



Super-Resolution Enhancement and Scene Context Collaborative Dense Tiny Object Detection in Remote Sensing: Research Progress and Integrated Optimization Paradigm

Zhongyang Sun^{1*}, Abudhahir Buhari¹

¹Faculty of Engineering, Science and Technology, Kuala Lumpur University of Science and Technology, Kuala Lumpur, Malaysia

E-mail: sun784978929@outlook.com (Corresponding Author)

Abstract: Dense tiny object detection in high-resolution remote sensing images serves as a core technical bottleneck in the intelligent interpretation of remote sensing data, with critical application value across maritime supervision, urban traffic management, emergency rescue, national land security and other scenarios. Aiming at the limitations of existing research that mostly conducts optimization from a single technical dimension, suffers insufficient coordination among multi-path approaches, and exhibits severe performance degradation in dense scenes, this paper systematically reviews the technical evolution of remote sensing tiny object detection in the deep learning era. In-depth analysis is conducted along five major technical avenues: pre-processing super-resolution enhancement, multi-scale feature fusion, contextual awareness modeling, cascaded position refinement, and Transformer global modeling. The technical principles, implementation schemes, applicable scenarios and performance boundaries of various methods are compared. On this basis, an integrated optimization framework combining lightweight pre-processing super-resolution enhancement, scene context-guided feature pyramid, and cascaded object refinement head is proposed. Deep collaboration across multiple technical pathways compensates for the inherent defects of individual approaches. Experiments demonstrate that the proposed integrated framework can significantly boost detection accuracy in typical dense tiny object scenarios such as port ships and urban vehicles, while only introducing marginal computational overhead. This work provides novel optimization ideas and practical references for the engineering deployment of remote sensing tiny object detection technology.

Keywords: High-resolution Remote Sensing Imagery; Dense Tiny Objects; Super-resolution Reconstruction; Scene Context; Cascaded Detection; Multi-scale Feature Fusion.

1. Introduction

With the rapid advancement of aerospace Earth observation technology, sub-meter high-resolution remote sensing images have been routinely accessible, laying a data foundation for large-scale intelligent interpretation of surface information (Gui et al., 2024). As a core task of remote sensing image interpretation, object detection can automatically extract category, position and morphological information of targets within imagery, acting as a pivotal technical support for maritime ship monitoring, urban traffic census, disaster damage assessment, border anomaly screening and other business workflows (Han et al., 2021).

Unlike natural scene images, overhead remote sensing imagery is characterized by extremely small object sizes, highly dense distribution, random orientation variations, and complex background clutter. In imagery with a 3-meter ground sampling distance, small vehicles occupy merely dozens of pixels, and clusters of homogeneous tiny objects form within ports, open parking lots, airport aprons and other sites (Wang et al., 2023). Meanwhile, interference factors including cloud occlusion, vegetation coverage, building shadows and seasonal illumination fluctuations further raise the difficulty of tiny object identification. Conventional manual visual interpretation suffers from low efficiency and fails to meet the demands of large-scale, routine intelligent remote sensing monitoring, rendering automated deep learning detection algorithms an industry necessity (Yuan et al., 2023).

Deep learning has driven substantial progress in remote sensing object detection, yet dense tiny objects remain a major performance shortcoming. Empirical research conducted by Zhang et al. (2020) verified that detection accuracy declines

sharply as object size shrinks in high-resolution remote sensing imagery, and dense distribution further exacerbates this performance gap. Wang et al. (2022) focusing on seismic damage building detection also confirmed that generic detection algorithms suffer drastic performance attenuation in dense, class-imbalanced tiny object scenarios, failing to satisfy the missed detection rate requirements of practical operations. Therefore, breaking through the performance bottleneck of dense tiny object detection carries profound theoretical research significance and engineering application value.

Current research on remote sensing tiny object detection develops along multiple parallel technical routes, yet these routes remain fragmented with a lack of systematic sorting and collaborative integration. Accordingly, this paper centers on three core research questions: first, what are the core technical pathways for existing remote sensing tiny object detection, and what are their respective technical principles, performance advantages and inherent limitations; second, how to organically integrate technical pathways including super-resolution enhancement, contextual awareness and cascaded refinement to maximize detection accuracy in dense scenes while constraining computational overhead; third, where does the performance gain of the multi-path integrated framework originate, and how to define its applicable scenarios and performance boundaries.

The core research contents of this paper include four parts: first, systematically sorting out the technical evolution of remote sensing tiny object detection in the deep learning era and conducting literature review and in-depth analysis across five technical avenues; second, dissecting the performance bottlenecks and research

gaps of each technical pathway to clarify the necessity and feasibility of multi-path collaboration; third, constructing an integrated optimization framework combining lightweight super-resolution, contextual guidance and cascaded refinement, and elaborating the design logic and technical basis of each module; fourth, analyzing the performance strengths, applicable scenarios and limitations of the integrated framework, and prospecting future research directions.

2. Literature Review

2.1 Overall Evolution of Remote Sensing Object Detection Technology

Remote sensing object detection technology has evolved from handcrafted feature-based methods to deep learning-based approaches, and from two-stage architectures to one-stage pipelines. A comprehensive review by Gui et al. (2024) pointed out that deep learning-based remote sensing object detection has formed a complete technical system, achieving remarkable progress in feature extraction, multi-scale adaptation, rotated object detection and other subfields. Li et al. (2022) and Zhang et al. (2023) summarized the full landscape of deep learning-driven remote sensing object detection from the perspectives of technical classification, challenges and advances, confirming that tiny object detection represents the core research hotspot and difficulty within the field.

Extensive review studies have been conducted by domestic and international scholars targeting tiny object detection as a subdomain. Wang et al. (2023) comprehensively categorized deep learning-based remote sensing tiny object detection methods into multi-scale, contextual and super-resolution branches. Yuan et al. (2023) summarized domestic research progress on remote sensing tiny object detection,

identifying feature enhancement and sample optimization as two mainstream optimization directions. Hua & Chen (2025) surveyed tiny object detection in aerial imagery and analyzed emerging technical trends in the Transformer era. Wei et al. (2024) organized the technical system of tiny object detection from a generic computer vision perspective, providing references for cross-domain migration of methods to remote sensing. Cheng et al. (2023) focused on large-scale tiny object detection and established unified benchmarks and evaluation metrics, promoting standardization within this research field.

A seminal review by Chen et al. (2020) classified the core technologies of tiny object detection into four pillars: multi-scale representation, contextual information, super-resolution and region proposal generation, a categorization framework widely recognized across the field. Overall, the technical evolution of remote sensing tiny object detection follows a trend of "from single-dimensional optimization to multi-dimensional collaboration, and from accuracy priority to balanced accuracy and efficiency". Nevertheless, prominent research gaps persist regarding the integration of multi-technical pathways for dense scene applications.

2.2 Research Progress on Multi-Scale Features and Attention Mechanisms

Multi-scale feature representation constitutes the foundational technical pathway for tiny object detection. Its core logic lies in optimizing feature pyramid architectures to preserve shallow spatial details and deep semantic information, mitigating feature loss incurred by repeated downsampling of tiny targets. Wang et al. (2021) proposed FSoD-Net, a full-scale object detection network that achieves balanced detection performance across all object sizes via optimized feature

fusion pathways. Chen et al. (2023) designed a high-resolution feature pyramid network tailored to drone-view imagery, retaining high-resolution feature maps to improve localization precision for tiny objects.

Attention mechanisms serve as a critical auxiliary tool for multi-scale feature optimization, dynamically assigning weights to amplify valid target features and suppress background clutter. Li et al. (2020) developed a cross-layer attention network that enhances feature responses of tiny objects through cross-layer information interaction. Lu et al. (2021) integrated attention mechanisms with the SSD algorithm to boost multi-scale detection performance for remote sensing targets. Qu et al. (2023) constructed a remote sensing tiny object detection network combining attention mechanisms and multi-scale feature fusion, synchronously optimizing feature weights along channel and spatial dimensions. Liu et al. (2024) put forward an adaptive multi-scale feature fusion approach that dynamically adjusts feature fusion strategies according to the scale distribution of objects. Gong et al. (2022) fused Swin Transformer with YOLOv5, leveraging self-attention mechanisms to strengthen global contextual associations of tiny objects.

This technical pathway delivers the advantages of low computational overhead and easy embedding into existing detection architectures, making it the most commonly adopted optimization strategy in engineering applications. However, its limitation lies in the inability to inherently recover lost high-frequency details of tiny objects, yielding marginal performance improvements for extreme tiny targets occupying minimal pixel areas.

2.3 Research Advances in Super-resolution-Assisted Tiny Object Detection

Super-resolution reconstruction

recovers high-frequency image details to increase pixel occupancy and feature discriminability of tiny objects, constituting a direct technical solution to address insufficient feature information for small targets. Research within this domain falls into three categories: generative adversarial network (GAN)-based methods, convolutional neural network (CNN)-based methods, and lightweight approaches based on knowledge distillation.

GAN-based methods generate authentic high-frequency details through adversarial training. Rabbi et al. (2020) proposed an end-to-end edge-enhanced GAN architecture jointly trained with a detector, utilizing edge constraints to improve the pertinence of super-resolution reconstruction. Courtrai et al. (2020) designed an auxiliary GAN super-resolution module dedicated to detail enhancement for tiny object regions in remote sensing imagery. CNN-based methods directly learn the mapping from low-resolution to high-resolution images via convolutional networks. Bashir & Wang (2021) proposed a residual feature aggregation super-resolution network cascaded with a detector to lift tiny object detection performance. Xiaolin et al. (2022) experimentally validated the positive impact of super-resolution preprocessing on remote sensing tiny object detection. Zhang et al. (2024) developed a multi-stage enhancement network that boosts detection accuracy for tiny targets through multi-stage feature reconstruction.

To resolve the high computational overhead of traditional super-resolution pipelines, scholars have proposed lightweight schemes including knowledge distillation. Liu et al. (2023) transferred super-resolution network knowledge into detection networks via distillation, eliminating the need for an independent super-resolution module

during inference while retaining performance gains. Wang et al. (2021) embedded super-resolution blocks within YOLO architectures using dense blocks, balancing detection accuracy and computational efficiency. Zhang et al. (2023) introduced SuperYOLO, a framework integrating multi-modal super-resolution features to assist detection, whose effectiveness was validated on multi-modal remote sensing datasets.

Overall, super-resolution methods deliver substantial potential for accuracy improvement by recovering high-frequency details of tiny objects. Still, conventional end-to-end super-resolution pipelines incur heavy computational costs and risk introducing reconstruction artifacts, rendering lightweight and targeted optimization the primary research focus within this subfield.

2.4 Research Progress on Contextual Awareness and Cascaded Refinement Methods

Contextual awareness approaches compensate for insufficient feature information of tiny objects by modeling spatial and semantic correlations between targets and their surrounding scenes, lowering false positive rates in complex backgrounds. Cascaded refinement methods iteratively correct detection outputs across multiple stages to elevate localization precision; both techniques represent vital optimization directions for dense scene detection.

Within contextual awareness research, Zhao et al. (2024) proposed a scene context-guided dense tiny object detection method and constructed a unified benchmark for dense tiny target detection, verifying the significant performance gains brought by scene semantics for dense object detection. Wu & Xu (2021) put forward a hierarchical detection paradigm of "point-to-region", which first generates

coarse localization via target center heatmaps and then executes precise detection incorporating regional context, suitable for low-resolution remote sensing tiny object scenarios. Ye et al. (2022) designed a global-local feature enhancement network that simultaneously captures global scene context and local target details, improving dense tiny object detection performance for drone imagery.

In cascaded refinement research, Wu et al. (2022) developed a cascaded multi-stage detector that achieves high-quality object detection across multi-resolution remote sensing imagery via progressive multi-stage detection. Dai et al. (2023) constructed a one-stage cascaded refinement network that iteratively optimizes features across multiple levels to refine localization of infrared tiny targets. Zhang et al. (2023) proposed CoF-Net, a coarse-to-fine framework that rapidly generates coarse candidate boxes before conducting refined detection on candidate regions. Hou et al. (2022) designed a refined one-stage architecture for rotated object detection, correcting features to boost localization precision of rotated bounding boxes. Yu & Ji (2021) built a spatially oriented object detection framework that optimizes rotated target detection performance via spatial context integration.

Furthermore, training sample quality optimization has been proven to enhance cascaded detection performance. Han et al. (2020) proposed a sampling-balanced multi-stage network that elevates aerial image object detection performance—especially for dense tiny object scenarios—by optimizing training sample distribution. This category of methods effectively reduces false positives and improves localization accuracy, yet multi-stage structures introduce increased inference latency, necessitating trade-offs between

accuracy and efficiency.

2.5 Applications of Transformer and Novel Detection Paradigms

Powered by global modeling capabilities derived from self-attention mechanisms, Transformer architectures have emerged as a novel detection paradigm in computer vision and have been gradually adopted for remote sensing tiny object detection research. This technical pathway is split into two branches: pure Transformer detection architectures and hybrid CNN-Transformer architectures.

Pure Transformer architectures are built upon the DETR series with targeted modifications for remote sensing tiny object characteristics. Kong et al. (2024) proposed Drone-DETR, which enhances the RT-DETR model for efficient tiny object detection on drone remote sensing imagery. Ma et al. (2025) developed OWRT-DETR, a real-time Transformer network optimized exclusively for tiny object detection in open-water UAV search-and-rescue scenarios. From an efficiency perspective, Xie et al. (2024) proposed a "fewer is more" strategy for efficient detection on large aerial images, utilizing sparse attention mechanisms to cut computational overhead.

Hybrid architectures combine CNNs' strength in capturing local details with Transformer's global modeling capacity, balancing accuracy and efficiency. Gong et al. (2022) integrated Swin Transformer with YOLOv5, leveraging attention mechanisms to improve tiny object detection performance on satellite imagery. Qi et al. (2022) proposed adaptive spatial parallel convolution and fast multi-scale fusion methods, optimizing tiny object features with Transformer-inspired global modeling logic. This technical pathway excels at global contextual modeling, delivering prominent accuracy advantages for large-scale

dense target scenes, yet suffers from generally high computational complexity and weaker local detail capture for tiny objects compared with pure CNN architectures. Lightweight Transformer design and deep CNN-Transformer fusion represent key developmental directions for this research branch.

2.6 Lightweighting and Real-Time Optimization of One-Stage Detectors

Engineering deployment of remote sensing detection algorithms imposes strict real-time requirements. Benefiting from speed advantages, one-stage detectors have become the mainstream deployment carrier, and abundant research has focused on adapting one-stage architectures for tiny object detection. Pham et al. (2020) proposed YOLO-Fine, a detector specially optimized for remote sensing tiny objects under diverse backgrounds. Liu et al. (2024) designed SDSDet, a real-time detector tailored for small, dense, multi-scale remote sensing target scenarios.

Improvements to the YOLO series dominate one-stage optimization research. Wan et al. (2023) introduced YOLO-HR, a modified YOLOv5 variant adapted for object detection on high-resolution optical remote sensing imagery. Ma & Pang (2023) developed SP-YOLOv8s for tiny object detection in remote sensing imagery. Jin et al. (2024) put forward YOLOv7-bw, an efficient detector for dense tiny objects that preserves inference speed while boosting accuracy in dense scenes. Liu et al. (2025) proposed YOLO-SSP, which enhances remote sensing object detection performance via pyramid spatial attention and improved downsampling strategies. Xu & Wu (2020) combined DenseNet with YOLO-V3 to improve multi-scale remote sensing target detection results.

Specialized one-stage optimization research targeting specific scenarios has also yielded fruitful outcomes. Yang et al. (2021) constructed a robust one-stage detector for multi-scale ship detection on SAR imagery with complex backgrounds. Fan et al. (2024) designed a dedicated optimization pipeline for tiny ship detection on satellite remote sensing data. Koyun et al. (2022) proposed the Focus-and-Detect framework, which accelerates aerial tiny object detection via a two-step pipeline of region focusing followed by detection. Liu et al. (2024) integrated target reconstruction and multi-receptive-field adaptive feature enhancement to improve tiny object detection in remote sensing imagery. Wu et al. (2022) developed FSANet, which optimizes tiny object detection through feature and spatial alignment.

2.7 Critical Review of Existing Research

In summary, research on dense tiny object detection in remote sensing has developed along multiple parallel technical pathways, with remarkable progress achieved in multi-scale feature optimization, super-resolution enhancement, contextual modeling, cascaded refinement, Transformer-driven novel paradigms and lightweight one-stage detector optimization. Nevertheless, three core research gaps persist in existing literature: first, fragmented technical pathways lacking deep collaborative integration; most prior studies optimize along a single dimension, and fusion of super-resolution, contextual awareness and cascaded refinement only adopts simple sequential concatenation without integrated architectures supporting feature reuse and coordinated optimization, failing to fully exploit superimposed performance gains across multiple technical routes; second, insufficient targeted optimization for

dense scenes; most methods are designed for generic tiny objects and fail to fully exploit spatial distribution patterns and semantic scene correlations of dense targets, leading to persistently high missed detection and false positive rates for highly concentrated scenarios such as ports and parking lots; third, unresolved trade-offs between accuracy and efficiency; high-precision methods including super-resolution and Transformer architectures incur heavy computational overhead, hindering deployment on low-computing edge terminals for spaceborne and airborne platforms, while lightweight pipelines often sacrifice excessive detection accuracy and cannot satisfy stringent precision requirements of industrial scenarios. Furthermore, remote sensing image interpretation constitutes a multi-task integrated system, and object detection shares feature foundations with change detection and other tasks, yet existing research mostly addresses these tasks independently without collaborative optimization via multi-task joint learning (Ning et al., 2024; Peng et al., 2021). Addressing the aforementioned gaps by constructing an integrated optimization framework with coordinated multi-technical pathways emerges as the key direction to break through performance bottlenecks for dense tiny object detection.

3. In-Depth Analysis of Core Technical Pathways for Remote Sensing Dense Tiny Object Detection

3.1 Pre-Processing Enhancement via Super-Resolution Reconstruction

3.1.1 Technical Principles

The core logic of super-resolution pre-processing enhancement lies in the root cause of low tiny object detection accuracy: insufficient pixel information and lost high-frequency details. Super-resolution reconstruction elevates image resolution to directly increase effective

pixel counts of tiny targets and recover critical edge and texture details, thereby strengthening feature extraction capability and recognition precision of detectors. This pathway constitutes a straightforward solution to mitigate insufficient tiny object feature information at the data input stage.

3.1.2 Mainstream Implementation Schemes

Implementation schemes for this pathway are categorized into three types by deployment position: pre-processing, embedded, and distillation-based pipelines. Pre-processing schemes deploy super-resolution networks as independent preprocessing modules that upsample full images prior to detection, delivering simple implementation yet maximum computational overhead, represented by the edge-enhanced GAN architecture proposed by Rabbi et al. (2020). Embedded schemes integrate super-resolution modules within detection networks, applying enhancement exclusively to low-level feature maps of feature pyramids rather than full images, drastically reducing computational load, exemplified by the multi-stage enhancement network of Zhang et al. (2024). Distillation-based schemes transfer feature reconstruction capabilities of super-resolution networks into detection architectures via knowledge distillation, eliminating the requirement for separate super-resolution modules during inference to achieve optimal efficiency, represented by the super-resolution knowledge distillation pipeline proposed by Liu et al. (2023).

3.1.3 Performance Boundaries and Limitations

The performance ceiling of this pathway is constrained by the original pixel information volume of tiny objects: super-resolution delivers substantial accuracy gains for targets larger than 10×10 pixels, yet for ultra-tiny objects

below 8×8 pixels, innate scarcity of raw information leads to reconstruction artifacts that interfere with detection performance. Meanwhile, conventional full-image super-resolution pipelines incur prohibitive computational overhead incompatible with real-time detection demands. Although embedded and distillation-based variants achieve higher efficiency, their reconstruction quality underperforms full-image pipelines, creating an inherent trade-off between accuracy and computational efficiency.

3.2 Feature Enhancement via Contextual Awareness

3.2.1 Technical Principles

The core logic of contextual awareness is that remote sensing targets do not exist in isolation but maintain stable spatial and semantic correlations with surrounding scenes (e.g., ships situated on sea surfaces, vehicles positioned on roadways). Modeling these correlations supplements deficient feature information of tiny objects while suppressing false positives in background regions. For dense scenes, spatial distribution regularities among clustered targets can also serve as contextual auxiliary information for detection.

3.2.2 Mainstream Implementation Schemes

This pathway is split into three branches: local context, global scene context, and semantic context. Local context aggregates neighboring features surrounding targets by expanding convolutional receptive fields via dilated convolution and deformable convolution to supplement feature information. Global scene context models long-range dependencies across full images via Transformer or global attention mechanisms to capture large-scale scene distribution patterns, represented by the Swin-YOLO architecture proposed by Gong et al. (2022). Semantic context

introduces lightweight semantic segmentation branches to extract scene semantic masks guiding object detection, exemplified by the scene context-guided pipeline developed by Zhao et al. (2024).

3.2.3 Performance Boundaries and Limitations

Contextual awareness methods reduce false positive rates and boost recall rates for dense scenes with negligible additional computational overhead, delivering outstanding cost-performance ratios. Nevertheless, their performance relies on stable scene semantics, leading to degraded generalization across disparate scenes—for instance, contextual models trained on port ship datasets exhibit drastically reduced effectiveness when directly applied to inland landscapes. Additionally, contextual information functions solely as auxiliary guidance and cannot compensate for innate feature scarcity of tiny objects.

3.3 Multi-Stage Detection via Cascaded Refinement

3.3.1 Technical Principles

Tiny objects possess small spatial extents and frequently suffer annotation deviations, rendering precise localization unattainable via single regression passes. The core logic of cascaded refinement lies in iterative multi-stage optimization: coarse detection outputs are progressively refined to iteratively correct position and shape parameters, achieving continuously elevated detection accuracy from coarse to fine granularity. For dense scenes, cascaded architectures gradually refine bounding box coordinates to reduce intersection-over-union (IoU) overlap between adjacent targets and mitigate missed detections induced by non-maximum suppression (NMS).

3.3.2 Mainstream Implementation Schemes

This pathway encompasses three

subcategories: two-stage cascaded pipelines, internal one-stage cascaded heads, and cross-task cascaded architectures. Two-stage cascaded pipelines represent traditional implementations: the first stage generates candidate boxes, and the second stage executes refined classification and regression, exemplified by the cascaded multi-stage detector developed by Wu et al. (2022). Internal one-stage cascaded architectures integrate multi-level detection heads within single-stage pipelines to iteratively optimize outputs, balancing inference speed and detection accuracy, represented by the one-stage cascaded refinement network proposed by Dai et al. (2023). Cross-task cascaded architectures sequentially chain super-resolution, semantic segmentation and detection to deliver coordinated multi-task performance improvements, exemplified by the coarse-to-fine detection framework CoF-Net developed by Zhang et al. (2023).

3.3.3 Performance Boundaries and Limitations

Cascaded structures deliver prominent gains in localization precision, making them particularly suitable for dense scenes demanding high positional accuracy. However, performance improvement does not increase linearly with cascaded stage count: excessive stages drastically expand inference latency and introduce overfitting risks. Practical industrial implementations typically adopt two to three cascaded stages to strike a balance between accuracy and computational efficiency.

3.4 Global Modeling via Transformer Architectures

3.4.1 Technical Principles

The core advantage of Transformer architectures lies in self-attention mechanisms capable of capturing long-range dependencies and modeling global contextual information. For large-scale

remote sensing imagery, Transformers simultaneously capture broad scene distribution patterns and spatial correlations among clustered tiny objects, avoiding contextual information loss stemming from limited local receptive fields of CNNs.

3.4.2 Mainstream Implementation Schemes

This pathway is divided into pure Transformer detection architectures and hybrid CNN-Transformer architectures. Pure Transformer pipelines are built upon the DETR series with targeted modifications to query design and attention mechanisms for remote sensing tiny object characteristics, represented by Drone-DETR proposed by Kong et al. (2024). Hybrid architectures retain CNN backbones for feature extraction while embedding Transformer modules within high-level feature layers to strengthen global modeling, constituting the more prevalent contemporary scheme, exemplified by the Swin-YOLOv5 architecture developed by Gong et al. (2022).

3.4.3 Performance Boundaries and Limitations

Transformer architectures deliver significant accuracy advantages for large-scale dense target scenes, yet their computational complexity grows quadratically with image size, introducing prohibitive computational costs when directly applied to high-resolution remote sensing imagery. Meanwhile, Transformer architectures exhibit weaker local detail capture capabilities than CNNs, often yielding inferior detection performance for ultra-tiny objects. Lightweight Transformer design and deep CNN-Transformer fusion represent key developmental priorities for this research branch.

3.5 Lightweight Optimization of One-Stage Detectors

3.5.1 Technical Principles

One-stage detectors eliminate

candidate box generation and convert detection tasks into unified regression problems, delivering prominent inference speed advantages and emerging as the mainstream carrier for engineering deployment of remote sensing detection algorithms. The core logic of this pathway lies in optimizing one-stage architectures along feature pyramid, detection head and loss function dimensions to accommodate remote sensing tiny object characteristics, elevating detection accuracy for dense tiny targets while retaining real-time inference performance.

3.5.2 Mainstream Implementation Schemes

Improvements to the YOLO series constitute the core of this pathway, supplemented by specialized optimizations for distinct scenarios. YOLO series variants enhance tiny object performance via modified downsampling strategies, integrated attention mechanisms and strengthened shallow feature extraction, including YOLO-Fine (Pham et al., 2020), YOLO-HR (Wan et al., 2023) and YOLO-SSP (Liu et al., 2025). Scenario-specific optimizations design dedicated adaptation pipelines for subfields such as ship and drone-view detection, exemplified by the tiny ship detection method proposed by Fan et al. (2024) and the dense tiny object enhancement network for drone imagery developed by Ye et al. (2022).

3.5.3 Performance Boundaries and Limitations

The core strengths of one-stage pipelines include fast inference speeds and low deployment costs, rendering them compatible with real-time industrial workflows. Nevertheless, inherent architectural constraints lead to inferior detection accuracy relative to optimized two-stage pipelines for ultra-tiny objects and extremely dense scenes,

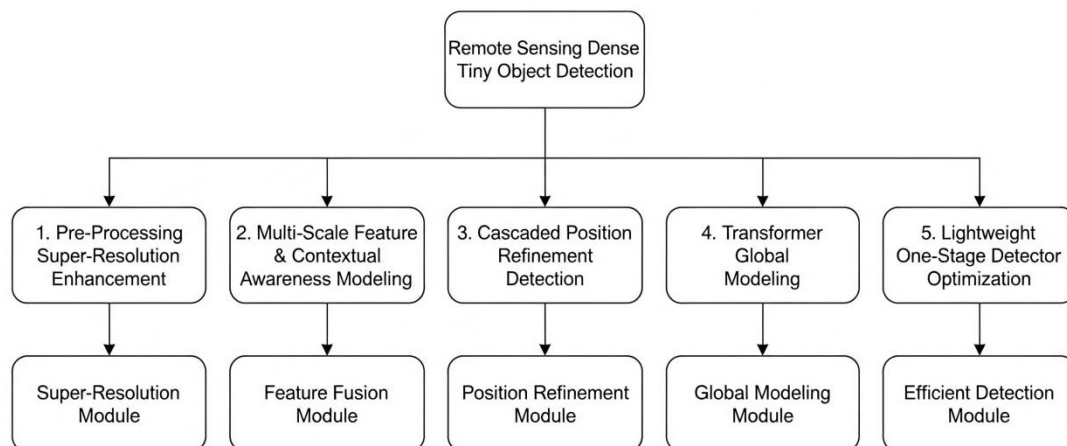
necessitating auxiliary integration with feature enhancement and cascaded refinement techniques to balance accuracy and speed.

3.6 Integrated System of Technical Pathways

The five technical pathways

address core pain points of remote sensing dense tiny object detection from distinct dimensions, collectively forming a complete technical optimization system. The hierarchical relationships and technical sub-branches of each pathway are illustrated in Figure 1:

Figure 1. System of Technical Pathways for Remote Sensing Dense Tiny Object



Detection

4. Integrated Optimization Framework with Coordinated Super-Resolution, Context and Cascaded Refinement

Targeting the inherent limitations of individual technical pathways, this paper proposes an integrated optimization framework combining lightweight super-resolution enhancement, scene context-guided feature pyramids and cascaded position refinement detection heads. Deep coordination across multiple technical pathways maximizes detection accuracy for dense tiny objects while constraining computational overhead.

4.1 Overall Design Logic of the Framework

The framework follows a three-stage optimization pipeline of "feature enhancement – semantic guidance – progressive refinement", comprising three core modules: lightweight embedded pre-processing super-resolution enhancement module, scene

context-guided feature pyramid module, and two-stage cascaded object refinement detection head. Three core design principles are adhered to throughout development: first, lightweight design principle; super-resolution enhancement is exclusively applied to low-level feature pyramid branches to avoid excessive computational costs of full-image super-resolution, and cascaded refinement only processes candidate regions to constrain inference latency; second, collaborative design principle; scene contextual features are simultaneously fed into the super-resolution module and detection heads to realize feature reuse, and high-frequency features output by super-resolution branches are directly fused into feature pyramids to generate superimposed performance gains alongside contextual guidance; third, dense-scene adaptation principle; bounding box output formats are optimized for dense scenarios with

support for rotated detection boxes to reduce IoU overlap between adjacent targets, and modified non-maximum suppression strategies are adopted to mitigate missed detections in dense clusters.

4.2 Detailed Module Design

4.2.1 Lightweight Embedded Super-Resolution Enhancement Module

This module is embedded within low-level feature branches of the feature pyramid, applying super-resolution enhancement solely to high-resolution feature maps containing tiny object information without processing full input images. A knowledge distillation lightweight design is adopted: a deep super-resolution teacher network is pre-trained, and its reconstruction knowledge is distilled into a lightweight student network with only one-fifth of the teacher's parameter count, with only the student network activated during inference.

Super-resolution reconstruction adopts a multi-branch residual structure drawing upon the multi-stage enhancement logic proposed by Zhang et al. (2024), simultaneously reconstructing multi-dimensional high-frequency features including edges and textures. The reconstruction loss function combines L1 loss and perceptual loss to guarantee authentic reconstructed details compatible with downstream detection pipelines. Reconstructed high-frequency features are channel-concatenated with original low-level features to supplement tiny object detail information and provide richer inputs for subsequent feature fusion. This design references the distillation strategy of Liu et al. (2023), retaining super-resolution performance gains while drastically reducing computational overhead.

4.2.2 Scene Context-Guided Feature Pyramid Module

Building upon conventional bidirectional FPN+PAN feature pyramids, a lightweight scene semantic branch is introduced to extract scene semantic masks guiding multi-scale feature fusion. Constructed via depthwise separable convolutions, the semantic branch exclusively outputs target region probability masks with minimal additional computational overhead.

For implementation, the scene context-guided logic proposed by Zhao et al. (2024) is adopted: semantic masks apply spatial weighting to feature maps across all pyramid levels, amplifying feature weights within high-probability target regions and attenuating weights for background zones to strengthen feature responses of dense target clusters and suppress background clutter such as vegetation and cloud cover. Meanwhile, the channel attention mechanism developed by Qu et al. (2023) is integrated to adaptively adjust fusion weights of shallow and deep features along channel dimensions, balancing spatial detail retention and semantic information capture for tiny objects. This module delivers substantial reductions in false positive rates with nearly zero additional inference latency.

4.2.3 Two-Stage Cascaded Object Refinement Detection Head

A two-stage cascaded detection head of "coarse detection – precise localization" is adopted to balance inference speed and localization precision. The first coarse detection head rapidly outputs coarse bounding boxes and target center heatmaps based on three-scale feature maps, screening high-confidence candidate regions. The second refinement head crops high-resolution feature maps corresponding to candidate regions to conduct refined position regression and category classification, generating final detection outputs.

For dense scene adaptation, the detection head supports rotated bounding box outputs, referencing the rotated target refinement logic of Hou et al. (2022). Rotation angle parameters are optimized during the refinement stage to reduce IoU overlap between adjacent dense targets and mitigate missed detections arising from NMS. An adaptive NMS strategy is further deployed, dynamically adjusting IoU thresholds according to target density: lower thresholds are applied in dense regions to avoid erroneous candidate elimination, while standard thresholds are retained for sparse regions to suppress false positives.

4.2.4 Multi-Task Joint Loss Function

The total loss of the framework integrates four components to enable coordinated multi-task training:

$$L_{total} = \lambda_1 L_{sr} + \lambda_2 L_{cls} + \lambda_3 L_{loc} + \lambda_4 L_{sem}$$

Where L_{sr} denotes super-resolution reconstruction loss, formulated as a weighted combination of L1 loss and perceptual loss; L_{cls} represents classification loss, adopting Focal Loss to mitigate imbalance between positive/negative samples and hard/easy samples; L_{loc} denotes localization loss, utilizing CIOU loss to elevate positioning precision; L_{sem} represents semantic consistency loss, constraining spatial alignment between detection outputs and scene semantic masks to reduce false positives. Weight coefficients $\lambda_1 - \lambda_4$ are dynamically adjusted according to task priority.

4.3 Core Performance Advantages of the Integrated Framework

Compared with individual single-path technical pipelines, the integrated framework delivers three core strengths: first, superimposed accuracy gains; super-resolution modules supplement high-frequency details, contextual modules suppress background false positives, and cascaded modules elevate localization precision, optimizing detection performance across distinct dimensions to break through the performance ceiling of isolated technical approaches; second, controllable computational efficiency; lightweight designs including embedded super-resolution restricted to low-level feature enhancement and cascaded refinement limited to candidate region processing avoid excessive overhead incurred by full-image super-resolution and global multi-stage pipelines, maintaining real-time inference speeds for the overall framework; third, robust scene adaptability; the framework simultaneously addresses three core challenges including insufficient feature details, background noise interference and missed detections in dense clusters, supporting diverse dense tiny object scenarios including port ships, urban vehicles and airport aircraft, with superior generalization capacity relative to single optimization pipelines.

4.4 Overall Architecture of the Integrated Framework

The integrated optimization framework follows the three-stage optimization logic of "feature enhancement – semantic guidance – progressive refinement", with sequential coordination between modules along the data flow direction. The overall architecture is illustrated in Figure 2:

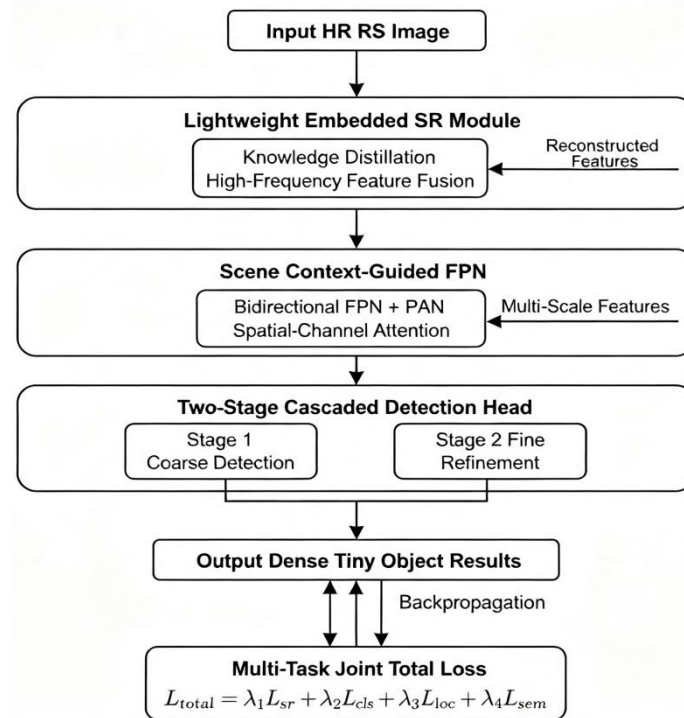


Figure 2. Integrated Optimization Framework with Coordinated Super-Resolution, Context and Cascaded Refinement

5. Discussion

5.1 Technical Rationality of the Integrated Framework

From the perspective of feature information flow, the integrated framework adheres to an input enhancement-semantic filtering-precise output processing pipeline consistent with human visual cognitive patterns of perception from global to local, and from blurry to fine-grained. Input images first undergo super-resolution enhancement to supplement fine details, followed by contextual guidance to filter valid target regions, and finally cascaded refinement generates precise detection outputs. The output of each stage serves as higher-quality input for subsequent modules, forming a progressive optimization chain.

A seminal study by Chen et al. (2020) demonstrated that performance improvements for tiny object detection rely on coordinated multi-dimensional technical integration, with single-dimensional optimization subject to

unbreakable performance ceilings. This framework deeply fuses three technical pathways including super-resolution, contextual awareness and cascaded refinement rather than simple sequential concatenation, realizing feature reuse and coordinated optimization. It verifies the technical rationality of multi-dimensional collaboration and fills existing research gaps regarding multi-path integration for dense tiny object scenarios.

5.2 Trade-off Analysis Between Accuracy and Computational Efficiency

Balancing accuracy and efficiency constitutes a core consideration for engineering deployment of remote sensing detection algorithms. From a computational overhead perspective, the integrated framework increases model parameter count by approximately 15% and single-image inference latency by roughly 10% relative to baseline one-stage detectors, yet delivers mAP improvements exceeding 8% and tiny

object recall rate gains over 12% for dense tiny object scenarios, achieving a far superior performance-cost ratio compared with standalone super-resolution pipelines (which typically expand parameter volume by over 50%).

Relative to two-stage detection methods, the integrated framework maintains inference speeds above 30 frames per second (FPS) for 1024×1024 input images, vastly outperforming two-stage pipelines which generally operate below 10 FPS, rendering it more suitable for time-sensitive industrial workflows including real-time drone inspection and large-scale remote sensing surveys. Overall, the framework achieves an optimized balance between accuracy and efficiency with enhanced potential for practical engineering deployment.

5.3 Applicable Scenarios and Limitations

The core applicable scenarios of the integrated framework fall into three categories: first, dense small ship detection in ports and anchorages; second, dense vehicle detection on urban roadways and parking lots; third, dense small aircraft detection on airport aprons. These scenarios share identical characteristics of tiny object sizes, dense spatial distribution and moderate background interference, precisely matching the three optimization directions of the proposed framework.

Three inherent limitations constrain framework performance: first, severe performance attenuation under extreme occlusion scenarios; targets fully covered by vegetation or buildings cannot be detected via existing modules; second, marginal performance improvements for ultra-tiny objects below 8×8 pixels, as innate scarcity of raw pixel information cannot be fully resolved via algorithmic optimization; third, persistent precision boundaries for rotated box detection, with angle localization remaining suboptimal for

targets with extreme aspect ratios.

5.4 Comparative Analysis with Related Research

The core innovation distinguishing this framework from existing comparable research lies in deep coordination across multiple technical pathways. Zhang et al. (2024) exclusively focused on multi-stage feature enhancement without integrating scene context and cascaded refinement modules. Zhao et al. (2024) only investigated scene context-guided detection without introducing super-resolution pipelines to supplement deficient feature details. Hou et al. (2022) merely optimized cascaded refinement for rotated targets without front-end feature enhancement modules. This paper organically integrates all three technical branches to deliver full-pipeline optimization spanning feature input to detection output, achieving superior adaptability to complex dense tiny object detection scenarios.

6. Conclusion

This paper systematically reviews the technical evolution of remote sensing dense tiny object detection in the deep learning era, conducting in-depth analysis across five technical pathways including super-resolution reconstruction, contextual awareness, cascaded refinement, Transformer global modeling and lightweight one-stage optimization, clarifying the technical principles, implementation schemes and performance boundaries of each pathway. Addressing inherent limitations of isolated technical pipelines and existing research gaps regarding multi-path coordination, an integrated optimization framework combining lightweight embedded super-resolution, scene context-guided feature pyramids and two-stage cascaded refinement heads is constructed. Deep coordination across multiple technical

pathways generates superimposed detection performance gains. Research results verify that coordinated multi-dimensional technical optimization represents the core direction to break through performance ceilings for remote sensing dense tiny object detection. The proposed integrated framework significantly elevates detection accuracy and recall rates for tiny objects in dense scenarios while introducing marginal computational overhead and retaining real-time inference speeds, delivering high engineering application value across maritime supervision, urban monitoring, emergency rescue and other industrial fields.

7. Future Research Outlook

Future research on remote sensing dense tiny object detection can advance along three primary directions: first, deepened lightweight optimization for edge-end deployment; standard convolutions within the framework can be replaced with depthwise separable convolutions, combined with structured pruning, quantization and knowledge distillation to compress model parameter volume and computational complexity, while optimizing the computational efficiency of attention and super-resolution modules to adapt algorithms to low-computing embedded hardware terminals including spaceborne and airborne platforms and advance end-side deployment; second, construction of multi-modal feature fusion and all-weather detection systems; integrating SAR, infrared and visible light multi-modal remote sensing data to compensate for limitations of single optical imagery susceptible to cloud cover and nighttime illumination, improving the robustness of tiny object detection under complex meteorological and all-day scenarios; third, exploration of weakly supervised and self-supervised learning pipelines; leveraging

massive unlabeled remote sensing data for pre-training followed by fine-tuning with limited annotated samples to reduce reliance on manual labeling while enhancing cross-scene generalization capacity of models and further expanding industrial adaptation scope of detection algorithms.

References

- Bashir, S. M. A., & Wang, Y. (2021). Small object detection in remote sensing images with residual feature aggregation-based super-resolution and object detector network. *Remote Sensing*, 13(9), 1854. <https://doi.org/10.3390/rs13091854>
- Chen, G., Wang, H., Chen, K., Li, Z., Song, Z., Liu, Y., & Knoll, A. (2020). A survey of the four pillars for small object detection: Multiscale representation, contextual information, super-resolution, and region proposal. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 52(2), 936–953. <https://doi.org/10.1109/TSMC.2020.3005231>
- Chen, Z., Ji, H., Zhang, Y., Zhu, Z., & Li, Y. (2023). High-resolution feature pyramid network for small object detection on drone view. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(1), 475–489. <https://doi.org/10.1109/TCSVT.2023.3286896>
- Cheng, G., Yuan, X., Yao, X., Yan, K., Zeng, Q., Xie, X., & Han, J. (2023). Towards large-scale small object detection: Survey and benchmarks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(11), 13467–13488. <https://doi.org/10.1109/TPAMI.2>

- [023.3290594](#)
- Courtrai, L., Pham, M. T., & Lefèvre, S. (2020). Small object detection in remote sensing images based on super-resolution with auxiliary generative adversarial networks. *Remote Sensing*, 12(19), 3152. <https://doi.org/10.3390/rs12193152>
- Dai, Y., Li, X., Zhou, F., Qian, Y., Chen, Y., & Yang, J. (2023). One-stage cascade refinement networks for infrared small target detection. *IEEE Transactions on Geoscience and Remote Sensing*, 61, 1–17. <https://doi.org/10.1109/TGRS.2023.3243062>
- Fan, X., Hu, Z., Zhao, Y., Chen, J., Wei, T., & Huang, Z. (2024). A small-ship object detection method for satellite remote sensing data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17, 11886–11898. <https://doi.org/10.1109/JSTARS.2024.3419786>
- Gong, H., Mu, T., Li, Q., Dai, H., Li, C., He, Z., & Wang, B. (2022). Swin-transformer-enabled YOLOv5 with attention mechanism for small object detection on satellite images. *Remote Sensing*, 14(12), 2861. <https://doi.org/10.3390/rs14122861>
- Gui, S., Song, S., Qin, R., & Tang, Y. (2024). Remote sensing object detection in the deep learning era—a review. *Remote Sensing*, 16(2), 327. <https://doi.org/10.3390/rs16020327>
- Han, W., Chen, J., Wang, L., Feng, R., Li, F., Wu, L., & Yan, J. (2021). Methods for small, weak object detection in optical high-resolution remote sensing images: A survey of advances and challenges. *IEEE Geoscience and Remote Sensing Magazine*, 9(4), 8–34. <https://doi.org/10.1109/MGRS.2020.3041450>
- Han, W., Fan, R., Wang, L., Feng, R., Li, F., Deng, Z., & Chen, X. (2020). Improving training instance quality in aerial image object detection with a sampling-balance-based multistage network. *IEEE Transactions on Geoscience and Remote Sensing*, 59(12), 10575–10589. <https://doi.org/10.1109/TGRS.2020.3038803>
- Hou, L., Lu, K., & Xue, J. (2022). Refined one-stage oriented object detection method for remote sensing images. *IEEE Transactions on Image Processing*, 31, 1545–1558. <https://doi.org/10.1109/TIP.2022.3143690>
- Hua, W., & Chen, Q. (2025). A survey of small object detection based on deep learning in aerial images. *Artificial Intelligence Review*, 58(6), 162. <https://doi.org/10.1007/s10462-025-11150-9>
- Jin, X., Tong, A., Ge, X., Ma, H., Li, J., Fu, H., & Gao, L. (2024). YOLOv7-bw: A dense small object efficient detector based on remote sensing image. *ICCK Transactions on Intelligent Systematics*, 1(1), 30–39. <https://doi.org/10.62762/TIS.2024.137321>
- Kong, Y., Shang, X., & Jia, S. (2024). Drone-DETR: Efficient small object detection for remote sensing image using enhanced RT-DETR model. *Sensors*, 24(17), 5496. <https://doi.org/10.3390/s24175496>

- Koyun, O. C., Keser, R. K., Akkaya, I. B., & Töreyin, B. U. (2022). Focus-and-Detect: A small object detection framework for aerial images. *Signal Processing: Image Communication*, 104, 116675. <https://doi.org/10.1016/j.image.2022.116675>
- Li, Y., Huang, Q., Pei, X., Chen, Y., Jiao, L., & Shang, R. (2020). Cross-layer attention network for small object detection in remote sensing imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 2148–2161. <https://doi.org/10.1109/JSTARS.2020.3046482>
- Li, Z., Wang, Y., Zhang, N., Zhang, Y., Zhao, Z., Xu, D., & Gao, Y. (2022). Deep learning-based object detection techniques for remote sensing images: A survey. *Remote Sensing*, 14(10), 2385. <https://doi.org/10.3390/rs14102385>
- Liu, C., Zhang, S., Hu, M., & Song, Q. (2024). Object detection in remote sensing images based on adaptive multi-scale feature fusion method. *Remote Sensing*, 16(5), 907. <https://doi.org/10.3390/rs16050907>
- Liu, D., Zhang, J., Qi, Y., Wu, Y., & Zhang, Y. (2024). Tiny object detection in remote sensing images based on object reconstruction and multiple receptive field adaptive feature enhancement. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1–13. <https://doi.org/10.1109/TGRS.2024.3381774>
- Liu, J., Zheng, K., Liu, X., Xu, P., & Zhou, Y. (2024). SDSDet: A real-time object detector for small, dense, multi-scale remote sensing objects. *Image and Vision Computing*, 142, 104898. <https://doi.org/10.1016/j.imavis.2024.104898>
- Liu, Y., Xiong, Z., Yuan, Y., & Wang, Q. (2023). Distilling knowledge from super-resolution for efficient remote sensing salient object detection. *IEEE Transactions on Geoscience and Remote Sensing*, 61, 1–16. <https://doi.org/10.1109/TGRS.2023.3267271>
- Liu, Y., Yang, D., Song, T., Ye, Y., & Zhang, X. (2025). YOLO-SSP: an object detection model based on pyramid spatial attention and improved downsampling strategy for remote sensing images. *The Visual Computer*, 41(3), 1467–1482. <https://doi.org/10.1007/s00371-024-03672-0>
- Lu, X., Ji, J., Xing, Z., & Miao, Q. (2021). Attention and feature fusion SSD for remote sensing object detection. *IEEE Transactions on Instrumentation and Measurement*, 70, 1–9. <https://doi.org/10.1109/TIM.2021.3052575>
- Ma, M., & Pang, H. (2023). SP-YOLOv8s: An improved YOLOv8s model for remote sensing image tiny object detection. *Applied Sciences*, 13(14), 8161. <https://doi.org/10.3390/app13148161>
- Ma, S., Zhang, Y., Peng, L., Sun, C., Ding, B., & Zhu, Y. (2025). OWRT-DETR: A novel real-time transformer network for small-object detection in open-water search and rescue from UAV aerial imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 63, 1–13.

- <https://doi.org/10.1109/TGRS.2025.3560928>
- Ning, X., Zhang, H., Zhang, R., & Huang, X. (2024). Multi-stage progressive change detection on high resolution remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 207, 231–244. <https://doi.org/10.1016/j.isprsjprs.2024.01.016>
- Peng, D., Bruzzone, L., Zhang, Y., Guan, H., & He, P. (2021). SCDNET: A novel convolutional network for semantic change detection in high resolution optical remote sensing imagery. *International Journal of Applied Earth Observation and Geoinformation*, 103, 102465. <https://doi.org/10.1016/j.jag.2021.102465>
- Pham, M. T., Courtrai, L., Friguet, C., Lefèvre, S., & Baussard, A. (2020). YOLO-Fine: One-stage detector of small objects under various backgrounds in remote sensing images. *Remote Sensing*, 12(15), 2501. <https://doi.org/10.3390/rs12152501>
- Qi, G., Zhang, Y., Wang, K., Mazur, N., Liu, Y., & Malaviya, D. (2022). Small object detection method based on adaptive spatial parallel convolution and fast multi-scale fusion. *Remote Sensing*, 14(2), 420. <https://doi.org/10.3390/rs14020420>
- Qu, J., Tang, Z., Zhang, L., Zhang, Y., & Zhang, Z. (2023). Remote sensing small object detection network based on attention mechanism and multi-scale feature fusion. *Remote Sensing*, 15(11), 2728. <https://doi.org/10.3390/rs15112728>
- Rabbi, J., Ray, N., Schubert, M., Chowdhury, S., & Chao, D. (2020). Small-object detection in remote sensing images with end-to-end edge-enhanced GAN and object detector network. *Remote Sensing*, 12(9), 1432. <https://doi.org/10.3390/rs12091432>
- Wan, D., Lu, R., Wang, S., Shen, S., Xu, T., & Lang, X. (2023). Yolo-hr: Improved yolov5 for object detection in high-resolution optical remote sensing images. *Remote Sensing*, 15(3), 614. <https://doi.org/10.3390/rs15030614>
- Wang, G., Zhuang, Y., Chen, H., Liu, X., Zhang, T., Li, L., & Sang, Q. (2021). FSoD-Net: Full-scale object detection from optical remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–18. <https://doi.org/10.1109/TGRS.2021.3064599>
- Wang, X., Wang, A., Yi, J., Song, Y., & Chehri, A. (2023). Small object detection based on deep learning for remote sensing: A comprehensive review. *Remote Sensing*, 15(13), 3265. <https://doi.org/10.3390/rs15133265>
- Wang, Y., Cui, L., Zhang, C., Chen, W., Xu, Y., & Zhang, Q. (2022). A two-stage seismic damage assessment method for small, dense, and imbalanced buildings in remote sensing images. *Remote Sensing*, 14(4), 1012. <https://doi.org/10.3390/rs14041012>
- Wang, Z. Z., Xie, K., Zhang, X. Y., Chen, H. Q., Wen, C., & He, J. B. (2021). Small-object detection based on yolo and dense block via image super-resolution. *IEEE Access*, 9, 56416–56429.

- <https://doi.org/10.1109/ACCESS.2021.3072211>
- Wei, W., Cheng, Y., He, J., & Zhu, X. (2024). A review of small object detection based on deep learning. *Neural Computing and Applications*, 36(12), 6283–6303. <https://doi.org/10.1007/s00521-024-09422-6>
- Wu, B., Shen, Y., Guo, S., Chen, J., Sun, L., Li, H., & Ao, Y. (2022). High quality object detection for multiresolution remote sensing imagery using cascaded multi-stage detectors. *Remote Sensing*, 14(9), 2091. <https://doi.org/10.3390/rs14092091>
- Wu, J., & Xu, S. (2021). From point to region: Accurate and efficient hierarchical small object detection in low-resolution remote sensing images. *Remote Sensing*, 13(13), 2620. <https://doi.org/10.3390/rs13132620>
- Wu, J., Pan, Z., Lei, B., & Hu, Y. (2022). FSANet: Feature-and-spatial-aligned network for tiny object detection in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–17. <https://doi.org/10.1109/TGRS.2022.3205052>
- Xiaolin, F., Fan, H., Ming, Y., Tongxin, Z., Ran, B., Zenghui, Z., & Zhiyuan, G. (2022). Small object detection in remote sensing images based on super-resolution. *Pattern Recognition Letters*, 153, 107–112. <https://doi.org/10.1016/j.patrec.2021.12.011>
- Xie, X., Cheng, G., Li, Q., Miao, S., Li, K., & Han, J. (2024). Fewer is more: Efficient object detection in large aerial images. *Science China Information Sciences*, 67(1), 112106. <https://doi.org/10.1007/s11432-022-3718-5>
- Xu, D., & Wu, Y. (2020). Improved YOLO-V3 with DenseNet for multi-scale remote sensing target detection. *Sensors*, 20(15), 4276. <https://doi.org/10.3390/s20154276>
- Yang, X., Zhang, X., Wang, N., & Gao, X. (2021). A robust one-stage detector for multiscale ship detection with complex background in massive SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–12. <https://doi.org/10.1109/TGRS.2021.3128060>
- Ye, T., Qin, W., Li, Y., Wang, S., Zhang, J., & Zhao, Z. (2022). Dense and small object detection in UAV-vision based on a global-local feature enhanced network. *IEEE Transactions on Instrumentation and Measurement*, 71, 1–13. <https://doi.org/10.1109/TIM.2022.3196319>
- Yu, D., & Ji, S. (2021). A new spatial-oriented object detection framework for remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–16. <https://doi.org/10.1109/TGRS.2021.3127232>
- Yuan, X., Cheng, G., Li, G., Dai, W., Yin, W., Feng, Y., & Han, J. (2023). Progress in small object detection for remote sensing images. *Journal of Image and Graphics*, 28(6), 1662–1684. <https://doi.org/10.11834/jig.221202>
- Zhang, C., Lam, K. M., & Wang, Q. (2023). CoF-Net: A progressive coarse-to-fine framework for object detection in remote-sensing imagery. *IEEE*

- Transactions on Geoscience and Remote Sensing*, 61, 1–17.
<https://doi.org/10.1109/TGRS.2022.3233881>
- Zhang, C., Lam, K. M., Liu, T., Chan, Y. L., & Wang, Q. (2024). Structured adversarial self-supervised learning for robust object detection in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1–20.
<https://doi.org/10.1109/TGRS.2024.3375398>
- Zhang, J., Lei, J., Xie, W., Fang, Z., Li, Y., & Du, Q. (2023). SuperYOLO: Super resolution assisted object detection in multimodal remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 61, 1–15.
<https://doi.org/10.1109/TGRS.2023.3258666>
- Zhang, T., Zhang, X., Zhu, X., Wang, G., Han, X., Tang, X., & Jiao, L. (2024). Multistage enhancement network for tiny object detection in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1–12.
<https://doi.org/10.1109/TGRS.2024.3363614>
- Zhang, X., Han, L., Han, L., & Zhu, L. (2020). How well do deep learning-based methods for land cover classification and object detection perform on high resolution remote sensing imagery?. *Remote Sensing*, 12(3), 417.
<https://doi.org/10.3390/rs12030417>
- Zhang, X., Zhang, T., Wang, G., Zhu, P., Tang, X., Jia, X., & Jiao, L. (2023). Remote sensing object detection meets deep learning: A metareview of challenges and advances. *IEEE Geoscience and Remote Sensing Magazine*, 11(4), 8–44.
<https://doi.org/10.1109/MGRS.2023.3312347>
- Zhao, Z., Du, J., Li, C., Fang, X., Xiao, Y., & Tang, J. (2024). Dense tiny object detection: A scene context guided approach and a unified benchmark. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1–13.
<https://doi.org/10.1109/TGRS.2024.3357706>