



DOI:10.12404/j.issn.1671-1815.2405587

引用格式: 邝先验, 王星星, 王龙锋, 等. 基于改进 YOLO 的多尺度聚合遥感图像小目标检测算法[J]. 科学技术与工程, 2025, 25(20): 8560-8570.

Kuang Xianyan, Wang Xingxing, Wang Longfeng, et al. Small target detection algorithm for multi-scale aggregate remote sensing images based on improved YOLO[J]. Science Technology and Engineering, 2025, 25(20): 8560-8570.

# 基于改进 YOLO 的多尺度聚合遥感图像 小目标检测算法

邝先验, 王星星\*, 王龙锋, 张祖梁

(江西理工大学电气与自动化学院, 赣州 341000)

**摘要** 针对目前遥感图像小目标检测任务中易出现漏检和误检的问题, 提出一种 SCS-YOLO[SMCA + CSC + SIoU (shape-aware intersection over union loss)-you only look once]的遥感图像小目标检测算法。首先, 针对遥感图像中目标小而聚集的问题, 构建空间多尺度卷积注意力(spatial multi-scale convolutional attention, SMCA), 提升模型对空间和通道信息的特征提取能力; 其次, 针对深层网络传递时小目标语义信息容易丢失的问题, 设计聚合亚像素卷积(concentrated sub-pixel convolution, CSC), 采用多尺度聚合特征提取方法, 增强了网络对语义信息的提取能力; 最后, 将 SIoU 损失函数替代原模型中的 CIoU (complete intersection over union loss) 损失函数, 加快了网络的收敛速度。SCS-YOLO 模型在 RSOD 和 NWPU VHR-10 数据集上, 平均精确率的平均值(mAP)分别达到 97% 和 90.9%, 相较于原模型分别提升了 2.2% 和 2.7%, 可见该方法在遥感图像小目标检测任务中的有效性。

**关键词** 遥感图像; SCS-YOLO; 小目标; 注意力; 聚合亚像素卷积; SIoU

**中图分类号** TP183; **文献标志码** A

## Small Target Detection Algorithm for Multi-scale Aggregate Remote Sensing Images Based on Improved YOLO

KUANG Xian-yan, WANG Xing-xing\*, WANG Long-feng, ZHANG Zu-liang

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

**[Abstract]** In order to solve the problems of missed detection and false detection in the current remote sensing image small target detection task, a SMCA + CSC + shape-aware intersection over union loss (SIoU)-you only look once (SCS-YOLO) remote sensing image small target detection algorithm was proposed. Firstly, in response to the problem of small and clustered targets in remote sensing images, a spatial multi-scale convolutional attention module (SMCA) was constructed to improve the model's feature extraction ability of spatial and channel information. Secondly, in order to solve the problem that the semantic information of small targets was easy to be lost during deep network transmission, the aggregation subpixel convolution module concentrated sub-pixel convolution (CSC) was designed, and the multi-scale aggregation feature extraction method was used to enhance the ability of the network to extract semantic information. Finally, the SIoU loss function was used to replace the complete intersection over union loss (CIoU) loss function in the original model, which accelerated the convergence speed of the network. The average of the average precision (mAP) of the SCS-YOLO model reaches 97% and 90.9% on the RSOD and NWPU VHR-10 datasets, respectively, which is 2.2% and 2.7% higher than that of the original model, which shows the effectiveness of the method in the small target detection task of remote sensing images.

**[Keywords]** remote sensing images; SMCA + CSC + SIoU you only look once (SCS-YOLO); small target; attention; aggregated sub-pixel convolution; SIoU

遥感技术在航空航海动态监控、生态资源环境监测和军事工程目标检测等多个方面发挥重要作用<sup>[1]</sup>。遥感图像中特定小目标的检测和精准定位可以给军事侦察、气象预测、导弹预警等领域带来

诸多帮助。然而, 由于拍摄遥感图像距离较远, 导致拍摄的图片背景复杂, 且目标分布小而密集。因此, 研究一种提高遥感图像小目标检测精度的算法具有重要意义。

收稿日期: 2024-07-25; 修订日期: 2025-04-21

基金项目: 国家自然科学基金(51268017, 72061016)

第一作者: 邝先验(1976—), 男, 汉族, 江西赣州人, 博士, 教授。研究方向: 深度学习和计算机视觉。E-mail: xianyankuang@163.com。

\*通信作者: 王星星(2000—), 男, 汉族, 江西抚州人, 硕士硕士生。研究方向: 深度学习和计算机视觉。E-mail: 1772650505@qq.com。

投稿网址: www.stae.com.cn

目前,针对遥感图像目标检测的研究方法主要分为传统方法和基于深度学习方法。传统方法主要有 AdaBoost<sup>[2]</sup>、支持向量机 (support vector machine, SVM)<sup>[3]</sup> 等,早期,其在图像检测中得到广泛应用,核心思想是将线性不可分的问题转换为线性可分的问题,然后通过内积来解决,但在检测速度和检测精度上还有待提高。随着深度卷积神经网络<sup>[4]</sup> (convolutional neural networks, CNN) 被应用于图像领域,并且取得了重大突破。受此影响,研究人员将深度学习迁移到遥感图像检测中。与传统的方法相比, CNN 独特的结构特点,不仅能够降低模型的参数量,也使得模型可以更好地学习特征,基于深度学习的方法显著提升了目标检测精度和检测效率。基于深度学习的遥感图像检测算法分为一阶段和二阶段两种主流算法。其中二阶段检测算法的代表有 Fast R-CNN<sup>[5]</sup>、Faster R-CNN<sup>[6]</sup>、SPP-Net<sup>[7]</sup> 等。以上算法先生成候选框,再对候选框进行目标分类与定位,能较准确检测出目标,但存在计算量和参数量大,检测效率低等问题。一阶段检测算法直接从输入图像中提取所有特征,预测目标的类别和位置。代表算法主要有:YOLO<sup>[8-11]</sup> (you only look once)、SSD<sup>[12]</sup> (single shot multibox detector)、RetinaNet<sup>[13]</sup>、CenterNet<sup>[14]</sup> 等。这些算法具有高检测精度的同时显著提高了检测速度,因此得到了广泛的应用。在一阶段算法中, YOLO 算法能更好地平衡检测精度和速度,但由于遥感图像中存在背景复杂,目标小而密集等特点,因此, YOLO 算法在遥感图像上的应用还有待改进。

闫钧华等<sup>[15]</sup>利用深度卷积神经网络逐级提取目标特征,得到高低层特征金字塔,然后在融合部分进行跨层级特征融合,在一定程度上能获取更多弱小目标的信息,但对小目标检测还存在不足。Shang 等<sup>[16]</sup>基于 YOLOv5s 模型,提出加权双向金字塔网络 (multi-level bidirectional feature pyramid network, Mul-BiFPN),增强了多尺度特征图之间的信息传递,同时提出注意力机制 M-SimAM 模块,突出关键点信息,改进后的模型降低了参数量,但检测精度有待提高。Zhao 等<sup>[17]</sup>提出了小目标检测模型 MS-YOLOv7,该模型通过多个检测头和卷积注意力来提取丰富的特征,但对于遮挡严重的目标检测效果不明显。孔建国等<sup>[18]</sup>在 YOLOv8 模型基础上设计了特征融合 C3 (feature fusion C3, FFC3) 模块,实现多尺度特征融合;设计 CSPPF (CBAM-spatial pyramid pooling fast) 模块,进一步提高了检测速度和精度;最后用数据集增强技术改良了数据集,改进后的模型提高了检测速度和精度,但对于背景复杂的

目标检测还存在不足。李红岩等<sup>[19]</sup>将引入 BiFPN 结构替代原有的 PANet 结构,提高对不同尺度特征的信息融合,但没有平衡好推理速度和检测精度。

以上方法在一定程度上提高了网络的检测精度,但对于背景复杂的小目标检测仍存在局限性。鉴于此,针对遥感图像小目标检测做出针对性改进,提出 SCS-YOLOSCS-YOLO [SMCA + CSC + shape-aware intersection over union loss (SIoU)-you only lookonce] 算法用于遥感图像小目标检测,针对遥感图像存在背景复杂,目标小而聚集的问题,设计空间多尺度卷积注意力 (spatial multi-scale convolutional attention, SMCA) 模块,将 SMCA 模块添加到主干网络特征提取部分,增强通道和空间语义信息的提取能力;针对从空间金字塔池化输出的照片分辨率低,存在着小目标信息丢失的问题,设计聚合亚像素卷积 (concentrated sub-pixel convolution, CSC) 模块,将 CSC 模块嵌入 SPPCSP 后面,不仅能获取更丰富的小目标语义信息的同时,同时增强了网络的特征融合能力;针对遥感图像目标形状多样化,引入 SIoU (shape-aware intersection over union loss) 损失函数替换网络中的 CIoU (complete intersection over union loss),加快网络的收敛速度,提高检测精度。遥感图像目标检测在军事领域和民用领域都发挥着重要的作用。

## 1 模型设计

SCS-YOLO 模型如图 1 所示,其中红色为改进部分。该模型主要由主干网络模块 (Backbone)、特征融合模块 (Neck) 和检测头模块 (Head) 组成,每个部分都有着相应的功能和作用。主干网络由多层 CBS 层、MP 模块以及多个高效的 E-ELAN 模块构成,主要任务为从输入的遥感图像中提取丰富的特征信息。颈部网络主要采用 PANet 网络结构,主要目的是进行特征融合。在检测头部分提出了辅助头的训练方法来提升训练精度,每个检测头针对不同尺度的特征图进行预测,以检测不同大小的目标。

针对背景信息复杂的遥感图像小目标存在漏检和误检的问题,改进工作包括:首先,构造 SMCA 模块并将其添加到主干网络部分,使得网络更加关注目标区域,忽略背景区域,有效增强主干网络对通道和空间语义信息的提取能力;其次,针对卷积增多导致小目标语义信息丢失的问题,设计了 CSC 模块并将其嵌入空间金字塔池化 (spatial pyramid pooling, SPP) 和跨阶段部分通道连接 (cross stage partial channel, CSPC) 后面,通过多尺度的特征融合增强网络的特征融合能力,提高了小目标检测精度;最后为了使预测框适应遥感图像目标形状的

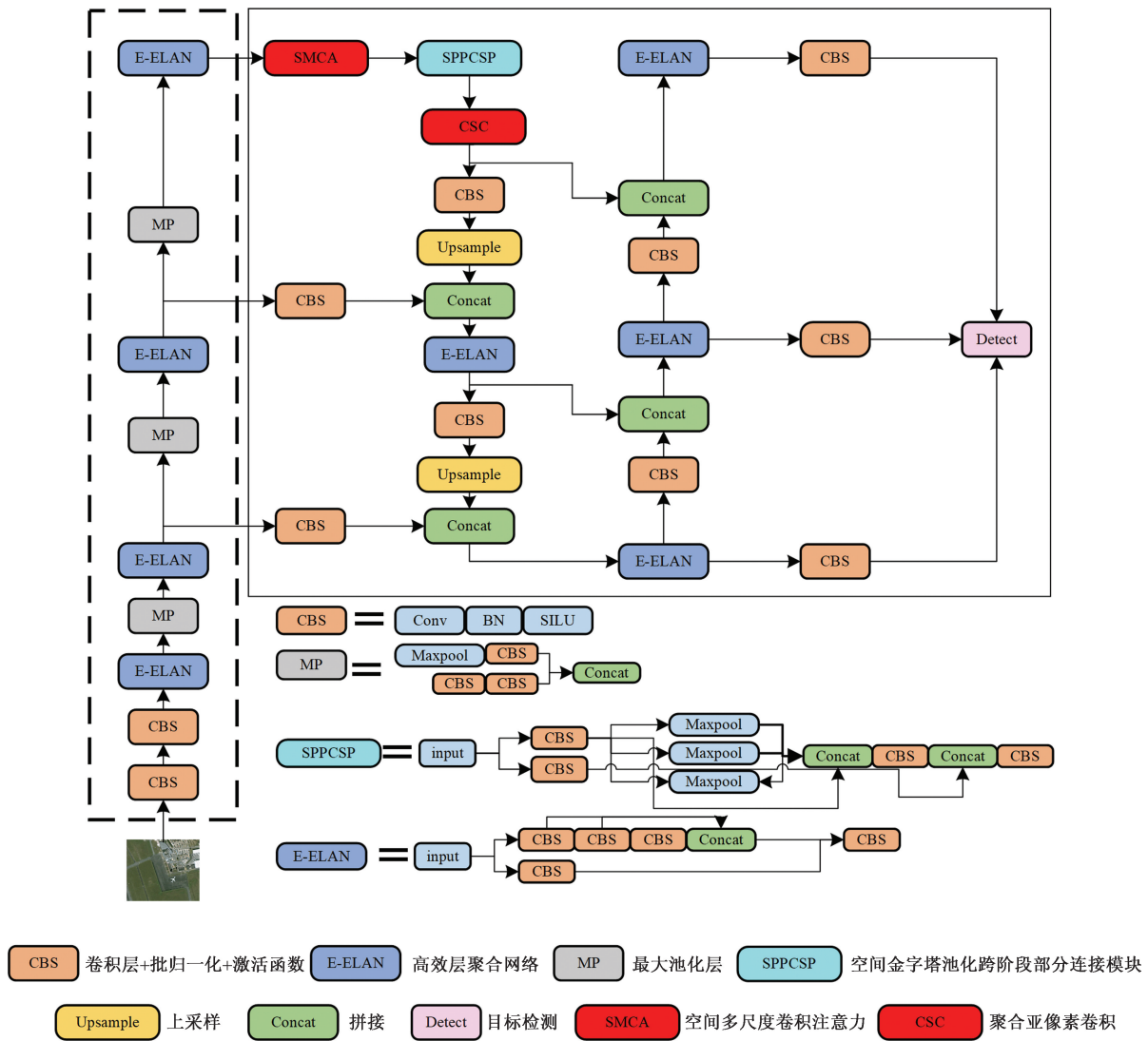


图1 SCS-YOLO 网络结构

Fig. 1 SCS-YOLO network structure

多样性,引入 Siou 替换原模型中的 CioU,提高网络的收敛速度。

### 1.1 SMCA 注意力机制

输入的遥感图像存在着背景复杂的问题,在经过多次特征提取网络卷积操作后,背景信息大量累加会产生大量冗余信息,给目标特征提取带来困难。因此,众多研究者尝试在主干网络部分融合注意力机制模块,注意力机制有弱化背景信息,强化特征信息的作用。常见的注意力机制有 SE (squeeze-and-excitation)<sup>[20]</sup>、CA (coordinate attention)<sup>[21]</sup>、CBAM (convolutional block attention module)<sup>[22]</sup> 等。SE 采用通道注意力和通道注意力融合的方式,抑制不重要通道的网络,但未涉及空间特征的关系;CA 考虑通道语义信息同时注意到方向相关的位置信息,但网络需要对整个特征图计算权重,显著增加了计算量;CBAM 同时涉

及通道维度和空间维度,但忽视了空间和通道的相互影响,导致在小目标检测任务中效果不佳。因此,设计 SMCA 注意力机制。

SMCA 注意力结合通道注意力和空间注意力,且根据它们的相互作用给予通道特征图和空间特征图对应的权重,提升网络的特征提取能力。利用深度可分离卷积捕捉特征之间的空间关系,使模型轻量化的同时,更好的聚集空间信息,强化遥感图像中小目标的特征,同时降低了计算量。SMCA 总体结构遵循着先通道(CA)后空间(SA)的顺序,其网络结构如图2所示。

SMCA 注意力机制的前半部分为通道注意力。首先,对输入特征图同时做平均池化和最大池化的处理,相当于对输入特征图的每个通道计算其平均特征值和最大特征值,计算结束后生成两个包含通道数的

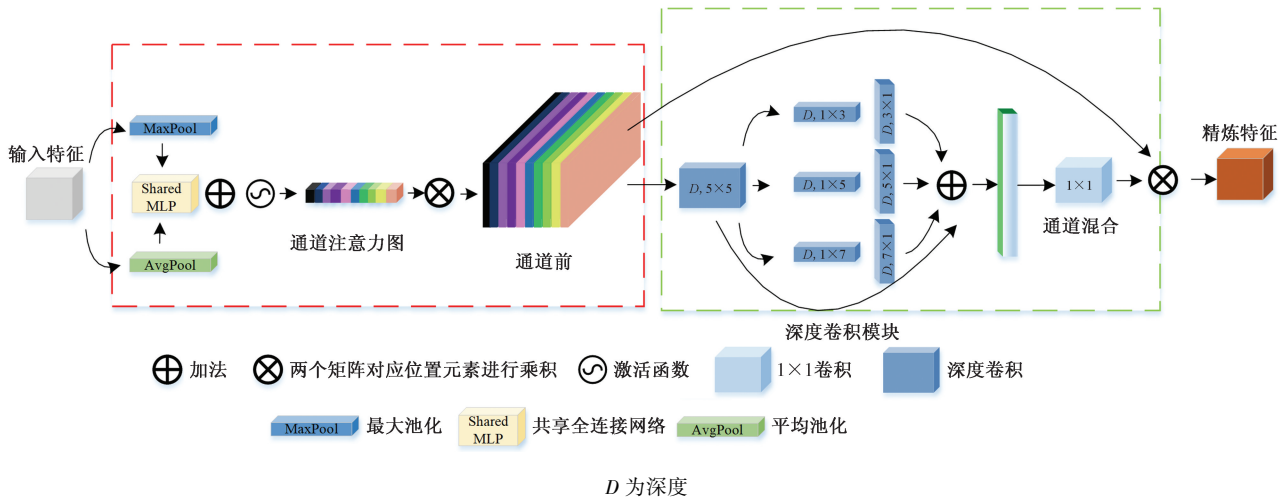


图2 SMCA 网络结构

Fig. 2 SMCA network structure

向量;其次,这两个向量经过一个共享全连接层 (MLP),这个全连接层计算每个通道的注意力权重,然后,为了使注意力权重不大于1,使用Sigmoid激活函数对每个通道的注意力权重进行处理;最后,将处理后的权重与原始特征图的每个通道相乘,得到通道特征图  $F_c$ 。这将强调对当前任务有帮助的通道,并抑制无关的通道,见式(1)~式(4)所示。

$$P(F) = \text{AvgPool}(F) + \text{MaxPool}(F) \quad (1)$$

$$X(F) = \text{MLP}[P(F)] \quad (2)$$

$$\text{CA} = \sigma[X(F)] \quad (3)$$

$$F_c = \text{CA}(F) \otimes F \quad (4)$$

式中:  $F \in \mathbf{R}^{C \times H \times W}$  为输入特征图,其中  $C$  为通道数;  $H$  为高度;  $W$  为宽度;  $\text{AvgPool}(F)$  为平均池化的结果;  $\text{MaxPool}(F)$  为最大池化的结果;  $P(F)$  为最大池化和平均池化相加的结果;  $X(F)$  为全连接层输出的结果;  $F_c$  为输出的通道特征图;  $\sigma$  为 Sigmoid 激活函数。

SMCA 注意力机制的后半部分为空间注意力。将通道注意力的输出  $F_c$  作为空间注意力的输入,首先进入一个  $5 \times 5$  的深度可分离卷积,捕捉特征之间的空间关系;其次,采用不同大小的卷积核卷积操作,提高网络捕获空间语义信息的能力;最后,使用一个  $1 \times 1$  的卷积改变 SA 的通道数,使输出的空间特征图 SA 与通道特征图  $F_c$  的通道数保持一致。这一过程如式(5)~式(7)所示。

$$Q(F) = \text{DwConv}(F_c) \quad (5)$$

$$Y(F) = \sum_{i=0}^3 \text{Branch}_i[Q(F_c)] \quad (6)$$

$$\text{SA}(F) = \text{Conv}_{1 \times 1}[Y(F_c)] \quad (7)$$

式中:  $Q(F)$  为输出的通道特征图进行深度卷积操作后的结果;  $Y(F)$  表示分别对  $Q(F)$  进行  $1 \times 3$ 、 $1 \times 5$ 、 $1 \times 7$  的深度卷积操作,再将  $Q(F)$  与这 3 个结果进

行拼接的结果;  $\text{Branch}_i$  为第  $i$  条支路,  $i \in \{0, 1, 2, 3\}$ ;  $\text{DwConv}$  为深度卷积;  $\text{Conv}_{1 \times 1}$  为  $1 \times 1$  卷积;  $\text{Branch}_0$  为第 0 条支路,即  $Q(F)$  直接连接。

将得到的空间特征图 SA 与通道特征图  $F_c$  的每个通道相乘,得到通道和空间维度上的分配权重,可表示为

$$R = \text{SA}(F_c) \otimes F_c \quad (8)$$

### 1.2 聚合亚像素卷积 (CSC) 模块

在深度学习特征提取过程中,伴随着卷积的增多,每经过一层网络都会损失一部分特征,输入特征图经过多次特征提取,从空间金字塔池化层输出的特征图分辨率仅为  $8 \times 8$ 。由于图像分辨率的减小,会导致图像细节的丢失,从而出现小目标语义信息丢失的问题,这会导致对小目标任务检测精度不高,出现漏检的情况。为了提高图像的分辨率,Shi 等<sup>[23]</sup>引入亚像素卷积,但由于目标特征分布是不均匀的,在提取边界时会出现伪影现象,导致输出结果的不平滑。因此,设计聚合亚像素卷积 (CSC) 模块,使用 focus 和亚像素卷积获得更丰富的语义特征信息,同时增强网络的特征融合能力,提高小目标检测精度。CSC 网络结构如图 3 所示。

focus 操作把输入的空间信息变为通道信息,在  $x$  倍下采样比例下可以将通道扩增为原来的  $x^2$  倍。亚像素卷积将通道信息变为空间信息,  $x^2$  倍通道缩减时可以实现  $x$  倍的上采样效果,亚像素卷积与 focus 搭配使用可以在不改变输入特征图分辨率的前提下实现特征融合。首先,当输入特征图大小为  $(C, H, W)$  时,使用 focus 和亚像素卷积获得 0.25、0.5、2、4 倍采样率下的输出特征图共 4 组,其大小分别为  $(16C, 0.25H, 0.25W)$ 、 $(4C, 0.5H, 0.5W)$ 、 $(0.25C, 2H, 2W)$ 、 $(0.0625C, 4H, 4W)$ ,

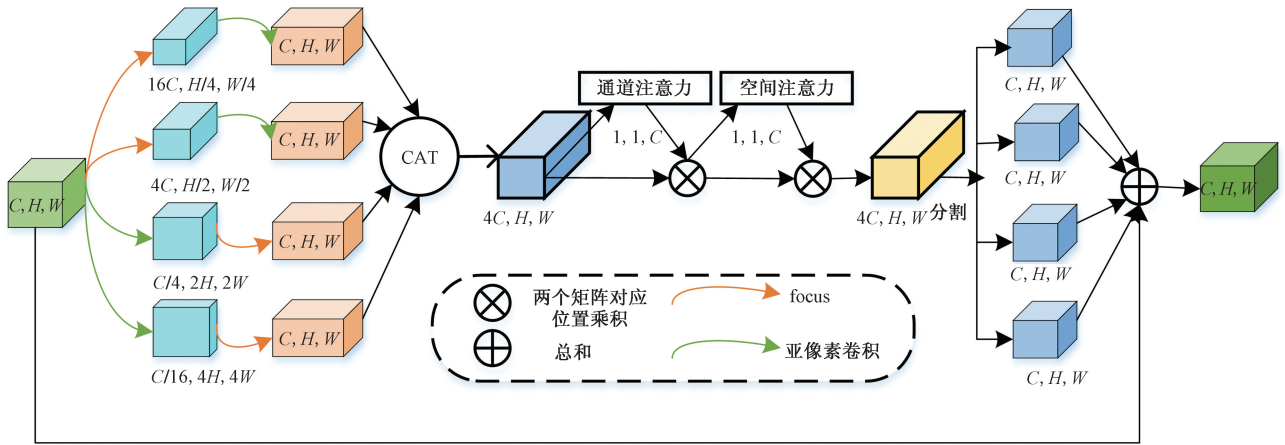


图3 CSC网络结构

Fig. 3 CSC network structure

多尺度输出特征图包含比单个特征图更丰富的特征信息;然后,分别使用亚像素卷积和 focus 将 4 组特征图的大小恢复到  $(C, H, W)$  并沿通道方向进行拼接,得到大小为  $(4C, H, W)$  的特征图;再然后,使用 GAM 注意力机制对特征图进行加权,使重要参数的表现更加突出;最后,沿通道方向进行分割并求和,使最后的输出特征图跟输入特征图大小一样。

### 1.3 SIoU 损失函数

损失函数 IoU 的定义为预测框和实际框的交集比上这两者并集。YOLOv7-tiny 网络模型中损失函数由定位损失、置信度损失和分类损失组成,可表示为

$$L_{total, Loss} = L_{loc, Loss} + L_{conf, Loss} + L_{class, Loss} \quad (9)$$

式(9)中: $L_{loc, Loss}$  为定位损失; $L_{conf, Loss}$  为置信度损失; $L_{class, Loss}$  为分类损失。

定位损失是根据 CIoU<sup>[24]</sup> 的计算公式得到的。CIoU 引入长宽比,然而当预测框和实际框的长宽比相等时,长宽比的惩罚项就发挥不了作用,CIoU 损失函数就无法发挥作用。

基于 CIoU 存在上述的缺陷,引入 SIoU<sup>[25]</sup> 损失函数替换原网络的损失函数。SIoU 损失函数包括角度成本、距离成本和形状成本。

SIoU 损失函数计算公式所需参数如图 4 所示。根据  $\alpha$  和  $\beta$  角度的大小,判断收敛过程首先是最小化  $\alpha$  还是  $\beta$ ,角度成本定义如式(10)所示。

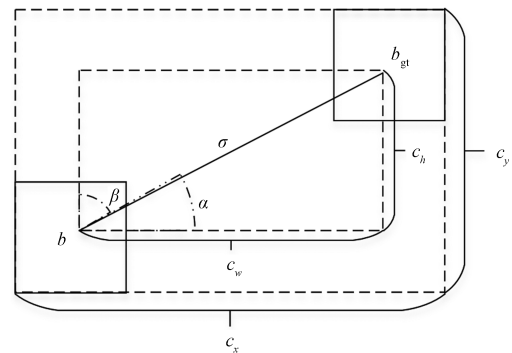
$$\Lambda = 1 - 2\sin^2\left(\arcsin x - \frac{\pi}{4}\right) \quad (10)$$

$$x = \frac{c_h}{\sigma'} = \sin\alpha \quad (11)$$

$$\sigma' = \sqrt{(b_{c_x, gt} - b_{c_x})^2 + (b_{c_y, gt} - b_{c_y})^2} \quad (12)$$

$$c_h = \max(b_{c_y, gt}, b_{c_y}) - \min(b_{c_y, gt}, b_{c_y}) \quad (13)$$

式中: $\Lambda$  为角度成本; $x$  为  $\alpha$  角度的正弦值; $c_h$  为预测框和真实框中心点的高度; $\sigma'$  为预测框和真实框中心



$c_x$  最小闭合区域的宽度;  $c_y$  最小闭合区域的高度;  $c_w$  预测框和真实框中心点的宽度;  $c_h$  预测框和真实框中心点的高度;  $\sigma$  预测框和真实框中心点的距离;  $\alpha$  预测框和真实框中心点连线与水平方向的夹角;  $\beta$  预测框和真实框中心点连线与垂直方向的夹角

图4 SIoU 损失函数的计算

Fig. 4 Calculation of SIoU loss function

点距离; $\alpha$  为预测框和真实框中心点连线与水平方向的夹角; $b_{c_x, gt}$  为预测框的横坐标; $b_{c_x}$  为真实框的横坐标; $b_{c_y, gt}$  为预测框的纵坐标; $b_{c_y}$  为真实框的纵坐标。

距离成本表示;预测框和真实框两者中心点的距离。根据定义的角度成本,SIoU 重新定义距离成本,如式(14)所示。

$$\Delta = \sum_{t=x,y} (1 - e^{-\rho_t}) \quad (14)$$

$$\rho_x = \left(\frac{b_{c_x, gt} - b_{c_x}}{c_w}\right)^2 \quad (15)$$

$$\rho_y = \left(\frac{b_{c_y, gt} - b_{c_y}}{c_h}\right)^2 \quad (16)$$

$$\gamma = 2 - \Lambda \quad (17)$$

式中: $\Delta$  为距离成本; $\gamma$  为调节系数; $\rho_t$  为真实框与预测宽度框中心点的距离差; $\rho_x$  为归一化的水平距离平方; $\rho_y$  为归一化的垂直距离平方。

由距离成本公式得出, $\alpha$  越小,距离成本的贡献越小。相反,当  $\alpha$  越接近  $\frac{\pi}{4}$ , 距离成本贡献越大。

随着角度增大,  $\gamma$  被赋予时间优先的距离值。形状成本定义为

$$\Omega = \sum_{t=w,h} (1 - e^{-\omega_t})^\theta \quad (18)$$

$$\omega_w = \frac{|w - w_{gt}|}{\max(w, w_{gt})} \quad (19)$$

$$\omega_h = \frac{|h - h_{gt}|}{\max(h, h_{gt})} \quad (20)$$

式中:  $\Omega$  为形状成本;  $w_t$  为真实框与预测框的尺寸差异;  $\theta$  为形状损失的关注程度, 将  $\theta$  设定为 1;  $w$  为预测框宽度;  $w_{gt}$  为真实框宽度;  $h$  为预测框高度;  $h_{gt}$  为真实框高度。

$$L_{\text{Siou}} = 1 - I_{\text{IoU}} + \frac{\Delta + \Omega}{2} \quad (21)$$

式(21)中:  $L_{\text{Siou}}$  为 Siou 损失函数值;  $I_{\text{IoU}}$  为交并比。

Siou 引入向量角度, 通过向量角度, 能够更好地反应目标的形状特征, 算法能更快地找到最优收敛方向。Siou 损失函数还重新定义了惩罚指标, 由于增加了向量角度, 惩罚项为 0, 概率就降低了。相比于原来的 CIoU 损失函数, 改进后的损失函数显著提高了收敛速度, 提高了检测的精度。

## 2 实验数据与分析

### 2.1 实验环境

模型训练环境为 Windows 11 操作系统, CPU 使用 13th Gen Intel(R) Core(TM) i7-13700H, 内存为 16 GB, GPU 为 8 G 的 NVIDIA GeForce RTX 4060 Laptop, 编程语言为 Python3. 10, 深度学习框架为 Pytorch1. 13. 0。

训练参数设置如下: 初始学习率 (learning rate) 为 0. 01, 训练周期 (epochs) 为 300, 批处理量 (batch-size) 为 16, 输入照片分辨率为  $640 \times 640$ 。

### 2.2 实验数据集

本实验数据集选取 RSOD 和 NWPU VHR-10。其中, RSOD 数据集<sup>[26]</sup> 是一个开放的遥感图像目标检测数据集, 数据集中包含飞机, 油箱, 运动场和立交桥四类。数据集共计 976 张图, 6 950 个目标。其中包含 4 658 个像素  $a$  为  $0 < a < 30^2$  的小目标样本, 764 个  $30^2 < a < 60^2$  的中目标样本, 以及 1 528 个  $a > 60^2$  的大目标样本, 实验将数据集按 8: 2 随机分为训练集、测试集。NWPU VHR-10<sup>[27]</sup> 也是一个公开且广泛使用的遥感图像目标检测数据集, 数据集中包含飞机、船、储罐等 10 种类别, 数据集一共有 800 张图, 3 651 个目标, 按 8: 2 随机分为训练集、测试集。

### 2.3 实验评估指标

实验的评估指标由精确率 ( $P$ )、召回率 ( $R$ ) 和各类平均精度的均值 (mAP) 组成。

精确率是指正确预测为正样本的数量占预测为正样本总数量的比例, 其中正确预测为正样本的总数量用 TP 表示, 错误预测为正样本的总数量用 FP 表示, 则精确率的表达式为

$$P = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \quad (22)$$

召回率指正确预测为正样本的数量占实际为正样本的总数量的比例, 其中正确预测为正样本的总数量用 TP 表示, 错误预测为正样本的总数量用 FP 表示, 则召回率的表达式为

$$R = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \quad (23)$$

计算 mAP 要先计算一种类别的平均精度 AP, AP 的定义为  $P$ - $R$  曲线下所包围的面积, 其计算公式为

$$\text{AP} = \int_0^1 P(t) dt \quad (24)$$

因此, 对于有  $M$  个类别的检测目标, mAP 的表达式为

$$\text{mAP} = \frac{\sum_{m=1}^M \text{AP}_m}{M} \times 100\% \quad (25)$$

理论上, 上述指标完全能够衡量模型的性能, 但在实际应用中还需要考虑模型的大小和检测速度。帧率 (frames per second, FPS) 常被用来衡量模型的检测速度, 每秒钟读取图片的多少的值即为 FPS, 数值越大则检测速度越快。

### 2.4 数据分析

为了验证 SMCA, CSC 模块和 Siou 损失函数对遥感图像小目标检测的效果, 在 RSOD 数据集上进行消融实验, 使用  $P$ 、 $R$  和 mAP@0. 5 (基于交并比阈值 0. 5 的平均精度均值) 作为性能评价指标, 实验结果如表 1 所示。

从表 1 可以看出, YOLOv7-tiny 模型在 RSOD 数据集上的 mAP@0. 5 精度为 94. 8%, 单独添加 SMCA 模块后的 mAP@0. 5 精度为 95. 9%, 说明 SMCA 模块通过调节通道和空间注意力的权重, 有

表 1 消融实验

组别	$P/\%$	$R/\%$	mAP@0. 5/ $\%$	FPS/ $\%$
YOLOv7-tiny	93. 6	95. 1	94. 8	96. 4
YOLOv7-tiny + CSC	94. 6	94. 8	96. 0	85. 6
YOLOv7-tiny + Siou	94. 0	94. 9	95. 6	<b>98. 7</b>
YOLOv7-tiny + SMCA	95. 2	93. 8	95. 9	92. 5
YOLOv7-tiny + CSC + Siou	94. 5	95. 2	96. 3	88. 4
YOLOv7-tiny + CSC + SMCA	95. 2	94. 8	96. 5	81. 8
YOLOv7-tiny + SMCA + Siou	95. 1	94. 5	96. 3	94. 0
YOLOv7-tiny + SMCA + CSC + Siou (SCS-YOLO)	95. 6	95. 3	<b>97. 0</b>	82. 9

注: 加粗数字表示对应指标效果最佳。

效地加强了网络对目标特征的捕获能力。在添加 SMCA 模块的基础上再引入 CSC 模块后精确度基本不变,召回率提升了约 1.0%,各类平均精度均值提高了约 0.6%,FPS 降低了 10.7 帧/s,说明添加 CSC 网络,提高了目标检测精度,但降低了检测速度,表明该模块能有效增强图像分辨率,减少小目标语义信息的丢失,并且加强网络特征融合能力,但也因此增加了模型的参数量。由于 SIOU 能加快收敛速度,理论上引入 SIOU 能提高检测速度,使 FPS 有所提高。在引入 SMCA 和 CSC 模块后添加 SIOU 损失函数,此时,精确度提升了约 0.4%,召回率提升了约 0.5%,各类平均精度均值提高了约 0.5%,FPS 提高了 1.1 帧/s,其他模型在引入 SIOU 后 FPS 也都有明显提高。

由实验数据分析可知,各种改进后的模块都对遥感图像小目标的检测有明显的效果,但由于模块的增加,模型的参数量也相应地增加,网络模型变大,使得 FPS 略有降低。

由消融实验的结果可知,SCS-YOLO 模型比其他消融实验的模型无论是精确度、召回率还是各类平均精度均值均有明显提高,FPS 值也相差不大,证明了该模型的可行性和有效性。

表 2 为 SCS-YOLO 算法和 YOLOv7-tiny 算法、YOLOv8 算法对 RSOD 数据集上每一类 mAP 结果对比。

从表 2 可以看出,SCS-YOLO 算法比 YOLOv7-tiny 算法各个类的 mAP 值都有所提升,尤其是对飞机提升最大,mAP 提升了 2.8%,对油箱提升最小,mAP 只提升了 1.4%;SCS-YOLO 算法对比 YOLOv8 算法,立交桥的 mAP 值低了 1.2%,但另外三类的 mAP 都要更高,总体呈现的各类平均精度均值要高于 YOLOv8,进一步证明了 SCS-YOLO 对遥感图像小目标检测效果比 YOLOv7-tiny 和 YOLOv8 更好。

为了验证 SMCA 注意力机制在遥感图像上的检测的优势,以 YOLOv7-tiny + CSC + SIOU (CS-YOLO) 为基础模型,引入主流的注意力机制做对比实验,实验结果如表 3 所示。可以看出,SE 通过融合通道注意力方式来抑制不重要的通道,但没有考虑空间

特征的关系,mAP 反而下降了 0.3%;CA 考虑了方向相关的位置信息,但缺乏空间维度的信息,mAP 上升了 0.1%;CBAM 通过学习通道特征和空间特征的权重,增强目标特征,mAP 提升了 0.4%,但增加了许多参数量;SMCA 模块沿用了 CBAM 模块通道注意力,空间注意力部分使用可分离卷积,在减少大量参数量的前提下,mAP 提升了 0.7%,验证了 SMCA 模块在遥感图像检测中的优势。

为了体现 SIOU 损失函数在遥感图像相比于其他损失函数的优势,以 YOLOv7-tiny + SMCA + CSC 为实验模型基础,引入不同的损失函数做对比实验,实验结果如表 4 所示。可以看出,DIoU 考虑了重叠面积和中心点距离,提高了收敛速度,但没有考虑到长宽比,因此 mAP 仅提高了 0.2%;EIoU 将高宽损失添加进惩罚项,进一步提高了收敛速度,但由于遥感图像中目标检测难易程度相差甚大,因此 mAP 反而下降了 0.5%;SIOU 重新定义了角度惩罚向量,它可以使预测框快速移动到距离最近的轴,然后只需要一个回归坐标,有效减少了自由度的总数,加快收敛速度的同时,mAP 提升了 0.5%,验证了 SIOU 在遥感图像检测上的优势。

为了进一步验证 SCS-YOLO 算法模型的性能,在同一台设备上将其 YOLOv3、YOLOv4、YOLOv5、YOLOv7-tiny、YOLOv8 做对比实验,仍然采用 RSOD 数据集,使用 mAP@0.5 和 FPS 为定量分析的定量指标,实验结果如表 5 所示。可以看出,SCS-YOLO 的各类平均精度均值相比于其他目标检测算法都有明显的提升。分析原因在于:YOLOv3 主干网络使用的是 darknet53,与其他深度学习网络相比,网

表 3 引入不同注意力机制的对比实验结果

Table 3 Comparative experimental results of introducing mainstream attention mechanisms

CS-YOLO	SE	CA	CBAM	SMCA	P/%	R/%	mAP/%	参数量
					94.5	95.2	96.3	6 548 674
					94.8	94.3	96.0	6 581 442
					95.0	94.7	96.4	6 574 322
					95.1	94.9	96.7	6 586 245
					95.6	95.3	97.0	6 552 463

注:表示在原模型基础上增加的注意力机制。

表 4 引入不同损失函数的对比实验结果

Table 4 Comparative experimental results of introducing different loss functions

损失函数	P/%	R/%	mAP/%
CIoU	95.2	94.8	96.5
DIoU	95.5	94.6	96.7
EIoU	95.0	94.3	96.0
SIOU	95.6	95.3	97.0

表 2 RSOD 数据集每个类 mAP 对比

Table 2 Comparison of mAP for each class in the RSOD dataset

类别	YOLOv7-tiny	YOLOv8	SCS-YOLO
飞机	0.932	0.952	0.960
油箱	0.969	0.974	0.983
立交桥	0.930	0.964	0.952
操场	0.968	0.976	0.985

络复杂,并且召回率差、定位不准确导致各类平均精度均值最低。YOLOv4 对输入的特征图采用 Mosaic数据增强,能更好地检测提取出目标特征,但由于没有利用全局上下文信息,对于目标密集的检测不稳定。YOLOv5 对小目标特征难以定位和分类,会出现漏检或误检的情况。

表 5 主流检测算法在 RSOD、NWPU VHR-10 数据集上的性能

Table 5 Performance of mainstream detection algorithms on RSOD, NWPU VHR-10 datasets

数据集	组别	mAP@0.5/%	FPS/(帧·s <sup>-1</sup> )
RSOD	YOLOv3	<b>85.8</b>	24.3
	YOLOv4	89.5	42.5
	YOLOv5	93.5	45.7
	YOLOv7-tiny	94.8	<b>96.4</b>
	YOLOv7	95.7	63.1
	YOLOv8	96.6	78.7
	SCS-YOLO	<b>97.0</b>	82.9
NWPU VHR-10	YOLOv3	<b>74.1</b>	<b>37.4</b>
	YOLOv4	85.9	64.5
	YOLOv5	87.5	69.7
	YOLOv7-tiny	88.2	<b>153.2</b>
	YOLOv7	89.9	95.6
	YOLOv8	90.5	120.7
	SCS-YOLO	<b>90.9</b>	132.5

注:加粗数字表示对应指标效果最佳。

YOLOv7 模型参数量大,计算复杂,导致检测速度略慢。YOLOv8 在目标检测的精度和检测效率方面都有不错的效果。

SCS-YOLO 算法,各类平均精度均值达到 97.0%,比其他算法都要高。与 YOLOv7-tiny 对比各类平均精度有明显的提高,而 FPS 值略微有所下降;与 YOLOv8 对比,无论是各类平均精度还是 FPS 都有所提升。因此,SCS-YOLO 算法综合性能对比其他算法更有优势。

为了验证 SCS-YOLO 算法的鲁棒性,在同一台设备上将其与 YOLOv3、YOLOv4、YOLOv5、YOLOv7-tiny、YOLOv7、YOLOv8 在 NWPU VHR-10 上继续做对比实验,使用 mAP@0.5 和 FPS 为定量分析的定量指标,实验结果如表 5 所示。

如表 5 所示,对于 NWPU VHR-10 数据集,主流目标检测算法 YOLOv3 算法各类平均精度均值和 FPS 值都最低,分别为 74.1% 和 37.4 帧/s;YOLOv7-tiny 的检测速度最快,FPS 达到 153.2 帧/s;SCS-YOLO 算法的各类平均精度均值最高为 97.0%。SCS-YOLO 比其他主流目标检测算法的综合性能都要好,证明了 SCS-YOLO 算法具有鲁棒性。

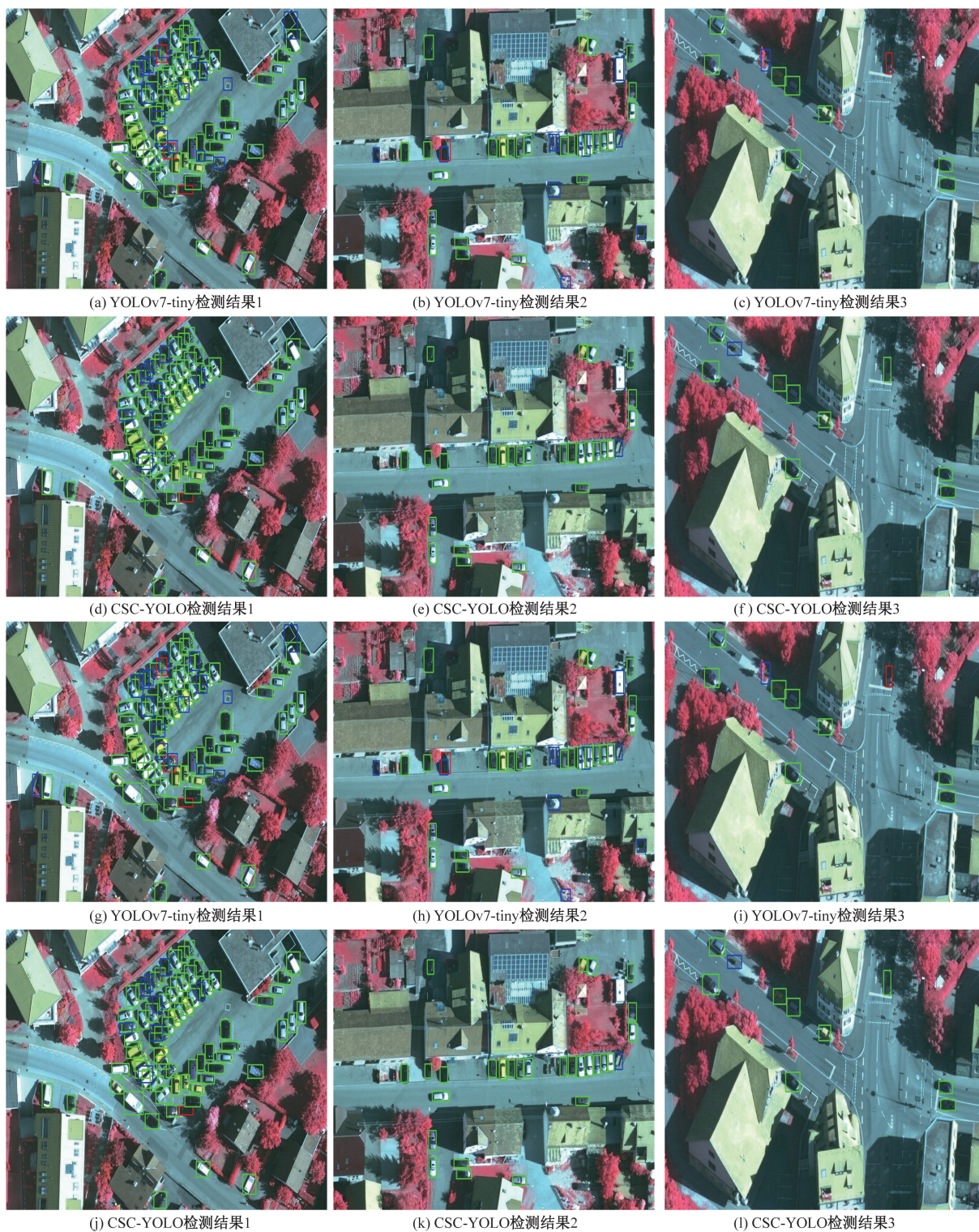
为了有一个直观的效果,将 SCS-YOLO 算法与原 YOLOv7-tiny 算法在 RSOD 和 NWPU VHR-10 数据集上的检测结果进行可视化展示。图 5 和图 6 为



绿色框区域表示检测目标正确;蓝色框区域表示检测目标错误;红色框区域表示漏检目标

图 5 RSOD 数据集上的可视化结果

Fig. 5 Visualization results on RSOD dataset



绿色框区域表示检测目标正确;蓝色框区域表示检测目标错误;红色框区域表示漏检目标

图6 NWPU VHR-10 数据集上的可视化结果

Fig. 6 Visualization results on NWPU VHR-10 dataset

YOLOv7-tiny 算法与 SCS-YOLO 算法在几个遥感图像典型样本的检测结果示例图。

从可视化结果(图 5、图 6)分析可知,SCS-YOLO能减少检测目标错误的数量并且还检测出原始 YOLOv7-tiny 漏检的遥感小目标,从而提高了遥

感图像小目标检测的精度。

### 3 结论

提出一种 SCS-YOLO 算法,有效降低了遥感图像中小目标漏检和误检的比例。首先,设计 SMCA

模块,考虑通道注意力与空间注意力相互作用,通过给予通道特征图和空间特征图对应的权重,提升网络对通道和空间语义信息提取能力,增强网络对小目标语义信息的提取聚焦的能力;其次,构建CSC模块,通过多尺度特征图的输出,提高网络特征融合的能力,从而改善小目标检测精度;最后,引入SIoU损失函数更好反应目标的形状特征,提升网络的收敛速度。在ROSD和NWPU VHR-10数据集上进行消融实验和对比实验,验证其有效性,mAP分别提升2.2%和2.7%。结果表明,本文算法对遥感图像目标检测有一定的应用价值。但也存在两点不足:一是由于SCS-YOLO算法嵌入两个模块,使模型增加了许多参数量,FPS较原有模型略有降低,后续可以进行轻量化的研究;二是没有考虑到被遮挡的遥感图像目标的检测,未来可以通过调整网络结构和优化新训练等方法进一步提高检测精度。

#### 参 考 文 献

- [1] Khanal S, Kc K, Fulton J P, et al. Remote sensing in agriculture: accomplishments, limitations and opportunities[J]. *Remote Sensing*, 2020, 12(22): 3783.
- [2] Wang W, Sun D. The improved AdaBoost algorithms for imbalanced data classification[J]. *Information Sciences*, 2021, 563: 358-374.
- [3] Pisner D A, Schnyer D M. Support vector machine[M]. Hoboken: John Wiley International Publishing Group, 2020: 101-121.
- [4] Zhang K, Zuo W, Gu S, et al. Learning deep CNN denoiser prior for image restoration[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2017: 3929-3938.
- [5] Bharati P, Pramanik A. Deep learning techniques: R-CNN to mask R-CNN: a survey[J]. *Computational Intelligence in Pattern Recognition: Proceedings of CIPR*, 2019, 20: 657-668.
- [6] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. *Advances in neural information processing systems*, 2012, 25: DOI: 10.1016/j.aasri.2012.05.013.
- [7] Li L, Yang Z, Jiao L, et al. High-resolution SAR change detection based on ROI and SPP net[J]. *IEEE Access*, 2019, 7: 177009-177022.
- [8] Jiang P, Ergu D, Liu F, et al. A review of YOLO algorithm developments[J]. *Procedia Computer Science*, 2022, 199: 1066-1073.
- [9] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2017: 7263-7271.
- [10] Farhadi A, Redmon J. YOLOv3: an incremental improvement[C]//Computer Vision and Pattern Recognition. Berlin: Springer, 2018: 1-6.
- [11] Lee Y H, Kim Y. Comparison of CNN and YOLO for object detection[J]. *Journal of the Semiconductor & Display Technology*, 2020, 19(1): 85-92.
- [12] Zhai S, Shang D, Wang S, et al. DF-SSD: an improved SSD object detection algorithm based on DenseNet and feature fusion[J]. *IEEE Access*, 2020, 8: 24344-24357.
- [13] Cheng M, Bai J, Li L, et al. Tiny-RetinaNet: a one-stage detector for real-time object detection[C]//11th International Conference on Graphics and Image Processing (ICGIP 2019). Bellingham: SPIE, 2020: 195-202.
- [14] Wang H, Xu Y, Wang Z, et al. CenterNet-auto: a multi-object visual detection algorithm for autonomous driving scenes based on improved CenterNet[J]. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2023, 3: DOI: 10.1109/TETCI.2023.3235381.
- [15] 闫钧华, 张琨, 施天俊, 等. 融合多层次特征的遥感图像地面弱小目标检测[J]. *仪器仪表学报*, 2022, 43(3): 221-229. Yan Junhua, Zhang Kun, Shi Tianjun, et al. Ground weak target detection in remote sensing images fused with multi-level features[J]. *Chinese Journal of Scientific Instrument*, 2022, 43(3): 221-229.
- [16] Shang J, Wang J, Liu S, et al. small target detection algorithm for UAV aerial photography based on improved YOLOv5s[J]. *Electronics*, 2023, 12(11): DOI: 10.3390/electronics12112434.
- [17] Zhao L L, Zhu M L. MS-YOLOv7: YOLOv7 based on multi-scale for object detection on UAV aerial photography[J]. *Drones*, 2023, 7(3): 188.
- [18] 孔建国, 张向伟, 赵志伟, 等. 基于改进YOLOv8的机场飞鸟实时目标检测方法[J]. *科学技术与工程*, 2024, 24(32): 13944-13952. Kong Jianguo, Zhang Xiangwei, Zhao Zhiwei, et al. Real-time target detection method for flying birds in airports based on improved YOLOv8[J]. *Science Technology and Engineering*, 2024, 24(32): 13944-13952.
- [19] 李红岩, 徐保庆, 张子扬, 等. 基于全局上下文信息的遥感图像小目标检测[J]. *光学学报*, 2024, 44(24): 205-212. Li Hongyan, Xu Baoqing, Zhang Ziyang, et al. Small object detection in remote sensing images based on global context information[J]. *Acta Optica Sinica*, 2024, 44(24): 205-212.
- [20] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 7132-7141.
- [21] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2021: 13713-13722.
- [22] Su H, Wang X, Han T, et al. Research on a U-Net bridge crack identification and feature-calculation methods based on a CBAM attention mechanism[J]. *Buildings*, 2022, 12(10): 1561.
- [23] Shi W, Caballero J, Huszar F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2016: 1874-1883.
- [24] Zheng Z, Wang P, Liu W, et al. Distance-IoU loss: faster and

- better learning for bounding box regression [C]//Proceedings of the AAAI Conference on Artificial Intelligence. Reston: AAAI, 2020: 12993-13000.
- [25] Gevorgyan Z. SLoU loss: more powerful learning for bounding box regression[J]. arXiv Preprint, 2022; arXiv: 2205.12740.
- [26] Sun W, Dai L, Zhang X, et al. RSOD: real-time small object detection algorithm in UAV-based traffic monitoring[J]. Applied Intelligence, 2022, 22: 1-16.
- [27] Chen X, Jiang J, Li Z, et al. An online continual object detector on VHR remote sensing images with class imbalance[J]. Engineering Applications of Artificial Intelligence, 2023, 117: 105549.