http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0509

轻量级多尺度特征融合增强的空间非合作小目标检测算法

刘振江[1]，张会娟[2,*]，姬淼鑫[1]，卢涛[1]，张弛[1]，余远金[2]

(1. 河南工业大学 电气工程学院，郑州 450001；2. 北京理工大学 自动化学院，北京 100081)

摘 要：针对深空背景下空间非合作小目标特征提取不充分而检测精度低、空间计算资源有限等问题，提出一种轻量级多尺度特征融合增强和上下文感知的空间目标检测算法。设计一种轻量级多特征下采样模块实现空间非合作目标的多尺度特征提取，加强对细粒度特征提取能力；设计快速双空间金字塔池化模块加强对目标全局上下文特征提取，并聚合局部显著特征和全局特征以提高特征泛化能力；设计跨维度三重特征增强模块（CTFEM）对不同维度信息进行全局上下文建模，进一步增强小目标特征信息，抑制易混淆背景信息，从而提高对小目标检测能力。使用VisDrone2019公开的小目标数据集和自建空间态势感知数据集（SSAD）对所提算法进行验证，实验结果表明：所提算法与基准模型YOLOv8n相比，模型参数量减少19.38%，在SSAD数据集上mAP50和mAP50-95分别提高0.2%和1.2%，在VisDrone2019数据集上mAP50和mAP50-95分别提高2.0%和1.4%。

关 键 词：空间非合作目标检测；小目标检测；特征融合；上下文信息；轻量级网络
中图分类号：V44; V52; TP391.41
文献标志码：A **文章编号**：1001-5965(2026)11-4040-11

随着空间目标数量的逐年增加，空间态势感知对保证在轨航天器安全起着重要的作用[1]，而空间目标检测是实现空间态势感知的基础。空间任务包括空间目标监测、回避和清除，需要尽早实时识别和监测空间非合作目标[2-3]，避免潜在碰撞，规划有效轨迹，以确保在轨服务卫星的正常运行[4-5]。因此，快速准确检测和识别远距离非合作航天器和空间碎片等小目标具有重要意义。

基于图像的深度学习目标检测[6-9]为空间态势感知带来诸多优势。Wu等[10]针对空间目标识别和定位效率低的问题，提出了两阶段卷积神经网络(two-stage convolutional neural network, T-SCNN)。Xiang等[11]提出了基于网格学习的快速空间碎片检测方法，将图像划分为14×14个网格，使用基于网格的快速神经网络(fast grid-based neural network, FGBNN)定位网格内的空间碎片。Wang等[12]在YOLOv3网络基础上改进了特征提取结构，融合浅层特征和深层特征，增强网络对不同尺度目标的检测能力。Jia等[13]将改进的Resnet-50骨干网络和特征金字塔网络融合到Faster R-CNN框架[14]，增强了模型的泛化能力。Aldahoul等[15]为提高对空间目标的定位和分类，提出了一种模型决策融合方法，利用EfficientNet-v2和EfficientNet-B4网络训练EfficientDet模型。Mahendrakar等[16]基于形状和

收稿日期：2024-07-03；**录用日期**：2024-09-13；**网络出版时间**：2024-09-20 10:37
网络出版地址：link.cnki.net/urlid/11.2625.V.20240919.1640.008
基金项目：国家资助博士后研究人员计划(GZC20233408)；国家自然科学基金(62222304,62173039,62201199)；河南工业大学创新基金支持计划专项资助(2021ZKCJ07)
* **通信作者**. E-mail：zhanghjqy@bit.edu.cn

引用格式：*刘振江，张会娟，姬淼鑫，等. 轻量级多尺度特征融合增强的空间非合作小目标检测算法[J]. 北京航空航天大学学报，2026，52(11)：4040-4050. Liu Z J，Zhang H J，Ji M X，et al. Lightweight multi-scale feature fusion enhancement algorithm for spatial non-cooperative small target detection[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(11)：4040-4050 (in Chinese).*

纹理的人类启发决策过程, 将形状检测器和纹理检测器与 YOLOv5 相结合, 增强安全决策能力, 提出了 SpaceYOLO 空间目标检测方法。Tang 等[17] 使用 RepPoint 替换 YOLOv7 中 ElAN 骨架的 3×3 卷积, 提高了检测空间目标组件的准确率。Ai 等[18] 采用 3 个一级特征图代替特征金字塔结构和解耦头, 增强对不同尺度空间目标的检测。Massimi 等[5] 将脉冲多普勒雷达系统与 YOLO 相结合, 提高了空间碎片探测能力。这些工作对于特征明显的目标具有较好的分类检测性能。

为了进一步提高对空间小目标的检测精度, Guo 等[19] 提出了基于空间图像特征的通道和空间注意力 U-Net 网络, 实现了单幅光学图像空间弱小目标检测。Jiang 等[20] 提出多通道直方图截断模块和中心差分卷积对特征进行增强, 构建了 SOD-YOLO 实现弱小空间目标检测。上述文献在简单背景和特征单一的小目标检测方面有一定的效果, 但复杂背景下的检测精度有待提高。Tao 等[21] 引入了增强的空间特征模块来增强小目标的空间细节, 构建空间碎片显著性检测算法 SDebrisNet。Yuan 等[22] 提出交叉上下文融合模块, 通过多分支并行学习不同尺度的上下文信息, 增强特征表达能力, 用于检测小型和微弱的空间目标。上述文献对小目标特征提取能力有一定的增强, 但模型复杂度在较高有限资源下应用受限。

上述工作提升了对空间目标检测的性能, 但是对于空间非合作目标检测, 深空图像具有杂散光、低信噪比等特性。现有的空间小目标检测方法在复杂背景中表现不佳, 尤其在特征提取能力和对小目标的优化方面存在不足, 检测精度下降。此外, 由于模型复杂度较高, 在星上资源受限的环境中难以实现实时检测。因此, 本文在 YOLOv8n 网络模型的基础上提出一种轻量级空间非合作小目标检测网络 S-YOLO, 其兼顾检测精度与模型轻量化, 具体创新点如下:

1) 为解决空间目标数据样本难获取的问题, 构建了以小尺度目标为主的空间态势感知数据集(spatial situational awareness detection, SSAD), 包含 22 个空间目标类别、50 563 张图片、68 925 个标签, 可为空间非合作目标探测提供大规模基准数据集;

2) 提出一种轻量级多特征下采样模块(light-weight multi-feature down sampling module, LMFD), 改善跨步标准卷积丢失目标信息问题, 增强对细粒度特征提取能力;

3) 提出快速双空间金字塔池化模块(double spatial pyramid pooling-fast, DSPPF), 加强对目标全局上下文特征提取, 融合目标局部显著特征, 提高特征表达能力, 进一步增强对小目标检测;

4) 提出跨维度三重特征增强模块(cross-dimension triplet feature enhance module, CTFEM), 在颈部跨通道进行全局上下文建模, 增强小目标弱特征表示。同时, 引入动态上采样模块(upsampling by dynamic sampling, DySample), 缓解特征丢失问题, 提高特征的表达能力和精确度, 增强小目标检测性能。

1　空间非合作小目标检测算法

YOLOv8 采用单阶段端到端架构[23–24], 包括骨干网络(Backbone)、颈部(Neck)和头部(Head)。骨干网络主要由 C2f 和 CBS 模块组成, 对输入图像执行四次下采样提取特征。C2f 模块在保证轻量级的同时实现了丰富梯度信息采集。颈部网络采用特征金字塔网络(feature pyramid networks, FPN)和路径聚合网络(path aggregation network, PAN)结构, 通过组合不同尺度特征, 增强网络特征融合能力。YOLOv8 头部采用解耦合结构, 包含 3 个不同尺度检测分支, 利用非极大值抑制得到最优检测框。

为提升空间非合作小目标检测能力, 关键在于增强模型对低信噪比、低光照、小尺度空间目标的特征提取能力, 同时还应考虑网络轻量化, 以满足实时性和稳定检测环境需求。因此, 本文提出基于 YOLOv8n 的空间非合作目标检测模型 S-YOLO。S-YOLO 的整体结构如图 1 所示。

1.1　轻量级多特征下采样模块

空间非合作小目标特征信息不明显, 使用跨步标准卷积实现下采样, 模型提取细节特征能力有限, 随着网络深度增加, 将导致细节信息丢失, 空间信息的丢失使模型难以识别检测小尺度目标。因此, 为了提取更丰富的空间非合作目标特征, 本文提出了 LMFD, 如图 2 所示。

首先, 使用 1×1 卷积将输入特征从通道 C_1 提高到 C_2, 不改变图像分辨率, 在增加非线性特征的同时获得更多通道特征。然后, 进行通道分割, 送入 2 个不同特征提取分支。其中, 上分支先进行 1×1 逐点卷积[25], 减少计算量和内存消耗, 在保持空间分辨率的同时实现不同通道信息融合, 有利于提升模型表达能力。接着采用 3×3, 步距为 2 的深度卷积[25], 在减少参数量的同时提取更多空间特征。最后, 再使用 1×1 逐点卷积进行特征融合。其中, 下分支直接使用 5×5, 步距为 2 的深度卷积。2 个不同尺寸的卷积核可扩展网络感受野, 更好地提取不同尺度空间目标特征, 提高模型对不同尺度目标的检测能力。2 个分支的输出在通道上进行拼接,

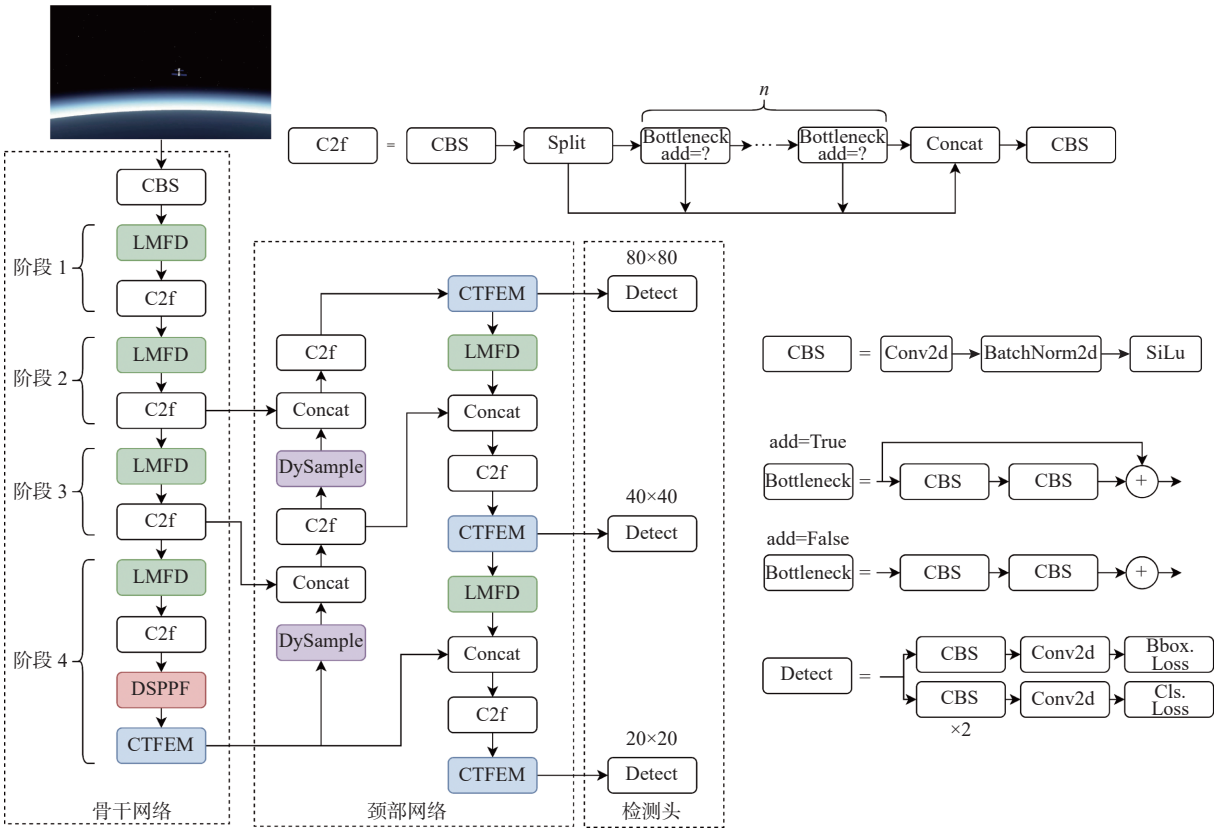


图 1 S-YOLO 网络结构

Fig. 1 Structure of S-YOLO network

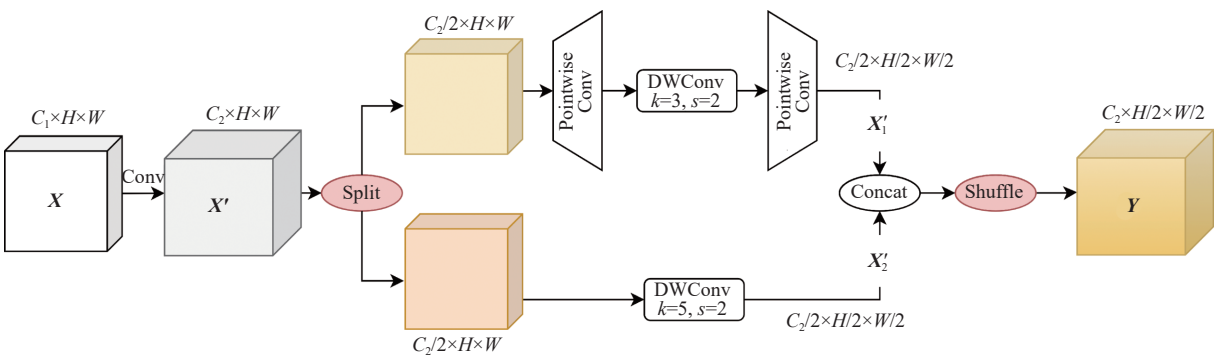


图 2 LMFD 结构

Fig. 2 Structure of LMFD

最终在通道上进行混洗[26], 可促进不同通道信息交流, 更好地融合不同通道信息, 进一步增强模型表达能力和鲁棒性。

LMFD 数学表达式如下:

$$\boldsymbol{X}' = f_{\text{Conv}}^{1\times1}(\boldsymbol{X}) \tag{1}$$

$$\boldsymbol{X}_1' = f_{\text{PWConv}}^{1\times1}(f_{\text{DWConv}}^{3\times3,2}(f_{\text{PWConv}}^{1\times1}(\text{Split}(\boldsymbol{X}')))) \tag{2}$$

$$\boldsymbol{X}_2' = f_{\text{DWConv}}^{5\times5,2}(\text{Split}(\boldsymbol{X}')) \tag{3}$$

$$\boldsymbol{Y} = \text{Shuffle}(\text{Cat}(\boldsymbol{X}_1', \boldsymbol{X}_2')) \tag{4}$$

式中: $f_{\text{Conv}}^{1\times1}$ 表示卷积核为 1×1 的标准卷积; $f_{\text{PWConv}}^{1\times1}$ 表示卷积核为 1×1 的逐点卷积; $f_{\text{DWConv}}^{3\times3,2}$ 和 $f_{\text{DWConv}}^{5\times5,2}$ 分别表示步距为 2, 卷积核为 3×3 和 5×5 的深度卷积; Split(·)表示特征图通道切分操作; Cat(·)表示特征图通道方向拼接操作; Shuffle(·)表示特征图通道混洗操作; $\boldsymbol{X}$表示输入特征图, $\boldsymbol{X}'$由式 (1) 得到; $\boldsymbol{X}_1'$和 $\boldsymbol{X}_2'$分别表示上分支和下分支输出; $\boldsymbol{Y}$表示 LMFD 输出特征图。

此外, 模块的轻量化设计主要通过逐点卷积和深度卷积实现。标准卷积参数量为$C_{\text{in}}\times C_{\text{out}}\times k\times k$, 而本文设计的 LMFD 卷积参数量为$C_{\text{in}}\times C_{\text{out}}+C_{\text{out}}/2\times(k_1\times k_1+k_2\times k_2+C_{\text{out}})$。假设$C_{\text{in}}=64$, $C_{\text{out}}=128$, $k=3$, $k_1=3$, $k_2=5$, 标准卷积参数量为 73 728, LMFD 的卷积参数量为 18 560, 则 LMFD 参数量减少了 74.83%。

这使得模型在保证多特征提取能力的同时, 极大地降低了计算复杂度。

1.2　快速双空间金字塔池化模块

针对空间非合作小目标检测, 不仅要关注局部特征信息, 也需要更多的全局上下文特征信息。SPP[27] 和 SPPF[28] 模块都采用最大池化提取不同尺度局部特征, 忽略了全局上下文特征的提取。因此, 为了聚合局部显著特征和全局上下文特征, 加强特征泛化能力, 本文提出 DSPPF, 如图 3 所示。

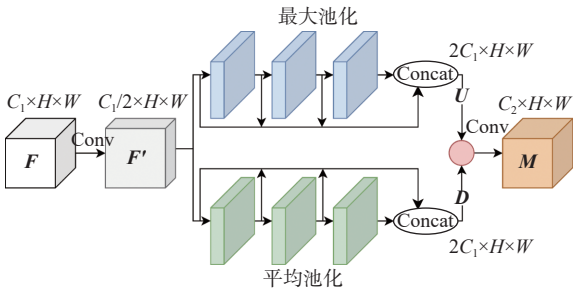


图 3　DSPPF 结构

Fig. 3　Structure of DSPPF

DSPPF 采用双分支特征提取, 分别使用最大池化和平均池化提取空间目标的局部特征和全局特征, 如图 3 所示。平均池化能够保留更多背景细节信息和上下文信息, 最大池化提取局部显著特征, 两者结合可生成更加丰富的特征, 这对空间目标检测非常有利。首先, 利用 1×1 标准卷积对输入特征图 $\boldsymbol{F}$ 降维操作, 得到特征图 $\boldsymbol{F}'$, 减少计算复杂度。然后, 对 $\boldsymbol{F}'$ 分别进行金字塔最大池化和平均池化, 得到特征图 $\boldsymbol{U}$ 和 $\boldsymbol{D}$。最后, 使用逐元素加法进行特征融合, 得到特征图 $\boldsymbol{M}$。具体公式如下:

$$\boldsymbol{U}=\mathrm{Cat}(P_{\mathrm{m}}(\boldsymbol{F}'),P_{\mathrm{m}}(P_{\mathrm{m}}(\boldsymbol{F}')),P_{\mathrm{m}}(P_{\mathrm{m}}(P_{\mathrm{m}}(\boldsymbol{F}'))),\boldsymbol{F}') \tag{5}$$

$$\boldsymbol{D}=\mathrm{Cat}(P_{\mathrm{a}}(\boldsymbol{F}'),P_{\mathrm{a}}(P_{\mathrm{a}}(\boldsymbol{F}')),P_{\mathrm{a}}(P_{\mathrm{a}}(P_{\mathrm{a}}(\boldsymbol{F}'))),\boldsymbol{F}') \tag{6}$$

$$\boldsymbol{M}=\boldsymbol{U}\oplus\boldsymbol{D} \tag{7}$$

式中: $P_{\mathrm{m}}(\cdot)$ 和 $P_{\mathrm{a}}(\cdot)$ 分别表示最大池化和平均池化; $\oplus$ 表示广播逐元素加法。

1.3　跨维度三重特征增强模块

针对小尺度的空间非合作目标, 当处于低信噪比和低光照的相似背景下时, 提取目标特征相对困难。在 LMFD 和 DSPPF 之后, 特征图已经考虑了局部显著特征、全局上下文信息, 并捕获了空间非合作目标不同尺度特征。在颈部建模小目标和背景之间的全局关系比在骨干网络更有效。全局上下文信息可用于表示跨空间像素之间的关系, 增强空间目标和背景之间的区分度, 帮助模型更好地实现目标的定位检测。因此, 本文提出 CTFEM, 具体结构如图 4 所示。

增强的通道注意力模块 (enhanced channel attention module, ECAM) 如图 4(a) 所示, 在不降维的情况下, 使用自适应最大池化[29] 和自适应平均池化[30] 聚合卷积特征后, 进行广播逐元素相乘。然后, 执行卷积核为 3 的 1D 卷积和 Sigmoid 函数来学习通道注意力权重 $\boldsymbol{\psi}$。最后, $\boldsymbol{\psi}$ 与输入特征 $\boldsymbol{F}_{\mathrm{c}}$ 进行广播逐元素相乘得到输出特征 $\boldsymbol{F}'_{\mathrm{c}}$。具体表达式如下:

$$\boldsymbol{\psi}=\mathrm{Sigmoid}(f^{3}_{\mathrm{conv1d}}(P_{\mathrm{am}}(\boldsymbol{F}_{\mathrm{c}})\odot P_{\mathrm{av}}(\boldsymbol{F}_{\mathrm{c}}))) \tag{8}$$

$$\boldsymbol{F}'_{\mathrm{c}}=\boldsymbol{\psi}\odot\boldsymbol{F}_{\mathrm{c}} \tag{9}$$

式中: $P_{\mathrm{am}}(\cdot)$ 和 $P_{\mathrm{av}}(\cdot)$ 分别表示自适应最大池化和自适应平均池化; f^{3}_{conv1d} 表示卷积核为 3 的 1D 卷积; Sigmoid(·) 表示 Sigmoid 函数; $\odot$ 表示广播逐元素相乘。

空间注意力模块 (spatial attention module, SAM) 如图 4(b) 所示, 利用特征之间的空间关系生成空间注意力图, 与通道注意力互补。首先, 将输入特征 $\boldsymbol{F}_{\mathrm{s}}$ 沿通道应用最大池化和平均池化, 并进行直接拼接生成一个有效的特征描述子。在保留丰富特征的同时, 收缩深度, 进一步降低计算复杂度。然后, 应用 7×7 标准卷积层生成空间注意力特征图。最后, 使用 Sigmoid 函数进行激活得到 $\boldsymbol{F}'_{\mathrm{s}}$。SAM 具体表达式如下:

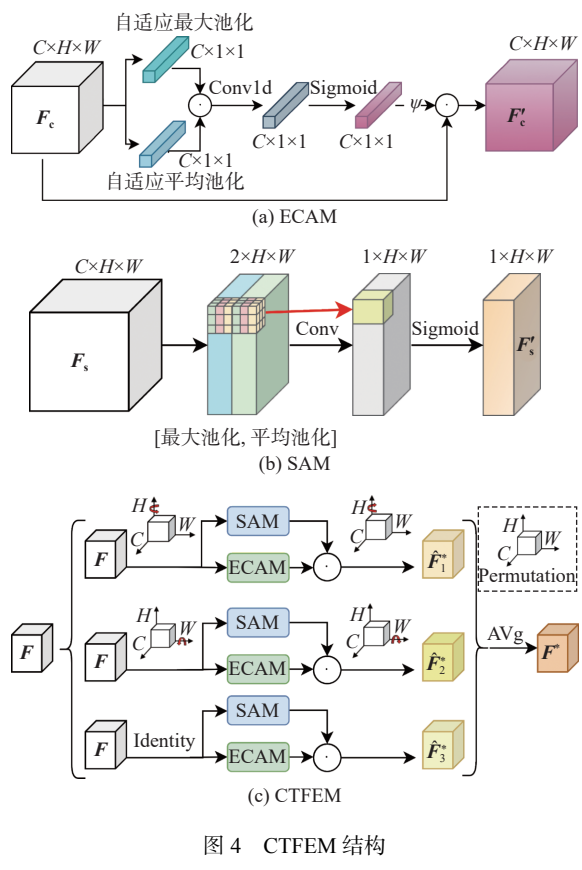


图 4　CTFEM 结构

Fig. 4　Structure of CTFEM

$$\boldsymbol{F}'_{\mathrm{s}}=\mathrm{Sigmoid}(f_{\mathrm{Conv}}^{7\times7}(\mathrm{Cat}(P_{\mathrm{m}}(\boldsymbol{F}_{\mathrm{s}}),P_{\mathrm{a}}(\boldsymbol{F}_{\mathrm{s}})))) \tag{10}$$

式中：$f_{\mathrm{Conv}}^{7\times7}$表示卷积核为7×7的标准卷积。

跨维度三重特征增强模块(CTFEM)具体结构如图4(c)所示，CTFEM利用三分支结构来捕获跨维度交互，以计算注意力权重。顶部分支负责计算通道维度C和空间维度H的注意力权重，中间分支计算通道维度C和空间维度W的注意力权重，最后一个分支捕获空间维度H和W之间的交互关系。当输入特征图$\boldsymbol{F}\in\mathbf{R}^{C\times H\times W}$，3个分支具体实现如下：

在顶部分支和中间分支，首先，输入特征图$\boldsymbol{F}$分别沿H轴和W轴逆时针旋转90°，分别建立维度W和通道C及维度H和通道C之间的交互，以获得旋转特征图$\hat{\boldsymbol{F}}_1\in\mathbf{R}^{W\times H\times C}$和$\hat{\boldsymbol{F}}_2\in\mathbf{R}^{H\times C\times W}$；然后，分别经过SAM和ECAM模块的输出$\boldsymbol{F}'_{\mathrm{c}}$和$\boldsymbol{F}'_{\mathrm{s}}$直接进行逐元素相乘获得特征图$\hat{\boldsymbol{F}}_1^*\in\mathbf{R}^{W\times H\times C}$和$\hat{\boldsymbol{F}}_2^*\in\mathbf{R}^{H\times C\times W}$；最后，分别沿$H$轴和$W$轴顺时针旋转90°以保持如输入特征图形状一致。对于底部分支，输入特征图$\boldsymbol{F}$直接分别经过SAM和ECAM模块，2个模块输出$\boldsymbol{F}'_{\mathrm{c}}$和$\boldsymbol{F}'_{\mathrm{s}}$直接进行逐元素相乘获得特征图$\hat{\boldsymbol{F}}_3^*\in\mathbf{R}^{C\times H\times W}$。通过简单平均来聚合由3个分支中的每个生成的形状为$(C\times H\times W)$的增强特征图$\boldsymbol{F}^*$，具体表达式如下：

$$\boldsymbol{F}^*=\frac{1}{3}(\hat{\boldsymbol{F}}_1^*+\hat{\boldsymbol{F}}_2^*+\hat{\boldsymbol{F}}_3^*) \tag{11}$$

通过结合跨维度信息交互，该方法解决了使用传统方法计算单个维度的注意力权重时出现的信息表达能力不足问题，能够在学习任务中捕获更复杂的关联和依赖关系，更有利于空间小尺度目标的检测。

1.4 动态上采样

YOLOv8目标检测算法应用最近邻插值法实现特征图的上采样操作，从而实现不同层之间的特征融合。虽然计算量小、运算速度快，但简单的最近邻插值法[31]会导致细节信息丢失，对于空间小尺度目标特征损失较严重，特征表达能力不足，随着网络深度传递会严重影响空间小目标检测精度。因此，本文引入了一种超轻量可学习的动态上采样(DySample)[32]，来缓解特征丢失问题，提高特征表达能力和精确度。具体结构如图5所示。

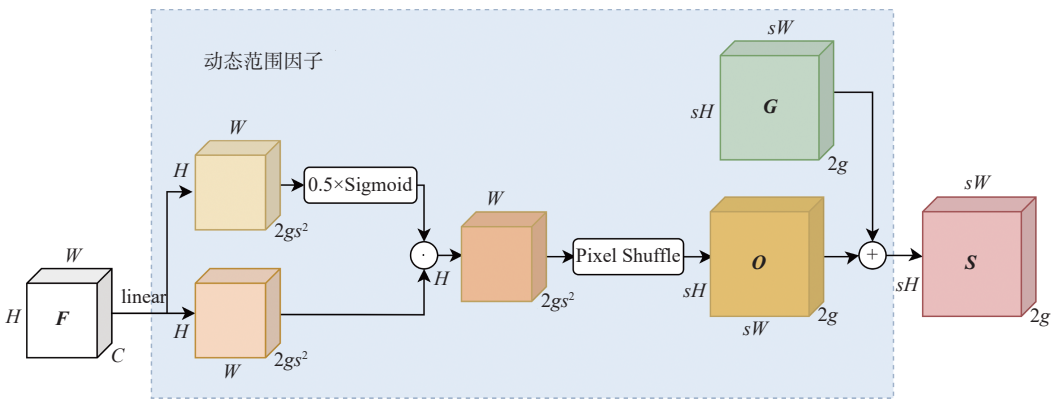


图5 动态上采样点生成器

Fig. 5 Dynamic upsampling point generator

给定上采样比例因子s和大小为$C\times H\times W$的特征图$\boldsymbol{F}$，沿着通道维度将特征图划分为$g(g=4)$组，可进一步降低计算量。经过一个线性层，生成大小为$2gs^2\times H\times W$的偏移。为了增加偏移的灵活性，通过使用Sigmoid函数和0.5静态因子进一步生成逐点的“动态范围因子”，动态范围取[0,0.5]范围内的值。最后，通过像素洗牌P_{s}将其重塑为$2g\times sH\times sW$的偏移$\boldsymbol{O}$，数学表达式如下：

$$\boldsymbol{O}=P_{\mathrm{s}}(0.5\times\mathrm{Sigmoid}(\mathrm{linear}(\boldsymbol{F})\odot\mathrm{linear}(\boldsymbol{F}))) \tag{12}$$

则采样集$\boldsymbol{S}$为偏移量$\boldsymbol{O}$与原始采样网格$\boldsymbol{G}$之和：

$$\boldsymbol{S}=\boldsymbol{G}+\boldsymbol{O} \tag{13}$$

2 实验与分析

本文使用的操作系统为Ubuntu 22.04，CUDA版本为12.1；实验在NVIDIA RTX 4090 (24 G) GPU和16 vCPU Intel(R) Xeon(R) Platinum 8352V CPU @ 2.10 GHz的计算机上进行；开发语言为Python 3.10和PyTorch 2.1.0。在训练过程中，部分关键参数设置如表1所示，其他参数使用YOLOv8默认参数。其中，优化器为随机梯度下降(stochastic gradient descent, SGD)。

本文采用准确率P、召回率R、mAP50、mAP50-95、APS、APM、APL、浮点运算量、检测速度、参数量、权重文件等指标作为模型评价指标。其中，浮点运算量、参数量和权重文件衡量目标检测模型大

表 1　网络模型的训练参数

Table 1　Training parameters of the network model

参数	数值
训练轮数	200
批次大小	32
输入图像尺寸/(像素×像素)	640×640
早停耐心值	100
初始学习率	0.01
最终学习率系数	0.01
动量系数	0.937
权重衰减系数	0.000 5
Mosaic数据增强系数	1.0

小, 检测速度代表模型每秒检测图像帧数, 衡量模型实时性。mAP50 表示 IoU 为 0.5 时所有类别的平均精度; mAP50-95 表示在 0.50～0.95 之间不同 IoU 阈值下计算得出的平均精度均值; APS 表示目标像素面积小于 32^2 所有类别的平均精度; APM 表示目标像素面积大于 32^2 并小于 96^2 所有类别的平均精度; APL 表示目标像素面积大于 96^2 所有类别的平均精度。

2.1　SSAD 数据集构建

为了构建空间态势感知数据集 SSAD, 本文采用网络图片、视频收集联合 STK(Satellite Tool Kit)软件仿真的方式获取大量的卫星和空间碎片图像进行扩展和标注。此外, 在制作 SSAD 数据集时采集相对正常运行航天器较远处的非合作目标, 此时空间非合作目标更多成像为小尺度目标。为丰富样本数据多样性, 利用 STK 仿真视频模拟了卫星不同姿态、尺度、背景等, 每个视频序列包含 1～2 类卫星。通过网络搜集的未知型号卫星在 SSAD 中统一归为 Satellite 类别。数据集类别及标签 ID 如表 2 所示。

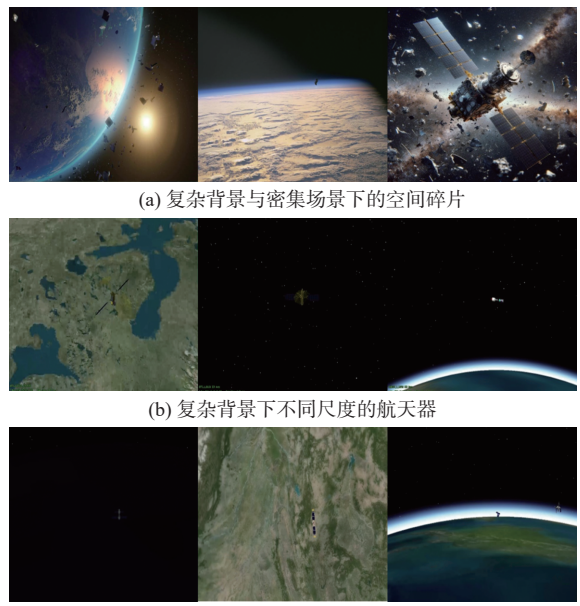
表 2　SSAD 数据集类别及标签统计

Table 2　SSAD dataset categories and labelling statistics

ID	类别名称	标签数量	ID	类别名称	标签数量
0	space debris	3 342	11	hubble	3 117
1	satellite	2 710	12	ikonos	3 221
2	aces	3 167	13	inmarsat4	3 070
3	apollo-soyuz	3 287	14	intelsat8	3 251
4	beidou	3 220	15	iss_stg3r	2 996
5	chandra	3 095	16	meteor	3 051
6	discoverer	3 123	17	mgs	3 287
7	dmsp	3 216	18	milstar	3 062
8	est-vil	3 149	19	mir	3 055
9	gps2	3 138	20	msx	3 271
10	gravityprobe	3 002	21	quickbird2	3 095

因此, SSAD 数据集共包含 22 种类别空间目标, 共 50 563 张图片, 68 925 个标签。仿真模拟生成的图像分辨率大小为 1 440×880, 为了保证数据的均衡性, 平均每类目标的标签约 3 000 个。数据使用 X-AnyLabeling 半自动标注软件完成, 首先每类人工标注 500 张, 使用 YOLOv8n 训练 300 个 epoch, 将训练权重导出为 ONNX 格式, 再导入 X-AnyLabeling 软件, 完成剩余图像标注并进行人工修正。

SSAD 数据集按照 6:2:2 的比例划分为训练集、验证集、测试集。图 6 为 SSAD 数据集中样本图像。



(a) 复杂背景与密集场景下的空间碎片

(b) 复杂背景下不同尺度的航天器

(c) 复杂背景下不同型号的航天器

图 6　SSAD 数据集中样本图像

Fig. 6　Sample images of SSAD dataset

2.2　消融实验

本文基于 YOLOv8n 提出了空间目标检测模型 S-YOLO, 使用 LMFD 替换跨步卷积下采样模块, 并使用 DSPPF、CTFEM 和 DySample 模块对该模型进行优化。为了评估每个优化模块的性能, 采用变量控制方法进行消融实验。在 VisDrone2019[33] 公开小目标数据集上用相同训练参数, 训练 200 个 epoch 的结果如表 3 所示, 表中加粗数值表示本文提出的 S-YOLO 模型的评价指标结果。

从表 3 中可观察到, 采用 LMFD 模块的模型 2 与模型 1(YOLOv8n) 相比, 在 mAP50 和 mAP50-95 上分别提高了 0.5% 和 0.6%, 但将网络模型参数数量从 3.2×10^6 减少到 2.57×10^6, 减少 19.69%, 浮点运算量降低了 2×10^8 次。模型 3 在模型 2 基础上增加 DSPPF 模块, 提高了准确率 P、召回率 R、mAP50 和 mAP50-95, mAP50 和 mAP50-95 分别提高 0.5% 和 0.2%。模型 4 在模型 3 基础上增加 CTFEM 模块, mAP50 和 mAP50-95 分别提高 0.4% 和 0.3%, 同时浮点运算量增加了 2×10^8 次, 但与模型 1 保持一致。最后, 模型 5(S-YOLO)在模型 4 基础上引入

表3 S-YOLO的消融实验结果

Table 3 Ablation experiment results of S-YOLO

模型	LMFD	DSPPF	CTFEM	DySample	参数量	浮点运算量/10^9次	P/%	R/%	mAP50/%	mAP50-95/%
1					3.20×10^6	8.2	43.2	32.4	32.1	18.3
2	√				2.57×10^6	8.0	43.8	32.8	32.6	18.9
3	√	√			2.57×10^6	8.0	43.9	34.0	33.1	19.1
4	√	√	√		2.57×10^6	8.2	44.3	**34.1**	33.5	19.4
5	√	√	√	√	2.58×10^6	8.2	**45.8**	33.7	**34.1**	**19.7**

了DySample组件，mAP50和mAP50-95分别提高了0.6%和0.3%，同时准确率P和召回率R也达到了最优，分别为45.8%和33.7%。此外，与基线模型YOLOv8n相比，S-YOLO模型在准确率P、召回率R、mAP50和mAP50-95分别提高2.6%、1.3%、2.0%和1.4%，达到45.8%、33.7%、34.1%和19.7%，而模型参数量仅为2.58×10^6，减少19.38%。消融实验结果证明了改进策略的有效性，特别是在提高小目标检测精度和减少参数量方面。

2.3 SSAD数据集实验结果

为验证本文算法的性能，本文将S-YOLO模型与最先进的目标检测方法YOLOv3-tiny[34]、YOLOv5n、YOLOv6n[35]、YOLOv7-tiny[36]、YOLOv8n[28]、LACTA[37]、YOLOv10n[38]、CS-YOLO[22]、在SADD数据集上进行比较，结果如表4所示，表中加粗数值表示各列评价指标的最优值。其中，CS-YOLO和LACTA是基于YOLO针对小目标改进的检测算法。对于空间目标探测任务，不仅要关注能否找到空间目标，还要关注空间目标的位置，因此，要关注mAP50和mAP50-95这2个重要指标。为了更好地衡量不同尺寸目标的检测性能，引入APS、APM、APL评估指标。

表4 不同检测算法在SSAD数据集上的实验结果

Table 4 Experimental results of different detection algorithms on SSAD dataset

模型	参数量	浮点运算量/10^9次	P/%	R/%	mAP50/%	mAP50-95/%	APS/%	APM/%	APL/%	检测速度/(帧·s^{-1})	模型权重/MB
YOLOv3-tiny*[34]	12.14×10^6	19.1	94.2	87.6	92.2	79.2	36.5	78.1	92.8	**240**	24.4
YOLOv5n*	2.51×10^6	**7.2**	96.8	94.6	96.0	83.1	48.5	80.3	93.0	138	5.3
YOLOv6n*[35]	4.24×10^6	11.9	96.4	94.4	95.7	82.4	46.5	80.1	93.0	163	8.7
YOLOv7-tiny*[36]	8.12×10^6	21.3	96.5	94.5	95.9	82.6	48.4	80.2	92.9	140	16.5
YOLOv8n[28]	3.20×10^6	8.2	96.9	94.7	96.2	83.8	48.6	80.6	93.9	150	6.3
CS-YOLO[22]	7.18×10^6	15.1	96.7	94.6	96.2	83.3	48.6	80.7	94.0	87	14.7
LACTA[37]	**1.20**$\times10^6$	9.6	96.2	94.2	95.9	81.8	48.5	80.4	93.0	157	**2.6**
YOLOv10n[38]	2.71×10^6	8.4	96.3	94.5	95.8	81.9	48.5	80.4	93.8	111	5.8
S-YOLO	2.58×10^6	8.2	96.9	94.9	**96.4**	**85.0**	**49.8**	**82.4**	**94.0**	77	5.5

注：*表示将网络模型的骨干网络移植到YOLOv8工程。

从表4可以得到，S-YOLO在mAP50、mAP50-95、APS、APM、APL指标上分别达到96.4%、85.0%、49.8%、82.4%、94.0%，优于基线YOLOv8n。与最新发布的YOLOv10n相比，mAP50、mAP50-95和APS分别提高0.6%、3.1%和1.3%。与YOLOv3-tiny相比，S-YOLO的mAP50、mAP50-95和APS指标分别提高4.2%、5.8%和13.3%。与YOLOv5n、YOLOv6n和YOLOv7-tiny相比，S-YOLO的mAP50分别提高0.4%、0.7%和0.5%，mAP50-95分别提高1.9%、2.6%和2.4%，APS分别提高了1.3%、3.3%和1.4%。与针对弱小空间目标检测的CS-YOLO相比，S-YOLO的mAP50、mAP50-95和APS分别提高0.2%、1.7%和1.2%，同时参数量减少64.07%，浮点运算量减少6.9×10^9次。与针对小目标检测的LACTA相比，mAP50、mAP50-95和APS分别提高了0.5%、3.2%和1.3%，但S-YOLO参数高于LACTA。虽然S-YOLO的检测速度只有77帧/s，但仍达到高于25帧/s实时性要求。因此，本文S-YOLO模型在保持轻量的同时，具有优越的检测精度和良好的整体性能。

2.4 VisDrone2019数据集实验结果

为进一步验证本文S-YOLO的优越性，在公开

小目标数据集 VisDrone2019[33] 上进行了测试, 该数据集包含 10 个类别, 超过 2.6×10^6 万个常用目标标注框, 具有场景可见性、对象类别和遮挡等重要属性, 提升了数据的可利用性。实验结果如表 5 所示, 表中加粗数值表示各列评价指标的最优值。与基线 YOLOv8n 相比, S-YOLO 在准确率 P、召回率 R、mAP50、mAP50-95、APS、APM 和 APL 指标上分别提高了 2.6%、1.2%、2.0%、1.4%、1.0%、1.2% 和 1.6%, 检测速度略有下降, 但仍可以达到 70 帧/s, 满足实时性需求。与最新发布的 YOLOv10n 相比, S-YOLO 的 mAP50、mAP50-95 和 APS 分别提高 2.3%、1.3% 和 1.1%。与 YOLOv3-tiny、YOLOv5n、YOLOv6n 和 YOLOv7-tiny 相比, S-YOLO 的 mAP50 分别提高 11.2%、2.1%、4.5% 和 1.8%, mAP50-95 分别提高 7.0%、1.7%、2.7% 和 1.6%, APS 分别提高了 5.8%、1.0%、2.2% 和 1.0%。与针对弱小空间目标检测的 CS-YOLO 相比, S-YOLO 的 mAP50、mAP50-95 和 APS 分别提高了 2.2%、1.4% 和 1.0%。与针对小目标检测的 LACTA 相比, S-YOLO 的 mAP50、mAP50-95 和 APS 分别提高了 1.7%、1.2% 和 0.7%。因此, S-YOLO 模型在小目标的检测中仍然表现出优越性能。

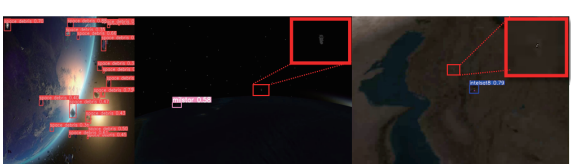
表 5　不同检测算法在 VisDrone2019 数据集上的实验结果

Table 5　Experimental results of different detection algorithms on VisDrone2019 dataset

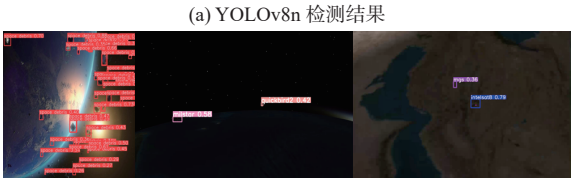
模型	参数量	浮点运算量/10^9次	P/%	R/%	mAP50/%	mAP50-95/%	APS/%	APM/%	APL/%	检测速度/(帧·s^{-1})	模型大小/MB
YOLOv3-tiny*[34]	12.14×10^6	19.1	37.7	23.6	22.9	12.7	3.8	19.7	35.5	**179**	24.4
YOLOv5n*	2.51×10^6	**7.2**	43.1	32.2	32.0	18.0	8.6	26.2	35.5	112	5.3
YOLOv6n*[35]	4.24×10^6	11.9	40.8	30.2	29.6	17.0	7.4	24.9	34.6	129	8.7
YOLOv7-tiny*[36]	8.12×10^6	21.3	42.3	32.5	32.3	18.1	8.6	27.7	35.7	116	16.5
CS-YOLO[22]	7.18×10^6	15.1	42.0	32.2	31.9	18.3	8.6	25.9	35.9	76	14.7
LACTA[37]	**1.20**$\times10^6$	9.6	42.3	33.1	32.4	18.5	8.9	26.6	35.7	128	**2.6**
YOLOv8n[28]	3.20×10^6	8.2	43.2	32.4	32.1	18.3	8.6	27.2	35.2	118	6.3
YOLOv10n[38]	2.71×10^6	8.4	43.3	31.9	31.7	18.4	8.5	27.1	35.2	113	5.8
S-YOLO	2.58×10^6	8.2	**45.8**	**33.6**	**34.1**	**19.7**	**9.6**	**28.4**	**36.8**	70	5.5

2.5　可视化结果分析

为更加直观地验证 S-YOLO 算法对小尺度目标的检测性能, 进行可视化分析。本文从自建 SSAD 测试集和 VisDrone2019 挑战性测试集中选取不同场景图片, 使用 YOLOv8n 和 S-YOLO 模型进行推理实验, 结果如图 7 和图 8 所示。



(a) YOLOv8n 检测结果



(b) S-YOLO 检测结果

图 7　SSAD 测试集检测结果

Fig. 7　Detection results of SSAD dataset

图 7 实验结果表明, S-YOLO 对密集场景、低光照、低对比度下的小目标具有最优的检测精度, 漏检率较低。图 8 实验结果表明, YOLOv8n 在不同场景下存在明显的漏检, 并有一定的误检率, S-YOLO 在密集场景下针对小目标检测的误检率和漏检率得到显著改善。实验结果表明, S-YOLO 在适应不同场景下对小目标的检测具有更高的检测精度, 具有更高的鲁棒性。



(a) YOLOv8n 检测结果　　(b) S-YOLO 检测结果

图 8　VisDrone 挑战性测试集检测结果

Fig. 8　Detection results of VisDrone dataset

3 结 论

1）本文构建了一个空间态势感知数据集SSAD，该数据集具有更大比例的小目标和多场景，为深入研究天基空间目标探测技术提供了一个大规模数据集。

2）本文设计了3个轻量级即插即用模块LMFD、DSPPF和CTFEM。LMFD具有双分支结构，采用不同大小的卷积核扩大模型感受野，提取不同尺度的空间目标特征。DSPPF通过双分支金字塔池化结构聚合目标局部和全局特征，获取更多背景细节信息和上下文信息。CTFEM跨通道交互学习通道之间的相关性，增强小目标弱特征表示，提高对空间小尺度目标的检测能力。

3）实验结果表明，本文S-YOLO在SSAD上的mAP50、mAP50-95和APS分别达到96.3%、85.0%和49.8%，检测速度为77帧/s。在模型参数量减少19.38%情况下，检测准确率仍然超过基准模型YOLOv8n。在VisDrone2019小目标数据集上，S-YOLO相比其他检测算法的检测精度同样具有明显优势。

虽然本文S-YOLO在空间非合作小目标检测任务中表现出色，可实现航天器实时目标检测，但是在硬件部署之前需要进一步优化模型推理速度。

参考文献（References）

[1] Zhang Z P, Deng C W, Deng Z Y. A diverse space target dataset with multidebris and realistic on-orbit environment[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2022, 15: 9102-9114.

[2] Miao L N, Zhou D, Li X. Spatial non-cooperative object detection based on deep learning[C]// Proceedings of the 2021 China Automation Congress. Piscataway: IEEE Press, 2022: 116-121.

[3] Wu A D, Zhang S Y, Shu L Z, et al. Non-cooperative spacecraft relative navigation based on monocular camera in space on-orbit servicing[C]// Proceedings of the 2022 International Conference on Service Robotics. Piscataway: IEEE Press, 2023: 133-138.

[4] Shan M H, Guo J, Gill E. Review and comparison of active space debris capturing and removal methods[J]. Progress in Aerospace Sciences, 2016, 80: 18-32.

[5] Massimi F, Ferrara P, Petrucci R, et al. Deep learning-based space debris detection for space situational awareness: a feasibility study applied to the radar processing[J]. IET Radar, Sonar & Navigation, 2024, 18(4): 635-648.

[6] Girshick R. Fast R-CNN[C]// Proceedings of the 2015 IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2016: 1440-1448.

[7] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]// Computer Vision-ECCV 2016. Cham: Springer International Publishing, 2016: 21-37.

[8] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 779-788.

[9] 王坤, 冯康威. 基于改进YOLOv5-s的交通场景小目标检测算法[J]. 北京航空航天大学学报, 2026, 52(4): 1015-1027.
Wang K, Feng K W. Small target detection algorithm for traffic scenes based on improved YOLOv5-s[J]. Journal of Beijing University of Aeronautics and Astronautics, 2026, 52(4): 1015-1027 (in Chinese).

[10] Wu T, Yang X, Song B, et al. T-SCNN: a two-stage convolutional neural network for space target recognition[C]// Proceedings of the 2019 IEEE International Geoscience and Remote Sensing Symposium. Piscataway: IEEE Press, 2019: 1334-1337.

[11] Xiang Y B, Xi J B, Cong M, et al. Space debris detection with fast grid-based learning[C]// Proceedings of the 2020 IEEE 3rd International Conference of Safe Production and Informatization. Piscataway: IEEE Press, 2021: 205-209.

[12] Wang G J, Lei N W, Liu H M. Improved-YOLOv3 network for object detection in simulated Space Solar Power Systems images[C]// Proceedings of the 2020 39th Chinese Control Conference. Piscataway: IEEE Press, 2020: 7177-7181.

[13] Jia P, Liu Q, Sun Y Y. Detection and classification of astronomical targets with deep neural networks in wide-field small aperture telescopes[J]. The Astronomical Journal, 2020, 159(5): 212.

[14] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[15] Aldahoul N, Karim H A, De Castro A, et al. Localization and classification of space objects using EfficientDet detector for space situational awareness[J]. Scientific Reports, 2022, 12: 21896.

[16] Mahendrakar T, White R T, Wilde M, et al. SpaceYOLO: a human-inspired model for real-time, on-board spacecraft feature detection[C]// Proceedings of the 2023 IEEE Aerospace Conference. Piscataway: IEEE Press, 2023: 1-11.

[17] Tang Q, Li X W, Xie M L, et al. Intelligent space object detection driven by data from space objects[J]. Applied Sciences, 2024, 14(1): 333.

[18] Ai H, Zhang H F, Ren L, et al. Detection and recognition of spatial non-cooperative objects based on improved YOLOX_L[J]. Electronics, 2022, 11(21): 3433.

[19] Guo X J, Chen T, Liu J C, et al. Dim space target detection via convolutional neural network in single optical image[J]. IEEE Access, 2022, 10: 52306-52318.

[20] Jiang Y X, Tang Y J, Ying C C. Finding a needle in a haystack: faint and small space object detection in 16-bit astronomical images using a deep learning-based approach[J]. Electronics, 2023, 12(23): 4820.

[21] Tao J, Cao Y F, Ding M. SDebrisNet: a spatial-temporal saliency network for space debris detection[J]. Applied Sciences, 2023, 13(8): 4955.

[22] Yuan Y M, Bai H Y, Wu P F, et al. An intelligent detection method for small and weak objects in space[J]. Remote Sensing, 2023,

15(12): 3169.

[23] 刘一诺, 张琪, 王蓉, 等. 针对航拍小目标检测的 YOLOv7 改进方法[J]. 北京航空航天大学学报, 2025, 51(7): 2506-2512.
Liu Y N, Zhang Q, Wang R, et al. Improved YOLOv7 method for aerial small target detection in aerial photography[J]. Journal of Beijing University of Aeronautics and Astronautics, 2025, 51(7): 2506-2512 (in Chinese).

[24] 邓天民, 余洋, 陈月田, 等. 基于自适应复合卷积的航拍小目标检测算法[J]. 北京航空航天大学学报, 2024: 1-14.
Deng T M, Yu Y, Cheng Y T, et al. Small object detection algorithm for aerial photography based on adaptive compound convolution[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024: 1-14 (in Chinese).

[25] Howard A G, Zhu M L, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17)[2024-07-03]. https://arxiv.org/abs/1704.04861.

[26] Zhang X Y, Zhou X Y, Lin M X, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 6848-6856.

[27] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2014: 580-587.

[28] Varghese R, M S. YOLOv8: a novel object detection algorithm with enhanced performance and robustness[C]// Proceedings of the 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems. Piscataway: IEEE Press, 2024: 1-6.

[29] Zeng N Y, Li X Y, Wu P S, et al. A novel tensor decomposition-based efficient detector for low-altitude aerial objects with knowledge distillation scheme[J]. IEEE/CAA Journal of Automatica Sinica, 2024, 11(2): 487-501.

[30] Wang Q L, Wu B G, Zhu P F, et al. ECA-net: efficient channel attention for deep convolutional neural networks[C]// Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 11531-11539.

[31] Wang J Q, Chen K, Xu R, et al. CARAFE: content-aware ReAssembly of FEatures[C]// Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 3007-3016.

[32] Liu W Z, Lu H, Fu H T, et al. Learning to upsample by learning to sample[C]// Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2024: 6004-6014.

[33] 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, 2022, 44(11): 7380-7399.

[34] Redmon J, Farhadi A. YOLOv3: an incremental improvement [EB/OL]. (2018-04-08)[2024-07-03]. https://arxiv.org/abs/1804.02767.

[35] Li C Y, Li L L, Jiang H L, et al. YOLOv6: a single-stage object detection framework for industrial applications[EB/OL]. (2022-09-07)[2024-07-03]. https://arxiv.org/abs/2209.02976.

[36] Wang C Y, Bochkovskiy A, Liao H M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]// Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2023: 7464-7475.

[37] Gao J, Zhang J X, Zhang F, et al. LACTA: a lightweight and accurate algorithm for cherry tomato detection in unstructured environments[J]. Expert Systems with Applications, 2024, 238: 122073.

[38] Wang A, Chen H, Liu L H, et al. YOLOv10: real-time end-to-end object detection[EB/OL]. (2024-05-23)[2024-07-03]. https://arxiv.org/abs/2405.14458.

Lightweight multi-scale feature fusion enhancement algorithm for spatial non-cooperative small target detection

LIU Zhenjiang[1], ZHANG Huijuan[2,*], JI Miaoxin[1], LU Tao[1], ZHANG Chi[1], YU Yuanjin[2]

(1. School of Electrical Engineering, Henan University of Technology, Zhengzhou 450001, China;
2. School of Automation, Beijing Institute of Technology, Beijing 100081, China)

Abstract: In deep space environments, the feature of the non-cooperative small targets cannot be fully extracted, so the detection accuracy of uncooperative small targets will decrease. Furthermore, the computational resources are limited for space applications. To solve the above-mentioned problems, a lightweight multi-scale feature fusion, enhancement and context-aware spatial target detection algorithm is proposed. Multi-scale characteristics of spatially non-cooperative targets are extracted using a lightweight multi-feature downsampling module, which may improve the capacity to obtain fine-grained features. A fast dual spatial pyramid pooling module is developed to enhance the extraction of global contextual features, and the locally salient and global features are aggregated to improve the ability of feature generalization. A cross-dimension triple feature enhancement module (CTFEM) is designed, which could model the global context in term of different dimensional information. Therefore, the CTFEM module could further enhance the feature information of small targets and suppress the confused background information. The VisDrone2019 dataset and a self-built spatial situational awareness dataset (SSAD) are used to validate the suggested method in this research. The experimental findings show that the proposed method decreases the model parameters by 19.38% when compared to the benchmark model YOLOv8n. On the SSAD dataset, the mAP50 and mAP50-95 of the proposed method are improved by 0.2% and 1.2%, respectively. On the VisDrone2019 dataset, the mAP50 and mAP50-95 of the proposed method are improved by 2.0% and 1.4%, respectively.

Keywords: space non-cooperative target detection; small target detection; feature fusion; context information; lightweight network

Received: 2024-07-03; **Accepted**: 2024-09-13; **Published Online**: 2024-09-20 10:37

URL: link.cnki.net/urlid/11.2625.V.20240919.1640.008

Foundation items: Postdoctoral Fellowship Program of CPSF (GZC20233408); National Natural Science Foundation of China (62222304,62173039, 62201199); Innovative Funds of Henan University of Technology(2021ZKCJ07)

* **Corresponding author**. E-mail: zhanghjqy@bit.edu.cn