

DOI: 10.12382/bgxb.2025.0938



面向空基侦察平台的轻量化多通道 军事小目标检测方法

赵阳, 王宪楠*, 郭策安, 张健

(沈阳理工大学 装备工程学院, 辽宁 沈阳 110159)

摘要: 空基侦察平台在复杂战场环境下军事目标通常具有尺寸小、分布密集、背景复杂及尺度变化显著等特征, 加之空基与无人机平台算力与载荷约束, 现有检测方法难以同时满足精度与实时性要求。为此, 提出一种面向空基与遥感场景的多通道轻量化小目标检测方法。通过改进网络主干结构, 引入灰度辅助通道构建彩色-灰度(Red Green Blue Gray, RGB-Gray)四通道输入, 以增强对小目标纹理与边缘细节的特征提取能力; 在特征融合阶段启用高分辨率的 P2 检测层, 并裁剪低分辨率的 P5 层, 在保证精度的同时降低模型复杂度; 设计基于门控的自适应轻量双分支 C3k2 模块(Gated-based Adaptive Lightweight Dual-branch Cross Stage Partial Bottleneck with 2 Convolutions, Conditional-C3k2)模块用于替代传统特征提取单元, 引入门控机制动态融合主分支与捷径路径, 实现特征信息的自适应分配; 进一步采用轻量化跨阶段聚合结构以提升多尺度特征整合效率; 此外, 改进动态检测头, 并提出面积加权损失(Varifocal-SIoU-Area-weighted Loss, VSA-Loss), 以优化边界框回归与分类性能。实验结果表明: 所提 GCSD-YOLOv11n 算法在 VisDrone2019 与军用车辆细粒度检测遥感图像数据集 (Military Vehicle Remote Sensing Dataset, MVRSD) 上平均精度均值较基准模型分别提升了 13.0%与 7.4%, 其中 MVRSD 全类精度达到 86.4%, 同时模型参数量大幅压缩 29.4%; 该算法在 MAR20 军事飞机数据集上亦展现出较强的泛化能力, 验证了其对于复杂军事成像场景下细粒度目标的探测鲁棒性。

关键词: 小目标检测; 轻量化网络; 多通道特征融合; 单阶段目标检测; 无人机侦察

中图分类号: TP391.41

文献标志码: A

文章编号: 1000-1093(202x)xx-250938-20

A Multi-Channel Lightweight Small Object Detection Algorithm for Low-Altitude Reconnaissance Scenarios

ZHAO Yang, WANG Xiannan*, GUO Ce'an, ZHANG Jian

(School of Equipment Engineering, Shenyang Ligong University, Shenyang 110159, Liaoning, China)

Abstract: Military targets in reconnaissance and remote sensing are characterized by small sizes, dense distributions, complex backgrounds, and significant scale variations, while limited computing resources of unmanned aerial vehicle (UAV) platforms hinder achieving both high accuracy and real-time performance. This paper proposes a multi-channel lightweight detection algorithm for such scenarios. A grayscale auxiliary channel is introduced to construct a red green blue gray (RGB-Gray) four-channel input, enhancing fine-grained texture and edge feature extraction. High-resolution P2 is enabled and low-resolution P5 is pruned to reduce model complexity. A gated-based adaptive lightweight dual-branch cross stage partial bottleneck with 2 convolutions (Conditional-C3k2) module with dynamic gating is designed for adaptive feature fusion, and a lightweight cross-stage aggregation structure is adopted to improve multi-scale integration efficiency. An enhanced dynamic head and a varifocal-SIoU-area-weighted loss (VSA-Loss) function with an area-weighted mechanism optimize performance. GCSD-YOLOv11n improves Mean Average Precision by 13.0% and 7.4% on VisDrone2019 and military vehicle remote sensing dataset (MVRSD), respectively, with MVRSD accuracy reaching 86.4% and parameters reducing by 29.4%. Strong generalization on the MAR20 dataset verifies robustness in complex military imaging, providing an effective solution for real-time airborne

reconnaissance tasks.

Keywords: small object detection; lightweight network; multi-channel feature fusion; you only look once; unmanned aerial vehicle reconnaissance

0 引言

随着无人机、导弹和星载平台在侦察与打击任务中的广泛部署,战场态势感知正逐渐向实时化、精细化方向发展^[1-3]。空基与低空侦察场景中,待识别目标普遍呈现尺寸微小、视角变化显著、背景复杂以及光照条件不稳定等特点,导致目标特征表达不足,易产生漏检与定位偏差。同时,平台受限于算力、能耗及载荷约束,复杂检测算法难以直接应用于机载、弹载等资源受限平台,使得高精度、轻量化与实时性之间的矛盾愈加突出。

深度学习的发展推动了目标检测技术的快速演进,但现有方法在军事小目标场景中仍面临局限。例如,当目标尺度仅占据数十像素时,边缘纹理极易被淹没于多尺度特征融合过程;在城市与战场环境中频繁出现的遮挡、阴影与密集场景,使得算法难以保持稳定的检测响应;在夜间与低照度条件下,目标亮度对比降低、噪声增多,进一步削弱了检测算法的判别能力。此外,由于军事侦察任务覆盖近距穿越机、低空巡飞至高空遥感等多种平台,图像呈现出显著的尺度跨度与成像差异,这对检测器的跨场景泛化能力提出了更高要求。

针对上述问题,研究者从特征增强、主干网络优化、多尺度融合与轻量化结构等方面提出了多种改进方案。然而,许多方法模型复杂度较高或依赖大型特征金字塔结构,难以兼顾军事任务中对机载实时处理的需求。因此,亟需一种能够同时提升小目标特征表达能力、强化多尺度关联并兼顾部署效率的检测框架。为此,本文构建了一种面向空基与

遥感场景的轻量化多通道小目标检测算法,以第11代单阶段目标检测纳型(You Only Look Once version 11 Nano, YOLOv11n)算法为基础,通过四通道输入增强细粒度特征响应、改进特征提取模块以增强语义表达能力,并采用轻量化特征融合结构与高效检测头,实现在复杂环境下对小目标的高精度识别。在保持轻量化与实时性的前提下,进一步提升算法对弱光、密集遮挡与尺度变化场景的适应性。如图1所示,本文的主要贡献如下:

1) 构建彩色-灰度(Red Green Blue Gray, RGB-Gray)四通道输入:通过引入灰度辅助通道与高分辨率检测层,强化小目标的纹理与边缘响应能力,提高弱光与密集场景的检测稳定性。

2) 设计基于门控的自适应轻量化C3k2 (Gated-based Adaptive Lightweight Dual-branch Cross Stage Partial Bottleneck with 2 Convolutions, Conditional-C3k2)结构:利用门控机制实现特征的动态分配,在不增加大量参数的情况下提升中层特征表达能力。

3) 提出轻量化多尺度融合结构与高效检测头:采用轻量化跨阶段聚合(Lightweight Cross-Stage Aggregation, LCSA)结构与动态检测头(Dynamic Head, DyHead)结合第4代可变形卷积网络(Dynamic Head with Deformable Convolutional Networks v4, DyHead-DCNv4),提升多尺度交互与小目标定位精度,同时显著降低推理延迟,适用于资源受限平台。

4) 构建适用于小目标的面积加权损失(Varifocal-SIoU-Area-weighted Loss, VSA-Loss):通过引入面积加权机制与Scylla交并比(Scylla Intersection over Union, Siou)回归项,提高小目标的定位精度与模型收敛稳定性。

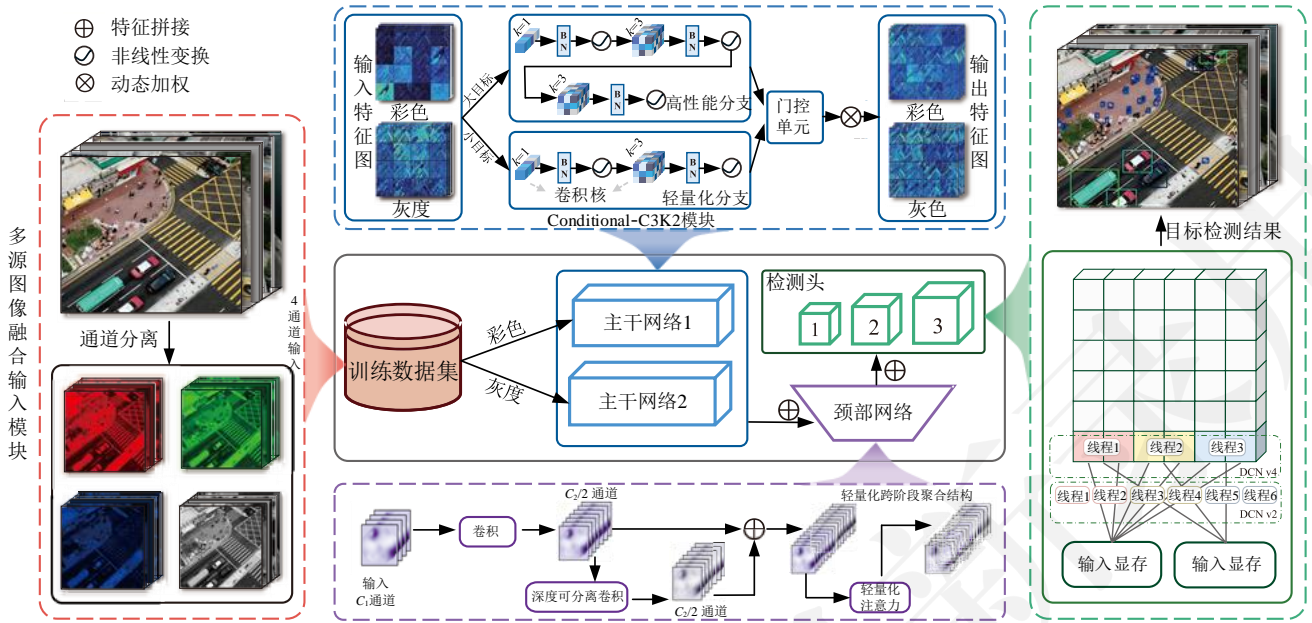


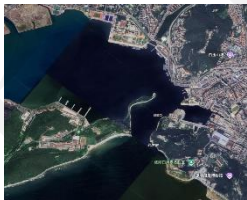
图1 算法整体框架图

Fig. 1 Diagram of the overall framework of the algorithm

在VisDrone2019、MVRSD和MAR20数据集上验证方法性能，并进一步分析其在军事侦察场景下的应用潜力。

1 面向密集小目标的轻量化检测算法

军事小目标在航拍、弹载与遥感平台中普遍呈现出如图2所示的尺寸小、易被遮挡、弱光下纹理不稳定以及密集分布等特性，使得浅层细节易丢失，中层语义表达不足，多尺度关联不充分。此外，空基与无人平台的算力与参数开销受限，使得传统基于大规模特征金字塔或重型注意力结构的检测器难以直接部署。因此，本文从细粒度特征增强、中层语义提升、跨阶段多尺度融合、动态检测头优化、小目标回归强化5个方面展开设计，构建轻量化小目标检测框架。



(a) 目标尺寸小
(a) Small target size



(b) 目标遮挡
(b) Target masking



(c) 弱光场景
(c) Low light environment



(d) 目标密集
(d) Dense targets

图2 不同情况下无人机航拍图像

Fig. 2 UAV aerial images under different situations

1.1 四通道主干设计

为增强军事航拍图像中小目标的纹理细节与边缘响应能力，尤其是在弱光条件下缓解亮度不足导致的特征退化问题，本文在主干网络前端构建了RGB与灰度组合的四通道输入结构，并设计SeparateChannel模块实现亮度与颜色信息的分离式编码。该设计的核心作用体现在以下3点：

1) 缓解小尺寸目标的特征模糊。小目标在图像中仅占据极少像素，导致边缘与纹理特征在RGB通道中易被掩盖。灰度通道能够突出亮度差异，使得细粒度的边缘信息更加清晰，从而提升网络对小尺寸目标的感知能力。

2) 增强结构信息的稳定性。在军事场景中，RGB通道在颜色偏移、阴影或局部遮挡情况下容易发生特征不完整，而灰度通道对形状与结构更敏感，可在颜色信息受损时提供更稳定的形状线索，提高算法的细粒度识别能力。

3) 提升弱光环境下的可检测性。在夜间或低照度条件下，彩色图像容易出现噪声增强、对比度下降的问题，导致目标轮廓不清晰。而灰度通道在亮度信息上更为稳定，能够保留关键的光强梯度变化，从而在弱光环境下提高小目标的检测鲁棒性。

基于以上设计，SeparateChannel模块将原始三通道图像拆分为RGB与灰度两个子分支，分别提取语义与结构特征，在后续网络中引入了基于中期融合的多通道特征增强机制。在P2、P3、P4等不同分辨率的特征图中，通过逐层的特征融合模块，灰度

通道与 RGB 通道的信息在特定层次上相互补充,以增强对小目标的细节捕捉能力。通过加权融合策略,不同通道的特征信息被动态调整,以便在中期阶段优先强调对小目标边缘和纹理的细粒度描述。

图 3 展示了经 `SeparateChannel` 模块处理后的 RGB 与灰度图像,其中灰度分支能够显著强化小目标的边缘细节、亮度梯度与结构轮廓,为浅层特征表达提供额外的判别信息。为了进一步说明灰度分支在小目标检测中的必要性,下面从图像形成机制与卷积神经网络的特征响应规律出发,对其有效性进行理论分析。

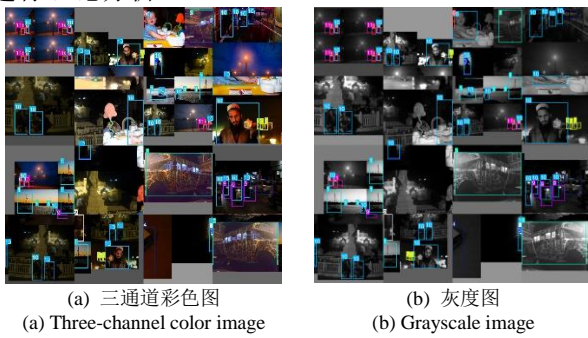


图 3 通道分离模块提取的图像

Fig. 3 Images extracted from the `SeparateChannel`

从图像形成模型与卷积神经网络的特征响应机制来看,灰度编码在军事小目标检测中具有更高的结构敏感性与信噪比优势。对于仅占据少量像素的小目标,RGB 各通道的色彩分量通常呈现强相关性,其差值往往不足以构成可分辨的颜色梯度,因此难以为网络提供稳定的语义支撑。而亮度作为反映能量分布的主要分量,其在成像过程中受色彩通道耦合影响较小,通过

$$I_{\text{gray}} = 0.299R + 0.587G + 0.114B \quad (1)$$

可得到信息密度更高的亮度通道。式(1)中: I_{gray} 为生成的灰度图像强度; R 、 G 、 B 分别为原始彩色图像在红、绿、蓝通道的亮度值。该线性组合在本质上是一种投影变换,将三维颜色空间压缩至单维强度空间,使局部亮度梯度、纹理变化及边界过渡等一阶结构特征得到强化。

在卷积神经网络中,浅层卷积核普遍具备类似 Sobel、Schar 或 Gabor 滤波器的性质,其最强响应来自亮度梯度而非颜色差异。因此,灰度通道能够提供更高的梯度幅值与方向一致性,使网络在小尺度区域内获得更具判别性的结构激活。尤其在军事航拍与夜间成像条件下,RGB 图像会因光照不足产生色彩漂移、噪声放大与信号失配,而亮度信号在

弱光条件下仍具有较高的稳定性与对噪声的鲁棒性,可维持目标边缘与轮廓的可追踪性。此外,灰度编码降低了跨通道的冗余性,使网络在浅层特征融合阶段能够更集中地学习目标的几何结构特征,从而显著增强小目标的纹理表达与可检测性。

1.2 Conditional-C3k2 模块

原生 YOLOv11 算法中的 C3k2 结构采用统一的两层 Bottleneck 进行特征转换,虽然具备良好的轻量化特性,但其固定深度的单路径模式难以同时兼顾浅层细节与深层语义,对无人机航拍图像中广泛存在的小目标、弱纹理及复杂背景环境适应性有限。尤其在小目标密集或受遮挡的场景下,固定的两层卷积可能导致局部边缘被过度平滑,从而削弱可分辨性;而对大尺度或背景复杂区域,单一路径又可能无法提供足够的语义表达能力。此外,原结构对所有通道进行同等深度处理,也导致计算资源无法根据特征难易程度进行自适应分配,限制了轻量化算法在资源受限平台上的进一步优化空间。

为解决上述不足,本文设计了一种基于门控机制的 Conditional-C3k2 模块,将原有的固定路径改造为由轻量分支、强语义分支与跳跃分支共同组成的双路径并联结构,使网络能够在浅层与深层特征表达之间实现按需选择。具体而言,输入特征首先通过 1×1 卷积进行通道压缩,随后分别进入轻量路径与强语义路径:轻量路径仅包含一个 Bottleneck,重点保留局部边缘、亮度梯度等细粒度信息,适用于对小目标至关重要的浅层结构特征;强语义路径则由两个串联 Bottleneck 构成,能够提供更强的语义聚合能力和上下文建模能力,使算法能够处理复杂背景或遮挡区域中的深层特征。同时保留的跳跃分支进一步确保了浅层输入特征能够不经破坏地传递至输出,使小目标所依赖的纹理与几何细节在深层卷积运算下不被削弱,整体结构因此形成了一种并联 1 层+2 层的可变深度特征变换方式,而非简单增加卷积层数。为了更清晰地展示本文提出的 Conditional-C3k2 的整体结构与数据流向,其模块化设计如图 4 所示。该模块由轻量路径($1 \times \text{Bottleneck}$)、强语义路径($2 \times \text{Bottleneck}$)与跳跃路径并行构成,并在分支输出后通过全局平均池化(Global Average Pooling, GAP)—Conv 1×1 —Sigmoid 激活函数形成通道级权重,实现对浅层与深层特征的自适应选择,从而获得可变深度的条件化特征表达能力。

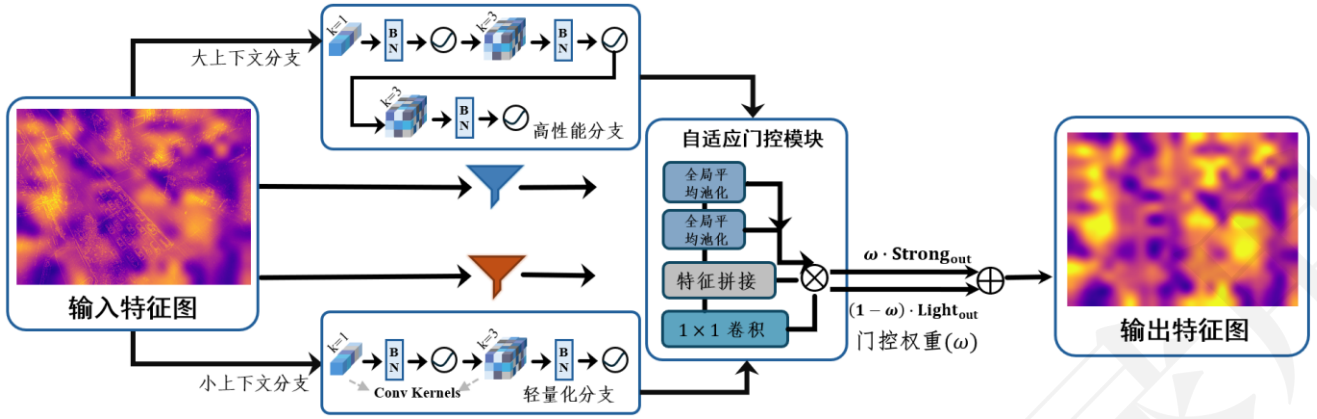


图4 Conditional-C3k2 模块

Fig. 4 Structure of the proposed Conditional-C3k2 module

为在浅层与深层特征表达之间实现自适应选择, 本文在轻量路径与强语义路径的输出之后引入门控机制。该门控网络以两条路径在卷积变换后的特征为输入, 通过对其全局统计信息进行建模, 生成用于调制不同路径贡献的门控权重。门控权重 ω 的生成过程可表示为

$$\omega = \sigma(\text{Conv}_{1 \times 1}(\text{GAP}(\text{Concat}(\mathbf{F}_s, \mathbf{F}_l)))) \quad (2)$$

式中: $\sigma(\cdot)$ 表示 Sigmoid 激活函数, 用以将门控权重归一化至 $[0,1]$ 区间; $\text{Conv}_{1 \times 1}(\cdot)$ 为 1×1 卷积映射, 用于压缩通道并建模通道间相关性; $\text{Concat}(\cdot)$ 表示在通道维度上的特征拼接; $\text{GAP}(\cdot)$ 表示全局平均池化操作, 用于提取各通道的全局统计特征, \mathbf{F}_s 为强语义路径输出特征; \mathbf{F}_l 为轻量化路径输出特征。

上述门控网络分别对轻量路径与强语义路径的输出特征进行全局平均池化, 并将其拼接后通过 1×1 卷积与 Sigmoid 激活生成通道级权重, 使算法能够依据当前特征复杂度自适应决定更偏向浅层还是深层路径。本文将门控机制作用于两条分支的输出特征, 使门控权重的生成依赖于不同路径在卷积变换后的实际表达, 从而在特征层面实现条件化融合, 而非基于输入进行静态路由。

最终输出由门控控制的两分支加权结果与跳跃分支共同构成。为从数学层面刻画 Conditional-C3k2 模块中门控机制对不同路径特征的调制作用, 其融合过程可形式化表示为

$$\mathbf{F}_o = \omega \odot \mathbf{F}_s + (1 - \omega) \odot \mathbf{F}_l + \mathbf{F}_k \quad (3)$$

式中: \mathbf{F}_o 为输出特征; ω 为通道级权重, 其计算见式(2), $\omega \in [0,1]$; \odot 表示逐通道加权操作; \mathbf{F}_k 为跳跃分支特征。

Conditional-C3k2 模块采用路径并联与门控调制相结合的结构形式, 在特征变换阶段引入细节保留分支, 使特征融合过程由单一路径的层级叠加转

变为多路径协同建模, 从结构上区别于通过简单增加网络深度实现表达增强的方式。在保持近似计算量与参数规模的前提下, 该模块能够有效增强算法对小目标、弱纹理与复杂场景的特征提取能力, 为后续多尺度检测头提供更具区分度的中层特征输入。

1.3 轻量化跨阶段聚合结构设计

在多尺度特征融合中, 高分辨率特征通常具有较大的空间尺寸, 其卷积计算量在四通道输入下进一步放大。传统多分支聚合结构, 如多分支跨阶段局部通道(Variety of View Cross Stage Partial Channel, VoVCSPC)具备较强的表达能力, 但其深度卷积堆叠与较高的通道扩张比例会显著增加参数量和每秒浮点运算次数(Floating Point Operations per Second, FLOPs), 使其难以在无人机(Unmanned Aerial Vehicle, UAV)航拍、弹载末段制导等实时军事场景中部署。轻量化检测框架亟需一种能够兼顾计算效率与聚合质量的浅层多尺度融合模块, 从而在不提高整体推理延迟的前提下提升特征交互能力。基于此需求, 本文从高效卷积组合、跨阶段特征流动与低开销注意力机制出发, 设计了轻量化跨阶段聚合结构, 旨在以极低的结构复杂度实现稳定而有效的多尺度融合。

设输入特征为

$$\mathbf{X} \in \mathbb{R}^{C_i \times H \times W}, \quad C_m = \lfloor 0.5C_o \rfloor \quad (4)$$

式中: $\mathbf{X} \in \mathbb{R}^{C_i \times H \times W}$ 表示输入特征图; C_i 为输入通道数; H 和 W 分别为特征图的高度与宽度, 表示空间尺寸; C_m 为中间瓶颈通道数; C_o 为输出通道数; $\lfloor \cdot \rfloor$ 表示向下取整操作。

LCSA 由跨阶段分流、轻量瓶颈变换、一次性聚合、轻量注意力、输出整合 5 个部分组成:

1) 跨阶段分流。通过两条轻量 1×1 卷积分流输入特征, 分别作为主干路径与跳连路径, 以减少深度卷积的冗余计算:

$$\begin{cases} \mathbf{X}_m = \text{Conv}_{1 \times 1}^{C_i \rightarrow C_n}(\mathbf{X}) \\ \mathbf{X}_k = \text{Conv}_{1 \times 1}^{C_i \rightarrow C_n}(\mathbf{X}) \end{cases} \quad (5)$$

式中： \mathbf{X}_m 与 \mathbf{X}_k 分别为主干路径特征与跳连路径特征； C_i 为输入通道数； C_n 为内部通道数，该结构在保持底层纹理的同时提高梯度流动效率。

2) 轻量瓶颈变换。主干路径采用深度可分离卷积与点卷积进行轻量化特征变换：

$$\mathbf{D} = \text{DWConv}_{3 \times 3}^{C_n \rightarrow C_n}(\mathbf{X}_m) \quad (6)$$

$$\mathbf{Z} = \text{PWConv}_{1 \times 1}^{C_n \rightarrow C_n}(\mathbf{D}) \quad (7)$$

式中： $\text{DWConv}(\cdot)$ 表示深度可分离卷积； PWConv 为逐点卷积； \mathbf{Z} 为逐点卷积输出特征，该结构无需多分支深堆叠即可获得稳定的聚合效果； \mathbf{D} 为深度卷积输出特征。深度可分离卷积显著降低计算量，而瓶颈结构避免通道冗余。

3) 一次性聚合卷积。瓶颈输出通过一层标准 3×3 卷积完成融合：

$$\mathbf{A} = \text{Conv}_{3 \times 3}^{C_n \rightarrow C_n}(\mathbf{Z}) \quad (8)$$

4) 轻量通道注意力。为增强特征自适应性，引入计算开销极低的通道注意力机制：

$$\begin{cases} \mathbf{s} = \text{GAP}(\mathbf{A}) \\ \alpha = \sigma(\text{Conv}_{1 \times 1}^{C_n \rightarrow C_n}(\mathbf{s})) \\ \mathbf{A}' = \mathbf{A} \odot \alpha \end{cases} \quad (9)$$

式中： \mathbf{s} 为通过全局平均池化提取的通道统计量； \mathbf{A}' 为调整后的输出特征； α 为生成的通道注意力权重。该模块仅包含一步 1×1 卷积，几乎不增加开销。

5) 跨阶段融合与输出映射：

$$\mathbf{Y} = \text{Conv}_{1 \times 1}^{2C_n \rightarrow C_o}([\mathbf{X}_k, \mathbf{A}']) \quad (10)$$

实现跳连特征与聚合特征的高效融合。

LCSA 的整体结构如图 5 所示。LCSA 通过跨阶段分流、深度可分离卷积与聚合卷积的组合构建了一种结构紧凑的多尺度特征融合单元。该结构在保持网络整体深度和通道规模不变的前提下，进一步降低了浅层融合过程的冗余计算量，使其能够在四通道输入与高分辨率特征图条件下维持较低的计算开销。与多分支聚合类结构相比，LCSA 在结构复杂度上具有更小的开销，有利于在轻量化框架中进行部署。在算法整体结构中，LCSA 的引入未改变特征金字塔的拓扑关系，其输出与原有特征融合流程保持一致。实验结果表明，采用 LCSA 后，算法的主要性能指标保持稳定，参数量与 FLOPs 得到有效控制，使网络在资源受限场景中的运行效率更具可预期性，为模型在轻量化条件下的算力分配提供了合理的基础。

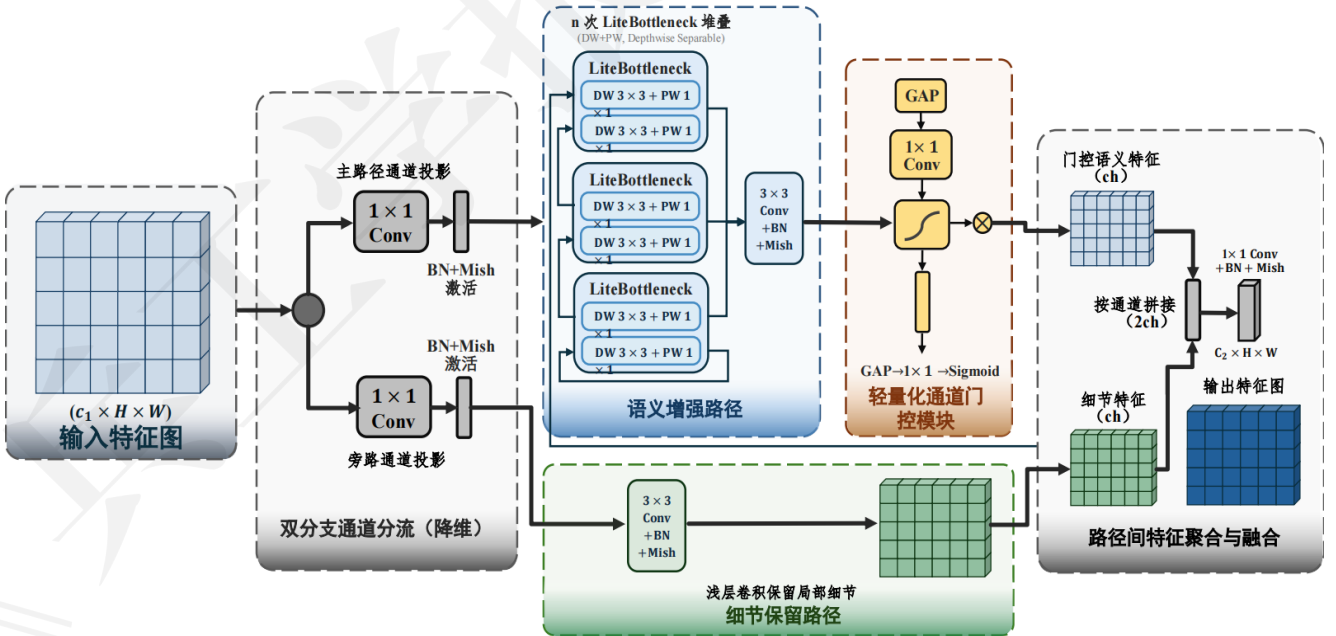


图 5 LCSA 结构

Fig. 5 Structure of LCSA

1.4 基于 DCNv4 改进的多尺度 DyHead

在无人机与空基军事侦察场景中，小目标通常具有空间尺度极小、目标边界不完整、局部形变显

著以及候选区域密集分布等特征，这对检测头阶段的特征对齐能力与空间建模能力提出了更高要求。现有 YOLO 系列检测头多采用固定采样位置的卷积结构，其感受野形态难以随目标尺度与几何形态动

态调整，在面对尺度剧烈变化或非刚性形变的小目标时，容易产生定位偏移与置信度不稳定等问题。

DyHead 通过在检测头中统一引入尺度注意力、空间注意力与任务感知注意力机制，实现了对多尺度特征的自适应加权与任务分支调制，在通用目标检测任务中已被证明能够有效提升多尺度目标的建模能力^[4]。然而，在小目标密集且空间分布高度不均的军事场景下，DyHead 内部的特征聚合过程仍主要依赖固定采样的卷积算子，其空间采样位置无法随目标局部几何结构自适应调整，从而限制了对微小目标局部形变与非完整边界的精细建模能力。

针对上述问题，本文以提升军事小目标检测头

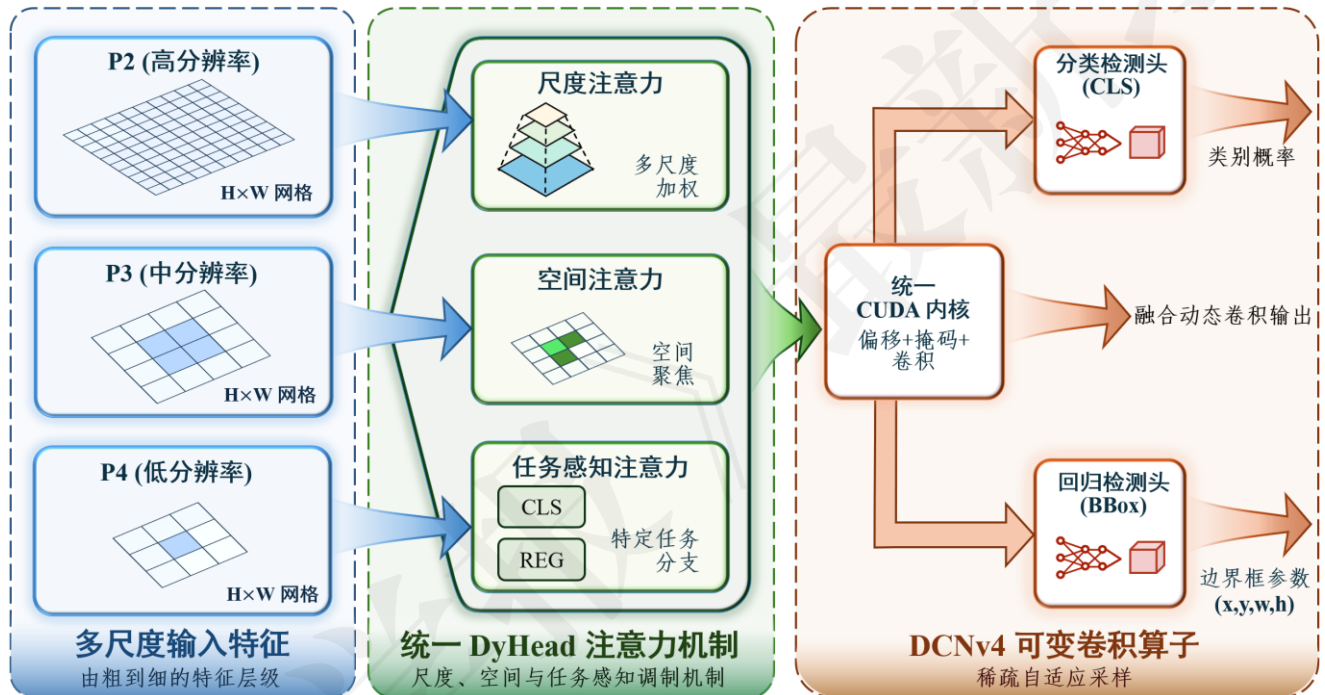


图 6 基于 DCNv4 的多尺度 DyHead 结构

Fig. 6 Structure of the multi-scale DyHead with DCNv4

从军事应用角度来看，所提出的 DyHead-DCNv4 结构在空基与无人机侦察场景中具有明显优势。针对军事目标尺度小、形态不规则且背景干扰复杂等特点，基于 DCNv4 的自适应采样机制能够更有效地对齐目标真实几何结构，有助于提升检测结果在复杂空基侦察场景下的整体鲁棒性。同时，军事侦察任务通常对实时性与平台算力具有严格约束，DCNv4 通过高效的算子融合与内存访问优化，在增强空间对齐能力的同时显著降低推理开销，使该结构更加契合空基平台对实时处理与持续运行的工程需求。

综上所述，DyHead-DCNv4 在继承 DyHead 多尺度注意力建模优势的基础上，通过针对性引入更高效的空间建模机制，进一步增强了检测头阶段对军事小目标复杂成像特性的适应能力。其在空间对

阶段的空间对齐精度与实时推理效率为目标，对 DyHead 中的核心卷积算子进行定向改造，引入 DCNv4 替换原有的 DCNv2 模块^[5]，构建面向小目标场景优化的 DyHead-DCNv4 结构。其具体结构如图 6 所示：来自 P2、P3 与 P4 的多尺度特征首先在 DyHead 框架内通过尺度注意力与空间注意力进行自适应加权与聚焦，并在任务感知注意力的调制下分别送入分类与回归分支；随后，在特征融合与预测阶段，引入 DCNv4 可变形卷积算子对关键区域进行稀疏自适应采样，使卷积感受野能够随小目标的几何形态与空间分布动态调整，从而有效缓解固定采样带来的定位偏移问题。

齐精度与推理效率之间实现了较为理想的平衡，使得检测结果在复杂空基侦察场景下更加稳定可靠，同时避免了对计算资源的过度消耗。上述特性使该检测头结构不仅在算法层面具备针对军事小目标的优化意义，也在工程层面为后续在无人机及其他空基平台上的实时部署提供了切实可行的技术支撑。

1.5 优化复合定位损失函数

针对无人机航拍场景中小目标尺度极小、边界模糊且样本尺度分布高度不均的问题，原 YOLOv11 所采用的完全交并比 (Complete Intersection over Union, CIoU) 与分布聚焦损失 (Distribution Focal Loss, DFL) 回归损失在方向对齐与尺度敏感性方面仍存在一定局限，表现为小目标回归精度不足、训练收敛偏慢等现象。为此，本文在保持 YOLOv11 算法整体损失框架不变的前提下，设计了一种融合面积加权机制的复合损失函数 VSA-Loss。该损失函数

以变焦 损失承担分类与置信度建模任务,在回归分支中采用 SIoU 损失替代传统 CIoU 损失,以增强对中心偏移方向与几何对齐关系的刻画能力;同时引入基于目标面积倒数的自适应加权策略,以缓解小目标在训练过程中的梯度弱化问题,从而提升算法在小目标密集场景下的回归精度与整体检测性能。

1.5.1 损失函数的原理

SIoU 损失通过传统 IoU 约束基础上引入方向感知与几何对齐机制,有效缓解了边界框回归过程中等距不同向的优化歧义问题。该损失函数综合考虑预测框与真实框之间的重叠度、中心偏移距离、宽高比例差异以及方向引导成本,将位置与形状约束统一建模,从而在回归初期即可提供更具判别性的梯度方向。相较于 CIoU 等仅基于距离与长宽比的损失形式,SIoU 在小目标和密集目标场景中表现出更快的收敛速度与更稳定的定位精度。

1.5.2 加权融合机制原理

在以 VisDrone 为代表的小目标密集场景中,训练样本在尺度分布上呈现显著不均衡特征。由于大目标在归一化损失中天然具有更大的回归尺度与重叠度,小目标样本在梯度反向传播过程中往往贡献有限,导致模型参数更新过程对小目标不敏感。针对这一问题,本文在复合损失函数中引入基于目标面积的自适应加权机制,通过对每个正样本赋予与其真实框面积成反比的权重系数,显式放大小目标在损失函数中的梯度占比,从优化层面缓解尺度不平衡带来的训练偏置。

首先设第 p 个目标的真实框面积为 $A_p = w_p \times$

h_p , 全体目标面积中位数为 A_m , 则该目标的面积加权系数可表示为

$$w_p = \left(\frac{A_m}{A_p + \varepsilon} \right)^\beta, \quad \beta \in [0.3, 1.0] \quad (11)$$

式中: ε 为防止除 0 的小常数; β 为面积加权指数,调节小目标的梯度放大强度。

在损失函数的具体实现中,所有正样本的分类损失与回归损失均乘以对应的面积权重 w_p , 再进行归一化聚合, 公式为

$$\mathcal{L}_c^w = \frac{\sum_p w_p \cdot \mathcal{L}_{VFL}^{(p)}}{\sum_p w_p}, \quad \mathcal{L}_b^w = \frac{\sum_p w_p \cdot \mathcal{L}_{SIoU}^{(p)}}{\sum_p w_p} \quad (12)$$

式中: \mathcal{L}_c^w 和 \mathcal{L}_b^w 分别表示加权后的分类损失与回归损失; $\mathcal{L}_{VFL}^{(p)}$ 与 $\mathcal{L}_{SIoU}^{(p)}$ 分别为第 p 个正样本原始的变焦

损失和 SIoU 损失; $\sum_p w_p$ 为所有正样本面积权重的累加, 用于对损失进行归一化处理。该机制本质上通过以小补大的方式, 显式提升了小目标在总损失中的梯度贡献, 有效缓解了原始训练过程中因尺度不平衡导致的小目标拟合弱、回归漂移大等问题。

为更直观地说明所提出的面积加权机制在分类与回归分支中的作用方式, 以及其与变焦损失和 SIoU 损失的协同关系, 图 7 给出了 VSA-Loss 的整体计算流程与各组成模块之间的交互示意。

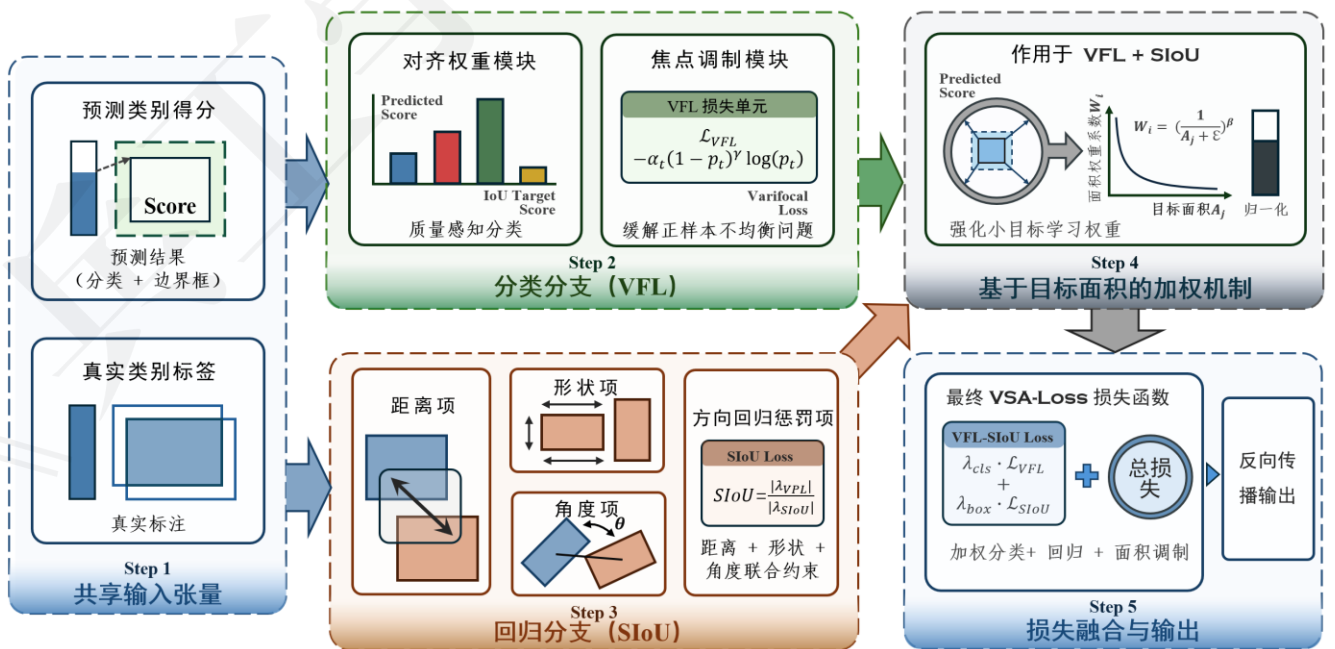


图 7 VSA-Loss 损失函数计算流程

Fig. 7 Computation Flow of the VSA-Loss

1.5.3 复合损失函数的构建策略与优化机制

本方法将分类与置信度优化统一由变焦损失^[6]承担,回归则采用 SIOU Loss^[7]建模目标框之间的重叠度、方向关系与形状偏差。为进一步提升小目标的优化强度,分类与回归两个分支均引入面积加权系数,使小目标在梯度计算中获得更高关注度。具体融合策略为:分别对两个分支损失按面积加权进行归一化,再按比例系数进行加权叠加,得到最终复合损失函数:

$$\mathcal{L}_{\text{final}} = \lambda_c \cdot \mathcal{L}_c^w + \lambda_b \cdot \mathcal{L}_b^w \quad (13)$$

式中: λ_c 与 λ_b 分别控制分类与置信度和回归分支的权重。面积加权项已在式(11)中定义,其作用在此阶段体现为梯度引导与优化偏移。

相较于原始 YOLOv11 所采用的变焦损失+ CIoU +DFL 的三支损失结构,本文提出的 VSA-Loss 在优化目标上更加集中,回归约束更具方向性,梯度分配机制更加合理。在不改变 YOLO 框架整体结构的前提下,该损失函数能够有效提升小目标的定位精度与检测稳定性,尤其适用于 VisDrone 等小目标密集的无人机航拍场景。

2 实验结果分析

2.1 实验数据与环境

2.1.1 实验数据

为全面评估所提 GSCD-YOLO 系列方法在多源军事感知场景下的适用性与泛化能力,本文选取 4 类公开基准数据集作为实验基础: VisDrone、MVRSD、MAR20 与 Exclusively Dark (ExDark)数据集。四者在成像视角、光照条件、目标尺度及任务属性上具有明显差异,能够从不同维度验证算法性能。

VisDrone 为大规模无人机航拍基准,覆盖城市与乡村环境,包含丰富的低空 UAV 视角图像与视频帧,具有目标小、数量密集、运动模糊和背景复杂等特点^[8]。MVRSD 是专门针对细粒度军事车辆研制的遥感检测数据集,包含小型军用车辆(Small Military Vehicle, SMV)、大型军用车辆(Large Military Vehicle, LMV)、装甲战斗车辆(Armored Fighting Vehicle, AFV)、通用车辆(Common Vehicle, CV)以及军事工程车辆(Military Construction Vehicle, MCV) 5 类典型目标。该数据集中的 SMV 与 AFV 类别在远距离遥感视角下像素占比极低,具有显著的微小目标特征,能够有效验证算法在真实战场态势感知中的细粒度识别性能^[9]。MAR20 则为专门构建的军事目标遥感识别数据集,涵盖多类军用飞机与中等尺度目标^[10],适用于验证算法在纯军事目标上的迁移与细粒度识别能力。

ExDark 是一类专门针对极低照度环境构建的可见光数据集,包含 12 类在微光、暗夜及复杂人工光源干扰下的目标图像。该数据集旨在模拟战场环境下的夜间隐蔽侦察及黎明/黄昏时段的低可见度场景,用以严谨验证本文提出的 RGB-Gray 四通道输入结构在色彩信息丢失、信噪比下降时的结构特征提取能力。

从应用背景看,近年来俄乌冲突中低成本无人机的广泛应用(见图 8),使低空、机动、分散的小目标逐渐成为现代战场态势感知与精确打击中的关键对象^[11]。由于无人机造价低廉,可通过群体化方式实施突防,而拦截所需的导弹或防空火力单价高、反应时间受限,难以实现全面覆盖,从而形成明显的成本—效能不对称问题。在此背景下,对低空无人机目标进行实时、可靠的视觉检测具有重要的军事意义。VisDrone 数据集在成像特征与目标分布上高度贴近该类应用场景,本文结合典型 FPV 视角示例对其应用背景进行了直观说明。



(a) 无人机对装甲与地面车辆实施近距离打击的第一视角示例

(a) FPV example of a UAV conducting close-range strike on armored/ground vehicles



(b) 无人机实施投弹与近距打击军事人员

(b) UAV bombing/close-range strike on military personnel

图 8 无人机近距打击作战示例

Fig. 8 Examples of UAV Close-Range Strike Operations

与此同时, MVRSD 与 MAR20 数据集体现了空基或卫星侦察平台在发现敌方机动装甲力量、航空兵力部署等关键军事资产方面的重要价值。相比于通用

遥感任务,军事感知要求算法能够捕获高价值装备特有的细粒度结构特征。

综上所述,本文构建了核心性能验证+典型场景补充的多维度实验体系:将 VisDrone 数据集与 MVRSD 作为主实验数据集,旨在通过高难度场景与细粒度军事类别,全面评估算法对军事小目标的检测精度与收敛性能;同时,将 ExDark 数据集与 MAR20 作为补充实验数据集,分别用于专项验证算法在极端黑暗环境下的稳健性,以及在优化小目标性能后对中大尺度目标的检测保持性。这种层级化的实验布局确保了方法在实战化态势感知任务中的全面性与高可靠性。

2.1.2 实验环境

为确保实验结果的有效性与可比性,本文各算法的训练超参数均保持一致。输入图像尺寸设为 640×640 ,IoU 阈值为 0.5,训练轮数为 100,初始学习率和终止学习率分别为 0.01 和 0.2,优化器采用随机梯度下降法,动量系数为 0.937,批次大小为 2。

在数据划分方面,VisDrone 数据集按照官方提供的划分方式,选取 6471 张图像作为训练集、548 张作为验证集、3190 张作为测试集;MVRSD,依照军事遥感目标检测规范将图像按 70%、20%与 10%的比例划分为训练集、验证集与测试集,旨在确保算法在面对 AFV 等细粒度军事类别时具备充足的特征学习样本。补充实验数据集 MAR20 同样采取 70%训练、20%验证与 10%测试的划分策略,用以验证大目标检测性能的保持性;针对低照度验证任务,ExDark 数据集按照 8:2 的比例随机划分为训练集与验证集,重点考察 GCSD-YOLOv11n 算法在极端暗光条件下对目标结构信息的感知能力。

本文实验所用的实验设备配置如表 1 所示。

表 1 实验环境配置信息

Table 1 Experimental environment configuration information

配置	版本
操作系统	Windows10
处理器	Intel Core i7-9750h
显卡	Nvidia Geforce RTX2070 Laptop (8G)

表 2 消融实验中各改进算法的性能测试评价指标

Table 2 Performance test evaluation results of each improved algorithms in ablation experiments

框架	Pytorch 2.0.1
开发软件	Pycharm 2024.3.1
开发语言	Python 3.9.21

为全面评估 GCSD-YOLOv11n 算法的性能,从分类准确性、模型轻量化与实时性能 3 个维度开展实验。分类准确性采用精确率 P 、召回率 R 以及平均精度均值(mean Average Precision, mAP)进行度量, mAP_{50} 表示当 IoU 阈值设为 0.5 时的平均精度, mAP_{50-95} 则表示 IoU 阈值在 0.50~0.95 之间(步长为 0.05)的 10 个平均精度值的平均数,该指标能更全面地反映模型在不同定位精度要求下的检测能力;模型轻量化以参数量与 FLOPs 作为评价标准;实时性能通过包含非极大值抑制(Non-Maximum Suppression, NMS)的推理时间来衡量。

2.1.3 速度测试方法

为客观评估各算法的实时性,本文统一采用单张图像延迟作为速度指标,其计算方式为模型前向推理与后处理 NMS 过程的平均耗时,不包含数据读取与可视化时间。实验环境与表 1 一致,推理阶段采用 FP32 精度,未使用 TensorRT、半精度或量化加速。输入尺寸固定为 640×640 ,批量大小设为 1,以反映在线单帧推理的响应延迟。测试时首先进行约 50 次预热以消除缓存与编译影响,随后连续推理 300 次,在每次前后调用 CUDA 同步函数以保证计时精度,取算术平均值作为单帧延迟。

本文所有对比算法均采用相同的硬件平台与计时流程进行测试,后续中的时间指标即基于上述流程得到的前向推理与 NMS 平均单帧耗时,不计入图像加载与显示时间。该方式能够在统一标准下衡量不同模型的推理性能,确保结果的公平性与可复现性。

2.2 消融实验

为验证所提改进在小目标检测中的实际效果,以 YOLOv11n 算法作为对照网络,按照模块化思路依次融入本文所提的改进方法并在 VisDrone2019 数据集上展开训练与评估。共设计了 12 组消融实验,结果汇总见表 2。

编号	算法	P/%	$mAP_{50}/\%$	$mAP_{50-95}/\%$	参数量/ 10^6	FLOPs/ 10^9	时间/ms
1	YOLOv11n 算法	42.9	32.1	18.5	2.62	6.6	10.8
2	YOLOv11n 算法+4 通道	49.6	41.1	24.7	1.42	15.6	13.7
3	YOLOv11n 算法+Conditional-C3k2 模块+4 通道	50.2	41.6	25.1	1.49	18.1	14.5
4	YOLOv11n 算法 Conditional-C3k2 模块+4 通道+ Dyhead-DCNv2	54.2	44.6	27.1	1.89	21.19	31.0
5	YOLOv11n 算法+Conditional-C3k2 模块+4 通道+ Dyhead-DCNv2+LCSA	54.6	44.8	27.2	1.90	20.40	34.2
6	YOLOv11n 算法+Conditional-C3k2 模块+4 通道 +Dyhead-DCNv4+ LCSA	54.6	44.9	27.3	1.85	22.50	17.6
7	YOLOv11n 算法+ConditionalC3k2 模块+4 通道 +Dyhead-DCNv4+ LCSA +VSA-Loss	55.3	45.1	27.4	1.85	22.50	17.6
8	YOLOv11n 算法+4 通道+Dyhead-DCNv2	54.5	44.0	26.7	1.83	18.9	30.8
9	YOLOv11n 算法+Conditional-C3k2 模块 +LCSA	46.1	34.7	20.3	2.73	8.4	18.7
10	YOLOv11n 算法+LCSA	43.3	32.0	18.6	2.49	5.9	16.9
11	YOLOv11n 算法+Conditional-C3k2	44.9	34.9	20.5	2.75	8.9	12.1
12	YOLOv11n 算法+ Dyhead-DCNv2	44.7	34.7	20.4	2.51	16.0	26.1
13	YOLOv11n 算法+ Dyhead-DCNv4	44.7	34.9	20.5	2.46	18.1	13.7

由表 2 实验结果可知, 基准算法 YOLOv11n 算法在小目标检测任务中的表现较弱, mAP_{50} 为 32.1%。第 2 组在其基础上引入 4 通道图像输入, 打开 P2 层并关闭了 P5 层, 统一检测头通道, 使模型能够利用更多的信息, 提升特征表达能力, mAP_{50} 提升至 41.1%, 较基线算法提升了 9.0%, 参数量也大幅下降; 第 3 组进一步加入 Conditional-C3k2 结构, 有效增强了语义建模能力, mAP_{50} 提升至 41.6%, 相较前一组再提升 0.5%; 第 4 组进一步在结构中加入 Dyhead-DCNv2, 增强了检测阶段对多尺度特征的适应能力, 使模型在候选框分类与回归任务中表现更为稳定, mAP_{50} 提升至 44.6%, 比前一组提升 3%。第 5 组加入 LCSA 模块以加强浅层信息融合, 进一步提升算法对小目标的检测效果, mAP_{50} 达到 44.8%, 比上一组提高 0.2%。第 6 组将 DynamicHead 中的 DCNv2 替换为 DCNv4, 大幅减少了推理时间。最终, 第 7 组融合 Conditional-C3k2 结构、4 通道输入、Dyhead-DCNv4 检测头、LCSA 模块, 并引入 VSA-Loss 损失函数构建完整模型, 在检测精度与模型复杂度之间取得良好平衡, 其主要性能指标为: P 为 55.3%, mAP_{50} 为 45.1%, mAP_{50-95} 为 27.4%, 参数量为 1.85×10^6 , 计算量为 22.50×10^9 。相比基线算法, mAP_{50} 提升了 13.0%, mAP_{50-95} 提升了 8.9%, 提升幅度高达 40.5%与 48.1%。

为从特征响应层面分析上述改进对模型关注区域的影响, 本文对基线算法与改进算法的中间特征进行可视化对比分析, 如图 9 所示。



(a) YOLOv11n 算法

(a) YOLOv11n algorithm



(b) GSCD-YOLOv11n 算法

(b) GSCD-YOLOv11n algorithm

图 9 YOLOv11n 与本文算法特征响应热力图对比

Fig. 9 Comparison of feature response heatmaps between YOLOv11n and the proposed algorithms

综上所述, 本文所提出的各项改进在不显著增加参数量的前提下, 均对检测精度产生了积极影响, 算法在精度与效率之间取得了优秀的权衡, 具备广阔的工程应用前景。

2.3 多场景适应性及跨数据集泛化分析

为进一步验证本文所提方法中关键结构设计的合理性与适用性, 在主消融实验与对比实验之外,

设计了一组补充实验，分别从复杂光照条件以及结构裁剪对中大尺度军事目标影响等方面，对核心设计动机进行针对性验证。相关实验旨在从工程与应用视角对各模块配置的合理性进行补充验证。

2.3.1 四通道输入在低照度场景下的有效性验证

低照度与复杂光照条件是无人机侦察与监视任务中常见且具有挑战性的成像环境，在此类场景下，目标边缘与纹理信息易受到噪声干扰而显著退化，进而导致检测算法出现漏检或置信度不稳定等问题。为验证本文所引入的 4 通道输入结构在复杂光照条件下的有效性，本文选用 ExDark 低照度数据集开展补充实验^[12]，从鲁棒性角度评估灰度通道对目标检测性能的影响。

在该实验中，保持网络结构、训练策略及推理配置一致，仅对输入形式进行调整，分别采用标准三通道 RGB 输入与本文提出的 RGB-Gray 四通道输入进行对比。为保证实验公平性，两种输入方式均在相同条件下进行训练与测试，评价指标采用 mAP_{50} ，用以衡量算法在低照度场景下的检测精度与漏检情况，结果如表 3 所示。

表 3 不同输入通道配置在 ExDark 数据集上的检测性能对比

输入配置	$mAP_{50}/\%$	$mAP_{50-95}/\%$	参数量/ 10^6	FLOPs/ 10^9	时间/ms
RGB	62.8	38.5	2.62	6.6	9.4
RGB-Gray	66.3	41.2	1.42	15.6	12.3

由表 3 可知，在 ExDark 数据集上引入四通道输入后，模型 mAP_{50} 指标取得稳步增长。这表明灰度信息在弱光条件下能为目标边缘与结构特征提供关键补充，有效降低了漏检风险。

为考察光照强度变化对目标检测结果稳定性的影响，本文对不同照度条件下的检测性能进行了对比分析。实验中采用富士 XH-2S 相机进行图像采集，为排除成像参数变化对亮度表征及目标特征提取的干扰，采集过程中固定相机光圈、快门速度与白平衡设置。其中，快门速度设为 $1/500$ s，光圈固定为 $f/4.0$ ，感光度保持恒定，白平衡模式采用固定值。环境照度通过德力西照度计在被测目标附近进行实时测量，并通过调节补光灯输出功率改变场景照度水平，从而构建不同光照条件下的图像数据集。上述设置保证了光照变化仅来源于环境亮度差异，而非相机成像参数调整带来的影响。

如表 4 所示，随着照度从 100 lx 降低至 10 lx，基于 RGB 输入的基准算法由于信噪比下降，暴露出明显的置信度跌落与局部漏检退化；相比之下，四通道输入算法在各照度梯度下均保持了更稳定的目标响应，其轮廓感知更为连续。实验证明，灰度信息的引入有效缓解了光照衰减引起的特征退化，显著提升了算法在极端弱光场景下的检测鲁棒性。

表 4 不同照度条件下 RGB 与 RGB-Gray 输入算法检测性能对比



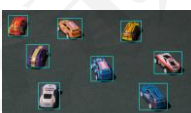

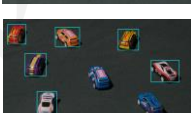
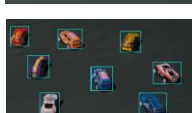
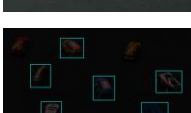
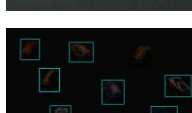
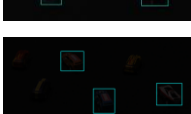
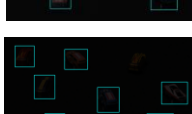
照度/lx	RGB 三通道	RGB+Gray 四通道
100		
50		
30		
20		
10		

图 10 为直观展示本文方法在低光照场景下的检测效果，给出了在 VisDrone 夜间航拍图像上的检测可视化结果，并与 YOLOv11n、YOLOv11s、UAV-DETR-ev2 算法进行对比。如图 10(b)所示，YOLOv11n 算法对小目标的检测存在明显漏检问题，部分边缘目标及低亮度区域目标未能正确识别；图 10(c)为 YOLOv11s 算法的检测结果，尽管检测框数量有所增加，但对密集目标的定位边界仍不精确，存在误检现象；图 10(d)为 UAV-DETR-ev2 算法的检测结果，与白天密集目标相比，尽管检测框重叠现象有所缓解，但漏检现象明显；图 10(a)展示了本文所提出的 GCSD-YOLOv11n 算法的检测效果，相较前两者在小目标检测数量、定位精度以及遮挡处理方面表现更为优异，尤其是在夜间复杂光照条件下，依然能够准确识别图像中大多数目标，体现出较强的鲁棒性和细粒度感知能力。

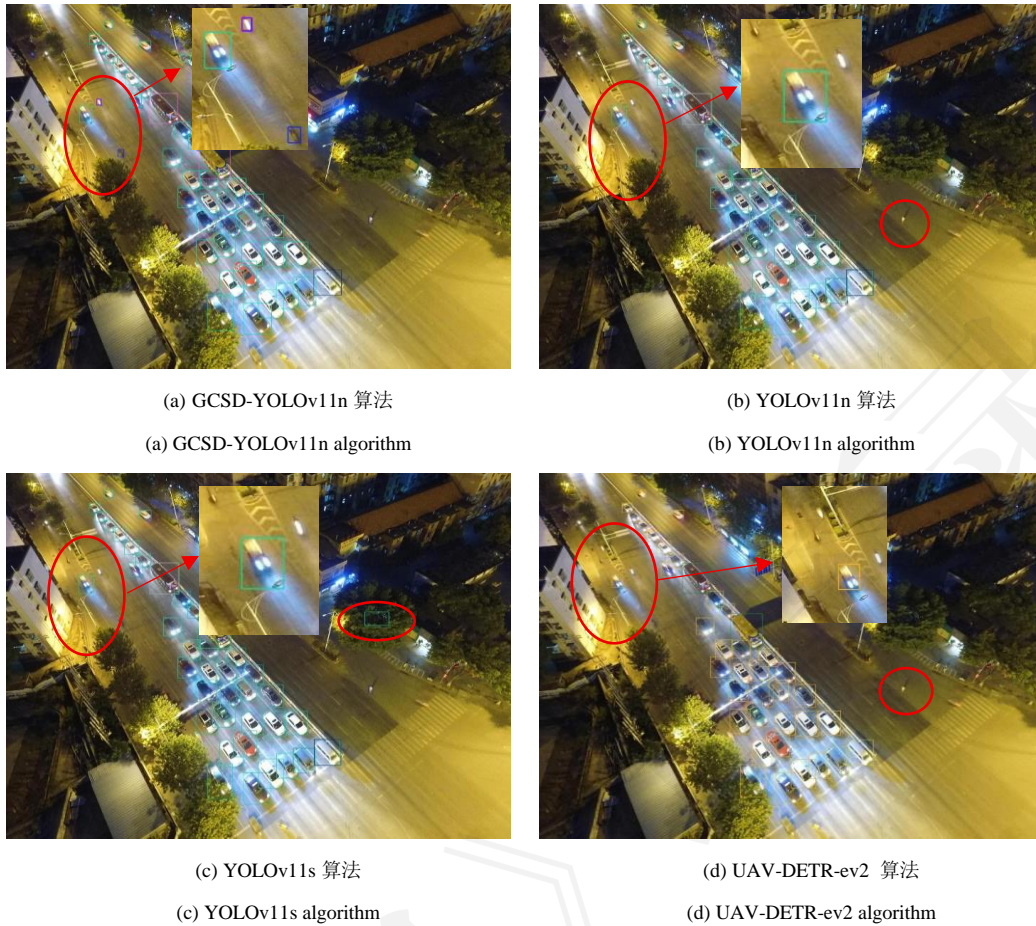


图 10 复杂背景下不同算法检测效果对比

Fig. 10 Comparison of detection results of different algorithms under complex backgrounds

综上所述，本文方法在低照度、高密度及多尺度混合场景下表现出更稳定的检测性能，体现出较好的实际应用潜力。该性能提升主要源于模型引入的多尺度感知与注意力增强设计：Conditional-C3k2与DyHead-DCNv4算法强化了小目标区域的特征响应能力，四通道图像输入为复杂背景下的边缘与纹理建模提供了有效补充。在Tricycle、Motor等低频类别上，改进算法仍保持较稳定的检出表现，表明其在类别不均衡场景中具备一定的泛化能力。综合定量与可视化结果可知，GCSO-YOLOv11n算法在检测精度与复杂场景适应性方面优于YOLOv11n算法，具有良好的工程应用潜力。

2.3.2 裁剪高层检测分支对中大目标的影响验证

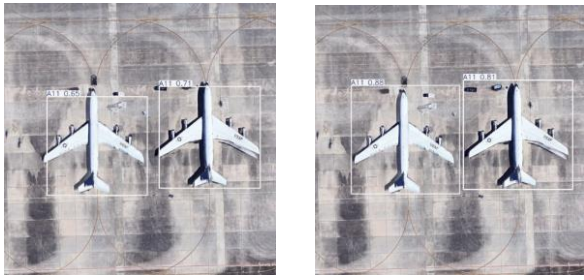
为验证本文所采用的高层检测分支裁剪策略在非小目标场景下的适用性，在MAR20军事目标遥感数据集上开展补充实验分析。考虑MAR20数据集以中大尺度军用飞机目标为主，该实验不以提升绝对检测精度为主要目标，而侧重于检验在裁剪高层检测分支、降低模型复杂度的情况下，所提方法在中大目标场景中是否仍能保持稳定的检测性能。

在该实验中，选取YOLOv11n算法作为基准算法，与本文改进算法在相同训练与推理配置下进行

对比。相关检测性能指标如表5所示。由表5可见，在MAR20数据集上，本文方法在 mAP_{50} 指标上并未落后于基准算法，在 mAP_{50} 指标上由96.1%提升至97.3%，同时在更严格的 mAP_{50-95} 指标上取得进一步提升，表明在去除高层检测分支后，算法对中大尺度目标的定位精度与回归质量未出现明显退化，检测结果如图11所示，与基准算法相比，本方法在该场景下同样能够实现对目标的有效检测。

表 5 不同算法在 MAR20 数据集上的检测性能对比
Table 5 Detection performance comparison of different models on the MAR20 dataset

算法	$mAP_{50}/\%$	$mAP_{50-95}/\%$	参数量 / 10^6	FLOPs / 10^9	时间 /ms
YOLOv11n 算法	96.1	76.4	2.62	6.6	7.5
YOLOv11s 算法	97.8	79.7	9.45	21.7	8.2
本文算法	97.3	79.5	1.85	22.50	13.9



(a) YOLOv11n 检测结果
(a) YOLOv11n detection results

(b) GSCD-YOLOv11n 检测结果
(b) GSCD-YOLOv11n detection results

图 11 YOLOv11n 与本文算法在 MAR20 上检测效果对比
Fig. 11 Comparison of detection performance of YOLOv11n and the proposed method on MAR20

为分析算法在训练过程中的稳定性，本文对 YOLOv11n 算法与改进算法在 MAR20 数据集上的 mAP 收敛曲线进行了对比，如图 12 所示。训练过程整体平稳，两种算法在各阶段均呈现相近的收敛趋势，且改进算法在训练中更高的 mAP，表明其在非小目标任务中同样具备良好的稳定性与泛化能力。

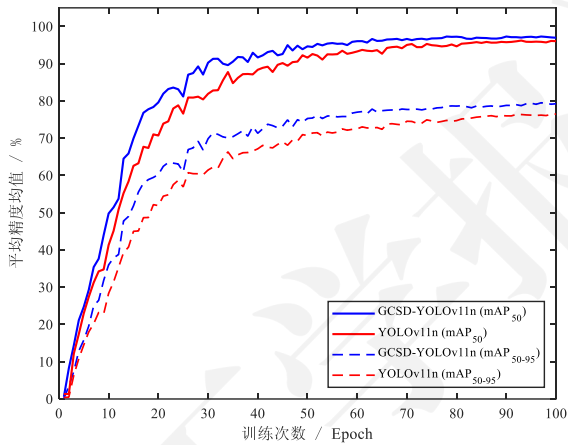


图 12 MAR20 数据集 mAP 收敛曲线对比
Fig. 12 Comparison of mAP convergence curves on the MAR20 dataset

2.4 对比实验

为了进一步验证本文方法的先进性与在复杂战场环境下的实战化潜力，实验选取了多类具有代表性的目标检测算法进行对比评估。实验在统一的软硬件环境下开展，分别基于通用小目标基准数据集 VisDrone2019 与细粒度军事车辆遥感数据集 MVRSD 进行性能测试。前者旨在验证方法在极高目标密度、复杂背景干扰下的底层鲁棒性；后者则重点考察方法对 AFV、SMV 等高价军事目标的识别性能。比较各模型在 mAP_{50} 、 mAP_{50-95} 、模型大小以及计算量等指标上的表现。VisDrone2019 数据集与 MVRSD 的实验结果分别汇总于表 6 与表 7 中。

表 6 不同算法在 VisDrone2019 数据集上的性能对比结果

Table 6 Performance comparison of different models on the VisDrone2019 dataset

算法	$mAP_{50}/\%$	$mAP_{50-95}/\%$	参数量/ 10^6	FLOPs/ 10^9	时间/ms
YOLOv5s	38.3	22.8	9.12	24.1	11.2
YOLOv8s	39.6	23.5	11.16	28.8	10.1
YOLOv11n	32.1	18.5	2.62	6.6	10.8
YOLOv11s	41.0	24.8	9.45	21.7	11.2
YOLOv11m	44.5	27.4	20.06	68.2	35.2
YOLOv12n ^[13]	32.4	18.5	2.57	6.5	9.5
GhostNetV2 ^[14]	25.5	14.3	6.06	7.2	14.0
ShuffleNetV2 ^[15]	22.7	12.4	1.70	4.1	6.1
MoblieNetV4 ^[16]	23.1	12.8	1.79	4.3	6.1
Gold-YOLO ^[17]	34.3	20.4	9.92	20.7	18.9
SBE-YOLO ^[18]	42.1	24.3	6.2	105.4	49.8
RT-DETR-1 ^[19]	16.9	9.2	32.82	108.0	62.1
UAV-DETR-ev2 ^[20]	43.5	26.3	12.1	33.3	46.9
本文算法	45.1	27.4	1.85	22.50	17.6

表 7 不同算法在 MVRSD 上的检测性能对比

Table 7 Detection performance comparison of different models on the MVRSD

算法	$mAP_{50}/\%$	$mAP_{50-95}/\%$	参数量/ 10^6	FLOPs/ 10^9	时间/ms
YOLOv8s	82.9	57.6	11.16	28.8	11.1
YOLOv11n	79.0	54.4	2.62	6.6	11.5
YOLOv11s	83.3	57.8	9.45	21.7	12.6
YOLOv11m	86.1	59.9	20.06	68.2	32.8
Gold-YOLO	81.3	55.6	9.92	20.7	16.9
SEB-YOLO	82.2	57.1	6.2	105.4	49.6
RT-DETR-1	80.2	56.5	32.82	108.0	58.2
UAV-DETR-ev2	79.9	56.0	12.1	33.3	43.8
本文算法	86.4	60.4	1.85	22.50	16.9

图 13 给出了 YOLOv11n 算法、UAV-DETR-ev2 算法与本文改进算法在 VisDrone2019 数据集上的 mAP 收敛曲线对比。由图 13 可以看出，改进算法在训练早期即表现出更快的收敛速度，并在收敛后取得更高的 mAP_{50} 与 mAP_{50-95} ，表明其在训练效率与检测精度方面均具有优势。

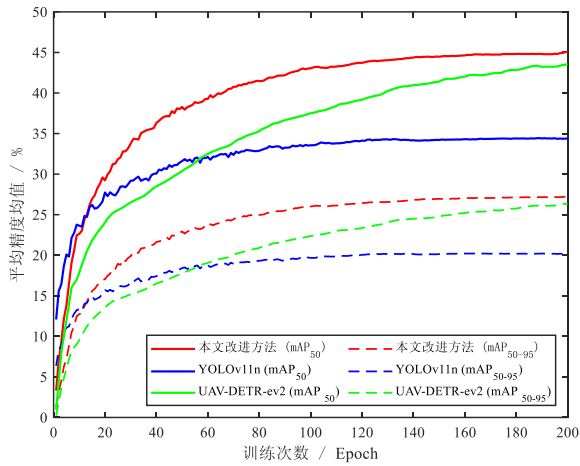


图 13 VisDrone2019 数据集 mAP 收敛曲线对比

Fig. 13 Comparison of mAP convergence curves on the VisDrone2019 dataset

在 MVRSD 对比实验中, 为进一步确立所提方法的先进性与实战化价值, 本文构建了更高强度的对比矩阵。借鉴在 VisDrone2019 通用基准上的实验结论, 为提升对比实验的针对性并剔除性能冗余, 本文放弃了表现欠佳的基础轻量化模型, 转而精选当前目标检测领域具有代表性的 SOTA 架构进行深度对标。通过在 MVRSD 上对 AFV、SMV 等 5 类细粒度军事目标的专项评测, 系统验证 GCSD-YOLO 在处理复杂战场环境下、具有伪装性与尺度极小特征的军事目标时的识别精度优势。实验重点评估算法在保持高性能的同时对硬件资源的极致压缩能力, 以契合空基侦察平台的实时化部署需求。实验结果如表 7 所示。同样地, 为验证本文方法在高分辨率极小目标场景下的训练表现, 绘制了在 MVRSD 上的 mAP 收敛曲线, 如图 14 所示。由图 14 可以看出, 尽管该数据集中的目标具有显著的细

粒度特征且背景纹理复杂, 改进算法依然保持了更快的收敛速度和更高的收敛精度。这说明本文方法在实战化军事遥感场景中同样具有良好的优化特性与稳定性, 能够精准捕捉具有伪装性与尺度极小特征的高价值军事目标。

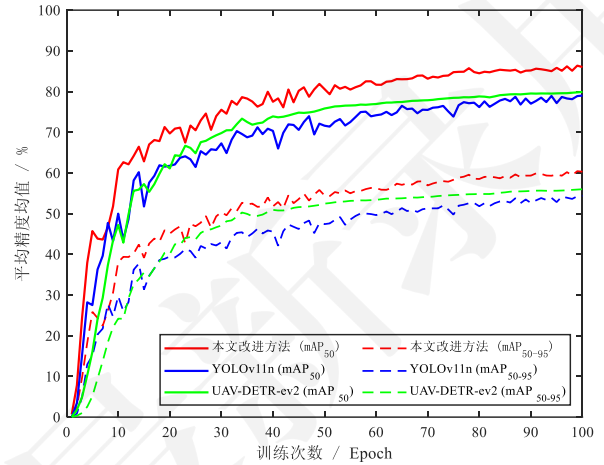


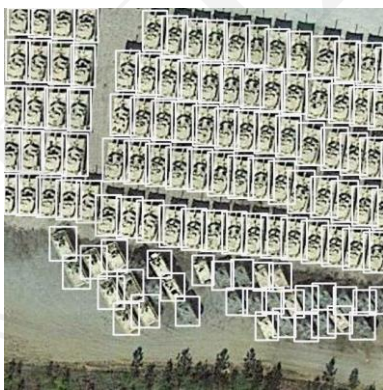
图 14 MVRSD mAP 收敛曲线对比

Fig. 14 Comparison of mAP convergence curves on the MVRSD

综上所述, GCSD-YOLOv11n 算法在保持轻量级模型复杂度的同时, 实现了精度的显著提升, 并兼顾推理速度, 在多目标检测尤其是小目标检测场景中具有良好的应用潜力。

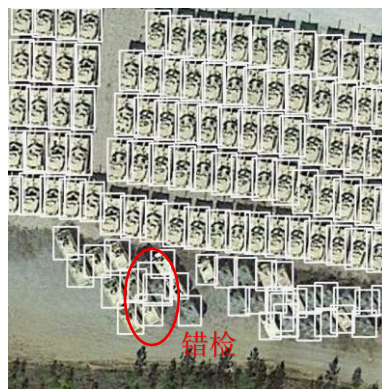
2.5 性能分析

为验证本文方法在真实军事遥感场景下的检测性能, 本文在 MVRSD 上对比评估了改进算法与基线 YOLOv11n 算法的检测精度。图 15 展示了在复杂战场背景下的检测效果, 表 8 给出了在不同军事目标类别上的详细检测精度指标。



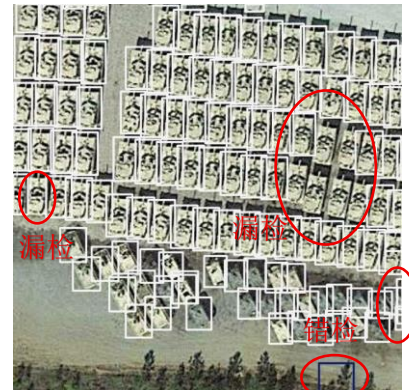
(a) GCSD-YOLOv11n 算法

(a) GCSD-YOLOv11n algorithm



(b) YOLOv11n 算法

(b) YOLOv11n algorithm



(c) UAV-DETR-ev2 算法

(c) UAV-DETR-ev2 algorithm



(d) GSCD-YOLOv11n 算法
(d) GSCD-YOLOv11n algorithm
(e) YOLOv11n 算法
(e) YOLOv11n algorithm
(f) UAV-DETR-ev2 算法
(f) UAV-DETR-ev2 algorithm

图 15 典型战术场景下不同算法检测效果对比

Fig. 15 Comparison of detection results of different algorithms under typical tactical scenarios

表 8 本文算法在 MVRSD 上各目标类别的检测精度

Table 8 Performance evaluation metrics of the proposed algorithm on MVRSD

类别	$P/\%$	$R/\%$	$mAP_{50}/\%$	$mAP_{50-95}/\%$
SMV	92.0	78.8	91.8	64.6
LMV	86.1	83.3	91.2	64.7
AFV	86.7	88.9	94.0	65.4
CV	94.9	94.6	97.9	73.6
MCV	56.5	62.1	57.0	33.8

从表 8 可以看出, GSCD-YOLOv11n 算法在各个类别上的检测精度均显著优于原始 YOLOv11n 算法, 体现出较强的普适提升能力。尤其是在军事目标细粒度特征显著且检测难度较大的类别中, 如 SMV、AFV 和 MCV, mAP_{50} 分别由 86.5%、88.5%、36.8% 提升至 91.8%、94.0%、57.0%, 分别提升了 5.3%、5.5%、20.2%, 提升幅度分别高达 6.1%、6.2% 和 54.9%, 显著缓解了原始模型在复杂战场背景下对微小目标及非典型装备的漏检与识别不准问题。

其中, 针对 SMV 类别, 改进算法引入的四通道输入与注意力增强结构有效提升了对细粒度军事小目标的辨识能力, 不仅提高了召回率 (R 从 71.9% 提升至 78.8%), 也减少了重复检测与错检现象, 表现出良好的稳定性与泛化性。在 LMV、CV 等典型目标类别上, 虽然原始基准已处于较高水平, 但改进算法依然实现了稳定提升。例如 LMV 类别的 mAP_{50} 从 85.9% 提升至 91.2%, CV 类别由 97.4% 提升至 97.9%, 表明本方法在面对不同尺度和数量分布时, 均能保持较强的目标识别能力与定位精度。此外, 在样本稀缺且特征复杂的低频目标 MCV 类别上, GSCD-YOLOv11n 算法的 mAP_{50} 从 36.8% 大幅跃升至 57.0%, 证明其具备极佳的小样本泛化能力, 能

够兼顾高频与长尾军事目标的检测需求。表 9 与图 15 展示了在相同测试条件下 (IoU 阈值设为 0.5, 置信度阈值设为 0.25), 基准算法 YOLOv11n、UAV-DETR-ev2 与本文改进算法在高密度同类集群(场景 A)与复杂背景异类分布 (场景 B) 两类典型场景下的检测目标数量统计。

表 9 对比实验中各算法的性能评价指标

Table 9 Comparison of the performance evaluation results of models in the experiment

算法	场景	目标总数	正确个数	漏检个数	误检个数
本文算法	A	119	119	0	0
YOLOv11n 算法	A	121	119	0	2
UAV-DETR 算法	A	128	112	7	16
本文算法	B	43	40	0	3
YOLOv11n 算法	B	46	35	5	11
UAV-DETR 算法	B	49	39	1	10

由表 9 可以看出, 在场景 A (高密度同类集群) 下, 由于军事目标排列极度密集, 基准算法表现出明显的虚警与性能退化。其中, YOLOv11n 算法虽然完成了全量检出, 但产生了 2 个误检目标; 而基于 Transformer 的 UAV-DETR-ev2 算法表现最差, 出现了 7 个目标的漏检且伴随 16 个误检。相比之下, 本文改进算法在该场景下实现了 119 个目标的精准全量检出, 且漏检数与误检数均为 0, 体现出极强的密集目标分离能力与特征分辨力。

在场景 B (复杂背景异类分布) 中, 针对散布的异类军事目标 (共计 40 个真实实例, 包括 21 个 SMV、12 个 LMV 和 7 个 AFV), 各算法在检测稳健性上展现出显著差异, 检测结果统计如表 9 所示。

如图 15 所示，基准算法 YOLOv11n 在该场景下表现出明显的不稳定性，共计检出 46 个目标，其中存在 5 个目标的漏检及 11 个误检。具体而言，其对 SMV 的检出数达 32 个（远高于实际 21 个），而对高价值目标 LMV 和 AFV 仅分别识别出 8 个与 6 个，漏检率较高。基于 Transformer 架构的 UAV-DETR-ev2 算法虽然具备一定的全局感知能力，共计检出 49 个目标，但其虚警问题严重，包含了 10 个误检且仍存在 1 个 LMV 目标的漏检，导致实战感知的可靠性受限。

与之形成鲜明对比的是，本文改进算法表现出卓越的探测可靠性，实现了 0 漏检的优异成绩。在总计检出的 43 个目标中，其对 LMV 和 AFV 均实现了全量精准识别，仅在 SMV 类别中产生了 3 个虚警。

综上所述，统计结果表明改进算法在面对战场复杂背景及细粒度目标识别时，具备更强的环境适配性，能够有效应对实战感知中密而不漏、疏而不误的严苛要求。

在该密集小目标场景中，UAV-DETR 等 Transformer 检测器由于采用多查询预测机制，在推理阶段可能对同一目标生成多个高度重叠的候选框，从而产生重复检测现象，如图 15 所示。该现象虽在定量评估中未必显著影响检测指标，但会降低检测结果的可读性并增加后处理复杂度。相比之下，本文改进的 YOLO 算法在保持检测精度的同时输出更加紧凑稳定，更适合无人机场景下的工程应用需求。

综上所述，本文方法在低照度、高密度及多尺度混合场景下表现出更稳定的检测性能，体现出较好的实际应用潜力。该性能提升主要源于算法引入的多尺度感知与注意力增强设计：Conditional-C3k2 与 DyHead-DCNv4 算法强化了小目标区域的特征响应能力，四通道图像输入为复杂背景下的边缘与纹理建模提供了有效补充。在 Tricycle、Motor 等低频类别上，改进算法仍保持较稳定的检出表现，表明其在类别不均衡场景中具备一定的泛化能力。综合定量与可视化结果可知，GCSD-YOLOv11n 算法在检测精度与复杂场景适应性方面优于 YOLOv11n 算法，具有良好的工程应用潜力。

3 工程部署与系统实现

3.1 多版本模型配置

为满足不同计算平台在指令集、并行能力与运行环境方面的差异化需求，本文在统一算法框架下设计了多种规模与配置的检测模型版本，以实现对不同工程应用场景的适配。具体而言，构建了 3 种不同复杂度的模型形式，分别记为 GCSD-

YOLOv11e、GCSD-YOLOv11n 和 GCSD-YOLOv11s。3 种模型在输入形式、主干网络结构以及特征增强策略上保持一致，均采用本文提出的多通道输入与改进特征提取框架；在此基础上，根据计算平台对并行算子与运行环境的支持能力，对检测头结构与网络容量进行差异化配置，从而在保证算法一致性的前提下，实现对工程可用性的有效拓展。

如表 10 所示，GCSD-YOLOv11-e 模型主要面向不具备 CUDA 加速环境的中低端计算平台，在保留主干特征提取结构与多通道输入形式的前提下，采用标准卷积检测头以替代依赖 GPU 并行算子的复杂检测头结构，从而确保模型在异构计算环境中的可移植性与稳定运行能力；GCSD-YOLOv11-n 模型作为本文的主要研究与验证版本，可部署于 Jetson 系列等边缘设备以及分体式检测设备，引入基于 DCNv4 的 DyHead 检测头结构，以增强对多尺度目标的空间建模与特征对齐能力，在检测精度、模型规模与推理效率之间取得较优平衡；GCSD-YOLOv11-s 模型在保持整体结构设计一致的前提下，进一步提升网络宽度与特征表达能力，适用于分体式检测架构下对检测精度要求更高的应用场景。

表 10 不同 GCSD-YOLOv11 模型在 Visdrone 2019 上的性能对比

Table 10 Performance comparison of different GCSD-YOLOv11 models on VisDrone 2019 dataset.

模型	mAP_{50}^o %	mAP_{50}^s %	$Par/10^6$	FLO Ps/ 10^9	时间/ms
GCSD-YOLOv11e 模型	41.6	25.1	1.49	18.1	14.5
GCSD-YOLOv11n 模型	45.1	27.4	1.85	22.5	17.6
GCSD-YOLOv11s 模型	50.4	30.8	6.73	76.3	48.3

通过上述多版本模型配置方式，本文算法在模型结构层面实现了对不同运行环境与算力条件的自适应扩展，在不改变核心算法思想的前提下，为后续面向多类计算平台的系统集成与应用提供了统一而灵活的模型基础。

3.2 检测与跟踪联合部署验证

为验证所提方法在实际应用场景中的部署效果与运行可行性，本文构建了基于无人机视频回传的部署验证平台，对所提出的 GCSD-YOLOv11e 与 GCSD-YOLOv11-n 两种模型配置进行实验验证。整体实验平台由无人机图像采集与图传模块、视频接收端以及计算处理端组成，其设备构成及视频数据流关系如图 16 所示。在该平台上，采用目标检测与

多目标跟踪相结合的处理方式，对无人机回传视频流进行连续处理，以验证模型在真实视频场景下的稳定运行能力。

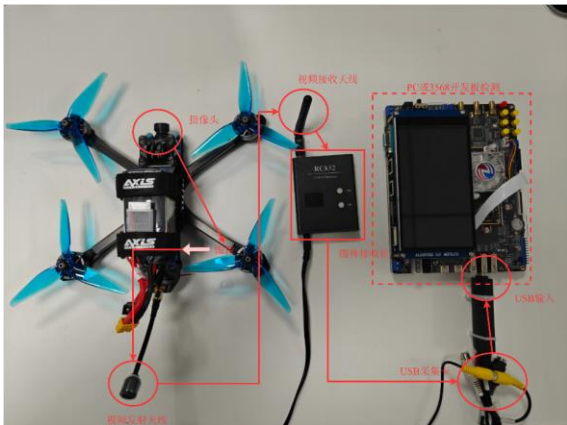


图 16 无人机视频回传检测与跟踪系统部署验证平台

Fig. 16 Architecture of the UAV Video-Based Detection and Tracking Deployment Verification Platform

其中，GCSD-YOLOv11e 模型的部署验证在基于 RK3568 的嵌入式计算平台上完成。为适配嵌入式运行环境，对本文算法进行 ONNX 导出与 RKNN 工具链转换，并完成量化编译，实现了无 CUDA 环境下的量化部署。经转换后的模型在嵌入式平台上对无人机回传视频流进行检测与跟踪处理，整体运行过程稳定，视频流处理流畅，能够满足实时处理需求，其检测结果如图 17 所示。

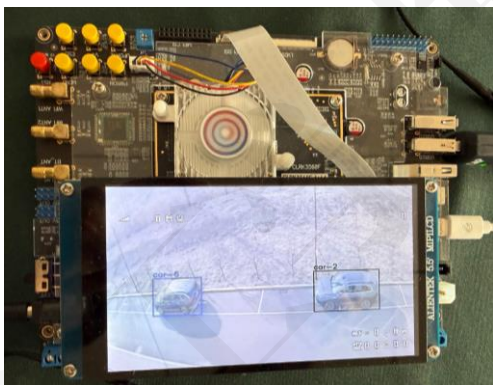


图 17 RK3568 平台检测结果

Fig. 17 RK3568 platform test results

GCSD-YOLOv11n 模型的部署验证在计算机端完成，依托具备 GPU 与 CUDA 支持的运行环境，对无人机回传视频进行检测与跟踪联合处理。实验过程中，模型能够持续输出稳定的检测结果，并与跟踪模块协同工作，如图 18 所示，实现目标在视频序列中的连续识别与状态更新，验证了该模型配置在计算机或 Jetson 系列开发板的可行性。



图 18 计算机检测结果

Fig. 18 PC test results

综合上述实验结果可知，所提模型在嵌入式与计算机两类典型计算平台上均能够稳定运行，在连续视频流条件下保持良好的处理连贯性与实时响应能力，表明本文模型具备较好的工程部署可行性与实际应用潜力。

4 结论

本文针对空基与星载军事侦察任务中小目标检测所面临的复杂环境与平台资源约束问题，提出一种融合多通道特征增强与轻量化结构设计的改进 YOLOv11n 算法，利用四通道图像增强表征能力，并在 VisDrone2019、MVRSD 与 MAR203 三个具有代表性的数据集上进行了验证，得出如下主要结论：

1) 本文基于图像四通道输入构建轻量化算法，使用灰度图作为辅助特征参与主干网络提取，有效提升了算法对细粒度纹理与细节的感知能力，增强了对军事小目标区域的聚焦效果。

2) 引入基于门控机制的 Conditional-C3k2 模块并结合 DyHead-DCNv4，实现了对多尺度目标特征的自适应建模与空间对齐优化，在参数量增长受控的前提下，提高了目标定位精度与检测置信度的稳定性。

3) 进一步采用轻量化的 LCSA 特征融合结构，以强化多尺度信息交互并优化高低层特征的流动路径，从而在保证检测精度的同时显著降低算法计算开销，实现了检测性能与部署效率的兼顾。

4) 在损失函数设计方面，引入 VSA-Loss 优化回归项，使算法在训练阶段更关注边界拟合质量与目标中心回归，有效提升了目标框的稳定性与预测精度。

综合实验结果表明，本文方法在多个数据集上均显著优于原始 YOLOv11n 算法。在通用小目标数据集 VisDrone2019 上， mAP_{50} 提升了 13.0%，整体提升幅度高达 40.5%；在军事遥感数据集 MVRSD 上， mAP_{50} 提升了 7.4%，达到 86.4%；在 mAP_{50-95} 指标上，最高提升幅度达 48.1%。此外，模型参数量大

幅压缩 29.4%，在保持轻量化优势的同时，具备极强的战场态势感知能力与工程应用潜力。未来工作将进一步结合模型量化与结构剪枝等压缩技术，持续提升算法在边缘端平台上的部署能力，并探索其在无人机、车载等资源受限场景中目标检测与跟踪任务中的应用价值。

参考文献 (References)

- [1] 张永梅,赖裕平,马健喆,等.基于视频的装甲车和飞机检测跟踪及轨迹预测算法[J].兵工学报,2021,42(3):545-554.
ZHANG Y M, LAI Y P, MA J Z, et al. Detection, tracking, and trajectory prediction algorithm for armored vehicles and aircraft based on video[J]. Acta Armamentarii, 2021, 42(3): 545–554. (in Chinese)
- [2] 赵海丽,包大洪,张从豪,等.基于改进 SDU-YOLOv8 的军事飞机目标检测算法[J].兵工学报, 2026, 47(1):250294.
ZHAO H L, BAO D Y, ZHANG C H, et al. Military aircraft target detection algorithm based on improved SDU-YOLOv8[J]. Acta Armamentarii, 2026, 47(1):250294. (in Chinese)
- [3] 刘克顺,左晓桐,张玉华,等.基于深度学习的无人机可见光目标检测研究综述[J].航空兵器,2024,31(6):23-35.
LIU K S, ZUO X T, ZHANG Y H, et al. Research review on UAV visible-light target detection based on deep learning[J]. Aeronautical Armament, 2024, 31(6): 23–35. (in Chinese)
- [4] DAI X Y, CHEN Y P, XIAO B, et al. Dynamic head: unifying object detection heads with attentions[C]// Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, TN, US:IEEE, 2021: 7369–7378.
- [5] XIONG Y W, LI Z Q, CHEN Y T, et al. Efficient deformable ConvNets: rethinking dynamic and sparse operator for vision applications[C]//Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, US:IEEE, 2024:5652-5661.
- [6] ZHANG H Y, WANG Y, DAYOUB F, et al. VarifocalNet: an IoU-aware dense object detector[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, TN, US:IEEE, 2021: 8510-8519.
- [7] GEVORGYAN Z. SiO loss: more powerful learning for bounding box regression: arXiv:2205.12740[R/OL]. Ithaca, NY, US: Cornell University, 2022(2022-05-25). <https://arxiv.org/abs/2205.12740>.
- [8] ZHU P F, WEN L Y, DU D W, et al. Detection and tracking meet drones challenge[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(11): 7380–7399.
- [9] 柏栋,于英,宋亮,等.面向军用车辆细粒度检测的遥感图像数据集构建与验证 [J]. 中国图象图形学报,2024,29(12):3564-3577.
BAI D, YU Y, SONG L, et al. Construction and validation of remote sensing image dataset for fine-grained detection of military vehicles[J]. Journal of Image and Graphics, 2024, 29(12): 3564-3577. (in Chinese)
- [10] 禹文奇,程琳,王美君,等.MAR20: 遥感图像军用飞机目标识别数据集[J].遥感学报,2023,27(12):2688-2696.
YU W Q, CHENG G, WANG M J, et al. MAR20: a benchmark for military aircraft recognition in remote sensing images[J]. Journal of Remote Sensing, 2023, 27(12): 2688–2696. (in Chinese)
- [11] 杨俭, 陈一剑, 耿樊骏喆. 乌克兰危机无人机作战运用及启示[C]// 第十三届中国指挥与控制大会论文集. 北京:兵器工业出版社, 2025: 205-208.
YANG J, CHEN Y J, GENG F J Z. Operational employment of unmanned aerial vehicles in the Ukraine crisis and its implications[C]// Proceedings of the 13th China Conference on Command and Control. Beijing: The Publishing House of Ordnance Industry,2025: 205–208. (in Chinese)
- [12] LOH Y P, CHAN C S. Getting to know low-light images with the Exclusively Dark dataset[J]. Computer Vision and Image Understanding, 2019, 178: 30-42.
- [13] TIAN Y J, Ye Q X, DOERMANN D. YOLOv12: Attention-centric real-time object detectors[C]//Advances in Neural Information Processing Systems (NeurIPS). La Jolla: Neural Information Processing Systems Foundation, 2025.
- [14] TANG Y H, HAN K, GUO J Y, et al. GhostNetV2: enhance cheap operation with long-range attention[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(8): 10144-10157.
- [15] MA N N, ZHANG X Y, ZHENG H T, et al. ShuffleNet V2: practical guidelines for efficient CNN architecture design[C]// Computer Vision – ECCV 2018. Cham, Switzerland:Springer, 2018, 11218: 122-138.
- [16] QIN D F, LEICHNER C, DELAKIS M, et al. MobileNetV4: Universal models for the mobile ecosystem[C]// Computer Vision – ECCV 2024. Cham, Switzerland:Springer, 2024, 15098: 78-96.
- [17] WANG C C, HE W, NIE Y, et al. Gold-YOLO: efficient object detection via gather-and-distribute mechanism: arXiv:2309.11331 [J/OL].Ithaca, NY, US: Cornell University, , 2023. <https://arxiv.org/abs/2309.11331>.

[18] 冯迎宾,郭泉尊,晏佳华.基于多尺度注意力机制的无人机小目标检测算法[J].兵工学报,2025,46(1):231124.

FENG Y B, GUO X Z, YAN J H. UAV small object detection algorithm based on multi-scale attention mechanism[J]. Acta Armamentarii, 2025, 46(1): 231124. (in Chinese)

[19]. ZHAO Y, LV W Y, XU S L, et al. DETRs beat YOLOs on

real-time object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2024: 22781-22790.

[20] ZHANG H X, LIU K, GAN Z X, et al. UAV-DETR: Efficient end-to-end object detection for unmanned aerial vehicle imagery[J/OL]. arXiv:2501.01855, 2025.