International Journal of Informatics and Computation (IJICOM)
Vol. 8, No. 1, 2026
ISSN: 2685-8711, E-ISSN: 2714-5263 DOI: 10.35842/ijicom

Small Object Detection in High-Resolution Images: A Systematic Literature Review

Nidya Sari Rahmawati[1], Achmad Yuyan[2], Chalvina Izumi Amalia[3]

Abstract

Detecting small objects in high-resolution imagery remains challenging due to extreme scale variation, feature degradation during down sampling, and complex background clutter. This study utilized a Systematic Literature Review (SLR) to analyze 55 deep learning studies on small object detection published between 2021 and 2026. The review aimed to identify dominant architectural approaches, methodological improvements, and remaining technical limitations in current detection frameworks. The analysis shows that YOLO-based architectures dominate the research landscape, accounting for 49.1% of the reviewed methods. The results indicate that multi-scale feature fusion and spatial-preservation techniques are essential for detecting objects smaller than 16 × 16 pixels. Methods such as Space-to-Depth down sampling, high-resolution P2 prediction heads, and coarse-to-fine detection strategies consistently improve feature retention and detection performance in high-resolution imagery. The review also finds that Transformer-based and hybrid CNN–Transformer architectures provide stronger contextual modeling in complex scenes; however, their computational complexity limits deployment in real-time edge environments. The findings highlight the need for more computationally efficient architectures and identify emerging directions such as Vision Mamba state-space models, temporal-aware detection using video data, and lightweight model distillation to improve scalability and cross-domain robustness.

Keywords:
Small Object Detection, High-Resolution Images, Object Detection, , Systematic Literature Review

1. Introduction

The rapid advancement of imaging sensors and data acquisition technologies has led to the widespread availability of high-resolution imagery across various application domains, including aerial surveillance, remote sensing, autonomous driving, intelligent transportation systems, and environmental monitoring. High-resolution images provide rich spatial details that enable fine-grained visual analysis; however, they also introduce significant computational and algorithmic challenges for object detection tasks. One of the most critical and persistent challenges in this context is the detection of small objects, where target objects occupy only a limited number of pixels relative to the image size and are often embedded in complex backgrounds.

Small object detection has attracted increasing attention in recent years due to its importance in real-world applications such as Unmanned Aerial Vehicle (UAV) monitoring, traffic sign recognition, maritime search and rescue, pest detection, and underwater exploration [1]–[5]. Unlike general object detection, small object detection is particularly sensitive to scale variation, feature degradation, occlusion, and background clutter. These issues are further exacerbated in high-resolution images, where downsampling operations

Corresponding Author: Nidya Sari Rahmawati (14240038@nusamandiri.ac.id)
1 Nidya Sari Rahmawati, Universitas Nusa Mandiri, 14240038@nusamandiri.ac.id
2 Achmad Yuyan, Universitas Nusa Mandiri, 14240037@nusamandiri.ac.id
3 Chalvina Izumi Amalia, Universitas Nusa Mandiri, 14240039@nusamandiri.ac.id

and deep network hierarchies can cause small object features to be lost during feature extraction and representation learning [6]–[8].

Early object detection frameworks were primarily designed for medium-to-large object scales and struggled to generalize effectively to small objects. With the emergence of deep learning–based detectors, especially single-stage models such as the You Only Look Once (YOLO) family, significant progress has been achieved in terms of detection accuracy and real-time performance. Consequently, a large body of recent research has focused on adapting and extending YOLO-based architectures to improve their sensitivity to small objects through multi-scale feature fusion, refined anchor strategies, and attention mechanisms [9]–[15].

Building on this foundation, recent implementations have further demonstrated the robustness of YOLO frameworks in dynamic and resource-constrained environments. For instance, the study in [16] successfully utilized the advanced YOLOv11 architecture for vehicle theft detection, proving the model's adaptability in recognizing specific targets like license plates within complex scenes. To address dataset challenges such as class imbalance, the research presented in [17] demonstrated that strategic transfer learning and layer freezing on YOLOv5s can significantly enhance detection precision in challenging vehicle datasets. Similarly, the work in [18] highlighted the efficacy of YOLO architectures in real-time robotics applications, further validating the framework's versatility in maintaining stable detection performance under dynamic operating conditions.

In parallel, feature fusion and multi-scale enhancement s have been extensively explored to address the loss of fine-grained information associated with small objects. By integrating low-level spatial details with high-level semantic representations, these methods aim to preserve discriminative features across multiple scales [19]–[23]. More recently, transformer-based and hybrid CNN–Transformer architectures have been introduced to small object detection tasks, leveraging self-attention mechanisms to capture long-range contextual dependencies that are often missing in convolutional-only models [24]–[28]. While such approaches show strong potential in high-resolution imagery, their computational complexity and data requirements remain open challenges.

To address these limitations, this paper presents a Systematic Literature Review (SLR) of recent advances in small object detection for high-resolution images. This study provides a comprehensive synthesis of research published between 2021 and 2026 by developing a structured taxonomy that categorizes small object detection approaches into three main groups: architectural modifications, feature enhancement strategies, and context-aware models. The review also analyzes how state-of-the-art methods address the vanishing feature problem in high-resolution imagery, particularly through the application of multi-scale feature fusion and attention mechanisms that help preserve fine-grained object information. In addition, the study identifies several open research gaps, including the trade-off between computational efficiency and detection accuracy, especially for deployment on edge devices, and highlights emerging research trends such as hybrid CNN–Transformer architectures that aim to improve detection robustness, scalability, and overall performance.

2. Related Works

Research on small object detection has evolved rapidly with the advancement of deep learning–based object detection frameworks and the increasing availability of high-resolution imagery. While modern detectors achieve high accuracy for mid-sized and large objects, detecting small objects remains challenging due to severe scale variation, feature degradation, and background clutter. These issues are particularly prominent in high-resolution images, where small targets occupy only a few pixels relative to the overall image size [1]–[5].

A large portion of existing studies focuses on YOLO-based small object detection, driven by the efficiency and real-time capability of single-stage detectors. Several works have demonstrated that tailored modifications to YOLO architectures can significantly improve detection performance for small objects in aerial and UAV imagery [1], [6], [10]. Common strategies include multi-scale feature pyramids, refined anchor design, and enhanced detection heads to compensate for the loss of fine-grained spatial information during downsampling [9], [12], [13]. Although these approaches are effective under real-time constraints, their performance often depends on the scale distribution and density of objects in the dataset.

In addition to architectural modifications, feature fusion and multi-scale enhancement methods have been widely explored to enhance small object representations. Adaptive feature fusion mechanisms integrate low-level spatial details with high-level semantic information, enabling more robust detection of small and dense targets [3], [21]. Pyramid-based approaches explicitly enhance small-scale features across multiple resolutions, showing improved performance in UAV and remote sensing scenarios [27], [28]. These studies emphasize that preserving spatial detail throughout the feature extraction pipeline is critical for effective small object detection.

More recently, transformer-based and hybrid CNN–Transformer approaches have been introduced to capture global contextual information that is often missing in convolution-only models. Several studies demonstrate that self-attention mechanisms can improve detection robustness in high-resolution and complex environments by modeling long-range dependencies [7], [29], [30]. Despite their promising performance, transformer-based detectors generally incur higher computational costs and require more training data, which limits their applicability in real-time and resource-constrained settings.

Another important research direction ocuses on lightweight and real-time optimized small object detection models, particularly for UAVs, autonomous driving, and edge devices. Lightweight architectures aim to balance detection accuracy and computational efficiency by reducing parameter count and optimizing feature fusion strategies [19], [31], [32]. These models are well-suited for deployment on embedded platforms; however, they often struggle in highly cluttered scenes or when objects are extremely small, highlighting an ongoing trade-off between efficiency and accuracy.

Several works further explore domain- and modality-specific small object detection, where detection strategies are tailored to particular sensing environments. Infrared-based methods exploit thermal characteristics to distinguish small targets from background noise [33], while RGB–event fusion approaches improve robustness in dynamic scenes [34]. Other studies focus on camouflaged or pest object detection by enhancing sensitivity to subtle structural cues [35], [36]. Although these approaches achieve strong domain-specific performance, their generalization across datasets and modalities remains limited.

Finally, comprehensive review studies have summarized existing small object detection and tracking methods, outlining key challenges and research trends [4]. However, such surveys often lack a systematic screening process and do not provide a structured categorization of recent deep learning–based methods specifically targeting high-

resolution imagery.

In contrast, this study presents a systematic literature review that rigorously screens recent works and organizes them into a coherent taxonomy of small object detection approaches for high-resolution images. By explicitly mapping selected studies to methodological categories and analyzing their strengths and limitations, this review provides a unified and reproducible perspective on current research trends and open challenges.

3. Research Method

This study employs a Systematic Literature Review (SLR) methodology to comprehensively analyze the state-of-the-art in small object detection within high-resolution imagery. The SLR approach was selected to provide a transparent, structured, and reproducible framework for synthesizing existing knowledge, minimizing selection bias, and identifying critical research gaps. The research design strictly adheres to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines.

3.1 Research Questions

To guide the systematic review and address the technological challenges identified in the introduction, four specific Research Questions (RQs) were defined:

1. RQ1: What are the dominant deep learning architectures and methodological frameworks currently employed for small object detection?
2. RQ2: How do state-of-the-art methods address the specific challenges of scale variation and feature degradation in high-resolution images?
3. RQ3: What are the trade-offs between detection accuracy and computational efficiency in recent implementations?
4. RQ4: What are the emerging trends, open challenges, and future research directions in this domain?

3.2 Literature Search Strategy

A comprehensive search was conducted across five major scientific databases: Scopus, Web of Science, IEEE Xplore, MDPI, and ScienceDirect. To ensure relevance to the rapidly evolving nature of deep learning, the search was restricted to studies published between January 2021 and January 2026.

The search query was constructed using Boolean operators combining three key concepts: *Target Object*, *Image Domain*, and *Methodology*. The generalized search string used was:

("Small Object Detection" OR "Tiny Object Detection") AND ("High-Resolution" OR "UAV" OR "Aerial Imagery" OR "Remote Sensing") AND ("Deep Learning" OR "YOLO" OR "Transformer" OR "CNN").

3.3 Inclusion and Exclusion Criteria

To ensure the quality and relevance of the reviewed literature, explicit inclusion and exclusion criteria were established before the screening process. These criteria are detailed in Table 1.

Table 1. Inclusion and Exclusion Criteria

Type	Description
Inclusion	(1) Studies explicitly focusing on small object detection; (2) Use of high-resolution datasets (e.g., UAV, satellite, or 4K surveillance); (3) Implementation of deep learning-based methods; (4) Peer-reviewed articles published in English (2021–2026); (5) Availability of quantitative evaluation metrics (e.g., mAP, FPS).
Exclusion	(1) Studies on general object detection without a specific focus on small scales; (2) Use of low-resolution or standard imagery; (3) Traditional image processing methods (non-deep learning); (4) Review papers, abstract-only records, or grey literature; (5) Studies lacking clear experimental results.

3.4 Study Selection and Screening (PRISMA Workflow)

The study selection process followed a three-stage screening protocol: (1) Title Screening; (2) Abstract Screening; and (3) Full-Text Assessment. The initial search across databases yielded a total of 255 records. After removing duplicates and strictly applying the inclusion/exclusion criteria, the dataset was reduced through progressive screening stages to ensure high relevance.

The quantitative attrition of the literature at each step is detailed in Table 2, which summarizes the number of records evaluated, excluded, and retained.

Table 2. Screening Results at Each Stage

Screening Stage	Records Evaluated	Records Excluded	Records Retained
Title Screening	255	110	145
Abstract Screening	145	40	105
Full-Text Assessment	105	50	**55**

To provide a comprehensive visual overview of the entire selection procedure, Fig. 1 illustrates the PRISMA flow diagram. This diagram illustrates the flow of information through the various phases of the systematic review, highlighting the specific points where studies were excluded.

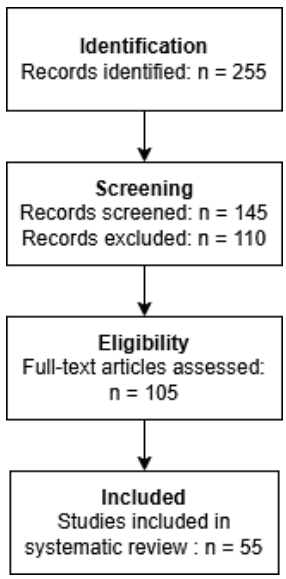

Fig. 1. PRISMA Flow Diagram

3.5 Data Extraction and Qualitative Synthesis

Data extraction was performed on the final set of 55 included studies using a standardized form. The extracted variables included: (1) Bibliometric data (author, year, publisher); (2) Methodological details (backbone architecture, detection heads, attention mechanisms, loss functions); (3) Experimental setup (datasets used, input resolution); and (4) Performance metrics (mAP, FPS, GFLOPs).

Instead of a statistical meta-analysis, which is often infeasible due to varying dataset protocols, this study employs a qualitative synthesis. The selected studies were taxonomically categorized into methodological families, including Anchor-based Detectors (YOLO series), Anchor-free Mechanisms, Transformer-based Architectures, and Feature Fusion Strategies to facilitate a structured comparison of their efficacy in handling small objects.

4. Overview of Included Studies

Since this study follows a Systematic Literature Review (SLR) protocol, this section presents the descriptive characteristics of the 55 selected papers. Unlike experimental research that reports raw data from a specific setup, the findings here summarize the bibliographic and categorical distribution of the reviewed literature to provide a factual landscape of the current research status.

4.1 Distribution of Studies by Publication Year

The temporal distribution of the selected studies reflects the rapidly growing interest in small object detection within high-resolution imagery. All included studies were published between 2021 and 2026, with a notable surge in publications in 2024 and 2025. This upward trend aligns with the rapid evolution of advanced real-time detectors (e.g., the transition from YOLOv8 to the more recent YOLOv10 and YOLOv11 architectures), the exploration of new paradigms like State Space Models, and the increasing accessibility of UAV datasets. Table 3 summarizes the annual distribution.

Table 3. Distribution of Reviewed Studies by Publication Year

Year	Number of Studies
2021	2
2022	10
2023	12
2024	14
2025	16
2026	1
Total	**55**

As shown in Table 3, the data illustrate consistent year-over-year growth in research output, culminating in a peak during 2025. This steady trajectory underscores that small object detection is not a saturated topic, but rather an actively expanding frontier. The sharp, continuous increase from 2022 onwards specifically mirrors the academic community's intense focus on resolving edge-case limitations in standard detection models when deployed for real-world, high-resolution tasks.

4.2 Application Domains

The reviewed studies span diverse application domains, demonstrating the broad applicability of small object detection. As detailed in Table 4, the majority of research focuses on UAV and Aerial Imaging (42%), followed by Remote Sensing (Satellite). This dominance highlights that the primary challenge in this field stems from the high altitude and wide scope of vision inherent in aerial surveillance.

Table 4. Application Domains of Reviewed Studies

Application Domain	Number of Studies
UAV and Aerial Imaging	23
Remote Sensing (Satellite)	15
Autonomous Driving and Transportation	9
Underwater Imaging	4
Infrared / Thermal Imaging	2
Pest and Camouflaged Object Detection	2
Total	**55**

The distribution in Table 4 clearly indicates that aerial and space-borne observations comprising UAVs, aerial imaging, and satellite remote sensing collectively account for nearly 70% of the reviewed literature. This heavy concentration suggests that the most pressing demand for small object detection lies in scenarios characterized by top-down perspectives and extreme camera-to-target distances. Conversely, emerging domains such as underwater and thermal imaging represent niche but critical areas where the challenge of feature degradation is further compounded by severe environmental noise and lack of optical texture.

4.3 Methodological Classification

To facilitate a structured analysis, the reviewed studies were classified into mutually exclusive methodological categories based on their primary detection framework. Each study was assigned to a single category according to its dominant contribution to avoid double counting.

Table 5. Classification of Reviewed Studies by Methodological Approach

Methodological Approach	Paper ID
YOLO-based Detection	[1], [2], [6], [9], [10], [12], [13], [16], [17], [18], [20], [21], [22], [25], [26], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48]
Feature Fusion and Multi-Scale Enhancement	[3], [5], [11], [23], [27], [28], [32], [35], [49], [50], [51]
Transformer-Based, Mamba, and Hybrid Models	[7], [8], [24], [29], [30], [52]
Lightweight and Real-Time Models	[15], [19], [31], [53], [54], [55]
Modality-Specific Detection	[33], [34], [36]
GAN-Assisted Detection	[14]
Review / Survey Studies	[4]

As summarized in Table 5, YOLO-based detection methods represent the largest group (27 papers), reflecting their widespread adoption due to the balance between speed and accuracy. Feature fusion strategies form the second-largest category, emphasizing the critical need for preserving fine-grained spatial features in high-resolution inputs. Transformer-based models, while powerful, account for a smaller portion, likely due to their higher computational demands.

4.4 Datasets Used in Reviewed Studies

The literature utilizes a mix of publicly available benchmarks and domain-specific proprietary datasets. UAV and aerial image datasets (such as VisDrone and UAVDT) remain the most frequently used benchmarks, reflecting the acute need for small object detection in aerial surveillance. Furthermore, the inclusion of custom datasets underscores the increasing adoption of these models for specific real-world applications. Table 6 presents the updated distribution of dataset types across the 55 reviewed studies.

Table 6. Types of Datasets Used in Reviewed Studies

Dataset Type	Number of Studies
UAV / Aerial Image Datasets	22
Remote Sensing Satellite Datasets	16
Traffic and Autonomous Driving Datasets	8
Underwater Imaging Datasets	4
Custom or Proprietary Datasets	3
Infrared / Thermal Datasets	2

The distribution in Table 6 clearly indicates that aerial and space-borne datasets, comprising UAV, aerial imaging, and satellite remote sensing, collectively account for nearly 70% of the reviewed literature. This heavy concentration confirms that the most pressing algorithmic challenges (e.g., extreme camera-to-target distances and dense background clutter) are predominantly found in top-down perspective imagery. Meanwhile, the emergence of custom and niche datasets (such as underwater and thermal imaging) demonstrates that small object detection frameworks are progressively being adapted to handle severe environmental noise and a lack of optical texture in specialized domains.

4.5 Evaluation Metrics Reported

Evaluation practices primarily focus on detection accuracy and inference speed. As shown in Table 7, Mean Average Precision (mAP) is the universal standard, reported in 100% of the selected studies. However, only a subset of studies (approximately 43%) explicitly report Frames Per Second (FPS). This indicates that while real-time performance is heavily discussed as a motivation, especially in YOLO-based and lightweight models, it is not always rigorously quantified across all architectural proposals.

Table 7. Evaluation Metrics Reported in Reviewed Studies

Evaluation Metric	Number of Studies
Mean Average Precision (mAP)	55
Precision	35
Recall	32
Frames Per Second (FPS)	24
F1-score	20

The data presented in Table 7 highlights a significant disparity in evaluation protocols. While mAP serves as the universal gold standard for measuring detection accuracy, enabling fair cross-study comparisons, the lack of consistent speed and efficiency provision is a notable limitation. With fewer than half of the studies quantifying computational speed (FPS), it becomes challenging to objectively verify the real-world deployability of models that claim to be "lightweight" or suitable for edge devices. This discrepancy emphasizes the critical need for future research to adopt a more comprehensive evaluation framework that rigorously measures both precision and operational efficiency.

4.6 Summary of Descriptive Findings

In summary, the descriptive analysis reveals a research landscape strongly dominated by YOLO-based architectures applied to aerial, UAV, and remote sensing imagery. The consistent use of mAP as a primary metric enables a fair baseline comparison of accuracy; however, the variability in datasets and the inconsistent reporting of computational speed (FPS) suggest a pressing need for more standardized benchmarking. These descriptive findings establish the factual foundation for the in-depth analytical discussion regarding scale-handling strategies and efficiency trade-offs presented in the following section.

5. Result and Analysis

This section presents the analytical findings derived from the systematic review. Unlike the descriptive statistics in Section 4, this section critically analyzes the methodological innovations, performance trade-offs, and emerging paradigms found in the selected studies. The analysis interprets these findings in relation to the specific challenges of feature degradation and background clutter discussed in the Introduction.

5.1 Methodological Distribution of Selected Studies (RQ1)

The classification of the 55 eligible studies reveals a clear dominance of single-stage detectors. As shown in Table 8, YOLO-based approaches constitute 49.1% of the literature. This dominance is not merely due to popularity but rather arises from the evolution of the YOLO architecture (from v5 to v11), which has increasingly incorporated features previously reserved for two-stage detectors, such as multi-scale fusion and attention heads.

Table 8. Methodological Distribution of Selected Studies

Method Category	Number of Studies	Percentage (%)
YOLO-Based Detection	27	49.1
Feature Fusion and Multi-Scale Enhancement	11	20.0
Transformer-Based, Mamba and Hybrid Models	6	10.9
Lightweight and Real-Time Models	6	10.9
Modality-Specific Detection	3	5.5
GAN-Assisted Detection	1	1.8
Review / Survey Studies	1	1.8

The data in Table 8 underscores a collective prioritization of real-time processing capabilities, with nearly half of the proposed solutions relying on YOLO variants. Conversely, the combined 30.9% share of Feature Fusion and Transformer/Mamba models indicates a parallel, albeit more computationally intensive, effort to maximize contextual awareness and spatial feature preservation for extremely small targets.

To provide a deeper understanding of these specific methodological contributions and their impact on performance, Table 9 details the proposed frameworks, key innovations, and quantitative findings of the most representative studies in this review.

Table 9. Summary of Key Methods and Findings in Representative Studies (2021–2026)

Paper ID	Core Framework	Key Innovation / Strategy	Dataset	Main Findings / Performance
[1]	ESOD-YOLO (v8)	Added P2 Prediction Head (high-resolution) and EMA attention for tiny object recovery.	VisDrone2019	mAP@0.5: 48.2%
[2]	Temporal-	Exploited temporal context (video	Custom Video	mAP increased from

	YOLOv8	frames) instead of single-image detection to spot moving small objects.		0.465 (baseline) to 0.839
[3]	MMF-YOLO	Proposed an adaptive multi-branch cross-scale feature fusion module with a fusion factor.	VisDrone	Improved detection of overlapping small objects.
[10]	SRM-YOLO (v8)	Utilized SPD-Conv (Space-to-Depth) to prevent feature information loss during downsampling.	VisDrone2019	mAP@0.5: 45.4%
[14]	MTGAN	Generative Adversarial Network (GAN) based super-resolution for ROI enhancement.	COCO / VOC	Significant gain on 'Small' category subset
[16]	YOLOv11	Applied latest YOLOv11 with dynamic anchor assignment for vehicle theft/plate detection.	Custom	Accuracy: 70% (Real-time application in complex scenes)
[19]	EMFE-YOLO	Lightweight model for UAVs using Enhanced Multi-Scale Feature Extraction module.	VisDrone	mAP@0.5: 42.1% (Highly suited for edge devices)
[22]	NATCA-YOLO	Combined Neighborhood Attention Transformer (NAT) with Coordinate Attention.	VisDrone2019	mAP@0.5: 32.7% (Test-dev)
[24]	EfficientV Mamba	First use of State Space Model (Mamba) combined with HRFPN to replace heavy Transformers.	VisDrone2019	mAP@0.5: 37.9% (Competitive with lower compute)
[26]	RHS-YOLOv8	Optimized for Underwater scenes using HGNetv2 backbone and EMA attention.	URPC2020	mAP@0.5: 84.8%, FPS: 158
[29]	MSO-DETR	Hybrid Encoder (CNN + Transformer) tailored for maritime search and rescue (SAR).	SeaDronesSee	mAP@0.5: 56.8%, FPS: 58
[34]	RGB-Event Fusion	Multi-modal: Fused standard RGB frames with Event Camera data to handle motion blur.	DSEC-MOD	mAP@0.5: 54.2% (Robust in high-speed scenes)
[36]	Boundary -Aware	Designed for Camouflaged Object Detection using boundary localization mechanisms.	Complex BG	Improved boundary accuracy in camouflaged scenes.
[44]	YOLOv8 (DAU-YOLO)	Integrated UniRepLKNet (Large Kernel) and DySample for lightweight dynamic upsampling.	VisDrone2019	mAP@0.5: 43.4%, Params: 2.6M (Highly Efficient)
[46]	ECAP-YOLO (v5)	Early work introducing Efficient Channel Attention Pyramid to reduce background noise.	VEDAI	mAP: 76.9%
[48]	BGF-YOLOv10	Incorporated BoTNet multi-head attention and a Patch Expanding Layer for small object contextual upsampling.	VisDrone, UAVDT	Significantly improved mAP while reducing parameters using GhostConv.
[51]	CAMS-AI	Proposed a coarse-to-fine framework utilizing RPN and DBSCAN clustering to focus on dense tiny object regions.	Custom/VHR	High efficiency in extremely high-resolution (4K+) remote sensing images.
[52]	Ghostformer	Two-stage Hybrid Transformer using GhostNet backbone to reduce computational complexity.	COCO 2017	Higher AP on small objects compared to standard DETR

As demonstrated in Table 9, the technological trajectory of small object detection is rapidly advancing beyond standard convolutional networks. The integration of specialized modules, such as Space-to-Depth (SPD-Conv) for feature conservation, State Space

Models (Mamba) for linear-complexity global receptive fields, and coarse-to-fine clustering frameworks, evidences a clear transition toward highly customized, context-aware detectors. Furthermore, the recurrent use of the VisDrone dataset across these leading studies highlights its status as the de facto benchmark for evaluating small object detection under severe scale variation and background clutter.

5.2 Analysis of Scale-Handling Strategies (RQ2)

The core challenge in high-resolution imagery is "feature degradation," where small objects (fewer than 16 × 16 pixels) lose their semantic information after multiple downsampling operations. The analysis of RQ2, summarized in Table 10, indicates that researchers have moved beyond simple Feature Pyramid Networks (FPN).

Table 10. Strategies for Handling Scale Variation (Summary)

Strategy	Number of Studies
Multi-Scale Feature Pyramids	29
Feature Fusion and Enhancement	24
Attention Mechanisms	18
Context Modeling (Global / Long-Range)	9
Data-Level Augmentation	6

Critical analysis of these strategies reveals two effective trends:

1. Feature Conservation: Recent studies, such as [10], utilize SPD-Conv (Space-to-Depth) layers instead of strided convolutions. This technique maps spatial information to the channel dimension, preventing the loss of fine-grained details that typically occur during downsampling.
2. High-Resolution Heads: Standard YOLO detects at P3, P4, and P5 scales. However, the architecture proposed in [1] demonstrated that adding a P2 prediction head (detecting at a larger scale) significantly improves recall for tiny objects, albeit with a slight increase in computational cost.

5.3 Accuracy vs. Efficiency Trade-offs (RQ3)

An important analytical outcome detailed in Table 11 is the identification of trade-offs between detection accuracy and computational efficiency.

Table 11. Optimization Objectives Reported in Reviewed Studies.

Optimization Focus	Number of Studies
Accuracy-Oriented Optimization	31
Efficiency-Oriented Optimization	14
Balanced Accuracy and Efficiency	5

Our analysis highlights a divergence in design philosophy:

1. Transformer-Based Models: Approaches such as the MSO-DETR architecture in [29] achieve superior accuracy in complex maritime environments by capturing global context. However, they often suffer from lower FPS, making them less suitable for on-board UAV processing.
2. Lightweight YOLO: Conversely, works such as [44] utilize Dynamic Upsampling and Large Kernel (UniRepLKNet) within a DAU-YOLO framework to expand the receptive field without the heavy computational burden of Transformers, achieving real-time speeds (>50 FPS) with competitive accuracy.

5.4 Application-Specific Analytical Findings (RQ4)

The breakdown across application domains, presented in Table 12, demonstrates that methodological choices are heavily dictated by specific environmental constraints and hardware limitations.

Table 12. Application Domains and Dominant Method Categories

Application Domain	Dominant Method Category
UAV and Aerial Imaging	YOLO-Based, Lightweight Models
Remote Sensing (Satellite)	Feature Fusion, Transformer-Based
Autonomous Driving	Lightweight YOLO Variants
Underwater Imaging	Transformer-Based, Modality-Specific
Infrared / Thermal Imaging	Modality-Specific Detection

As illustrated in Table 12, the operational environment directly dictates the architectural priorities of the detection models. For instance, studies focusing on Infrared/Thermal imagery, such as the architecture proposed in [33], prioritize background suppression and modality-specific designs because thermal images inherently lack spatial texture and fine color cues. Similarly, underwater detection tasks frequently rely on hybrid or Transformer-based models to counteract severe light scattering and color distortion.

Conversely, Aerial and UAV applications are predominantly driven by Lightweight YOLO variants; this trend reflects the strict necessity to optimize inference speed and minimize parameter counts due to the limited battery capacity and on-board computing power of drones. Furthermore, the strong reliance on Feature Fusion and coarse-to-fine frameworks in Satellite Remote Sensing highlights the unique algorithmic challenge of processing extremely high-resolution (e.g., 4K+) imagery, where target objects appear extraordinarily small against vast, complex geographic backgrounds.

5.5 Emerging Trends and Analytical Implications

The PRISMA-based analysis reveals an important shift in the 2024–2025 literature that extends beyond conventional CNN and Transformer architectures. Recent studies highlight the emergence of State Space Model (SSM)–based approaches, particularly Vision Mamba architectures, as a promising direction for small object detection. The study in [24] demonstrates that the EfficientVMamba model captures long-range dependencies similar to Transformer-based models while maintaining linear computational complexity, making it more suitable for edge-based applications. In addition, another emerging trend involves the utilization of temporal information in detection frameworks. While most existing approaches rely on single-image analysis, the study in [2] shows that Temporal-YOLO, which analyzes video sequences, improves the detection of small dynamic objects by leveraging motion consistency across frames. This temporal context helps address the static occlusion problem and enhances detection reliability in dynamic environments.

6. Conclusion

Detecting small objects in high-resolution imagery remains a significant challenge due to extreme scale variation, feature degradation during downsampling, and complex background clutter. To address this problem, we utilize a Systematic Literature Review (SLR) approach to systematically examine recent advances in deep learning–based small object detection. This study analyzes 55 state-of-the-art studies published between 2021 and 2026 to identify dominant architectural trends, performance characteristics, and technical limitations. The synthesis of the reviewed literature provides a comprehensive understanding of how modern detection frameworks attempt to preserve fine-grained spatial information and improve detection reliability in high-resolution visual environments.

The analysis reveals that YOLO-based architectures dominate the current research landscape, representing approximately 49.1% of the reviewed solutions. This study obtains consistent evidence that multi-scale feature fusion and spatial-preservation mechanisms play a crucial role in improving detection accuracy for objects smaller than 16 × 16 pixels. Techniques such as Space-to-Depth downsampling, high-resolution P2 prediction layers, and coarse-to-fine detection strategies significantly improve feature retention during the early stages of convolutional processing. In addition, this finding shows that hybrid detection pipelines combining feature pyramid networks, attention modules, and lightweight architectural adjustments achieve strong performance in aerial and surveillance scenarios. However, the review also identifies a key limitation in the current literature, namely the strong dependence on a limited set of benchmark datasets such as VisDrone, which restricts the generalizability of many proposed methods across different application domains.

This study also evaluates the growing adoption of Transformer-based architectures for contextual reasoning in complex visual scenes. The reviewed studies demonstrate that Transformer and hybrid CNN–Transformer models provide superior global context modeling, particularly in environments with dense clutter or occlusion. Nevertheless, this study finds that the computational complexity of pure Transformer models remains a major barrier for deployment in real-time edge environments such as unmanned aerial vehicles and autonomous systems. These findings show that future research should prioritize computationally efficient architectures, including emerging state-space models such as Vision Mamba, temporal-aware detection systems that utilize video information, and knowledge distillation from large foundation models into lightweight detectors. Such directions may significantly improve the scalability, adaptability, and cross-domain robustness of small object detection systems in real-world high-resolution applications.

Acknowledgment

The author would like to express sincere appreciation to the academic supervisors and lecturers at Nusa Mandiri University for their guidance and valuable feedback throughout this research. Special thanks are also extended to peers who contributed ideas during discussions, and to family members for their continued support and encouragement. All individuals acknowledged have agreed to be mentioned in this section.

References

[1] J. Luo, Z. Liu, Y. Wang, A. Tang, H. Zuo, and P. Han, “Efficient Small Object Detection You Only Look Once: A Small Object Detection Algorithm for Aerial Images,” *Sensors*, vol. 24, no. 21, 2024, doi: 10.3390/s24217067.

[2] M. C. van Leeuwen, E. P. Fokkinga, W. Huizinga, J. Baan, and F. G. Heslinga, “Toward Versatile Small Object Detection with Temporal-YOLOv8,” *Sensors*, vol. 24, no. 22, 2024, doi: 10.3390/s24227387.

[3] Q. Zhang, H. Zhang, and X. Lu, “Adaptive Feature Fusion for Small Object Detection,” *Appl. Sci.*, vol. 12, no. 22, pp. 1–20, 2022, doi: 10.3390/app122211854.

[4] B. Mirzaei, H. Nezamabadi-pour, A. Raoof, and R. Derakhshani, “Small Object Detection and Tracking: A Comprehensive Review,” *Sensors*, vol. 23, no. 15, 2023, doi: 10.3390/s23156887.

[5] S. H. Kang and J. S. Park, “Aligned Matching: Improving Small Object Detection in SSD,” *Sensors*, vol. 23, no. 5, 2023, doi: 10.3390/s23052589.

[6] Z. Wang *et al.*, “Improved Small Object Detection Algorithm CRL-YOLOv5,” *Sensors*, vol. 24, no. 19, pp. 1–12, 2024, doi: 10.3390/s24196437.

[7] X. Xu, H. Zhang, Y. Ma, K. Liu, H. Bao, and X. Qian, “TranSDet: Toward Effective Transfer Learning for Small-Object Detection,” *Remote Sens.*, pp. 1–21, 2023.

[8] M. Madan and C. Reich, “Strengthening Small Object Detection in Adapted RT-DETR Through Robust Enhancements,” *Electron.*, vol. 14, no. 19, 2025, doi: 10.3390/electronics14193830.

[9] Z. Lin, W. Chen, L. Su, Y. Chen, and T. Li, “HS-YOLO: Small Object Detection for Power Operation Scenarios,” *Appl. Sci.*, vol. 13, no. 19, 2023, doi: 10.3390/app131911114.

[10] B. Yao *et al.*, “SRM-YOLO for Small Object Detection in Remote Sensing Images,” *Remote Sens.*, vol. 17, no. 12, 2025, doi: 10.3390/rs17122099.

[11] X. Xiao, X. Xue, Z. Zhao, and Y. Fan, “A Recursive Prediction-Based Feature Enhancement for Small Object Detection,” *Sensors*, vol. 24, no. 12, 2024, doi: 10.3390/s24123856.

[12] M. Li *et al.*, “MST-YOLO: Small Object Detection Model for Autonomous Driving,” *Sensors*, vol. 24, no. 22, 2024, doi: 10.3390/s24227347.

[13] X. Chen, L. Deng, C. Hu, T. Xie, and C. Wang, “Dense Small Object Detection Based on an Improved YOLOv7 Model,” *Appl. Sci.*, vol. 14, no. 17, pp. 1–18, 2024, doi: 10.3390/app14177665.

[14] H. Wang, J. Wang, K. Bai, and Y. Sun, “Centered Multi-Task Generative Adversarial Network for Small Object Detection,” *Sensors*, vol. 21, no. 15, 2021, doi: 10.3390/s21155194.

[15] H. Zhou, A. Ma, Y. Niu, and Z. Ma, “Small-Object Detection for UAV-Based Images Using a Distance Metric Method,” *Drones*, vol. 6, no. 10, pp. 1–19, 2022, doi: 10.3390/drones6100308.

[16] B. A. Wisesa, M. H. Wathan, E. Faristasari, S. Andreanto, and J. Duli, “Vehicle Theft Detection Using YOLO Based on License Plates and Vehicle Ownership,” *Int. J. Informatics Comput.*, vol. 7, no. 1, 2025, doi: 10.35842/ijicom.

[17] A. N. Y. Rafi and M. Yusuf, “Improving Vehicle Detection in Challenging Datasets : YOLOv5s and Frozen Layers Analysis,” *Int. J. Informatics Comput.*, vol. 5, no. 2, 2023, doi: 10.35842/ijicom.

[18] A. Khumaidi, M. A. Yaqin, R. Y. Aditya, and S. M. Irsyad, “Object Recognition in Robosoccer on Wheeled Using,” *Int. J. Informatics Comput.*, vol. 7, no. 2, 2025, doi: 10.35842/ijicom.

[19] C. Yang, Y. Shen, and L. Wang, “EMFE-YOLO: A Lightweight Small Object Detection Model for UAVs,” *Sensors*, vol. 25, no. 16, 2025, doi: 10.3390/s25165200.

[20] S. Li, S. Wang, and P. Wang, “A Small Object Detection Algorithm for Traffic Signs Based on Improved YOLOv7,” *Sensors*, vol. 23, no. 16, 2023, doi: 10.3390/s23167145.

[21] F. Zhao, J. Zhang, and G. Zhang, “FFEDet: Fine-Grained Feature Enhancement for Small Object Detection,” *Remote Sens.*, vol. 16, no. 11, 2024, doi: 10.3390/rs16112003.

[22] Y. Zhu, Z. Ai, J. Yan, S. Li, G. Yang, and T. Yu, “NATCA YOLO-Based Small Object Detection for Aerial Images,” *Inf.*, vol. 15, no. 7, 2024, doi: 10.3390/info15070414.

[23] G. Xing, Z. Xu, Y. He, H. Ning, M. Sun, and C. Wang, “ECAN-Detector : An Efficient Context-Aggregation Network for Small-Object Detection,” no. 618, pp. 1–26, 2025.

[24] S. Wu, X. Lu, C. Guo, and H. Guo, “Accurate UAV Small Object Detection Based on HRFPN and EfficentVMamba,” *Sensors*, 2024.

[25] J. Zhou, T. Su, K. Li, and J. Dai, “Small Target-YOLOv5: Enhancing the Algorithm for Small Object Detection in Drone Aerial Imagery Based on YOLOv5,” *Sensors*, vol. 24, no. 1, 2024,

doi: 10.3390/s24010134.
[26] Y. Wei, J. Tao, W. Wu, D. Yuan, and S. Hou, “RHS-YOLOv8: A Lightweight Underwater Small Object Detection Algorithm Based on Improved YOLOv8,” *Appl. Sci.*, vol. 15, no. 7, 2025, doi: 10.3390/app15073778.
[27] J. Sun, H. Gao, X. Wang, and J. Yu, “Scale Enhancement Pyramid Network for Small Object Detection from UAV Images,” *Entropy*, vol. 24, no. 11, pp. 1–18, 2022, doi: 10.3390/e24111699.
[28] Z. Quan and J. Sun, “A Feature-Enhanced Small Object Detection Algorithm Based on Attention Mechanism,” *Sensors*, vol. 25, no. 2, 2025, doi: 10.3390/s25020589.
[29] J. Li, Y. Hua, and M. Xue, “MSO-DETR: A Lightweight Detection Transformer Model for Small Object Detection in Maritime Search and Rescue,” *Electron.*, vol. 14, no. 12, 2025, doi: 10.3390/electronics14122327.
[30] G. Chen, Z. Mao, K. Wang, and J. Shen, “HTDet: A Hybrid Transformer-Based Approach for Underwater Small Object Detection,” *Remote Sens.*, vol. 15, no. 4, pp. 1–22, 2023, doi: 10.3390/rs15041076.
[31] S. Y. Jhong, H. C. Hsu, H. C. Lin, and Y. Y. Chen, “ELNet: An Efficient and Lightweight Framework for Small Object Detection in UAV Images,” *Int. Conf. Adv. Robot. Intell. Syst. ARIS*, pp. 1–26, 2025, doi: 10.1109/ARIS66143.2025.11163440.
[32] H. Wu, Y. Zhu, and L. Wang, “A Dense Small Object Detection Algorithm Based on a Global Normalization Attention Mechanism,” *Appl. Sci.*, vol. 13, no. 21, 2023, doi: 10.3390/app132111760.
[33] X. Xi, J. Wang, F. Li, and D. Li, “IRSDet: Infrared Small-Object Detection Network Based on Sparse-Skip Connection and Guide Maps,” *Electron.*, vol. 11, no. 14, pp. 1–16, 2022, doi: 10.3390/electronics11142154.
[34] Y. Shi, M. Li, N. Chen, Y. Luo, S. He, and W. An, “Sparse-Gated RGB-Event Fusion for Small Object Detection in the Wild,” *Remote Sens.*, vol. 17, no. 17, 2025, doi: 10.3390/rs17173112.
[35] X. Huang, J. Dong, Z. Zhu, D. Ma, F. Ma, and L. Lang, “TSD-Truncated Structurally Aware Distance for Small Pest Object Detection,” *Sensors*, vol. 22, no. 22, pp. 1–14, 2022, doi: 10.3390/s22228691.
[36] G. Zhang, Z. Yang, Y. Wang, Y. Chen, and D. Miao, “Camouflaged Object Detection with boundary localization in complex backgrounds,” *Eng. Appl. Artif. Intell.*, vol. 161, pp. 1–24, 2025, doi: 10.1016/j.engappai.2025.112047.
[37] S. Liu *et al.*, “ERF-RTMDet: An Improved Small Object Detection Method in Remote Sensing Images,” *Remote Sens.*, vol. 15, no. 23, pp. 1–18, 2023, doi: 10.3390/rs15235575.
[38] T. Gao, M. Wushouer, and G. Tuerhong, “DMS-YOLOv5: A Decoupled Multi-Scale YOLOv5 Method for Small Object Detection,” *Appl. Sci.*, vol. 13, no. 10, 2023, doi: 10.3390/app13106124.
[39] S. Bin Amir and K. Horio, “YOLOv8s-NE: Enhancing Object Detection of Small Objects in Nursery Environments Based on Improved YOLOv8,” *Electron.*, vol. 13, no. 16, pp. 1–16, 2024, doi: 10.3390/electronics13163293.
[40] B. Yan, J. Li, Z. Yang, X. Zhang, and X. Hao, “AIE-YOLO: Auxiliary Information Enhanced YOLO for Small Object Detection,” *Sensors*, vol. 22, no. 21, 2022, doi: 10.3390/s22218221.
[41] H. Lou *et al.*, “DC-YOLOv8: Small-Size Object Detection Algorithm Based on Camera Sensor,” *Electron.*, vol. 12, no. 10, pp. 1–14, 2023, doi: 10.3390/electronics12102323.
[42] J. Ni, S. Zhu, G. Tang, C. Ke, and T. Wang, “A Small-Object Detection Model Based on Improved YOLOv8s for UAV Image Scenarios,” *Remote Sens.*, vol. 16, no. 13, pp. 1–19, 2024, doi: 10.3390/rs16132465.
[43] Y. Yang, S. Yang, and Q. Chan, “LEAD-YOLO: A Lightweight and Accurate Network for Small Object Detection in Autonomous Driving,” *Sensors*, vol. 25, no. 15, 2025, doi: 10.3390/s25154800.
[44] Z. Wan, Y. Lan, Z. Xu, K. Shang, and F. Zhang, “DAU-YOLO: A Lightweight and Effective Method for Small Object Detection in UAV Images,” *Remote Sens.*, vol. 17, no. 10, 2025, doi: 10.3390/rs17101768.
[45] J. Song, C. Han, and C. Wu, “A Small-Scale Object Detection Algorithm in Intelligent Transportation Scenarios,” *Entropy*, vol. 26, no. 11, 2024, doi: 10.3390/e26110920.
[46] M. Kim, J. Jeong, and S. Kim, “ECAP-YOLO: Efficient Channel Attention Pyramid YOLO for Small Object Detection in Aerial Image,” *Remote Sens.*, vol. 13, no. 23, pp. 1–20, 2021, doi: 10.3390/rs13234851.
[47] X. Zhao *et al.*, “MSUD-YOLO: A Novel Multiscale Small Object Detection Model for UAV Aerial Images,” *Drones*, vol. 9, no. 6, 2025, doi: 10.3390/drones9060429.
[48] J. Mei and W. Zhu, “BGF-YOLOv10 : Small Object Detection Algorithm from,” *Sensors*, 2024.

[49] C. Peng, M. Zhu, H. Ren, and M. Emam, “Small Object Detection Method Based on Weighted Feature Fusion and CSMA Attention Module,” *Electron.*, vol. 11, no. 16, 2022, doi: 10.3390/electronics11162546.
[50] T. Shi *et al.*, “Feature-Enhanced CenterNet for Small Object Detection in Remote Sensing Images,” *Remote Sens.*, vol. 14, no. 21, 2022, doi: 10.3390/rs14215488.
[51] Z. Chen, Z. Chen, B. Yang, Q. Guo, H. Wang, and X. Zeng, “CAMS-AI : A Coarse-to-Fine Framework for Efficient Small Object Detection in High-Resolution Images,” *Remote Sens.*, pp. 1–22, 2026.
[52] S. Li, F. Sultonov, J. Tursunboev, J. H. Park, S. Yun, and J. M. Kang, “Ghostformer: A GhostNet-Based Two-Stage Transformer for Small Object Detection,” *Sensors*, vol. 22, no. 18, pp. 1–9, 2022, doi: 10.3390/s22186939.
[53] Q. Dong, T. Han, G. Wu, B. Qiao, and L. Sun, “RSNet: Compact-Align Detection Head Embedded Lightweight Network for Small Object Detection in Remote Sensing,” *Remote Sens.*, vol. 17, no. 12, pp. 1–27, 2025, doi: 10.3390/rs17121965.
[54] H. Liu, F. Sun, J. Gu, and L. Deng, “SF-YOLOv5: A Lightweight Small Object Detection Algorithm Based on Improved Feature Fusion Mode,” *Sensors*, vol. 22, no. 15, pp. 1–14, 2022, doi: 10.3390/s22155817.
[55] L. Meng, L. Zhou, and Y. Liu, “SODCNN: A Convolutional Neural Network Model for Small Object Detection in Drone-Captured Images,” *Drones*, vol. 7, no. 10, pp. 1–22, 2023, doi: 10.3390/drones7100615.