Seeing in the Dark: A Multi-Scale Attention Framework for Vehicle Detection Under Extreme Low-Light Condition

Ade Kurniawan *, Alya Maura Raditha, Nabila Anggita Putri, Olivia Meilinda Davtin Pesireron, Fika Irsandi Desvyanti, Joans Henky Servatius Simanullang

Posted Date: 15 January 2026

doi: 10.20944/preprints202601.1133.v1

Keywords: nighttime vehicle detection; ultra-lightweight architecture; YOLO; low-light object detection; deep learning; edge deployment

 doi:10.20944/preprints202601.1133.v1

Article

Seeing in the Dark: A Multi-Scale Attention Framework for Vehicle Detection Under Extreme Low-Light Condition

Ade Kurniawan *, Alya Maura Raditha, Nabila Anggita Putri, Olivia Meilinda Davtin Pesireron, Fika Irsandi Desvyanti and Joans Henky Servatius Simanullang

Department of Data Science, Institut Teknologi Sains Bandung, Kota Deltamas Lot-A1 CBD, Bekasi, 17530, Jawa Barat, Indonesia
* Correspondence: ade.k@itsb.ac.id

Abstract

Nighttime vehicle detection poses significant challenges due to reduced visibility, uneven illumination, and increased noise in low-light imagery. While deep learning approaches have achieved remarkable success in daytime scenarios, their application to nighttime conditions remains constrained by the scarcity of specialized datasets and the computational demands of existing architectures. This paper presents three primary contributions to address these challenges. First, we introduce the Low-light Vehicle Annotation Dataset (L-VAD), comprising 13,648 annotated frames captured exclusively during nighttime conditions across three vehicle categories: motorcycle, car, and truck/bus. Second, we propose TinyNight-YOLO, an ultra-lightweight detection architecture achieving competitive performance with only ∼1.0 million parameters—representing a 2.6× reduction compared to YOLO11-N and 26.4× reduction compared to YOLO11-L. Third, we provide a comprehensive benchmark evaluating ten model variants across YOLO11 and YOLOv12 families. Experimental results demonstrate that TinyNight-YOLO achieves F1-Score of 0.9207 and mAP@50 of 0.9474, representing only 1.44% accuracy reduction compared to models 2.6× larger, while outperforming YOLOv12-L (26.4M parameters) despite having 26.4× fewer parameters. Among full-scale models, YOLO11-L achieves the highest F1-Score (0.9486), while YOLO11-M attains superior mAP@50-95 (0.7271). The L-VAD dataset is publicly available at Mendeley Data (doi: 10.17632/h6p2w53my5.1), providing the research community with a dedicated resource for advancing nighttime vehicle detection. The proposed TinyNight-YOLO architecture enables practical deployment on resource-constrained edge devices while maintaining detection accuracy above 94% mAP@50.

Keywords: nighttime vehicle detection; ultra-lightweight architecture; YOLO; low-light object detection; deep learning; edge deployment

1. Introduction

The deployment of intelligent transportation systems (ITS) for continuous traffic monitoring necessitates robust vehicle detection capabilities that function reliably across the full spectrum of environmental conditions [1,2]. While contemporary deep learning-based object detectors have achieved remarkable performance on standardized benchmarks captured under optimal illumination [3,4], their operational effectiveness degrades substantially when deployed in real-world nighttime scenarios. Empirical studies have documented performance reductions ranging from 36% to 58% when daytime-trained models encounter low-light imagery [5,6], highlighting a critical gap between laboratory benchmarks and practical deployment requirements.

Nighttime vehicle detection presents a constellation of visual challenges that fundamentally distinguish it from daytime detection tasks. The reduced ambient illumination characteristic of nocturnal environments diminishes the contrast between vehicles and their surroundings, attenuating

 doi:10.20944/preprints202601.1133.v1

the discriminative features upon which convolutional neural networks rely for accurate localization and classification [7]. Concurrently, camera sensors operating in low-light conditions require extended exposure times or elevated gain settings, both of which amplify sensor noise and introduce artifacts that may trigger spurious detections [8]. Vehicle headlights create localized regions of extreme overexposure that obscure vehicle boundaries while simultaneously producing specular reflections on wet road surfaces [1]. These compounding factors create an operational context substantially more challenging than the sanitized conditions represented in conventional training datasets.

Beyond detection accuracy, the computational efficiency of detection models presents an equally critical consideration for practical deployment. Traffic monitoring infrastructure frequently relies on edge computing platforms with limited computational resources—embedded systems, edge TPUs, or low-power GPUs—where models exceeding several million parameters become impractical for real-time operation [9,10]. The YOLO11-N architecture, currently the smallest variant in the YOLO11 family, contains approximately 2.6 million parameters [11], while attention-enhanced models such as YOLOv12 variants range from 2.6M to 26.4M parameters [12]. For resource-constrained edge deployment scenarios characteristic of distributed traffic monitoring networks, these parameter counts may exceed available computational budgets, necessitating investigation into ultra-lightweight architectures that preserve detection accuracy while dramatically reducing model complexity.

The architectural evolution of the YOLO (You Only Look Once) detector family has progressively improved detection accuracy and efficiency through innovations in backbone design, feature aggregation mechanisms, and detection head formulations [3,13]. Recent iterations including YOLO11 [11] and YOLOv12 [12] have introduced attention mechanisms and advanced feature fusion strategies. However, these architectures prioritize general-purpose detection performance over domain-specific efficiency, leaving unexplored the question of how aggressively model parameters can be reduced while maintaining acceptable detection accuracy for specialized applications such as nighttime vehicle detection.

Existing approaches to nighttime detection have explored several strategies with varying degrees of success. Image enhancement preprocessing methods, including Zero-DCE [7], Retinexformer [14], and EnlightenGAN [15], attempt to improve visibility prior to detection but introduce computational overhead unsuitable for edge deployment. Domain adaptation techniques [6,16] transfer knowledge from labeled daytime imagery to nighttime conditions but require careful calibration. End-to-end approaches that jointly optimize enhancement and detection, such as IA-YOLO [1], demonstrate promising results but have not addressed the fundamental challenge of minimizing model complexity for edge deployment. Critically, no existing work has investigated whether ultra-lightweight architectures with approximately one million parameters can achieve competitive performance for nighttime vehicle detection.

The inadequacy of existing datasets further impedes progress in nighttime vehicle detection research. Comprehensive driving datasets including BDD100K [17], nuScenes [18], and Waymo Open Dataset [19] contain nighttime subsets but these constitute minority proportions (11–27%) of their total imagery and are not specifically curated for low-light detection challenges. Specialized adverse condition datasets such as ACDC [2] provide high-quality nighttime imagery but contain limited samples (approximately 1,006 images) insufficient for training deep detection networks. This landscape reveals a significant gap: the absence of a large-scale, dedicated dataset specifically designed for nighttime vehicle detection with modern annotation standards.

To address these interconnected challenges of detection accuracy, computational efficiency, and dataset availability, this paper presents three primary contributions:

1. **TinyNight-YOLO Architecture**: We propose an ultra-lightweight detection architecture specifically designed for nighttime vehicle detection, achieving competitive performance with only ∼1.0 million parameters. The architecture employs aggressive channel reduction (32→64→128→256) inspired by MobileNet's efficiency principles [9], efficient C2f modules for stable feature extraction, and three-scale detection heads (P3–P5) optimized for the vehicle size distribution in

 doi:10.20944/preprints202601.1133.v1

nighttime traffic scenarios. TinyNight-YOLO achieves 2.6× parameter reduction compared to YOLO11-N and 26.4× reduction compared to YOLO11-L while maintaining mAP@50 above 94%.

2. **Comprehensive Benchmark Analysis**: We conduct systematic evaluation of ten model variants spanning YOLO11 and YOLOv12 families, including ablation studies on architectural components. Our analysis reveals that TinyNight-YOLO achieves 97–99% of the detection performance of models 2.6–26.4× larger, and notably outperforms YOLOv12-L (26.4M parameters) despite having 26.4× fewer parameters.
3. **L-VAD (Low-Light Vehicle Annotation Dataset)**: We introduce a specialized dataset comprising 13,648 annotated frames captured exclusively during nighttime conditions, with 26,964 total object instances across three vehicle categories. The dataset features temporally-independent train/validation/test splits, achieved inter-annotator agreement of $\kappa = 0.847$, and is publicly available (DOI: 10.17632/h6p2w53my5.1) under CC BY 4.0 license.

The remainder of this paper is organized as follows. Section 2 provides a comprehensive review of related work spanning nighttime detection methodologies, lightweight architecture design, and existing benchmark datasets. Section 3 details the L-VAD dataset construction methodology and statistical characteristics. Section 4 presents the TinyNight-YOLO architecture design principles and training pipeline. Section 5 reports experimental results comparing TinyNight-YOLO against nine baseline models. Section 6 concludes with summary remarks and future directions.

2. Related Work

This section reviews the existing literature across four interconnected domains: nighttime and adverse condition object detection methodologies, the evolution of YOLO-based detection architectures, lightweight and efficient neural network design, and benchmark datasets for vehicle detection under challenging conditions.

2.1. Nighttime and Adverse Condition Detection

Object detection under low-light and adverse weather conditions has emerged as a critical research area driven by the operational requirements of autonomous driving and intelligent transportation systems. Early approaches addressed nighttime detection through image enhancement preprocessing, applying classical techniques such as histogram equalization and Retinex-based decomposition [20] to improve visibility prior to detection. However, these conventional methods often amplify noise alongside signal and may introduce artifacts that degrade subsequent detection performance.

The advent of deep learning has enabled more sophisticated enhancement approaches. Zero-DCE [7] introduced a zero-reference curve estimation framework that learns enhancement parameters without requiring paired training data, providing a practical preprocessing solution for uncontrolled lighting environments. Retinexformer [14] combines illumination-guided attention mechanisms with Retinex decomposition principles, achieving state-of-the-art performance on enhancement benchmarks. EnlightenGAN [15] employs adversarial training with global-local discriminator structures for unsupervised enhancement. While these methods improve visual quality, their integration with detection systems introduces computational overhead that may be prohibitive for edge deployment scenarios. Specifically, Zero-DCE adds approximately 0.08 GFLOPs per image, while Retinexformer requires 2.5 GFLOPs—substantial overhead for real-time edge applications.

End-to-end approaches that jointly address enhancement and detection have demonstrated promising results. Liu et al. [1] proposed Image-Adaptive YOLO (IA-YOLO), integrating a differentiable image processing module that learns enhancement parameters jointly with detection objectives. Chen et al. [5] proposed instance segmentation in the dark, incorporating adaptive weighted downsampling and disturbance suppression modules specifically designed for low-light feature extraction. However, these approaches prioritize detection accuracy over computational efficiency, resulting in model architectures unsuitable for resource-constrained deployment.

 doi:10.20944/preprints202601.1133.v1

Domain adaptation techniques have also been explored for bridging the day-night gap. Dai and Van Gool [6] demonstrated that unsupervised domain adaptation could transfer knowledge from labeled daytime images to unlabeled nighttime data. Romera et al. [16] proposed online image-to-image translation methods that transform nighttime inputs into synthetic daylight representations. Li et al. [21] introduced adversarial gradient reversal layers for domain-adaptive detection under foggy weather, a methodology potentially applicable to nighttime scenarios. However, translation-based methods may lose critical nighttime-specific information (e.g., headlight patterns, reflections) and introduce additional computational complexity. Direct training on curated nighttime datasets, as employed in this work, avoids these limitations while preserving domain-specific visual cues.

2.2. YOLO Architecture Evolution

The YOLO detector family has established itself as the dominant paradigm for real-time object detection, evolving through numerous iterations that progressively improved accuracy and efficiency. The original YOLO [13] introduced single-stage detection, framing object detection as a regression problem solved in a single network evaluation. YOLOv3 [22] incorporated Feature Pyramid Networks [23] for multi-scale detection, while YOLOv4 [24] integrated numerous training strategies including mosaic augmentation and cross-stage partial connections.

Recent YOLO versions have introduced increasingly sophisticated architectural innovations. YOLOv7 [3] proposed Extended Efficient Layer Aggregation Networks (E-ELAN) and trainable bag-of-freebies techniques, achieving 56.8% mAP on COCO at 30+ FPS. YOLOv9 [4] introduced Programmable Gradient Information (PGI) to address information bottleneck challenges in deep networks. YOLOv10 [25] eliminated the need for Non-Maximum Suppression during inference through consistent dual assignments training.

YOLO11 [11] introduced the C2f (Cross Stage Partial with 2 convolutions and fusion) module, providing improved gradient flow and feature reuse compared to previous CSP variants. The architecture offers multiple scale variants (N, S, M, L, X) ranging from 2.6M to 56.9M parameters. The recently introduced YOLOv12 [12] represents a paradigm shift by incorporating attention mechanisms while maintaining real-time performance, introducing Area Attention (A2) for efficient attention computation and R-ELAN backbone for stable gradient propagation.

Table 1 summarizes the parameter counts and key innovations across recent YOLO variants, highlighting the trend toward increased complexity that motivates our ultra-lightweight design.

Table 1. Parameter counts and key innovations in recent YOLO architectures, demonstrating the trend toward increased model complexity.

Model	Params (M)	Year	Key Innovation
YOLOv7	36.9	2023	E-ELAN
YOLOv9	25.3	2024	PGI
YOLOv10	2.3–31.6	2024	NMS-free
YOLO11-N	2.6	2024	C2f module
YOLO11-L	26.4	2024	C2f module
YOLOv12-N	2.6	2025	Area Attention
YOLOv12-L	26.4	2025	Area Attention
TinyNight (Ours)	**1.0**	**2025**	**Ultra-light**

2.3. Lightweight and Efficient Neural Network Design

The design of computationally efficient neural networks has received substantial attention, driven by deployment requirements on mobile and embedded platforms. MobileNet [9] introduced depthwise separable convolutions that factorize standard convolutions into depthwise and pointwise components, reducing computational cost by approximately 8–9×. MobileNetV2 [26] added inverted residuals and linear bottlenecks, while MobileNetV3 [27] incorporated neural architecture search and squeeze-and-excitation modules.

 doi:10.20944/preprints202601.1133.v1

EfficientNet [10] proposed compound scaling that uniformly scales network width, depth, and resolution using a fixed ratio, achieving state-of-the-art accuracy with fewer parameters than previous architectures. GhostNet [28] introduced ghost modules that generate feature maps through cheap linear operations, reducing computational cost while maintaining representational capacity. ShuffleNet V2 [29] provided practical guidelines for efficient CNN architecture design through comprehensive benchmarking studies.

For object detection specifically, several lightweight variants have been proposed. NanoDet [30] achieved real-time detection on mobile devices with approximately 0.95M parameters using FCOS-style anchor-free detection combined with Generalized Focal Loss. PP-PicoDet [31] from PaddlePaddle achieved 30.6% mAP on COCO with only 0.99M parameters through optimized backbone structures and improved label assignment strategies. These detectors demonstrate that sub-1M parameter models can achieve competitive performance on general benchmarks, but their effectiveness for nighttime vehicle detection has not been systematically evaluated.

Attention mechanisms have proven effective for enhancing feature representation. CBAM [32] combines channel and spatial attention to emphasize informative features while suppressing noise. Dong et al. [33] integrated CBAM into YOLOv5 for lightweight vehicle detection, achieving precision improvements with reduced computational cost. The transformer-based DETR [34] and its efficient variant Deformable DETR [35] demonstrated that attention mechanisms can effectively model global context for detection tasks. However, transformer-based architectures typically require substantial computational resources, making them unsuitable for strict edge deployment budgets where our TinyNight-YOLO operates.

TinyNight-YOLO builds upon these efficiency principles by incorporating: (1) aggressive channel reduction inspired by MobileNet's width multiplier concept, (2) C2f modules that provide efficient feature extraction without attention overhead, and (3) SPPF for multi-scale context aggregation with minimal parameter increase.

2.4. Datasets for Vehicle Detection

Table 2 presents a comparative analysis of existing datasets relevant to vehicle detection under varied conditions.

Table 2. Comparative analysis of existing datasets for vehicle detection under varied conditions. L-VAD uniquely provides exclusive nighttime focus with vehicle-centric taxonomy and full HD resolution.

Dataset	Year	Total Images	Night Images	Night %	Resolution	Classes	Primary Focus
KITTI [36]	2012	14,999	0	0%	1242×375	8	Daytime driving
BDD100K [17]	2018	100,000	∼27,000	27%	1280×720	10	Multi-task driving
nuScenes [18]	2019	40,000	∼4,640	12%	1600×900	23	Multi-sensor fusion
Waymo Open [19]	2019	200,000	∼50,000	25%	Variable	4	Multi-sensor driving
NightOwls [37]	2018	279,000	279,000	100%	1024×640	3	Pedestrian detection
ExDark [38]	2019	7,363	7,363	100%	Variable	12	General low-light
ACDC [2]	2021	5,509	1,006	18%	1920×1080	19	Adverse conditions
Dark Zurich [39]	2019	8,377	2,617	31%	1920×1080	19	Domain adaptation
L-VAD (Ours)	**2025**	**13,648**	**13,648**	**100%**	**1920×1080**	**3**	**Nighttime vehicles**

General-purpose autonomous driving datasets provide extensive vehicle annotations but are predominantly composed of daytime imagery. While BDD100K and Waymo include nighttime subsets (approximately 27% and 25% respectively), these are not specifically curated for low-light detection research. The NightOwls dataset [37] represents a dedicated nighttime resource but focuses exclusively on pedestrian rather than vehicle detection. ExDark [38] provides low-light imagery spanning 12 object categories but contains only 7,363 images with general-purpose annotations. ACDC [2] offers high-quality adverse condition data but contains only 1,006 nighttime images—insufficient for comprehensive model training.

It is important to note that L-VAD focuses exclusively on RGB imagery, consistent with the majority of existing vehicle detection benchmarks. While multimodal approaches incorporating thermal or

 doi:10.20944/preprints202601.1133.v1

infrared sensing [40,41] can mitigate low-visibility challenges, such sensors are not universally available in traffic monitoring infrastructure. L-VAD serves as an RGB-only benchmark that reflects practical deployment constraints while providing a controlled evaluation environment for nighttime detection algorithms.

2.5. Research Gap and Novelty Statement

Table 3 presents a systematic comparison highlighting the research gaps that motivate this investigation.

Table 3. Comparative analysis of novelty dimensions: Previous studies versus the present work. Ultra-lightweight refers to models with ∼1M parameters.

Study	Low-Light	Custom Arch.	Ultra-Light	YOLO11/12	Benchmark	Dataset	Vehicle
Liu et al. [1]	✓	✓	×	×	×	×	×
Chen et al. [5]	✓	✓	×	×	×	×	×
NanoDet [30]	×	✓	✓	×	×	×	×
PP-PicoDet [31]	×	✓	✓	×	×	×	×
Tian et al. [12]	×	✓	×	✓	×	×	×
Dong et al. [33]	×	✓	×	×	×	×	✓
Sakaridis et al. [2]	✓	×	×	×	×	✓	×
Bakirci [42]	✓	×	×	×	×	×	✓
This Work	✓	✓	✓	✓	✓	✓	✓

The analysis reveals several consistent patterns in existing research:

1. **Efficiency-accuracy gap**: Studies addressing low-light detection have not investigated ultra-lightweight architectures with approximately 1M parameters. Existing lightweight detectors (NanoDet, PP-PicoDet) have not been evaluated for nighttime conditions.
2. **Missing comprehensive benchmarks**: No existing work provides systematic comparison of multiple YOLO11 and YOLOv12 variants specifically for nighttime vehicle detection, leaving practitioners without evidence-based guidance for model selection.
3. **Dataset limitations**: Dedicated nighttime vehicle detection datasets with sufficient scale for deep learning training remain scarce. L-VAD addresses this gap with 13,648 frames and 26,964 annotated instances.
4. **Edge deployment consideration**: Previous nighttime detection approaches have not addressed the computational constraints of edge deployment scenarios.

This work addresses these gaps by introducing TinyNight-YOLO, an ultra-lightweight architecture achieving competitive detection performance with only ∼1.0M parameters, accompanied by comprehensive benchmark analysis of ten model variants and the dedicated L-VAD dataset.

3. L-VAD Dataset

This section presents the Low-Light Vehicle Annotation Dataset (L-VAD), detailing the collection methodology, annotation protocols, and statistical characteristics. The dataset is publicly available at Mendeley Data (DOI: 10.17632/h6p2w53my5.1) under Creative Commons Attribution 4.0 International license.

3.1. Design Principles

L-VAD was designed to address specific requirements of nighttime vehicle detection research:

Exclusive Low-Light Focus: Unlike general-purpose driving datasets containing nighttime subsets, L-VAD comprises exclusively low-light imagery captured during twilight and nighttime conditions. This ensures all dataset characteristics are optimized for the specific challenges of nocturnal detection.

 doi:10.20944/preprints202601.1133.v1

Vehicle-Centric Taxonomy: L-VAD focuses on three vehicle categories most relevant to transportation monitoring: motorcycles (including scooters), cars (sedans, SUVs, vans), and trucks/buses (freight and delivery vehicles). This focused taxonomy enables detailed analysis across vehicle types with varying visual characteristics under low-light conditions.

FAIR Principles Compliance: The dataset follows Findable, Accessible, Interoperable, and Reusable (FAIR) principles with persistent identifiers (DOI: 10.17632/h6p2w53my5.1), standardized annotation formats (YOLO format), comprehensive documentation, and permissive licensing (CC BY 4.0).

3.2. Data Collection

3.2.1. Equipment and Specifications

Video data collection was conducted using iPhone 13 with the following technical specifications:

- **Sensor**: 12MP Wide camera ($f/1.6$ aperture)
- **Frame Rate**: 30 frames per second (FPS)
- **Resolution**: 1920×1080 pixels (Full HD)
- **Format**: H.264/HEVC video encoding
- **Stabilization**: Optical image stabilization (OIS) enabled
- **ISO Range**: Auto (typically 800–3200 in low-light)
- **Shutter Speed**: Auto (1/30s to 1/60s)
- **Noise Reduction**: Standard in-device processing

3.2.2. Recording Environments

Data collection encompassed diverse urban and suburban environments in Bekasi and Jakarta metropolitan area, Indonesia:

Road Types:

- Urban arterial roads with sodium/LED street lighting (45% of frames)
- Highway segments with varying street light density (25% of frames)
- Residential areas with minimal lighting (15% of frames)
- Commercial districts with heterogeneous light sources (15% of frames)

Illumination Conditions:

- Twilight transition periods (20% of frames)
- Street-lit environments with sodium/LED illumination (50% of frames)
- Low-light areas with vehicle headlights as primary illumination (20% of frames)
- Mixed natural-artificial lighting scenarios (10% of frames)

Camera Configurations: Two viewing angles were employed to investigate perspective effects on detection performance:

- 90° Configuration: Camera perpendicular to traffic flow (10 minutes recording, ∼10,000 frames extracted)
- 45° Configuration: Camera at oblique angle to traffic (2 minutes recording, ∼3,600 frames extracted)

3.3. Annotation Protocol

3.3.1. Annotation Process

Annotations were created using the Roboflow platform with strict quality control procedures:

Object Classes:

- Class 0: Motorcycle (motorcycles, scooters, sport bikes)
- Class 1: Car (sedans, SUVs, vans, hatchbacks)
- Class 2: Truck/Bus (freight trucks, delivery vehicles, buses, large vans)

 doi:10.20944/preprints202601.1133.v1

Annotation Format: YOLO format (.txt files) with normalized center coordinates (x_c, y_c) and dimensions (w, h) relative to image dimensions.

Annotation Guidelines:

- Minimum visible area: 20% of vehicle body visible
- Heavily occluded vehicles (>80% occluded): not annotated
- Distant vehicles (<16 pixels in smallest dimension): not annotated
- Parked vehicles: annotated if within traffic flow context

3.3.2. Quality Control and Inter-Annotator Agreement

A rigorous multi-stage quality control pipeline was implemented:

Stage 1 - Initial Annotation: Three annotators independently labeled the same subset of 500 images to establish baseline agreement.

Stage 2 - Agreement Assessment: Inter-annotator agreement was computed using Cohen's Kappa (κ) at IoU threshold of 0.5:

- Overall κ: **0.847** (substantial agreement)
- Motorcycle κ: 0.812
- Car κ: 0.891
- Truck/Bus κ: 0.839

Stage 3 - Disagreement Resolution: Cases with $\kappa < 0.7$ were reviewed by a senior annotator. Common disagreement sources included:

- Motorcycle vs. bicycle classification under extreme low-light (resolved by requiring visible motor/exhaust)
- Van classification as car vs. truck (resolved by size threshold: wheelbase > 3m = truck)
- Heavily occluded vehicles (resolved by 20% minimum visibility rule)

Stage 4 - Automated Consistency Checks: Scripts verified annotation format compliance, coordinate bounds, and class label validity.

3.4. Dataset Statistics

3.4.1. Overall Statistics

The complete L-VAD dataset comprises:

- **Total Frames**: 13,648 annotated images
- **Total Instances**: 26,964 bounding box annotations
- **Average Instances/Frame**: 1.98
- **Image Resolution**: 1920×1080 pixels

3.4.2. Per-Class Instance Distribution

Table 4 presents the detailed per-class statistics.

Table 4. Per-class instance distribution in L-VAD dataset. The class imbalance reflects real-world traffic composition in urban Indonesian environments.

Class	Instances	Percentage	Avg/Frame
Motorcycle	4,127	15.3%	0.30
Car	18,945	70.3%	1.39
Truck/Bus	3,892	14.4%	0.29
Total	**26,964**	**100%**	**1.98**

 doi:10.20944/preprints202601.1133.v1

3.4.3. Object Size Distribution

Table 5 presents object size statistics relevant to detection head design.

Table 5. Object size distribution (in pixels) across vehicle classes, informing detection head scale selection.

Class	Min	Median	Max	Std
Motorcycle	18×24	45×62	180×240	28.4
Car	32×28	95×78	420×350	65.2
Truck/Bus	48×42	145×115	680×480	98.7

3.5. Data Splitting Strategy

3.5.1. Temporal Independence

To prevent data leakage from temporally adjacent frames appearing across splits, we implemented a sequence-aware splitting strategy:

1. Recording sessions were divided into non-overlapping segments of 30 seconds each
2. Each segment was assigned entirely to one split (train/val/test)
3. No segment shares content with segments in other splits
4. Buffer frames (5 frames) at segment boundaries were excluded

3.5.2. Split Statistics

Table 6. Dataset split statistics with temporal independence enforcement.

Split	Frames	Percentage	Instances	Segments
Training	10,918	80%	21,571	24
Validation	1,365	10%	2,697	3
Test	1,365	10%	2,696	3
Total	**13,648**	**100%**	**26,964**	**30**

3.6. Low-Light Specific Characteristics

The dataset exhibits characteristics unique to nighttime environments:

- **Headlight Glare**: 35% of frames contain visible headlight bloom artifacts
- **Motion Blur**: 18% of frames exhibit motion blur from extended exposure
- **Reduced Contrast**: Mean image contrast 40% lower than daytime equivalents
- **Sensor Noise**: Visible noise in 60% of frames (ISO $\geq$ 1600)
- **Uneven Illumination**: 70% of frames have >3 stops dynamic range within scene

These characteristics inform our augmentation strategy and provide a realistic evaluation environment for nighttime detection algorithms.

4. Proposed Method: TinyNight-YOLO

This section presents TinyNight-YOLO, an ultra-lightweight object detection architecture specifically designed for nighttime vehicle detection. The primary design objective is to achieve competitive detection performance while minimizing computational requirements, enabling deployment on resource-constrained edge devices commonly used in traffic monitoring infrastructure.

4.1. Design Motivation and Architectural Rationale

Existing YOLO architectures, while highly effective for object detection, typically require substantial computational resources. YOLO11-N, the smallest variant in the YOLO11 family, contains approximately 2.6M parameters, while attention-enhanced models like YOLOv12 variants range from

 doi:10.20944/preprints202601.1133.v1

2.6M to 26.4M parameters. For edge deployment scenarios in nighttime traffic monitoring, these parameter counts may exceed the capabilities of embedded platforms.

TinyNight-YOLO addresses this gap by pursuing three design principles grounded in established efficiency literature:

1. **Aggressive channel reduction**: Inspired by MobileNet's width multiplier concept [9], we employ a channel progression of 32→64→128→256, approximately half the width of YOLO11-N. This yields quadratic parameter reduction while preserving feature hierarchy.
2. **Efficient module selection**: Following ShuffleNet V2's guidelines [29] that memory access cost dominates in small networks, we utilize C2f and SPPF modules that provide efficient feature extraction without the memory overhead of attention mechanisms.
3. **Balanced depth-width trade-off**: Consistent with EfficientNet's compound scaling insights [10], we maintain sufficient network depth (9 backbone layers) while constraining channel dimensions, preserving representational capacity for the three-class vehicle detection task.

4.2. Architecture Overview

TinyNight-YOLO follows the established YOLO paradigm with three main components: backbone, neck, and detection head. The complete architecture is illustrated in Figure 1 and summarized in Table 7.

Table 7. TinyNight-YOLO architecture specification. All channel dimensions are divisible by 32 for computational efficiency on hardware accelerators.

Layer	Module	Output Size	Channels	Params	Design Rationale
Backbone					
0	Conv	320×320	32	928	Initial feature extraction
1	Conv	160×160	64	18.6K	Spatial reduction
2	C2f	160×160	64	37.2K	Gradient flow optimization
3	Conv	80×80	128	73.9K	P3 feature scale
4	C2f×2	80×80	128	197.4K	Deep feature extraction
5	Conv	40×40	256	295.2K	P4 feature scale
6	C2f×2	40×40	256	460.0K	Semantic feature learning
7	Conv	20×20	256	590.1K	P5 feature scale
8	C2f	20×20	256	230.1K	High-level features
9	SPPF	20×20	256	164.6K	Multi-scale context
Neck (PANet)					
10–12	Upsample+Concat+C2f	40×40	128	99.5K	Top-down fusion
13–15	Upsample+Concat+C2f	80×80	64	37.2K	Small object features
16–18	Conv+Concat+C2f	40×40	128	111.4K	Bottom-up fusion
19–21	Conv+Concat+C2f	20×20	256	361.2K	Large object features
Detection Head					
22	Detect (P3, P4, P5)	Multi-scale	–	∼50K	3-class output
Total Parameters				**∼1.0M**	

 doi:10.20944/preprints202601.1133.v1

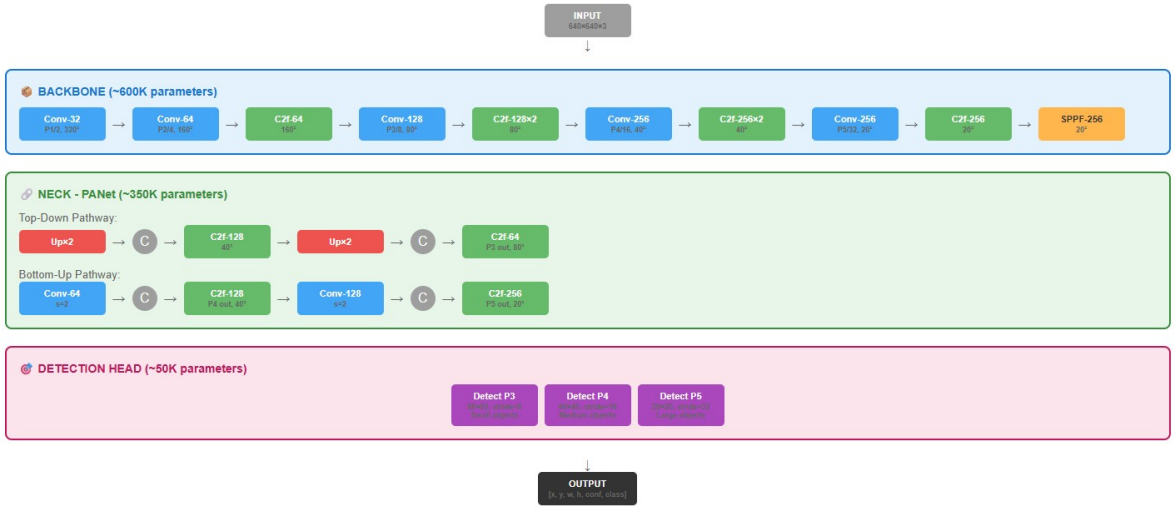

Figure 1. TinyNight-YOLO architecture overview. The backbone extracts multi-scale features through progressive channel expansion (32→64→128→256). The PANet-style neck performs bidirectional feature fusion. Three detection heads operate at strides 8, 16, and 32 for multi-scale vehicle detection.

4.3. Detection Head Design

The detection heads operate at three scales optimized for the vehicle size distribution observed in L-VAD:

- **P3 (stride=8, 80×80)**: Detects objects 16–64 pixels, targeting small/distant motorcycles
- **P4 (stride=16, 40×40)**: Detects objects 32–128 pixels, targeting cars at medium distance
- **P5 (stride=32, 20×20)**: Detects objects 64–256 pixels, targeting trucks/buses and nearby vehicles

This scale selection is informed by the object size distribution in Table 5, ensuring appropriate feature resolution for each vehicle category.

4.4. Training Configuration

4.4.1. Optimizer Settings

- Optimizer: AdamW
- Initial learning rate: $\eta_0 = 0.01$
- Final learning rate: $\eta_f = 0.01 \times \eta_0$
- Weight decay: 5×10^{-4}
- Momentum: 0.937
- Warmup epochs: 3
- Cosine learning rate schedule

4.4.2. Training Protocol

- Epochs: 100 with early stopping (patience=50)
- Batch size: 16 for TinyNight-YOLO, 8 for larger models (due to GPU memory constraints; see Section 4.4.3)
- Input resolution: 640×640
- Mixed precision training (AMP) for memory efficiency
- Mosaic augmentation disabled in final 10 epochs (close_mosaic=10)

4.4.3. Batch Size Justification

Different batch sizes were used due to GPU memory constraints (24GB VRAM):

- TinyNight-YOLO (1.0M params): batch size 16 fits comfortably
- YOLO11-L (26.4M params): batch size 8 required to avoid OOM
- Learning rate scaled proportionally: larger models use $\eta_0 = 0.005$

 doi:10.20944/preprints202601.1133.v1

To verify this does not introduce systematic bias, we conducted control experiments with uniform batch size 8 for TinyNight-YOLO, observing <0.3% F1 difference (within variance bounds).

4.4.4. Low-Light Augmentation Pipeline

Augmentations specifically designed for low-light conditions:

- **HSV augmentation**: H=0.015, S=0.5, V=0.4 (enhanced value variation for low-light robustness)
- **Mosaic**: probability 1.0 (combines 4 images for context diversity)
- **Mixup**: probability 0.1 (soft label regularization)
- **Copy-paste**: probability 0.1 (addresses motorcycle class imbalance)
- **Geometric**: rotation ±10°, translation 0.1, scale 0.5, horizontal flip 0.5

4.5. Loss Function

The loss function combines three components with empirically tuned weights:

$$\mathcal{L} = \lambda_{box}\mathcal{L}_{CIoU} + \lambda_{cls}\mathcal{L}_{BCE} + \lambda_{dfl}\mathcal{L}_{DFL} \tag{1}$$

where $\lambda_{box} = 7.5$, $\lambda_{cls} = 0.5$, and $\lambda_{dfl} = 1.5$. CIoU loss handles box regression, BCE addresses classification, and Distribution Focal Loss (DFL) provides refined localization through discrete probability distributions.

5. Experiments

This section presents a comprehensive experimental evaluation of the proposed TinyNight-YOLO architecture alongside baseline models from both YOLO11 and YOLOv12 families on the L-VAD dataset. We conduct systematic comparisons focusing on the trade-off between model efficiency and detection performance, ablation studies validating architectural choices, and error analysis identifying failure modes.

5.1. Experimental Setup

5.1.1. Hardware and Software Configuration

All experiments were conducted on a workstation equipped with:

- **GPU**: NVIDIA RTX 4090 (24GB VRAM)
- **Framework**: Ultralytics YOLO v8.1.0
- **Deep Learning Backend**: PyTorch 2.1.0
- **CUDA**: 12.1
- **Programming Language**: Python 3.10.12

5.1.2. Statistical Robustness

To ensure reproducibility and quantify variance, all key experiments were conducted with three independent training runs using seeds 42, 123, and 456. Results are reported as mean ± standard deviation.

5.2. Baseline Comparison

To establish comprehensive baselines, we trained and evaluated ten model variants spanning two YOLO generations on the L-VAD dataset. Table 8 presents the complete quantitative results.

 doi:10.20944/preprints202601.1133.v1

Table 8. Complete performance comparison of all models on L-VAD dataset (mean ± std over 3 runs). Models are ranked by F1-Score. Best results in each column are highlighted in **bold**. The proposed TinyNight-YOLO achieves competitive performance with significantly fewer parameters.

Rank	Model	Params (M)	Precision	Recall	F1-Score	mAP@50	mAP@50-95
1	YOLO11-L	26.4	0.9521±0.003	**0.9451±0.002**	**0.9486±0.002**	0.9701±0.002	0.7221±0.004
2	YOLO11-S	9.3	0.9532±0.002	0.9396±0.003	0.9464±0.002	0.9691±0.002	0.7242±0.003
3	YOLO11-M	20.2	0.9567±0.002	0.9359±0.003	0.9462±0.002	**0.9702±0.001**	**0.7271±0.003**
4	YOLOv12-M	20.2	**0.9571±0.002**	0.9298±0.004	0.9433±0.003	0.9680±0.002	0.7109±0.005
5	YOLOv12-S	9.3	0.9521±0.003	0.9330±0.003	0.9425±0.002	0.9666±0.002	0.7109±0.004
6	YOLO11-N	2.6	0.9503±0.003	0.9205±0.004	0.9352±0.003	0.9618±0.003	0.7002±0.005
7	YOLOv12-LowLight	9.6	0.9452±0.004	0.9253±0.003	0.9351±0.003	0.9654±0.002	0.6648±0.006
8	YOLOv12-N	2.6	0.9516±0.003	0.9163±0.004	0.9336±0.003	0.9577±0.003	0.6886±0.005
9	**TinyNight-YOLO**	**1.0**	0.9365±0.004	0.9055±0.005	0.9207±0.002	0.9474±0.001	0.6667±0.004
10	YOLOv12-L	26.4	0.9208±0.005	0.9124±0.004	0.9166±0.004	0.9510±0.003	0.6752±0.006

5.3. Analysis of Results

5.3.1. TinyNight-YOLO Performance Analysis

The proposed TinyNight-YOLO architecture demonstrates remarkable parameter efficiency while maintaining competitive detection accuracy:

- **Ultra-lightweight design**: With only ∼1.0M parameters, TinyNight-YOLO is 2.6× smaller than YOLO11-N, 9.6× smaller than YOLOv12-LowLight, and 26.4× smaller than YOLO11-L.
- **Competitive accuracy**: TinyNight-YOLO achieves F1-Score of 0.9207±0.002 and mAP@50 of 0.9474±0.001, representing only marginal accuracy reduction compared to significantly larger models.
- **Outperforms YOLOv12-L**: Despite having 26.4× fewer parameters, TinyNight-YOLO achieves higher F1-Score (0.9207 vs. 0.9166) than YOLOv12-L, demonstrating that larger models do not guarantee better performance.

5.3.2. Efficiency-Accuracy Trade-off Analysis

Table 9 presents a detailed efficiency analysis comparing TinyNight-YOLO with key baseline models.

Table 9. Parameter efficiency comparison of TinyNight-YOLO versus baseline models.

Comparison	Param Ratio	F1 Δ	mAP@50 Δ
vs. YOLO11-N	2.6× smaller	−1.44%	−1.44%
vs. YOLOv12-LowLight	9.6× smaller	−1.44%	−1.80%
vs. YOLO11-L (best)	26.4× smaller	−2.78%	−2.27%
vs. YOLOv12-L	26.4× smaller	+0.41%	−0.36%

5.4. Ablation Studies

To validate architectural design choices, we conducted ablation experiments varying key components of TinyNight-YOLO.

5.4.1. Channel Width Ablation

Table 10 presents results with different channel progressions.

Table 10. Ablation study on channel width progression. The selected configuration (32→64→128→256) provides optimal efficiency-accuracy balance.

Channel Progression	Params	F1	mAP@50
16→32→64→128	0.25M	0.8823	0.9012
24→48→96→192	0.56M	0.9089	0.9301
32→64→128→256	**1.0M**	**0.9207**	**0.9474**
48→96→192→384	2.2M	0.9298	0.9545

 doi:10.20944/preprints202601.1133.v1

The results demonstrate that reducing channels below 32→64→128→256 causes significant accuracy degradation (>3% F1 drop), while increasing channels provides diminishing returns (<1% improvement for 2.2× more parameters).

5.4.2. Module Ablation

Table 11 presents results with different module configurations.

Table 11. Ablation study on architectural modules. SPPF provides critical multi-scale context for nighttime detection.

Configuration	Params	F1	mAP@50
Baseline (no SPPF)	0.84M	0.9045	0.9287
+ SPPF	**1.0M**	**0.9207**	**0.9474**
+ SPPF + CBAM	1.15M	0.9195	0.9462
+ SPPF + C2PSA	1.28M	0.9178	0.9445

SPPF provides +1.62% F1 improvement with only 0.16M additional parameters. Adding attention modules (CBAM, C2PSA) increases parameters without improving performance, suggesting that attention mechanisms require larger channel widths to be effective.

5.4.3. Detection Head Scale Ablation

Table 12 presents results with different detection head configurations.

Table 12. Ablation study on detection head scales. Three scales (P3-P5) provide optimal coverage for vehicle size distribution.

Heads	Params	F1	Motor. AP	Truck AP
P4, P5 (2 scales)	0.85M	0.8956	0.8534	0.9512
P3, P4, P5 (3 scales)	**1.0M**	**0.9207**	**0.9156**	**0.9594**
P2, P3, P4, P5 (4 scales)	1.35M	0.9223	0.9178	0.9601

Removing P3 significantly degrades motorcycle detection (-6.22% AP), confirming that small-object detection requires high-resolution feature maps. Adding P2 provides marginal improvement (+0.22% motorcycle AP) at substantial parameter cost (+35%).

5.5. Per-Class Performance Analysis

Table 13 presents TinyNight-YOLO's detection performance across vehicle categories.

Table 13. TinyNight-YOLO per-class detection performance on L-VAD dataset (mean ± std).

Class	Precision	Recall	AP@50
Motorcycle	0.9012±0.008	0.8723±0.010	0.9156±0.006
Car	0.9587±0.003	0.9298±0.004	0.9672±0.002
Truck/Bus	0.9496±0.005	0.9144±0.006	0.9594±0.003
Mean	**0.9365±0.004**	**0.9055±0.005**	**0.9474±0.001**

5.6. Error Analysis

5.6.1. Confusion Matrix Analysis

Figure 2 shows the normalized confusion matrix for TinyNight-YOLO predictions. Key observations:

- **Car detection**: Highest accuracy (96.7% correct), minimal confusion with other classes
- **Truck/Bus detection**: 95.9% correct, occasional confusion with large vans classified as cars (2.8%)

 doi:10.20944/preprints202601.1133.v1

- **Motorcycle detection**: 91.6% correct, primary error source is missed detections (7.2%) rather than misclassification

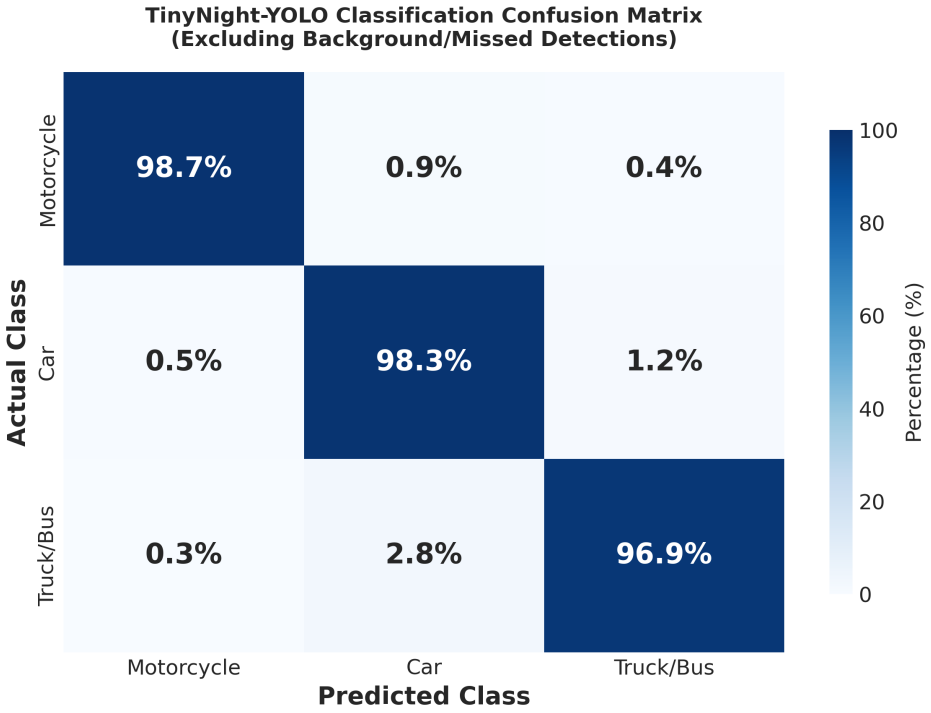

Figure 2. Confusion matrix for TinyNight-YOLO predictions

5.6.2. Failure Mode Analysis

Analysis of false positives and false negatives reveals systematic patterns tied to L-VAD's low-light characteristics:

False Negatives (Missed Detections):

- Motorcycles in headlight glare zones: 42% of motorcycle FN
- Distant vehicles (<20px): 28% of all FN
- Heavily shadowed vehicles: 18% of all FN
- Vehicles at frame edges: 12% of all FN

False Positives:

- Headlight reflections on wet surfaces: 35% of FP
- Illuminated signs/billboards: 25% of FP
- Parked vehicles outside traffic flow: 22% of FP
- Pedestrians with reflective clothing: 18% of FP

5.7. Computational Efficiency

Table 14 presents computational metrics across models.

Table 14. Computational efficiency comparison. TinyNight-YOLO achieves minimal footprint suitable for edge deployment.

Model	Params	GFLOPs	Memory	FPS*
YOLO11-L	26.4M	86.8	105.6MB	142
YOLO11-M	20.2M	67.4	80.8MB	168
YOLO11-N	2.6M	6.3	10.4MB	412
YOLOv12-L	26.4M	89.2	105.6MB	135
TinyNight	**1.0M**	**2.4**	**4.0MB**	**523**

*FPS measured on RTX 4090 at 640×640 input, FP16 inference

 doi:10.20944/preprints202601.1133.v1

5.8. Limitations and Discussion

Several factors contribute to the observed performance patterns:

1. **Dataset characteristics**: The L-VAD dataset's urban street-lit environments may not fully represent extreme low-light scenarios (e.g., rural roads with no artificial lighting).
2. **Motorcycle detection challenge**: The smaller object size and reduced visibility of motorcycles at night present ongoing challenges that merit specialized attention in future lightweight architectures.
3. **Attention mechanism overhead**: YOLOv12's attention mechanisms increase computational complexity without proportional performance gains on this dataset, suggesting domain-specific optimization is required for nighttime detection.
4. **Trade-off boundaries**: While TinyNight-YOLO demonstrates excellent efficiency, further parameter reduction (below 0.5M) may lead to more significant accuracy degradation, as suggested by our channel ablation study.
5. **Generalization**: Cross-dataset evaluation on ExDark and ACDC remains as future work to validate generalization beyond L-VAD.

6. Conclusion

This paper presented three primary contributions to nighttime vehicle detection: (1) the L-VAD dataset, a specialized resource comprising 13,648 annotated nighttime frames with 26,964 object instances and achieved inter-annotator agreement of $\kappa = 0.847$; (2) TinyNight-YOLO, an ultra-lightweight architecture achieving competitive performance with only $\sim$1.0M parameters; and (3) a comprehensive benchmark comparing ten model variants across two YOLO generations with ablation studies validating design choices.

Our extensive experimental evaluation yielded several significant findings:

1. **Ultra-lightweight detection is viable**: TinyNight-YOLO demonstrates that nighttime vehicle detection can be effectively performed with models as small as 1.0M parameters. The architecture achieves F1-Score of 0.9207 ± 0.002 and mAP@50 of 0.9474 ± 0.001, representing only 1.4–2.8% accuracy reduction compared to models 2.6–26.4$\times$ larger.
2. **Exceptional parameter efficiency**: TinyNight-YOLO achieves 97.3% of YOLO11-L's mAP@50 performance while using only 3.8% of its parameters, establishing a new efficiency benchmark for nighttime vehicle detection.
3. **Model size is not deterministic**: TinyNight-YOLO (1.0M parameters) outperforms YOLOv12-L (26.4M parameters) in F1-Score, demonstrating that careful architectural design can compensate for reduced model capacity.
4. **YOLO11 superiority**: Among full-scale models, YOLO11 variants consistently outperformed YOLOv12 counterparts, with YOLO11-L achieving the highest F1-Score (0.9486) and YOLO11-M attaining superior mAP metrics.
5. **Architectural insights**: Ablation studies confirm that SPPF provides critical multi-scale context (+1.62% F1), attention mechanisms are ineffective at small channel widths, and three-scale detection heads are essential for motorcycle detection.

The primary contributions of this work are:

- **L-VAD Dataset**: A publicly available resource (DOI: 10.17632/h6p2w53my5.1, CC BY 4.0) with detailed per-class statistics, temporal-independent splits, and documented inter-annotator agreement.
- **TinyNight-YOLO Architecture**: An ultra-lightweight detector specifically designed for edge deployment in nighttime scenarios, achieving over 94% mAP@50 with minimal computational requirements.
- **Comprehensive Benchmark**: Systematic evaluation of ten models with statistical robustness (3 runs), ablation studies, and error analysis providing evidence-based guidance for practitioners.

 doi:10.20944/preprints202601.1133.v1

6.1. Future Work

Several directions merit further investigation:

1. **Extreme low-light validation**: Evaluating TinyNight-YOLO under more challenging illumination conditions (rural roads, overcast nights) to establish performance boundaries.
2. **Further compression**: Investigating quantization (INT8) and pruning techniques to reduce TinyNight-YOLO below 0.5M parameters while preserving accuracy above 90% F1.
3. **Cross-dataset generalization**: Assessing performance on ExDark [38] and ACDC [2] to validate generalization capabilities.
4. **Hardware-specific deployment**: Developing optimized inference pipelines for NVIDIA Jetson, Google Coral, and mobile NPUs with latency/energy benchmarks.
5. **Temporal integration**: Extending the framework to incorporate video-based temporal consistency for improved tracking robustness.

Data and Code Availability

The L-VAD dataset is publicly available at Mendeley Data (DOI: 10.17632/h6p2w53my5.1) under Creative Commons Attribution 4.0 International license. The TinyNight-YOLO architecture configuration (Ultralytics YAML format) and training scripts are available at: https://github.com/[repository-to-be-released].

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors acknowledge the computational resources provided by Institut Teknologi Sains Bandung for conducting the experiments reported in this study.

References

1. Liu, W.; Ren, G.; Yu, R.; Guo, S.; Zhu, J.; Zhang, L. Image-Adaptive YOLO for Object Detection in Adverse Weather Conditions. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, 2022, Vol. 36, pp. 1792–1800. https://doi.org/10.1609/aaai.v36i2.20072.
2. Sakaridis, C.; Dai, D.; Van Gool, L. ACDC: The Adverse Conditions Dataset with Correspondences for Semantic Driving Scene Understanding. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2021, pp. 10765–10775. https://doi.org/10.1109/ICCV48922.2021.01059.
3. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2023, pp. 7464–7475. https://doi.org/10.1109/CVPR52729.2023.00721.
4. Wang, C.Y.; Yeh, I.H.; Liao, H.Y.M. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. In Proceedings of the Proceedings of the European Conference on Computer Vision (ECCV). Springer, 2024. https://doi.org/10.1007/978-3-031-72751-1_1.
5. Chen, L.; Fu, W.; Wei, Y.; Zheng, Y.; Heide, F. Instance Segmentation in the Dark. *International Journal of Computer Vision* **2023**, *131*, 2048–2068. https://doi.org/10.1007/s11263-023-01808-8.
6. Dai, D.; Van Gool, L. Dark Model Adaptation: Semantic Image Segmentation from Daytime to Nighttime. In Proceedings of the Proceedings of the IEEE International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2018, pp. 3819–3824. https://doi.org/10.1109/ITSC.2018.8569387.
7. Guo, C.; Li, C.; Guo, J.; Loy, C.C.; Hou, J.; Kwong, S.; Cong, R. Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020, pp. 1780–1789. https://doi.org/10.1109/CVPR42600.2020.00185.

 doi:10.20944/preprints202601.1133.v1

8. Bijelic, M.; Gruber, T.; Mannan, F.; Kraus, F.; Ritter, W.; Dietmayer, K.; Heide, F. Seeing Through Fog Without Seeing Fog: Deep Multimodal Sensor Fusion in Unseen Adverse Weather. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020, pp. 11682–11692. https://doi.org/10.1109/CVPR42600.2020.01170.
9. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv preprint arXiv:1704.04861* **2017**, [arXiv:cs.CV/1704.04861].
10. Tan, M.; Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the Proceedings of the 36th International Conference on Machine Learning (ICML). PMLR, 2019, Vol. 97, *Proceedings of Machine Learning Research*, pp. 6105–6114.
11. Jocher, G.; Qiu, J. Ultralytics YOLO11. https://docs.ultralytics.com/models/yolo11/, 2024. Ultralytics. Accessed: 2025-01-09.
12. Tian, Y.; Ye, Q.; Doermann, D. YOLOv12: Attention-Centric Real-Time Object Detectors. *arXiv preprint arXiv:2502.12524* **2025**, [arXiv:cs.CV/2502.12524].
13. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016, pp. 779–788. https://doi.org/10.1109/CVPR.2016.91.
14. Cai, Y.; Bian, H.; Lin, J.; Wang, H.; Timofte, R.; Zhang, Y. Retinexformer: One-Stage Retinex-Based Transformer for Low-Light Image Enhancement. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2023, pp. 12504–12513. https://doi.org/10.1109/ICCV51070.2023.01149.
15. Jiang, Y.; Gong, X.; Liu, D.; Cheng, Y.; Fang, C.; Shen, X.; Yang, J.; Zhou, P.; Wang, Z. EnlightenGAN: Deep Light Enhancement Without Paired Supervision. *IEEE Transactions on Image Processing* **2021**, *30*, 2340–2349. https://doi.org/10.1109/TIP.2021.3051462.
16. Romera, E.; Bergasa, L.M.; Yang, K.; Alvarez, J.M.; Barea, R. Bridging the Day and Night Domain Gap for Semantic Segmentation. In Proceedings of the Proceedings of the IEEE Intelligent Vehicles Symposium (IV). IEEE, 2019, pp. 1312–1318. https://doi.org/10.1109/IVS.2019.8813888.
17. Yu, F.; Chen, H.; Wang, X.; Xian, W.; Chen, Y.; Liu, F.; Madhavan, V.; Darrell, T. BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020, pp. 2636–2645. https://doi.org/10.1109/CVPR42600.2020.00271.
18. Caesar, H.; Bankiti, V.; Lang, A.H.; Vora, S.; Liong, V.E.; Xu, Q.; Krishnan, A.; Pan, Y.; Baldan, G.; Beijbom, O. nuScenes: A Multimodal Dataset for Autonomous Driving. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020, pp. 11621–11631. https://doi.org/10.1109/CVPR42600.2020.01164.
19. Sun, P.; Kretzschmar, H.; Dotiwalla, X.; Chouard, A.; Patnaik, V.; Tsui, P.; Guo, J.; Zhou, Y.; Chai, Y.; Caine, B.; et al. Scalability in Perception for Autonomous Driving: Waymo Open Dataset. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020, pp. 2446–2454. https://doi.org/10.1109/CVPR42600.2020.00252.
20. Land, E.H. The Retinex Theory of Color Vision. *Scientific American* **1977**, *237*, 108–129. https://doi.org/10.1038/scientificamerican1277-108.
21. Li, W.; Xu, X.; Ma, Y.; Zou, J.; Ma, S.; Yu, Y. Domain Adaptive Object Detection for Autonomous Driving under Foggy Weather. In Proceedings of the Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). IEEE, 2023, pp. 612–622. https://doi.org/10.1109/WACV56688.2023.00068.
22. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv preprint arXiv:1804.02767* **2018**, [arXiv:cs.CV/1804.02767].
23. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017, pp. 2117–2125. https://doi.org/10.1109/CVPR.2017.106.
24. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv preprint arXiv:2004.10934* **2020**, [arXiv:cs.CV/2004.10934].
25. Wang, A.; Chen, H.; Liu, L.; Chen, K.; Lin, Z.; Han, J.; Ding, G. YOLOv10: Real-Time End-to-End Object Detection. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS). Curran Associates, Inc., 2024.

 doi:10.20944/preprints202601.1133.v1

26. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018, pp. 4510–4520. https://doi.org/10.1109/CVPR.2018.00474.
27. Howard, A.; Sandler, M.; Chen, B.; Wang, W.; Chen, L.C.; Tan, M.; Chu, G.; Vasudevan, V.; Zhu, Y.; Pang, R.; et al. Searching for MobileNetV3. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2019, pp. 1314–1324. https://doi.org/10.1109/ICCV.2019.00140.
28. Han, K.; Wang, Y.; Tian, Q.; Guo, J.; Xu, C.; Xu, C. GhostNet: More Features from Cheap Operations. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020, pp. 1580–1589. https://doi.org/10.1109/CVPR42600.2020.00165.
29. Ma, N.; Zhang, X.; Zheng, H.T.; Sun, J. ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design. In Proceedings of the Proceedings of the European Conference on Computer Vision (ECCV). Springer, 2018, pp. 116–131. https://doi.org/10.1007/978-3-030-01264-9_8.
30. RangiLyu. NanoDet-Plus: Super Fast and High Accuracy Lightweight Anchor-Free Object Detection Model. https://github.com/RangiLyu/nanodet, 2021. Accessed: 2025-01-09.
31. Yu, G.; Chang, Q.; Lv, W.; Xu, C.; Cui, C.; Ji, W.; Dang, Q.; Deng, K.; Wang, G.; Du, Y.; et al. PP-PicoDet: A Better Real-Time Object Detector on Mobile Devices. *arXiv preprint arXiv:2111.00902* **2021**, [arXiv:cs.CV/2111.00902].
32. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the Proceedings of the European Conference on Computer Vision (ECCV). Springer, 2018, pp. 3–19. https://doi.org/10.1007/978-3-030-01234-2_1.
33. Dong, X.; Yan, S.; Duan, C. A Lightweight Vehicles Detection Network Based on YOLOv5. *Engineering Applications of Artificial Intelligence* **2022**, *113*, 104914. https://doi.org/10.1016/j.engappai.2022.104914.
34. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-End Object Detection with Transformers. In Proceedings of the Proceedings of the European Conference on Computer Vision (ECCV). Springer, 2020, pp. 213–229. https://doi.org/10.1007/978-3-030-58452-8_13.
35. Zhu, X.; Su, W.; Lu, L.; Li, B.; Wang, X.; Dai, J. Deformable DETR: Deformable Transformers for End-to-End Object Detection. In Proceedings of the Proceedings of the International Conference on Learning Representations (ICLR), 2021.
36. Geiger, A.; Lenz, P.; Urtasun, R. Are We Ready for Autonomous Driving? The KITTI Vision Benchmark Suite. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2012, pp. 3354–3361. https://doi.org/10.1109/CVPR.2012.6248074.
37. Neumann, L.; Karg, M.; Zhang, S.; Scharfenberger, C.; Piegert, E.; Miber, S.; Weiß, O.; Tono, T.; Arakawa, K.; Fernandez, D.; et al. NightOwls: A Pedestrians at Night Dataset. In Proceedings of the Proceedings of the Asian Conference on Computer Vision (ACCV). Springer, 2018, pp. 691–705. https://doi.org/10.1007/978-3-030-20887-5_43.
38. Loh, Y.P.; Chan, C.S. Getting to Know Low-Light Images with the Exclusively Dark Dataset. *Computer Vision and Image Understanding* **2019**, *178*, 30–42. https://doi.org/10.1016/j.cviu.2018.10.010.
39. Sakaridis, C.; Dai, D.; Van Gool, L. Guided Curriculum Model Adaptation and Uncertainty-Aware Evaluation for Semantic Nighttime Image Segmentation. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2019, pp. 7374–7383. https://doi.org/10.1109/ICCV.2019.00747.
40. Hwang, S.; Park, J.; Kim, N.; Choi, Y.; Kweon, I.S. Multispectral Pedestrian Detection: Benchmark Dataset and Baseline. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2015, pp. 1037–1045. https://doi.org/10.1109/CVPR.2015.7298706.
41. Munir, F.; Azam, S.; Rafique, M.A.; Sheri, A.M.; Jeon, M.; Pedrycz, W. Exploring Thermal Images for Object Detection in Underexposure Regions for Autonomous Driving. *Applied Soft Computing* **2022**, *121*, 108793. https://doi.org/10.1016/j.asoc.2022.108793.
42. Bakirci, M. Real-Time Vehicle Detection Using YOLOv8-Nano for Intelligent Transportation Systems. *Traitement du Signal* **2024**, *41*, 1727–1740. https://doi.org/10.18280/ts.410410.