YOLO Algorithm for Irregular Small Target Detection: A Survey

www.surveyx.cn

Abstract

The YOLO (You Only Look Once) algorithm has revolutionized object detection by offering a unified architecture that excels in real-time processing, making it a state-of-the-art solution for detecting small and irregular targets across diverse domains. This survey paper explores the significance of YOLO in various applications, such as surveillance, UAV technology, and medical imaging, highlighting its adaptability and efficiency. The survey addresses persistent challenges in detecting small and dim targets, particularly in infrared imagery, where factors like low thermal contrast and cluttered backgrounds complicate detection. It evaluates the performance of YOLO and other detection algorithms within existing benchmarks, emphasizing the need for models capable of learning causal features incrementally. Recent advancements, including innovative network architectures and loss functions, have further enhanced YOLO's capabilities. Comparative analyses underscore YOLO's superior speed and competitive accuracy, reinforcing its applicability in real-time scenarios. The survey also identifies future research directions, such as integrating advanced technologies and modalities, exploring new application domains, and developing advanced evaluation metrics. By documenting these advancements and challenges, this survey aims to guide future research and foster innovation necessary to overcome existing limitations in object detection, ensuring YOLO's continued prominence in the field of computer vision.

1 Introduction

1.1 Significance of YOLO in Object Detection

The YOLO (You Only Look Once) algorithm has revolutionized object detection by providing a unified architecture that enables real-time processing with high accuracy, outpacing traditional methods. Variants such as YOLO-LITE reduce computational demands while maintaining detection effectiveness, making them ideal for resource-limited settings [1]. This efficiency is crucial for detecting irregular small targets, which are challenging due to their limited pixel representation and indistinct features [2].

In aerial imagery, the detection of small objects is complicated by scale variance and weak target visibility, challenges that the original YOLO algorithm and its successors aim to address [3]. YOLO's importance is further underscored in Open World Object Detection (OWOD) applications, where it successfully identifies unknown objects, thereby expanding the capabilities of conventional detection frameworks [4].

YOLO's adaptability is evident in infrared small object detection, where it has shown enhanced accuracy in real-time scenarios [5]. This is particularly vital in surveillance and security, where rapid detection of small and dim targets is essential [6]. Challenges in infrared small target detection, such as sensor resolution limits and atmospheric scattering, are areas where YOLO's advancements provide significant benefits [7].

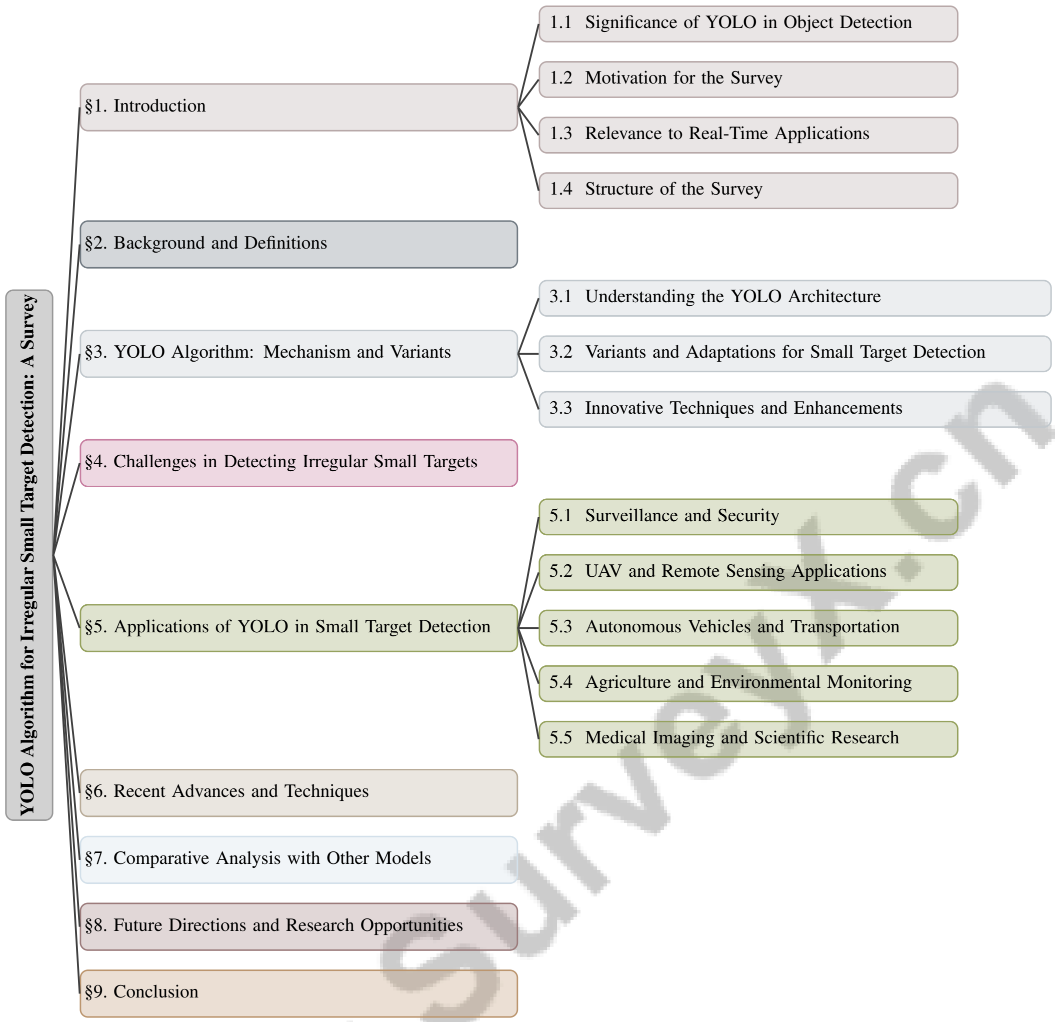


Figure 1: chapter structure

Moreover, YOLO's ability to balance precision and recall is highlighted in specialized tasks like infrared small-target detection, where innovative network structures, such as CourtNet, have been developed to enhance performance [8]. The algorithm's effectiveness across diverse applications emphasizes its critical role in advancing object detection capabilities for irregular small targets.

1.2 Motivation for the Survey

This survey is motivated by ongoing challenges in detecting small and dim targets, particularly in infrared (IR) imagery, where low thermal contrast against cluttered backgrounds complicates accurate detection [6]. This issue is exacerbated in aerial images due to occlusion, scale variations, and blurriness, which significantly hinder detection accuracy [3]. The survey aims to evaluate the performance of various object detection algorithms, including CNN-based methods and YOLO, within existing benchmarks [9].

Traditional object detection models often struggle to adapt to new tasks without compromising performance on previously learned categories, highlighting the need for models capable of incrementally learning causal features [10]. Additionally, the lack of fully annotated training sets poses a significant barrier to effectively deploying deep learning models in object detection tasks [11]. This survey explores advancements in overcoming these limitations, particularly through Few-Shot Object Detection (FSOD) algorithms that manage novel classes with limited labeled samples.

Furthermore, the survey emphasizes the necessity of developing detection methods suitable for complex RF environments, which could enhance capabilities in spectrum sensing and battlefield

AI-generated, for reference only.

command and control [12]. The need for effective detection methods for small infrared targets, often less than 10 pixels in size, remains a critical focus area [7].

By documenting advancements and challenges in the evolution of object detection techniques, this survey traces the transition from traditional methods based on handcrafted features to modern deep learning approaches utilizing sophisticated architectures like convolutional neural networks (CNNs). It reviews benchmark datasets, evaluation metrics, and performance comparisons across various models, underscoring the transformative impact of deep learning and identifying promising future research directions [13, 14, 15]. Additionally, it aims to provide insights into architectural improvements of single-stage detectors, contributing to the development of more effective and efficient detection systems across various sectors. Through this analysis, the survey aspires to guide future research and foster innovation necessary to address existing limitations in object detection.

1.3 Relevance to Real-Time Applications

The proficiency of the YOLO algorithm in real-time applications is evident through its exceptional processing speed and high detection accuracy, making it indispensable across numerous domains. Its integration with class-agnostic detection algorithms significantly enhances its utility in security monitoring and tracking, where immediate responsiveness is essential [16]. In agriculture, the YOLO-Tomato model exemplifies improved detection accuracy for harvesting robots, facilitating efficient operations in challenging environments [17].

In traffic management, YOLO is employed in real-time wrong-way vehicle detection systems, providing timely solutions for traffic control and safety [18]. The SL-YOLO model further demonstrates effectiveness in complex environments such as drone operations, where real-time accuracy is critical [19].

The construction industry benefits from the YOLO-EA model, which maintains real-time processing speeds while achieving high detection accuracy, thereby enhancing safety monitoring on construction sites [20]. During the COVID-19 pandemic, YOLO's application in mask detection provided a state-of-the-art solution balancing speed and accuracy, proving vital for practical implementation [21].

In autonomous driving, YOLO-facilitated image detection significantly influences decision-making and safety performance, underscoring its critical role in vehicle automation [22]. Furthermore, YOLO-LITE's ability to function effectively on devices without GPUs broadens the accessibility of object detection technologies in everyday applications [1].

Comparative analyses highlight YOLO's superior speed and competitive accuracy, reinforcing its applicability in practical real-time scenarios [9]. Additionally, methodologies like DASSF, which improve detection accuracy for small and occluded targets, demonstrate YOLO's relevance in dynamic real-time applications [3].

The ICOD model's emphasis on causal features over data-bias features enhances performance in real-time applications, further illustrating advancements in YOLO's architecture and its integration with various techniques [10]. These developments not only enhance speed and efficiency but also extend applicability across diverse fields, solidifying YOLO's position as a cornerstone in contemporary object detection frameworks.

1.4 Structure of the Survey

This survey is meticulously organized to provide a thorough examination of the YOLO algorithm's application in detecting irregular small targets. It begins with an *Introduction*, establishing the significance of YOLO in object detection, the motivation for the survey, and its relevance to real-time applications. Following this, the *Background and Definitions* section elaborates on essential concepts in computer vision and object detection, defines irregular small targets, and introduces the YOLO algorithm, highlighting its evolution and core features.

The third section, *YOLO Algorithm: Mechanism and Variants*, explores YOLO's architecture, working mechanism, and reviews various versions and adaptations, emphasizing improvements for small target detection. In *Challenges in Detecting Irregular Small Targets*, we identify and discuss challenges such as scale variance, occlusion, and data constraints.

AI-generated, for reference only.

The survey then examines *Applications of YOLO in Small Target Detection*, assessing its effectiveness across diverse domains, including surveillance, UAV technology, and medical imaging. The section titled *Recent Advances and Techniques* provides an overview of the latest research and methodologies aimed at enhancing YOLO's performance, highlighting significant advancements in network architecture and innovative approaches, including the integration of YOLO with UAV technology, which improves real-time object detection and classification across various applications such as engineering, transportation, and agriculture. It also discusses the evolution of YOLO's architecture, particularly advancements in YOLOv5 and YOLOv6, which demonstrate superior accuracy and speed, making them suitable for diverse industrial applications [23, 24, 25, 26, 27].

A *Comparative Analysis with Other Models* evaluates YOLO's performance against other object detection models, highlighting the strengths and weaknesses of each approach. The survey concludes with *Future Directions and Research Opportunities*, emphasizing potential areas for further exploration and innovation within the rapidly evolving field of YOLO-based UAV technology. It underscores the significance of interdisciplinary applications, particularly in agriculture, engineering, and automation, providing a comprehensive summary of key findings that highlight implications for future research and the enhancement of productivity through advanced technological integration [28, 25, 23].The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Key Concepts in Object Detection and Computer Vision

Object detection, a fundamental task in computer vision, involves identifying and localizing objects within images by assigning them to predefined categories and providing spatial coordinates [15]. The methodologies are primarily divided into two frameworks: region proposal-based methods (two-stage detectors) and regression/classification-based methods (single-stage detectors), with YOLO exemplifying the latter by predicting object classes and bounding boxes in a single pass for enhanced efficiency.

In multi-object tracking (MOT) systems, the tracking-by-detection approach is crucial, allowing simultaneous tracking and detection of multiple objects through distinct detection and embedding models, particularly in dynamic environments [29]. Additionally, detecting objects and estimating their poses require advancements in algorithmic approaches and computational efficiency [30].

Online Open World Object Detection (OLOWOD) models mimic human learning by detecting both known and unknown categories while incrementally learning new objects without forgetting previously acquired knowledge, addressing the low recall issues of traditional methods for unknown objects [31, 4]. The high cost of dense annotation, requiring a bounding box and class label for each instance, presents another significant challenge, necessitating methods that function effectively with limited annotated data [32].

These object detection concepts have broad applications, including traffic sign detection and recognition [33] and agricultural object detection for monitoring and automation [28], highlighting the widespread impact of object detection technologies.

2.2 Defining Irregular Small Targets

Irregular small targets, often less than 10 pixels in size, pose significant detection challenges against complex backgrounds [7]. These targets, common in aerial and infrared imaging, suffer from indistinct texture and shape features and are prone to occlusion and scale variance [3]. In aerial imagery, small targets like vehicles or wildlife may be blurred and occluded, complicating accurate classification [6], while infrared imaging further complicates detection due to poor visibility and background noise.

The reliance on data-bias features that do not generalize across tasks hinders detection, necessitating models capable of incremental learning and adaptation to new scenarios [10]. This is especially critical in Few-Shot Object Detection (FSOD) tasks, where limited labeled samples for novel classes pose significant generalization challenges [34]. In Open World Object Detection (OWOD) contexts, the lack of labeled data for unknown objects necessitates probabilistic approaches for identification and learning [4].

Underwater environments, such as coral reefs, present unique challenges in detecting species like the crown of thorns starfish (COTS) due to weak light and low resolution, requiring robust models that function effectively in compromised visibility conditions [35]. Advanced detection models must discern small targets amidst environmental noise and clutter, leveraging both RGB and infrared modalities for improved accuracy, despite potential modality imbalance during inference [36]. Real-time detection frameworks that process data in a streaming manner are essential for applications requiring immediate responsiveness [37].

2.3 Introduction to the YOLO Algorithm

The YOLO (You Only Look Once) algorithm revolutionizes object detection by framing it as a single regression problem, allowing simultaneous prediction of bounding boxes and class probabilities from entire images. This approach significantly enhances processing speed and accuracy, making YOLO effective for real-time applications in dynamic environments like autonomous driving and surveillance [22]. Unlike traditional two-stage detectors, YOLO's single-stage framework facilitates real-time detection, offering substantial practical advantages.

Since its inception, YOLO has evolved through several iterations, each introducing architectural enhancements to boost detection capabilities. YOLO-LITE, for instance, optimizes performance for non-GPU environments by employing a shallow network architecture to achieve approximately 21 frames per second (FPS) [1]. Another variant, YOLO-S, utilizes a lightweight convolutional neural network with a Darknet20 backbone and a single output scale, specifically designed for small target detection in aerial imagery [2].

Further advancements include integrating multi-scale attention mechanisms and feature fusion methods, as demonstrated in models using the InfraTiny dataset, which is rich in small bounding boxes [5]. The Dynamic-Attention Scale-Sequence Fusion (DASSF) method exemplifies enhancements that improve detection accuracy for small targets through dynamic attention mechanisms and scale-sequence feature fusion [3].

YOLO's adaptability is also evident in its application to diverse environments, such as the Probabilistic Objectness Open World Detection Transformer (PROB), which enhances detection capabilities in open-world contexts [4]. This adaptability extends to Few-Shot Object Detection (FSOD) methods, categorized into data-oriented, model-oriented, and algorithm-oriented types, providing a comprehensive understanding of their contributions [34].

YOLO's evolution reflects a commitment to addressing the inherent challenges of object detection. Its innovative architecture and adaptability ensure relevance across various domains, from theoretical research to practical implementations. The algorithm's ability to achieve high detection accuracy while operating efficiently in real-time scenarios, as demonstrated by advancements such as the YOLOv5 model and optimized variants like YOLO-LITE, underscores its significant contribution to the evolution of computer vision technologies. These developments enhance the algorithm's applicability in fields such as security, autonomous vehicles, and medical imaging, allowing it to function effectively on devices with limited computational resources and broadening access to real-time object detection applications [38, 1, 39, 25, 40].

In recent years, the development of object detection algorithms has significantly evolved, with the YOLO (You Only Look Once) framework emerging as a leading approach due to its efficiency and accuracy. To better understand the complexities and innovations within this framework, Figure **??** provides a comprehensive visual representation. This figure illustrates the hierarchical structure of the YOLO algorithm, encompassing its architectural framework, recent advancements, and various adaptations for small target detection. Specifically, the diagram categorizes the YOLO architecture, highlighting its transformation of object detection into a regression problem. It also details recent model enhancements and innovations aimed at improving efficiency in resource-constrained environments. Furthermore, the figure emphasizes the adaptations seen in YOLOv5 and YOLOv7 for small target detection, showcasing advanced features such as GhostNet-based modules. Lastly, it presents innovative techniques, including SpeechYOLO and IoU-Net, along with their applications across diverse domains, thereby enhancing our understanding of the current landscape in object detection technologies.

Figure 2: This figure illustrates the hierarchical structure of the YOLO algorithm, its architectural framework, recent advancements, variants for small target detection, and innovative techniques. The diagram categorizes the YOLO architecture, highlighting its transformation of object detection into a regression problem, recent model enhancements, and innovations for efficiency in resource-constrained environments. It further details the adaptations in YOLOv5 and YOLOv7 for small target detection, emphasizing advanced features like GhostNet-based modules. Lastly, it showcases innovative techniques such as SpeechYOLO and IoU-Net, and their applications in various domains.

3 YOLO Algorithm: Mechanism and Variants

3.1 Understanding the YOLO Architecture

The YOLO (You Only Look Once) architecture is a transformative framework in object detection, conceptualized as a regression problem that simplifies detection and classification by dividing images into grids. Each grid cell predicts bounding boxes and class probabilities, thus optimizing real-time processing and enhancing detection speed [21]. This unified approach reduces computational demands while maintaining high accuracy, making it ideal for applications in autonomous driving and surveillance [1].

Recent advancements have refined YOLO for improved detection of small and occluded objects. The YOLO-EA model incorporates the Efficient Channel Attention (ECA) mechanism and Enhanced Intersection over Union (EIoU) loss to heighten spatial awareness and small object detection [20]. Additionally, the C2f_RFAConv module with Triplet Attention enhances feature extraction in autonomous driving scenarios [22].

YOLO's adaptability is further exemplified by the YOLO-S model, which achieves high detection accuracy through multi-layer feature fusion while maintaining a lightweight structure [2]. The Dynamic-Attention Scale-Sequence Fusion (DASSF) method uses CSPDarknet53 as a backbone, introducing a dynamic head for small object detection [3].

Noteworthy innovations include probabilistic objectness prediction heads in deformable models, as demonstrated by the PROB method [4]. YOLO-LITE exemplifies the framework's efficiency in resource-constrained environments, enabling rapid processing with acceptable accuracy [1].

In underwater target detection, enhancements to YOLOv5 via attention mechanisms and multi-stage architectures have improved feature extraction and detection precision [35]. The patch-wise modality agnostic module facilitates learning a common representation across infrared and visible modalities, reducing inference overhead [36].

The continuous evolution of YOLO, characterized by cutting-edge modules and techniques, underscores its adaptability and relevance in real-time object detection. YOLO consistently enhances detection accuracy and inference speed, making it a preferred choice across industries. Recent advancements, such as YOLOv6, have achieved remarkable metrics, including a 43.5

As shown in Figure 3, the figure illustrates the hierarchical structure of the YOLO architecture advancements, methodologies, and applications. The first example highlights YOLOv6's performance and its quantized variants on the COCO dataset, illustrating quantization's impact on Average Precision (AP) scores and the trade-offs between model size and detection accuracy. The second example demonstrates YOLO's application in video streams, showcasing its ability to process and detect objects in real-time, with a flowchart detailing the algorithm's process from image extraction to non-maximum suppression. These examples reflect YOLO's adaptability and robustness, affirming its status in computer vision applications from image analysis to real-time video processing [24, 27]. The recent advancements, including models like YOLO-EA and YOLO-S, focus on improved detection capabilities, while methodologies such as Efficient Channel Attention and Dynamic-Attention Fusion highlight innovative techniques enhancing YOLO's performance across various domains, including autonomous driving, underwater detection, and real-time video processing, showcasing its versatility and efficiency in diverse environments.

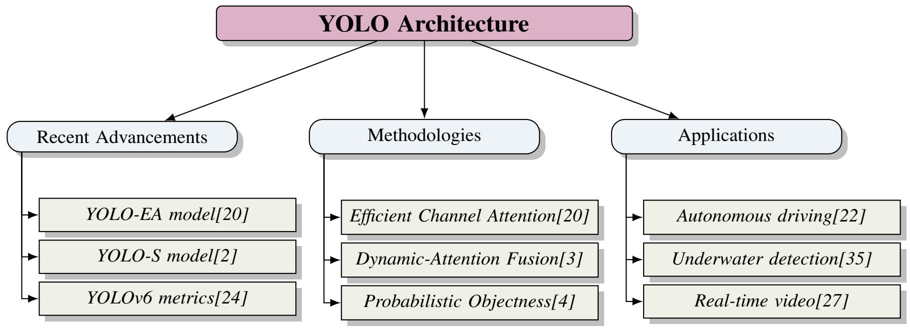


Figure 3: This figure illustrates the hierarchical structure of the YOLO architecture advancements, methodologies, and applications. The recent advancements include models like YOLO-EA and YOLO-S, focusing on improved detection capabilities. Methodologies such as Efficient Channel Attention and Dynamic-Attention Fusion highlight innovative techniques enhancing YOLO's performance. Applications span various domains, including autonomous driving, underwater detection, and real-time video processing, showcasing YOLO's versatility and efficiency in diverse environments.

3.2 Variants and Adaptations for Small Target Detection

The YOLO algorithm has significantly evolved, yielding adaptations enhancing small target detection. YOLOv5 stands out for its improvements in speed and accuracy, particularly in small object detection, using the Focus module to extract rich features [41]. This variant balances computational efficiency and detection performance, making it suitable for real-time applications.

YOLOv7 introduces the Extended Efficient Layer Aggregation Network (E-ELAN) for enhanced feature aggregation, crucial for capturing small target details. Compound model scaling and planned re-parameterization convolution improve detection accuracy and efficiency, providing a robust framework for small target detection [41].

The OL-NFA detection head in YOLOv7-tiny exemplifies adaptations for small target detection, refining objectness score estimation based on false alarm statistics, enhancing the model's ability to identify small objects against cluttered backgrounds [42]. Such innovations are vital for precision in small target detection applications like surveillance and aerial imaging.

These variants underscore YOLO's continuous evolution, enhancing small target detection capabilities. By incorporating advanced architectural features like GhostNet-based convolutional modules, RepGFPN-based Neck module optimization, and various attention mechanisms, these models significantly improve detection precision and reliability. This enhancement facilitates effective application in challenging environments, as demonstrated by studies including CourtNet's implementation, which balances precision and recall rates in infrared small-target detection, and LR-Net, achieving state-of-the-art performance in resource-limited scenarios. Furthermore, integrating synthetic data generation and image enhancement techniques in automatic detection systems for search and rescue operations illustrates the practical benefits of these advanced models [8, 43, 44, 40].

3.3 Innovative Techniques and Enhancements

Innovative techniques and enhancements have been pivotal in advancing YOLO's performance, particularly for small target detection. SpeechYOLO integrates detection and localization tasks in a single framework, enhancing both accuracy and efficiency [45]. This dual-task approach exemplifies YOLO's adaptability across various modalities.

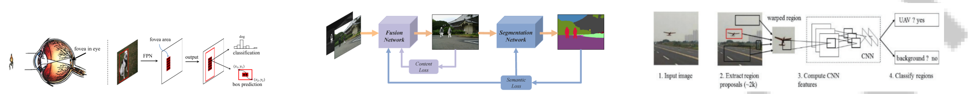
The Brain-Inspired Streaming Dual-Level Perturbation (BSDP) method mitigates catastrophic forgetting by incorporating feature-level and data-level perturbations, enhancing YOLO's ability to retain knowledge while learning new categories, thus maintaining high detection accuracy over time [31]. Such advancements are crucial for applications requiring continuous learning and adaptation.

The IoU-Net enhances localization confidence, improving non-maximum suppression (NMS) and bounding box refinement, which is particularly beneficial for small target detection where accurate localization is critical [46].

In infrared imagery, the normalized-cross-correlational (NCC) layer addresses small target detection requirements by normalizing inputs and filters, enhancing detection effectiveness while reducing convergence data needs [47]. This approach, through supervised training, computes optimal filters for target detection, providing tailored solutions for small targets in infrared environments.

Additionally, the EDGSP method employs a single-point prompt to guide pseudo label generation, leveraging energy distribution and embedding prompts at multiple stages of the detection network, thereby enhancing YOLO's ability to generate accurate labels with minimal supervision [48].

These enhancements demonstrate the ongoing evolution of the YOLO algorithm, highlighting its adaptability and potential for further advancements in object detection. By integrating advanced methodologies and architectural innovations, the YOLO framework consistently achieves superior detection accuracy and efficiency, solidifying its status as a preeminent choice in computer vision. The latest iteration, YOLOv6, exemplifies this trend with impressive performance metrics, achieving 43.5



(a) Eye Anatomy and Fovea Area in a Dog Image[49]

(b) Fusion Network and Segmentation Network for Person Re-identification[50]

(c) UAV Detection Using Region Proposal Network[9]

Figure 4: Examples of Innovative Techniques and Enhancements

As shown in Figure 4, the YOLO algorithm has revolutionized object detection with its innovative approach to real-time processing and accuracy. The "Eye Anatomy and Fovea Area in a Dog Image" example illustrates the precision required in distinguishing intricate details, akin to the fovea's role in human vision, crucial for applications requiring high detail recognition. Advancements in person re-identification are demonstrated through the "Fusion Network and Segmentation Network," which integrates multiple input images to enhance individual identification accuracy. Lastly, the "UAV Detection Using Region Proposal Network" example underscores the application of region proposal networks in detecting unmanned aerial vehicles, highlighting the identification process in complex environments. Together, these examples reflect the diverse enhancements and innovative methodologies evolving within YOLO and its applications in computer vision [49, 50, 9].

4 Challenges in Detecting Irregular Small Targets

4.1 Scale Variance and Small Target Representation

Detecting irregular small targets is significantly challenged by scale variance, which complicates their identification and localization within limited pixel spaces. CNN-based models often struggle with bounding box issues, leading to ineffective classification and high false positive rates, particularly in small infrared object detection [9, 5]. The inability to effectively extract features across scales results in information loss and computational overhead, especially in remote sensing [3]. Additionally, class imbalance in datasets, where small targets are overshadowed by larger elements, impedes effective feature learning [7]. In dynamic settings like autonomous driving, data-bias features exacerbate catastrophic forgetting, complicating small target representation [10]. The high similarity of certain targets to their surroundings, such as the crown of thorns starfish, further complicates detection [35]. Balancing precision and recall rates is crucial for small infrared target detection, necessitating models that integrate scale-aware mechanisms to enhance accuracy and efficiency in cluttered environments [8].

4.2 Occlusion and Background Clutter

Occlusion and background clutter present significant obstacles in detecting irregular small targets, complicating accurate identification and tracking. Occlusion, where targets are obscured, challenges consistent tracking and recognition, particularly in multi-target tracking that requires robust association mechanisms [51]. Detection methods often yield false readings due to environmental factors and occlusion [18]. Dense clutter exacerbates these issues, as small targets can be lost in noisy

AI-generated, for reference only.

environments. Advanced algorithms must distinguish small targets from clutter, using temporal and spatial cues to enhance tracking. Recent advancements, including bi-level adversarial frameworks and YOLOv5 enhancements, have improved detection through synthetic data generation and refined architectures [43, 52, 53, 40, 54].

4.3 Data and Computational Constraints

Data and computational constraints significantly impact the detection of small targets. The scarcity of dedicated datasets for small target detection hampers model development and generalization [11]. Computational constraints hinder complex model deployment, especially in resource-limited environments like UAVs [23]. Lightweight architectures and efficient scaling techniques are needed for real-time processing without sacrificing accuracy [55].

This figure illustrates the key challenges and solutions related to data and computational constraints in small target detection, highlighting data scarcity, computational limitations, and detection challenges Figure 5. Catastrophic forgetting in non-stationary data streams further complicates detection, as models must retain knowledge while adapting to new data [37]. Efficient multi-scale feature fusion remains challenging, often leading to computational overhead and reduced performance [55]. Quantization methods reduce model size but often decrease accuracy, highlighting the need for solutions that optimize computational resources while maintaining performance [56]. Strategies addressing data and computational constraints, incorporating efficient data use and continuous learning, are crucial for enhancing detection capabilities. Recent advancements in UAV-based systems and lightweight networks exemplify efforts to balance performance with resource use [44, 57, 40].

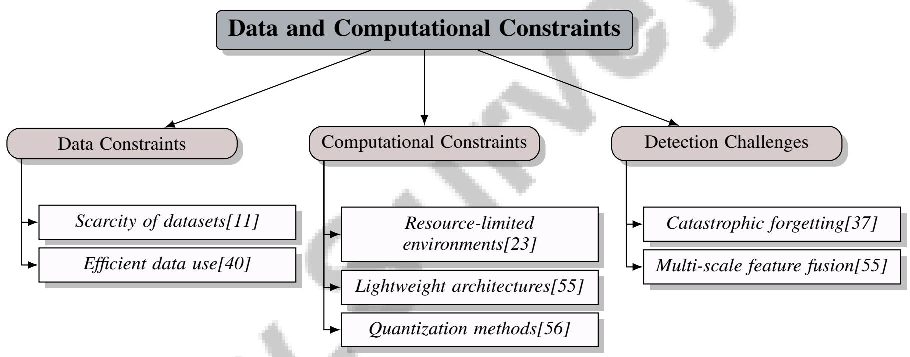


Figure 5: This figure illustrates the key challenges and solutions related to data and computational constraints in small target detection, highlighting data scarcity, computational limitations, and detection challenges.

4.4 Environmental and Operational Challenges

Environmental and operational challenges affect small target detection, particularly in dynamic settings. Variability in lighting and weather conditions impacts visibility and contrast, complicating real-time detection [47]. Sensor and optics choices also influence detection, as filter and sensor designs may limit applicability to specific bands, constraining model generalizability [47]. Sensor noise and limited fields of view further impede detection, especially in small, dim target detection in infrared imagery [58, 28, 57, 59, 6]. Addressing these challenges requires robust algorithms that adapt to varying conditions and configurations. Implementing advanced models resilient to environmental and operational variations enhances detection precision. Innovations like optimized convolutional modules and attention mechanisms improve accuracy in complex scenarios, such as search and rescue, ensuring effective performance across applications [54, 43, 57, 40].

5 Applications of YOLO in Small Target Detection

The YOLO algorithm is increasingly utilized across various domains, notably enhancing detection accuracy and operational efficiency. This section explores how YOLO is applied in surveillance,

AI-generated, for reference only.

security, UAVs, remote sensing, autonomous vehicles, agriculture, and medical imaging, highlighting its significant contributions.

5.1 Surveillance and Security

YOLO has substantially improved small target detection and tracking in surveillance and security, offering enhanced operational efficiency and accuracy. Its ability to process data at over 30 fps is critical for real-time applications, enabling swift responses and decision-making [2]. In dynamic environments, YOLO's high accuracy is beneficial for detecting moving targets [3]. Models like YOLO-S effectively address occlusion challenges in pedestrian detection, making them valuable for surveillance systems and autonomous driving. The lightweight design of YOLO networks ensures high detection accuracy across multi-scale samples [2]. CourtNet has outperformed traditional methods in early-warning systems for small infrared target detection, demonstrating superior capabilities in challenging conditions [8].

The MiPa method maintains competitive performance on LLVIP and FLIR datasets, ensuring modality invariance without additional inference time, crucial for flexible surveillance applications [36]. Models like TBC-Net, evaluated on real infrared sequences from drones, further demonstrate YOLO's effectiveness in UAV-based surveillance [7]. YOLO's deployment in surveillance systems highlights its adaptability and precision across applications, including urban traffic management and construction site safety. The YOLO-EA model, a refined version of YOLOv5, achieves a precision of 98.9

5.2 UAV and Remote Sensing Applications

YOLO has significantly enhanced small target detection in UAV and remote sensing technologies, crucial for environmental and agricultural monitoring. Ag-YOLO effectively detects palm trees for precise pesticide application, demonstrating its utility in precision agriculture [60]. YOLO-based methods excel in detecting sparse objects, vital for rapid identification in UAV operations [39]. Advanced attention mechanisms and efficient architectures promise to further enhance YOLO's performance in these applications [43]. The InfraYOLO model emphasizes the importance of benchmarking real-time capabilities in UAV systems, ensuring robust operation under computational constraints [5].

Recent advancements, such as the lightweight YOLO-S network, show significant improvements in speed and accuracy for real-time detection in mobile environments, enhancing operational capabilities and fostering interdisciplinary integration of UAV technology [25, 23, 2].

As illustrated in Figure 6, which depicts the hierarchical organization of UAV and remote sensing applications utilizing YOLO-based methods, key enhancements include Ag-YOLO, ODGI, and YOLOv5s-Optimized, all focusing on improvements in small target detection. The figure also highlights benchmark advancements for infrared small object detection through the InfraYOLO model and the InfraTiny dataset. Real-time detection capabilities are exemplified by the YOLO-S network and DDL-ST, which enhance speed and accuracy in mobile and search scenarios. The system processes 4K images by dividing them into smaller patches for efficient analysis, utilizing SSD modules to identify objects. Contrast enhancement further improves detection accuracy, supported by a sophisticated deep learning architecture for high-resolution image management [40, 61].

5.3 Autonomous Vehicles and Transportation

In autonomous vehicles and transportation, YOLO advances the detection of small targets like pedestrians and traffic signs, critical for safety and operational efficiency. Models such as YOLOv5 and YOLO-EA have demonstrated improved accuracy and speed in real-time object detection, enhancing compliance with safety measures in complex environments like railway construction sites [33, 25, 20, 2]. YOLO's real-time processing enables rapid obstacle identification in dynamic driving environments, essential for autonomous navigation.

The Threat Detection Model has achieved an accuracy of 82.65

YOLO's continuous evolution ensures its relevance in the transportation sector, significantly enhancing safety and reliability through real-time object detection. With an accuracy of approximately 93

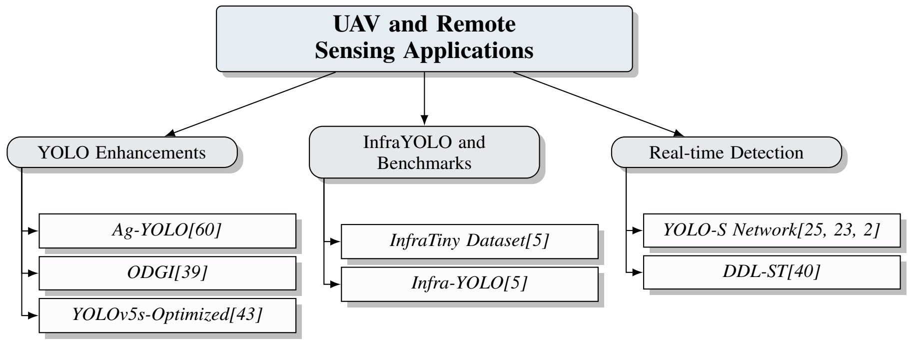


Figure 6: This figure illustrates the hierarchical organization of UAV and remote sensing applications utilizing YOLO-based methods. Key enhancements include Ag-YOLO, ODGI, and YOLOv5s-Optimized, focusing on small target detection improvements. InfraYOLO and the InfraTiny dataset highlight benchmark advancements for infrared small object detection. Real-time detection is exemplified by the YOLO-S network and DDL-ST, enhancing speed and accuracy in mobile and search scenarios.

5.4 Agriculture and Environmental Monitoring

YOLO is crucial in agriculture and environmental monitoring, providing real-time detection capabilities that enhance efficiency and productivity. Its ability to detect small targets like fruits and pests improves automated monitoring and management practices, leading to better resource allocation and crop yield [28]. The integration of generative AI with YOLO has further enhanced its capabilities, as seen in studies utilizing generative AI for fruit detection and quality assessment [62].

In environmental monitoring, YOLO is essential for detecting small targets obscured by complex backgrounds, such as wildlife and environmental hazards. Its efficiency in processing large datasets in real-time makes it invaluable for disaster response and environmental protection efforts, emphasizing its relevance for timely decision-making [61].

YOLO's versatility and high performance in these applications underscore its role in enhancing operational efficiency and accuracy. By enabling real-time detection across various tasks, YOLO significantly contributes to advancements in automated systems, supporting sustainable agricultural practices and improved environmental stewardship while maximizing resource utilization. The integration of YOLO with UAV technology further amplifies its potential to address agricultural and ecological challenges [28, 25, 20, 63, 23].

5.5 Medical Imaging and Scientific Research

YOLO enhances detection and analysis in medical imaging and scientific research, improving diagnostic accuracy and research outcomes. In medical imaging, YOLO's rapid processing and high detection accuracy are crucial for early malignancy detection, with models like MSDet significantly reducing false positives [64]. The F-YOLOv3 model effectively detects and segments melanoma lesions, showcasing YOLO's adaptability to various medical imaging challenges [65].

YOLO's versatility extends across scientific research domains, including agriculture, underwater exploration, security, and medical imaging, facilitating monitoring and surveillance, terrain surveys, and advanced diagnostics. Recent advancements like YOLOv5 and YOLO-S enhance its capabilities in challenging environments, making it a critical asset for real-time detection tasks [43, 28, 2, 25, 66]. YOLO's efficiency in processing large datasets makes it invaluable for high-throughput analysis and precise target identification, enabling researchers to achieve greater insights in complex detection scenarios.

The integration of YOLO in medical imaging and scientific research significantly enhances detection accuracy and operational efficiency, supporting improved diagnostic capabilities and accelerating scientific investigations through faster inference times compared to traditional methods [25, 26].

6 Recent Advances and Techniques

Category	Feature	Method
Advancements in Network Architecture	Efficiency Enhancements	DASSF[3], YOLO-EA[20], YOLO-S[2]
Innovative Loss Functions	Robustness to Data Quality	AY[60], DLD[67]
Integration of Attention Mechanisms	Reliability Enhancement Attention Mechanism Integration	CFARnet[59] C2f$_R FAConv - TA$[22]
Techniques for Resource-Constrained Environments	Resource Optimization	FH[68], OY[69], AAE-YOLOv3[70]

Table 1: This table provides a comprehensive summary of recent advancements in the YOLO algorithm, categorizing them into network architecture improvements, innovative loss functions, attention mechanism integration, and techniques for resource-constrained environments. It highlights specific features and methods, along with their respective references, showcasing the progress in enhancing detection capabilities and efficiency across various applications.

The evolution of the YOLO algorithm is marked by significant improvements in network architecture, loss functions, and attention mechanisms, enhancing its detection capabilities across diverse applications. Table 4 presents a detailed comparison of recent advancements in the YOLO algorithm, emphasizing improvements in network architecture, innovative loss functions, and the integration of attention mechanisms. Additionally, Table 1 presents a detailed summary of the recent advances and techniques in the YOLO algorithm, emphasizing the improvements in network architecture, loss functions, attention mechanisms, and resource optimization methods.

6.1 Advancements in Network Architecture

Method Name	**Architectural Enhancements**	**Efficiency and Lightweight Design**	**Performance Metrics**
YOLO-EA[20]	Eca Attention Mechanism	Minimal Increase Parameters	Precision, Recall, Map
C2f_RFAConv-TA[22]	C2f_rfaconv	Reduced Size	Map Values
DASSF[3]	Dynamic-attention Scale-sequence	Reduced Computational Overhead	Mean Average Precision
YOLO-S[2]	Feature Fusion	Lightweight Backbone	Higher Accuracy Rates

Table 2: Comparative Analysis of Architectural Enhancements and Performance Metrics in Recent YOLO Models. This table presents a detailed comparison of various YOLO-based methods, highlighting their architectural enhancements, efficiency in design, and key performance metrics such as precision, recall, and mean average precision (MAP). The table underscores the advancements in lightweight design and computational efficiency across different models.

Recent architectural enhancements have bolstered YOLO's efficacy in detecting small and irregular targets. The YOLO-EA model exemplifies this with an efficient attention mechanism that enhances detection without significantly increasing parameters, crucial for real-time applications [20]. The YOLOv8 model, featuring the C2f_RFAConv module and Triplet Attention, achieves superior Mean Average Precision (MAP) and Precision-Recall (PR) metrics [22]. Similarly, DASSF's Dynamic Scale-Sequence Feature Fusion (DSSFF) module improves small target detection [3]. The YOLO-S model, with its reduced size and computational load, underscores the importance of lightweight designs in resource-constrained settings [2]. Models like LSTD and EfficientDet demonstrate innovative approaches to mitigate overfitting and reduce computational demands while maintaining accuracy [11, 55]. CourtNet and ICOD models further highlight advancements in balancing precision and recall, and feature retention in dynamic environments [8, 10]. Table 2 provides a comprehensive overview of recent advancements in YOLO network architectures, illustrating the enhancements in efficiency and performance metrics achieved by various models.

As illustrated in Figure **??**, recent advancements in network architecture highlight enhancements in YOLO models, innovative modules, and key performance metrics in object detection. The progression of YOLO, particularly YOLOv5, signifies substantial progress in object detection, outperforming its predecessors in speed and accuracy [9, 24, 71, 26, 27].

Figure 7: This figure illustrates recent advancements in network architecture, highlighting enhancements in YOLO models, innovative modules, and key performance metrics in object detection.

6.2 Innovative Loss Functions

The development of innovative loss functions is pivotal in enhancing YOLO's robustness, especially in noisy environments like remote sensing. The Dynamic Loss Decay (DLD) method exemplifies this by improving resilience against noisy labels [67]. Ag-YOLO integrates specialized loss functions with hardware accelerators, enhancing detection in resource-limited settings such as precision agriculture [60]. Continuous refinement of these functions is essential for maintaining high accuracy and competitiveness with traditional detectors [9, 26, 27].

6.3 Integration of Attention Mechanisms

Attention mechanisms significantly enhance YOLO's detection accuracy by focusing on relevant features and filtering noise. Mechanisms like Efficient Channel Attention (ECA) and Triplet Attention improve performance in complex environments, crucial for small target detection [22]. These mechanisms ensure robust detection across various scenarios by maintaining constant false alarm rates and adjusting classification loss [59]. The integration of spatial and channel-wise attention addresses occlusion and background clutter without additional computational costs [72, 69, 73, 74]. YOLOv6's impressive metrics affirm its adaptability and solidify its position as a leading object detection framework [24, 25, 26].

6.4 Techniques for Resource-Constrained Environments

Method Name	Optimization Strategies	Efficiency Techniques	Application Domains
OY[69]	Attention Mechanisms	Data Augmentation	Ore Sorting
FH[68]	Error Rate Optimization	Framehopper, Quantization	Uav Technology
AAE-YOLOv3[70]	Automatic Labeling	Data Augmentation	Aerial Imagery

Table 3: Overview of optimization strategies, efficiency techniques, and application domains for various methods in resource-constrained environments. The table highlights the diverse approaches employed to enhance detection accuracy and computational efficiency across different technological domains.

Optimizing YOLO for resource-constrained environments involves enhancing detection accuracy while minimizing computational demands. Table 3 provides a comprehensive summary of methods optimized for resource-constrained environments, detailing their respective optimization strategies, efficiency techniques, and application domains. Data augmentation strategies improve performance in limited-resource settings [69]. FrameHopper reduces redundant processing by leveraging temporal correlations in video applications [68]. Quantization techniques, like those in LBW-Net, reduce computational complexity and memory usage while preserving feature capture [56]. Future research should focus on enhancing robustness through fine-tuning with localized datasets and exploring multi-modal sensor fusion [75]. Techniques like rotated anchors can further enhance performance by accommodating diverse object orientations [70]. The ongoing refinement of YOLO algorithms emphasizes the balance between speed and accuracy, particularly in agriculture, UAV technology, and aerial imagery [28, 2, 24, 25, 23]. By focusing on data augmentation, efficient processing, and robust architectures, YOLO continues to evolve as a versatile tool in object detection across various applications.

Feature	Advancements in Network Architecture	Innovative Loss Functions	Integration of Attention Mechanisms
Performance Metric	Map And PR	Resilience IN Noise	Detection Accuracy
Optimization Focus	Small Target Detection	Noisy Environments	Feature Focus
Application Domain	Real-time Applications	Remote Sensing	Complex Environments

Table 4: This table provides a comparative analysis of recent advancements in the YOLO algorithm, focusing on improvements in network architecture, innovative loss functions, and the integration of attention mechanisms. It highlights key performance metrics, optimization focuses, and application domains associated with each feature, illustrating the diverse applications and enhanced capabilities of YOLO in various environments.

7 Comparative Analysis with Other Models

Comparative analysis of object detection models is essential for discerning their capabilities, particularly in small target detection. This section evaluates the YOLO algorithm's architecture and efficiency, contextualizing its role among detection models. The following subsection examines YOLO's advantages and challenges in small target detection across various applications.

7.1 Effectiveness of YOLO in Small Target Detection

YOLO's single-stage detection framework combines speed and accuracy, making it effective for small target detection and real-time processing. The YOLO-Tomato model exemplifies this with high detection accuracy under varying illumination, surpassing traditional methods [17]. Deep learning approaches, including YOLO, generally outperform traditional techniques, especially in challenging tasks [6]. The YOLO-EA model, with precision and recall rates of 98.9% and 94.7

However, YOLO's performance can decline under specific conditions, such as detecting small or obscured objects in snowy environments, highlighting areas for improvement [75]. Despite these challenges, YOLO's speed advantage remains significant, outperforming traditional CNN-based methods, which lack the necessary speed for dynamic environments [68]. In infrared target detection, models like ALCNet outperform state-of-the-art methods, showcasing the potential of attention mechanisms to enhance YOLO's capabilities [72]. These advancements underscore YOLO's adaptability across applications, from agriculture to security surveillance.

YOLO's balance of speed and accuracy makes it a formidable tool for small target detection, offering significant advantages in real-time applications. Its ongoing evolution enhances its relevance across detection scenarios, including concealed and arbitrary-oriented object detection. Robust methodologies like Search Identification Network (SINet) and Dynamic Anchor Learning (DAL) show superior performance in challenging environments, solidifying YOLO's role in advancing the field [76, 58].

7.2 Single-Stage vs. Two-Stage Detectors

The comparison between single-stage and two-stage detectors is crucial for understanding object detection model advancements and limitations, particularly for small target detection. Single-stage detectors like YOLO are known for their exceptional processing speeds, predicting bounding boxes and class probabilities in one pass. This efficiency allows YOLO to achieve inference times up to 300 times faster than two-stage detectors like Fast R-CNN while maintaining competitive accuracy on benchmarks like the COCO dataset. YOLO's evolution, including YOLOv6, enhances its applicability across industrial scenarios, balancing speed and accuracy demands [9, 24, 25, 26, 27].

Two-stage detectors, such as Faster R-CNN, involve generating region proposals and classifying them, enhancing accuracy but resulting in longer inference times. While they often achieve higher accuracy, their computational demands and processing times can be drawbacks in scenarios requiring immediate detection and response [77, 38, 26].

For small target detection, YOLO's grid-based prediction mechanism maintains high accuracy for small and irregular targets by leveraging efficient feature extraction and attention mechanisms. However, challenges like scale variance and background clutter can affect performance, necessitating enhancements in feature representation and network design [3]. Conversely, two-stage detectors benefit from refining region proposals, improving accuracy in small target detection, albeit with increased processing time and resource consumption [13].

The choice between single-stage and two-stage detectors depends on task-specific factors, such as the need for accuracy versus inference speed. While two-stage detectors like Faster R-CNN typically offer superior accuracy, single-stage detectors like YOLO provide faster inference times, making them suitable for real-time applications despite occasionally lower accuracy [77, 28, 26, 57]. For real-time processing, single-stage detectors like YOLO present a compelling advantage, while two-stage detectors may be more appropriate where detection accuracy is paramount. Ongoing research into optimizing both approaches for small target detection ensures their continued relevance in computer vision.

7.3 Comparative Analysis of Deep Learning Methods

The landscape of deep learning methods for small target detection includes diverse approaches with unique strengths and challenges. YOLO stands out for its single-stage architecture, facilitating real-time processing, especially in dynamic environments [1]. Alternative methods like Faster R-CNN and SSD emphasize accuracy and precision through different strategies.

Faster R-CNN, a two-stage detector, achieves high accuracy by generating and refining region proposals. While it excels in precision-critical scenarios, its computational demands and slower processing speeds can limit real-time applications [13]. SSD merges the speed of single-stage detectors with two-stage models' accuracy by predicting bounding boxes and class scores directly from feature maps of varying scales [15].

Hybrid models integrate features from both single-stage and two-stage detectors. EfficientDet employs a compound scaling method to balance accuracy and efficiency, achieving competitive performance with reduced computational overhead [55]. This approach highlights the potential of combining architectural innovations to enhance small target detection across applications.

Attention mechanisms and feature fusion techniques enrich deep learning methods for small target detection. Models like ALCNet use attention mechanisms to enhance feature extraction and localization, improving infrared target detection performance [72]. Dynamic feature fusion methods, as in the DASSF model, emphasize adapting feature representations to enhance detection accuracy for small and occluded targets [3].

The comparative analysis of deep learning methods for small target detection reveals a complex interplay between speed, accuracy, and computational efficiency. While YOLO remains a prominent choice for real-time detection due to its balance of speed and accuracy, ongoing advancements foster specialized solutions for specific detection challenges across fields like security, healthcare, and autonomous driving. This evolution enhances existing models like YOLO and encourages new frameworks, such as YOLOv5 and YOLOv6, tailored for industrial applications with improved accuracy and efficiency in real-world scenarios [24, 25, 26, 27]. The integration of innovative architectural features and attention mechanisms continues to drive advancements, enhancing detection model capabilities across various domains.

7.4 Performance Metrics and Model Evaluation

Evaluating object detection models, especially for small target detection, requires comprehensive metrics to ensure accuracy, precision, and efficiency. Mean Average Precision (mAP) is a primary metric, aggregating precision across recall levels to assess a model's ability to detect objects accurately across scales and conditions [78]. mAP is significant for models like YOLO, where balancing speed and precision is critical for real-time applications.

Intersection over Union (IoU) measures the overlap between predicted and ground truth bounding boxes, directly assessing localization accuracy. High IoU values indicate precise object localization, crucial for evaluating models detecting small and irregular targets [46]. IoU is often used alongside precision and recall metrics to provide a holistic view of a model's performance, particularly in challenging detection scenarios involving small targets.

Model evaluation also considers computational efficiency, measured by floating point operations per second (FLOPs) and inference time. These metrics are vital for determining a model's suitability for deployment in resource-constrained environments [55]. Models like YOLO, which prioritize speed, are evaluated based on their ability to maintain high detection accuracy while minimizing computational demands, making them suitable for edge devices and real-time applications.

Additionally, robustness is assessed through performance under varying environmental conditions and occlusion scenarios, tested across diverse datasets to ensure adaptability in real-world applications [72]. The ability to maintain consistent performance across modalities, such as infrared and visible light, is critical, emphasizing modality invariance in detection tasks [36].

A comprehensive evaluation approach for object detection models, particularly in small target detection, encompasses multiple dimensions, including accuracy, efficiency, and adaptability. This is crucial given the challenges posed by complex backgrounds and the need for precise recognition of tiny targets, as demonstrated by advancements in models like YOLOv5, which have shown significant improvements through multi-module optimization strategies. The development of large-scale datasets, such as COD10K, highlights the importance of robust annotations and diverse real-world scenarios in enhancing model performance in concealed object detection tasks [43, 58]. Employing a comprehensive set of performance metrics ensures the development of robust and effective detection models capable of meeting the demands of diverse applications.

8 Future Directions and Research Opportunities

The dynamic field of computer vision requires a strategic focus on future directions and research opportunities to advance object detection algorithms. This section identifies pivotal areas for enhancing YOLO's capabilities through the integration of advanced technologies and modalities, addressing challenges in Few-Shot Object Detection (FSOD), and adopting innovative methodologies to improve detection accuracy and efficiency.

8.1 Integration with Advanced Technologies and Modalities

Integrating YOLO with cutting-edge technologies and multiple modalities presents significant opportunities to enhance detection capabilities across domains. Future research should aim to improve YOLO's generalization in FSOD tasks by addressing domain shifts and integrating it with other computer vision tasks [34]. This is crucial for applications like aquatic species detection, where environmental variability is a challenge.

Optimizing techniques like Dynamic-Attention Scale-Sequence Fusion (DASSF) will be essential for improving YOLO's real-world efficiency [3]. Enhancements to feature decomposers and strategies to mitigate data-bias impacts are also needed to bolster model robustness and adaptability [10]. Advanced architectures and refined models, such as CourtNet, can enhance robustness, particularly in infrared small target detection [8]. Research will also focus on diverse underwater environments to improve the capabilities of underwater robots for ecological protection, leveraging YOLO's adaptability in challenging conditions [35].

Strategies like initial pre-training and curriculum learning will be explored to boost model performance across various scenarios [36]. Implementing these advanced integration strategies will ensure YOLO's continued relevance and effectiveness in applications such as real-time object detection, security, and automated driving, as evidenced by improvements in YOLOv5 [71, 25, 26, 27].

8.2 Exploring New Application Domains

Exploring new application domains for YOLO in small target detection offers significant potential for capability enhancement. In agriculture, combining generative AI with YOLO can improve detection accuracy and efficiency, with future research focusing on generative AI applications across agricultural domains and utilizing synthetic datasets for sophisticated model development [62]. This can enhance agricultural monitoring and management, boosting productivity and sustainability.

YOLO's potential in semantic segmentation is another promising area. Refining scaling methods in models like EfficientDet could enable YOLO to tackle complex tasks such as segmenting objects in cluttered environments [55], which is crucial in medical imaging and autonomous navigation.

Expanding training datasets to include diverse scenarios will significantly enhance YOLO's generalization capabilities. Incorporating a wide range of environmental conditions and object types will refine algorithms to improve accuracy and processing speed [79]. These expansions ensure YOLO models remain robust and effective across various applications, from environmental monitoring to industrial automation.

The exploration of new application domains for YOLO in small target detection highlights the algorithm's versatility and potential for innovation. By leveraging advancements in generative AI, optimizing scaling techniques, and broadening training datasets, researchers can significantly enhance YOLO's performance and adaptability, particularly in its latest iteration, YOLOv5. These improvements address challenges related to detecting small targets and complex backgrounds, ensuring YOLO's relevance in computer vision applications, including security, healthcare, and autonomous driving [43, 25].

AI-generated, for reference only.

8.3 Advanced Evaluation Metrics and Benchmarking

Benchmark	Size	Domain	Task Format	Metric
SIRST[80]	427	Infrared Small Target Detection	Instance Segmentation	nIoU, IoU
SIRST-V2[81]	1,024	Infrared Imaging	Object Detection	mNoCoAP
YOLOv6[24]	123,000	Object Detection	Object Detection	AP, FPS
UAV-Benchmark[79]	10,000	Image Processing	Object Detection	Mean Average Precision, Intersection over Union
LimitIRSTD[57]	15,000	Infrared Small Target Detection	Weakly Supervised Detection	IoU, Pd
GAI-MFQ[62]	700	Agriculture	Fruit Detection	PSNR, SSIM
MARITIME-YOLO[82]	56,400	Maritime Surveillance	Object Detection	mAP, AP
DUT[83]	24,804	Uav Detection	Object Detection	mAP50, IoU

Table 5: This table presents a comprehensive summary of various benchmarks used in the evaluation of object detection models, highlighting their size, domain, task format, and the specific metrics employed. These benchmarks span multiple domains, including infrared imaging, agriculture, and maritime surveillance, providing a broad spectrum of challenges for assessing detection capabilities. The inclusion of diverse task formats and metrics underscores the complexity and specificity required for accurate performance evaluation in different environments.

Advancing evaluation metrics and benchmarking strategies is crucial for refining object detection models in small target detection. Traditional metrics like Average Precision (AP) and mean Average Precision (mAP) have been foundational in assessing detection algorithms' precision and recall capabilities [9]. However, these metrics may not fully capture the complexities of detecting small and irregular targets in diverse environments, necessitating future research to refine them for a more nuanced understanding of model performance.

Innovative evaluation frameworks should incorporate multi-frame analysis, as seen in the Local Patch Network (LPNet), which integrates temporal features to enhance detection capabilities [84]. Future metrics should account for temporal dynamics to improve detection outcomes in video-based applications. Optimizing models like LCAE-Net, which aims to reduce false alarm rates, highlights the need for evaluation frameworks that consider practical deployment capabilities [74].

Expanding datasets to include a wide array of object types and environmental conditions is essential for developing robust evaluation metrics. Future research should prioritize comprehensive dataset creation encompassing various scenarios, such as different drone types and aquatic environments, enhancing model generalization and applicability. This expansion will facilitate precise assessments of model performance and robustness through advanced methodologies, including multi-module optimization, computation reallocation strategies, and uncertainty quantification techniques [43, 38, 54, 24, 85].

Moreover, future research could explore efficient memory management strategies and alternative network architectures to improve speed and performance in online object detection [37]. Investigating dynamic chunk sizes for multi-stream videos and optimizing reinforcement learning agents could further enhance detection capabilities in video surveillance applications [68].

Standard evaluation metrics for multi-target tracking, such as MOTA (Multiple Object Tracking Accuracy) and MOTP (Multiple Object Tracking Precision), remain critical for assessing tracking performance [51]. Future research should integrate these metrics with advanced detection frameworks to provide comprehensive evaluations of model performance in complex tracking scenarios. Enhancing models' generalization across different classes and improving benchmarks for Open World Object Detection (OWOD) methods are also vital areas for exploration [4].

The advancement of object detection research relies heavily on developing sophisticated evaluation metrics and benchmarking strategies. Table 5 provides an extensive overview of representative benchmarks crucial for advancing evaluation metrics and benchmarking strategies in object detection research. These tools are essential for assessing the performance of various detection algorithms, including those in the YOLO series and FSOD methods, and for ensuring that new techniques, such as Soft-NMS and Montage pre-training, can be effectively compared and optimized across diverse applications and datasets [24, 86, 34, 87]. By focusing on these areas, researchers can ensure the continued evolution of detection models, enhancing their performance and applicability across a wide range of applications.

9 Conclusion

The survey illustrates the profound influence of the YOLO algorithm in advancing small target detection, emphasizing its remarkable speed and accuracy compared to traditional networks. Innovations such as BTP-yoloV3 exemplify significant progress over conventional methodologies, suggesting substantial potential for future research and applications in embedded systems. YOLO's adeptness is further highlighted in traffic sign detection, effectively addressing challenges like sign degradation and geographical variability.

Exploration into anchor-free frameworks, such as FoveaBox, offers a promising direction for enhancing detection accuracy and adaptability, setting a strong foundation for ongoing object detection research. Moreover, the Multitask-Net model demonstrates exceptional accuracy in face detection and head pose estimation, underscoring the promise of multitask learning in practical applications.

Despite these advancements, challenges remain in the precise identification of objects under adverse conditions, such as winter environments. Future endeavors should focus on deep learning-based feature representation and weakly supervised learning to overcome these obstacles. The survey underscores YOLO's pivotal role in advancing object detection, particularly for small targets, while paving the way for future innovations that could further augment its capabilities and applications across various fields.

AI-generated, for reference only.

References

[1] Rachel Huang, Jonathan Pedoeem, and Cuixian Chen. Yolo-lite: a real-time object detection algorithm optimized for non-gpu computers. In *2018 IEEE international conference on big data (big data)*, pages 2503–2510. IEEE, 2018.

[2] Alessandro Betti. A lightweight and accurate yolo-like network for small target detection in aerial imagery, 2022.

[3] Haodong Li and Haicheng Qu. Dassf: Dynamic-attention scale-sequence fusion for aerial object detection, 2024.

[4] Orr Zohar, Kuan-Chieh Wang, and Serena Yeung. Prob: Probabilistic objectness for open world object detection, 2022.

[5] Zhonglin Chen, Anyu Geng, Jianan Jiang, Jiwu Lu, and Di Wu. Infra-yolo: Efficient neural network structure with model compression for real-time infrared small object detection, 2024.

[6] Nikhil Kumar and Pravendra Singh. Small and dim target detection in ir imagery: A review, 2023.

[7] Mingxin Zhao, Li Cheng, Xu Yang, Peng Feng, Liyuan Liu, and Nanjian Wu. Tbc-net: A real-time detector for infrared small target detection using semantic constraint, 2019.

[8] Jingchao Peng, Haitao Zhao, Kaijie Zhao, Zhongze Wang, and Lujian Yao. Courtnet for infrared small-target detection, 2023.

[9] Yong-Hwan Lee and Youngseop Kim. Comparison of cnn and yolo for object detection. *Journal of the semiconductor & display technology*, 19(1):85–92, 2020.

[10] Zhenwei He and Lei Zhang. Learning causal features for incremental object detection, 2024.

[11] Hao Chen, Yali Wang, Guoyou Wang, and Yu Qiao. Lstd: A low-shot transfer detector for object detection, 2018.

[12] Xiwen Kang, Hua mei Chen, Genshe Chen, Kuo-Chu Chang, and Thomas M. Clemons. Joint detection and classification of communication and radar signals in congested rf environments using yolov8, 2024.

[13] Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu, and Xindong Wu. Object detection with deep learning: A review. *IEEE transactions on neural networks and learning systems*, 30(11):3212–3232, 2019.

[14] Syed Sahil Abbas Zaidi, Mohammad Samar Ansari, Asra Aslam, Nadia Kanwal, Mamoona Asghar, and Brian Lee. A survey of modern deep learning based object detection models. *Digital Signal Processing*, 126:103514, 2022.

[15] Li Liu, Wanli Ouyang, Xiaogang Wang, Paul Fieguth, Jie Chen, Xinwang Liu, and Matti Pietikäinen. Deep learning for generic object detection: A survey. *International journal of computer vision*, 128:261–318, 2020.

[16] Zhuang He, Qi Li, Huajun Feng, and Zhihai Xu. Class agnostic moving target detection by color and location prediction of moving area, 2021.

[17] Guoxu Liu, Joseph Christian Nouaze, Philippe Lyonel Touko Mbouembe, and Jae Ho Kim. Yolo-tomato: A robust algorithm for tomato detection based on yolov3. *Sensors*, 20(7):2145, 2020.

[18] Zillur Rahman, Amit Mazumder Ami, and Muhammad Ahsan Ullah. A real-time wrong-way vehicle detection based on yolo and centroid tracking, 2022.

[19] Defan Chen and Luchan Zhang. Sl-yolo: A stronger and lighter drone target detection model, 2025.

[20] Hao Liu and Xue Qin. Target detection of safety protective gear using the improved yolov5, 2024.

[21] Ren Liu and Ziang Ren. Application of yolo on mask detection task, 2021.

[22] Zhipeng Ling, Qi Xin, Yiyu Lin, Guangze Su, and Zuwei Shui. Optimization of autonomous driving image detection based on rfaconv and triplet attention, 2024.

[23] Chunling Chen, Ziyue Zheng, Tongyu Xu, Shuang Guo, Shuai Feng, Weixiang Yao, and Yubin Lan. Yolo-based uav technology: A review of the research and its applications. *Drones*, 7(3):190, 2023.

[24] Chuyi Li, Lulu Li, Hongliang Jiang, Kaiheng Weng, Yifei Geng, Liang Li, Zaidan Ke, Qingyuan Li, Meng Cheng, Weiqiang Nie, et al. Yolov6: A single-stage object detection framework for industrial applications. *arXiv preprint arXiv:2209.02976*, 2022.

[25] Haokang Wen, Fengzhi Dai, and Yasheng Yuan. A study of yolo algorithm for target detection. In *26th international conference on artificial life and robotics (ICAROB)*, volume 26, pages 622–625, 2021.

[26] Tausif Diwan, G Anirudh, and Jitendra V Tembhurne. Object detection using yolo: Challenges, architectural successors, datasets and applications. *multimedia Tools and Applications*, 82(6):9243–9275, 2023.

[27] Viswanatha V, Chandana R K, and Ramachandra A. C. Real time object detection system with yolo and cnn models: A review, 2022.

[28] Chetan M Badgujar, Alwin Poulose, and Hao Gan. Agricultural object detection with you look only once (yolo) algorithm: A bibliometric and systematic literature review, 2024.

[29] Zhongdao Wang, Liang Zheng, Yixuan Liu, Yali Li, and Shengjin Wang. Towards real-time multi-object tracking, 2020.

[30] Daniel Oñoro-Rubio, Roberto J. López-Sastre, Carolina Redondo-Cabrera, and Pedro Gil-Jiménez. The challenge of simultaneous object detection and pose estimation: a comparative study, 2018.

[31] Yu Chen, Liyan Ma, Liping Jing, and Jian Yu. Bsdp: Brain-inspired streaming dual-level perturbations for online open world object detection, 2024.

[32] Semi-weakly supervised object detection by sampling pseudo ground-truth boxes.

[33] Marco Flores-Calero, César A Astudillo, Diego Guevara, Jessica Maza, Bryan S Lita, Bryan Defaz, Juan S Ante, David Zabala-Blanco, and José María Armingol Moreno. Traffic sign detection and recognition using yolo object detection algorithm: A systematic review. *Mathematics*, 12(2):297, 2024.

[34] Tianying Liu, Lu Zhang, Yang Wang, Jihong Guan, Yanwei Fu, Jiajia Zhao, and Shuigeng Zhou. Recent few-shot object detection algorithms: A survey with performance comparison, 2023.

[35] Jingyao Wang and Naigong Yu. Utd-yolov5: A real-time underwater targets detection method based on attention improved yolov5, 2022.

[36] Heitor R. Medeiros, David Latortue, Eric Granger, and Marco Pedersoli. Mipa: Mixed patch infrared-visible modality agnostic object detection, 2024.

[37] Manoj Acharya, Tyler L. Hayes, and Christopher Kanan. Rodeo: Replay for online object detection, 2020.

[38] Feng Liang, Chen Lin, Ronghao Guo, Ming Sun, Wei Wu, Junjie Yan, and Wanli Ouyang. Computation reallocation for object detection, 2019.

[39] Amelie Royer and Christoph H. Lampert. Localizing grouped instances for efficient detection in low-resource scenarios, 2020.

[40] Kyongsik Yun, Luan Nguyen, Tuan Nguyen, Doyoung Kim, Sarah Eldin, Alexander Huyen, Thomas Lu, and Edward Chow. Small target detection for search and rescue operations using distributed deep learning and synthetic data generation, 2019.

[41] Ismat Saira Gillani, Muhammad Rizwan Munawar, Muhammad Talha, Salman Azhar, Yousra Mashkoor, M Uddin, and U Zafar. Yolov5, yolo-x, yolo-r, yolov7 performance comparison: A survey. *Artificial Intelligence and Fuzzy Logic System*, pages 17–28, 2022.

[42] Alina Ciocarlan, Sylvie Le Hégarat-Mascle, Sidonie Lefebvre, Arnaud Woiselle, and Clara Barbanson. *A Contrario* paradigm for yolo-based infrared small target detection, 2024.

[43] Qingyang Li, Yuchen Li, Hongyi Duan, JiaLiang Kang, Jianan Zhang, Xueqian Gan, and Ruotong Xu. Improvement and enhancement of yolov5 small target recognition based on multi-module optimization, 2023.

[44] Chuang Yu, Yunpeng Liu, Jinmiao Zhao, and Zelin Shi. Lr-net: A lightweight and robust network for infrared small target detection, 2024.

[45] Yael Segal, Tzeviya Sylvia Fuchs, and Joseph Keshet. Speechyolo: Detection and localization of speech objects, 2019.

[46] Borui Jiang, Ruixuan Luo, Jiayuan Mao, Tete Xiao, and Yuning Jiang. Acquisition of localization confidence for accurate object detection. In *Proceedings of the European conference on computer vision (ECCV)*, pages 784–799, 2018.

[47] H. Seçkin Demir and Erdem Akagunduz. Filter design for small target detection on infrared imagery using normalized-cross-correlation layer, 2020.

[48] Shuai Yuan, Hanlin Qin, Renke Kou, Xiang Yan, Zechuan Li, Chenxu Peng, and Abd-Krim Seghouane. Beyond full label: Single-point prompt for infrared small target label generation, 2024.

[49] Tao Kong, Fuchun Sun, Huaping Liu, Yuning Jiang, Lei Li, and Jianbo Shi. Foveabox: Beyound anchor-based object detection. *IEEE Transactions on Image Processing*, 29:7389–7398, 2020.

[50] Zhipeng Chang, Ruiling Ma, and Wenliang Jia. Pedestrain detection for low-light vision proposal, 2023.

[51] Sanam Nisar Mangi. Multi-target tracking for video surveillance using deep affinity network: a brief review, 2021.

[52] Iván García, Rafael Marcos Luque, and Ezequiel López. Improved detection of small objects in road network sequences, 2021.

[53] Syed Muhammad Aamir, Hongbin Ma, Malak Abid Ali Khan, and Muhammad Aaqib. Real-time object detection in occluded environment with background cluttering effects using deep learning, 2024.

[54] Zhu Liu, Zihang Chen, Jinyuan Liu, Long Ma, Xin Fan, and Risheng Liu. Enhancing infrared small target detection robustness with bi-level adversarial framework, 2023.

[55] Mingxing Tan, Ruoming Pang, and Quoc V Le. Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10781–10790, 2020.

[56] Penghang Yin, Shuai Zhang, Yingyong Qi, and Jack Xin. Quantization and training of low bit-width convolutional neural networks for object detection, 2017.

[57] Boyang Li, Xinyi Ying, Ruojing Li, Yongxian Liu, Yangsi Shi, and Miao Li. The first competition on resource-limited infrared small target detection challenge: Methods and results, 2024.

[58] Deng-Ping Fan, Ge-Peng Ji, Ming-Ming Cheng, and Ling Shao. Concealed object detection. *IEEE transactions on pattern analysis and machine intelligence*, 44(10):6024–6042, 2021.

[59] Tzvi Diskin, Uri Okun, and Ami Wiesel. Learning to detect with constant false alarm rate, 2022.

[60] Zhenwang Qin, Wensheng Wang, Karl-Heinz Dammer, Leifeng Guo, and Zhen Cao. A real-time low-cost artificial intelligence system for autonomous spraying in palm plantations, 2021.

[61] Guangcun Shan, Hongyu Wang, Wei Liang, Congcong Liu, Qizi Ma, and Quan Quan. An efficient target detection and recognition method in aerial remote-sensing images based on multiangle regions-of-interest, 2022.

[62] Seungri Yoon, Yunseong Cho, and Tae In Ahn. Melon fruit detection and quality assessment using generative ai-based image data augmentation, 2024.

[63] Bohan Liang, Shangxi Wu, Kaiyuan Xu, and Jingyu Hao. Butterfly detection and classification based on integrated yolo algorithm, 2020.

[64] Guohui Cai, Ruicheng Zhang, Hongyang He, Zeyu Zhang, Daji Ergu, Yuanzhouhan Cao, Jinman Zhao, Binbin Hu, Zhinbin Liao, Yang Zhao, et al. Msdet: Receptive field enhanced multiscale detection for tiny pulmonary nodule. *arXiv preprint arXiv:2409.14028*, 2024.

[65] Mohamad Taghizadeh and Karim Mohammadi. The fast and accurate approach to detection and segmentation of melanoma skin cancer using fine-tuned yolov3 and segnet based on deep transfer learning, 2023.

[66] Chenjie Zhang and Pengcheng Jiao. Yolo series target detection algorithms for underwater environments, 2023.

[67] Guozhang Liu, Ting Liu, Mengke Yuan, Tao Pang, Guangxing Yang, Hao Fu, Tao Wang, and Tongkui Liao. Dynamic loss decay based robust oriented object detection on remote sensing images with noisy labels, 2024.

[68] Md Adnan Arefeen, Sumaiya Tabassum Nimi, and Md Yusuf Sarwar Uddin. Framehopper: Selective processing of video frames in detection-driven real-time video analytics, 2022.

[69] Junjiang Zhen and Bojun Xie. Fused attention mechanism-based ore sorting network, 2024.

[70] Siyan Li, Yue Xiao, Yuhang Zhang, Lei Chu, and Robert C. Qiu. Learning efficient representations for enhanced object detection on large-scene sar images, 2022.

[71] Wvo.

[72] Yimian Dai, Yiquan Wu, Fei Zhou, and Kobus Barnard. Attentional local contrast networks for infrared small target detection. *IEEE transactions on geoscience and remote sensing*, 59(11):9813–9824, 2021.

[73] Xiyang Dai, Yinpeng Chen, Bin Xiao, Dongdong Chen, Mengchen Liu, Lu Yuan, and Lei Zhang. Dynamic head: Unifying object detection heads with attentions. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7373–7382, 2021.

[74] Peichao Wang, Jiabao Wang, Yao Chen, Rui Zhang, Yang Li, and Zhuang Miao. Paying more attention to local contrast: improving infrared small target detection performance via prior knowledge, 2024.

[75] Stefan Schoder. First qualitative observations on deep learning vision model yolo and detr for automated driving in austria, 2024.

[76] Qi Ming, Zhiqiang Zhou, Lingjuan Miao, Hongwei Zhang, and Linhao Li. Dynamic anchor learning for arbitrary-oriented object detection, 2020.

[77] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

[78] Haichao Zhang, Jiashi Li, Xin Xia, Kuangrong Hao, and Xuefeng Xiao. Multi-objective evolutionary for object detection mobile architectures search, 2022.

[79] Alessandro Palmas and Pietro Andronico. Deep learning computer vision algorithms for real-time uavs on-board camera image processing, 2022.

[80] Yimian Dai, Yiquan Wu, Fei Zhou, and Kobus Barnard. Asymmetric contextual modulation for infrared small target detection. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 950–959, 2021.

[81] Yimian Dai, Xiang Li, Fei Zhou, Yulei Qian, Yaohong Chen, and Jian Yang. One-stage cascade refinement networks for infrared small target detection, 2022.

[82] Alessandro Betti, Benedetto Michelozzi, Andrea Bracci, and Andrea Masini. Real-time target detection in maritime scenarios based on yolov3 model, 2020.

[83] Mohssen E. Elshaar, Zeyad M. Manaa, Mohammed R. Elbalshy, Abdul Jabbar Siddiqui, and Ayman M. Abdallah. Performance evaluation of deep learning-based quadrotor uav detection and tracking methods, 2024.

[84] Fang Chen, Chenqiang Gao, Fangcen Liu, Yue Zhao, Yuxi Zhou, Deyu Meng, and Wangmeng Zuo. Local patch network with global attention for infrared small target detection, 2021.

[85] Marius Schubert, Karsten Kahl, and Matthias Rottmann. Metadetect: Uncertainty quantification and prediction quality estimates for object detection, 2020.

[86] This iccv paper is the open acce.

[87] Dongzhan Zhou, Xinchi Zhou, Hongwen Zhang, Shuai Yi, and Wanli Ouyang. Cheaper pre-training lunch: An efficient paradigm for object detection, 2020.

AI-generated, for reference only.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.

AI-generated, for reference only.