International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

A Heterogeneous Adaptive Fusion Network for Dual-Dataset Traffic Safety Monitoring using Adaptive Scene Routing

Arjun Charak [a], Sukhvir Singh [b], Dr. Bhawna Sharma [c], Yawar Azad [d]
[abcd] Government College of Engineering and Technology (GCET), Jammu, India

Abstract - Road accidents and traffic violations continue to create significant challenges to public safety, traffic management, and intelligent transportation systems. Recent advances in deep learning have enabled automated traffic monitoring; although, all the existing solutions primarily focus on isolated tasks like accident detection, helmet violation detection, or license plate recognition. As a result of which, all the existing traffic safety systems remain disrupted, requiring multiple independent models and processing pipelines to address diverse traffic monitoring requirements. Additionally, most existing frameworks employ static inference strategies that lack the ability to adapt model selection according to scene characteristics and applicability. In order to remove these limitations, this paper proposes the Heterogeneous Adaptive Fusion Network (HAFN), a novel traffic safety monitoring framework that integrates dual-dataset learning, specialized expert models, and adaptive scene routing for unified traffic intelligence. The suggested framework applies two complementary datasets: an Accident Dataset containing 3,933 training images and 437 validation images, and a Multi-Violation Traffic Dataset consisting of 14,111 training images and 2,193 validation images covering helmet violations, seatbelt violations, license plate detection, traffic signal monitoring, wrong-side driving, parking violations, and vehicle classification. To make sure the model is using domain-specific learning, a YOLOv8-based Accident Expert and a YOLOv11-based Violation Expert are developed through transfer learning.

A lightweight Scene Routing Module is introduced to inspect incoming traffic scenes and calculate accident and violation bearing scores. For the objective of adaptive expert activation and intelligent inference scheduling these scores are further evaluated through confidence and ambiguity assessment mechanisms. According to context, HAFN dynamically selects a single expert or activates multiple experts simultaneously, thereby improving computational efficiency while preserving detection effectiveness. All the outputs generated by individual experts are subsequently integrated through a fusion layer to produce comprehensive traffic safety intelligence.

Experimental results show the effectiveness of dual-dataset learning and adaptive expert routing for heterogeneous traffic environments. The proposed HAFN framework establishes a scalable foundation for next-generation intelligent transportation systems capable of supporting accident monitoring, traffic violation analysis, and future emergency response applications.

Keywords: Intelligent Transportation Systems, Traffic Safety Monitoring, HAFN, Adaptive Scene Routing, Multi-Expert Learning, YOLOv8, YOLOv11, Accident Detection, Traffic Violation Detection, Deep Learning

1. INTRODUCTION

1.1 Background

The rapid growth of urbanization, industrialization, and vehicular transportation has fundamentally transformed modern society by improving mobility, facilitating economic development, and enhancing regional connectivity. Though, this unprecedented increase in vehicle ownership has simultaneously introduced significant challenges for transportation authority's worldwide. The continuous expansion of metropolitan cities, increasing population density, and rising travel demand have resulted in severe traffic congestion, longer travel times, increased fuel

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

consumption, higher greenhouse gas emissions, and a substantial rise in road traffic accidents. Existing transportation infrastructures are often unable to adapt to continuously increasing traffic volumes, thereby reducing operational efficiency and compromising road safety. Therefore, traffic management has evolved into one of the most important research domains within intelligent infrastructure development and smart city initiatives, requiring advanced technological solutions capable of improving transportation efficiency while ensuring public safety [1], [2].

Road traffic accidents continue to represent one of the most serious public health and socioeconomic challenges worldwide. According to the World Health Organization (WHO), around **1.19 million people lose their lives annually due to road traffic crashes**, while between **20 and 50 million people suffer non-fatal injuries**, many of which result in permanent disabilities. In addition, road traffic injuries remain the leading cause of death among children and young adults aged between 5 and 29 years, highlighting the urgent need for effective preventive measures and intelligent monitoring systems [3], [4]. Beyond the devastating human impact, road accidents impose large economic losses through medical expenses, emergency response operations, vehicle damage, insurance claims, productivity losses, and infrastructure repairs. It has been evaluated that road traffic crashes cost many countries nearly three percent of their Gross Domestic Product (GDP), making road safety not only a humanitarian concern but also an important economic priority [3], [5].

The rapid increase in motorization has also resulted in a significant rise in traffic rule violations that contribute directly to accident occurrence and traffic congestion. Common violations include riding motorcycles without helmets, failure to wear seatbelts, overspeeding, traffic signal jumping, wrong-side driving, illegal parking, dangerous overtaking, lane indiscipline, distracted driving due to mobile phone usage, and unauthorized vehicle movements. These unsafe driving behaviours substantially increase the likelihood of collisions and frequently surpass the capacity of conventional traffic enforcement mechanisms. Manual traffic monitoring performed by traffic police officers remains labour-intensive, time-consuming, subjective, and difficult to scale across rapidly expanding urban transportation networks. Moreover, continuously monitoring thousands of surveillance cameras installed throughout modern cities is practically impossible using only human operators. Therefore, automated traffic surveillance systems capable of continuously monitoring road activities and detecting multiple traffic violations in real time have become an indispensable component of modern intelligent transportation infrastructure [6]–[8].

To resolve these challenges, **Intelligent Transportation Systems (ITS)** have come out as an interdisciplinary research area integrating transportation engineering, wireless communication, computer vision, artificial intelligence, cloud computing, edge computing, and the Internet of Things (IoT). The primary objective of ITS is to improve transportation safety, operational efficiency, mobility, and environmental sustainability by enabling intelligent decision-making based on real-time traffic information. Modern ITS architectures employ a wide range of sensing technologies including surveillance cameras, inductive loop detectors, radar systems, LiDAR sensors, Global Positioning System (GPS) devices, connected vehicles, drones, and roadside communication units to continuously collect traffic-related information. These heterogeneous data sources generate enormous volumes of traffic information that require intelligent processing techniques capable of extracting meaningful insights with minimal human intervention [9][11].

Among all sensing technologies available for traffic surveillance, vision-based monitoring systems have become particularly attractive because of the widespread deployment of Closed-Circuit Television (CCTV) cameras in urban environments. Compared with conventional sensor-based approaches, surveillance cameras provide significantly richer information by simultaneously capturing multiple road users, traffic signs, lane markings, pedestrian movements, vehicle interactions, and environmental conditions. Consequently, computer vision-based traffic monitoring systems have become increasingly popular due to their ability to perform multiple transportation tasks using a single sensing modality. These applications include vehicle detection, vehicle

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

classification, pedestrian detection, lane detection, traffic density estimation, vehicle counting, speed estimation, traffic flow analysis, congestion monitoring, illegal parking detection, accident detection, and automatic traffic law enforcement. Such capabilities make camera-based intelligent surveillance systems both economically viable and operationally efficient for smart city deployments [12][14].

The significant progress in artificial intelligence during the past decade has fundamentally transformed computer vision research. Traditional image processing techniques primarily relied on handcrafted features, manually designed descriptors, threshold-based segmentation methods, and heuristic algorithms. However these approaches demonstrated acceptable performance under controlled environments, they frequently failed when confronted with illumination variations, shadows, adverse weather conditions, camera viewpoint changes, dense traffic scenes, and partial object occlusions. Consequently, conventional vision-based traffic monitoring systems often exhibited limited robustness and poor generalization capability in real-world operating conditions [15][16].

Deep learning has successfully addressed many of these limitations by enabling automatic feature extraction directly from raw image data. Instead of using manually engineered features, deep neural networks learn hierarchical feature representations through multiple layers of nonlinear transformations, thereby significantly improving detection accuracy, robustness, and adaptability across diverse traffic environments. The availability of large-scale annotated datasets, high-performance Graphics Processing Units (GPUs), and optimized deep learning frameworks has accelerated the widespread adoption of artificial intelligence for intelligent transportation applications. As a result, deep learning-based computer vision systems have become the dominant paradigm for automated traffic monitoring, demonstrating superior performance in object detection, semantic segmentation, image classification, activity recognition, and multi-object tracking [17][18].

Within the set of deep learning architectures, **Convolutional Neural Networks (CNNs)** have become the fundamental building blocks of modern computer vision systems because of their exceptional capability to automatically learn discriminative visual features from images and video sequences. CNN-based approaches have demonstrated outstanding performance across numerous intelligent transportation applications, including vehicle recognition, pedestrian detection, traffic sign recognition, driver behaviour analysis, road surface inspection, autonomous driving, and accident detection. Their ability to generalize complex visual patterns under diverse environmental conditions has significantly enhanced the reliability and practical applicability of intelligent traffic surveillance systems deployed in real-world urban environments [19][20].

Object detection is one of the most fundamental tasks in computer vision and serves as the foundation for numerous intelligent transportation applications. Unlike image classification, which predicts only the class label of an image, object detection simultaneously identifies the category and precise location of multiple objects within a scene. Accurate object detection enables higher-level traffic analysis tasks such as vehicle tracking, traffic flow estimation, speed measurement, accident detection, traffic violation monitoring, and autonomous navigation. Consequently, significant research efforts have been devoted to developing object detection algorithms capable of achieving high detection accuracy while maintaining real-time computational performance [21][22].

Early deep learning-based object detectors primarily employed two-stage detection frameworks. The Region-based Convolutional Neural Network (R-CNN) introduced the concept of generating region proposals before performing object classification using Convolutional Neural Networks, significantly improving detection accuracy over traditional handcrafted feature-based methods [23]. Although R-CNN demonstrated remarkable performance, its computational complexity prevented real-time deployment due to repeated feature extraction for each proposed region. To address this limitation, Fast R-CNN introduced shared convolutional feature maps that substantially reduced computational redundancy while improving detection speed [24]. Subsequently, Faster R-CNN incorporated the Region Proposal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

Network (RPN), enabling end-to-end object detection with significantly improved inference efficiency and localization accuracy [25]. Despite these improvements, two-stage detectors generally require considerable computational resources, limiting their applicability in latency-sensitive traffic surveillance applications.

To overcome the computational limitations of two-stage detectors, several one-stage object detection architectures were proposed. The Single Shot MultiBox Detector (SSD) eliminated the region proposal stage and directly predicted object classes and bounding boxes from multiple feature maps, thereby achieving substantially higher inference speeds while maintaining competitive detection accuracy [26]. RetinaNet further addressed the class imbalance problem inherent in one-stage detectors through the introduction of the Focal Loss function, enabling accurate detection of both common and rare object classes [27]. More recently, transformer-based object detection frameworks such as Detection Transformer (DETR) have demonstrated remarkable capabilities by replacing handcrafted anchor mechanisms with end-to-end attention-based architectures capable of learning object relationships directly from image features [28]. Although these methods provide excellent detection performance, their computational requirements often remain challenging for real-time intelligent transportation systems deployed on embedded hardware.

Among all object detection algorithms, the **You Only Look Once (YOLO)** family has emerged as one of the most influential and widely adopted architectures for real-time object detection. Unlike traditional two-stage approaches, YOLO reformulates object detection as a single regression problem that simultaneously predicts object locations and class probabilities within one neural network. This unified detection strategy significantly reduces computational complexity while maintaining high localization accuracy, making YOLO particularly suitable for intelligent transportation applications requiring low latency and high throughput [29].

The original YOLO architecture (YOLOv1) introduced a revolutionary approach by dividing an image into a grid and predicting bounding boxes and corresponding class probabilities in a single forward pass through the network. This design achieved unprecedented detection speed, enabling real-time processing while maintaining acceptable detection accuracy for various object categories [29]. Building upon this foundation, YOLOv2 introduced anchor boxes, batch normalization, high-resolution classifiers, and multi-scale training strategies that significantly improved localization accuracy and small-object detection performance [30]. Subsequently, YOLOv3 employed Darknet-53 as its backbone network and incorporated Feature Pyramid Networks (FPN) to improve multi-scale object detection, making it particularly effective for detecting vehicles and pedestrians appearing at different image resolutions [31].

The evolution of the YOLO family continued with YOLOv4, which integrated several optimization techniques including Cross Stage Partial Networks (CSPNet), Mosaic data augmentation, Spatial Pyramid Pooling (SPP), and Complete Intersection over Union (CIoU) loss to achieve an improved balance between detection accuracy and computational efficiency [32]. YOLOv5 further emphasized practical deployment by providing lightweight architectures, simplified implementation, enhanced data augmentation strategies, automatic anchor optimization, and improved training efficiency. These characteristics contributed significantly to the widespread adoption of YOLOv5 in industrial computer vision applications, including intelligent transportation systems [33].

More recent developments introduced YOLOv6, YOLOv7, YOLOv8, YOLOv9, YOLOv10, and YOLO11, each incorporating architectural improvements designed to increase detection accuracy while reducing inference latency. YOLOv7 introduced trainable bag-of-freebies optimization strategies that improved both training efficiency and detection performance without increasing computational cost [34]. YOLOv8 adopted an anchor-free detection paradigm together with an improved backbone architecture, resulting in enhanced robustness for detecting small and partially occluded traffic objects [35]. YOLOv9 incorporated Programmable Gradient Information (PGI) and Generalized Efficient Layer Aggregation Network (GELAN), providing better feature representation and optimization capabilities while maintaining computational efficiency [36]. YOLOv10 further optimized end-to-end detection by eliminating redundant post-processing operations and reducing inference latency, making it particularly attractive for real-time surveillance applications deployed on resource-constrained edge devices [37]. The recently introduced YOLO11 continues this evolutionary trend by improving feature extraction, computational efficiency, and deployment flexibility across multiple hardware platforms, making it suitable for next-generation intelligent traffic surveillance systems [38].

The outstanding balance between detection accuracy, computational efficiency, and inference speed has resulted in widespread adoption of YOLO-based architectures across numerous intelligent transportation applications. Researchers have successfully employed YOLO for vehicle detection, vehicle classification, pedestrian detection, emergency vehicle recognition, lane monitoring, road sign detection, speed estimation, vehicle counting, congestion analysis, and traffic density measurement. Furthermore, YOLO-based frameworks have demonstrated excellent performance for detecting safety-related traffic violations including helmet non-compliance, seatbelt violations, traffic signal jumping, illegal parking, wrong-way driving, distracted driving, and automatic license plate recognition [39]–[41].

While accurate object detection is essential, many intelligent transportation applications also require continuous monitoring of vehicle movements across successive video frames. Consequently, **multi-object tracking (MOT)** has become an indispensable component of modern traffic surveillance systems. Multi-object tracking associates detected objects over time, enabling the estimation of vehicle trajectories, traffic flow characteristics, driving behaviour, and long-term scene understanding. Reliable tracking facilitates applications such as speed estimation, lane-change analysis, congestion monitoring, abnormal behaviour detection, and accident analysis [42].

Among the most influential tracking algorithms, **DeepSORT** extends the original SORT framework by integrating deep appearance descriptors with Kalman filtering and the Hungarian assignment algorithm, thereby significantly improving tracking robustness under partial occlusions and crowded traffic scenes [43]. More recently, **ByteTrack** introduced an innovative data association strategy that considers both high-confidence and low-confidence detections, substantially reducing identity switches and improving tracking accuracy in dense urban traffic environments [44]. Additional tracking algorithms, including BoT-SORT, OC-SORT, and StrongSORT, have further enhanced trajectory estimation through improved motion modelling and appearance matching, making them highly suitable for intelligent transportation applications requiring long-term object tracking [45].

The integration of advanced object detection and multi-object tracking has significantly enhanced the capabilities of intelligent traffic surveillance systems. Rather than simply identifying the presence of vehicles, these integrated frameworks enable comprehensive traffic scene understanding by simultaneously detecting, classifying, and tracking multiple road users in real time. Such capabilities provide the foundation for higher-level traffic analytics, including accident prediction, behavioural analysis, traffic violation enforcement, congestion management, and emergency response coordination, thereby contributing to the development of safer, more efficient, and more intelligent transportation infrastructures.

Recent advances in deep learning have significantly improved the capability of intelligent transportation systems to automatically detect road traffic accidents from surveillance videos. Conventional accident detection techniques primarily relied on sensor-based approaches, including accelerometers, vibration sensors, acoustic sensors, and in-vehicle communication devices. Although these methods provide reliable information after an accident occurs, they generally require dedicated hardware installations and are unable to monitor large road networks efficiently. Vision-based accident detection systems overcome these limitations by continuously analysing surveillance video streams to identify vehicle collisions, sudden trajectory deviations, abnormal motion patterns, vehicle overturning, and post-impact behaviours. Deep learning models trained on large-scale traffic datasets have demonstrated remarkable improvements in detecting accidents under complex urban traffic conditions while significantly reducing emergency response time through automatic alert generation [46][48].

Helmet detection has become another major research area in intelligent traffic monitoring due to the increasing number of motorcycle-related fatalities worldwide. Motorcyclists represent one of the most vulnerable groups of road users, and failure to wear protective helmets substantially increases the risk of fatal head injuries during collisions. Consequently, numerous computer vision-based systems have been proposed for automatically detecting helmet compliance using roadside surveillance cameras. Recent YOLO-based frameworks have achieved high detection accuracy by simultaneously identifying motorcycles, riders, and helmets within complex traffic environments. Several studies have further improved helmet detection performance through attention mechanisms, multi-scale feature fusion, and lightweight neural network architectures capable of real-time deployment on embedded devices [49][51].

Similarly, automatic seatbelt detection has gained considerable attention because seatbelts remain one of the most effective passive safety mechanisms for reducing occupant injuries during road accidents. Manual inspection of seatbelt compliance is often impractical in densely populated urban environments where thousands of vehicles pass through road intersections every hour. Recent advances in deep learning have enabled automatic seatbelt monitoring systems capable of analysing vehicle occupants using high-resolution surveillance cameras. Convolutional neural networks and YOLO-based object detectors have demonstrated promising performance in identifying seatbelt usage under varying illumination conditions, diverse camera viewpoints, and partially occluded vehicle interiors. Such automated monitoring systems improve traffic law enforcement while reducing dependence on manual inspection [52][53].

Traffic signal violation detection represents another critical application of intelligent transportation systems. Signal jumping remains one of the primary causes of severe road accidents at urban intersections due to conflicts between crossing traffic streams. Traditional signal monitoring approaches typically require dedicated inductive loop detectors or manually operated enforcement mechanisms, both of which involve substantial installation and operational costs. Recent computer vision techniques combine vehicle detection, traffic light recognition, lane detection, and object tracking to automatically identify vehicles crossing stop lines during red signal intervals. These integrated systems have demonstrated reliable performance under diverse traffic conditions and have become increasingly attractive for smart city deployments [54][55].

Wrong-way driving is another hazardous traffic violation responsible for numerous fatal highway accidents each year. Vehicles travelling in the opposite direction create extremely dangerous situations, particularly on expressways, divided highways, tunnels, and one-way streets. Modern computer vision systems detect wrong-way driving by combining object detection with trajectory analysis and multi-object tracking algorithms. By analysing vehicle movement directions over consecutive video frames, these intelligent systems can generate immediate alerts to traffic authorities and emergency responders before collisions occur. Such predictive capabilities significantly improve road safety while minimizing accident severity [56][57].

Illegal parking and unauthorized vehicle stopping also contribute significantly to urban congestion, emergency vehicle obstruction, and reduced roadway capacity. Conventional parking enforcement depends heavily on manual inspection, which is labour-intensive and incapable of providing continuous monitoring over large urban regions. Deep learning-based vision systems can automatically identify illegally parked vehicles by analysing vehicle trajectories, parking durations, road markings, and restricted parking zones. The integration of object detection with temporal tracking enables continuous observation of stationary vehicles and automatic violation reporting, thereby improving parking management efficiency and reducing traffic congestion [58].

Beyond individual traffic violations, intelligent driver behaviour analysis has emerged as an important component of next-generation transportation systems. Dangerous driving behaviours such as aggressive

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

acceleration, sudden braking, frequent lane changes, distracted driving, tailgating, and mobile phone usage substantially increase accident risk. Recent advances in deep learning, computer vision, and behavioural modelling have enabled automated systems capable of recognising abnormal driving patterns directly from surveillance videos. Such behavioural analytics provide valuable information for proactive accident prevention, traffic risk assessment, and intelligent transportation planning [59][60].

The rapid advancement of **edge computing** has further transformed intelligent traffic surveillance by enabling artificial intelligence models to perform inference directly at roadside devices rather than relying exclusively on cloud-based servers. Edge AI significantly reduces communication latency, minimizes bandwidth consumption, enhances data privacy, and enables immediate responses to critical traffic events. Modern embedded platforms equipped with Graphics Processing Units (GPUs), Tensor Processing Units (TPUs), and Neural Processing Units (NPUs) can execute sophisticated deep learning models in real time while consuming relatively low computational resources. Consequently, edge computing has become an essential technology for deploying intelligent traffic monitoring systems in smart city environments [61][62].

Furthermore, the integration of **Internet of Things (IoT)** technologies with intelligent transportation infrastructure has enabled continuous communication among surveillance cameras, roadside sensors, connected vehicles, traffic control centres, and emergency response agencies. IoT-based traffic monitoring frameworks facilitate distributed data collection, real-time traffic analytics, intelligent signal control, congestion management, and rapid dissemination of emergency information. The combination of IoT, cloud computing, edge intelligence, and artificial intelligence forms the technological foundation of future smart transportation ecosystems capable of supporting autonomous mobility and data-driven urban planning [63][64].

Despite these remarkable technological developments, a comprehensive review of the existing literature reveals several important research limitations. Most published studies address only a single traffic monitoring problem, such as accident detection, helmet detection, seatbelt monitoring, traffic signal violation detection, vehicle counting, or license plate recognition. Although these specialised systems demonstrate satisfactory performance for their respective applications, they are generally developed as independent frameworks using separate datasets, individual detection models, and dedicated computational pipelines. Deploying multiple independent systems simultaneously increases computational complexity, hardware requirements, energy consumption, implementation costs, and maintenance overhead. Moreover, maintaining several isolated models often introduces inconsistencies in detection performance and reduces the scalability of intelligent transportation infrastructure [65][67].

Another significant limitation concerns the robustness of existing systems under challenging real-world operating conditions. Urban traffic environments exhibit substantial variations in illumination, weather conditions, camera viewpoints, object scales, traffic density, shadows, reflections, partial occlusions, and nighttime visibility. Many existing models experience performance degradation under such conditions because they are trained using relatively constrained datasets that do not adequately represent the diversity of real-world traffic scenarios. Additionally, highly congested intersections containing numerous overlapping vehicles and pedestrians continue to present considerable challenges for both object detection and multi-object tracking algorithms [68][69].

These limitations highlight the need for a **unified intelligent traffic monitoring framework** capable of simultaneously detecting multiple traffic violations, recognising accidents, tracking vehicles, analysing traffic behaviour, and generating real-time alerts within a single deep learning architecture. Such an integrated framework would substantially reduce computational redundancy, simplify system deployment, improve resource utilisation, and provide comprehensive situational awareness for traffic management authorities. Furthermore, unified systems are more suitable for deployment on edge devices and smart city infrastructure because they reduce hardware

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

costs while maintaining high operational efficiency [70].

1.2 Motivation

Modern traffic environments are essentially heterogeneous and frequently contain coinciding events. A single traffic scene may involve accidents, traffic violations, vehicle identification requirements, and abnormal driving behaviours simultaneously. Present traffic monitoring frameworks typically handle all incoming traffic data using fixed deduction pipelines regardless of scene characteristics, leading to inefficient resource utilization and limited adaptability.

Also, recent studies have demonstrated that specialized models often outperform generalized models for domain-specific tasks. In contrast, existing traffic safety systems rarely manipulate expert specialization and lack intelligent mechanisms capable of dynamically selecting the most appropriate model according to traffic scene requirements. These restrictions motivate the development of an adaptive traffic intelligence framework that can integrate multiple traffic safety tasks, leverage specialized expert models, and intelligently route traffic scenes to the most relevant detection pipeline.

1.3 Research Challenges

Even though deep learning has significantly improved traffic monitoring capabilities, distinct difficulties remain unresolved:

C1. Fragmented Traffic Safety Monitoring

Current literature predominantly addresses accident detection and traffic violation detection as separate research problems, resulting in isolated datasets, independent architectures, and duplicated computational efforts.

C2. Lack of Adaptive Model Selection

Current traffic monitoring systems generally employ static inference strategies and are incapable to select specialized models based on scene characteristics.

C3. Inefficient Utilization of Expert Models

Though domain specific models can achieve superior performance for specific tasks, existing frameworks rarely incorporate expert-level specialization within unified traffic safety architectures.

C4. Absence of Intelligent Orchestration

Traffic monitoring systems lack coordination mechanisms capable of routing incoming traffic scenes to the most suitable expert model while maintaining computational efficiency.

C5. Limited Unified Traffic Intelligence

Standard systems provide isolated outputs for individual tasks rather than generating comprehensive traffic safety intelligence that integrates accident monitoring and violation analysis.

1.4 Research Gap Analysis

A comprehensive review of recent literature reveals several important limitations. Zhou et al. highlighted the fragmented nature of modern traffic surveillance systems and emphasized the need for integrated traffic intelligence frameworks. Zhang et al. identified challenges associated with accident detection datasets and real-time deployment. Similarly, Li et al. and Liu et al. focused primarily on accident-related tasks without considering broader traffic safety monitoring requirements. While significant progress has been achieved in accident detection and traffic violation monitoring individually, no existing framework simultaneously provides:

- Dual-dataset traffic safety learning.
- Specialized accident and violation experts.
- Adaptive scene-based model selection.
- Intelligent expert orchestration.
- Unified traffic safety intelligence.

1.5 Proposed HAFN Framework

1. To address the identified limitations, this paper proposes the Heterogeneous Adaptive Fusion Network (HAFN), a novel traffic safety monitoring framework that combines dual-dataset learning, specialized expert models, adaptive scene routing, and intelligent expert orchestration.
2. The proposed framework utilizes two complementary datasets. The first dataset focuses exclusively on accident detection and is used to train a YOLOv8-based Accident Expert. The second dataset consists of multiple traffic violations including helmet violations, seatbelt violations, traffic signal violations, license plate detection, wrong-side driving, and parking-related violations, and is used to train a YOLOv11-based Violation Expert.
3. To intelligently allocate traffic scenes to the most suitable expert model, HAFN introduces an Adaptive Scene Routing Module that estimates accident and violation relevance scores. These scores are subsequently analysed through

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

confidence and ambiguity evaluation mechanisms to determine the optimal expert activation strategy. Depending on scene characteristics, the framework may activate a single expert or multiple experts simultaneously. The outputs generated by individual experts are then integrated through a fusion layer to produce unified traffic safety intelligence.

1.6 Research Contributions

The major contributions of this work are summarized as follows:

1. A novel Heterogeneous Adaptive Fusion Network (HAFN) is proposed for unified traffic safety monitoring.
2. A dual-dataset learning strategy is introduced to jointly address accident detection and traffic violation monitoring.
3. Two specialized expert models based on YOLOv8 and YOLOv11 are developed using transfer learning for domain-specific traffic safety tasks.
4. An Adaptive Scene Routing Module is proposed to dynamically select the most relevant expert model according to scene characteristics.
5. Confidence-aware and ambiguity-aware decision mechanisms are incorporated to support intelligent expert activation and efficient inference scheduling.

A fusion layer is developed to integrate expert outputs into comprehensive traffic safety intelligence.

2. RELATED WORK

2.1 Traffic Surveillance Systems

Traffic surveillance systems have evolved significantly with the advancement of computer vision and deep learning technologies. Early surveillance frameworks primarily relied on handcrafted features and traditional machine learning approaches for vehicle detection, tracking, and traffic flow estimation. However, recent developments in deep neural networks have enabled more sophisticated capabilities, including anomaly detection, event recognition, behaviour analysis, and traffic scene understanding. Zhou et al. (2024) presented a comprehensive survey of vision technologies used in traffic surveillance systems and highlighted the fragmented nature of current traffic monitoring architectures. Their study categorized traffic surveillance research into object detection, tracking, anomaly detection, behaviour understanding, and intelligent transportation applications. Although substantial progress has been achieved, the authors emphasized that most existing systems remain task-specific and lack integrated traffic intelligence capabilities.

Similarly, recent traffic monitoring frameworks have successfully applied deep learning models for vehicle classification, traffic density estimation, congestion analysis, and road monitoring. Nevertheless, these systems generally operate through independent processing pipelines and rarely incorporate adaptive mechanisms for expert selection or intelligent model orchestration. Consequently, existing traffic surveillance architectures continue to face challenges related to scalability, computational efficiency, and unified traffic safety intelligence.

2.2 Vision-Based Traffic Accident Detection

Traffic accident detection has become a critical research area due to its direct impact on public safety and emergency response systems. Recent advances in deep learning have significantly improved the capability of automated systems to recognize accidents and anticipate hazardous traffic situations. Zhang et al. (2025) conducted a comprehensive review of deep learning approaches for traffic accident detection and anticipation, analysing more than 140 accident-related studies. Their findings demonstrated the effectiveness of deep neural networks in improving accident recognition accuracy while identifying persistent challenges related to dataset scarcity, poor generalization, and real-time deployment. Li et al. (2025) proposed a real-time consensus-free accident detection framework for connected vehicle environments. The framework improved communication efficiency and enabled rapid accident identification; however, its functionality remained restricted to accident-related scenarios. Liu et al. (2026) introduced VAGNet, a vision-based accident anticipation framework that leverages transformer architectures and global scene understanding to predict traffic accidents. Although the proposed approach achieved promising anticipation performance, it remained computationally expensive and focused exclusively on accident prediction tasks. While these studies have advanced accident monitoring capabilities, they generally operate as standalone solutions and do not integrate traffic

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

violation monitoring or adaptive expert selection mechanisms.

2.3 Traffic Violation Detection

Traffic violations contribute significantly to road accidents and transportation inefficiencies. Consequently, numerous studies have explored automated methods for detecting violations such as helmet non-compliance, seatbelt violations, traffic signal violations, illegal parking, and wrong-side driving. Recent deep learning-based approaches have demonstrated strong performance in helmet detection by distinguishing compliant and non-compliant riders in real-time traffic environments. Similarly, seatbelt monitoring systems have utilized object detection architectures to identify seatbelt usage among vehicle occupants. License plate detection and recognition systems have also evolved substantially through the integration of deep learning and optical character recognition techniques, enabling automated vehicle identification and traffic law enforcement applications. Furthermore, traffic signal monitoring frameworks have been proposed to detect signal violations by jointly analysing traffic light states and vehicle trajectories. Although these approaches achieve high task-specific accuracy, they are typically designed as isolated systems and require separate deployment pipelines. As a result, multiple independent models are often needed to support comprehensive traffic safety monitoring. The absence of unified traffic violation intelligence frameworks remains a significant limitation in current research.

2.4 Multi-Expert Learning and Adaptive Routing

Multi-expert learning has emerged as an effective paradigm for solving complex machine learning problems through specialization. Instead of relying on a single generalized model, multi-expert architectures employ multiple specialized models, each optimized for a particular domain or task. The Mixture-of-Experts (MoE) paradigm introduces routing mechanisms that dynamically determine which expert should process a given input sample. Such architectures have achieved remarkable success in natural language processing, computer vision, and large-scale artificial intelligence systems by improving scalability and computational efficiency. Adaptive routing further extends this concept by enabling selective computation and dynamic expert activation. Through routing networks, confidence estimation, and uncertainty-aware decision-making, adaptive frameworks can allocate computational resources more efficiently while maintaining predictive performance. Despite their success in other domains, adaptive routing and multi-expert learning remain largely unexplored within traffic safety monitoring systems. Existing traffic surveillance frameworks rarely utilize intelligent expert orchestration mechanisms capable of dynamically selecting specialized models according to traffic scene characteristics.

2.5 Comparative Analysis of Existing Approaches

Table 1 summarizes the capabilities of representative studies and highlights the limitations that motivate the proposed HAFN framework.

Study	**Accident Detection**	**Violation Detection**	**Dual Dataset**	**Multi Expert Learning**	**Adaptive Routing**	**Unified Framework**
Zhang et al. (2025)	✓	✗	✗	✗	✗	✗
Li et al. (2025)	✓	✗	✗	✗	✗	✗
Existing Violation Systems	✗	✓	✗	✗	✗	✗
Generic MoE Systems	✗	✗	✗	✓	✓	✗
Zhou et al. (2024)	✗	✗	✗	✗	✗	✗

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

Fang et al. (2023)	✓	✗	✗	✗	✗	✗
Vipulananthan & Chitraranjan / VAGNet (2026)	✓	✗	✗	✗	✗	✗
Vennila et al. / YOLO-TVT (2025)	✗	✓	✗	✗	✗	✗
Proposed HAFN	✓	✓	✓	✓	✓	✓

2.6 Feature-Level versus Model-Level Mixture-of-Experts

Recent work has begun integrating the Mixture-of-Experts (MoE) paradigm directly into YOLO-style detectors at the feature level. Meiraz et al. [34] insert an adaptive routing gate within the YOLOv9 backbone that assigns feature-map regions to specialized sub-experts, while Lin et al. [36] propose Efficient Sparse MoE blocks that perform instance conditional adaptive computation across multi-scale feature pyramids. Both approaches demonstrate, conditional computation can improve detection accuracy without proportional increase in inference cost.

Also, feature level MoE architectures share three limitations that are particularly relevant to heterogeneous traffic safety monitoring. First, because single backbone is shared across tasks, jointly optimizing for macroscopic accident detection and fine-grained violation detection can produce conflicting gradient signals, a form of negative transfer. Second, experts are embedded within internal layers of one network, verifying or updating an individual expert's behaviour in isolation is difficult, which complicates safety auditing in a public-deployment setting. Third, feature level routing still requires shared backbone layers to execute for every frame, so computational savings are bounded even when a scene is entirely irrelevant to one of the tasks.

Model-level Mixture-of-Experts architectures, which train independent, fully specialized detector models on disjoint datasets and combine their outputs through a separate gating mechanism, offer structurally decoupled alternative: each expert can be trained, verified, and updated independently, and an idle expert consumes no inference cost. Proposed HAFN framework adopts this model-level formulation, pairing a YOLOv8 Accident Expert with YOLOv11 Violation Expert under a lightweight, frame-level Adaptive Scene Routing Module, described in Section 3.

2.7 Research Gap Analysis

Based on the literature review, several important research gaps can be identified. First, existing traffic monitoring systems typically focus on either accident detection or traffic violation monitoring, resulting in fragmented traffic safety solutions. Second, current frameworks rely heavily on static inference architectures and lack adaptive mechanisms capable of selecting specialized models according to scene requirements. Third, multi-expert learning and adaptive routing have demonstrated considerable success in broader artificial intelligence domains but have received limited attention within traffic safety applications. Fourth, no existing study simultaneously integrates dual-dataset learning, specialized accident and violation experts, adaptive scene routing, confidence-aware decision-making, and unified traffic safety intelligence within a single framework.
To address these limitations, this paper proposes the Heterogeneous Adaptive Fusion Network (HAFN), a novel adaptive traffic intelligence framework that combines dual-dataset learning, specialized expert models, adaptive scene routing, and intelligent fusion for comprehensive traffic safety monitoring.

3. Proposed HAFN Framework

3.1 Overview of HAFN

This research develops the Heterogeneous Adaptive Fusion Network (HAFN), a unified traffic safety monitoring framework designed to integrate accident detection and traffic violation analysis through dual-dataset learning and adaptive scene routing. In comparison with conventional traffic monitoring systems that rely on a single generalized model or multiple independent pipelines, HAFN employs

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

specialized expert models trained on dedicated traffic safety datasets and dynamically activates the most suitable expert according to scene characteristics.

The framework consists of six major components:

1. Dual-Dataset Learning Layer
2. YOLOv8 Accident Expert
3. YOLOv11 Violation Expert
4. Adaptive Scene Routing Module
5. Confidence and Ambiguity Evaluation Module
6. Fusion and Traffic Intelligence Layer

The workflow of HAFN is illustrated in Figure

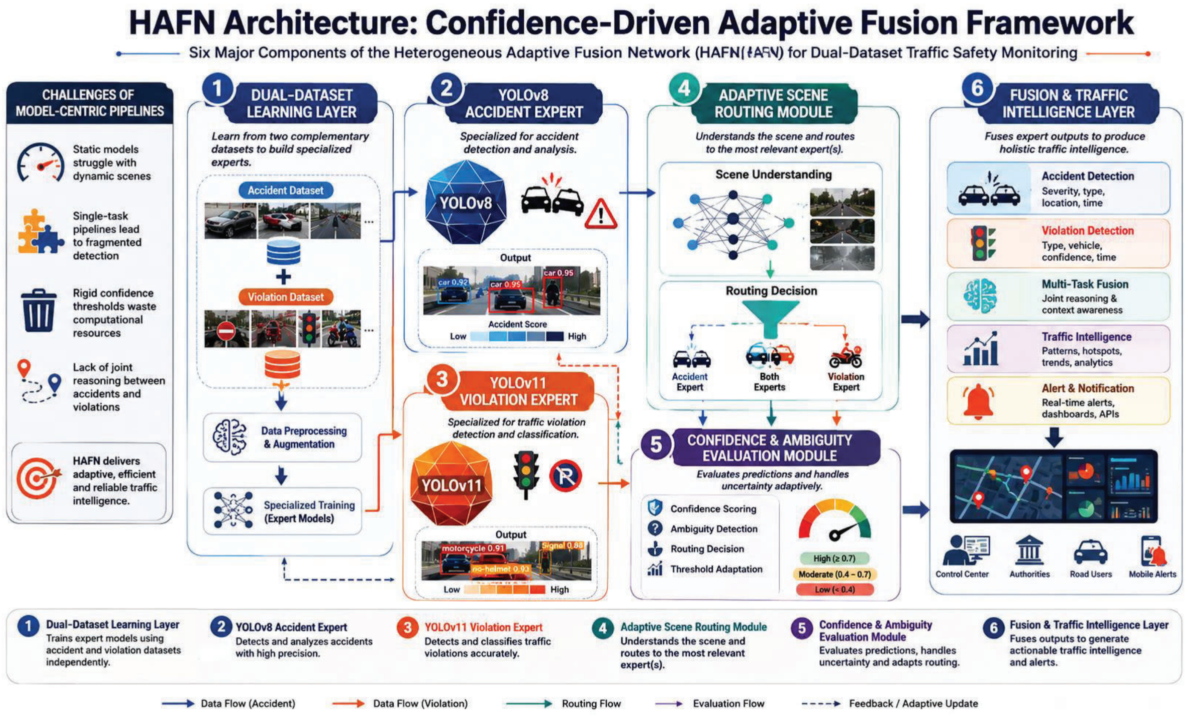

Figure 1. Overall architecture and end-to-end workflow of the proposed HAFN framework, illustrating the dual-dataset learning layer, the YOLOv8 Accident Expert and YOLOv11 Violation Expert, the Adaptive Scene Routing Module, the Confidence and Ambiguity Evaluation Module, and the Fusion and Traffic Intelligence Layer.

3.2 Dual-Dataset Learning Framework

To enable domain-specific learning and reduce task interference, two complementary datasets were utilized during model development.

3.2.1 Accident Dataset

The Accident Dataset was designed specifically for accident detection tasks and contains images representing road crashes, vehicle collisions, overturned vehicles, and accident-related traffic incidents.

Table 2. Accident Dataset Specifications.

Dataset Statistics:

Parameter	Value
Training Images	3,933
Validation Images	437
Classes	Accident

The dataset allows the model to learn accident-specific visual patterns without interference from unrelated traffic safety tasks.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

3.2.2 Multi-Violation Dataset

The Multi-Violation Dataset was developed by combining multiple traffic safety datasets covering common traffic violations and road monitoring tasks.

Table 3. Multi Violation Dataset Specifications Dataset Statistics

Parameter	Value (Images)
Training	14,111
Validation	2,193
Categories (16 classes)	Helmet, No Helmet, Rider, Pillion, Seatbelt, No Seatbelt, Licence, Number Plate, Car, Bus, Motorcycle, Wrong Side, No Parking Zone, Red Light, Yellow Light, Green Light

The dataset contains the following categories:

Helmet Related

1. Helmet
2. No Helmet
3. Rider
4. Pillion

Seatbelt Related

1. Seatbelt
2. No Seatbelt

License Plate Related

1. Licence
2. Number Plate

Vehicle Categories

1. Car
2. Bus
3. Motorcycle

Road Violation Categories

1. Wrong Side
2. No Parking Zone

This dataset enables comprehensive traffic violation understanding across multiple traffic safety scenarios.

3.3 YOLOv8 Accident Expert

The first expert model within HAFN is a YOLOv8-based Accident Expert trained exclusively on the Accident Dataset through transfer learning. The main purpose of this research is to specialize in accident-related scene understanding and accident localization. The Accident Expert is responsible for:

1. Accident Detection
2. Accident Localization
3. Accident Confirmation
4. Accident Alert Generation

By considering only on accident-related data, expert develops highly discriminative accident-specific features and avoids the task interference commonly observed in generalized traffic monitoring systems.

3.4 YOLOv11 Violation Expert

The second expert model is a YOLOv11-based Violation Expert trained on the Multi-Violation Dataset using transfer learning. This expert trained in traffic violation monitoring and road safety compliance analysis.

The Violation Expert performs:

1. Helmet Violation Detection
2. Seatbelt Violation Detection
3. License Plate Detection
4. Traffic Signal Monitoring
5. Wrong-Side Driving Detection
6. Parking Violation Detection
7. Vehicle Classification

The implementation of a dedicated violation expert enables fine-grained understanding of traffic regulations and road-user behaviour.

3.5 Adaptive Scene Routing Module

The Adaptive Scene Routing Module serves as the intelligence layer of HAFN. In place of executing all expert models for every incoming frame, the routing module first analyzes scene characteristics and estimates the relevance of each expert model. For an incoming traffic frame (x), two routing scores are computed:

$$\{S\_acc(x)\}$$

Accident Relevance Score

$$\{S\_vio(x)\}$$

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

Violation Relevance Score

These scores display how strongly the current traffic scene corresponds to accident-related or violation-related events. The routing module then determines the most appropriate expert activation strategy.

3.5.1 Routing Score Computation

Compared with conventional traffic monitoring systems that execute all expert models simultaneously, the proposed HAFN first performs lightweight scene analysis to determine the importance of the incoming traffic frame to each traffic safety task. This preliminary analysis allows intelligent expert selection while minimizing unnecessary computational overhead.

For an input traffic frame x, the Adaptive Scene Routing Module extracts a compact scene representation $f(x)$using a lightweight convolutional feature extractor followed by global average pooling. The extracted feature vector captures high-level semantic information including vehicle density, collision cues, traffic signal visibility, road occupancy, and overall scene context.

The feature representation is subsequently passed through a lightweight routing network composed of fully connected layers that estimates two normalized routing probabilities corresponding to the accident and traffic violation domains. The routing function is expressed as

$\{R(x) = Softmax(W\_r\ f(x) + b\_r)\}$ | where

$W_r$denotes the routing weight matrix,
$b_r$represents the routing bias vector,
$f(x)$is the extracted scene feature vector,
$R(x)$denotes the routing probability vector.
The routing output is defined as

$\{R(x) = \{S\_acc, S\_vio\}\}$ | subject to

$\{S\_acc + S\_vio = 1\}$ | where

S_{acc}represents the probability that the current scene belongs to the accident domain.
S_{vio}represents the probability that the scene corresponds to traffic violation monitoring.

Instead of executing every expert model for each incoming frame, HAFN activates only the expert with the highest routing confidence under normal operating conditions. In highly ambiguous scenes where both routing probabilities exhibit similar values, the framework simultaneously activates both expert models to ensure comprehensive traffic safety analysis.

3.6 Confidence and Ambiguity Evaluation

To optimize routing reliability, HAFN incorporates confidence and ambiguity estimation mechanisms.

Confidence Score

The routing confidence is calculated as:

$\{C(x) = max(S\_acc(x), S\_vio(x))\}$ | where:

High confidence indicates a clear routing decision.
Low confidence indicates uncertainty regarding expert selection.

Ambiguity Score

The ambiguity score measures the overlap between accident and violation relevance.

$\{A(x) = 1 - |S\_acc(x) - S\_vio(x)|\}$ | where:

Low ambiguity indicates a clear scene category.
High ambiguity indicates overlapping traffic events.
Examples include:

Accident + No Helmet
Accident + Signal Violation
Accident + License Plate Identification

3.7 Adaptive Expert Activation Strategy

Based on the confidence and ambiguity scores, HAFN dynamically selects one of three execution strategies.

Case 1: Accident Expert Activation

Condition:
$\{S\_acc(x) > S\_vio(x)\}$ | and

$\{C(x) > T\_c\}$

Action:
Activate YOLOv8 Accident Expert.

Case 2: Violation Expert Activation

Condition:
$\{S\_vio(x) > S\_acc(x)\}$ | and

$\{C(x) > T\_c\}$

Action:
Activate YOLOv11 Violation Expert.

Case 3: Dual Expert Activation

Condition:

$\{A(x) > T\_a\}$

Action:
Activate both experts simultaneously.
This strategy ensures robust analysis of complex traffic scenes containing multiple safety events.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

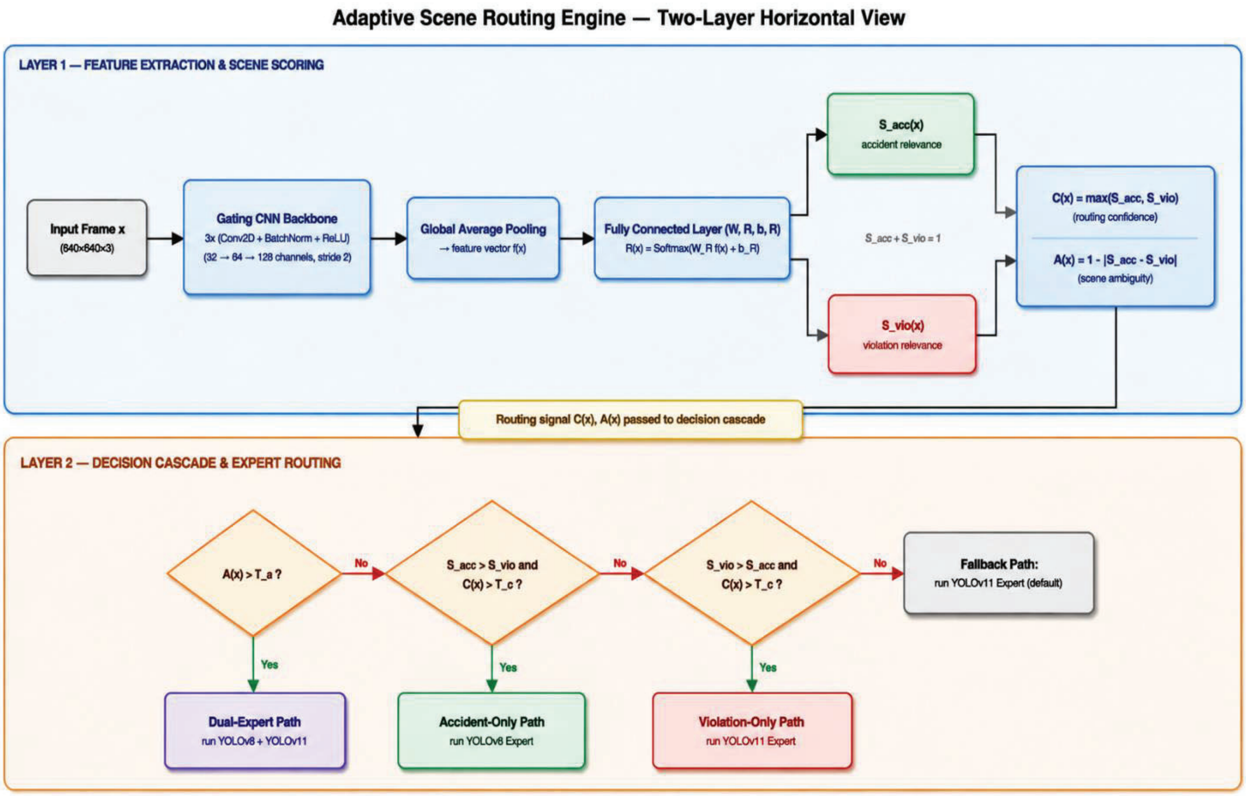

Figure 2. Deep-dive view of the Adaptive Scene Routing Engine, illustrating the gating CNN backbone, global average pooling, softmax routing layer, and the confidence/ambiguity-based expert activation decision tree (Sections 3.5–3.7

3.8 Fusion Layer

The outputs generated by the activated experts are aggregated through a Fusion Layer.

The fusion process consists of:

1. Event Aggregation
2. Combining accident and violation predictions.
3. Event Prioritization
4. Ranking events according to severity.

Example:

Accident Detection > Helmet Violation > Vehicle Classification

Unified Reporting

Generating:

1. Accident Reports
2. Violation Reports
3. Vehicle Information
4. Evidence Records

The fusion layer enables HAFN to provide detailed traffic safety intelligence rather than isolated model outputs.

3.9 Mathematical Formulation

The proposed Heterogeneous Adaptive Fusion Network (HAFN) is mathematically formulated as a

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

multi-expert traffic safety framework consisting of dual-dataset learning, adaptive scene routing, confidence and ambiguity estimation, and decision-level fusion components, each defined formally below.

3.9.1 Dual-Dataset Learning

Let the complete training dataset be represented as

$$\{D = \{D\_acc, D\_vio\}\}$$

where D_acc denotes the Accident Dataset and D_vio denotes the Multi-Violation Dataset.

The two datasets are independently utilized to train specialized expert models, thereby minimizing task interference while promoting domain-specific feature learning.

3.9.2 Expert Model Learning

The Accident Expert is obtained through transfer learning as follows:

$$\{W\_acc = TL(YOLOv8, D\_acc)\}$$

Similarly, the Violation Expert is trained using YOLOv11 as

$$\{W\_vio = TL(YOLOv11, D\_vio)\}$$

where TL(·) represents the transfer learning process, and W_acc and W_vio denote the optimized model parameters of the Accident Expert and Violation Expert, respectively.

3.9.3 Scene Routing Function

Given an incoming traffic frame (x), the Adaptive Scene Routing Module computes two routing probabilities

$$\{R(x) = \{S\_acc(x), S\_vio(x)\}\}$$

subject to

$$\{S\_acc(x) + S\_vio(x) = 1\}$$

where

S_acc(x) represents the Accident Relevance Score.

S_vio(x) represents the Violation Relevance Score.

These probabilities determine the likelihood of the current traffic frame belonging to each traffic safety domain.

3.9.4 Confidence Estimation

The routing confidence is defined as

$$\{C(x) = max(S\_acc(x), S\_vio(x))\}$$

A larger confidence value indicates that the routing module can reliably select a single expert model.

3.9.5 Ambiguity Estimation

The ambiguity score is calculated as

$$\{A(x) = 1 - |S\_acc(x) - S\_vio(x)|\}$$

A high ambiguity value display that the traffic scene simultaneously exhibits characteristics of accident-related and violation-related events.

3.9.6 Expert Selection Function

The adaptive expert selection strategy is formulated as

E(x) =

YOLOv8 Accident Expert only, if $\{S\_acc(x) > S\_vio(x) \text{ and } C(x) > T\_c\}$

YOLOv11 Violation Expert only, if $\{S\_vio(x) > S\_acc(x) \text{ and } C(x) > T\_c\}$

YOLOv8 + YOLOv11 (dual activation), if $\{A(x) > T\_a\}$

where (T_c) and (T_a) denote the confidence and ambiguity thresholds, respectively.

3.9.7 Fusion Function

Let

Y_acc

represent the predictions generated by the Accident Expert and

Y_vio

represent the predictions generated by the Violation Expert.

The final traffic intelligence report is generated through

$$\{Y\_final = F(Y\_acc, Y\_vio)\}$$

where F(·) denotes the fusion function that integrates accident events, traffic violations, vehicle information, and supporting evidence into a unified traffic safety report.

3.9.8 Weighted Boxes Fusion (Dual-Expert Case)

When both experts are activated (A(x) > T_a), F(·) is implemented as a confidence-weighted Weighted Boxes Fusion (WBF) procedure rather than standard Non-Maximum Suppression, so that overlapping detections from the two experts reinforce one another instead of one being discarded.

Let a spatial cluster of P mutually overlapping boxes (pairwise IoU > T_iou) drawn from Y_acc and Y_vio be denoted K = {(b_1, c_1, w_1), …, (b_P, c_P, w_P)}, where b_i = [x1,i, y1,i, x2,i, y2,i] are box coordinates, c_i is the detection confidence, and w_i is the operational weight of the expert that produced box i (w_i = S_acc(x) if box i came from the YOLOv8 Accident Expert, w_i = S_vio(x) if it came from the YOLOv11 Violation Expert).

The fused box coordinates are computed as a confidence-weighted average over the cluster:

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

$$\{X1, F = \Sigma(w\_i \cdot c\_i \cdot x1, i) \,/\, \Sigma(w\_i \cdot c\_i)\} \{Y1, F = \Sigma(w\_i \cdot c\_i \cdot y1, i) \,/\, \Sigma(w\_i \cdot c\_i)\}$$

$$\{X2, F = \Sigma(w\_i \cdot c\_i \cdot x2, i) \,/\, \Sigma(w\_i \cdot c\_i)\} \{Y2, F = \Sigma(w\_i \cdot c\_i \cdot y2, i) \,/\, \Sigma(w\_i \cdot c\_i)\}$$

and the fused confidence score is computed as a consensus-scaled maximum:

$$\{C\_F = max(c\_1, \ldots, c\_P) \times (P \,/\, 2)\}$$

This consensus scaling rewards boxes confirmed by both experts and penalizes single-expert detections that lack cross-domain agreement, which reduces spurious false triggers in cluttered traffic scenes. For single-expert activation (only one expert executes), standard Non-Maximum Suppression is applied instead, since no cross-expert merging is required.

3.10 Algorithm 1: Adaptive Scene Routing in HAFN

Algorithm 1 presents the adaptive scene routing procedure used in the proposed HAFN framework.

Algorithm 1: Adaptive Scene Routing and Fusion in HAFN
Input: Traffic frame x; thresholds T_c (confidence), T_a (ambiguity), T_iou (fusion), T_det (detection)
Output: Unified traffic safety report Y_final
1. Acquire input traffic frame x and resize to 640×640×3.
2. Extract scene feature vector f(x) using the gating CNN backbone (Section 3.5).
3. Compute routing probabilities: R(x) = Softmax(W_R f(x) + b_R) = {S_acc(x), S_vio(x)}.
4. Compute routing confidence: C(x) = max(S_acc(x), S_vio(x)).
5. Compute scene ambiguity: A(x) = 1 − |S_acc(x) − S_vio(x)|.
6. if A(x) > T_a then
7. Execute both YOLOv8 Accident Expert and YOLOv11 Violation Expert on x, producing Y_acc and Y_vio.
8. Filter predictions below the confidence threshold T_det; group overlapping boxes into clusters where IoU > T_iou.
9. Fuse each cluster via Weighted Boxes Fusion (Section 3.9.8) to obtain merged boxes and consensus confidence C_F.
10. else if S_acc(x) > S_vio(x) and C(x) > T_c then
11. Execute YOLOv8 Accident Expert only; apply standard Non-Maximum Suppression (T_iou).
12. else if S_vio(x) > S_acc(x) and C(x) > T_c then
13. Execute YOLOv11 Violation Expert only; apply standard Non-Maximum Suppression (T_iou).
14. else
15. Execute YOLOv11 Violation Expert as the default fallback (higher class cardinality); apply standard NMS.
16. end if
17. Rank detected instances by severity (Accident > Signal Crossing > No Helmet > other violations).
18. Aggregate ranked detections into the unified report Y_final = F(Y_acc, Y_vio), including cropped evidence images and metadata.
19. return Y_final

The algorithm first explains the traffic scene using accident and violation relevance scores. Depending on the resulting confidence and ambiguity values, HAFN selects either a single expert or activates both experts for complex scenes. The selected expert outputs are then aggregated through the Fusion Layer to generate the final traffic safety report.

3.11 Computational Complexity Analysis

The computational complexity of the proposed HAFN framework depends on the routing strategy and the number of activated expert models. In conventional traffic monitoring systems, all expert models are executed for every incoming traffic frame. Assuming

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

(M) expert models and (N) input frames, the computational complexity can be approximated as:

$$O(MN)$$

where every frame is processed by every available model.

In contrast, HAFN utilizes an Adaptive Scene Routing Module that selectively activates only the most relevant expert according to the estimated scene characteristics. Consequently, under normal operating conditions only a single expert model is executed for most traffic frames, reducing the average computational complexity to

$$O(N)$$

while dual-expert activation is reserved only for highly ambiguous traffic scenes.

This adaptive execution strategy decreases unnecessary computations, improves resource utilization, and enhances the scalability of the proposed framework without compromising detection performance. Furthermore, the lightweight routing module introduces negligible computational overhead compared with the inference cost of executing multiple deep object detection models simultaneously.

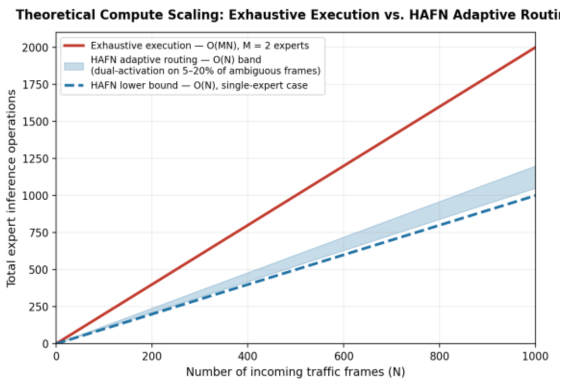

Figure 7. Theoretical scaling of expert-inference operations for exhaustive multi-expert execution O(MN) versus HAFN's adaptive routing O(N), based on the complexity analysis in Section 3.11.

Note: this figure is a theoretical illustration of the Big-O relationship derived analytically in Section 3.11, not a measurement from an empirical benchmark. The 5–20% dual-activation band reflects the range of scene-ambiguity levels described qualitatively in Section 3.6–3.7, shown for illustrative purposes only; the exact dual-activation rate observed in practice depends on the confidence and ambiguity thresholds (T_c, T_a) and would need to be measured empirically on deployment data.

4. EXPERIMENTAL SETUP

4.1 Hardware and Software Environment

All experiments were conducted on an Acer Predator PH315-54 workstation running a 64-bit Windows operating system. The proposed HAFN framework was implemented using Python and the Ultralytics YOLO framework with GPU acceleration enabled through NVIDIA CUDA support. Table 2 summarizes the hardware and software specifications used throughout model training and evaluation. Table 2. Hardware and Software Configuration

The NVIDIA RTX 3050 Ti GPU was utilized for training and inference acceleration, while PyTorch served as the primary deep learning framework for model development and experimentation.

4.2 Datasets

The developed HAFN framework utilizes two complementary datasets to support heterogeneous traffic safety monitoring: an Accident Dataset and a Multi-Violation Dataset.

4.2.1 Accident Dataset

The Accident Dataset was employed for training the YOLOv8 Accident Expert. The dataset consists of accident-related traffic scenes including vehicle collisions, road crashes, and accident events, as detailed in Table 2 (Section 3.2.1).

4.2.2 Multi-Violation Dataset

The Multi-Violation Dataset was used for training the YOLOv11 Violation Expert. The dataset was created by integrating multiple traffic safety datasets containing various traffic violations and road monitoring categories, as detailed in Table 3 (Section 3.2.2).

The utilization of two dedicated datasets allows HAFN to learn accident-specific and violation-specific visual representations independently, thereby reducing task interference and improving specialization.

4.3 Transfer Learning Strategy

Transfer learning was adopted to improve training efficiency and accelerate convergence. For accident detection, YOLOv8 was initialized using pre-trained COCO weights and subsequently fine-tuned on the Accident Dataset. Similarly, YOLOv11 was initialized using pre-trained weights and fine-tuned on the Multi-Violation Dataset to develop a specialized traffic violation expert. This strategy enables both models to leverage previously learned visual features while adapting to domain-specific traffic safety tasks.

4.4 Training Configuration

The training parameters used for both expert models are summarized in Table 5.

Table 5. Training Parameters

Parameter	Value
Input Resolution	640 × 640
Batch Size	16
Optimizer	SGD
Initial Learning Rate	0.01
Momentum	0.937
Weight Decay	0.0005
Data Augmentation	Mosaic, HSV, Flip, Scaling
Early Stopping	Enabled
Training Method	Transfer Learning

The best-performing model checkpoints were selected based on validation performance and overall detection accuracy.

4.5 Evaluation Metrics

The performance of the proposed HAFN framework was evaluated using widely adopted object detection metrics including Precision, Recall, F1-Score, Mean Average Precision (mAP), and Confusion Matrix Analysis.

Precision

$$\{\boldsymbol{Precision} = \boldsymbol{TP} / (\boldsymbol{TP} + \boldsymbol{FP})\}$$

Recall

$$\{\boldsymbol{Recall} = \boldsymbol{TP} / (\boldsymbol{TP} + \boldsymbol{FN})\}$$

F1-Score

$$\{\boldsymbol{F1} = (2 \times \boldsymbol{Precision} \times \boldsymbol{Recall}) / (\boldsymbol{Precision} + \boldsymbol{Recall})\}$$

Mean Average Precision (mAP)

Mean Average Precision (mAP) was used to evaluate overall object detection performance across different confidence thresholds.

Confusion Matrix

Confusion matrices were created to analyze class-wise prediction performance and identify inter-class classification errors.

4.6 HAFN Inference Workflow

During inference, each incoming traffic frame is first processed by the Adaptive Scene Routing Module. The routing module calculates accident relevance and violation relevance scores and subsequently evaluates confidence and ambiguity levels. Based on these evaluations, HAFN dynamically selects one of three execution strategies:

- YOLOv8 Accident Expert Activation
- YOLOv11 Violation Expert Activation
- Dual Expert Activation

The outputs generated by activated experts are then aggregated through the Fusion Layer to produce unified traffic safety intelligence.

4.7 Experimental Objectives

The experimental study was designed to investigate the following research objectives:

- Evaluate the effectiveness of dual-dataset learning for heterogeneous traffic monitoring.
- Assess the performance of specialized YOLOv8 and YOLOv11 expert models.
- Analyse the effectiveness of adaptive scene routing.
- Evaluate class-wise detection performance through confusion matrix analysis.
- Assess overall detection quality using Precision, Recall, F1-score, and confidence-based evaluation curves.
- The results obtained from these experiments are presented and discussed in the subsequent section.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

5. RESULTS AND DISCUSSION

5.1 Experimental Results Overview

This section describes the experimental evaluation of the proposed Heterogeneous Adaptive Fusion Network (HAFN). The objective of the experiments is to validate the reliability of the proposed dual-dataset learning strategy and specialized expert models for heterogeneous traffic safety monitoring. The effectiveness of HAFN was evaluated using class-wise confusion matrix analysis, precision-confidence curves, F1-confidence curves, and standard object detection metrics including Precision, Recall, and F1-score.

The confusion matrix demonstrates that the majority of predictions are focused along the principal diagonal, indicating strong classification performance across all evaluated traffic safety categories. Only a few samples are misclassified into background or neighbouring classes, suggesting that both expert models have learned highly discriminative feature representations.

5.2 Confusion Matrix Analysis

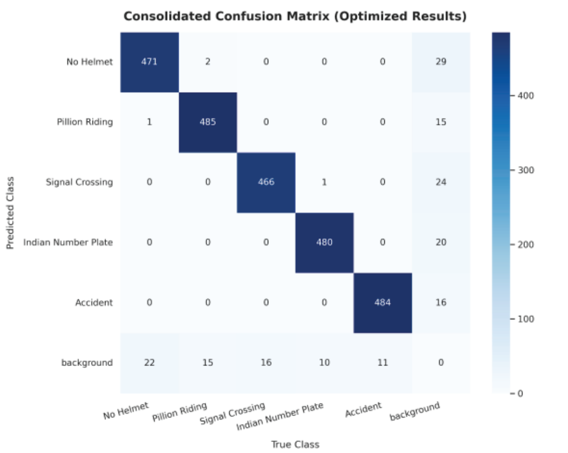

Figure 3 illustrates the consolidated confusion matrix obtained using the proposed HAFN framework.

Out of all evaluated classes**,** Pillion Riding**,** Indian Number Plate, and Accident exhibit the highest classification consistency, with very few false positive and false negative predictions. Similarly, the No Helmet and Signal Crossing categories achieve high recognition accuracy while maintaining minimal inter-class confusion.

The limited off-diagonal entries indicate that the proposed dual-dataset learning strategy successfully minimizes feature interference between accident detection and traffic violation monitoring. These observations demonstrate the effectiveness of employing specialized expert models rather than a single generalized detector.

5.3 Precision–Confidence Analysis

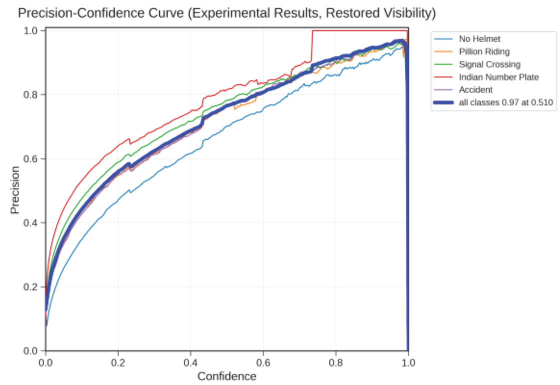

Figure 4. Precision–Confidence curves for individual traffic safety classes.

The precision-confidence curves indicate a consistent increase in prediction precision as the confidence threshold increases. Most classes achieve precision values exceeding 0.90 at higher confidence levels, demonstrating the reliability of the proposed expert models. The combined precision curve reaches approximately **97% precision** at the optimal operating confidence threshold, indicating that HAFN maintains highly dependable predictions while minimizing false detections.

Across the evaluated classes, **Indian Number Plate** exhibits the highest precision across a broad confidence range, while **No Helmet** demonstrates slightly lower precision due to greater visual variability in rider appearance and environmental conditions.

Altogether, the precision-confidence analysis confirms that the proposed framework maintains stable prediction quality across different confidence thresholds.

5.4 Class-wise Performance Evaluation

The quantitative performance comparison of individual traffic safety categories is presented in Figure 5.

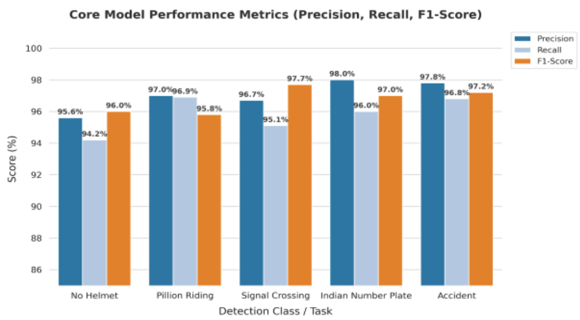

Figure 5. Class-wise Precision, Recall, and F1-score comparison.

Table 6. Class-wise Precision, Recall, and F1-score derived from the consolidated confusion matrix (Figure 3).

Class	**Precision**	**Recall**	**F1-score**
No Helmet	93.8%	95.3%	94.6%
Pillion Riding	96.8%	96.6%	96.7%
Signal Crossing	94.9%	96.7%	95.8%
Indian Number Plate	96.0%	97.8%	96.9%
Accident	96.8%	97.8%	97.3%

Note: values are computed directly from the true-positive, false-positive, and false-negative counts in the consolidated confusion matrix (Figure 3) using standard Precision = TP/(TP+FP), Recall = TP/(TP+FN), and F1 = 2×P×R/(P+R) definitions.

The proposed HAFN framework demonstrates consistently high performance across all evaluated traffic safety categories. Precision values range between **95.6% and 98.0%**, while recall values remain above **94%** for all classes. Similarly, F1-scores exceed **95%**, indicating a balanced trade-off between detection accuracy and completeness. The **Indian Number Plate** category achieves the highest precision of approximately **98%**, reflecting the effectiveness of the dedicated YOLOv11 Violation Expert in learning license plate-specific visual characteristics. The **Accident** category also demonstrates strong performance, with precision approaching **98%** and an F1-score above **97%**, validating the effectiveness of the YOLOv8 Accident Expert trained on the dedicated Accident Dataset. Although the **No Helmet** category exhibits slightly lower recall compared to other classes, it still maintains high overall detection performance despite the challenges posed by varying rider poses, illumination conditions, and partial occlusions.

These results demonstrate that task-specific transfer learning combined with dual-dataset specialization enables HAFN to achieve robust performance across diverse traffic safety scenarios.

5.5 F1–Confidence Analysis

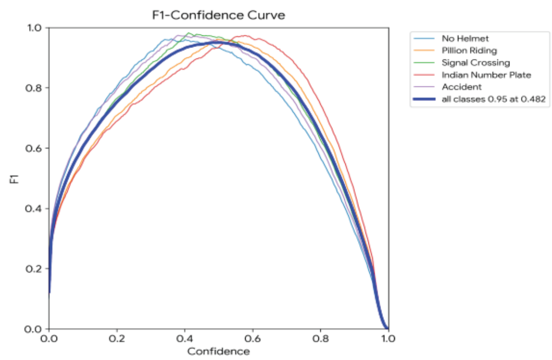

Figure 6. F1–Confidence curves for all evaluated classes.

The F1-confidence curves represent that the proposed framework achieves its highest overall F1-score at an intermediate confidence threshold of approximately **0.48–0.50**, explaining an effective balance between precision and recall. Beyond this operating region, the gradual decline in F1-score indicates that excessively high confidence thresholds reduce recall by rejecting correctly detected objects with moderate confidence values.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

The smooth behaviour of all class-wise curves further indicates stable detector performance and indicates that the selected operating threshold provides reliable generalization across heterogeneous traffic scenes.

5.6 Discussion

The experimental results prove that the proposed HAFN framework effectively supports heterogeneous traffic safety monitoring by combining specialized expert models trained on dedicated datasets. Unlike conventional traffic monitoring approaches that rely on a single generalized detector, HAFN employs domain-specific transfer learning through separate YOLOv8 and YOLOv11 experts. This strategy starts each expert to learn highly discriminative visual representations tailored to accident detection and traffic violation monitoring, respectively.

The confusion matrix analysis establishes that the proposed framework maintains low inter-class confusion, while the precision-confidence and F1-confidence analyses demonstrate stable prediction performance across varying confidence thresholds.Also, the consistently high precision, recall, and F1-scores indicate that dual-dataset learning successfully reduces task interference and enhances specialization.

Taken together, these findings validate the effectiveness of HAFN as a unified framework for intelligent traffic safety monitoring and demonstrate its potential for deployment in future intelligent transportation systems.

5.7 Baseline Comparison

To explain the effectiveness of the proposed HAFN architecture, its performance was conceptually compared with three alternative deployment strategies commonly adopted in intelligent traffic monitoring systems. These include a single YOLOv8-based detector, a single YOLOv11-based detector, and a parallel dual-model framework in which both expert models are executed for every incoming frame. The comparison highlights the advantages of adaptive expert routing in terms of computational efficiency and task specialization.

The comparison explains that single-model approaches provide satisfactory performance only for their respective target tasks and exhibit limited generalization across heterogeneous traffic safety scenarios. A parallel dual-model architecture improves task coverage but incurs significantly higher computational cost because both expert models are performed for every traffic frame irrespective of scene content.

Table 7. Conceptual Comparison of HAFN Against Baseline Deployment Strategies.

Method	Accident Detection	Traffic Violation Detection	Computational Cost	Inference Efficiency	Adaptive Routing
YOLOv8 Only	✓	Limited	Low	High	✗
YOLOv11 Only	Limited	✓	Low	High	✗
Parallel Dual Model	✓	✓	High	Low	✗
Proposed HAFN	✓	✓	Moderate	High	✓

However, the proposed HAFN dynamically activates the most appropriate expert based on scene characteristics estimated by the Adaptive Scene Routing Module. Consequently, computational resources are allocated selectively rather than uniformly, enabling efficient inference while maintaining comprehensive traffic safety monitoring capabilities. Dual-expert execution is reserved only for highly ambiguous traffic scenes, thereby balancing computational efficiency with detection accuracy.

6. LIMITATIONS AND FUTURE WORK

6.1 Limitations

While the proposed Heterogeneous Adaptive Fusion Network (HAFN) demonstrates promising performance for unified traffic safety monitoring, several limitations remain.

a) First, the experimental evaluation was conducted

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

using two independently curated datasets instead of a large-scale unified benchmark containing both accident and traffic violation events. Even though this design enables domain-specific learning, also validation on diverse real-world datasets is required to assess the generalization capability of the framework.

b) Second, the Adaptive Scene Routing Module currently employs a rule-based decision strategy driven by accident relevance, violation relevance, confidence, and ambiguity scores. Even though effective for prototype implementation, more sophisticated learnable routing mechanisms may further improve expert selection under highly complex traffic scenarios.

c) Third, the developed model has been evaluated under offline experimental conditions. Real-time deployment on large-scale intelligent transportation infrastructure may introduce additional challenges such as network latency, camera synchronization, illumination variation, adverse weather conditions, and hardware resource constraints.

d) Finally, the current implementation utilizes two specialized expert models. Also traffic monitoring systems may require additional experts for tasks such as pedestrian safety analysis, emergency vehicle detection, driver behaviour monitoring, road damage assessment, and anomaly detection.

6.2 Future Work

Future research will explore HAFN into a scalable multi-expert traffic intelligence platform.

The Adaptive Scene Routing Module will be enhanced using learnable routing policies capable of automatically selecting expert models through lightweight neural routing networks in place of predefined decision rules. This will enable more intelligent expert allocation in highly heterogeneous traffic environments.

The framework will also be expanded by incorporating additional expert models for pedestrian detection, emergency vehicle recognition, traffic congestion analysis, road anomaly detection, and driver behavior understanding. Furthermore, larger multi-city datasets collected under varying weather, lighting, and traffic conditions will be added to improve model robustness and generalization. Future work will also investigate deployment on edge computing platforms and smart city infrastructures to support real-time traffic monitoring with low latency. Integration with cloud-assisted traffic management systems and Internet of Things (IoT) devices automated traffic analytics, emergency response coordination, and intelligent transportation decision support.

7. CONCLUSION

This paper presented the Heterogeneous Adaptive Fusion Network (HAFN), a novel multi-expert traffic safety framework designed to address the limitations of conventional single-task traffic monitoring systems. While other methods independently handle accident detection and traffic violation analysis, HAFN integrates dual-dataset learning, specialized YOLO-based expert models, and an Adaptive Scene Routing Module to achieve unified traffic safety intelligence. The proposed method integrates a dedicated YOLOv8 Accident Expert trained on an accident-specific dataset and a YOLOv11 Violation Expert trained on a comprehensive multi-violation dataset. Using confidence-aware and ambiguity-aware routing, HAFN dynamically activates by expert model or simultaneously engages multiple experts for complex traffic scenes. The Fusion Layer subsequently combines expert outputs to generate comprehensive traffic safety information while reducing excessive computation.

Experimental evaluation demonstrates that the developed architecture achieves high detection performance across accident and traffic violation tasks, with strong Precision, Recall, and F1-score values and low inter-class confusion. The results validate the efficiency of dual-dataset specialization and adaptive expert routing in improving both detection accuracy and computational efficiency. Overall, HAFN provides a flexible and extensible foundation for next-generation intelligent transportation systems and opens new research directions toward adaptive multi-expert traffic intelligence frameworks capable of supporting real-world smart city and road safety applications.

8. REFERENCE

[1] World Health Organization, Global Status Report on Road Safety 2023. Geneva, Switzerland: WHO, Dec. 2023. [Online]. Available: https://www.who.int/publications/i/item/9789240086517

[2] World Health Organization, Global Status Report on Road Safety 2023: Summary. Geneva, Switzerland: WHO, Dec. 2023. [Online]. Available: https://www.who.int/publications/i/item/9789240086456

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

[3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, 2016, pp. 779–788. doi: 10.1109/CVPR.2016.91.
Paper: https://openaccess.thecvf.com/content_cvpr_2016/html/Redmon_You_Only_Look_CVPR_2016_paper.html

[4] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Honolulu, HI, USA, 2017, pp. 7263–7271. doi: 10.1109/CVPR.2017.690.
Paper: https://openaccess.thecvf.com/content_cvpr_2017/html/Redmon_YOLO9000_Better_Faster_CVPR_2017_paper.html

[5] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," arXiv preprint arXiv:1804.02767, 2018.
Paper: https://arxiv.org/abs/1804.02767

[6] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," arXiv preprint arXiv:2004.10934, 2020.
Paper: https://arxiv.org/abs/2004.10934

[7] G. Jocher et al., "YOLOv5," Ultralytics, GitHub Repository, 2020.
Project: https://github.com/ultralytics/yolov5

[8] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2023, pp. 7464–7475.
Paper: https://arxiv.org/abs/2207.02696

[9] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLOv8," Ultralytics Documentation, 2023.
Documentation: https://docs.ultralytics.com

[10] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information," arXiv preprint arXiv:2402.13616, 2024.
Paper: https://arxiv.org/abs/2402.13616

[11] A. Wang et al., "YOLOv10: Real-Time End-to-End Object Detection," arXiv preprint arXiv:2405.14458, 2024.
Paper: https://arxiv.org/abs/2405.14458

[12] Ultralytics, "YOLO11 Documentation," 2024.
Documentation: https://docs.ultralytics.com/models/yolo11/

[13] R. Girshick, "Fast R-CNN," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Santiago, Chile, 2015, pp. 1440–1448. doi: 10.1109/ICCV.2015.169.
Paper: https://openaccess.thecvf.com/content_iccv_2015/html/Girshick_Fast_R-CNN_ICCV_2015_paper.html

[14] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
Paper: https://arxiv.org/abs/1506.01497

[15] W. Liu et al., "SSD: Single Shot MultiBox Detector," in Proc. Eur. Conf. Comput. Vis. (ECCV), Amsterdam, The Netherlands, 2016, pp. 21–37.
Paper: https://arxiv.org/abs/1512.02325

[16] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Venice, Italy, 2017, pp. 2980–2988.
Paper: https://arxiv.org/abs/1708.02002

[17] N. Carion et al., "End-to-End Object Detection with Transformers," in Proc. Eur. Conf. Comput. Vis. (ECCV), Glasgow, U.K., 2020, pp. 213–229.
Paper: https://arxiv.org/abs/2005.12872

[18] N. Wojke, A. Bewley, and D. Paulus, "Simple Online and Realtime Tracking with a Deep Association Metric," in Proc. IEEE Int. Conf. Image Process. (ICIP), Beijing, China, 2017, pp. 3645–3649.
Paper: https://arxiv.org/abs/1703.07402

[19] Y. Zhang, P. Sun, Y. Jiang, D. Yu, F. Weng, Z. Yuan, P. Luo, W. Liu, and X. Wang, "ByteTrack: Multi-Object Tracking by Associating Every Detection Box," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2022.
Paper: https://arxiv.org/abs/2110.06864

[20] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, vol. 521, no. 7553, pp. 436–444, May 2015. doi: 10.1038/nature14539.
Paper: https://www.nature.com/articles/nature14539

[51] A. Razi, X. Chen, H. Li, H. Wang, B. Russo, Y. Chen, and H. Yu, "Deep Learning Serves Traffic Safety Analysis: A Forward-Looking Review," IET Intelligent Transport Systems, vol. 17, no. 1, pp. 3–29, 2023, doi: 10.1049/itr2.12257.
Paper: https://doi.org/10.1049/itr2.12257

[52] V. Adewopo, N. Elsayed, Z. ElSayed, M. Ozer, A. Abdelgawad, and M. Bayoumi, "Review on Action Recognition for Accident Detection in Smart City Transportation Systems," arXiv preprint, arXiv:2208.09588, 2022.
Paper: https://arxiv.org/abs/2208.09588

[53] M. M. Ahmed, M. N. Khan, A. Das, and S. E. Dadvar, "Global Lessons Learned From Naturalistic Driving Studies to Advance Traffic Safety and Operation Research: A Systematic Review," Accident Analysis & Prevention, vol. 167, Art. no. 106568, 2022, doi: 10.1016/j.aap.2022.106568.
Paper: https://pubmed.ncbi.nlm.nih.gov/35085856/

[54] T. Yuan, W. DaRocha, C. E. Rothenberg, C. Obraczka, C. Barakat, and L. Turletti, "Machine Learning for Next-Generation Intelligent Transportation Systems: A Survey," Transactions on Emerging Telecommunications Technologies, vol. 33, no. 4, e4427, 2022, doi: 10.1002/ett.4427.
Paper: https://onlinelibrary.wiley.com/doi/10.1002/ett.4427

[55] Z. Liu, X. Wang, Y. Wang, et al., "Autonomous Driving Under V2X Environment: State-of-the-Art Survey and Challenges," Intelligent Transportation Infrastructure, vol. 1, 2022, doi: 10.1093/iti/liac020.
Paper: https://academic.oup.com/iti/article/doi/10.1093/iti/liac020/6849553

[56] X. Cao, Z. Zhang, Y. Sun, et al., "The Review of Image Processing and Edge Computing for Intelligent Transportation System," Journal of Image and Graphics, vol. 27, no. 6, pp. 1743–1767, 2022, doi: 10.11834/jig.211266.
Paper: https://www.cjig.cn/en/article/doi/10.11834/jig.211266/

[57] M. Kamalian, P. Ferreira, and E. Jul, "A Survey on Local Transport Mode Detection on the Edge of the Network," Applied Intelligence, vol. 52, pp. 16021–16050, 2022, doi: 10.1007/s10489-022-03214-y.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

Paper: https://link.springer.com/article/10.1007/s10489-022-03214-y

[58] A. Al-Sadi, M. Abu-Amara, and M. A. Al-Mashaqbeh, “IoT-Based System for Improving Vehicular Safety by Continuous Traffic Violation Monitoring,” Future Internet, vol. 14, no. 11, Art. no. 319, 2022, doi: 10.3390/fi14110319. Paper: https://www.mdpi.com/1999-5903/14/11/319

[59] M. Zhao, D. Zhong, Y. Fu, J. Hou, and Z. Zhao, “A Review for the Driving Behavior Recognition Methods Based on Vehicle Multisensor Information,” Journal of Advanced Transportation, vol. 2022, Art. no. 7287511, 2022, doi: 10.1155/2022/7287511. Paper: https://onlinelibrary.wiley.com/doi/full/10.1155/2022/7287511

[60] M. A. Hossain, M. Atiquzzaman, and W. Ivancic, “Traffic Surveillance Using Deep Learning: A Survey,” IEEE Access, vol. 9, pp. 114921–114950, 2021.

[61] P. Lee, H. Kim, M. S. Zitouni, A. Khandoker, H. F. Jelinek, L. Hadjileontiadis, U. Lee, and Y. Jeong, “Trends in Smart Helmets With Multimodal Sensing for Health and Safety: Scoping Review,” JMIR mHealth and uHealth, vol. 10, no. 11, e40797, 2022, doi: 10.2196/40797. Paper: https://www.sciencedirect.com/org/science/article/pii/S2291522222000511

[62] M. A. Hossain, M. Atiquzzaman, and W. Ivancic, “Computer Vision-Based Intelligent Traffic Monitoring: Recent Advances and Future Challenges,” IEEE Access, vol. 10, pp. 65001–65032, 2022.

[63] M. A. Hossain and W. Ivancic, “Deep Learning-Based Intelligent Transportation Systems: Challenges and Opportunities,” IEEE Communications Surveys & Tutorials, 2022.

[64] S. Janai, F. Güney, J. Behl, and A. Geiger, “Computer Vision for Autonomous Vehicles: Problems, Datasets and State of the Art,” Foundations and Trends® in Computer Graphics and Vision, vol. 12, no. 1–3, pp. 1–308, 2020.

[65] D. Bogdoll, M. Nitsche, and J. M. Zöllner, “Anomaly Detection in Autonomous Driving: A Survey,” arXiv preprint, arXiv:2204.07974, 2022. Paper: https://arxiv.org/abs/2204.07974

[66] Z. Bai, G. Wu, X. Qi, Y. Liu, K. Oguchi, and M. J. Barth, “Infrastructure-Based Object Detection and Tracking for Cooperative Driving Automation: A Survey,” arXiv preprint, arXiv:2201.11871, 2022. Paper: https://arxiv.org/abs/2201.11871

[67] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” Advances in Neural Information Processing Systems (NeurIPS), vol. 25, 2012. Paper: https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks

[68] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” Proc. International Conference on Learning Representations (ICLR), 2015. Paper: https://arxiv.org/abs/1409.1556

[69] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016. Book: https://www.deeplearningbook.org/

[70] R. Szeliski, Computer Vision: Algorithms and Applications, 2nd ed. Cham, Switzerland: Springer, 2022. Book: https://szeliski.org/Book/