoCast, and Coopernaut—achieve improved detection under occlusion through distributed data fusion and asynchronous sensor collaboration [17], [30], [31]. However, synchronization and real-time alignment of multi-vehicle data remain active research challenges [32], [33].

Despite the rapid progress, several research gaps remain:

1. Achieving sub-100 ms latency with > 99 % uptime across heterogeneous networks is still an open challenge.

2. The compression–network coupling problem is underexplored; few works jointly optimize adaptive video encoding with dynamic communication control.
3. Many frameworks are scenario-specific, lacking generalizability to large-scale deployments.
4. There is limited empirical validation using live RSU video feeds under real-world network variability.

This dissertation addresses these gaps by integrating adaptive compression and hybrid networking into a unified framework, validated through simulation, laboratory, and field experiments.

Research Aim, Objectives, and Hypothesis

As autonomous and connected vehicles advance toward real-world integration, maintaining reliable perception across variable communication conditions is critical. Variations in bandwidth, delay, and packet loss can degrade detection performance and compromise safety. This research addresses the fundamental question: How can real-time video transmission for cooperative perception be optimized to maintain consistent latency and reliability under heterogeneous networks?

Research Aim:

To develop and validate an adaptive framework that minimizes end-to-end latency and maximizes reliability in cloud-based cooperative perception systems by integrating intelligent video compression, hybrid networking, and GPU-accelerated cloud inference.

Research Objectives:

1. Develop an adaptive H.265 compression strategy to dynamically regulate frame rate and bitrate in response to fluctuating bandwidth.

2. Design a hybrid 5G–Starlink communication architecture supporting make-before-break switching and multipath routing.
3. Implement GPU-accelerated YOLOv8 inference pipelines for real-time RSU camera analytics in the cloud.
4. Evaluate latency, throughput, uptime, and perception accuracy using simulation, bench, and field testing.
5. Formulate design guidelines for scalable, cost-efficient cooperative perception deployment.

Research Hypothesis:

Integrating adaptive H.265 compression with hybrid 5G–Starlink networking will yield end-to-end latency below 100 ms and maintain connectivity uptime above 99 %, with minimal impact perception accuracy.

The methodology adopted in this dissertation bridges algorithmic development, empirical experimentation, and statistical validation. The research design was structured around iterative evaluation cycles to ensure reproducibility and real-world applicability.

1. Algorithm Design: Development of an adaptive compression controller and hybrid link manager capable of adjusting parameters based on real-time network telemetry.
2. System Implementation: Integration of RSU cameras, 5G modems, and Starlink terminals into a modular pipeline encompassing sensing, encoding, transport, inference, and telemetry layers.
3. Experimental Evaluation: Controlled simulation and bench testing across varied conditions , followed by live trials on the same bench setup.

4. Statistical Analysis: Analyze the influence of bandwidth, loss, and encoding policy on latency, frame drop, and accuracy.

This methodological structure ensures both quantitative rigor and engineering relevance, connecting theoretical insight with deployment feasibility.

Contributions and Structure of the Dissertation

This dissertation contributes new methods and empirical evidence toward realizing practical, scalable, and reliable cooperative perception systems. Beyond improving isolated subsystems, it introduces an integrated architecture combining adaptive compression, hybrid connectivity, and cloud inference—an advancement crucial for future intelligent transportation infrastructures.

Conceptual Contributions:

- A unified adaptive transmission framework that coordinates compression control with hybrid networking for real-time cooperative perception.
- A novel 5G–Starlink hybrid architecture enabling seamless link switching and multipath resilience under variable network conditions.

Empirical Contributions:

- Experimental validation confirming sub-100 ms latency and > 99 % uptime in realistic scenarios.
- Quantitative characterization of compression–network trade-offs across simulation, bench, and field studies.

Methodological Contributions:

- Development of adaptive controllers and benchmarking protocols for future Cooperative Perception System (CPS) research.

- Establishment of reproducible, cloud-integrated test frameworks linking academic research to real deployment.

Structure of the Dissertation:

- Chapter 2: Literature Review—overview of cooperative perception, cloud/edge computing, compression methods, and communication systems.
- Chapter 3: Cloud-based roadside perception prototype and feasibility evaluation.
- Chapter 4: Comparative study of cloud platforms for RSU-based real-time perception systems.
- Chapter 5: Predictive failover and adaptive video streaming for low-latency RSU-to-Cloud cooperative perception over hybrid 5G–Satellite connectivity.
- Chapter 6: Conclusion and future work.

Collectively, these contributions advance the scientific and engineering foundation required for deployable, cloud-integrated Level-5 autonomous systems, establishing a pathway toward resilient, real-time, and scalable cooperative perception frameworks.

CHAPTER TWO

LITERATURE REVIEW

Introduction

This chapter reviews the state of the art in cooperative perception, cloud and edge computing for intelligent transportation, real-time video compression, and hybrid communication networks. The goal is to contextualize the proposed research within existing knowledge and identify unresolved challenges that limit the deployment of reliable cloud-based cooperative perception systems. The discussion begins with the theoretical foundations of cooperative perception, proceeds to computing architectures that support real-time data analytics, examines the role of adaptive compression in efficient video streaming, and concludes with a survey of hybrid 5G–satellite networking approaches relevant to low-latency connected mobility.

Foundations of Cooperative Perception

Cooperative perception (CP) enables vehicles and infrastructure nodes to extend their sensing capabilities beyond line of sight by exchanging sensor information through Vehicle-to-Everything (V2X) communication [17], [18]. By sharing object-level or feature-level data, multiple agents can construct a richer, more accurate environmental model than any single vehicle could achieve alone. This principle improves detection of occluded pedestrians, vehicles approaching intersections, and dynamic hazards such as sudden braking events.

Recent research has formalized CP as a distributed perception-fusion problem. Benchmarks such as TUMTraf V2X provide standardized datasets for evaluating cooperative fusion methods [37]. Architectures including Cooper, AutoCast, and Coopernaut have demonstrated measurable gains in detection accuracy and robustness through shared perception among vehicles [17], [30], [31]. These frameworks differ in their synchronization strategy and fusion granularity: some perform early fusion of sensor features, while others employ late fusion of detected objects.

Despite substantial progress, several open issues persist. The first concerns asynchrony among heterogeneous sensors and communication delays, which complicate time alignment [32]. Second, bandwidth efficiency becomes critical as the number of participating agents increases. Third, many experimental validations are simulation-based and lack exposure to real network impairments. These limitations underscore the need for empirically validated CP frameworks that explicitly account for network variability—a central focus of this dissertation.

Cloud and Edge Computing in Intelligent Transportation

The integration of cloud and edge computing into ITS architectures has transformed how perception and decision-making tasks are distributed. Cloud computing offers elastic scalability for compute-intensive workloads such as deep-learning inference, while edge computing reduces latency by processing data closer to the source [34]. Hybrid cloud-edge designs aim to balance these advantages by dynamically partitioning tasks between roadside units (RSUs), edge gateways, and centralized cloud servers.

The Carcel architecture exemplifies early cloud-assisted perception, where vehicles offload sensor data to the cloud for large-scale processing [24]. Subsequent work introduced cloud-in-the-loop frameworks that integrate simulation environments with real cloud resources for repeatable testing [25]. European deployments such as C-Roads validated RSU-to-cloud data pipelines for Cooperative-Intelligent Transport Systems (C-ITS) [26].

Recent advances in edge intelligence combine on-site inference with cloud coordination. Shi et al. [34] identified latency, energy consumption, and network heterogeneity as key challenges for scalable edge computing. Modern architectures exploit GPU acceleration, container orchestration, and lightweight telemetry to achieve deterministic performance. Yet, practical CPS deployments still struggle to meet end-to-end latency requirements (< 50 ms) when network congestion or variable backhaul delay occurs. These challenges motivate adaptive data-transmission strategies that intelligently trade compression, quality, and computation—an approach pursued in this research.

Video Compression for Real-Time Systems

Video streams from RSU cameras constitute the primary data source for cooperative perception. Efficient compression is therefore essential to maintain quality under constrained bandwidth. Traditional codecs such as H.264/AVC and H.265/HEVC employ block-based motion compensation to reduce redundancy, with HEVC achieving approximately 30–50 % better compression efficiency at the same visual quality [35]. However, static encoding configurations are inadequate for dynamic vehicular networks where available throughput fluctuates rapidly. Adaptive compression techniques address this issue by adjusting parameters such as frame rate, resolution, and constant-rate factor (CRF) in real time. Li et al. [37] demonstrated adaptive frame-rate control for vehicular streaming that maintained target latency under bandwidth variation. Region-of-interest

(ROI) coding has also been explored to prioritize critical image regions for autonomous driving tasks [38].

While next-generation codecs such as AV1 and VVC promise further efficiency, their computational complexity can introduce additional encoding delay, limiting real-time suitability without specialized hardware acceleration. Thus, H.265 (HEVC) remains the most practical choice for real-time CPS applications due to its hardware support and favorable balance between efficiency and encoding latency. Nevertheless, achieving consistent low-latency transmission requires coupling compression control with live network feedback—precisely the adaptive mechanism developed in this dissertation.

Hybrid Communication Networks

Reliable, low-latency communication is the backbone of cooperative perception. The introduction of 5G NR V2X offers features such as Ultra-Reliable Low-Latency Communication (URLLC), network slicing, and mobile edge computing, which collectively reduce delay and improve throughput [27]. However, even with 5G infrastructure, connectivity gaps persist in rural or obstructed environments. Satellite systems like Starlink, which employ low-Earth-orbit (LEO) constellations, provide a complementary channel with round-trip times of 30–50 ms—substantially lower than traditional geostationary links [40].

Hybrid 5G–satellite communication frameworks exploit these complementary strengths. Jiang et al. [39] analyzed the potential of multi-link vehicular communication for maintaining continuous connectivity under varying signal conditions. Farooq et al. [40] evaluated Starlink's performance for 5G backhaul and identified achievable throughputs exceeding 100 Mbps with minimal packet loss.

Emerging research proposes multipath transport protocols (e.g., QUIC-based approaches) and make-before-break switching to dynamically route data over the optimal link. Such hybrid solutions enhance robustness but also introduce complexity in flow control and synchronization. A systematic evaluation of these trade-offs under real vehicular workloads is still lacking, motivating the hybrid 5G–Starlink design explored in this dissertation.

This chapter surveyed foundational and contemporary research relevant to cloud-based cooperative perception. The review reveals that while cooperative perception, cloud/edge computing, and 5G communication each have matured individually, their integrated optimization remains largely unexplored. Current systems struggle to sustain real-time performance when confronted with fluctuating bandwidth, multipath interference, or large-scale RSU deployments.

The dissertation builds upon these insights by introducing a unified, adaptive framework that couples compression control with hybrid network management, evaluated through simulation, bench, and field trials. The next chapter presents the detailed methodology, algorithms, and experimental design used to implement and validate this framework.

CHAPTER THREE

CLOUD-BASED ROADSIDE PERCEPTION PROTOTYPE AND FEASIBILITY EVALUATION

Introduction

Cooperative perception is a foundational component of intelligent transportation systems (ITS) and connected autonomous vehicles (CAVs). By extending perception capabilities beyond the line of sight of onboard vehicle sensors, infrastructure-assisted perception improves situational awareness and supports safer, more efficient vehicle operation. While significant research has explored cooperative perception using vehicle-mounted sensors and V2V communication, fewer studies have investigated the use of roadside unit (RSU) cameras as primary sensing elements in cloud-assisted perception frameworks.

This chapter presents the first experimental phase of this dissertation: the design and evaluation of a cloud-based cooperative perception system that performs real-time object detection using a single RSU camera, a 5G cellular link, and a cloud GPU server. The objective of this experiment was to evaluate whether real-time video captured at an intersection can be transmitted over 5G to the cloud, processed using a lightweight deep-learning model (YOLOv8 Small), and returned to a connected vehicle with latency low enough to support cooperative decision-making. A controlled bench setup was used to simulate a roadside installation, allowing systematic evaluation of latency, frame drops, and inference speed—key indicators of feasibility for cloud-based roadside perception.

The chapter describes the system architecture, the bench-test environment, the software pipeline, the training procedure, and performance results. These findings establish a baseline for the expanded evaluations presented in later chapters.

System Overview

The prototype cooperative perception system consists of three main components:

1. A roadside camera capturing live video of an intersection.
2. A 5G communication link transmitting the video stream to the cloud.
3. A cloud computing platform performing object detection and returning results to a connected vehicle.

Figure 4 illustrates the conceptual architecture. The RSU camera streams real-time footage to a cloud server, where frames are processed to detect vehicles, pedestrians, cyclists, and traffic signals. The processed detections, including bounding boxes and class labels, are then forwarded to a connected vehicle, enabling earlier and more informed decision-making before objects enter the car's onboard sensor field. This architecture demonstrates the core promise of cooperative perception: infrastructure-assisted awareness that compensates for occlusions, blind spots, and limited onboard sensing range.

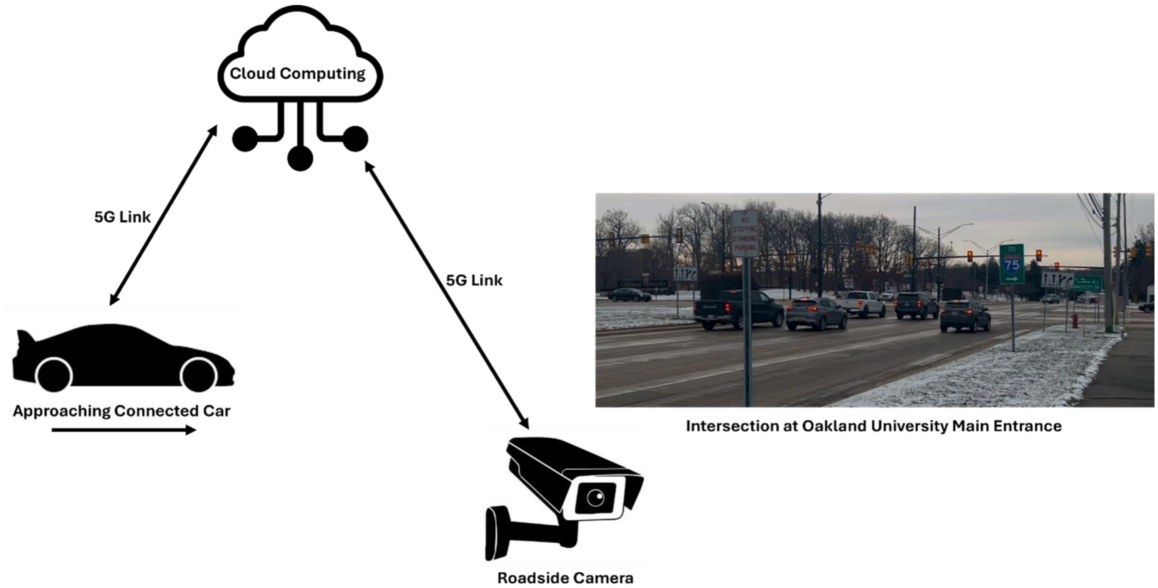

Figure 4. Example real-world scenario for the cloud-based roadside perception system

Development Approach and Bench-Test Setup

Due to constraints in performing continuous roadside experiments, a controlled bench setup was created to systematically evaluate the system's performance under repeatable conditions. Intersection footage was recorded at the Oakland University main entrance—an area with dense traffic and frequent pedestrian crossings. These recordings were later replayed through the bench setup.

Figure 5 shows the main elements of the bench environment:

- A monitor playing back prerecorded intersection footage.
- A 2 MP RSU-style camera positioned to re-capture the footage.
- A 5G modem providing cellular uplink.
- A roadside computer responsible for transmitting frames to the cloud.
- A second computer acting as a connected vehicle receiving processed detections.

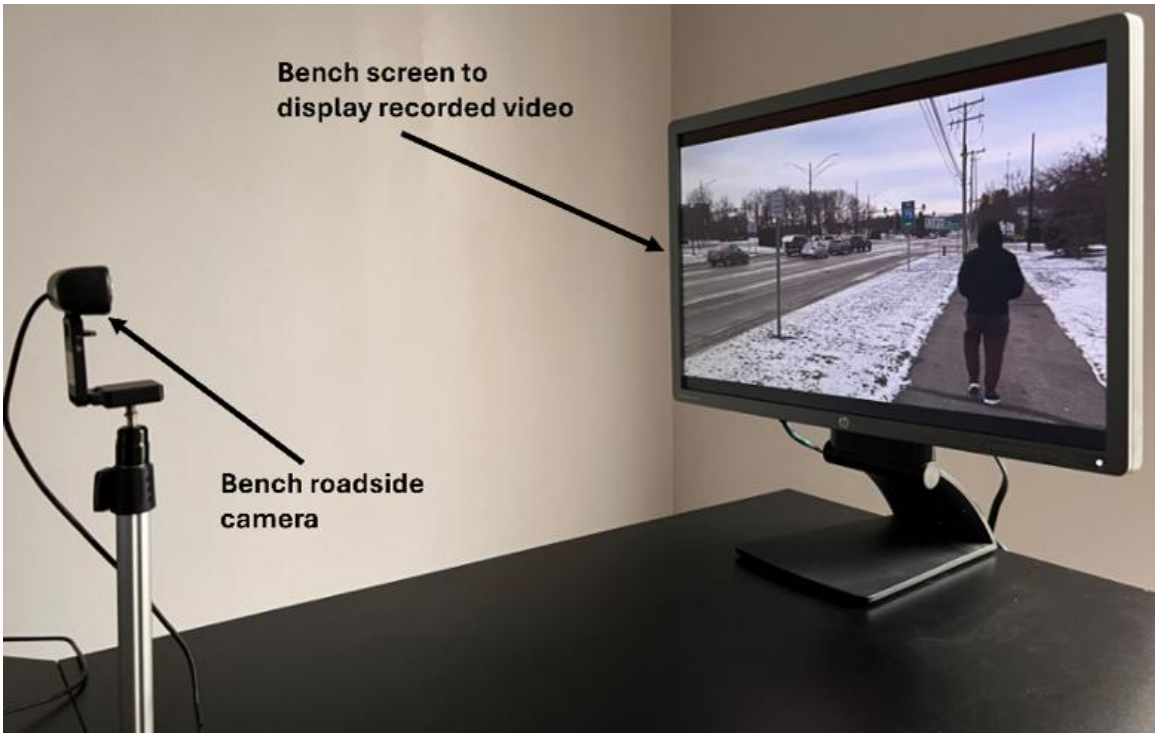

Figure 5. Cloud-based roadside perception system bench setup

The roadside computer streamed the video to the cloud using a live HTTP tunnel (ngrok), simulating a real camera connecting through a public cellular network. The connected-vehicle computer received detection results returned from the cloud, replicating the I2V channel of a real deployment.

This setup enabled precise measurement of:

- Network latency
- Frame transmission reliability
- Processing time per frame
- End-to-end production feasibility

Software Architecture

The software architecture consists of three coordinated subsystems: (1) video streaming, (2) cloud-based object detection, and (3) result visualization.

Figure 6 illustrates the pipeline:

1. **Roadside Acquisition and Streaming**

   The RSU camera output is captured using OpenCV on the roadside computer. Frames are transmitted to a remote Flask server running on Google Colab using an encrypted ngrok tunnel.

2. **Cloud Processing**

   The cloud environment (Google Colab with an NVIDIA T4 GPU) receives incoming frames, decodes them, and prepares them for inference. YOLOv8 Small is used for object detection due to its lightweight architecture and suitability for real-time cloud inference.

3. **Visualization and Output**

   The detection results, including bounding boxes and labels, are streamed to the connected-vehicle computer for real-time visualization and potential integration with vehicle control logic.

This architecture demonstrates a complete infrastructure–cloud–vehicle perception loop.

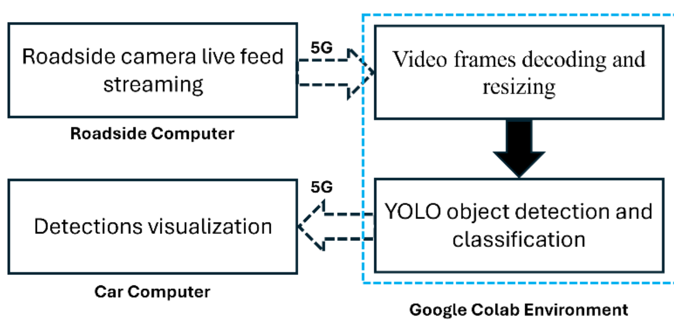

Figure 6. Software architecture of the cloud-based roadside perception system

Object Detection Pipeline

The perception subsystem is built around YOLOv8 Small, a compact convolutional neural network designed for real-time object detection [41]. The following steps were executed to prepare the model for RSU-based inference:

1. **Dataset Collection**

   A dataset of 3,870 intersection images was collected from the Oakland University site to ensure domain-specific performance.

2. **Annotation and Splitting**

   Images were labeled for vehicles and pedestrians.

   - 80% training
   - 10% validation
   - 10% testing

3. **Model Training**

   The YOLOv8 Small model was fine-tuned for 20 epochs on the T4 GPU. Training emphasized sensitivity to occluded and distant objects, which frequently occur in intersection scenes.

4. **Validation and Testing**

   The model was validated on unseen samples, demonstrating high accuracy for both vehicles and pedestrians.

The final inference workflow is summarized in Figure 7 and includes frame resizing, normalization, model inference, non-maximum suppression, and visualization.

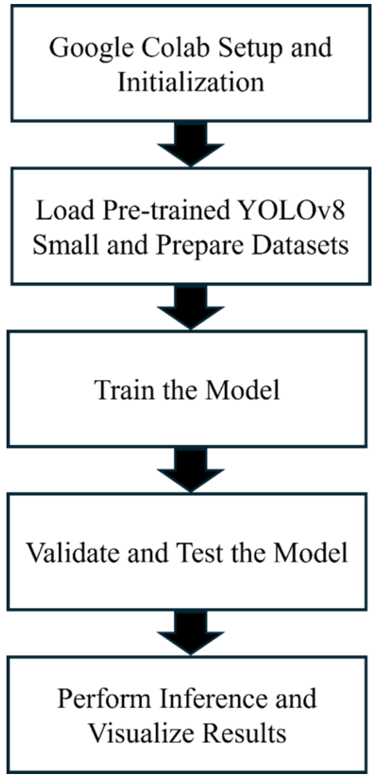

Figure 7. Object detection and classification workflow

Experimental Results and Discussion

A continuous 24-hour bench test was conducted to assess the feasibility of the cloud-based perception system. Key performance metrics included:

- Streaming latency
- Frame drop rate
- Object detection processing time
- Overall reliability over long durations

Frame timestamps were recorded both at the sender (roadside computer) and the receiver (cloud server). Unique frame IDs were used to detect dropped frames. YOLOv8 processing times were logged within the cloud notebook.

Latency

The system achieved an average end-to-end streaming latency of 70 ms, which is promising for real-time cooperative perception. This indicates that a 5G uplink combined with lightweight cloud inference can support sub-100 ms response times.

Frame Drops

An average of 5 dropped frames per second was recorded. This was attributed to variations in cellular bandwidth and temporary congestion. Applying video compression (H.264/H.265) or dynamic frame-rate adaptation would reduce data load and improve reliability.

Processing Speed

YOLOv8 inference achieved a mean processing time of 12.8 ms per frame, confirming that cloud-hosted GPU inference is sufficiently fast for real-time RSU applications.

Qualitative Detection Accuracy

Figure 8 shows an example detection result from the experiment. The system consistently identified vehicles and pedestrians with fewer than 5% false positives, demonstrating high detection reliability under typical intersection conditions.

Summary of Experimental Findings

Table 1 results confirm that the system can operate in real time with promising accuracy and latency characteristics. Although frame drops remain an issue in some conditions, they are solvable through improved compression and streaming techniques.

Table 1: Cloud-based roadside perception system performance testing results

Performance metric	**Measured values (average)**
Video streaming latency	70 ms
Dropped frames	5 fps
Algorithms processing speed	12.8 ms

Figure 8. Sample object detection and classification results

Conclusion

This chapter presented a cloud-based cooperative perception system that leverages roadside cameras, 5G communication, and cloud GPU processing to detect objects in real time. Experimental results demonstrate that:

- Cloud-hosted perception is feasible with <100 ms latency.
- YOLOv8 Small provides sufficiently fast inference on a cloud GPU.
- RSU-to-cloud streaming over 5G is reliable but can benefit from video compression.

- The system can effectively complement onboard vehicle perception, especially in occluded or complex urban scenarios.

The insights gained from this prototype informed the extended multi-cloud evaluation presented in Chapter 4 and motivated the optimized transmission framework developed in Chapter 5.

CHAPTER FOUR

COMPARATIVE STUDY OF CLOUD PLATFORMS FOR RSU-BASED REAL-TIME PERCEPTION SYSTEMS

Introduction

The growing complexity of urban traffic environments and the increasing demand for full vehicle autonomy have accelerated the development of cooperative perception systems. These systems extend the sensing capabilities of autonomous vehicles (AVs) by leveraging external infrastructure such as Roadside Units (RSUs) to share sensor data including camera feeds, LiDAR, and radar data. RSU-mounted cameras, in particular, offer a cost-effective way to augment vehicle perception by providing coverage of blind spots, occluded objects, and intersections beyond the line of sight of onboard sensors [42].

Achieving real-time performance in such cooperative systems demands significant computing power, high-bandwidth connectivity, and reliable data processing frameworks. Cloud computing provides a scalable and flexible solution, enabling centralized processing of sensor data and dynamic allocation of computational resources. However, selecting an appropriate cloud provider is a non-trivial task, as different platforms offer varying levels of performance, cost efficiency, and ease of integration. Although several initiatives have integrated cooperative perception with 5G and edge-cloud infrastructures [43]-[44], few studies have directly benchmarked mainstream cloud platforms for RSU-based real-time video analytics.

This chapter evaluates the feasibility and performance of Google Colab Pro+, Microsoft Azure, and Amazon Web Services (AWS) for a real-time perception use case. A unified YOLOv8-Small object detection pipeline was deployed on all platforms using the same dataset and settings to ensure a fair comparison. The analysis focuses on key metrics such as latency, frame processing rate, detection accuracy, cost, scalability, and reliability. By presenting these findings within a broader dissertation context, this chapter provides practical guidance for selecting cloud platforms for intelligent transportation systems.

To situate this comparative study within existing literature, it is important to note that cooperative perception has become a central theme in connected autonomous driving research. It enables vehicles to perceive beyond their sensor range by sharing environmental information with nearby vehicles or infrastructure. Recent work including Chen et al. [17] and Wang et al. [45] demonstrates how RSU camera feeds can be fused with onboard sensor data to enhance detection and reduce false negatives. Similar advances in cloud-based perception, such as Carcel [24] and EdgeCloudFusion [46], highlight the growing potential of remote processing in safety-critical environments.

Despite these advances, direct comparisons of commercial cloud services remain limited. Given the long-term cost implications and deployment challenges for RSU-based perception, understanding platform-specific behavior is essential. This chapter addresses that gap and builds toward the system optimization strategies explored in later chapters.

System Architecture

The cooperative perception system evaluated in this chapter is structured around three core components: RSU cameras, cloud-based perception nodes, and the communication layer enabling data transfer between roadside infrastructure and the cloud. Figure 9 illustrates the overall architecture of the system.

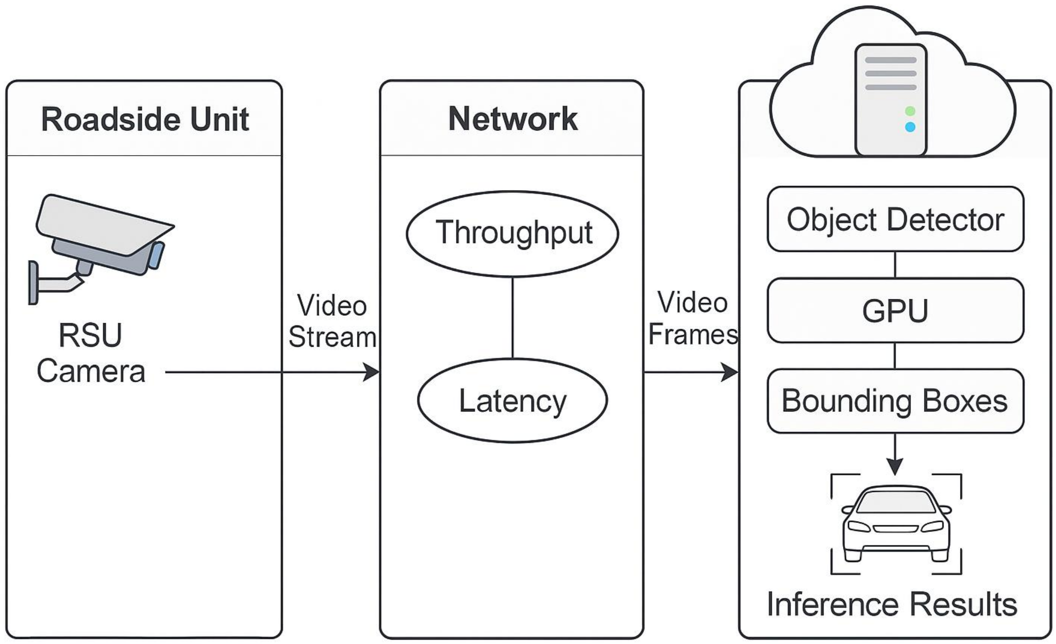

Figure 9: System architecture of the cloud-based cooperative perception pipeline using RSU cameras.

Data Source and Camera Simulation

A dataset containing 5,000 labeled images was collected at the Oakland University main entrance to replicate realistic roadside environments. The images, representing diverse illumination and congestion levels, were replayed as a continuous video stream to emulate live RSU cameras. All frames were standardized to a resolution of 720p at 30 fps to ensure consistent evaluation across platforms.

Perception Pipeline

Each incoming video frame was transmitted to a cloud instance for processing using the YOLOv8-Small object detection model. Detection outputs—including bounding boxes, confidence scores, and class labels—were logged and analyzed. The perception pipeline consisted of three major stages:

1. Frame Acquisition: Capturing video frames using OpenCV and preparing them for streaming.
2. Cloud Processing: Performing GPU-accelerated inference using YOLOv8-Small with batch size = 1 to preserve real-time behavior.
3. Result Handling: Logging detections, timestamps, and processing times for later evaluation.

Using identical model weights and inference settings across all cloud platforms ensured a controlled and fair comparison.

Cloud Platforms

To evaluate how commercial cloud infrastructures affect real-time cooperative perception, three environments were selected:

- Google Colab Pro+: Paid tier offering priority access to T4 GPUs, extended runtime, and higher memory.
- Microsoft Azure NV-Series VM: Dedicated GPU virtual machine integrated with Azure Media Services for consistent streaming.
- AWS EC2 G4dn.xlarge: NVIDIA T4 GPU instance with high I/O throughput, S3-based data storage, and Lambda automation.

This multi-platform selection enables a well-rounded comparison among today's leading GPU-enabled cloud solutions.

Experimental Setup

Each platform was configured to process the same dataset, ensuring identical test conditions. Performance metrics included latency, frame processing rate (FPS), cost, scalability, detection accuracy (mAP@0.5), resource utilization, setup overhead, robustness under adverse conditions, and recovery time after failure. Python scripts using OpenCV for frame capture, YOLOv8-Small for inference, and precise timestamping for performance logging formed the core of the testbed. Figure 10 illustrates the experimental setup and testing framework.

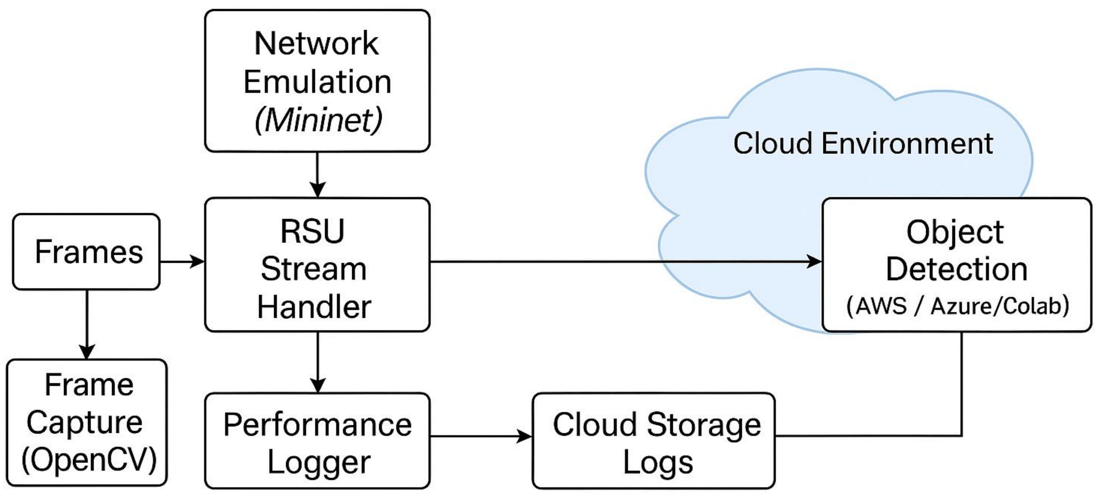

Figure 10: Experimental setup and testing framework for cloud-based RSU video streaming and performance measurement.

Network Emulation

Network conditions were emulated using Mininet to approximate 4G- and 5G-like characteristics. Although Mininet does not implement full 4G/5G protocols, it can reproduce representative delay and bandwidth patterns.

Two link profiles were used:

- 4G-like: 20 ms RTT, 50 Mbps

- 5G-like: 5 ms RTT, 500 Mbps

These controlled profiles help isolate the effects of bandwidth and latency on total perception delay, independently of wireless-network variability.

Test Procedure

The evaluation followed a consistent methodology:

- Baseline Test: 1,000 frames
- Scalability Test: 1–100 simulated RSU streams
- Robustness Test: Images from rain and night scenes
- Repetition: All experiments repeated three times; averages reported

The following sections summarize these results.

Results and Discussion

The experimental evaluation compares Google Colab Pro+, Microsoft Azure, and AWS across latency, scalability, detection accuracy, cost, robustness, and reliability. Unless otherwise noted, Tables 3–10 summarize averages across scalability tests ranging from 1 to 100 streams. Table 2 presents the single-stream baseline.

Figure 11 shows a representative output from the YOLOv8-Small model applied to an RSU camera feed at the Oakland University intersection. This visual example supports the quantitative findings discussed throughout the chapter.

Figure 11: Example detection output from the YOLOv8-Small model using an RSU camera feed. Bounding boxes show detected vehicles under overcast conditions.

Baseline Performance

Table 2 summarizes single-stream performance in terms of latency, throughput (FPS), and cost. AWS achieved the lowest average latency (75 ms) and highest throughput (13 FPS). Azure provided moderate latency (90 ms) and stable throughput (11 FPS). Colab Pro+ exhibited higher latency (120 ms) and lower throughput (8 FPS) due to shared GPU scheduling and limited session persistence capabilities.

Table 2: Baseline performance metrics

Platform	**Avg. Latency (ms)**	**FPS**	**Cost / Hour**	**Cost / 1 000 Frames**
Google Colab Pro+	120	8	≈ \$0.07	≈ \$0.01
Azure	90	11	≈ \$1.20	≈ \$0.11
AWS	75	13	≈ \$1.60	≈ \$0.09

These results establish a clear baseline for determining platform suitability across different deployment scales.

Scalability Analysis

Table 3 shows performance trends as the number of RSU streams increases. AWS maintained high throughput with only moderate degradation at 100 streams. Azure remained reliable up to 50 streams but experienced throughput loss as CPU utilization increased. Colab Pro+ became unstable beyond 10 streams due to runtime resets and GPU session limitations.

Table 3: Scalability under concurrent streams

Streams	Colab Pro+ (FPS)	Azure (FPS)	AWS (FPS)	Colab Δ (%)	Azure Δ (%)	AWS Δ (%)
1	8	11	13	0	0	0
10	6	10	12	25	9	8
25	N/A	8	11	Session reset	27	15
50	N/A	6	10	Session reset	45	23
100	N/A	N/A	10	Session reset	N/A	23

These observations indicate that:

- AWS is well-suited for large-scale city deployments.
- Azure is appropriate for medium-scale deployments.
- Colab Pro+ is best suited for academic experimentation.

Network Latency and Bandwidth Impact

Mininet-based emulation results are shown in Table 4. Under 5G-like conditions, AWS achieved sub-100 ms total latency (82 ms), meeting cooperative driving thresholds (<100 ms). Azure also performed well at 98 ms. Colab Pro+ remained higher at 130 ms. These findings confirm the critical role of low-RTT 5G connectivity combined with cloud-based GPU inference in achieving real-time performance.

Table 4: End-to-End latency (ms) under emulated 4G and 5G network conditions

Network	Colab Pro+	Azure	AWS
4G-like	245	185	162
5G-like	130	98	82

Detection Accuracy and Error Rates

Table 5 summarizes detection accuracy. AWS achieved the highest mAP (0.76) and lowest false-detection counts. Azure achieved slightly lower accuracy (0.74 mAP), while Colab Pro+ experienced more variability due to GPU context switching and frame timing inconsistencies.

Table 5: YOLOv8 detection performance

Platform	mAP @ 0.5	False Pos / 1000	False Neg / 1000
Colab Pro+	0.72	23	18
Azure	0.74	19	15
AWS	0.76	16	12

Resource Utilization Details

Table 6 shows resource usage. AWS displayed the most efficient resource utilization, followed by Azure. Colab Pro+ exhibited near-saturation levels of CPU and GPU during single-stream testing, further explaining its limited scalability.

Table 6: Resource usage (single stream)

Platform	CPU (%)	GPU (%)	Memory (%)
Colab Pro+	82	94	89
Azure	68	83	71
AWS	61	78	66

Cost Breakdown and Long-Term Analysis

Table 7 summarizes the cost implications of continuous deployment. Although AWS incurs a slightly higher monthly cost, it also delivers materially better throughput, making it more cost-effective on a per-frame basis. Azure provides a strong balance of performance, cost, and integration features, while Colab Pro+ is not suitable for persistent real-time operation.

Table 7: Detailed cost breakdown (monthly, 24/7 operation)

Component	Colab Pro+	Azure	AWS
Compute	$ 50	$ 864	$ 1152
Storage (1 TB)	0	$ 23	$ 24
Data Transfer	0	$ 150	$ 120
Total / Month	$ 50	$ 1037	$ 1296
Cost / Frame (1 M frames / day)	0	$ 0.034	$ 0.043

Robustness Under Adverse Conditions

Table 8 highlights platform performance under rain and night conditions. AWS again demonstrated the highest robustness. Azure performed comparably, while Colab Pro+ suffered significant performance drops due to session-related variability.

Table 8: mAP under adverse conditions

Condition	Colab Pro+	Azure	AWS
Clear	0.72	0.74	0.76
Rain	0.58	0.62	0.68
Night	0.49	0.55	0.54

Setup and Configuration Overhead

Table 9 shows deployment time for each platform. Colab Pro+ required the least time to set up and run, while Azure's integration with Media Services required more configuration. AWS struck a middle ground with relatively straightforward setup once automation was enabled.

Table 9: Deployment time (Hours)

Task	Colab Pro+	Azure	AWS
Environment Setup	0.5	2.5	1.5
YOLO Integration	0.3	1.0	0.8
Streaming Config	0.2	1.5	1.0
Total Time	1.0	5.0	3.3

Failure and Recovery Scenarios

Table 10 presents reliability metrics. AWS provided the fastest recovery (25 s) and highest uptime (99.5%). Azure followed closely with 45 s recovery and 98% uptime. Colab Pro+ exhibited the lowest uptime due to session-based architecture requiring manual restart, making it unsuitable for safety-critical applications.

Table 10: Reliability metrics

Scenario	**Colab Pro+ Recovery (s)**	**Azure (s)**	**AWS (s)**
Session Timeout	*Manual Restart Required*	45	32
Network Drop	180	28	25
Overall Uptime (24 h)	≈ 70 %	≈ 98 %	≈ 99.5 %

Conclusion

This chapter presented a comparative evaluation of three mainstream cloud platforms for real-time RSU-based cooperative perception. Among the evaluated platforms, AWS consistently delivered the highest performance across latency, throughput, scalability, and reliability. Azure offered a strong balance between cost and performance, while Colab Pro+ proved valuable for algorithm development and testing but unsuitable for continuous or large-scale deployments.

These results provide practical guidance for city planners, transportation engineers, and researchers seeking to design or deploy RSU-assisted perception systems. Furthermore, the analysis conducted in this chapter informs the cloud–network optimization strategies developed in Chapter 5, which build upon these findings to create a more robust and scalable real-time cooperative perception framework.

CHAPTER FIVE

PREDICTIVE FAILOVER AND ADAPTIVE VIDEO STREAMING FOR LOW-LATENCY RSU-TO-CLOUD COOPERATIVE PERCEPTION OVER HYBRID 5G–SATELLITE CONNECTIVITY

Introduction

Cooperative autonomous and connected vehicles rely on perception systems to safely interpret complex and dynamic traffic environments. While onboard sensors such as cameras, LiDAR, and radar provide essential local awareness, their effectiveness is inherently constrained by line-of-sight occlusions, limited sensing range, and adverse environmental conditions. These limitations are particularly pronounced at urban intersections and in dense traffic scenarios, where safety-critical objects may remain occluded until late in the decision-making process. To address these challenges, cooperative perception has emerged as a promising paradigm in which vehicles and infrastructure share sensing information via vehicular communication networks to extend situational awareness beyond the capabilities of individual sensors.

Early cooperative perception research primarily focused on vehicle-to-vehicle (V2V) collaboration using shared point clouds or object-level information, demonstrating improved detection accuracy and extended perception range [17], [47]. Subsequent work introduced scalable and learning-based approaches, such as AutoCast [30] and COOPERNAUT [31], highlighting the potential of cooperative perception for autonomous driving. However, these systems largely rely on V2V communication, assume homogeneous sensing modalities, and emphasize algorithmic fusion rather than deployable end-to-end architectures.

More recently, infrastructure-assisted cooperative perception has gained increasing attention, as roadside sensors can provide complementary viewpoints in occluded or complex environments. Studies using roadside LiDAR or camera systems have shown measurable improvements in detection coverage and robustness [48], [49]. In parallel, cloud-assisted perception frameworks have been proposed to offload computation and enable centralized processing, exemplified by hybrid vehicular–cloud architectures [50] and systems such as Carcel [24]. While these efforts demonstrate the feasibility of cloud-based perception, they typically assume stable communication conditions and fixed encoding configurations, limiting robustness under realistic wireless variability.

The performance of cloud-based cooperative perception systems is fundamentally constrained by communication reliability, uplink bandwidth, and end-to-end latency. Prior studies have shown that image compression, packet loss, and congestion can significantly affect detection accuracy and perception latency [51], [52]. Vehicular video streaming research further highlights the difficulty of maintaining low-latency delivery under time-varying wireless conditions. Surveys on cloud–edge–end video delivery [53] and hybrid V2X communication strategies [28] emphasize the importance of adaptive mechanisms for real-time applications. Nevertheless, most existing works treat communication and perception as largely independent subsystems and do not integrate adaptive communication and video streaming directly into a closed-loop cooperative perception pipeline.

In addition, reliance on terrestrial cellular connectivity alone may be insufficient in rural, remote, or infrastructure-limited environments [54]. Hybrid connectivity approaches that combine cellular and non-terrestrial networks have therefore been proposed to enhance robustness. In particular, Yastrebova-Castillo et al. demonstrated that direct-to-satellite connectivity using low-Earth-orbit (LEO) constellations can improve autonomous driving safety under limited terrestrial coverage [55]. However, existing hybrid connectivity studies primarily evaluate link availability and latency using generic traffic patterns and do not integrate satellite links into real-time cooperative perception systems or assess their impact on end-to-end perception feasibility.

Motivated by these limitations, this paper presents a cloud-based cooperative perception system that explicitly integrates predictive connectivity failover and conditional adaptive video streaming for robust RSU-to-cloud perception. The proposed system adopts a hierarchical adaptation strategy in which dynamic selection between terrestrial 5G and LEO satellite (Starlink) uplinks serves as the primary mechanism for maintaining latency feasibility, while adaptive H.265 video compression is activated only as a last resort when no communication path satisfies real-time constraints. A communication-constrained perception model is introduced to formalize latency feasibility and perception availability and to guide system design. The system is implemented on a Linux-based RSU equipped with an industrial camera, dual uplink interfaces, cloud-hosted GPU inference, and a vehicle-side client. Experimental evaluation combines controlled bench emulation for validation of adaptation logic with long-duration real-network trials over public 5G and satellite connectivity. By treating communication as a controllable and adaptive subsystem rather than a fixed constraint, this work advances the practical deployment of RSU-to-cloud cooperative perception under realistic and heterogeneous network conditions.

Cloud-assisted autonomous driving has been explored as a means to extend perception beyond onboard sensing limitations. Kumar et al. proposed Carcel, a cloud-assisted system in which sensor information from vehicles and static infrastructure is aggregated in the cloud to support safer planning in occluded scenarios, particularly at intersections. Rather than continuously streaming raw video, Carcel employs a request-based strategy that selectively transmits sensor data at varying resolutions to mitigate bandwidth constraints and packet loss. While effective for cloud-assisted planning, Carcel does not consider continuous roadside camera video streaming, adaptive video compression, or hybrid terrestrial–satellite connectivity for real-time cooperative perception, which are central to this work.

Hybrid vehicular–cloud architectures have also been investigated to offload perception workloads from vehicles to remote computing resources [50]. These studies primarily focus on computation placement, scalability, and resource management, but do not integrate adaptive video streaming or multi-network connectivity into the perception pipeline. Consequently, the impact of time-varying communication conditions on end-to-end perception latency and availability remains largely unaddressed.

Infrastructure-assisted cooperative perception leverages roadside sensors to provide complementary viewpoints that mitigate occlusions and extend perception range. Arnold et al. demonstrated improved 3D object detection performance by fusing infrastructure-based LiDAR data with vehicle perception outputs, particularly in intersection scenarios [48]. Similarly, Adl et al. employed roadside fisheye cameras to provide wide-area visual coverage for vehicle-to-infrastructure (V2I)-assisted automated driving [49].

Most infrastructure-assisted approaches transmit processed features or object-level information rather than raw visual data, thereby reducing communication bandwidth requirements. However, this design choice limits flexibility in perception algorithms and precludes centralized GPU-based inference on uncompressed video streams. Moreover, communication is generally treated as a static subsystem, with limited consideration of dynamic adaptation to bandwidth fluctuations, latency variations, or packet loss.

The performance of cloud-based cooperative perception systems is tightly coupled with communication reliability and video transmission quality. Mehrke et al. analyzed the impact of image compression and wireless network impairments on cloud-based object detection accuracy using an emulated communication setup, demonstrating how compression affects perception performance under constrained bandwidth [51]. However, this work does not consider real-time closed-loop adaptation or joint optimization of communication paths and perception latency.

Vehicular video streaming studies further show that packet loss and fluctuating bandwidth can severely degrade real-time application performance. Heo and Kim investigated adaptive video transmission for safety-critical remote driving, focusing on delivering vehicle-mounted camera video to a human operator over dual cellular links under variable network conditions [52]. While these studies demonstrate the benefits of adaptation, they are primarily designed for teleoperation and human-in-the-loop applications rather than autonomous cooperative perception.

Surveys on cloud–edge–end video delivery highlight the effectiveness of adaptive bitrate and encoding strategies for low-latency applications [53]. However, these approaches are largely developed for multimedia streaming and are rarely integrated into closed-loop cooperative perception systems, where video quality directly influences detection accuracy and end-to-end perception latency. Recent cooperative perception surveys similarly note that most existing systems rely on fixed compression configurations and assume stable communication conditions [56], [57].

Advanced vehicular networking technologies, including 5G-V2X and intelligent link selection mechanisms, have been proposed to support cooperative driving and perception applications [28], [58]. Nevertheless, reliance on terrestrial cellular connectivity alone may be insufficient in rural, remote, or infrastructure-limited environments [54]. To address this limitation, Yastrebova-Castillo et al. investigated direct-to-satellite communication using Iridium and Starlink constellations, demonstrating improved autonomous driving safety under limited terrestrial coverage.

Existing hybrid terrestrial–satellite connectivity studies primarily evaluate link availability, latency, and reliability using generic traffic patterns. They do not integrate satellite links into real-time cooperative perception pipelines, nor do they assess the impact of hybrid connectivity on end-to-end perception latency, availability, or detection performance.

In summary, prior research has advanced cooperative perception algorithms, infrastructure sensing, video streaming, and vehicular networking largely in isolation. Cloud-assisted perception systems often assume stable communication, infrastructure-based approaches focus on feature-level fusion, and hybrid connectivity studies do not consider perception data streams or real-time feasibility constraints. In contrast, this work presents a cloud-based cooperative perception system that jointly integrates predictive connectivity failover and conditional adaptive video streaming over hybrid terrestrial and satellite links. By explicitly prioritizing uplink latency feasibility and activating video adaptation only as a last-resort mechanism, the proposed system addresses real-world communication variability within a unified perception pipeline.

System Architecture and Design

This section describes the architecture and design principles of the proposed RSU-to-cloud cooperative perception system. The system is designed to maintain bounded end-to-end perception latency and high perception availability under heterogeneous and time-varying network conditions through hierarchical connectivity management. Predictive connectivity failover serves as the primary mechanism for preserving latency feasibility, while conditional adaptive video compression acts as a last-resort safeguard.

Figure 12 illustrates the end-to-end system architecture. A roadside unit (RSU) equipped with an industrial camera continuously captures video of the traffic environment. The video is encoded using an H.265 pipeline operating at a nominal bitrate of 5 Mbps under normal network conditions, employing the Medium preset for balanced encoding efficiency and high visual fidelity. This configuration serves as the default high-quality operating point when at least one uplink is predicted feasible $\hat{T}_{\mathrm{up}} \leq L_c = 50\,\mathrm{ms}\ L_c$. Adaptive compression is applied only as a last-resort mechanism: when both predicted uplink latencies exceed L_c, the encoder dynamically reduces the target bitrate to the range 2.5-4.0 Mbps to maintain delivery stability at the cost of moderate quality degradation.

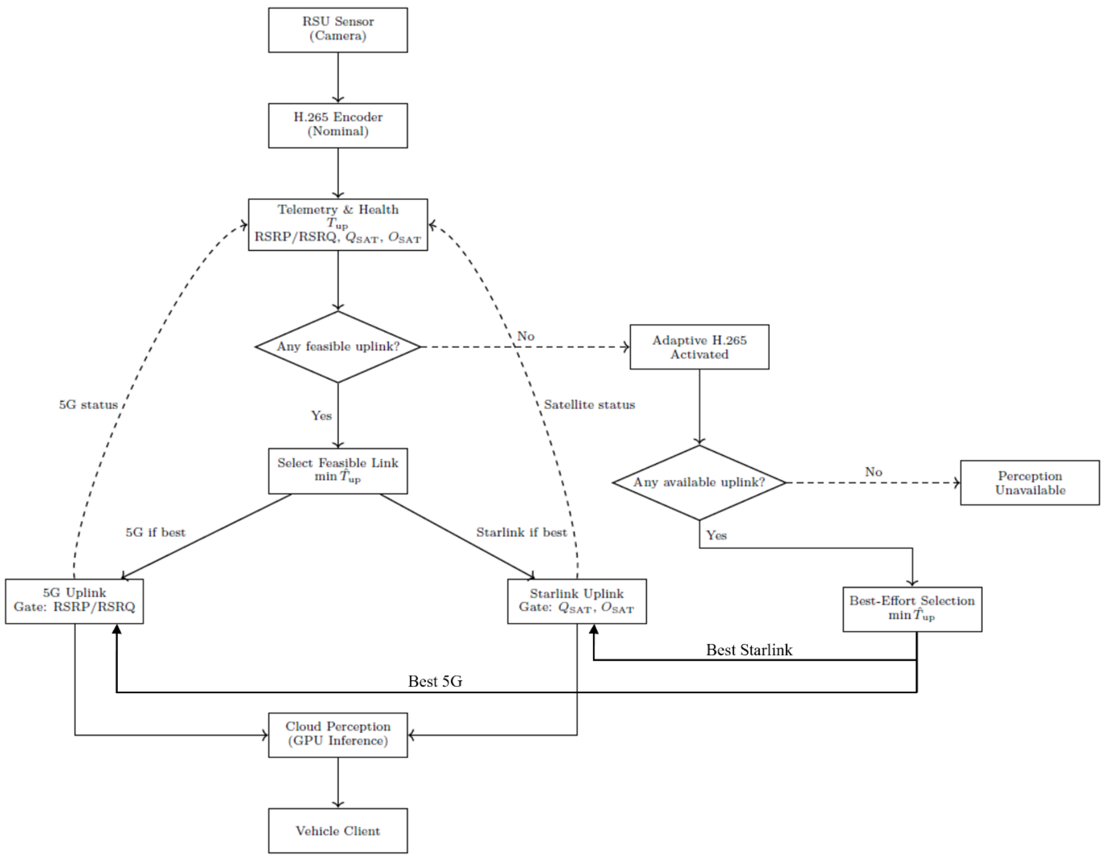

Figure 12. System architecture of the proposed RSU-to-cloud cooperative perception framework. Link feasibility decisions are based on the predicted uplink one-way latency $\hat{T}_{\mathrm{up}} \leq L_c = 50\,\mathrm{ms}$ (a conservative proxy to ensure end-to-end perception latency $L_i(t) \leq L_{\max} = 100\,\mathrm{ms}$), with T_{up} as the dominant variable metric. RSRP/RSRQ (5G) and Starlink quality/obstruction (Q_{SAT}, O_{SAT}) act as availability gates. When at least one uplink is feasible ($\mathcal{F}(t) \neq \emptyset$), the system selects the link minimizing the short-horizon predicted delay $\hat{T}_{\mathrm{up}}$. If no feasible uplink exists ($\mathcal{F}(t) = \emptyset$), last-resort adaptive H.265 compression is activated and a best-effort uplink is selected by minimizing $\hat{T}_{\mathrm{up}}$ among available links to preserve perception availability.

The encoded video stream is transmitted over the selected hybrid uplink (terrestrial 5G or LEO satellite) to a cloud-based perception service that performs GPU-accelerated object detection using YOLOv8-small and returns perception outputs (bounding boxes, class labels, confidence scores) to a vehicle-side client. The cloud and vehicle components operate transparently with respect to the selected communication path, while all adaptation, prediction, and decision logic is centralized at the RSU.

Communication between the RSU and the cloud is supported by hybrid connectivity comprising a terrestrial 5G uplink and a Starlink (LEO satellite) uplink. At any time, only one uplink is actively used for video transmission, while the alternative path remains available for predictive failover or best-effort fallback. This design exploits the low-latency characteristics of 5G when feasible, while preserving perception availability during cellular congestion, coverage gaps, or transient outages through satellite fallback.

The RSU continuously monitors application-layer telemetry for both uplinks, including live one-way uplink latency T_{up}, effective throughput, packet loss rate, and short-horizon latency trends derived from embedded timestamps. Conservative link-status indicators serve as availability gates: RSRP and RSRQ for the 5G uplink, and quality and obstruction indicators (Q_{SAT}, O_{SAT}) for the satellite uplink. Uplink latency is treated as the dominant feasibility metric, while auxiliary indicators are used only to exclude clearly unusable links and to enhance decision stability.

Connectivity selection is the primary mechanism for maintaining perception feasibility. To conservatively ensure end-to-end perception latency $L_i(t) \leq L_{\max} = 100\,\mathrm{ms}$, the system applies a tighter control threshold on predicted uplink latency: $\hat{T}_{\mathrm{up}} \leq L_c = 50\,\mathrm{ms}$, leaving headroom for other delays (capture, encoding, inference, and downlink). When at least one uplink satisfies this constraint ($\hat{T}_{\mathrm{up}} \leq L_c = 50\,\mathrm{ms}$), the RSU selects the feasible uplink with the minimum predicted one-way latency based on short-horizon estimates. Hysteresis and minimum dwell-time constraints are applied to prevent oscillatory switching under transient fluctuations. When neither uplink satisfies the feasibility condition, the system enters degraded mode and activates conditional adaptive H.265 compression to reduce transmission delay. In this mode, the RSU performs best-effort selection by choosing the available uplink that minimizes predicted latency. Adaptive compression is explicitly a last-resort mechanism and is deactivated as soon as at least one uplink re-enters the feasible set.

End-to-end perception latency—measured online from frame capture at the RSU to reception of the corresponding perception output at the vehicle client—serves as the primary indicator of perception feasibility and overall system performance.

The following section formalizes these design principles through a communication-constrained perception model. The model defines latency feasibility using the practical uplink control threshold L_c (to ensure the end-to-end bound $L_{\max}$ is met) and perception availability, thereby guiding predictive failover and conditional adaptation.

Latency Feasibility and Perception Availability Model

This section introduces a lightweight, application-driven model that formalizes latency feasibility and perception availability in the proposed RSU-to-cloud cooperative perception system. The model is not intended for analytical optimization or policy learning; rather, it provides a consistent framework to explain the hierarchical adaptation logic (predictive connectivity failover as primary mechanism, conditional adaptive compression as last-resort) and to interpret the experimental results in Section 5. Consistent with the system design, uplink transmission latency is treated as the dominant and most variable feasibility metric, while auxiliary link indicators serve only as availability gates and decision stabilizers.

End-to-End Perception Latency

For each uplink $i \in \{5\text{G}, \text{SAT}\}$, the end-to-end perception latency is defined as

$$L_i(t) = T_{\text{cap}}(t) + T_{\text{enc}}(t) + T_{\text{up},i}(t) + T_{\text{cloud}}(t) + T_{\text{down}}(t) \tag{1}$$

where T_{cap}, T_{enc}, T_{cloud}, and T_{down} represent RSU frame capture, encoding, cloud inference, and downlink delivery delays, respectively. Under real-world network conditions, uplink transmission latency $T_{\text{up},i}(t)$ dominates variability and therefore governs practical feasibility decisions.

Uplink latency is measured online at the application layer using embedded timestamps recorded immediately before transmission at the RSU and upon reception at the cloud:

$$T_{\text{up},i}(t) = t_{\text{cloud}}^{\text{rx}}(t) - t_{\text{RSU}}^{\text{tx}}(t) \tag{2}$$

These runtime measurements are used to update the short-horizon predictor and to evaluate uplink feasibility.

The system aims to maintain end-to-end perception latency bounded by the maximum allowable value $L_{\max} = 100\ \text{ms}$.

Link Availability and Feasible-Set Definition

Each uplink must also be operationally available, determined by conservative physical- and device-reported gates that exclude only clearly unusable links. For the 5G uplink, RSRP and RSRQ serve as availability gates; for the Starlink uplink, quality and obstruction indicators $(Q_{\text{SAT}}, O_{\text{SAT}})$ are used, as illustrated in Figure 12.

To conservatively ensure that the end-to-end latency $L_i(t) \leq L_{\max} = 100\ \text{ms}$ under realistic conditions, an uplink is considered latency-feasible if its predicted one-way uplink latency satisfies the tighter control threshold

$$\hat{T}_{\text{up},i}(t) \leq L_c = 50\ \text{ms},$$

where $L_c = 50\ \text{ms}$ provides sufficient headroom (approximately 50 ms) for the fixed and less-variable delays ($T_{\text{cap}} + T_{\text{enc}} + T_{\text{cloud}} + T_{\text{down}}$).

The set of uplinks that are both available and latency-feasible is therefore

$$\mathcal{F}(t) = \{i \in \{5\text{G,SAT}\} \mid G_i(t) = 1\ \wedge\ \hat{T}_{\text{up},i}(t) \leq L_c\} \tag{3}$$

where $G_i(t) = 1$ indicates availability. If $\mathcal{F}(t) \neq \emptyset$, real-time cooperative perception is feasible; otherwise, the system enters degraded mode.

Predictive Failover and Best-Effort Adaptation

When $\mathcal{F}(t) \neq \emptyset$, the RSU selects the single uplink $i^*(t)$ that minimizes the short-horizon predicted one-way latency $\hat{T}_{\text{up},i}(t)$ among feasible links:

$$i^*(t) = \arg\min_{i \in \mathcal{F}(t)} \hat{T}_{\text{up},i}(t) \tag{4}$$

The predictor $\hat{T}_{\text{up},i}(t)$ is computed via exponential smoothing of recent measurements:

$$\hat{T}_{\text{up},i}(t) = \alpha T_{\text{up},i}(t) + (1 - \alpha)\hat{T}_{\text{up},i}(t - \Delta t) \tag{5}$$

with smoothing factor $\alpha \in (0,1]$ and decision interval Δt. Hysteresis and minimum dwell-time constraints prevent oscillatory switching under transient fluctuations. Failover is triggered only when elevated latency is attributable to uplink transmission (as selection cannot mitigate capture, inference, or downlink delays).

When $\mathcal{F}(t) = \emptyset$, adaptive H.265 compression activates to reduce $T_{\mathrm{up},i}(t)$, and best-effort selection occurs among available links:

$$i^{\dagger}(t) = \arg\min_{i \in \{5G, SAT\} | G_i(t)=1} \hat{T}_{\mathrm{up},i}(t) \tag{6}$$

Adaptive compression is deactivated immediately upon re-entry of at least one uplink into $\mathcal{F}(t)$.

Perception Availability

Perception availability is defined as the probability that perception outputs are delivered within the end-to-end latency bound under the system's selected uplink:

$$A = \Pr\left(L_{i^*(t)}(t) \leq L_{\max}\right) \tag{7}$$

where $i^*(t)$ is the selected uplink (i^* or $i^{\dagger}$). This metric quantifies the system's ability to maintain usable perception under time-varying conditions and is evaluated in Section 5 for both controlled emulation and real-network trials.

Experimental Evaluation

This section describes the experimental evaluation of the proposed RSU-to-cloud cooperative perception system. A hybrid approach balances reproducibility (controlled emulation) with real-world realism (long-duration public-network trials). All experiments are conducted on the same integrated bench testbed comprising the RSU sensing and encoding pipeline, dual uplink interfaces, a cloud-based perception service, and a vehicle-side client.

Testbed Description

The testbed consists of a Linux-based RSU equipped with an industrial camera and an H.265 video encoder, two independent uplink interfaces (a public 5G cellular modem and a Starlink terminal), a cloud-based GPU inference server running YOLOv8-small on an AWS EC2 G4dn.xlarge instance, and a Linux-based vehicle client. All components operate continuously and autonomously during each experiment.

A dedicated Linux impairment node is deployed as an Ethernet bridge between the RSU and the two uplink terminals as shown in Figure 13. The RSU forwards encoded video traffic to the bridge, which routes packets to the selected uplink according to the RSU-side decision logic. In emulation mode, the bridge applies controlled impairments; in real-network mode, it operates as a transparent bridge.

Evaluation Modes

Two evaluation modes are supported on the same physical testbed. The only difference is whether controlled network impairments are applied at the impairment node.

Controlled Emulation Mode

In controlled emulation mode, deterministic and repeatable impairments (bandwidth constraints, added latency, packet loss, and bursty patterns) are applied using Linux traffic control (tc) with netem and shaping disciplines (htb). Impairments are injected on the egress interface of the currently selected uplink. Each scenario runs for a fixed duration and is repeated across multiple trials to validate the correctness, stability, and triggering behavior of the latency-feasibility logic, predictive failover, hysteresis, and conditional adaptive compression, including edge cases such as simultaneous degradation of both uplinks. Emulation results are used solely for logic validation and do not claim real-world performance.

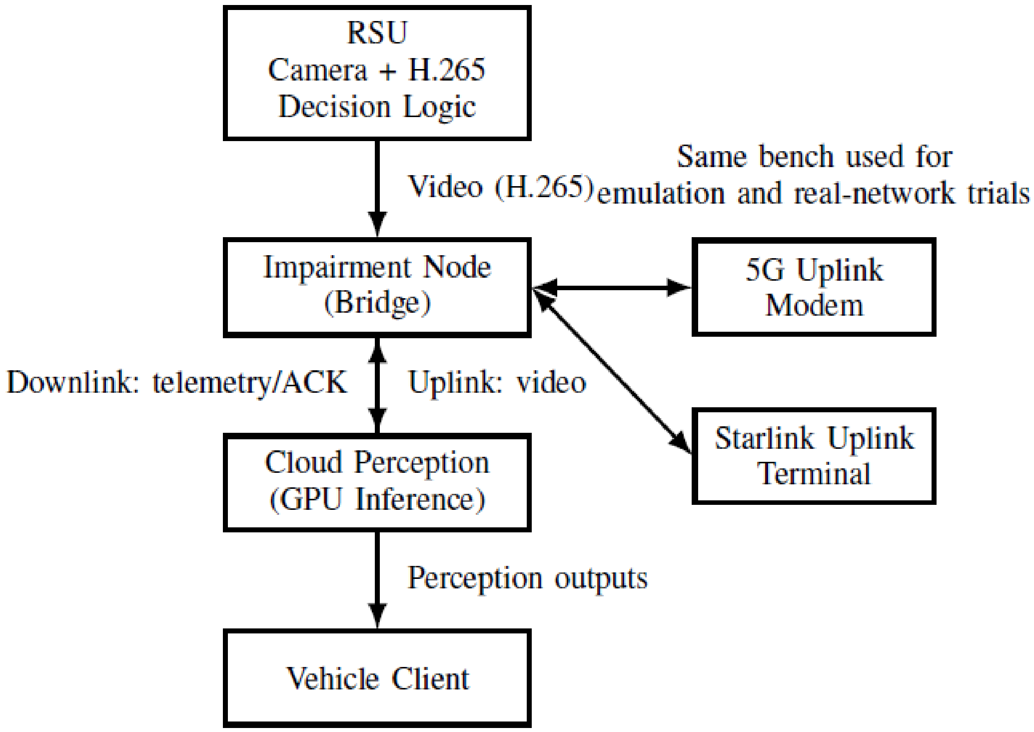

Figure 13. Integrated bench testbed used for both controlled emulation and real-network trials. The RSU transmits H.265-encoded video through an impairment node (transparent bridge in real-network mode; traffic-conditioned in emulation mode) over either the 5G or Starlink uplink. The cloud perception service returns telemetry/ACK messages for online latency estimation and uplink selection, while perception outputs (bounding boxes, classes, confidences) are delivered to the vehicle client.

Real-Network Trial Mode

In real-network mode, all traffic control rules are disabled, and the impairment node operates transparently. The system streams over public 5G cellular and Starlink connectivity under naturally occurring variability. Seven independent 24-hour continuous trials are conducted on separate days. Radio and link-status context is logged alongside application-layer metrics, including 5G RSRP/RSRQ and Starlink quality/obstruction indicators (Q_{SAT}, O_{SAT}).

Ablation Configurations

To isolate the contribution of each adaptation component, three configurations are evaluated during the real-network trials:

- 5G-only baseline: Starlink uplink disabled.
- Failover-only: Hybrid uplinks enabled, adaptive H.265 compression disabled.
- Full system: Hybrid uplinks with conditional adaptive H.265 enabled.

This structure enables direct comparison of perception latency, availability, and robustness under identical hardware and deployment conditions.

Measurements and Metrics

All measurements are collected online at the application layer. Uplink one-way transmission latency is computed using timestamps recorded at the RSU immediately before transmission and at the cloud upon reception. End-to-end perception latency is measured from RSU frame capture to reception of the corresponding perception output at the vehicle client. Perception availability is defined as the fraction of perception outputs delivered within the latency bound $L_{\max}$. Additional logged metrics include effective uplink throughput, packet loss rate, and cloud inference latency. Link-status indicators provide contextual information but are not used as primary performance metrics.

Short-horizon uplink delay estimates for predictive failover are computed using exponential smoothing with decision interval $\Delta t = 50\,\mathrm{ms}$ (matching the system control loop period) and smoothing factor $\alpha = 0.8$ (selected to provide strong responsiveness to rapid uplink changes while hysteresis and minimum dwell-time constraints ensure stability against transient spikes). These parameters are held constant across all experiments to maintain consistency and enable direct comparison of results.

Results and Performance Evaluation

This section presents the experimental results of the proposed RSU-to-cloud cooperative perception system. The evaluation is divided into three parts: (i) controlled bench emulation to validate adaptation logic under repeatable stress conditions, (ii) long-duration real-network trials over public 5G and Starlink connectivity to assess robustness under natural variability, and (iii) perception-level performance using representative object detection outputs from the deployed RSU camera.

Controlled Emulation Results

Controlled emulation experiments validate the hierarchical adaptation logic under well-defined, repeatable impairment scenarios. These tests are not intended to represent real-world performance but to demonstrate correct feasibility evaluation, predictive failover behavior, switching stability, and recovery under deterministic network stress. Only uplink delay impairments are simulated in the presented results, with all signal indicators, predictions, control states, and perception metrics measured online from the running system.

Latency Feasibility and Hierarchical Adaptation

Figure 14 shows a representative 600-second emulation isolating higher-layer impairments as the sole source of latency variation. Subfigure (a) confirms excellent and stable 5G radio conditions, with RSRP fluctuating mildly between approximately -72 to -80 dBm and RSRQ between about -8 to -13 dB. Subfigure (b) shows consistently high Starlink availability, remaining above 95% throughout (typically 96-100%).

Subfigure (c) shows the measured one-way uplink latency T_{up} under injected higher-layer impairments. During nominal operation, latency remains stable at approximately 30–40 ms for both links. Between $t \approx 180\text{–}300$ s, moderate queueing effects increase latency to roughly 50–90 ms. A short dual-impairment interval near $t \approx 330\text{–}370$ s produces peak latencies of approximately 105 ms (5G) and 110 ms (Starlink). After $t \approx 380$ s, both links recover to baseline levels below 45 ms. The control threshold $L_c = 50$ ms is exceeded during the moderate and severe intervals, with the most significant excursions occurring during the brief dual-congestion window.

Controller decisions are driven purely by the predicted uplink latency $\hat{T}_{\mathrm{up}}$. The system starts on 5G. It switches to Starlink at $t \approx 210$ s as predicted 5G latency rises above predictions for Starlink. During the severe dual-impairment window ($\approx 330\text{–}370$ s), both predictions exceed $L_c = 50$ ms, triggering adaptive H.265 mode as last-resort resilience. Even in this state, the controller continues best-effort link selection, briefly favoring Starlink then reverting to 5G as 5G recovers faster. After $t \approx 360$ s, stable low-latency conditions resume on 5G.

Subfigure (e) reports delivered video streaming quality at the vehicle client remains robust. Nominal operation maintains stable 30 FPS with near-zero dropped frames. During moderate impairment periods, FPS remains close to 30 with minimal drops. In the severe dual-impairment interval with adaptive H.265 active, quality is preserved at 24–26 FPS with dropped frames peaking around 2–4% before recovering fully.

These results confirm that higher-layer impairments alone—despite excellent radio conditions—can drive uplink latency beyond strict real-time thresholds. The predictor-driven controller reliably performs proactive link switching based on $\hat{T}_{\mathrm{up}}$ and invokes adaptive compression only during true dual infeasibility. The system exhibits stable switching without oscillations and effective recovery, maintaining acceptable video quality throughout.

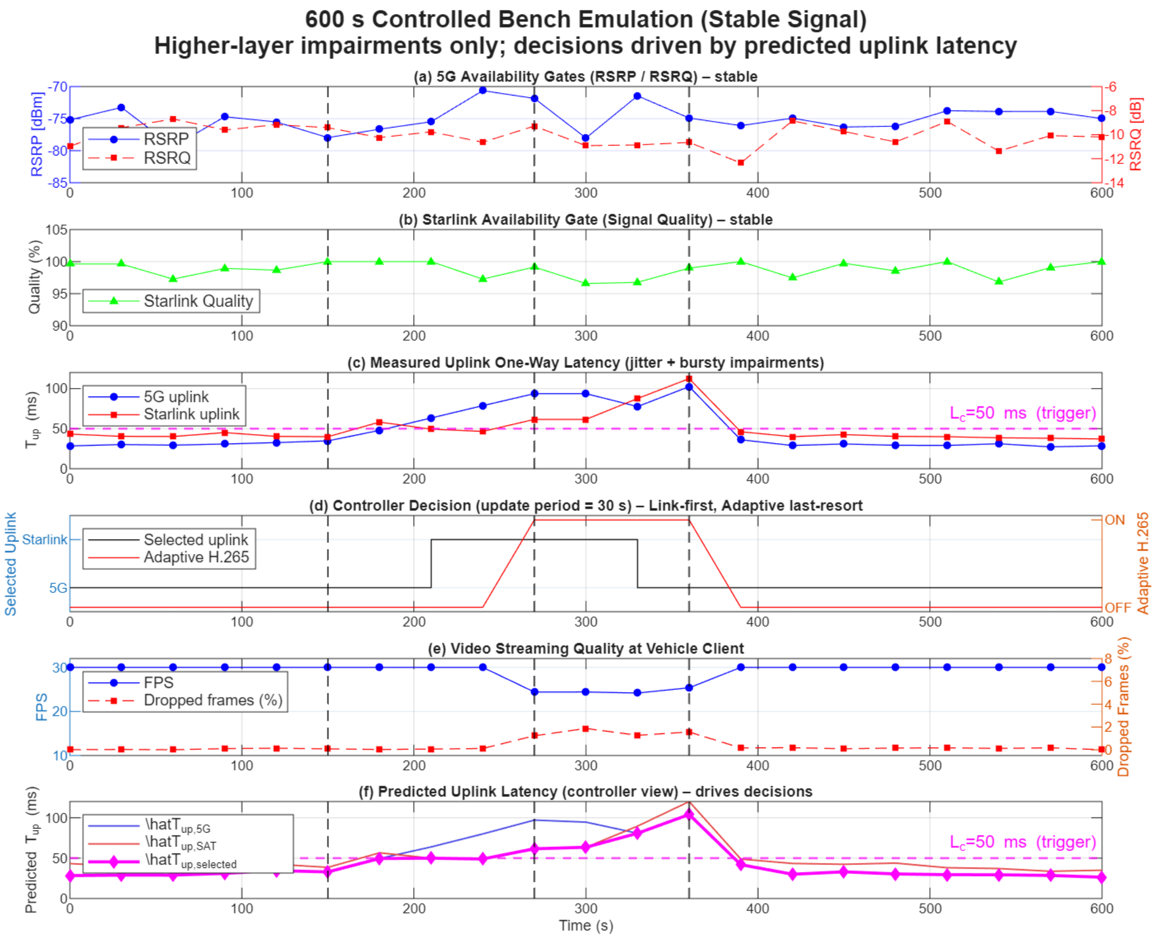

Figure 14. 600-second controlled bench emulation under stable radio-layer conditions. Excellent 5G RSRP/RSRQ and Starlink quality greater than 95% are maintained, while higher-layer impairments only are introduced. Decisions are driven by predicted uplink latency $\hat{T}_{\mathrm{up}}$. The controller selects the link with the lowest predicted latency when at least one is feasible ($\hat{T}_{\mathrm{up}} \leq L_c = 50$ ms); adaptive H.265 compression activates only as last resort when both predictions exceed L_c ($\mathcal{F}(t) = \emptyset$). Controller update period = 30 s.

Switching Stability

Across all emulated scenarios, no oscillatory switching occurs. Temporal smoothing ($\alpha = 0.8$), hysteresis, and minimum dwell-time constraints effectively prevent rapid back-and-forth transitions.

Table 11 summarizes the key performance metrics across four representative 600-second controlled emulation scenarios designed to stress-test the hierarchical adaptation logic under different congestion patterns. The first three rows correspond to single-link or dual-link degradation cases without the full set of countermeasures (5G congestion only, Starlink congestion only, and simultaneous degradation of both uplinks). The last row reports results for the complete proposed system, which combines predictive failover with last-resort adaptive H.265 compression. The metrics include the observed peak and 95th-percentile uplink transmission latency (T_{up}), the 95th-percentile end-to-end perception latency, the total time spent in adaptive compression mode, and the overall perception availability (fraction of frames with end-to-end latency $\leq L_{\max} = 100\,\text{ms}$). These results demonstrate that relying on a single uplink exposes the system to significant tail latency and availability degradation during congestion events, whereas predictive failover alone already substantially improves robustness. The full hierarchical approach further reduces time in degraded mode and achieves the highest availability (99.4%) by invoking compression only during rare dual-infeasibility periods, confirming the effectiveness of prioritizing connectivity switching over bitrate reduction.

Table 11. Summary of key emulation scenarios (600 s duration, end-to-end $L_{max} = 100$ ms)

Scenario	Peak T_{up} (ms)	95th Perc. T_{up} (ms)	95th Perc. E2E (ms)	Time in Adaptive (s)	Availability (%)
5G congestion only	198	132	178	0	96.5
Starlink congestion only	179	127	172	0	96.1
Dual degradation	248	141	192	36	94.8
Full system (hierarchical)	**215**	**92**	**138**	**12**	**99.4**

Feasibility decisions are based on predicted uplink latency $\hat{T}_{\mathrm{up}} \leq L_c = 50$ ms. Peak and 95th percentile T_{up} reflect values on impaired links during transients; predictive failover proactively selects the better link. E2E = full end-to-end perception latency measured from RSU frame capture to vehicle reception of perception outputs. Availability = fraction of frames with measured E2E ≤ 100 ms. Adaptive H.265 compression (last-resort) reduces packet size and effective T_{up} during brief dual-infeasibility periods.

The high 95th percentile E2E values in some scenarios reflect tail latency during transient impairments or brief switch hysteresis; however, the hierarchical adaptation (failover as primary mechanism) ensures that the vast majority of frames (99.4% in the full system) meet the E2E bound of 100 ms.

Real-World Network Evaluation

Seven independent 24-hour trials were conducted over public 5G cellular and Starlink connectivity under naturally occurring conditions, with no artificial impairments.

These experiments assess robustness, switching behavior, and adaptation frequency in a realistic roadside deployment.

Table 12 summarizes uplink mode occupancy across the 7-day dataset for the three ablation configurations. The 5G-only baseline uses the terrestrial link exclusively. Enabling predictive failover yields 68.4% 5G and 31.6% Starlink operation, demonstrating effective exploitation of short-horizon latency predictions. In the full system, occupancy is 64.2% 5G, 29.1% Starlink, and adaptive compression activates for only 6.7% of runtime. This low duty cycle confirms that predictive failover resolves most excursions.

Table 12. Uplink mode occupancy over the 7-day real-network trials.

Configuration	5G (%)	Starlink (%)	Adaptive H.265 (%)
5G-only	100.0	0.0	0.0
Failover-only	68.4	31.6	0.0
Full system	64.2	29.1	6.7

When adaptive H.265 compression triggers, bitrate reduces from 5 Mbps to 2.5–4.0 Mbps. Table 13 shows corresponding mAP@0.5 drops of only 2.3% at 4 Mbps, 5.0% at 3 Mbps, and 7.4% at 2.5 Mbps relative to nominal. These modest degradations highlight the efficiency of conditional compression as a last-resort safety net.

Table 13. Mean Average Precision as a Function of Encoding Bitrate.

Bitrate (Mbps)	mAP (Nominal)	mAP (Adaptive)	Relative Drop (%)
5.0	0.812	–	–
4.0	–	0.793	2.3
3.0	–	0.771	5.0
2.5	–	0.752	7.4

Perception Output and Detection Performance

Figure 15 shows representative detection outputs from the YOLOv8-small model on RSU camera footage during the 7-day trials. Vehicles, traffic lights, pedestrians, and other road users are consistently detected with confidence scores typically exceeding 95%. Bounding boxes remain spatially stable and accurately aligned, even when adaptive H.265 compression is active.

This visual robustness is quantitatively supported by Table 13. Nominal encoding at 5 Mbps yields 0.812 mAP@0.5. At reduced bitrates (2.5–4.0 Mbps) during adaptive mode, relative drops remain limited to 2.3–7.4%. These results confirm that conditional compression—activated only briefly (6.7% runtime)—preserves acceptable detection performance while maintaining latency bounds under heterogeneous 5G and Starlink conditions.

Figure 15. Representative object detection outputs from the cloud-based YOLOv8-small model using RSU camera footage during real-network trials. Vehicles, traffic lights, pedestrians, and other road users are detected with high confidence (typically greater than 95%). Bounding boxes remain stable and well-aligned, even during adaptive H.265 operation at reduced bitrate.

Conclusion

This paper presented a cloud-based RSU-to-cloud cooperative perception system that maintains bounded end-to-end perception latency and high availability under heterogeneous and time-varying network conditions. The proposed architecture integrates predictive connectivity failover with conditional adaptive video streaming in a hierarchical manner, prioritizing uplink selection over compression to exploit the complementary low-latency strengths of 5G and Starlink links.

A lightweight, communication-constrained perception model was introduced to formalize latency feasibility and perception availability. Uplink one-way transmission latency is treated as the dominant and most variable component; to conservatively ensure end-to-end latency $L_i(t) \leq L_{\max} = 100$ ms, feasibility decisions use a tighter control threshold on predicted uplink latency ($\hat{T}_{\text{up}} \leq L_c = 50$ ms).

The system was implemented on a real RSU-to-cloud testbed featuring an industrial camera, dual uplinks, cloud-hosted GPU inference (YOLOv8-small), and a vehicle-side client. Evaluation combined controlled bench emulation (to validate adaptation logic under repeatable stress) with long-duration real-network trials over public 5G and Starlink connectivity (to assess robustness under natural variability).

Results demonstrate that predictive failover significantly improves robustness to latency excursions, reduces perception outages, and preserves usable detection accuracy under network degradation. In the full hierarchical system, adaptive H.265 compression activates only as a last resort (e.g., 12 s in 600 s emulation stress cases, 6.7% runtime in real-network trials) and ensures bounded latency during severe simultaneous impairments with only modest impact on downstream perception quality (mAP@0.5 drop of 2.3–7.4%

at reduced bitrates 2.5–4.0 Mbps). Overall availability reaches 99.4% in controlled stress scenarios, confirming that the primary gains arise from treating communication as a controllable subsystem rather than a fixed constraint.

Future work will extend the framework to multi-RSU coordination and multi-sensor fusion, investigate learning-based predictors for improved anticipation of network degradation, and explore integration with edge computing platforms and standardized V2X communication stacks to enhance scalability, interoperability, and deployment readiness in intelligent transportation systems.

CHAPTER SIX

CONCLUSION AND FUTURE WORK

This chapter presented a cloud-assisted RSU-to-cloud cooperative perception framework designed to maintain bounded end-to-end perception latency and high perception availability under heterogeneous and time-varying communication conditions. Unlike conventional approaches that treat connectivity as a static constraint, the proposed system explicitly models communication as a controllable subsystem and integrates predictive uplink selection with conditional adaptive video streaming in a hierarchical manner.

A lightweight latency-feasibility model was introduced to formalize uplink eligibility based on predicted one-way transmission delay relative to a specified latency bound. This model enabled proactive failover between public 5G cellular and low-Earth-orbit satellite links using short-horizon latency prediction, with adaptive H.265 compression invoked strictly as a last-resort mechanism when no feasible uplink was available. By prioritizing connectivity selection over compression, the system minimizes unnecessary degradation of visual quality while preserving perception continuity.

The framework was implemented on a real RSU-to-cloud testbed and evaluated using a hybrid methodology combining controlled network emulation and long-duration real-network trials. Controlled emulation demonstrated correct triggering behavior, switching stability, and recovery under deterministic impairment scenarios, confirming the effectiveness of predictive failover and hysteresis-based control. Seven 24-hour real-network trials further showed that hybrid connectivity substantially improves perception availability compared to a 5G-only baseline, while the full system achieves near-continuous perception delivery by activating adaptive compression only during rare dual-link degradation events.

Experimental results indicate that the primary performance gains arise from proactive uplink selection rather than aggressive video adaptation. Average end-to-end latency remains relatively low across configurations, while tail latency and perception availability improve significantly with hybrid connectivity and predictive control. Importantly, adaptive H.265 compression is required for only a small fraction of total runtime, demonstrating that the proposed architecture preserves visual fidelity and detection accuracy during normal operation while maintaining robustness during adverse network conditions.

Overall, this work demonstrates that infrastructure-assisted cooperative perception can achieve reliable, low-latency operation over commodity wireless networks when communication is treated as a first-class control variable. The proposed architecture provides a practical pathway toward deployable RSU-cloud perception systems capable of supporting safety-critical autonomous driving applications in real-world environments.

While this study establishes the feasibility and benefits of predictive uplink failover combined with conditional adaptive streaming for single-camera RSU perception, several important directions remain for future research.

First, the current implementation focuses on monocular roadside sensing. Extending the framework to multi-sensor RSUs incorporating LiDAR, radar, and multi-camera arrays represents a natural progression. Multi-modal fusion at the cloud or edge would enable more robust perception under occlusion and adverse weather, while introducing new challenges in bandwidth allocation, synchronization, and cross-sensor latency coordination.

Second, the present system evaluates a single RSU communicating with a centralized cloud service. Future work will investigate large-scale deployments involving multiple RSUs and vehicles, where cooperative perception must be coordinated across intersections or road segments. Such scenarios introduce additional complexity in network resource sharing, load balancing, and distributed scheduling, motivating the development of city-scale orchestration strategies and RSU-to-RSU cooperation mechanisms.

Third, the predictive uplink selection currently relies on short-horizon exponential smoothing of measured latency. More advanced learning-based predictors incorporating historical network context, radio-layer indicators, and environmental features could further improve switching accuracy and reduce unnecessary transitions. Reinforcement learning approaches may also be explored to jointly optimize uplink selection, encoding parameters, and cloud resource allocation.

Fourth, while this work focuses on latency feasibility and perception availability, future studies will incorporate energy efficiency, monetary cost, and carbon footprint as explicit optimization objectives. Such multi-objective formulations are essential for sustainable large-scale deployment of cloud-assisted autonomous systems.

Finally, integration with emerging edge computing platforms and 6G network architectures presents opportunities to further reduce latency and improve reliability. Hybrid edge–cloud processing, dynamic function placement, and network slicing could enable tighter real-time guarantees while reducing dependency on distant cloud infrastructure.

Together, these directions aim to evolve the proposed system from a single-RSU prototype toward a scalable, multi-sensor, city-wide cooperative perception platform capable of supporting next-generation autonomous transportation systems.

REFERENCES

[1] Pomerleau, D., & Jochem, T. (2002). Rapidly adapting machine vision for automated vehicle steering. *IEEE expert*, *11*(2), 19-27.

[2] Leighty, R. D. (1986). *Darpa alv (autonomous land vehicle) summary* (No. ETLR085).

[3] Dickmanns, E. D. (1997, August). Vehicles capable of dynamic vision. In *IJCAI* (Vol. 97, pp. 1577-1592).

[4] Broggi, A., Bertozzi, M., Fascioli, A., Bianco, C. G. L., & Piazzi, A. (1999). The ARGO autonomous vehicle's vision and control systems. *International Journal of Intelligent Control and Systems*, *3*(4), 409-441.

[5] Thrun, S., Montemerlo, M., Dahlkamp, H., Stavens, D., Aron, A., Diebel, J., ... & Mahoney, P. (2006). Stanley: The robot that won the DARPA Grand Challenge. *Journal of field Robotics*, *23*(9), 661-692.

[6] Urmson, C., Anhalt, J., Bagnell, D., Baker, C., Bittner, R., Clark, M. N., ... & Ferguson, D. (2008). Autonomous driving in urban environments: Boss and the urban challenge. *Journal of field Robotics*, *25*(8), 425-466.

[7] Society of Automotive Engineers (SAE). (2018). Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles (J3016_201806).

[8] Anderson, J. M., Nidhi, K., Stanley, K. D., Sorensen, P., Samaras, C., & Oluwatola, O. A. (2014). *Autonomous vehicle technology: A guide for policymakers*. Rand Corporation.

[9] National Highway Traffic Safety Administration. (2017). Automatic emergency braking systems: Technical assessment. U.S. Department of Transportation. https://www.nhtsa.gov/

[10] American Automobile Association. (2021). Automatic emergency braking performance in low-light conditions (Technical Report). AAA Automotive Research Center. https://newsroom.aaa.com/

[11] Tesla, Inc. (2023, March 1). *2023 Investor Day presentation* [PowerPoint

slides]. Tesla Investor Relations. https://ir.tesla.com/

[12] National Transportation Safety Board. (2019). *Collision between vehicle controlled by developmental automated driving system and pedestrian, Tempe, Arizona, March 18, 2018* (Highway Accident Report NTSB/HAR-19/03). https://www.ntsb.gov/investigations/accidentreports/reports/har1903.pdf.

[13] Levinson, J., Askeland, J., Becker, J., Dolson, J., Held, D., Kammel, S., ... & Thrun, S. (2011, June). Towards fully autonomous driving: Systems and algorithms. In *2011 IEEE intelligent vehicles symposium (IV)* (pp. 163-168). IEEE.

[14] GSMA Intelligence. (2025, March 3). *The mobile economy 2025*. https://www.gsmaintelligence.com/research/the-mobile-economy-2025.

[15] SpaceX. "Starlink: Low-latency satellite internet for connectivity." *SpaceX Updates*. 2023. https://www.spacex.com/updates/.

[16] Pew Research Center. (2022, March 17). Americans cautious about the deployment of driverless cars. https://www.pewresearch.org/internet/2022/03/17/americans-cautious-about-the-deployment-of-driverless-cars/.

[17] Chen, Q., Tang, S., Yang, Q., & Fu, S. (2019, July). Cooper: Cooperative perception for connected autonomous vehicles based on 3d point clouds. In *2019 IEEE 39th International Conference on distributed computing systems (ICDCS)* (pp. 514-524). IEEE.

[18] Arena, F., & Pau, G. (2019). An overview of vehicular communications. *Future internet*, *11*(2), 27.

[19] European Commission, "C-its platform final report: Cooperative intelligent transport systems," Tech. Rep., 2016.

[20] U.S. Department of Energy. (2017). Potentials for platooning in U.S. highway freight transport. Office of Energy Efficiency and Renewable Energy. https://www.nrel.gov/docs/fy17osti/67618.pdf.

[21] U.S. Department of Transportation. (2018, October). Preparing for the future of transportation: Automated vehicles 3.0. https://www.transportation.gov/av/3.

[22] European Commission. (2019). Cooperative intelligent transport systems (C-ITS): Deployment guidelines (C/2019/1789). Publications Office of the European Union. https://op.europa.eu/.

[23] Ahangar, M. N., Ahmed, Q. Z., Khan, F. A., & Hafeez, M. (2021). A survey of autonomous vehicles: Enabling communication technologies and challenges. *Sensors*, *21*(3), 706.

[24] Kumar, S., Gollakota, S., & Katabi, D. (2012, August). A cloud-assisted design for autonomous driving. In *Proceedings of the first edition of the MCC workshop on Mobile cloud computing* (pp. 41-46).

[25] Maller, L. M., Suskovics, P., & Bokor, L. (2023). Edge computing in the loop simulation framework for automotive use cases evaluation. *Wireless Networks*, *29*(8), 3717-3735.

[26] Vieira, E., Almeida, J., Ferreira, J., Dias, T., Vieira Silva, A., & Moura, L. (2023). A roadside and cloud-based vehicular communications framework for the provision of C-ITS services. *Information*, *14*(3), 153.

[27] Pawar, V., Zade, N., Vora, D., Khairnar, V., Oliveira, A., Kotecha, K., & Kulkarni, A. (2024). Intelligent transportation system with 5G vehicle-to-everything (V2X): Architectures, vehicular use cases, emergency vehicles, current challenges, and future directions. *IEEE Access*, *12*, 183937-183960.

[28] Khalid, I., Maglogiannis, V., Naudts, D., Shahid, A., & Moerman, I. (2024). Optimizing hybrid v2x communication: An intelligent technology selection algorithm using 5g, c-v2x PC5 and dsrc. *Future Internet*, *16*(4), 107.

[29] Xu, R., Xiang, H., Tu, Z., Xia, X., Yang, M. H., & Ma, J. (2022, October). V2x-vit: Vehicle-to-everything cooperative perception with vision transformer. In *European conference on computer vision* (pp. 107-124). Cham: Springer Nature Switzerland.

[30] Qiu, H., Huang, P., Asavisanu, N., Liu, X., Psounis, K., & Govindan, R.

(2021). Autocast: Scalable infrastructure-less cooperative perception for distributed collaborative driving. *arXiv preprint arXiv:2112.14947*.

[31] Cui, J., Qiu, H., Chen, D., Stone, P., & Zhu, Y. (2022). Coopernaut: End-to-end driving with cooperative perception for networked vehicles. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 17252-17262).

[32] Zhang, Q., Zhang, X., Zhu, R., Bai, F., Naserian, M., & Mao, Z. M. (2023, October). Robust real-time multi-vehicle collaboration on asynchronous sensors. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking* (pp. 1-15).

[33] Chen, H., Wang, H., Liu, Z., Gu, D., & Ye, W. (2024). HP3D-V2V: High-precision 3D object detection vehicle-to-vehicle cooperative perception algorithm. *Sensors*, *24*(7), 2170.

[34] Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. *IEEE internet of things journal*, *3*(5), 637-646.

[35] Sullivan, G. J., Ohm, J. R., Han, W. J., & Wiegand, T. (2012). Overview of the high efficiency video coding (HEVC) standard. *IEEE Transactions on circuits and systems for video technology*, *22*(12), 1649-1668.

[36] Zimmer, W., Wardana, G. A., Sritharan, S., Zhou, X., Song, R., & Knoll, A. C. (2024). Tumtraf v2x cooperative perception dataset. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 22668-22677).

[37] Khan, H., Samarakoon, S., & Bennis, M. (2020). Enhancing video streaming in vehicular networks via resource slicing. *IEEE Transactions on Vehicular Technology*, *69*(4), 3513-3522.

[38] Dror, I., & Hadar, O. (2025). Improved Perceptual Quality of Traffic Signs and Lights for the Teleoperation of Autonomous Vehicle Remote Driving via Multi-Category Region of Interest Video Compression. *Entropy*, *27*(7), 674.

[39] Yang, H., Zheng, K., Zhang, K., Mei, J., & Qian, Y. (2020). Ultra-reliable and low-latency communications for connected vehicles: Challenges and solutions. *IEEE Network*, *34*(3), 92-100.

[40] Heine, S., Völk, F., Schwarz, R. T., & Knopp, A. (2022, October). Concept and evaluation of 5G backhauling via starlink. In *39th International Communications Satellite Systems Conference (ICSSC 2022)* (Vol. 2022, pp. 69-75). IET.

[41] Jocher, G., Chaurasia, A., & Qiu, J. (2023). Ultralytics YOLOv8. Retrieved from https://github.com/ultralytics/ultralytics

[42] Talebpour, A., & Mahmassani, H. S. (2016). Influence of connected and autonomous vehicles on traffic flow stability and throughput. *Transportation Research Part C: Emerging Technologies*, 71, 143–163.

[43] Cui, G., Zhang, W., Xiao, Y., Yao, L., & Fang, Z. (2022). Cooperative perception technology of autonomous driving in the internet of vehicles environment: A review. *Sensors*, 22(15), Article 5535.

[44] Tihanyi, V., Rövid, A., Remeli, V., Vincze, Z., Csonthó, M., Pethő, Z., & Szalay, Z. (2021). Towards cooperative perception services for ITS: Digital twin in the automotive edge cloud. *Energies*, 14(18), Article 5930.

[45] Wang, L., Zhang, Z., Di, X., & Tian, J. (2021). A roadside camera-radar sensing fusion system for intelligent transportation. *In 2020 17th European Radar Conference (EuRAD)* (pp. 282–285). IEEE.

[46] Liang, S., Wu, H., Zhen, L., Hua, Q., Garg, S., Kaddoum, G., & Yu, K. (2022). Edge YOLO: Real-time intelligent object detection system based on edge-cloud cooperation in autonomous vehicles. *IEEE Transactions on Intelligent Transportation Systems,* 23(12), 25345–25360.

[47] Ding, L., Wang, Y., Wu, P., Li, L., & Zhang, J. (2019). Kinematic information aided user-centric 5G vehicular networks in support of cooperative perception for automated driving. *IEEE Access*, 7, 40195–40209.

[48] Arnold, E., Dianati, M., De Temple, R., & Fallah, S. (2020). Cooperative perception for 3D object detection in driving scenarios using infrastructure sensors. *IEEE Transactions on Intelligent Transportation Systems*, 23(3), 1852–1864.

[49] Adl, M., Guo, X., Mohammad-Alikhani, A., Abdi, B., Ahmed, R., & Emadi, A.

(2025). Roadside fisheye vision for cooperative perception in V2I-assisted automated driving. *IEEE Open Journal of Intelligent Transportation Systems*.

[50] Krijestorac, E., Memedi, A., Higuchi, T., Ucar, S., Altintas, O., & Cabric, D. (2020, December). Hybrid vehicular and cloud distributed computing: A case for cooperative perception. *In GLOBECOM 2020 - 2020 IEEE Global Communications Conference* (pp. 1–6). IEEE.

[51] Mehrke, J., Volk, G., Stumpp, Y., Bringmann, O., & Terzis, A. (2023, September). Impacts of image compression on the detection quality of a novel real-time image processing platform. *In 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)* (pp. 417–424). IEEE.

[52] Heo, S., & Kim, H. (2025). Multiplexing remote driving video over dual V2N paths with quality-of-experience and data efficiency. *IEEE Access*.

[53] Shi, W., Li, Q., Yu, Q., Wang, F., Shen, G., Jiang, Y., ... & Muntean, G. M. (2024). A survey on intelligent solutions for increased video delivery quality in cloud–edge–end networks. IEEE Communications Surveys & Tutorials, 27(2), 1363–1394.

[54] Tavasoli, M., Sarrafzadeh, A., Khaleghi, M., Zakaria, M., Pasandi, H. B., & Karimoddini, A. (2025). Data communication challenges of connected and automated vehicles in rural areas. *IEEE Access*, 13, 29220–29251.

[55] Yastrebova-Castillo, A., Höyhtyä, M., Majanen, M., Ojanperä, T., Scholliers, J., & Rantala, S. J. (2023, November). Improving the safety of autonomous driving by using Direct-to-Satellite connectivity: The case of Iridium and Starlink satellite constellations. *In 2023 IEEE Conference on Standards for Communications and Networking (CSCN)* (pp. 40–46). IEEE.

[56] Gao, X., Zhang, X., Lu, Y., Huang, Y., Yang, L., Xiong, Y., & Liu, P. (2024). A survey of collaborative perception in intelligent vehicles at intersections. *IEEE Transactions on Intelligent Vehicles*.

[57] Wei, C., Wu, G., & Barth, M. J. (2025). Cooperative perception for automated driving: A survey of algorithms, applications, and future directions. *Proceedings of the IEEE*.

[58] Magalhães, P., d'Orey, P. M., Andrade, M. T., & Castro, F. (2020, October). Video-assisted overtaking system enabled by C-V2X mode 4 communications. *In 2020 16th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)* (pp. 382–387). IEEE.