vehicles

MDPI

Article

DSGF-YOLO: A Lightweight Deep Neural Network for Traffic Accident Detection and Severity Classifications

Weijun Li, Huawei Xie and Peiteng Lin *

Department of Forensic Science, Fujian Police College, Fuzhou 350007, China; xjlwj@fjpsc.edu.cn (W.L.); xiehuawei@fjpsc.edu.cn (H.X.)
* Correspondence: linpeiteng@outlook.com

Abstract

Traffic accidents pose unpredictable and severe social and economic challenges. Rapid and accurate accident detection, along with reliable severity classification, is essential for timely emergency response and improved road safety. This study proposes DSGF-YOLO, an enhanced deep learning framework based on the YOLOv13 architecture, developed for automated road accident detection and severity classification. The proposed methodology integrates two novel components: the DS-C3K2-FasterNet-Block module, which enhances local feature extraction and computational efficiency, and the Grouped Channel-Wise Self-Attention (G-CSA) module, which strengthens global context modeling and small-object perception. Comprehensive experiments on a diverse traffic accident dataset validate the effectiveness of the proposed framework. The results show that DSGF-YOLO achieves higher precision, recall, and mean average precision than state-of-the-art models such as Faster R-CNN, DETR, and other YOLO variants, while maintaining real-time performance. These findings highlight its potential for intelligent transportation systems and real-world accident monitoring applications.

Keywords: road accident detection; severity classification; global context modeling; feature extraction

check for updates

Academic Editors: Dinh-Thuan Do, Vitor Fialho, Luis Pires, Francisco Rego, Ricardo Santos and Vasco Velez

Received: 25 September 2025
Revised: 2 December 2025
Accepted: 3 December 2025
Published: 5 December 2025

Citation: Li, W.; Xie, H.; Lin, P. DSGF-YOLO: A Lightweight Deep Neural Network for Traffic Accident Detection and Severity Classifications. *Vehicles* **2025**, *7*, 153. https://doi.org/10.3390/vehicles7040153

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Traffic accident detection and severity classification are critical components of modern traffic management and public safety systems. The rapid expansion of road networks and increasing vehicle density have heightened the need for not only accurate accident detection but also timely assessment of accident severity, which is essential for prioritizing emergency response and resource allocation [1]. Traditional traffic monitoring methods, relying on manual observation, are limited by human error, low scalability, and delayed response times, making both detection and severity assessment inefficient and unreliable [2]. These limitations have driven the development of automated systems within intelligent transportation systems (ITSs) to provide comprehensive, real-time accident monitoring.

Early automated approaches primarily employed traditional machine learning techniques, which required extensive image preprocessing and manual feature extraction, including shape, texture, and color descriptors, followed by classifier training to identify accidents [3]. While effective under controlled conditions, these methods are highly sensitive to environmental variations, lighting changes, occlusions, and overlapping objects [4,5]. Moreover, most of these methods focused solely on accident detection and lacked the

capability to automatically categorize the severity of the incidents, which restricts their utility for real-time emergency prioritization.

The advent of deep learning has provided a transformative solution to these limitations [6]. Deep learning models can process raw image data in an end-to-end manner, automatically learning hierarchical feature representations, eliminating the need for manual feature engineering, and reducing error propagation inherent in multi-stage pipelines. Importantly, these models can be extended to simultaneously detect accidents and classify their severity, enabling more intelligent decision-making in real-world traffic management [7,8]. Deep learning models also demonstrate strong scalability and achieve superior performance when trained on large-scale datasets, allowing for deployment across diverse and complex traffic scenarios [9].

Although promising, many existing deep learning-based traffic accident detection systems still have notable limitations. Most models are developed and evaluated under simplified or idealized conditions, often overlooking real-world complexities such as varying illumination, weather changes, occlusions, overlapping objects, and diverse traffic scenes [10,11]. Their generalization ability is also constrained by limited dataset diversity and scale. Moreover, few studies integrate severity classification, reducing their practical value for timely emergency response [12]. These limitations highlight the need for unified and efficient models capable of achieving both accurate accident detection and reliable severity assessment in complex and dynamic traffic environments.

To address these limitations, we propose DSGF-YOLO, an enhanced model built upon the YOLOv13 framework, designed for real-time traffic accident detection and severity classification. The main contributions of this work are as follows:

Comprehensive Dataset Construction: A diverse dataset was developed to include multiple accident types, varied environmental conditions, and complex traffic scenarios. This facilitates robust feature learning and enhances model generalization, ensuring reliable performance in real-world applications.

Enhanced Backbone with DS-C3K2-G-CSA Modules: The backbone architecture substitutes the original DSC3K2 modules with DS-C3K2-G-CSA, enabling a unified enhancement of local feature extraction and global context modeling. This design allows the network to capture hierarchical and context-aware representations, potentially improving the detection of small, overlapping, and structurally complex accident objects.

Optimized Detection Head with DS-C3K2-FasterNet: To further strengthen the detection capability, the head network incorporates DS-C3K2-FasterNet, enabling a more efficient balance between lightweight computation and expressive feature learning. This integration enhances the robustness of the detection process, leading to improved accuracy and generalization.

This integration enhances the overall robustness and stability of the detection process, resulting in higher localization precision, improved detection accuracy, and better generalization across diverse scenarios. Moreover, the proposed design maintains high computational efficiency, making it well-suited for real-time deployment in intelligent transportation systems.

2. Related Works

Recent research has increasingly emphasized the application of advanced machine learning and deep learning techniques to address the challenges of traffic accident detection and severity prediction. A comprehensive review by Somawanshi and Shah [13] surveys the algorithms employed for severity prediction, underscoring the effectiveness of convolutional neural networks (CNNs) and other deep learning architectures. To tackle the critical issue of data imbalance in real-world applications, Zhang et al. [14] demonstrate the

effectiveness of integrating deep learning with clustering techniques, thereby improving predictive accuracy.

Further efforts have focused on the development of novel deep learning models for both accident detection and severity classification. For instance, Mohana Deepthi et al. [15] introduce the RFCNN model, which employs decision-level fusion of machine learning and deep learning approaches to enhance severity classification accuracy. In parallel, Zhu et al. [16] propose an accident detection framework based on comprehensive traffic flow feature extraction, highlighting the pivotal role of robust feature engineering in achieving reliable detection performance.

Expanding on these efforts, Lian et al. [17] developed an attention feature fusion network to enhance small-object detection performance in complex traffic scenes, effectively improving recognition accuracy under occlusion and cluttered backgrounds. Li et al. [18] proposed a spatio-temporal deep learning framework based on a ConvLSTM–Transformer architecture for traffic accident severity prediction, enabling more accurate modeling of temporal dependencies and spatial correlations in accident data. Wang et al. [19] explored hybrid CNN–Transformer feature representations to strengthen contextual feature extraction, which can be extended to multi-vehicle accident detection tasks requiring both local and global feature understanding. Furthermore, Liu et al. [20] introduced a multimodal deep learning approach that integrates crash records, road geometry, and textual descriptions to enhance the robustness and interpretability of accident severity prediction.

Collectively, these studies reflect a growing trend toward leveraging sophisticated machine learning and deep learning [21,22]. The integration of such technologies demonstrates significant potential for improving road safety and optimizing emergency response systems. Nonetheless, this evolving field continues to demand further exploration of advanced algorithmic strategies and resilient data analysis techniques to effectively address the complexities inherent in real-world traffic environments [23,24].

3. Materials and Methods

3.1. Dataset Construction

A comprehensive and diverse image dataset was constructed by aggregating data from multiple publicly available online sources to support the training and evaluation of the proposed DSGF-YOLO model for traffic accident detection and severity classification. The dataset was carefully curated to capture the heterogeneity of real-world traffic scenarios, thereby ensuring robust and reliable model performance across diverse environments. In total, 2265 images were collected, exhibiting substantial variation in resolution, image quality, object scale, and illumination. To approximate practical operational environments, the dataset integrates varied and complex conditions that impose additional challenges on visual recognition, covering both urban and highway accident scenes, with urban environments being more prominently represented due to their higher traffic density and visual complexity. As the primary focus of this study is vehicle-related traffic accidents, the dataset predominantly covers common road vehicles such as cars, buses, and trucks, which represent the majority of real-world accident scenarios. Other traffic participants such as motorcycles, bicycles, and pedestrians are not included in the current dataset, thereby defining the scope of the present work while leaving opportunities for future extensions.

Each image was meticulously annotated to indicate the presence of an accident and was further assigned a severity level—moderate or severe—determined by comprehensive indicators such as the degree of vehicle damage, the intensity of collision impact, and surrounding contextual scene information. To better capture the variability inherent in real-world traffic environments, additional contextual attributes were also documented, thereby enhancing the diversity and representativeness of the dataset. It is worth noting,

however, that these supplementary attributes were not directly utilized during the model training process but remain valuable for potential extensions and broader analyses.

For model development and evaluation, the dataset was partitioned into training (n = 1892), validation (n = 262), and testing (n = 111) subsets. The split was designed to ensure balanced representation across scene types and severity levels, thereby optimizing parameter estimation and enabling reliable performance assessment. To further enhance robustness and mitigate overfitting, a three-stage data augmentation pipeline was applied. Specifically, random 90-degree rotations were introduced to simulate variations in camera orientation, Gaussian blurring was employed to approximate motion-induced degradation, and stochastic noise injection was used to emulate artifacts caused by low-quality sensors or unstable transmission channels. Through this strategy, the training dataset was expanded to encompass a richer variety of visual patterns and perturbations—such as noise, occlusion, and distortion—thereby providing a more comprehensive representation of real-world traffic accident scenes. Through this augmentation strategy, the diversity of the training data was significantly enriched, enabling DSGF-YOLO to achieve stronger feature generalization and greater resilience to noise, occlusion, and distortion—attributes that are essential for accurate and dependable traffic accident detection and severity classification.

3.2. DSGF-YOLO

Real-world traffic environments pose significant challenges for object detection owing to high vehicle density, frequent occlusions, small object sizes, and diverse environmental conditions, including variable illumination and adverse weather. Accurate accident detection coupled with reliable severity classification therefore requires a model that combines robust feature extraction, effective multiscale reasoning, and efficient inference to support real-time decision-making in intelligent transportation systems (ITS). To address these challenges, we propose DSGF-YOLO, an enhanced variant of YOLOv13 specifically tailored for traffic accident detection and severity classification [25,26]. Building upon the canonical three-stage YOLO structure—backbone, neck, and head—the framework introduces novel modules that enhance contextual reasoning, refine multiscale feature representation, and ensure computational efficiency. An overview of the architecture is presented in Figure 1.

In the backbone, standard depthwise separable convolutions and bottleneck blocks are replaced with DS-C3k2-G-CSA and A2C2f modules. The DS-C3k2-G-CSA module enables global context-aware feature extraction, thereby enhancing the detection of small, overlapping, and partially occluded vehicles that are prevalent in dense accident scenes. The lightweight A2C2f blocks further facilitate hierarchical multiscale learning, reinforcing the integration of spatial and semantic features while preserving computational efficiency. The proposed improvements enable the backbone to model fine-grained accident-related features, thereby supporting the representation of traffic scenarios spanning moderate to severe levels of severity.

The neck employs DS-C3k2-FasterNet in combination with FullPAD Tunnel and HyperACE modules to achieve enhanced multiscale feature fusion across P3, P4, and P5. By systematically integrating upsampling, downsampling, and lateral skip connections, the neck combines low-level spatial detail with high-level semantic information. The design provides mechanisms that facilitate more precise localization and severity-aware detection when operating under adverse conditions, including low illumination, rain, and fog.

The head retains the standard YOLO detection module, configured to generate predictions at three scales (P3, P4, and P5). This multi-scale detection strategy ensures reliable identification of both large-scale collisions and subtle accident details. For each bounding box, severity classification is performed, yielding outputs that may support traffic management authorities and emergency responders in prioritizing interventions.

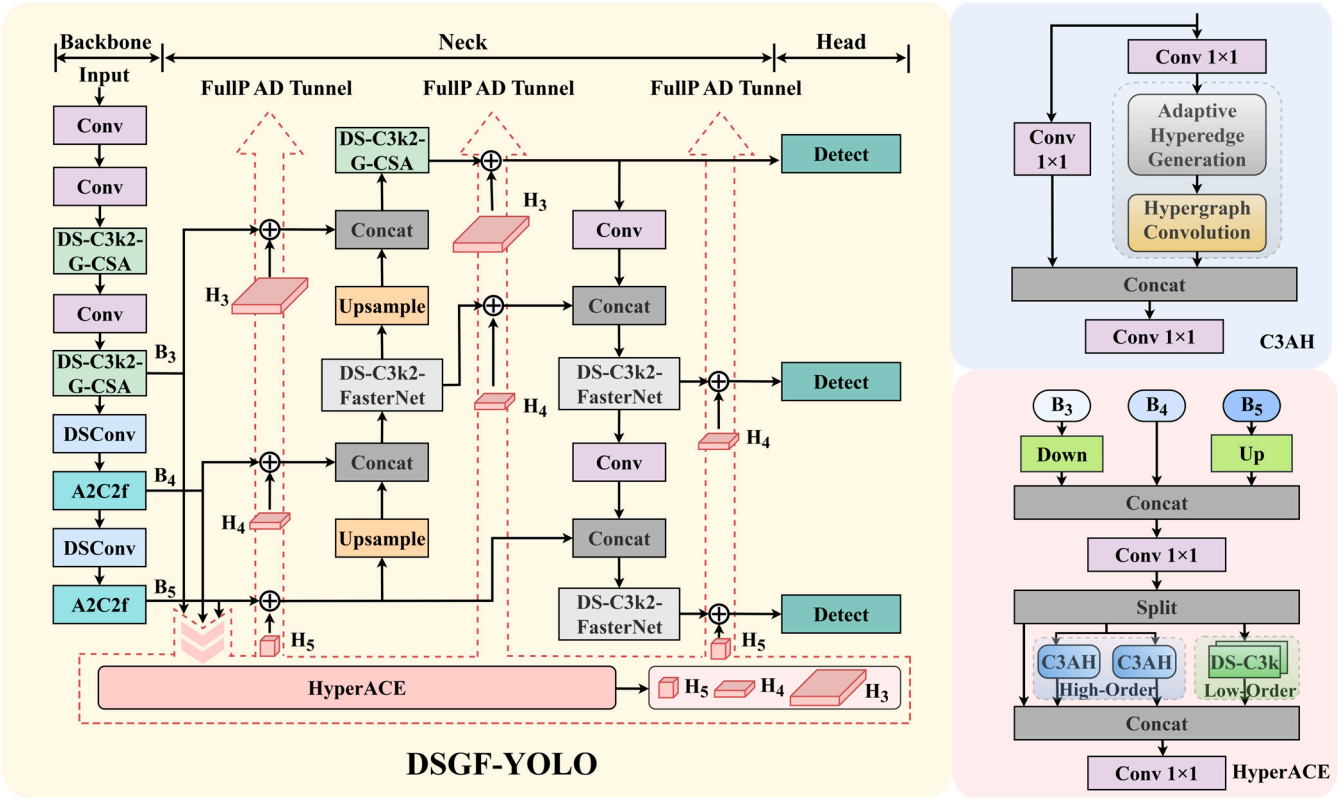


Figure 1. Overall architecture of the proposed DSGF-YOLO model, which consists of a Backbone, Neck, and Head. The Backbone extracts multi-scale features, the Neck fuses contextual information, and the Head performs detection and severity classification. The right panels illustrate key modules that enhance feature fusion and context modeling.

Overall, DSGF-YOLO achieves a favorable trade-off between accuracy and efficiency. By integrating enhanced backbone and neck modules with a robust multi-scale detection head, the model delivers precise localization, reliable severity classification, and strong generalization across heterogeneous traffic environments. These characteristics make DSGF-YOLO a compelling candidate for real-time ITS applications, where timely and accurate accident detection is critical to reducing emergency response times and mitigating secondary risks.

3.2.1. Improvements with the G-CSA

The G-CSA module is a specialized architectural enhancement designed to improve feature representation in lightweight object detection frameworks derived from the YOLO family [27,28]. The design is informed by the progressive development of convolutional modules within modern object detection architectures, most notably the C3 structure first introduced in YOLOv5 and subsequently refined. While traditional C3 modules rely on multiple bottleneck blocks to balance accuracy and efficiency, DS-C3K2 extended this structure by introducing depthwise separable convolution (DSConv) and kernel grouping (K2). As illustrated in Figure 2, the proposed G-CSA further integrates a Grouped Channel–Spatial Attention (G-CSA) mechanism into DS-C3K2, yielding a more expressive and context-aware feature extraction unit.

DSConv reduces computational cost by decomposing standard convolution into depthwise and pointwise operations. This design lowers FLOPs while preserving representational power, which makes the module well-suited for real-time applications. The kernel grouping strategy (K2) further partitions convolutional kernels into multiple groups to capture diverse receptive field patterns with fewer parameters, thereby allowing the network to

effectively model multi-scale object features without the need for large kernels. In addition, the G-CSA mechanism adaptively recalibrates features along both channel and spatial dimensions. By emphasizing salient regions and suppressing redundant background information, G-CSA enhances discriminative learning in complex traffic scenes involving small, overlapping, or occluded objects.

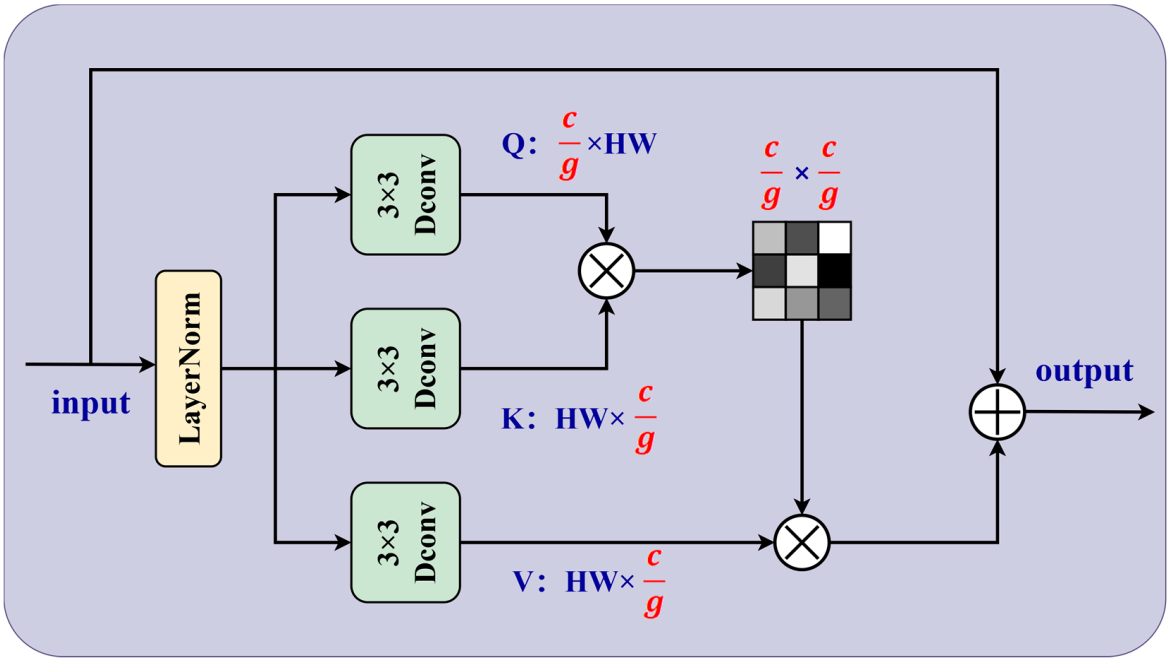


Figure 2. Architecture of the Global Channel-Spatial Attention (G-CSA) module. It enhances feature representation by capturing long-range dependencies through joint channel and spatial attention mechanisms.

The integration of DS-C3K2-G-CSA into the backbone of YOLO-style detectors provides a mechanism for strengthening feature representation and improving efficiency within the detection pipeline. The synergy of DSConv and kernel grouping provides a lightweight yet powerful feature extractor, while the incorporation of G-CSA enhances robustness against challenging visual conditions. Empirical evaluations confirm that detectors equipped with DS-C3K2 and G-CSA attain superior mean average precision (mAP) with reduced latency compared to baseline C3-based designs, validating their suitability for real-time applications such as intelligent transportation systems, autonomous driving, UAV vision, and edge AI deployments.

3.2.2. Improvements with the FasterNet-Block

The original YOLOv13 backbone employs DS-C3K2 modules based on depthwise separable convolutions to reduce computational complexity. While these operators effectively decrease the theoretical number of floating-point operations (FLOPs), their practical efficiency is often compromised by excessive memory access and limited parallel utilization [29]. Therefore, despite their low FLOPs, DS-C3K2-based networks can exhibit a gap between theoretical efficiency and the computational constraints of deployment environments such as edge devices and other resource-constrained platforms.

To address this limitation, we introduce the FasterNet-Block, specifically designed to bridge the gap between theoretical efficiency and real-world throughput, as illustrated in Figure 3. At its core lies the Partial Convolution (PConv) operator, which applies convolution to only a subset of channels (Cp), while propagating the remaining channels through identity mapping. This selective computation effectively reduces redundant operations

and memory traffic, thereby lowering FLOPs while simultaneously improving effective FLOPS [30]. Compared with depthwise and grouped convolutions, PConv achieves more efficient hardware utilization and yields measurable latency reductions in practice.

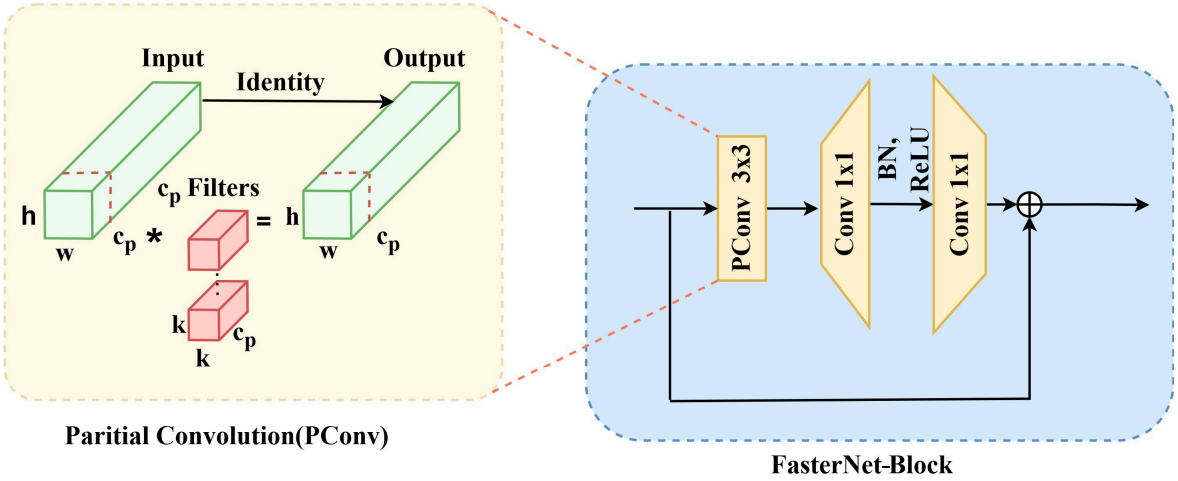


Figure 3. Structure of the FasterNet-Block using Partial Convolution (PConv) for efficient feature extraction. This design reduces computational cost while maintaining strong feature representation capability.

Structurally, the FasterNet-Block adopts an inverted residual design. An initial PConv layer captures informative channel subsets, followed by two pointwise convolutions (PWConv, 1×1). The first PWConv expands the channel dimension and incorporates batch normalization and non-linear activation to enhance representational capacity, while the second projects features back to the original dimensionality. A residual shortcut connects the input and output, facilitating stable gradient propagation and effective feature reuse. This architectural refinement not only improves training stability but also accelerates inference, making the FasterNet-Block a hardware-friendly alternative to traditional DS-C3K2 modules.

By replacing the DS-C3K2 modules with the proposed FasterNet-Block, YOLOv13 attains a more hardware-friendly backbone architecture. Empirical results demonstrate that this design consistently achieves higher throughput and lower latency across a wide range of platforms, including GPUs, CPUs, and ARM processors, while maintaining—or even surpassing—the detection accuracy of depthwise separable convolution–based counterparts. Consequently, the integration of the FasterNet-Block represents a principled step toward bridging the gap between FLOPs and real-time inference, thereby enabling YOLOv13 to scale more effectively and reliably across heterogeneous deployment environments.

3.2.3. Evaluation Indicators

A comprehensive evaluation of the proposed DSGF-YOLO model for traffic accident detection and severity classification requires not only conventional object detection metrics but also considerations of robustness and real-time performance. Since traffic accidents are safety-critical events, the evaluation framework must capture the ability of the model to achieve high accuracy, minimize false alarms, avoid missed detections, and respond in real time to ensure practical applicability in intelligent transportation systems [31].

Precision (P) and recall (R) are two fundamental indicators. Precision reflects the proportion of true accident detections among all predicted accidents, emphasizing the system's ability to reduce false alarms:

$$P = \frac{TP}{TP + FP}$$

where TP denotes true positives and FP false positives. High precision ensures that normal driving scenes are not misclassified as accidents, which is crucial for maintaining public trust and operational efficiency. Recall measures the proportion of actual accidents successfully detected by the model:

$$R = \frac{TP}{TP + FN}$$

where FN represents false negatives. In accident detection, recall is equally important, as a missed severe accident could delay rescue and exacerbate consequences.

To unify these complementary measures, average precision (AP) is computed as the area under the precision–recall curve:

$$AP = \int_0^1 P\,(r)\,dr$$

where P(r) is precision as a function of recall. For multiple accident categories (mild, moderate, severe), mean average precision (mAP) is adopted:

$$mAP = \frac{1}{n}\sum_{n=1}^{n} AP_n$$

with n as the number of severity levels. These measures collectively provide a rigorous quantification of detection and classification accuracy.

Fβ-Score is a generalized form of the F1-Score that provides a weighted harmonic mean of Precision and Recall. It is defined as follows:

$$F_\beta = \left(1 + \beta^2\right) \cdot \frac{P \cdot R}{\beta^2 \cdot P + R}$$

β is a non-negative parameter that controls the trade-off between Precision and Recall.

Specifically, when β = 1, the measure reduces to the conventional F1-Score, which equally weights Precision and Recall. For β > 1, the metric places greater emphasis on Recall, making it particularly suitable for tasks where false negatives are more costly than false positives, such as accident detection, medical diagnosis, or safety-critical applications. Conversely, β < 1 emphasizes Precision, which is preferable in scenarios where false alarms must be minimized.

In this study, the F2-Score (β = 2) is adopted to highlight Recall, ensuring that the evaluation metric penalizes missed detections more heavily than false alarms. This choice is aligned with the objective of traffic accident detection, where failure to identify a true accident is considered more critical than mistakenly raising an alert [32].

In summary, these indicators provide a multidimensional evaluation framework that assesses not only the accuracy of DSGF-YOLO but also its robustness and practical feasibility for deployment in real-world accident detection and severity classification scenarios.

3.3. Accident Alert Transmission Mechanism

To support practical deployment within intelligent transportation systems (ITS), the proposed framework integrates an accident alert transmission mechanism designed to deliver accident information to emergency and rescue units in real time. Once the DSGF-YOLO model detects an accident and determines its severity level, the corresponding event metadata—including location, timestamp, and severity classification—is automatically

transmitted through existing wired monitoring infrastructures or via 5G-enabled wireless communication networks.

Leveraging the high bandwidth, low latency, and reliable connectivity offered by 5G communication, the system ensures that accident alerts are delivered promptly to traffic management centers and emergency responders, thereby enabling rapid intervention and reducing the risk of secondary incidents. Although the present study primarily focuses on model development and performance evaluation, this alert transmission component forms a critical part of our broader system integration efforts and will be fully implemented in subsequent deployment stages.

4. Results

4.1. Experimental Environment and Parameter Settings

To ensure fairness and reproducibility in both training and evaluation, a comprehensive traffic accident image dataset was constructed and randomly partitioned into training, validation, and testing subsets in an approximate ratio of 8:1:1. Specifically, the dataset consisted of 1892 training images, 262 validation images, and 111 testing images. The training set encompassed diverse real-world accident scenarios, ranging from normal traffic flow to minor collisions and severe crashes, thereby enhancing the model's generalization capability across varying levels of accident severity.

All experiments were conducted in a high-performance cloud environment equipped with an NVIDIA A100 GPU (80 GB memory). The detailed system configuration is summarized in Table 1, while the selected hyperparameter settings are listed in Table 2.

Table 1. Experimental environment configuration.

Environment Configuration	Value
Operating system	Ubuntu 20.04 (Cloud Environment)
CPU	Intel Xeon @ 2.20 GHz
GPU	NVIDIA A100 (NVIDIA A100-SXM4)
RAM	80 GB
Development environment	Cloud-based Jupyter
Programming language	Python 3.12

Table 2. Hyperparameter.

Hyperparameter	Value
Epochs	500
Batch size	64
Initial learning rate	0.01
Optimizer	SGD
Input image size	640 × 640

To evaluate detection performance under complex real-world traffic conditions, the mean Average Precision (mAP) at an IoU threshold of 0.5 was employed as the primary metric. This measure effectively captures both detection accuracy and robustness across varying levels of accident severity. The final hyperparameter configuration was selected to achieve an optimal trade-off between convergence speed, detection accuracy, and real-time inference efficiency, thereby ensuring the model's practical applicability in traffic accident detection and severity classification tasks.

4.2. Comparison with Mainstream Methods

To establish a reliable baseline architecture for subsequent improvements, a systematic comparison was conducted across representative object detection models, including Faster R-CNN, DETR, YOLOv5n, YOLOv8n, YOLO11n, and YOLOv13n, as summarized in Table 3.

Table 3. Accuracy and Efficiency Trade-off Across Detection Models.

Algorithm	Batch Size	Precision/%	Recall/%	mAP50/%	mAP50-95/%	GFLOPS
DETR	64	91.0	83.0	92.8	82.6	36.8
Faster R-CNN	64	90.2	83.0	92.2	80.6	37.5
YOLOv5n	64	95.1	82.7	90.2	79.2	7.2
YOLOv8n	64	95.9	85.0	90.5	80.7	8.2
YOLO11n	64	95.5	84.0	90.9	82.0	6.4
YOLOv13n	64	92.8	86.0	92.6	84.6	6.4
DSGF-YOLO	64	92.7	90.8	94.8	86.8	6.6

Two-stage detectors, including Faster R-CNN and DETR, demonstrated high detection performance, achieving precision above 90% and mAP50 greater than 92%. Nevertheless, their computational requirements exceeded 36 GFLOPs, thereby limiting their suitability for real-time traffic accident detection and severity classification, where efficiency is equally important as accuracy.

In contrast, lightweight one-stage YOLO variants exhibited substantially lower computational costs while maintaining competitive accuracy. YOLOv5n and YOLOv8n achieved high precision, reaching 95.1% and 95.9%, respectively, yet their overall performance was constrained by lower mAP50-95 values of 79.2% and 80.7%. YOLO11n showed moderate improvements, attaining a recall of 84.0% and an mAP50-95 of 82.0%, but it still did not achieve an optimal balance across precision, recall, and multi-scale detection accuracy.

Within the category of lightweight detectors, YOLOv13n provided the most favorable balance between performance and efficiency. The model attained 92.8% precision, 86.0% recall, and an mAP50-95 of 84.6%, while requiring 6.4 GFLOPs. Taken together, these attributes make YOLOv13n a suitable reference baseline for accident detection and severity classification.

Most importantly, the proposed DSGF-YOLO achieved consistent improvements over all baseline models. With a precision of 92.7%, it substantially increased recall to 90.8%, thereby reducing missed detections in critical accident scenarios. In addition, the model attained the highest mAP50 of 94.8% and an mAP50-95 of 86.8%, while requiring only 6.6 GFLOPs. These outcomes indicate that the integration of depthwise separable convolution, spatial–channel attention, and gradient fusion strategies enhances detection accuracy and reinforces robustness under diverse real-world conditions. The improved recall is especially relevant for traffic accident detection, where reducing false negatives contributes to safer and faster emergency response.

To evaluate the optimization dynamics and stability of the proposed DSGF-YOLO model, Figure 4 illustrates the training and validation curves over 500 epochs. The results demonstrate consistent convergence across all loss components. In particular, the box loss, classification loss, and distribution focal loss steadily decrease throughout training, with both training and validation losses showing strong consistency, suggesting the absence of overfitting. At the same time, the performance metrics such as precision, recall, mAP50, and mAP50-95 improve continuously and stabilize at high values in the later epochs. Notably, recall and mAP50-95 approach saturation, indicating that the model generalizes reliably across diverse accident scenarios.

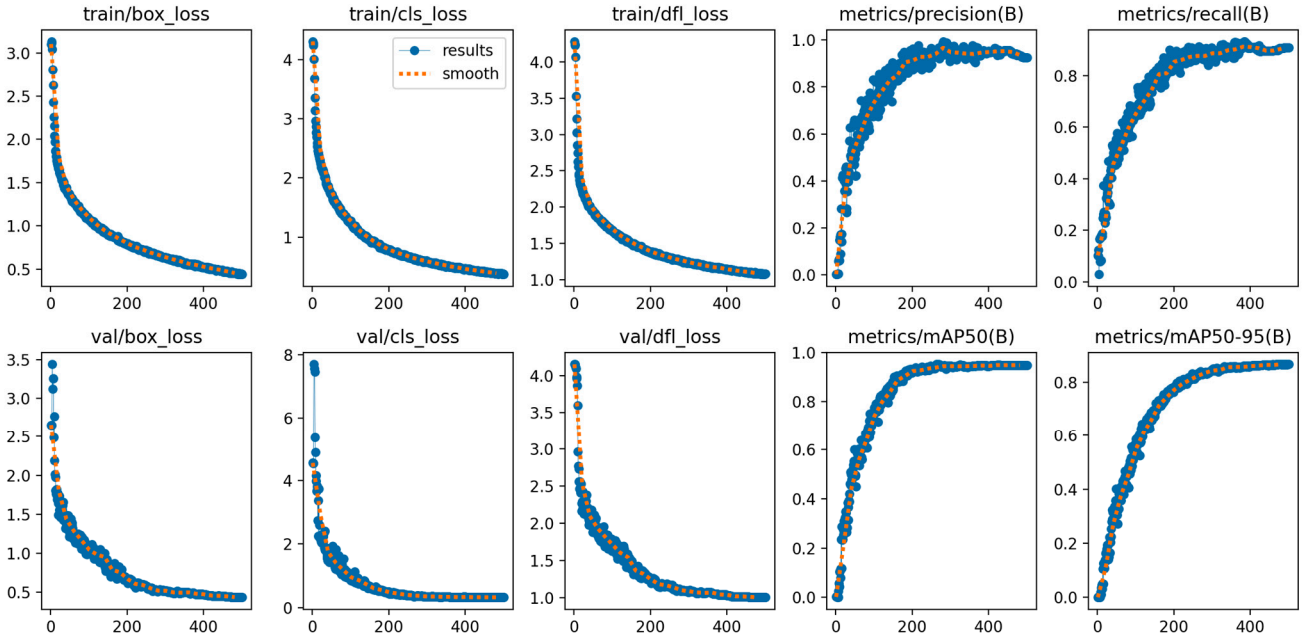


Figure 4. Training and validation convergence curves of the proposed DSGF-YOLO model.

These convergence patterns further support the quantitative comparisons reported in Table 3 and the qualitative detection results presented in Figure 4. Collectively, they provide evidence that DSGF-YOLO surpasses baseline detectors in accuracy and robustness, while also maintaining efficient and stable training dynamics. This combination underscores its potential for deployment in intelligent transportation systems.

Building upon the convergence analysis in Figure 4, additional qualitative experiments were conducted to assess the practical effectiveness of DSGF-YOLO in real-world traffic accident detection and severity classification. Although the convergence curves indicate stable optimization and reliable generalization during training, it is equally important to examine whether these quantitative improvements are reflected in robust performance under diverse operational conditions. To this end, visual inspection of detection outcomes was carried out on challenging accident scenarios, enabling a more comprehensive evaluation of the model's strengths and limitations in environments that closely approximate practical deployment.

Figure 5 provides representative examples of the detection and severity classification results produced by the proposed DSGF-YOLO model. As illustrated, the model accurately localizes accident regions and assigns appropriate severity labels (moderate or severe) with consistently high confidence scores. More importantly, DSGF-YOLO demonstrates strong robustness across heterogeneous and complex traffic conditions, including nighttime low illumination, multi-vehicle collisions on highways, single-vehicle crashes in urban environments, overturned vehicles, congestion-induced incidents, and severe head-on collisions. These qualitative results highlight the model's ability to capture subtle visual cues such as vehicle deformation, collision angles, and contextual scene information, which are critical for distinguishing between varying levels of accident severity.

Taken together, the results in Figure 5 demonstrate that DSGF-YOLO not only surpasses existing detectors in quantitative metrics but also provides reliable and interpretable predictions in real-world applications. By integrating enhanced backbone and neck modules, the model simultaneously leverages global contextual reasoning and fine-grained feature extraction, ensuring robust detection across diverse environmental and traffic conditions. This capability is particularly valuable for real-time accident monitoring, where accurate accident identification and severity classification are critical for minimizing re-

sponse delays and improving traffic safety. To further explore the interpretability and internal decision-making process of the model, Figure 6 presents a visualization comparison between mainstream detectors and DSGF-YOLO using Gradient-weighted Class Activation Mapping (Grad-CAM), a technique that highlights the image regions most influential to the model's predictions.



Figure 5. Detection and severity classification results of the proposed model under various challenging scenarios. The model accurately detects accidents and classifies their severity levels across different lighting, weather, and occlusion conditions.

As illustrated in Figure 6, DETR, Faster R-CNN, and lightweight YOLO variants including YOLOv5n, YOLOv8n, YOLO11n, and YOLOv13n often display scattered or diluted activation regions. In particular, these models tend to focus excessively on irrelevant background elements such as vegetation, road boundaries, or non-damaged parts of vehicles, while neglecting the critical regions that reflect accident severity. Such dispersed attention undermines the reliability of severity classification, especially under complex conditions involving occlusion, cluttered backgrounds, or multi-vehicle collisions.

In contrast, DSGF-YOLO produces activation patterns that are both highly concentrated and semantically meaningful. The Grad-CAM visualizations demonstrate that its attention is closely aligned with accident-relevant regions, such as vehicle deformation, collision boundaries, and affected road infrastructure. This alignment suggests that the integration of the G-CSA and FasterNet-block modules contributes not only to improved detection performance but also to enhanced interpretability of the decision process. By prioritizing critical accident cues, DSGF-YOLO provides more reliable severity classification, supporting its potential for deployment in safety-critical intelligent transportation systems.

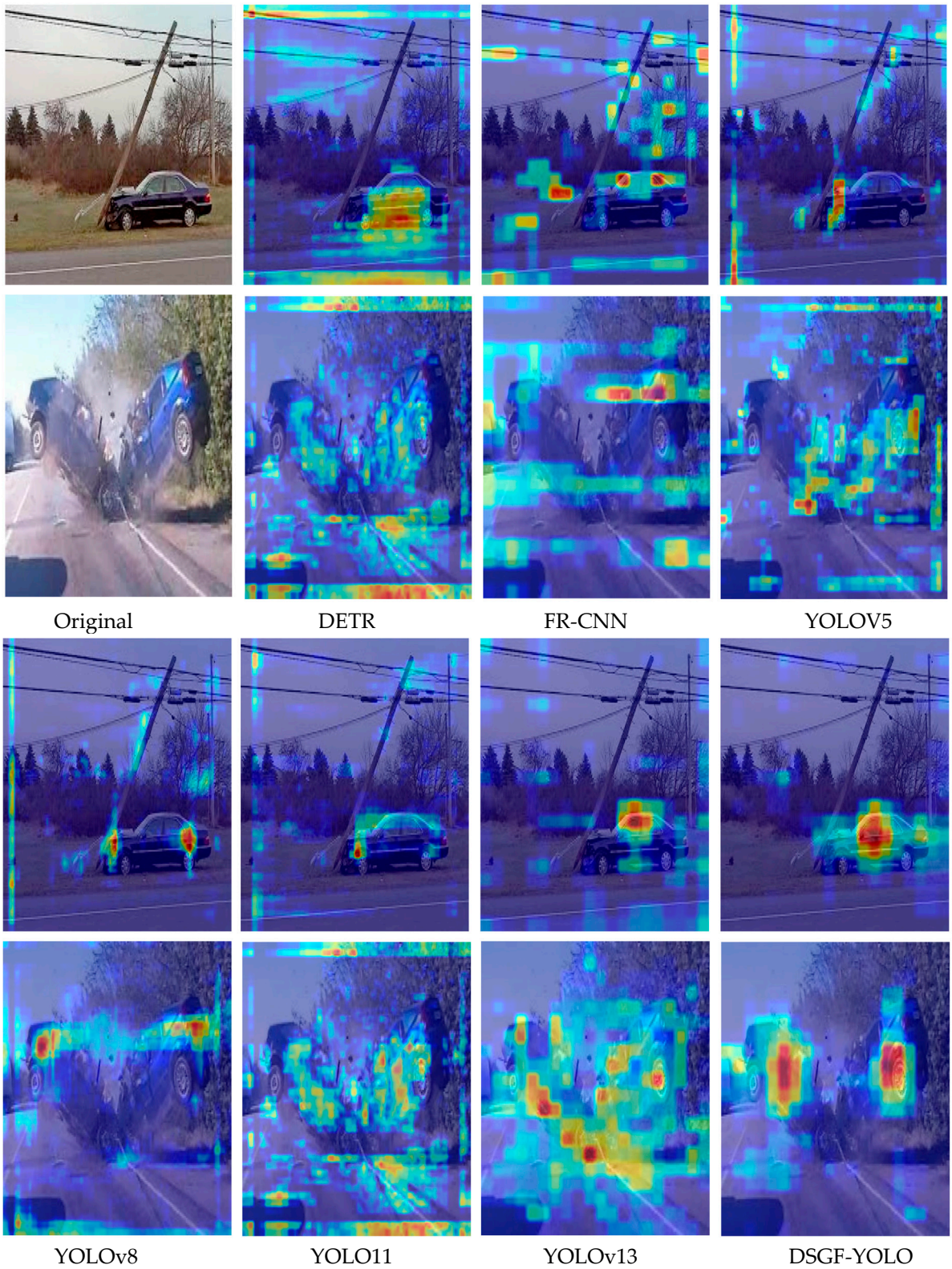


Figure 6. Grad-CAM visualization comparison across mainstream detectors and the proposed model. DSGF-YOLO demonstrates stronger attention on critical accident regions, indicating better feature focus and interpretability.

4.3. Ablation Experiment

To investigate the effectiveness of the proposed architectural components, a series of ablation experiments were performed on the baseline YOLOv13n model by selectively en-

abling the Global Channel–Spatial Attention (G-CSA) module and the FasterBlock module. The quantitative results are summarized in Table 4.

Table 4. Ablation results of modules in DSGF-YOLO performance.

G-CSA	FasterBlock	Precision/%	Recall/%	mAP50/%	mAP50-95/%	F2-Score/%
-	-	92.8	86.0	92.6	84.6	89.3
√	-	91.6	88.5	93.7	86.1	90.0
-	√	96.7	85.7	92.5	85.2	90.9
√	√	92.7	90.8	94.8	86.8	91.7

When neither module was applied, the baseline model achieved a precision of 92.8%, a recall of 86.0%, an mAP50 of 92.6%, an mAP50-95 of 84.6%, and an F2-score of 89.3%. With the inclusion of G-CSA alone, the recall increased from 86.0% to 88.5% and the mAP50-95 improved from 84.6% to 86.1%, demonstrating its ability to strengthen global context modeling and enhance robustness under diverse conditions. This improvement, however, was accompanied by a slight reduction in precision, which declined to 91.6%. By contrast, the integration of FasterBlock alone produced the highest precision of 96.7% while also yielding a relatively balanced performance across the other indicators, highlighting its effectiveness in enhancing fine-grained feature extraction.

Most notably, the simultaneous incorporation of G-CSA and FasterBlock achieved the most balanced trade-off across all evaluation metrics, with precision of 92.7%, recall of 90.8%, mAP50 of 94.8%, mAP50-95 of 86.8%, and an F2-score of 91.7%. These outcomes underscore the complementary functions of the two modules: G-CSA enhances contextual reasoning and improves recall, whereas FasterBlock refines feature learning and strengthens precision. Their joint integration provides a robust foundation for accident detection and severity classification, ensuring both stability and generalization.

To provide a more comprehensive perspective, Figure 7 depicts the convergence behavior of different YOLO variants during training, including YOLOv13n (baseline model), G-CSA-YOLO (YOLOv13n enhanced with the Global Channel–Spatial Attention module), FasterBlock-YOLO (YOLOv13n enhanced with the FasterBlock module), and DSGF-YOLO (YOLOv13n with both modules integrated). The curves illustrate the reduction in losses together with the progression of recall and mAP values over 500 epochs. Models incorporating either G-CSA or FasterBlock converge more rapidly and achieve higher recall and mAP compared with the baseline YOLOv13n. The complete DSGF-YOLO configuration consistently demonstrates superior performance throughout training, reflecting the synergistic advantages of combining both modules. As shown in Figure 7, this integration not only accelerates optimization but also yields the most stable and reliable detection outcomes.

As illustrated in Figure 7, the training dynamics of different model variants highlight the contributions of the proposed modules. The Total Loss Curve shows that DSGF-YOLO converges more rapidly and with greater stability compared to the baseline YOLOv13n and the single-module variants. The consistently lower loss values across epochs indicate more efficient optimization and improved generalization when G-CSA and FasterBlock are jointly applied. The Recall Curve further illustrates the complementary roles of the two modules: G-CSA strengthens global contextual perception, leading to higher recall, while FasterBlock contributes to more stable recall behavior. DSGF-YOLO achieves the highest recall among all configurations, underscoring its capacity to reduce missed detections under challenging conditions.

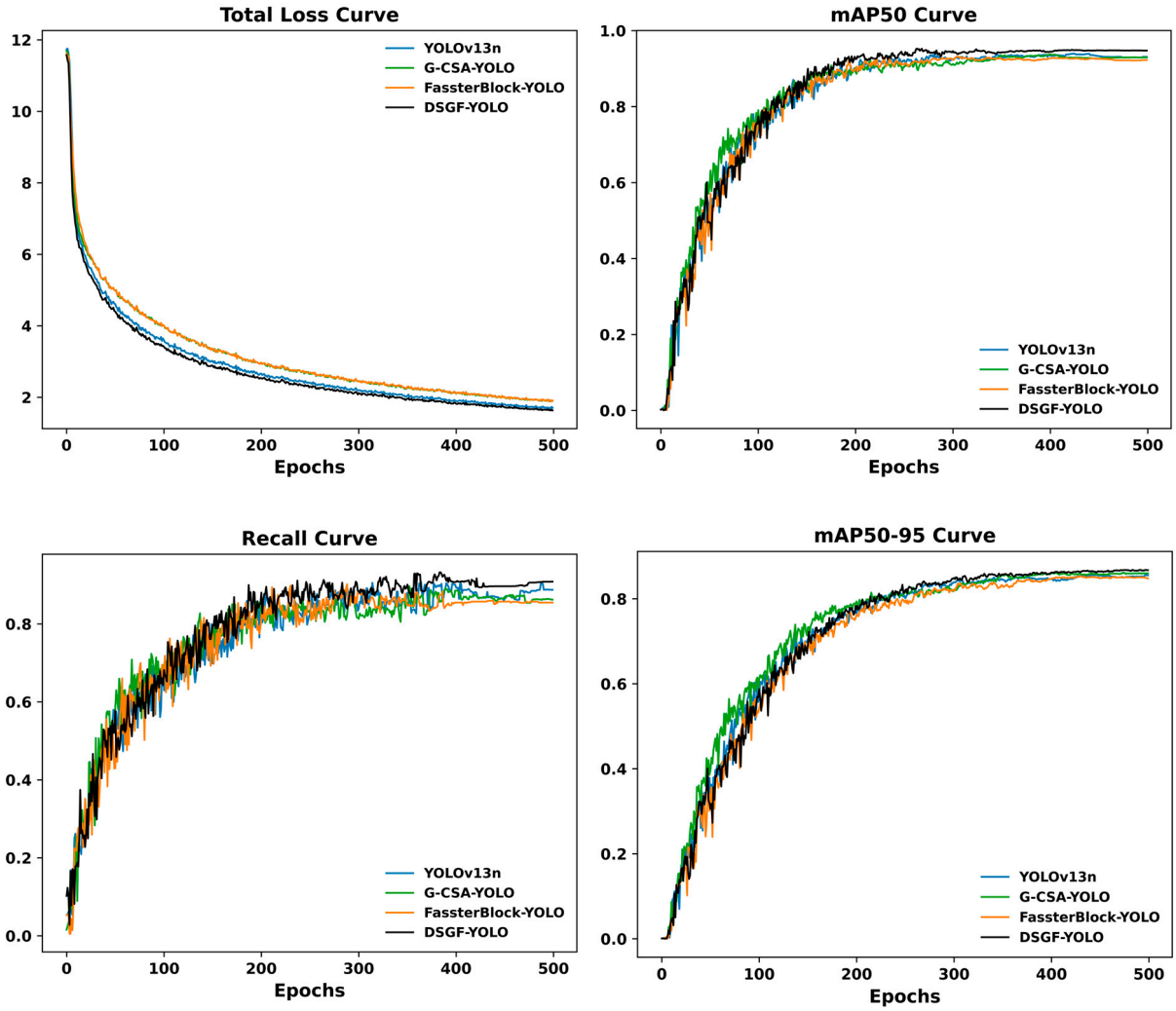


Figure 7. Comparative training curves of YOLO variants across loss, recall, and mAP.

The mAP50 and mAP50-95 curves further highlight the advantages of the proposed architecture. While G-CSA-YOLO, which augments YOLOv13n with the Global Channel–Spatial Attention module, and FasterBlock-YOLO, which incorporates the FasterBlock module, each enhance detection accuracy to a certain extent, their integration within DSGF-YOLO produces the most substantial gains, particularly under stricter IoU thresholds. This outcome suggests that the complementary interaction between global context reasoning and fine-grained feature extraction is crucial for achieving robust detection. Taken together, the results demonstrate that DSGF-YOLO attains faster convergence, higher accuracy, and greater stability, thereby reinforcing its reliability for practical deployment in traffic accident detection and severity classification.

5. Discussion

This study introduced DSGF-YOLO, a novel object detection framework specifically designed to improve the accuracy and robustness of traffic accident detection and severity classification. The key innovation of DSGF-YOLO lies in the synergistic integration of the Global Channel-Spatial Attention (G-CSA) and FasterBlock modules into the baseline YOLOv13n architecture. Extensive experiments demonstrated that DSGF-YOLO achieved notable performance improvements over state-of-the-art detectors such as DETR, Faster R-CNN, and other YOLO variants. In particular, the model attained a precision of 92.7%, recall of 90.8%, mAP50 of 94.8%, and mAP50-95 of 86.8%, thereby establishing a new

performance benchmark for this critical application domain. These gains can be attributed to the complementary mechanisms of the two modules: G-CSA enhances global contextual reasoning, enabling the model to capture accident-relevant cues more effectively, while FasterBlock refines feature extraction, thereby boosting both precision and recall.

The ablation studies further validated the contributions of the proposed modules. When applied independently, G-CSA primarily improved recall by strengthening contextual perception, whereas FasterBlock increased precision through more discriminative feature learning. When combined, these modules yielded the most substantial gains, highlighting their complementary strengths and synergistic effects. This was further supported by the convergence curves of loss and mAP metrics, which showed that DSGF-YOLO not only converged faster but also consistently outperformed the baseline and single-module variants throughout training.

Beyond quantitative improvements, qualitative analyses further demonstrated the effectiveness of the proposed framework. Detection case studies showed that DSGF-YOLO consistently localized accident regions and accurately classified severity across diverse conditions such as low-light nighttime environments, multi-vehicle collisions, and congested traffic scenes. Grad-CAM visualizations also reinforced interpretability, revealing that the model's attention aligned closely with accident-relevant regions. The ability to concentrate on critical visual cues highlights the robustness and practical reliability of DSGF-YOLO for real-world deployment in intelligent transportation systems.

Although the proposed approach demonstrates promising outcomes, several limitations remain. The dataset, despite its comprehensiveness, is relatively limited in size with 1892 training images and restricted geographic coverage, which may constrain generalization to unseen environments. Expanding the dataset to encompass a wider variety of accident scenarios, weather conditions, and global traffic contexts would enhance robustness. While the computational cost of 6.6 GFLOPs is reasonable for real-time deployment, efficiency could be further optimized through techniques such as model compression, pruning, and hardware acceleration. Finally, severity classification in this work was limited to two categories, moderate and severe. Extending the classification granularity to represent a broader spectrum of accident severity would generate more informative outputs for emergency response and intelligent transportation systems.

In terms of technology readiness, the proposed system is currently at the research and prototype stage, yet it demonstrates strong potential for real-world deployment. The framework can be adapted for edge-computing platforms or integrated into intelligent traffic management infrastructures, and pilot trials of similar systems in several cities indicate growing operational feasibility. To enhance practical deployment, the system incorporates an accident alert transmission mechanism that delivers detected events and severity information in real time via existing monitoring networks or 5G wireless communication systems, supporting timely emergency response. Deployment on existing infrastructures requires minimal additional investment, while costs associated with new hardware depend on specific performance and environmental requirements. By emphasizing compatibility with current monitoring systems, the proposed framework provides a cost-efficient pathway toward implementation. Future work will focus on improving model stability, strengthening system integration, and ensuring reliable large-scale deployment in next-generation intelligent transportation systems.

6. Conclusions

This study presents DSGF-YOLO, an advanced object detection framework designed for traffic accident detection and severity classification. By incorporating the G-CSA and FasterBlock modules into the YOLOv13n baseline, the proposed model achieves state-of-

the-art performance, delivering substantial gains in precision, recall, and mean average precision. Its demonstrated capability to localize accident regions with high accuracy and to classify severity levels reliably under complex real-world conditions highlights its strong potential for deployment in intelligent transportation systems. In addition, analyses of loss convergence, mAP evolution, and Grad-CAM visualizations provide a comprehensive understanding of the model's optimization behavior and interpretability, thereby offering valuable insights to guide subsequent research in the field. The framework also demonstrates applicability across both urban and highway environments, further underscoring its versatility for real-world traffic monitoring applications.

Future research will be directed toward addressing the limitations outlined in this study. Key priorities include expanding the dataset to encompass more diverse accident scenarios and traffic environments, incorporating multi-modal data sources to enrich contextual information, enhancing computational efficiency through optimization techniques, and refining severity classification to achieve greater granularity. Another promising direction is the extension of DSGF-YOLO to related tasks such as anomaly detection and incident prediction, which could further improve its utility in intelligent transportation systems. Moreover, as the current system remains at the prototype stage, future work will also focus on improving technology readiness, facilitating integration with existing monitoring infrastructures, and reducing deployment costs. In the longer term, the overarching goal is to develop a comprehensive and reliable framework capable of proactively identifying and mitigating traffic accidents, thereby reducing fatalities and injuries. Beyond traffic accident detection, the proposed approach also demonstrates strong potential for broader applications in intelligent surveillance and autonomous driving, where robust object detection and scene understanding are critical. This research thus provides a solid foundation for advancing the role of artificial intelligence in enhancing traffic safety and supporting emergency response systems.

Author Contributions: Conceptualization, W.L.; methodology, W.L.; software, P.L.; validation, W.L. and P.L.; formal analysis, H.X.; investigation, H.X.; resources, W.L.; data curation, P.L.; writing—original draft preparation, W.L.; writing—review and editing, P.L.; visualization, P.L.; supervision, H.X.; project administration, W.L.; funding acquisition, W.L. All authors have read and agreed to the published version of the manuscript.

Funding: Funds for the Innovation of Policing Science and Technology, Fujian province (Grant number: 2024Y0072).

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Xie, S.; Zhao, Z.; Shangguan, Q.; Fu, T.; Wang, J.; Wu, H. The existence and impacts of sequential traffic conflicts: Investigation of traffic conflict in sequences encountered by left-turning vehicles at signalized intersections. *Accid. Anal. Prev.* **2025**, *215*, 108015. [CrossRef] [PubMed]
2. Katsamenis, I.; Bakalos, N.; Lappas, A.; Protopapadakis, E.; Martín-Portugués Montoliu, C.; Doulamis, A.; Doulamis, N.; Rallis, I.; Kalogeras, D. DORIE: Dataset of Road Infrastructure Elements—A Benchmark of YOLO Architectures for Real-Time Patrol Vehicle Monitoring. *Sensors* **2025**, *25*, 6653. [CrossRef]
3. Hussain, F.; Ali, Y.; Li, Y.; Haque, M.M. Revisiting the hybrid approach of anomaly detection and extreme value theory for estimating pedestrian crashes using traffic conflicts obtained from artificial intelligence-based video analytics. *Accid. Anal. Prev.* **2024**, *199*, 107517. [CrossRef]

4. Naser, Z.S.; Marouane, H.; Fakhfakh, A. Multi-Object-Based Efficient Traffic Signal Optimization Framework via Traffic Flow Analysis and Intensity Estimation Using UCB-MRL-CSFL. *Vehicles* **2025**, *7*, 72. [CrossRef]
5. Khaled, L.B.; Rahman, M.; Ebu, I.A.; Ball, J.E. FlashLightNet: An End-to-End Deep Learning Framework for Real-Time Detection and Classification of Static and Flashing Traffic Light States. *Sensors* **2025**, *25*, 6423. [CrossRef]
6. Yan, H.; Pan, S.; Zhang, S.; Wu, F.; Hao, M. Sustainable utilization of road assets concerning obscured traffic signs recognition. *Proc. Inst. Civ. Eng. Eng. Sustain.* **2025**, *178*, 124–134. [CrossRef]
7. Zhao, S.; Gong, Z.; Zhao, D. Traffic signs and markings recognition based on lightweight convolutional neural network. *Vis. Comput.* **2024**, *40*, 559–570. [CrossRef]
8. Liu, T.; Meidani, H. End-to-end heterogeneous graph neural networks for traffic assignment. *Transp. Res. Part C Emerg. Technol.* **2024**, *165*, 104695. [CrossRef]
9. Talaat, F.M.; ZainEldin, H. An improved fire detection approach based on YOLO-v8 for smart cities. *Neural Comput. Appl.* **2023**, *35*, 20939–20954. [CrossRef]
10. Elassy, M.; Al-Hattab, M.; Takruri, M.; Badawi, S. Intelligent transportation systems for sustainable smart cities. *Transp. Eng.* **2024**, *16*, 100252. [CrossRef]
11. Liang, R.; Jiang, M.; Li, S. YOLO-DPDG: A Dual-Pooling Dynamic Grouping Network for Small and Long-Distance Traffic Sign Detection. *Appl. Sci.* **2025**, *15*, 10921. [CrossRef]
12. Yang, L.; He, Z.; Zhao, X.; Fang, S.; Yuan, J.; He, Y.; Li, S.; Liu, S. A Deep Learning Method for Traffic Light Status Recognition. *J. Intell. Connect. Veh.* **2023**, *6*, 173–182. [CrossRef]
13. Hollósi, J. YOLO-Based Object and Keypoint Detection for Autonomous Traffic Cone Placement and Retrieval for Industrial Robots. *Appl. Sci.* **2025**, *15*, 10845. [CrossRef]
14. Zhang, M.; Li, J.; Zhang, C.; Wang, B.; Tang, X. Integrating deep learning and clustering techniques to address imbalanced data in traffic accident severity prediction. *J. Transp. Saf. Secur.* **2025**, *17*, 1167–1194. [CrossRef]
15. Mohana Deepti, M.; Badrinath, K.; Venkat, P.; Saideep, S. RFCNN: Traffic Accident Severity Prediction based on Decision Level Fusion of Machine and Deep Learning Model. *Int. J. Adv. Res. Comput. Commun. Eng.* **2025**, *14*, 389–404. [CrossRef]
16. Zhu, L.; Wang, B.; Yan, Y.; Guo, S.; Tian, G. A novel traffic accident detection method with comprehensive traffic flow features extraction. *Signal Image Video Process.* **2023**, *17*, 305–313. [CrossRef] [PubMed]
17. Lian, J.; Yin, Y.; Li, L.; Wang, Z.; Zhou, Y. Small Object Detection in Traffic Scenes Based on Attention Feature Fusion. *Sensors* **2021**, *21*, 3031. [CrossRef]
18. Li, K.; Guo, K.; Liang, Z.; Zhang, J.; Zhao, F. Traffic Accident Severity Prediction and Analysis via Spatio-Temporal Deep Learning: A ConvLSTM–Transformer Approach. *Chaos Solitons Fractals* **2025**, *200*, 117047. [CrossRef]
19. Wang, Y.; Qiu, Y.; Cheng, P.; Liu, Z.; Zhang, X. Hybrid CNN–Transformer Features for Visual Place Recognition. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *33*, 1109–1122. [CrossRef]
20. Liu, Y.; Gao, Z.; Ge, H.; Sun, P.; Zhao, L. A Multimodal Deep Learning Approach for Predicting Traffic Accident Severity Using Crash Records, Road Geometry, and Textual Descriptions. *Comput.-Aided Civ. Infrastruct. Eng.* **2025**, *40*, 3773–3793. [CrossRef]
21. Zhou, Q.; Zhang, D.; Liu, H.; He, Y. KCS-YOLO: An Improved Algorithm for Traffic Light Detection under Low Visibility Conditions. *Machines* **2024**, *12*, 557. [CrossRef]
22. Sundaresan Geetha, A.; Alif, M.A.R.; Hussain, M.; Allen, P. Comparative Analysis of YOLOv8 and YOLOv10 in Vehicle Detection: Performance Metrics and Model Efficacy. *Vehicles* **2024**, *6*, 1364–1382. [CrossRef]
23. Xing, C.; Sun, H.; Yang, J. A Lightweight Traffic Sign Detection Model Based on Improved YOLOv8s for Edge Deployment in Autonomous Driving Systems Under Complex Environments. *World Electr. Veh. J.* **2025**, *16*, 478. [CrossRef]
24. Ashraf, I.; Hur, S.; Kim, G.; Park, Y. Analyzing Performance of YOLOx for Detecting Vehicles in Bad Weather Conditions. *Sensors* **2024**, *24*, 522. [CrossRef] [PubMed]
25. Zhang, Y.; Lu, Y.; Huo, Z.; Li, J.; Sun, Y.; Huang, H. USSC-YOLO: Enhanced Multi-Scale Road Crack Object Detection Algorithm for UAV Image. *Sensors* **2024**, *24*, 5586. [CrossRef]
26. Feng, Z.; Liu, F. Balancing Feature Symmetry: IFEM-YOLOv13 for Robust Underwater Object Detection Under Degradation. *Symmetry* **2025**, *17*, 1531. [CrossRef]
27. Luan, T.; Zhou, S.; Zhang, Y.; Pan, W. Fast Identification and Detection Algorithm for Maneuverable Unmanned Aircraft Based on Multimodal Data Fusion. *Mathematics* **2025**, *13*, 1825. [CrossRef]
28. Wei, F.; Wang, W. SCCA-YOLO: A Spatial and Channel Collaborative Attention Enhanced YOLO Network for Highway Autonomous Driving Perception System. *Sci. Rep.* **2025**, *15*, 6459. [CrossRef]
29. Chai, Y.; Yao, X.; Chen, M.; Shan, S. FPFS-YOLO: An Insulator Defect Detection Model Integrating FasterNet and an Attention Mechanism. *Sensors* **2025**, *25*, 4165. [CrossRef]
30. Wu, L.; Li, X.; Ma, P.; Cai, Y. Research on a Dense Pedestrian-Detection Algorithm Based on an Improved YOLO11. *Future Internet* **2025**, *17*, 438. [CrossRef]

31. Li, Z. Mamba with split-based pyramidal convolution and Kolmogorov-Arnold network-channel-spatial attention for electroencephalogram classification. *Front. Sens.* **2025**, *6*, 2673–5067. [CrossRef]
32. Merolla, D.; Latorre, V.; Salis, A.; Boanelli, G. Improving Road Safety with AI: Automated Detection of Signs and Surface Damage. *Computers* **2025**, *14*, 91. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.