CN 51-1346/O4 ISSN 1003-501X (印刷版) ISSN 2094-4019 (网络版)



## 融合视觉中心机制和并行补丁感知的遥感图像检测算法

梁礼明,陈康泉,王成斌,冯耀,龙鹏威

## 引用本文:

梁礼明,陈康泉,王成斌,等.融合视觉中心机制和并行补丁感知的遥感图像检测算法[J].光电工程,2024,**51**(7): 240099.

Liang L M, Chen K Q, Wang C B, et al. Remote sensing image detection algorithm integrating visual center mechanism and parallel patch perception[J]. *Opto-Electron Eng*, 2024, **51**(7): 240099.

## https://doi.org/10.12086/oee.2024.240099

收稿日期: 2024-05-01; 修改日期: 2024-07-10; 录用日期: 2024-07-10

## 相关论文

## 特征协同与细粒度感知的遥感图像小目标检测

肖振久,张杰浩,林渤翰 光电工程 2024, **51**(6): 240066 doi: 10.12086/oee.2024.240066

## 面向遥感图像检索的级联池化自注意力研究 吴刚, 葛芸, 储珺, 叶发茂

光电工程 2022, 49(12): 220029 doi: 10.12086/oee.2022.220029

结合遥感卫星及深度神经决策树的夜间海雾识别

李涛,金炜,符冉迪,李纲,尹曹谦 光电工程 2022, **49**(9): 220007 doi: 10.12086/oee.2022.220007

## 基于多尺度特征融合的遥感图像小目标检测

马梁,苟于涛,雷涛,靳雷,宋怡萱 光电工程 2022, **49**(4): 210363 doi: 10.12086/oee.2022.210363

## 更多相关论文见光电期刊集群网站



http://cn.oejournal.org/oee





Website



DOI: 10.12086/oee.2024.240099

# 融合视觉中<mark>心</mark>机制和并行补丁感知 的遥感图像检测算法



梁礼明,陈康泉\*,王成斌,冯 耀,龙鹏威 江西理工大学电气工程与自动化学院,江西赣州 341000

摘要:针对遥感图像存在复杂背景干扰、目标多尺度差异和微小目标提取难的问题,本文基于YOLOv7-tiny 模型提 出一种融合视觉中心机制和并行补丁感知的遥感图像检测算法。该算法一是引入显式视觉中心机制,构建像素点间的 长距离依赖关系,丰富图像的整体语义信息,同时提升对目标纹理的提取性能;二是改进并行补丁感知模块,调整特 征提取感受野,以适应不同目标尺度;三是设计多尺度特征融合模块,实现对多层特征的高效融合,提升模型推理速 度。在公共数据集 RSOD 上进行实验,所提算法的准确率、召回率和平均准确率均值相较 YOLOv7-tiny 分别提升 1.5%、2.4% 和 2.4%,此外在 NWPU VHR-10 和 DOTA 数据集上进行泛化性验证,结果表明本文算法具备较强的泛 化性能。通过与不同算法对比分析,进一步体现本文算法性能的优越性。 关键词:遥感图像;目标检测;YOLOv7-tiny;显式视觉中心机制;并行补丁感知

梁礼明,陈康泉,王成斌,等.融合视觉中心机制和并行补丁感知的遥感图像检测算法 [J]. 光电工程,2024, **51**(7): 240099 Liang L M, Chen K Q, Wang C B, et al. Remote sensing image detection algorithm integrating visual center mechanism and parallel patch perception[J]. *Opto-Electron Eng*, 2024, **51**(7): 240099

## Remote sensing image detection algorithm integrating visual center mechanism and parallel patch perception

## Liang Liming, Chen Kangquan<sup>\*</sup>, Wang Chengbin, Feng Yao, Long Pengwei

School of Electrical Engineering and Automation, Jiangxi University of Science and Technology, Ganzhou, Jiangxi 341000, China

**Abstract:** To address the challenges of complex background interference, multi-scale differences in targets, and the difficulty in extracting small targets from remote sensing images, this paper proposes a remote sensing image detection algorithm based on the YOLOv7-tiny model that integrates the visual center mechanism and parallel patch perception. Firstly, the algorithm introduces an explicit visual center mechanism to establish long-distance dependencies between pixels, enriching the overall semantic information of the image and improving the extraction performance of target textures. Secondly, it improves the parallel patch perception module by adjusting the feature extraction receptive fields to adapt to different target scales. Thirdly, a multi-scale feature fusion module is designed

基金项目: 国家自然科学基金资助项目 (51365017, 61463018); 江西省自然科学基金资助项目 (20192BAB205084); 江西省教育厅科 学技术研究青年项目 (GJJ2200848) \*通信作者: 陈康泉, 1136344152@gq.com。

版权所有©2024 中国科学院光电技术研究所

收稿日期: 2024-05-01; 修回日期: 2024-07-10; 录用日期: 2024-07-10

#### 梁礼明,等.光电工程,2024,51(7):240099

to efficiently fuse multi-layer features, thereby improving the model's inference speed. Experimental results on the RSOD dataset show that the proposed algorithm achieves improvements over YOLOv7-tiny in terms of precision, recall, and mean average precision by 1.5%, 2.4%, and 2.4%, respectively. Additionally, validation on the NWPU VHR-10 and DOTA datasets confirms the strong generalization performance of the proposed algorithm. Comparative analysis with other algorithms further demonstrates the superior performance of the proposed approach.

Keywords: remote sensing images; object detection; YOLOv7-tiny; explicit visual center mechanism; parallel patch perception

## 1 引 言

光学遥感图像目标检测旨在准确定位和分类感兴 趣的目标,适用于情报侦察、目标监视和灾害救援等 领域[1-2]。在不同领域的应用场景中,对于多尺度目标 和微小目标的定义各有不同。本文基于绝对尺度,将 小于 32 pixel×32 pixel 的目标定义为微小目标,将大 于等于 32 pixel×32 pixel 且小于等于 96 pixel×96 pixel 的目标定义为中目标,将大于 96 pixel×96 pixel 的目 标定义为大目标,将具有不同大小的目标统称为多尺 度目标。由于遥感图像存在复杂背景、密集分布和尺 度多变等挑战,亟需设计高效准确的检测算法<sup>[3-4]</sup>。在 传统的遥感图像目标检测算法中,一种是基于模板匹 配,计算输入图像中特定区域的特征向量与模板特征 向量的匹配度;另一种是基于人工先验规则,获取候 选区域以建立目标的特征表示。这两种检测算法理论 完备、检测精度较高,但难以对多样化任务场景及目 标进行充分特征表示,且滑动窗口式效率低下,导致 其目标检测的精度和速度难以满足实际需求。随着深 度学习的发展,光学遥感图像目标检测取得重大进展。 基于深度学习的遥感图像目标检测算法可分为基于候 选区域的算法和回归分析的算法。前者又称双阶段算 法,第一阶段用于生成可能包含目标的候选区域,第 二阶段对候选区域进行分类及边界框回归。其代表算 法有 R-CNN<sup>[5]</sup>、Fast R-CNN<sup>[6]</sup>、Faster R-CNN<sup>[7]</sup>和 Mask R-CNN<sup>[8]</sup>等,该类算法检测精度较高但计算量 大导致速度难以满足实时性需求。而后者又称单阶段 算法,从输入图像的多个位置直接回归分析出目标的 边界框和类别。其典型算法有 SSD<sup>[9]</sup> 和 YOLO<sup>[10-12]</sup> 系 列等,更好地兼顾检测精度和速度。

早期的遥感目标检测器通常基于卷积神经网络 (convolutional neural network, CNN)<sup>[13]</sup>,其性能受到卷 积操作固有局部性的严重限制,只能定位局部最具有 代表性的目标区域,且计算复杂度较高。Gao 等<sup>[14]</sup>提 出用于在卷积神经网络中进行高效的感受野 (receptive field, RF) 搜索算法 RF-Next, 该算法通过全 局到局部的搜索方案,寻找更好的感受野组合,以提 高目标检测的性能,但搜索更好的感受野组合需要更 多的计算资源和时间,尤其是当感受野范围较大时。 针对复杂背景干扰和小物体特征提取难的问题, 文 献 [15] 构建一种针对遥感目标检测的全局到局部尺度 感知检测网络 GLSANet, 引入全局语义学习交互模 块来挖掘和强化深度特征图中的高级语义学习,缓解 前景对象上复杂背景的障碍;引入局部注意力金字塔 抑制较浅特征图中的背景和噪声,突出小物体的特征 表示,但该网络对大尺度目标的表现效果欠佳。 Zhang 等<sup>[16]</sup> 在 YOLOv5s 算法<sup>[17]</sup> 基础上构建一种快速 的目标检测算法 SuperYOLO, 该算法通过简单的辅助 超分辨率分支来学习高分辨率目标检测,取得较好的 检测效果,但在处理复杂背景的遥感图像时仍存在微 小目标特征信息丢失的问题。因此,为解决遥感图像 目标检测当前存在的技术挑战,本文以 YOLOv7-tiny 为基线模型,提出一种融合视觉中心机制和并行补丁 感知的遥感图像检测算法,提升遥感图像目标检测的 准确性,主要工作如下:

1)利用显式视觉中心机制模块 (explicit visual center, EVC)<sup>[18]</sup>,建立全局长距离依赖关系,以捕获 上下文信息的中心特征,同时聚合层内局部区域信息,以捕获局部具有代表性的特征表示,进一步提升对目 标纹理的提取性能。

2) 优化并行补丁感知模块 (parallelized patchaware module, PPA), 动态调整特征提取感受野, 捕获 不同尺度的特征信息, 有效地适应和处理广泛的背景。

3) 设计多尺度特征融合模块 (multi-scale feature fusion module, MFFM),高效融合来自不同层次的特征,全面捕捉目标的多样化表征,提升检测

速度。

## 2 本文算法

## 2.1 YOLOv7-tiny 算法

本研究基于 YOLOv7-tiny 算法进行改进。该算法 是一种轻量级的目标检测算法,主要由三个核心组件 构成: Backbone 网络、Neck 网络和 Head 层。其中, Backbone 网络包括卷积计算单元、高效层聚合网络 (efficient layer aggregation networks, ELAN) 和最大池 化层,用于从输入图像中提取特征。Neck 网络用于 融合 Backbone 提取的多尺度特征。Head 层则负责根 据融合特征层进行目标检测的分类和回归预测。在此 基础上,本文对 YOLOv7-tiny 算法进行改进,以提高 其在遥感图像目标检测中的性能。

## 2.2 算法设计

为了增强网络全局信息感知能力以及聚合丰富的 细节信息,针对遥感图像复杂背景、目标多尺度和微 小目标的特点,本文基于 YOLOv7-tiny 模型提出一种 融合视觉中心机制和并行补丁感知的遥感图像检测算 法,其整体结构如图 1 所示。首先引入显式视觉中心 机制,通过轻量级多层感知机 (lightweight multilayer perceptron, LMLP) 模块实现全局长距离遥感建模,着 眼于全局的关键特征,同时使用可学习的视觉中心机 制 (learnable visual center, LVC) 模块聚合输入图像的 局部关键区域,捕获局部极具区分性的特征表示,提 升对目标纹理的提取性能;其次改进并行补丁感知模 块,动态调整特征提取感受野,丰富上下文特征语义 信息,获取多尺度特征信息;最后设计多尺度特征融 合模块,实现对多层特征的高效融合,不仅有效降低 计算时间,而且满足高精度的检测需求。

## 2.3 显式视觉中心机制模块

标准 CNN 骨干网络在目标检测方面取得初步成 功,但受限于有限的感受野,仅能定位局部特征区域。 为解决 CNN 中有限局部特征的问题,引入显式视觉 中心机制模块以建立图像全局长距离依赖关系,同时 关注层内局部区域特征,其结构如图 2 所示,EVC 模块主要包括 LMLP 块和 LVC 块两条分支。对于给 定输入特征X<sub>in</sub>,该模块首先使用 Stem 模块提取初始 特征并划分,然后将划分后的特征图X<sup>i</sup><sub>in</sub>和X<sup>2</sup><sub>in</sub>输入到 LMLP 块和 LVC 块这两条分支进行特征优化,最后 利用拼接操作和1×1的标准卷积融合两个分支的结果, 其计算过程分别如下:

$$X_{\rm in}^1, X_{\rm in}^2 = Split(X_{\rm in}), \qquad (1)$$

$$X_{\rm o} = C_{1\times 1}(Cat(LLMP(X_{\rm in}^1), LVC(X_{\rm in}^2))), \qquad (2)$$

其中: *LMLP*(·)表示轻量级 MLP, *LVC*(·)表示可学习 的视觉中心机制, *Cat*(·)表示拼接操作, *C*<sub>1×1</sub>(·)表示 1×1卷积操作, *X*<sub>0</sub>表示 EVC 模块的输出。

LMLP 块主要强调空间中每个像素点之间的远程



图 1 融合视觉中心机制和并行补丁感知的遥感图像检测模型

Fig. 1 Remote sensing image detection model integrating visual center mechanism and parallel patch perception

## 梁礼明,等.光电工程,2024,51(7):240099

#### https://doi.org/10.12086/oee.2024.240099





依赖关系,填补局部细节信息生成分支的语义空白。 给定输入特征*X*<sup>in</sup>,首先对其进行组归一化操作消除 尺度差异,并使用深度可分离卷积操作实现降维;其 次采用通道缩放增强特征的多样性;再次经过正则化 减少过拟合风险;最后将正则化输出特征与给定输入 特征*X*<sup>in</sup>进行逐元素加法操作,以增强特征交互。上 述输出特征经过组归一化、通道 MLP、通道缩放和 正则化操作,并与该特征进行逐元素加法操作。其过 程分别表述为

$$X_{\rm in}^{1'} = DP(SC(DC(GN(X_{\rm in}^{1})))) \oplus X_{\rm in}^{1}, \qquad (3)$$

 $LMLP_{o} = DP(SC(CMLP(GN(X_{in}^{i'})))) \oplus X_{in}^{i'}, \quad (4)$ 

其中: GN(·)表示组归一化, DC(·)表示1×1深度卷 积操作, SC(·)表示通道缩放, DP(·)表示正则化, ⊕ 表示逐元素加法, CMLP(·)表示通道 MLP 操作, LMLP。表示 LMLP 块的输出。

LVC 块分支主要是在训练过程中通过反向传播 算法对参数进行优化,从给定特征图中挖掘感兴趣的 局部细节特征。首先对其经过一个1×1、一个3×3和 一个1×1卷积操作调整特征维度,并进行批归一化和 激活函数结合操作以消除梯度消失问题;然后采用编 码映射捕获数据之间的关系;再后经过批归一化、正 则化和均值池化组合操作,以提升特征表达能力,并 通过全连接层和1×1卷积整合特征信息;最后将上述 输出特征与给定输入特征进行逐元素乘法和加法操作。 其具体表达式分别为

$$X_{\rm in}^{2'} = CBR(C_{1\times 1}(C_{3\times 3}(C_{1\times 1}(X_{\rm in}^2)))), \qquad (5)$$

$$e_{k} = \sum_{i=1}^{n} \frac{\mathrm{e}^{-s_{k} \left\| |x_{i}^{z'} - b_{k} \right\|^{2}}}{\sum^{K} \mathrm{e}^{-s_{k} \left\| |x_{i}^{z'} - b_{k} \right\|^{2}}} (x_{i}^{2'} - b_{k}), \qquad (6)$$

$$e = C_{1 \times 1} \left( FC\left(\sum_{k=1}^{K} \phi(e_k)\right) \right), \tag{7}$$

$$LVC_{o} = X_{in}^{2} \oplus (X_{in}^{2} \otimes e), \qquad (8)$$

其中:  $C_{3\times3}(\cdot)$ 表示3×3卷积操作,  $CBR(\cdot)$ 表示批归一 化和激活函数结合操作,  $X_{in}^{2r}$ 表示第 i 个像素点,  $b_k$ 表 示第 k 个码字,  $s_k$ 表示第 k 个缩放因子,  $x_i^2 - b_k$ 表示 像素位置对应的码字信息, K表示可视化中心的总数,  $FC(\cdot)$ 表示全连接层操作,  $\otimes$ 表示逐元素乘法操作,  $LVC_o$ 表示 LVC 模块的输出。

## 2.4 并行补丁感知模块

为解决微小目标定位难和识别难的挑战,引入并 行补丁感知模块。该模块由并行多分支特征提取模 块 (parallel multi-branch feature extraction module, PMFE)<sup>[19]</sup>和大型选择核模块 (large selective kernel module, LSK)<sup>[20]</sup>组成。PPA 模块使用并行多分支特征 提取策略,捕获不同尺度及层次特征信息,从而提高 小目标检测的准确性;采用大型选择核模块自适应增 强小目标的特征表示,保留下采样后的关键信息。

## 2.4.1 并行多分支特征提取模块

PMFE 模块包含全局、局部和串行卷积分支,每 个分支负责不同尺度的特征提取,结构如图 3 所示。 该模块首先通过逐点卷积将给定输入特征张量  $F \in R^{H' \times W \times C}$ 调整为 $F' \in R^{H' \times W' \times C'}$ ;然后通过3×3卷积 计算得到串行卷积特征张量 $F_{conv} \in R^{H \times W' \times C'}$ ;其次通 过不同大小的补丁感知子块实现全局和局部特征信息 的提取与交互,分别得到全局特征张量 $F_{global} \in$  $R^{H' \times W' \times C'}$ 和局部特征张量 $F_{local} \in R^{H' \times W' \times C'}$ ;最后将并行 多分支结果叠加得到输出特征张量 $\tilde{F} \in R^{H' \times W' \times C'}$ 。其 计算式为

$$\tilde{F} = F_{\text{global}} + F_{\text{local}} + F_{\text{conv}} , \qquad (9)$$

其中:补丁感知子块的全局和局部特征信息提取与交 互通过空间维度聚合与移位非重叠补丁实现。该过程 首先展开并重塑,将F'划分为空间连续的补丁  $p \times p, \frac{H'}{r}, \frac{W'}{r}, C';然后经过通道方向平均化处理,得$ 

到 $p \times p, \frac{H'}{p}, \frac{W'}{p};$ 其次使用前馈神经网络进行线性计

算,应用激活函数 Softmax 获取特征的概率分布,并 调整其权重, Softmax 激活函数计算式为

$$Softmax(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}},$$
 (10)

其中: *Softmax*(*z<sub>i</sub>*)表示第*i*个元素的概率值, e表示自 然对数的底, *K*表示向量的长度; 最后使用特征选择 从标记和通道中捕获重要特征。

## 2.4.2 大型选择核模块

LSK 模块的主要任务是动态调整特征提取感受野, 处理多样化的背景,提高小目标的检测性能。在模块 中添加跳跃连接,进一步丰富特征语义信息,其结构 如图 4 所示。对于给定输入特征 *X*,首先利用两个大 核选择块 (LK selection, LK)<sup>[21]</sup>扩大感受野,然后将不 同感受野的特征拼接以扩大感受野覆盖范围,最后经 过平均和最大池化及卷积处理丰富特征多样性。其计 算式分别为

 $\widetilde{U}$ 

$$U_0 = X, \ \widetilde{U}_1 = F_1^{1 \times 1}(U_0), \ \widetilde{U}_2 = F_2^{1 \times 1}(U_0), \ (11)$$

$$= [\widetilde{U}_1; \widetilde{U}_2], \qquad (12)$$





## 图 3 并行多分支特征提取模块

Fig. 3 Parallel multi-branch feature extraction module





$$SA_{\text{avg}} = P_{\text{avg}}(\widetilde{U}), SA_{\text{max}} = P_{\text{max}}(\widetilde{U}),$$
 (13)

$$\widehat{SA} = F^{1 \times 1}([SA_{\text{avg}}; SA_{\text{max}}]), \qquad (14)$$

$$\widetilde{SA}_1 = \sigma(\widehat{SA}_1), \quad \widetilde{SA}_2 = \sigma(\widehat{SA}_2), \quad (15)$$

其中: $F_1^{1\times 1}(\cdot)$ 、 $F_2^{1\times 1}(\cdot)$ 和 $F^{1\times 1}(\cdot)$ 均表示1×1卷积操作,  $P_{avg}(\cdot)$ 和 $P_{max}(\cdot)$ 分别表示平均和最大池化, $\sigma(\cdot)$ 表示 sigmoid函数。

将上述输出特征图  $SA_1$ 和  $SA_2$ 分别与  $\tilde{U}_1$ 和  $\tilde{U}_2$ 进行 空间逐元素乘法,然后通过卷积处理,最后将输出特 征与给定输入特征进行逐元素乘法和加法操作。其具 体过程分别为

$$S = F^{1 \times 1} \left( \sum_{i=1}^{2} \left( \widetilde{SA}_{i} \otimes \widetilde{U}_{i} \right) \right), \tag{16}$$

$$Y = X \oplus (X \otimes S) . \tag{17}$$

## 2.5 多尺度特征融合模块

设计高效、高质量的网络架构是深度学习领域<sup>[22]</sup>中一项重要的研究课题。本文在 ELAN 模块的基础上,设计一种多尺度特征融合模块,结构如图 5 所示。



图 5 多尺度特征融合模块 Fig. 5 Multi-scale feature fusion module

MFFM 模块实现对多层特征的高效融合,加速梯 度传播和模型训练。该模块首先利用并行多分支卷积 捕获不同尺度的特征信息;然后通过Concat融合多层 特征信息;最后进行卷积处理以增强网络特征学习能 力。其具体过程可分别表述为

$$C_{i\times i}(P_1) = Cat \begin{bmatrix} C_{1\times 1}(P_1), C_{1\times 1}(P_1), \\ C_{3\times 3}(C_{1\times 1}(P_1)) \end{bmatrix}, \quad (18)$$

$$MFFM_{\rm O} = C_{1\times 1}(C_{i\times i}(P_{\rm I})), \qquad (19)$$

其中: *P*<sub>1</sub>表示多尺度特征融合模块的输入; *Cat*[·]表示多层特征融合操作; *MFFM*<sub>0</sub>表示多尺度特征融合模块的输出。

## 3 实 验

### 3.1 实验环境及参数设置

本文实验基于 64 位 Windows 11 操作系统,使用 Python 编程语言在 Pytorch 框架中搭建,主要硬件

配置如下:处理器为 AMD Ryzen 9 7945HX with Radeon Graphics、显卡为 NVIDIA GeForce RTX 4060、显存 8 GB。实验参数设置如表 1 所示。

表 1 参数设置 Table 1 Parameter setting

参数	参数值
输入图像分辨率	640×640
初始学习率	0.01
动量参数	0.937
权重衰减系数	0.0005
训练轮次	300
批量大小	16

#### 3.2 实验数据集

本文实验使用 RSOD、NWPU VHR-10 和 DOTA 数据集。其中 RSOD 数据集<sup>[23]</sup>由武汉大学标注,包 含 976 张遥感图像,涵盖飞机 (aircraft)、油桶 (oil tank)、操场 (playground) 和立交桥 (overpass) 四种不 同类型目标。随机抽取 781 张图像作为训练集,195 张图像用作验证集。

NWPU VHR-10 数据集<sup>[24]</sup>由西北工业大学标注, 包含 800 张遥感图像,其中含目标图像共 650 张,仅 含背景图像共 150 张。该数据集共 10 个目标类别, 包括: 飞机 (airplane)、船只 (ship)、油罐 (storage)、 棒球场 (baseball diamond)、网球场 (tennis court)、篮 球场 (basketball court)、田径场 (ground track field)、 港口 (harbor)、桥梁 (bridge) 和车辆 (vehicle)。随机抽 取 640 张图像作为训练集, 160 张图像用作验证集。

DOTA 数据集<sup>[25]</sup> 由清华大学标注,选取 3717 张 遥感图像,其中包含飞机 (plane)、棒球场 (baseballdiamond)、桥梁 (bridge)、田径场 (ground-track-field)、 小车 (small-vehicle)、大车 (large-vehicle)、船只 (ship)、 网球场 (tennis-court)、篮球场 (basketball-court)、油 罐 (storage-tank)、足球场 (soccer-ball-field)、环岛 (roundabout)、海港 (harbor)、游泳池 (swimming-pool) 和直升机 (helicopter)。随机抽取 2974 张图像作为训 练集,743 张图像用作验证集。

#### 3.3 实验评价指标

目标检测中使用准确率 (precision, P)、召回率 (recall, R)、平均准确率 (average precision, AP)、参数 量 (params, Par) 和平均准确率均值 (mean average precision, mAP) 等指标来评价检测结果,其计算式分

别为

$$P = \frac{TP}{TP + FP} , \qquad (20)$$

$$R = \frac{TP}{TP + FN} , \qquad (21)$$

$$AP = \int_{0}^{1} P(R) \mathrm{d}R , \qquad (22)$$

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i , \qquad (23)$$

其中: TP 表示将真实正样本预测为正样本的数量, TN 表示将真实负样本预测为负样本的数量, FP 表示 将真实负样本预测为正样本的数量, FN 表示将真实 正样本预测为负样本的数量。此外,使用帧速度 (frames per second, FPS)评估检测的速度,即每秒处 理的图片数量。

## 3.4 LSK 模块实验

## 3.4.1 不同注意力对比实验

本文将 PPA 模块中原有的 CBAM 注意力模块替 换为 LSK 模块,并结合 PMFE 模块重新构建了 PPA 模块。为验证 LSK 模块的优越性,将其置于并行补 丁感知模块中,在 RSOD 数据集上与 CBAM、SE、 EMA、CA 注意力进行对比,实验结果如表 2 所示, 其中加粗字体表示该项指标的最优值。

表 2 不同注意力对比实验 Table 2 Experiments on contrasting attentional differences

_				
	注意力	参数量/M	FPS	mAP@0.5/%
	CBAM	11.3	92	96.1
	SE	11.1	90	93.4
	CA	11.1	92	96.2
	EMA	11.1	93	95.6
	LSK	11.5	92	96.5

分析表 2 可知,综合考虑三个指标,在增加极少 参数量的情况下,LSK 的 mAP@0.5 值最优,达到 96.5%,比原注意力 CBAM 提升 0.4%,同时,其 FPS 值仅次于 EMA,与 CBAM 和 CA 平齐,高于 SE。 实验结果表明,LSK 模块的整体性能优秀。

## 3.4.2 不同大小感受野的有效性验证

为获取优异的检测性能,本文通过调整 LSK 模 块中深度可分离卷积的大小扩展有效感受野,以提升 多尺度目标识别精度。将大内核分解为两个深度可分 离卷积,在 RSOD 数据集上进行实验,其结果如表 3

所示,	其中k <sub>i</sub> 表示	第i个卷积核的大小,	$d_i$ 表示第 $i$ 个
卷积核	(的扩张率,	加粗字体表示该项指标	际的最优值。

表3	大内核分解为两个深度可分离卷积的有效性
Table 3	Effectiveness of decomposing a large kernel into two
	sequences of denth-wise senarable kernels

$(k_1, d_1)$	$(k_2, d_2)$	RF	FPS	mAP@0.5/%	
(3, 1)	(5, 2)	11	104	92.0	
(5, 1)	(7, 3)	23	105	95.0	
(7, 1)	(9, 4)	39	88	94.7	

由表 3 分析可知,过小或过大的感受野会限制模型的性能,大小为 23 的感受野同时具有良好的检测 速度和检测精度。因此,本文算法采用感受野大小 为 23 的 LSK 模块做进一步改进。

#### 3.5 MFFM 模块实验

本文在 ELAN 模块的基础上,去掉一个卷积层, 构成 MFFM 模块。为验证 MFFM 模块的有效性和合 理性,在 RSOD 数据集上与 ELAN 模块进行对比, 实验结果如表 4 所示,其中加粗字体表示该项指标的 最优值。

表 4 MFFM 与 ELAN 对比实验 Table 4 Comparison of experiments between MFFM and ELAN

模块	参数量/M	FPS	mAP@0.5/%
ELAN	6.0	88	94.6
MFFM	4.8	126	94.2

分析表 4 可知,与 ELAN 相比,MFFM 的 mAP 指标仅减少 0.4%,但其 Par 值降低 1.2 M,同时 FPS 值提高 38 f/s,高达 126 f/s。综合考虑三个指标, MFFM 在牺牲极少量精度的情况下,换取更低的参数 量和更快的检测速度。

#### 3.6 消融实验

为证明本文模型的优越性,使用 RSOD 数据集进 行消融实验,其结果如表 5 所示,其中加粗字体表示 此项指标的最优值。M1 表示基线模型 YOLOv7-tiny; M2 表示在 M1 基础上引入 EVC 模块; M3 表示在 M2 基础上添加 PPA 模块; M4 表示设计 MFFM 模块 替换 M3 模型中 ELAN 模块,即本文模型。

分析表 5 可知,基线模型 YOLOv7-tiny 在遥感图像目标检测方面表现出色,平均准确率均值达到94.6%。其中飞机、油桶和操场的平均准确率较高,但立交桥的平均准确率较低,仅 85.0%。此外,在模型结构固定的情况下,其准确率和召回率仍有提升空

Table 5   Ablation experimental data							
齿刑	准确室 <b>P</b> IO	刀同变 <b>尸</b> 间。		平均准	平均准确率均值		
侠堂	作出9月4平 <b>日</b> 7%	台凹伞 <b>⊼</b> /%	飞机	油桶	立交桥	操场	mAP@0.5/%
M1	90.3	93.1	97.9	97.8	85.0	97.7	94.6
M2	92.6	91.2	97.7	98.5	88.5	98.4	95.8
M3	92.0	95.2	97.8	98.8	91.4	99.5	96.9
M4	91.8	95.5	97.7	98.6	93.0	98.8	97.0

表5 消融实验数据

间。因此,本文在初始模型的基础上进行改进。通过 在初始模型 M1 的 Neck 网络中引入 EVC 模块, 整体 指标有所提升,其中平均准确率均值提升1.2%,准 确率达到最优值,说明 EVC 模块不仅能够建立像素 点间的长距离依赖关系,而且有效关注局部区域特征, 对遥感图像目标检测具有较强的多尺度特征捕捉能力。 在 M2 的 Backbone 网络底下添加 PPA 模块,平均准 确率均值提升1.1%,各目标类别的平均准确率均有 所提升,其中油桶和操场的平均准确率达到各项指标 的最优值,说明 PPA 模块可以充分提取和丰富特征 语义信息。最后设计 MFFM 模块替换 M3 中 ELAN 模块, 召回率和立交桥的平均准确率达到各项指标的 最优值,其中立交桥的平均准确率相较于初始模型提 升 8.0%, 说明 MFFM 模块能够充分融合多尺度特征 信息,有效检测复杂背景下的立交桥遥感目标。虽然 卷积层的去除,导致前一层信息的丢失,可能会导致 模型的训练不足,进而造成飞机、油桶和操场的平均 准确率略微降低,但是可以简化网络结构,减少网络 的层数,以减少模型的参数量和计算复杂度,从而减 少模型过拟合的风险,提高模型在新数据上的表现稳 定性,使得模型取得更好的泛化能力。消融实验结果 表明,本文所提模型的合理性和有效性。

## 3.7 多种模型检测结果与分析

## 3.7.1 定量分析

为了验证本文算法的优越性,将Faster R-CNN、 SSD、YOLOv3-tiny、YOLOv4-tiny、YOLOv5s、YOLOv5m、 YOLOv7-tiny、YOLOv8s、YOLOv8m 和本文算法置 于同一实验环境进行对比实验,其实验结果如表6所 示,其中加粗字体表示该项指标的最优值。

分析表 6 可知,本文模型相较于 Faster R-CNN、 SSD、YOLOv3-tiny、YOLOv5m 和 YOLOv8m 算法, 在参数量、FPS 和 mAP@0.5 这三个指标上表现优异, 其中 mAP@0.5 达到此项指标的最优值,FPS 值仅次 于 YOLOv3-tiny,高达 85 帧,满足实时性需求。与 YOLOv5s 和 YOLOv8s 算法相比,在参数量略增的情 况下,本文算法的 mAP@0.5 分别提高 1.5% 和 2.7%。 与 YOLOv4-tiny 和 YOLOv7-tiny 算法比较,本文算 法的参数量表现略微逊色,但 FPS 相较 YOLOv7tiny 基本保持不变,相较 YOLOv4-tiny 大幅增长, mAP@0.5 分别提升 14.6% 和 2.4%。

Table 6         Comparison of detection data from different algorithms							
構刊	模型    参数量/M	500		平均准	平均准确率均值		
侠堂		FPS	飞机	油桶	立交桥	操场	mAP@0.5%
Faster R-CNN	72.0	10	71.0	98.0	85.0	100.0	88.5
SSD	24.4	43	79.0	98.0	73.0	100.0	87.5
YOLOv3-tiny	12.1	104	94.2	96.4	76.9	98.5	91.5
YOLOv4-tiny	6.1	50	70.7	97.3	61.7	99.1	82.4
YOLOv5s	9.1	90	97.4	97.8	87.4	99.3	95.5
YOLOv5m	25.0	56	97.0	96.8	89.4	99.2	95.6
YOLOv7-tiny	6.0	88	97.9	97.8	85.0	97.7	94.6
YOLOv8s	11.1	97	97.6	97.2	82.8	99.4	94.3
YOLOv8m	25.8	53	97.2	98.1	84.3	99.5	94.8
ours	11.5	85	97.7	98.6	93.0	98.8	97.0

表 6 不同算法检测数据对	比
---------------	---

### 梁礼明, 等. 光电工程, 2024, 51(7): 240099

对比不同类别的平均准确率可知,本文模型在包含多尺度和微小目标的飞机类别检测中表现出色, AP 高达 97.7%,仅次于 YOLOv7-tiny;在油桶类别 检测中,平均准确率为 98.6%,达到该项最优指标; 在复杂背景下的立交桥类别检测中表现卓越,平均准 确率高达 93.0%,远优于其他算法。

## 3.7.2 定性分析

为直观显示本文模型的检测性能,将本文模型同 其他目标检测算法在遥感图像上进行预测,可视化结 果分别展示原图、Faster R-CNN、SSD、YOLOv3tiny、YOLOv5s、YOLOv5m、YOLOv7-tiny、YOLOv8s、 YOLOv8m 和本文算法的检测效果,如图 6 所示。其 中图 6(a, d) 表示微小目标遥感图像;图 6(b, e) 表示 背景复杂遥感图像;图 6(c, f) 表示多尺度目标遥感图 像。

观察图 6(a, d) 可知,在微小目标的检测结果中, SSD 算法存在重叠框问题,即对同一物体进行重复预 测; YOLOv3-tiny 算法存在目标漏检的情况; Faster R-CNN、SSD、YOLOv5s、YOLOv5m、YOLOv7tiny、YOLOv8s 和 YOLOv8m 算法均出现非目标误检 的现象,其中 Faster R-CNN、YOLOv5s 和 YOLOv8s 算法误检率较高;而本文算法表现出众,有效实现对 微小目标的准确预测。

在图 6(b, e) 复杂背景目标的检测结果中, Faster R-CNN、 SSD、 YOLOv3-tiny、 YOLOv5s、 YOLOv5m、YOLOv7-tiny、YOLOv8s 和 YOLOv8m 算法表现欠佳,均出现立交桥目标类别漏检的现象; Faster R-CNN 算法存在非目标误检的状况; SSD 算法



Original image Faster R-CNN SSD YOLOv3-tiny YOLOv5s





Fig. 6 Remote sensing target detection results of different algorithms

存在重叠框问题;而本文算法具有较强的背景噪声抗 干扰能力,有效识别复杂背景下的立交桥目标。

在图 6(c, f) 多尺度目标检测结果中,针对图像中 较大尺度和较小尺度的物体,Faster R-CNN、SSD、 YOLOv3-tiny、 YOLOv5s、 YOLOv5m、 YOLOv8s 和 YOLOv8m 算法表现不佳,均出现不同程度的目标 漏检情况;针对图像中较小尺度的物体,SSD 和 YOLOv7-tiny 算法均存在无关物体误检的现象;而本 文算法有效地检测出图像中包含的全部目标。

综上所述,本文算法在面对遥感图像中微小目标、 复杂背景和多尺度目标时,取得优秀的检测成绩,尤 其是针对复杂背景下的立交桥目标类别检测。

### 3.8 泛化性验证

本文采用涵盖更多不同类型和场景的 NWPU VHR-10 和 DOTA 数据集对所提算法进行泛化性验 证,NWPU VHR-10 和 DOTA 数据集实验结果分别 如表 7 和表 8 所示,其中加粗字体表示相应指标的 最优值。

分析表 7 和表 8 可知,与基线模型 YOLOv7-tiny 相比,尽管本文算法的参数量略有增加,但 FPS 基本 保持不变,满足实时性需求。同时,本文算法在 NWPU VHR-10 和 DOTA 数据集上的 mAP@0.5 指标 分别提升 3.0% 和 1.3%,显示出高精度和高效率的 特点。

实验结果表明,本文算法不仅在 RSOD 数据集中 表现优异,而且在包含更多不同类型和场景的数据集 上也取得出色的成绩,展示了良好的泛化能力。

## 4 结 语

针对遥感图像存在复杂背景干扰、目标多尺度差

异和微小目标提取难问题,本文基于 YOLOv7-tiny 模型提出一种融合视觉中心机制和并行补丁感知的遥感 图像检测算法,以提高遥感图像目标的检测性能。本 文算法在 RSOD、NWPU VHR-10 和 DOTA 数据集上 均具有较高的 mAP@0.5 值,相较于基线算法 YOLOv7-tiny 分别提高 2.4%、3.0% 和 1.3%,表明本文算法对 遥感图像目标检测的有效性和泛化性。在后续研究中,本文将朝着低参数的方向对模型做进一步改进和优化,使其能更有效地应用于遥感图像目标检测。

利益冲突:所有作者声明无利益冲突

## 参考文献

- Ma L, Gou Y T, Lei T, et al. Small object detection based on multi-scale feature fusion using remote sensing images[J]. *Opto-Electron Eng*, 2022, **49**(4): 210363.
   马梁, 苟于涛, 雷涛, 等. 基于多尺度特征融合的遥感图像小目标 检测[J]. 光电工程, 2022, **49**(4): 210363.
- [2] Yuan J H, Zhang N F, Ruan J S, et al. Detection of prohibited items in X-ray images based on modified YOLOX algorithm[J]. *Laser Technol*, 2023, **47**(4): 547–552. 袁金豪, 张南峰, 阮洁珊, 等. 基于改进 YOLOX 算法的 X 射线图 像违禁品检测方法[J]. 激光技术, 2023, **47**(4): 547–552.
- [3] Ming Q, Miao L J, Zhou Z Q, et al. CFC-Net: a critical feature capturing network for arbitrary-oriented object detection in remote-sensing images[J]. *IEEE Trans Geosci Remote Sens*, 2022, 60: 5605814.
- [4] Cong R M, Zhang Y M, Fang L Y, et al. RRNet: relational reasoning network with parallel multiscale attention for salient object detection in optical remote sensing images[J]. *IEEE Trans Geosci Remote Sens*, 2022, **60**: 5613311.
- [5] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, 2014: 580–587. https://doi.org/10.1109/CVPR.2014.81.
- [6] Girshick R. Fast R-CNN[C]//Proceedings of 2015 IEEE International Conference on Computer Vision, Santiago, 2015: 1440–1448. https://doi.org/10.1109/ICCV.2015.169.

表7 NV	VPU VHR-10	数据集上检测结果对比
-------	------------	------------

Table 7	Comparison	of detection	results on	NWPU	VHR-10	dataset
---------	------------	--------------	------------	------	--------	---------

模型	准确率 <b>P</b> /%	召回率 <b>R</b> /%	参数量/M	FPS	mAP@0.5/%
YOLOv7-tiny	88.7	88.4	6.0	83	90.7
Ours	92.5	87.6	11.5	79	93.7

#### 表 8 DOTA 数据集上检测结果对比

Table 8 Comparison of dete	ection results on DOTA dataset
----------------------------	--------------------------------

模型	准确率 <b>P</b> /%	召回率 <b>R</b> /%	参数量/M	FPS	mAP@0.5/%
YOLOv7-tiny	78.2	70.4	6.0	82	74.7
Ours	80.0	71.2	11.5	77	76.0

### 梁礼明, 等. 光电工程, 2024, 51(7): 240099

- [7] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards realtime object detection with region proposal networks[J]. *IEEE Trans Pattern Anal Mach Intell*, 2017, **39**(6): 1137–1149.
- [8] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN[C]// Proceedings of 2017 IEEE International Conference on Computer Vision, Venice, 2017: 2961–2969. https://doi.org/10.1109/ICCV.2017.322.
- [9] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[C]//Proceedings of the 14th European Conference, Amsterdam, 2016: 21–37. https://doi.org/10.1007/978-3-319-46448-0\_2.
- [10] Zhao L Q, Li S Y. Object detection algorithm based on improved YOLOv3[J]. *Electronics*, 2020, 9(3): 537.
- [11] Gai R L, Chen N, Yuan H. A detection algorithm for cherry fruits based on the improved YOLO-v4 model[J]. *Neural Comput Appl*, 2023, **35**(19): 13895–13906.
- [12] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: trainable bagof-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, 2023: 7464–7475. https://doi.org/10.1109/CVPR52729.2023.00721.
- [13] Salehi A W, Khan S, Gupta G, et al. A study of CNN and transfer learning in medical imaging: advantages, challenges, future scope[J]. *Sustainability*, 2023, **15**(7): 5930.
- [14] Gao S H, Li Z Y, Han Q, et al. RF-Next: efficient receptive field search for convolutional neural networks[J]. *IEEE Trans Pattern Anal Mach Intell*, 2023, 45(3): 2984–3002.
- [15] Gao T, Niu Q Q, Zhang J, et al. Global to local: a scale-aware network for remote sensing object detection[J]. *IEEE Trans Geosci Remote Sens*, 2023, 61: 5615614.
- [16] Zhang J Q, Lei J, Xie W Y, et al. SuperYOLO: super resolution assisted object detection in multimodal remote sensing imagery[J]. *IEEE Trans Geosci Remote Sens*, 2023, 61:

#### 作者简介



梁礼明(1967-),男,硕士,教授,硕士生导师, 主要研究方向为机器学习、医学影像和系统建 模等公开发表学术论文百余篇,其中被 SCI、 EI、ISTP 收录论文二十余篇。获得中国发明专 利六项(排名第一)、出版研究生教材一部。

E-mail: lianglm67@163.com

5605415.

- [17] Wang L, Liu X B, Ma J T, et al. Real-time steel surface defect detection with improved multi-scale YOLO-v5[J]. *Processes*, 2023, **11**(5): 1357.
- [18] Quan Y, Zhang D, Zhang L Y, et al. Centralized feature pyramid for object detection[J]. *IEEE Trans Image Process*, 2023, **32**: 4341–4354.
- [19] Xu S B, Zheng S C, Xu W H, et al. HCF-Net: hierarchical context fusion network for infrared small object detection[Z]. arXiv: 2403.10778, 2024. https://arxiv.org/abs/2403.10778.
- [20] Li Y X, Li X, Dai Y M, et al. LSKNet: a foundation lightweight backbone for remote sensing[Z]. arXiv: 2403.11735, 2024. https://arxiv.org/abs/2403.11735.
- [21] Li X, Wang W H, Hu X L, et al. Selective kernel networks[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, 2019: 510–519. https://doi.org/10.1109/CVPR.2019.00060.
- [22] Liang L M, Zhan T, Lei K, et al. Multi-resolution fusion input U-shaped retinal vessel segmentation algorithm[J]. *J Electron Inf Technol*, 2023, **45**(5): 1795–1806.
  梁礼明, 詹涛, 雷坤, 等. 多分辨率融合输入的 U 型视网膜血管分 割算法[J]. 电子与信息学报, 2023, **45**(5): 1795–1806.
- [23] Chen Y X, Lin M W, He Z, et al. Consistency-and dependenceguided knowledge distillation for object detection in remote sensing images[J]. *Expert Syst Appl*, 2023, **229**: 120519.
- [24] Zhao D W, Shao F M, Liu Q, et al. A small object detection method for drone-captured images based on improved YOLOv7[J]. *Remote Sens*, 2024, **16**(6): 1002.
- [25] Xia G S, Bai X, Ding J, et al. DOTA: a large-scale dataset for object detection in aerial images[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, 2018: 3974–3983. https://doi.org/10.1109/CVPR.2018.00418.



【通信作者】陈康泉(1995-),男,硕士研究生, 主要研究方向为机器学习、模式识别与图像 处理。

E-mail: 1136344152@qq.com



## Remote sensing image detection algorithm integrating visual center mechanism and parallel patch perception

Liang Liming, Chen Kangquan<sup>\*</sup>, Wang Chengbin, Feng Yao, Long Pengwei



Remote sensing image detection model integrating visual center mechanism and parallel patch perception

**Overview:** In response to challenges posed by complex background interference, multi-scale variations of targets, and difficulties in extracting small targets in remote sensing images, this paper proposes a novel remote sensing image detection algorithm based on the YOLOv7-tiny model. The algorithm integrates a visual centering mechanism and parallel patch perception to enhance target detection performance. The algorithm introduces three main innovations. Firstly, it introduces an explicit visual centering mechanism that uses a lightweight multi-layer perceptron to establish long-distance dependencies between pixels, focusing on capturing central features of contextual information to enrich the overall semantic information of images, including scene structures and contextual details. Simultaneously, a trainable visual centering mechanism aggregates local area information within layers to capture locally representative feature representations, thereby further improving the extraction performance of target textures. This approach effectively extracts and utilizes the overall semantic information of images, accurately capturing global features of targets to enhance recognition of target textures and shapes during detection. Secondly, the algorithm improves the parallel patch perception module by dynamically adjusting the feature extraction receptive field to adapt to different target scales and capture diverse scale feature information, effectively handling varied backgrounds. In practical applications, targets in remote sensing images often exhibit different scales and complex environmental backgrounds, where traditional methods may struggle to distinguish or ignore these differences. By dynamically adjusting the receptive field, the algorithm flexibly perceives targets of different scales while maintaining high accuracy and low error rates in complex background scenarios. Finally, the algorithm designs a multi-scale feature fusion module to efficiently integrate multilevel and multi-scale feature information, comprehensively capturing diverse representations of targets and further enhancing model inference speed while meeting high-precision detection requirements. This fusion method significantly enhances the algorithm's effectiveness in static image detection tasks. Experimental results on the RSOD dataset demonstrate improvements in accuracy, recall, and mean average precision by 1.5%, 2.4%, and 2.4%, respectively, compared to YOLOv7-tiny. Additionally, generalization validation on the NWPU VHR-10 and DOTA datasets shows commendable results, with average precision mean values increasing by 3.0% and 1.3%, respectively, compared to baseline models. These findings illustrate the algorithm's outstanding performance not only on the RSOD dataset but also on datasets encompassing diverse types and scenes, highlighting its robust generalization capability. Through comparative analysis with different algorithms, the superiority of the proposed algorithm's performance is further underscored.

Liang L M, Chen K Q, Wang C B, et al. Remote sensing image detection algorithm integrating visual center mechanism and parallel patch perception[J]. *Opto-Electron Eng*, 2024, **51**(7): 240099; DOI: 10.12086/oee.2024.240099

\* E-mail: 1136344152@qq.com

Foundation item: Project supported by National Natural Science Foundation of China (51365017, 61463018), Natural Science Foundation of Jiangxi Province (20192BAB205084), Jiangxi Provincial Department of Education Science, and Technology Research Youth Project (GJJ2200848)

School of Electrical Engineering and Automation, Jiangxi University of Science and Technology, Ganzhou, Jiangxi 341000, China