

DOI: 10.13382/j.jemi.B2306892

基于 DFECANet 的遥感图像飞机目标检测方法*

单慧琳^{1,2} 吕宗奎² 付相为² 胡宇翔¹ 段修贤¹ 张银胜^{1,2}

(1. 无锡学院电子信息工程学院 无锡 214105; 2. 南京信息工程大学电子与信息工程学院 南京 210044)

摘要:针对现有的遥感图像目标检测方法中对小尺寸飞机目标的检测精度不高、特征信息传递不准确、信息交互不充分等问题,提出了一种基于可辨别特征提取和上下文感知的遥感图像飞机目标检测方法。设计了以可辨别特征提取模块为主体的主干网络,用以加强对多尺度飞机目标的特征提取;引入自适应特征增强模块,选择性关注小目标、优化特征信息的传递与信息交互;并设计了特征融合上采样模块对特征图进行上采样操作,用以提升高层语义信息的准确性。在 DOTAv1 数据集上的检测精度达到了 95.2%,相较于 YOLOv5s,SCRDet,ASSD 等主流算法,飞机目标的检测精度提高了 3.7%~18%。此外,该方法的检测速度以及模型参数量分别为 147 fps 和 13.4 M,相较于当前主流算法具备较强的竞争力,满足在遥感背景下对飞机目标的实时检测需求。

关键词: 图像处理;目标检测;多尺度特征融合;遥感图像;特征上采样

中图分类号: TP751.2;TN919.5 **文献标识码:** A **国家标准学科分类代码:** 520.20

Aircraft target detection in remote sensing images based on DFECANet

Shan Huilin^{1,2} Lyu Zongkui² Fu Xiangwei² Hu Yuxiang¹ Duan Xiuxian¹ Zhang Yinsheng^{1,2}

(1. School of Electronic and Information Engineering, Wuxi University, Wuxi 214105, China; 2. School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China)

Abstract: Aiming at the existing remote sensing image target detection methods with low detection accuracy for small-size aircraft targets, inaccurate feature information transfer and insufficient information interaction, a remote sensing image aircraft target detection method based on discriminative feature extraction and context-awareness is proposed. A backbone network with a discriminative feature extraction module is designed to enhance feature extraction for multi-scale aircraft targets; an adaptive feature enhancement module is introduced to selectively focus on small targets and optimize the transfer of feature information and information interaction; and a feature fusion up-sampling module is designed to perform up-sampling operations on the feature maps to improve the accuracy of high-level semantic information. The detection accuracy on the DOTAv1 dataset reaches 95.2%, which is 3.7% to 18% higher than that of mainstream algorithms such as YOLOv5s, SCRDet, ASSD. In addition, the detection speed and the number of model parameters of the proposed method are 147 frames per second and 13.4 M, respectively. Compared with the current mainstream algorithms, the proposed method has strong competitiveness and meets the real-time detection requirements of aircraft targets in the background of remote sensing.

Keywords: image processing; target detection; multi-scale feature fusion; remote sensing images; feature upsampling

0 引言

目标检测作为计算机视觉领域的核心任务之一,近年来得到了前所未有的发展^[1]。与此同时,随着遥感技

术的进步,卫星遥感图像因具备分辨率高、覆盖面积广等特点,使得图像中的信息量越来越丰富,在军事和民用领域的作用也越来越明显。飞机作为重要的军事装备和交通工具,利用目标检测算法对其进行精准定位识别,在军事侦察^[2]、情报分析、军事决策以及机场监控管理等应用

场景具有重要意义。

然而,由于遥感图像一般由卫星或无人机从高空中以鸟瞰视角获取,因此存在复杂的空间场景和大量不同的目标物。与普通图像相比,遥感图像中的目标往往尺度更小、排列更密集、像素占比更少^[3],而且很难与图像背景分离,这加大了特征提取难度,导致检测精度较低。不仅如此,现有的一些方法因模型过于庞大,难以在现实场景中有效应用。因此,探索和应用遥感图像目标检测算法是一个意义重大且充满挑战性的课题,相关领域研究人员已进行了大量的工作^[4-6]。传统目标检测中,具有代表性的方法为模板匹配法^[7],但这类方法普遍存在比较明显的缺陷,不仅易受光照等自然环境因素影响,导致检测精度下降,还需要人工设定规则或模板,且这些规则和模板往往受到应用场景的限制,泛化性能较差。

近年来,基于深度学习的目标检测方法^[8]在目标检测领域进展显著,相比于传统检测算法,深度学习算法通过学习数据集的特征实现自动提取图像中目标的特征信息与位置信息,相应的检测器性能得到大幅度提升。基于深度学习的目标检测算法分为两阶段和一阶段^[9]。两阶段算法集成了卷积神经网络(convolutional neural network, CNN)等深度学习技术,检测性能出色,其中 CNN 通常被用作特征提取器提取图像特征。具有代表性的两阶段算法有 R-CNN^[10]、Faster R-CNN^[11],此类方法的核心思想是区域提取,通过在图像中提取候选区域,再对这些区域进行特征提取,并通过非极大值抑制获取最终结果。与两阶段目标检测算法不同,一阶段算法将目标检测任务转化为端到端的分类和回归问题,以实现快速检测和识别图像中的目标,代表性算法有 YOLOv5^[12]、SSD^[13]、RetinaNet^[14]等。

然而自然场景下的目标检测算法在遥感级别的目标检测任务中表现欠佳,研究人员已从多个角度出发寻求解决方案。Li 等^[15]在检测网络的特征融合部分引入高分辨率特征图 P2 以及引入 Swin 变压器作为检测头等方法,以解决模型对重要特征的处理和融合能力不足的问题;Yu 等^[16]提出了一种基于全局和局部多尺度特征融合的遥感飞机检测方法。该方法采用编码器-解码器结构,分别通过网络编码器部分提取目标的局部特征,解码器部分提取目标的全局特征,以提升模型检测精度以及召回率。Zong 等^[17]设计了一种反向特征金字塔结构和跨尺度位移网络以提高模型推理速度,同时有效缓解来自非相邻层的语义在特征金字塔中被稀释的情况。Qiao 等^[18]认为传统的特征金字塔网络中不同的特征域的浅层和深层功能抑制模型的性能,同时上层特征的积累会导致下层特征出现混叠现象,从而干扰对小目标的特征表示。为此,利用高分辨率图像和原始图像的深度特征来形成一个统一的特征域,并采用通道注意力融合模块

来学习上层特征,以改善小目标检测性能,但该方法大大增加了计算开销。Xu 等^[19]设计了 VOVDarkNet 作为特征提取骨干、AFC-PAN 作为特征融合网络,使模型能够学习到不同尺度物体的特征信息和位置信息,提高网络对小物体的感知能力。Wang 等^[20]设计了特征增强模块,以捕获更多小目标的判别特征,设计空间感知模块过滤空间信息,通过可形变卷积更好地聚焦前景对象。

上述方法在一定程度上提升了遥感图像目标检测器的性能,然而鉴于遥感图像中飞机目标的特殊性,仍存在如下几个问题亟待解决:1) 遥感图像中飞机目标像素有限,所含特征信息较少,而且方向多变、形状不一,现有的算法难以精准提取特征信息;2) 现有的特征金字塔网络(feature pyramid networks, FPN)通过横向连接进行跨尺度特征融合时,浅层小目标特征表达能力相对较弱,受背景信息干扰严重,导致模型对小目标关注度不足。另外,由于浅层特征可能包含一些噪声和不可靠的区域细节信息,这将导致低层特征信息缺失。同时,多尺度目标具有不同大小的尺度,其特征通常分布在不同的尺度特征图中,现有 FPN 的通过简单的横向连接进行特征融合时不能充分进行信息交互;3) 特征金字塔结构中普通上采样操作会造成特征细节丢失从而引入误差,影响高层语义信息的准确性。

为了解决上述问题,本文提出了一种基于可辨别特征提取和上下文感知(discriminative feature extraction and context awareness, DFECA)的光遥感图像飞机目标检测方法,主要贡献内容如下:1) 设计了可辨别特征提取模块(discriminable feature extraction module, DFEM),并以此搭建主干网络用于特征提取。DFEM 模块引入梯度路径的设计策略,采用多尺度特征提取分支以及非对称卷积分支强调对不同尺度、方向多变的飞机目标进行充分的特征提取;2) 设计了特征细节增强模块(feature detail enhancement module, FDEM),通过自适应的选择卷积核大小强调对小目标的关注,并解决 FPN 在横向连接过程中特征信息传递不准确、高底层级信息交互不充足的问题;3) 设计了特征聚合上采样模块(feature aggregation upsampling module, FAUM)用来取代上采样操作,在减少特征细节信息丢失的同时,也丰富了上下文语义信息。

1 研究方法

1.1 DFECANet 网络结构

本文提出的 DFECANet 网络的结构如图 1 所示,首先将待检测遥感图像输入到由 DFEM 模块和普通卷积层组建的主干网络进行特征提取,以获取高质量的多尺度特征信息。然后依据 FPN 结构,将提取的特征信息传入 FDEM 模块并结合 FAUM 模块辅助特征信息融合,获取

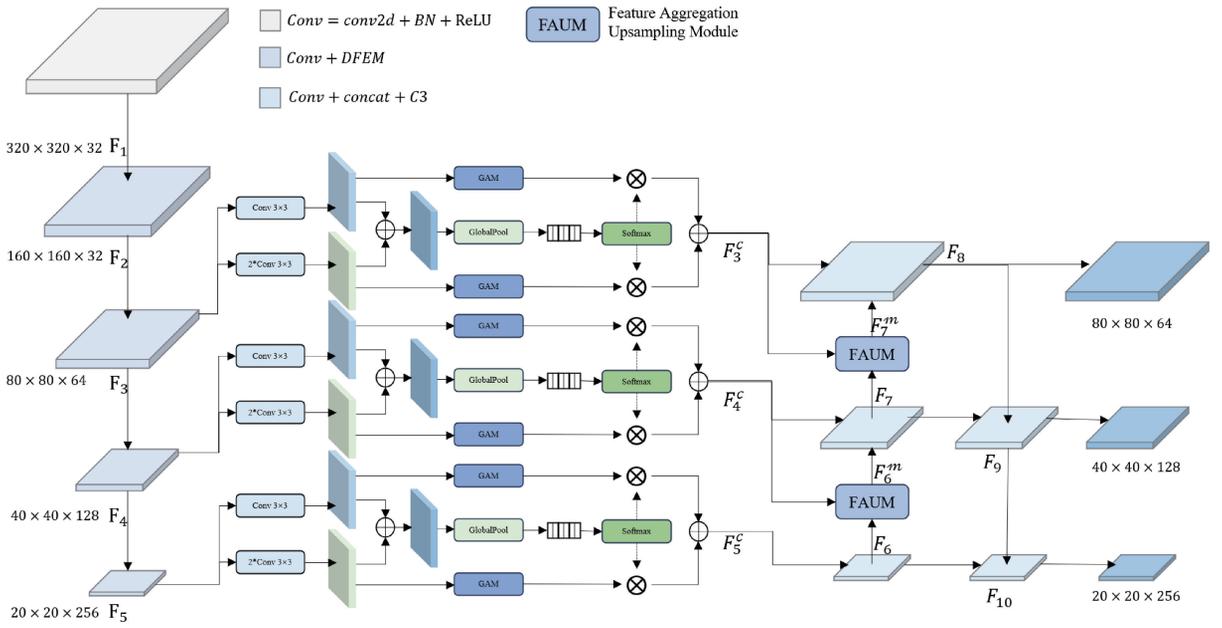


图 1 DFECANet 整体网络结构

Fig. 1 DFECANet overall network structure

丰富的上下文特征信息。最后通过检测头实现飞机目标检测。

1.2 可辨别特征提取模块

当前，热点的特征提取网络有 Mobilenet、CSPDarknet53、ResNet 等，其中 Mobilenet 网络采用的深度可分离卷积模块会对输入的特征图进行下采样操作，并通过降低卷积核的参数数量和计算量来提高计算效率，这降低了模型对细节信息的敏感度；而 CSPDarknet53 重

复使用 CSP 结构进行卷积操作，导致小目标的信息过多地丢失，从而影响网络对小目标的检测效果；而 ResNet 作为特征提取网络由于其参数量较大，增加了硬件部署的难度。

针对上述问题，本文设计了一种全新的 DFEM 模块构建特征提取网络。DFEM 模块遵循高效的网络设计策略^[21]，从硬件特性和梯度传播路径等角度来设计网络^[22]。DFEM 模块结构如图 2 所示。

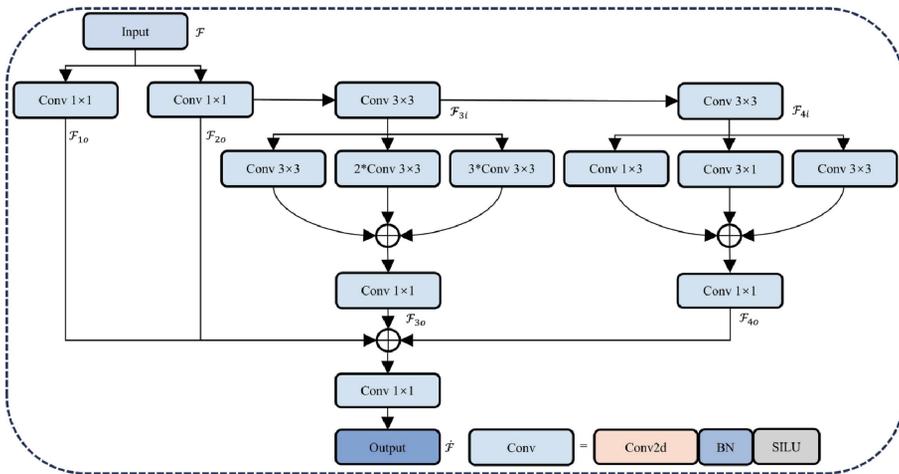


图 2 DFEM 模块结构

Fig. 2 DFEM module structure diagram

对于普通卷积层,其内存访问开销为(memory access cost, MAC)^[23]:

$$MAC = HW(c_1 + c_2) + c_1c_2 \quad (1)$$

其中, H 和 W 分别表示特征图的高和宽, HWc_1 和 HWc_2 分别表示输入、输出特征图内存访问成本,式(1)中,根据均值不等式,当且仅当 $c_1 = c_2$ 时, MAC 取值最小。因此本文的 DFEM 模块设定每条支路的各个卷积层的输入、输出特征矩阵通道数相等,以获得最小的内存访问开销。

对于 DFEM 整体结构而言,本文从梯度路径设计角度出发,设计了梯度分流策略,通过控制最短和最长的梯度路径,使模型能够学习到更多的特征,增强其鲁棒性。其中 DFEM 第 3 分支中分别采取了卷积核大小为分别为 3×3 、 5×5 、 7×7 的卷积层进行多尺度特征提取,在实际实验中,将 5×5 和 7×7 的卷积核用多个 3×3 的卷积核替换,这不仅可以更全面的描述目标特征,减少目标信息的丢失,还能减少模型中的参数数量以及增加网络的非线性。最后通过 concat 操作进行特征融合,丰富了上下文多尺度信息。具体表示为:

$$\mathcal{F}_{3i} = \text{Conv}_{3 \times 3}(\mathcal{F}_{2o}) \quad (2)$$

$$\mathcal{F}_{3o} = \text{Conv}_{1 \times 1} \{ \text{concat} [\text{Conv}_{3 \times 3}(\mathcal{F}_{3i}), 2 * \text{Conv}_{3 \times 3}(\mathcal{F}_{3i}), 3 * \text{Conv}_{3 \times 3}(\mathcal{F}_{3i})] \} \quad (3)$$

DFEM 的第 4 分支采用了非对称卷积^[24],本文考虑到卷积核为 $n \times n$ 的普通卷积层对目标的尺度和形状不够敏感,不能很好的适应不同的尺度、形状和方向的目标,可能导致目标特征丢失或信息冗余。而非对称卷积可以强化水平和垂直方向上的平方卷积核,更好的适应了目标在方向上的差异性,并在不同的方向上学习目标特征,

突出局部关键特征,提高小目标的特征判别性。此外,非对称卷积结构相较于普通卷积可以对模型进行压缩和加速,亦可减少模型参数量和计算量。本节采用 $\text{Conv}_{1 \times 1}$ 、 $\text{Conv}_{3 \times 3}$ 、 $\text{Conv}_{3 \times 1}$ 和 $\text{Conv}_{1 \times 3}$ 构建非对称卷积模块进行特征提取,并将其拼接进行特征融合操作,其中每个卷积模块后面都加上批量归一化和 ReLU 激活函数,可以更好地拟合通道间复杂的相关性,以避免图像在反向传播时候发生梯度消失等问题。具体可表示为:

$$\mathcal{F}_{4i} = \text{Conv}_{3 \times 3}(\mathcal{F}_{3i}) \quad (4)$$

$$\mathcal{F}_{4o} = \text{Conv}_{1 \times 1} \{ \text{concat} [\text{Conv}_{3 \times 3}(\mathcal{F}_{4i}), \text{Conv}_{3 \times 1}(\mathcal{F}_{3i}), \text{Conv}_{1 \times 3}(\mathcal{F}_{3i})] \} \quad (5)$$

DFEM 最终得到的输出特征 \mathcal{F} 表示为:

$$\mathcal{F} = \text{Conv}_{1 \times 1}(\text{concat} [\mathcal{F}_{1o}, \mathcal{F}_{2o}, \mathcal{F}_{3o}, \mathcal{F}_{4o}]) \quad (6)$$

1.3 特征细节增强模块

上下文信息对于特征融合尤为重要,特别是对于遥感这样复杂场景。FPN 通过自顶向下的路径和横向连接将高层语义和低层细节相结合,以获得多尺度上下文信息。但由于低层级的特征图分辨率较低,且遥感图像中飞机目标较小、方向和尺度多变,同时容易受背景信息、噪声影响,造成目标特征信息细节丢失,影响横向连接中特征信息的准确性,不能很好的信息交互。现有的方法 BiFPN^[25]、GFPN^[26]在一定程度上可以缓解上述问题,但同时引入了更长的融合路径,增加了模型计算的复杂度,无法在检测精度和推理速度上取得平衡。对此,本节提出 FDEM 模块来强调低层级中的小目标并优化特征细节信息,如图 3 所示。

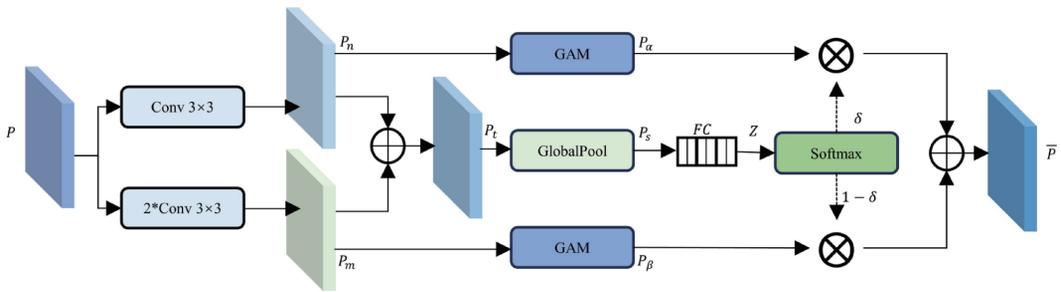


图 3 FDEM 模块结构

Fig. 3 FDEM module structure diagram

从局部来看,该模块首先对输入特征图 P 使用大小为 3×3 、 5×5 的卷积核分别进行卷积操作得到两个输出 P_n 和 P_m ,每个卷积层后依次进行批量归一化和 ReLU 激活操作,其中 5×5 的卷积核由两个 3×3 的卷积核提升替换以减少模型参数数量。随后先将上述的两个输出分别通过 GAM 注意力模块^[27]以增强模型对全局信息的感知并减少信息丢失,得到特征 P_α 和 P_β 。再对 P_n 和 P_m 进行逐元素相加操作,得到新的输出特征图 P_t :

$$P_t = \delta(B(\text{Conv}_{3 \times 3}(P_n))) \oplus \delta(B(2 * \text{Conv}_{3 \times 3}(P_m))) \quad (7)$$

式中: $B(\cdot)$ 为批量归一化, $\delta(\cdot)$ 为 ReLU 激活函数。接着通过全局平均池化进行降维处理得到一个具有全局信息的特征向量 P_s , P_s 经过全连接层后得到一个新特征向量 Z :

$$P_s = f_{\text{sp}}(P_t) \quad (8)$$

$$\mathbf{Z} = FC(\mathbf{P}_\alpha) \quad (9)$$

式中: $f_{sp}(\cdot)$ 表示全局池化操作, $FC(\cdot)$ 表示全连接操作。然后通过 softmax 函数计算不同卷积核权重并引导其与上下两个分支进行加权融合, 得到两个加权特征图, 此过程可以自适应地选择不同尺度的信息进行融合。最后再对各分支进行逐元素相加得到最终特征图 \bar{P} :

$$\bar{P} = (P_\alpha \otimes \delta) \oplus (P_\beta \otimes (1 - \delta)) \quad (10)$$

从整体来看, 该模块通过自适应地调整卷积核的感受野大小, 可以选择性关注遥感图像中小目标区域以及选择性的关注不同尺度的特征图, 在横向连接过程中促

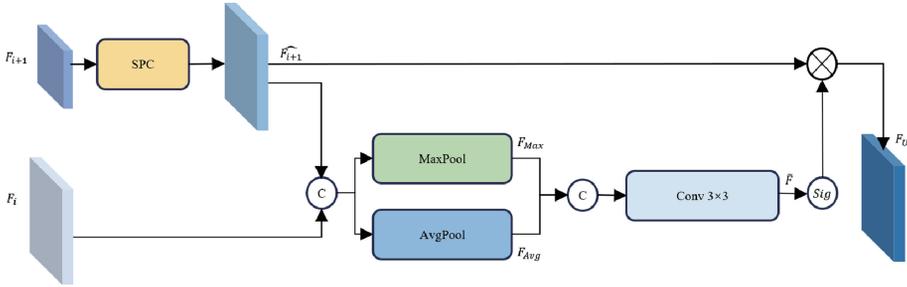


图4 FAUM 模块结构

Fig. 4 FAUM module structure diagram

FAUM 模块利用亚像素卷积 (sub-pixel convolution, SPC)^[29] 处理较高级特征图 F_{i+1} 得到 \widehat{F}_{i+1} , 使其与较低层级特征图 F_i 大小一致, SPC 可以有效缓解因通道降维带来的信息丢失。接着对 \widehat{F}_{i+1} 和 F_i 进行拼接操作对特征信息进行整合。拼接后的特征图分别进行最大池化和平均池化操作, 生成二维特征映射 F_{Max} 和 F_{Avg} , 具体表示为:

$$F_{Max} = Map(\text{concat}[SPC_{\mu=2}(\widehat{F}_{i+1})], F_i) \quad (11)$$

$$F_{Avg} = Avp(\text{concat}[SPC_{\mu=2}(\widehat{F}_{i+1})], F_i) \quad (12)$$

式中: $Map(\cdot)$ 和 $Avp(\cdot)$ 分别表示最大池化和平均池化操作, $SPC_{\mu=2}(\cdot)$ 表示放大系数为 2 的亚像素卷积。接着再将特征映射进行拼接, 并用卷积核大小为 3×3 卷积模块整合拼接后的特征:

$$\tilde{F} = \delta(B(\text{Conv}_{3 \times 3}(\text{concat}[F_{Max}, F_{Avg}]))) \quad (13)$$

式中: $B(\cdot)$ 为批量归一化, $\delta(\cdot)$ 为 ReLU 激活函数。最后, 整合后的特征经过激活操作生成空间权重, 并将空间权重和经过亚像素卷积输出的高级特征相乘, 使较高级别的特征包含更丰富的上下文语义信息, 最终得到的输出特征 F_U 表示为:

$$F_U = Sig(\tilde{F}) \otimes \widehat{F}_{i+1} \quad (14)$$

式中: $Sig(\cdot)$ 为 Sigmoid 函数。

进高、低层级特征更好的进行信息交互。

1.4 特征聚合上采样模块

上采样在 FPN 中有着显著的作用, 当前应用较为广泛的上采样操作方法主要是插值法, 包括临近上采样和双线性上采样^[28], 但上述方法都只关注图像的局部特征, 相对较小而感受野限制了它们对全局图像特征的表达, 而且插值法上采样会导致原始像素之间的信息丢失, 进而影响高层语义信息的准确性。针对上述方法中存在的问题, 本研究设计了 FAUM 模块来解决。FAUM 具体结构图如图 4 所示。

2 数据结果处理与分析

2.1 数据集介绍

本文采用经典的大规模遥感目标检测数据集 DOTA v1^[30] 对算法模型进行评估, 其中 DOTA v1 包含 2 806 张尺寸在 800~4 000 的遥感图像, 该数据集包含 15 个类别共计 18 882 个实例, 包括飞机、棒球场、桥梁、田径场、小型车辆、大型车辆、船舶、网球场、篮球场、储油罐、足球场、环形交叉路口、港口、游泳池、直升机, 采用水平目标标注框作为标签注释。为减少大像素图像输入网络时在调整图片大小步骤造成图像信息损失, 本文在原数据集的基础上对 DOTA v1 数据集中的图像进行图像切割操作, 经切割后的图像大小为 $1 024 \times 1 024$ 。数据集分为训练集和验证集, 其中训练集包含随机选取的 15 749 张图片, 验证集包含随机选取的 5 297 张图片。切割后的 DOTA v1 数据集部分样本如图 5 所示。

2.2 实验环境及参数配置

实验环境如下: 操作系统为 Windows 10, 实验设备中显卡采用 Nvidia GeForce RTX 3090 GPU, 显存为 32 G, CPU 型号为 12th Gen intel (R) Core(TM) i9-12900KF @ 3.19 GHz, 深度学习框架采用 Anconda + Python3.8 + Pytorch1.12.1。在模型训练过程中, 本文采取随机梯度下降优化器和余弦函数动态调整学习率策略, 初始学习



图 5 DOTA v1 数据集样本示例

Fig. 5 Sample DOTA v1 dataset

率为 1×10^{-3} , 动量参数设置为 0.937, 权重衰减系数为 5×10^{-4} , 批量大小为 16, 每次训练轮数为 300 轮。

2.3 评价指标

为了验证本文所提出的算法对遥感图像飞机目标的检测性能, 本文引入精确率 (precision, P)、召回率 (recall, R) 以及帧率 (frames per second, FPS) 作为评价指标。其中 P 用来表示预测正确的飞机数占所有飞机目标数的比例; R 表示预测结果中检测到的正确飞机数量, 也称为查全率; FPS 用来衡量检测速度, 表示算法在一秒内处理的帧数。此外, 由于该遥感图像检测场景需要考虑到将模型有效部署到硬件设备上, 本文也考虑了网络模型的大小问题。

P 、 R 、 AP 、 FPS 指标的计算方法如下:

$$P = \frac{TP}{TP + FP} \quad (15)$$

$$R = \frac{TP}{TP + FN} \quad (16)$$

$$AP = \frac{dy}{dx} \int_0^1 P(R) dR \quad (17)$$

$$FPS = \frac{FrameNum}{ElapsedTime} \quad (18)$$

其中, TP 表示正确预测为飞机的数数量, TN 表示正确预测为背景的数据, FP 表示错误预测为飞机的数量, FN 表示错误预测为背景的数量。 N 为检测目标类别总数, $FrameNum$ 为检测到的图像个数, $ElapsedTime$ 为图像处理时间。

2.4 实验结果分析

1) DFECANet 模型检测结果展示

图 6 展示了本文 DFECANet 模型在测试集上对 4 个不同场景图像的检测性能, 由检测结果可以看出, 本文检测模型在小目标场景、飞机密集区域、背景复杂区域以及飞机坟场 (包含残缺飞机) 等场景下不仅可以精确检测出相对较大的飞机目标, 对小飞机目标也有较高的检测精度, 实验结果证明了 DFECANet 对不同尺度地面飞机目标具有较强的鲁棒性, 同时, 较高的检测精度也验证了其性能的优越。

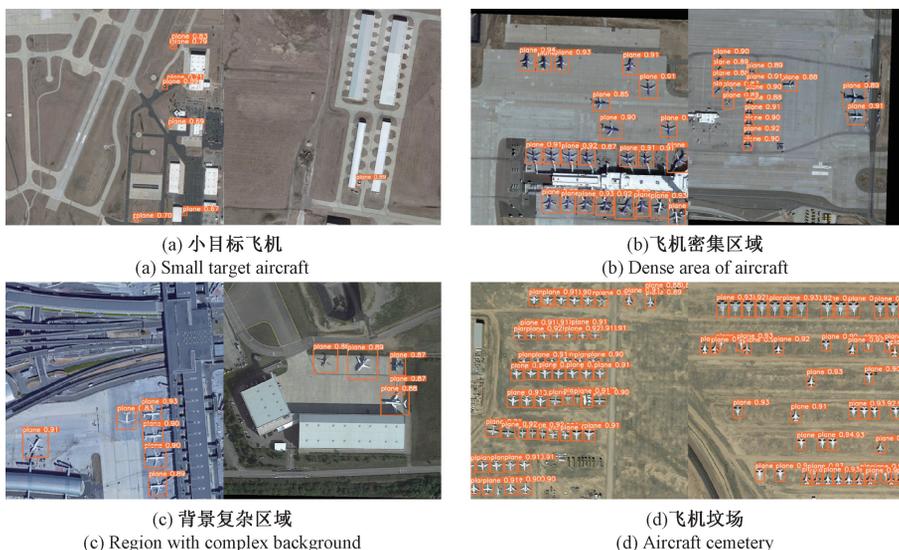


图 6 DOTA v1 数据集部分检测结果展示

Fig. 6 DOTA v1 dataset part of the detection results show

2) 与其他算法对比

为了更好的评估 DFECANet 模型对遥感图像中飞机目标的检测性能, 本节选择了现有的九种先进方法与 DFECANet 进行比较, 包括两阶段算法 SCRDet^[31]、Faster RCN^[11] 和一阶段算法 FMSS^[32]、ASSD^[33]、S2Ane^[34]、

YOLOv3^[35]、YOLOv4^[36]、YOLOv5, 从检测精度、帧率、模型参数量 3 个角度进行对比。

如表 1 所示, 从检测精度上来看, DFECANet 对飞机目标的检测精度达到 95.2%, 与基于 ResNet101 的 SCRDet 算法相比提升了 5%, 与基于 VGG16 的 FMSSD

和 ASSD 算法相比分别提升了 6.1% 和 5.9%, 与基于 CSPDarkNet53 的 YOLOv5 算法相比提升了 3.7%, 相较于另外 4 种算法, 本文提出的 DFECANet 算法检测精度具有更大的优势; 从检测速度上来看, DFECANet 算法的帧率为 147, 比 YOLOv5s 算法低 31.6, 略低于 YOLOv5s 算法的 178.6, 与其他以 ResNet101、VGG16 等作为主干网络的算法相比, 本文算法在检测效率上具备较大的优势,

满足实时检测的需求; 从模型参数量来看, DFECANet 算法的模型参数量为 13.4 M, 仅低于 YOLOv5s 的 7.1 M, 且远远小于其他算法的模型参数量, 适合部署到硬件设备上。综合 DFECANet 算法的检测精度、检测速度以及模型参数量 3 个客观指标, 该方法相较于其他算法的综合性能更优。

表 1 本文算法与另外 8 种算法检测性能对比

Table 1 Comparison of the detection performance of the algorithm in this paper with eight other algorithms

Method	Size	Backbone	AP/%	FPS	Parameters/M
SCRDet	800×800	ResNet101	90.2	12	34.1
Faster RCNN	800×600	VGG16	77.2	1.9	41.4
FMSSD	640×640	VGG16	89.1	16	99.7
ASSD	512×512	VGG16	89.3	21	30.2
S2Anet	416×416	R-50-FPN	88.9	57	38.6
YOLO v3	416×416	DarkNet53	79	13	61.5
YOLO v4	416×416	CSPDarkNet53	88.5	43	63.7
YOLOv5s	640×640	CSPDarkNet53	91.5	178.6	7.1
本文	640×640	DFEMS	95.2	147	13.4

3) 算法对比图

本节采取了表 1 中的 ASSD、SCRDet、YOLOv5s 3 种检测精度相对较高的模型与本文提出的 DFECANet 进行定性比较。对于常规尺寸但背景复杂的飞机目标, 4 种模型的检测精度相近。对于方向多变的飞机目标, DFECANet 模型的检测精度均超过 90%, 优于另外 3 种对比算法, 这得益于 DFEM 模块中非对称卷积强化了水平和垂直方向上的平方卷积核, 更好的适应了目标在方向上的差异性。对于小尺寸飞机目标, ASSD 出现了漏检现象, 而 SCRDet 和 YOLOv5s 的检测精度也都低于 75%, DFECANet 对小尺寸飞机目标的检测精度达到了 88%, 其中 DFEM 中的多尺度特征提取分支以及 FDEM 中的选择性关注小目标发挥了重要作用。图 7 充分体现了 DFECANet 的检测性优越性。

4) 消融实验

为了评估本文算法提出的关键模块的有效性和鲁棒性, 本节通过添加各个模块进行消融实验。由表 2 可知, 由 DFEM 模块组建的主干网络对遥感图像中的飞机目标的检测精度达到了 92.8%, 这得益于 DFEM 模块中的多尺度特征提取部分具有不同的感受野并且非对称卷积模

块可以有效的针对目标方向多变等问题, 保证了出色的特征提取能力。以 DFEM 为基准, 分别加入 FDEM 和 FAUM 模块时, 检测精度分别提高了 0.5% 和 1.1%, 这得益于 FDEM 可以通过自适应地调整卷积核的感受野大小选择不同尺度信息进行融合, 而 FAUM 通过低级特征聚合高级特征进行上采样强调特征细节, 从而带来精度的提升。当所有模块共同作用时, 模型检测精度达到了 95.2%, 充分验证了关键模块的有效性。此外, 以 DFEM 模块为主干网络组建算法, 其模型复杂度为 17GFLOPs, 其中 DFEM 模块通过保持输入输出通道数一致来减小内存访问开销, 并采用梯度路径传播角度设计模块, 从而降低网络模型复杂度。加入 FAUM 模块后为 17.3GFLOPs, 复杂度几乎不变, 这是由于其仅采用简单的池化、卷积等操作, 与普通上采样模块相比并不涉及更复杂的计算。引入 FDEM 模块后为 19.1GFLOPs, 模型复杂度提升约 12.3%, 这是由于在高底层级之间加入 FDEM 模块增强特征细节, 不可避免带来额外的计算量。所有模块共同作用时, DFECANet 算法的模型复杂度仅为 20.9GFLOPs, 仅提升 22%。

表 2 消融实验

Table 2 Ablation experiment

实验	DFEM	FDEM	FAUM	AP/%	FPS	Parameters/M	GFLOPs
1	√			92.8	172.4	9.1	17.0
2	√	√		93.3	151.5	13.4	19.1
3	√		√	93.9	157.3	9.1	17.3
4	√	√	√	95.2	147	13.4	20.9

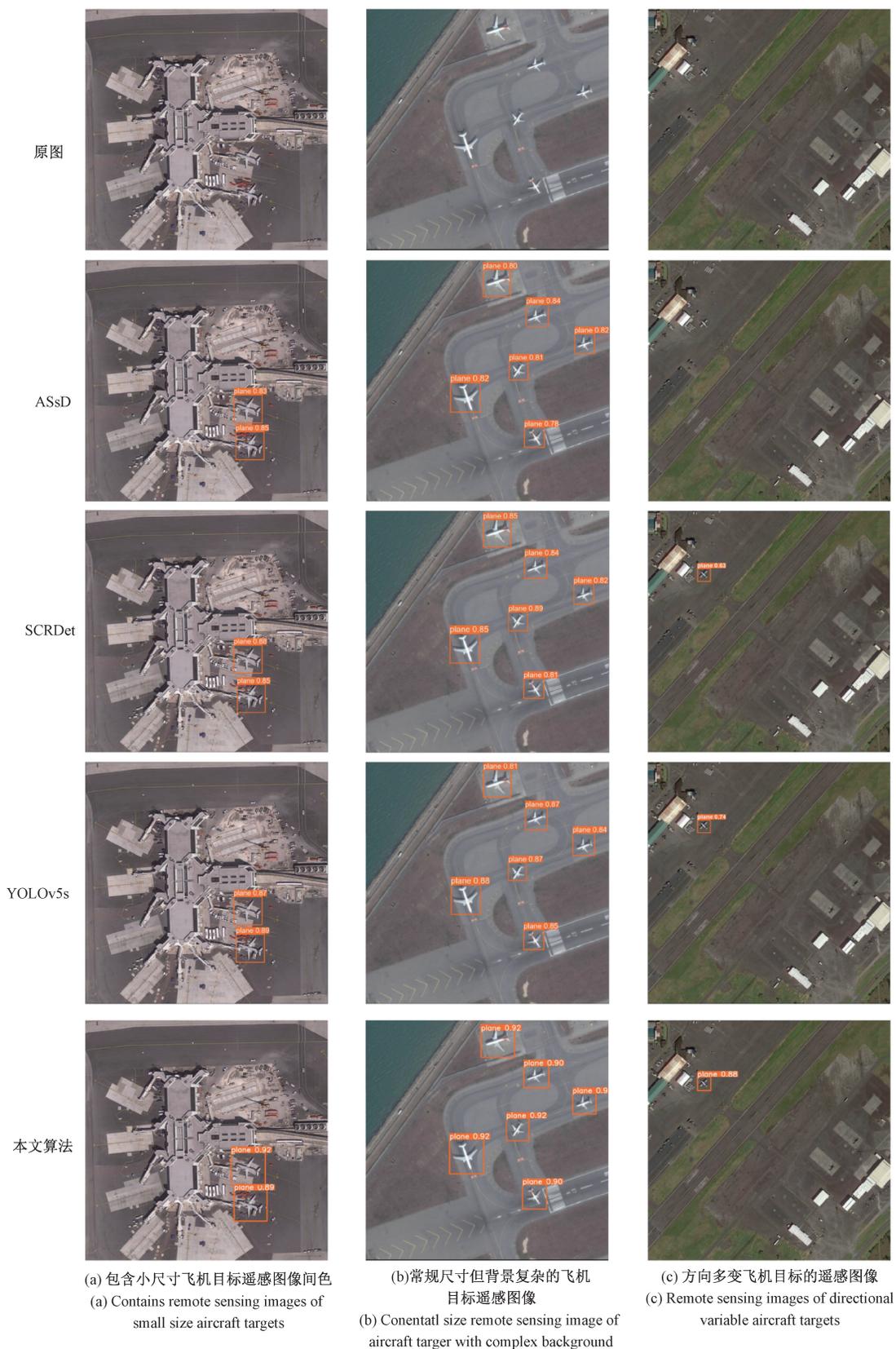


图 7 4 种算法对比

Fig. 7 Comparison chart of the four algorithms

5) 特征图可视化

为了进一步验证 FAUM 模块的有效性,本节分别对采用 Upsample 和 FAUM 进行上采样操作的前后特征图进行对比,如图 8 所示,图 8(a)、(b)和(e)、(f)分别展示了 YOLOv5s 先后经过两次上采样的前后特征图对比,图 8(c)、(d)和(g)、(h)分别展示了本文模型经过 FAUM 进行上采样的前后特征图对比。显然,FAUM 模块通过利用相对低层特征聚合相对高层特征来引导上采样,帮助模型生成更加细致和真实的特征信息,且更加关注感兴趣的目标区域。

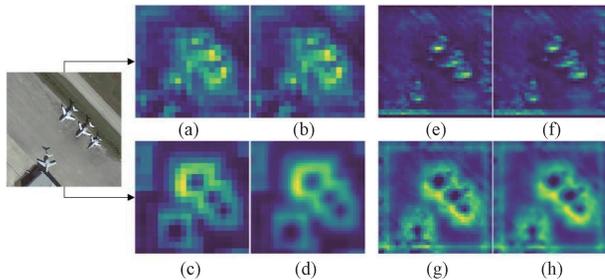


图 8 上采样前后对比

Fig. 8 Before and after upsampling

6) 热力图可视化

为了进一步体现本文模型的优越性能,本节分别对 YOLOv5s 和本文算法的 3 个不同阶段预测结果的特征图进行热力图可视化,如图 9 所示,图 9(1)为 YOLOv5s 的热力图可视化结果,图 9(2)为本文模型的热力图可视化结果。不难看出,图 9(1)相比于(2)本文模型对输入图像中的目标区域关注更加精准,对检测决策的贡献更大,此外,本文算法对相对较小的飞机目标特征区域响应更高、更均匀。

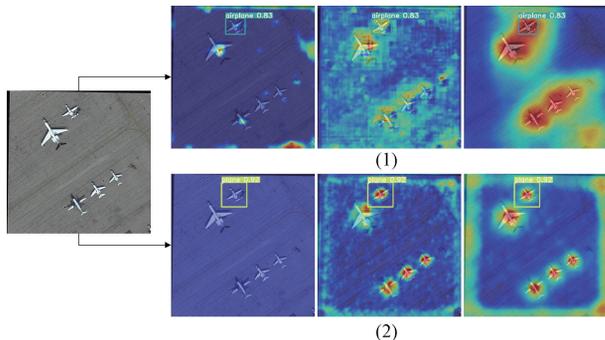


图 9 热力图可视化对比

Fig. 9 Heat map visualization comparison

3 结论

本文结合遥感图像空间场景复杂、目标尺寸小且方

向多变、排列密集等特点,以改善遥感图像中飞机小目标检测精度低等问题为首要目标,根据实际应用场景需求设计了一种用于遥感图像飞机目标检测特征金字塔网络 DFECANet。所提出的解决方案以特征金字塔架构为主体,主要包括 DFEM、FDEM、FAUM 3 个关键模块。其中 DFEM 为构建主干网络的主要模块,DFEM 从梯度路径设计角度出发,通过非对称卷积分支以及多尺度特征提取分支加强了对不同尺度、方向多变的飞机目标进行特征提取,有效地获取了更多目标特征信息。FDEM 在横向连接过程中通过自适应选择卷积核大小,灵活的选取不同尺度的信息进行融合,实现加强对小目标的关注度,促进高、底层级特征充分融合。FAUM 选取亚像素卷积进行上采样操作并聚合低层级特征,使高层级的特征包含更丰富的上下文语义信息,有效获取更多细节信息。DFECANet 通过在 DOTA v1 数据集上与其他先进算法对比分析、消融实验以及特征图对比等方法验证了本文算法在遥感图像飞机目标检测任务中的可靠性,为遥感图像飞机目标检测提供了一种新的解决方案。

参考文献

- [1] 陈朋磊,王江涛,张志伟,等. 基于特征聚合与多元协同特征交互的航拍图像小目标检测[J]. 电子测量与仪器学报, 2023, 37(10): 183-192.
CHEN P L, WANG J T, ZHANG ZH W, et al. Small target detection in aerial images based on feature aggregation and multiple collaborative feature interaction [J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(10): 183-192.
- [2] 李广帅,苏娟,李义红. 基于改进 Faster R-CNN 的 SAR 图像飞机检测算法[J]. 北京航空航天大学学报, 2021, 47(1): 159-168.
LI G SH, SU J, LI Y H. Improved Faster R-CNN-based aircraft detection algorithm for SAR images [J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(1): 159-168.
- [3] 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 J H, ZHANG K, SHI T J, et al. Ground dim small target detection in remote sensing images by fusing multi-level features [J]. Chinese Journal of Scientific Instrument, 2022, 43(3): 221-229.
- [4] ZHUANG W, TANG X G, YANG G, et al. Remote sensing image object detection based on rotatable bounding box [C]. 2021 International Conference on Digital Society and Intelligent Systems (DSInS). IEEE, 2021: 172-177.
- [5] CHENG X, ZHANG C. C 2-YOLO: Rotating object

- detection network for remote sensing images with complex backgrounds [C]. 2022 International Joint Conference on Neural Networks (IJCNN). IEEE, 2022: 1-8.
- [6] LIU Z, GAO Y, DU Q, et al. YOLO-extract: Improved YOLOv5 for aircraft object detection in remote sensing images [J]. IEEE Access, 2023, 11: 1742-1751.
- [7] 廖育荣, 王海宁, 林存宝, 等. 基于深度学习的光学遥感图像目标检测研究进展 [J]. 通信学报, 2022, 43(5): 190-203.
- LIAO Y R, WANG H N, LING C B, et al. Research progress on target detection in optical remote sensing images based on deep learning [J]. Journal of Communication, 2022, 43(5): 190-203.
- [8] 程旭, 宋晨, 史金钢, 等. 基于深度学习的通用目标检测研究综述 [J]. 电子学报, 2021, 49(7): 1428-1438.
- CHENG X, SONG CH, SHI J G, et al. A review of deep learning-based generic target detection research [J]. Journal of Electronics, 2021, 49(7): 1428-1438.
- [9] CHEN S, ZHAO J, ZHOU Y, et al. Info-FPN: An informative feature pyramid network for object detection in remote sensing images [J]. Expert Systems with Applications, 2023, 214: 119132.
- [10] GIRSHICK R, DONAHUE J, DARRELL T, et al. Region-based convolutional networks for accurate object detection and segmentation [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 38(1): 142-158.
- [11] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. Advances in Neural Information Processing Systems, 2015, 28.
- [12] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [13] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector [C]. Computer Vision - ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11 - 14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.
- [14] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980-2988.
- [15] LI R, YING S. YOLOS-R: A deep learning method for small target detection in infrared remote sensing images based on super-resolution and YOLO [J]. Signal Processing, 2023, 208: 0165-1684.
- [16] YU L, HU H, ZHONG Z, et al. GLF-Net: A target detection method based on global and local multiscale feature fusion of remote sensing aircraft images [J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 1-5.
- [17] ZONG Z, CAO Q, LENG B. RCNet: Reverse feature pyramid and cross-scale shift network for object detection [C]. Proceedings of the 29th ACM International Conference on Multimedia, 2021: 5637-5645.
- [18] QIAO Z, SHI D, YI X, et al. UEFPN: Unified and enhanced feature pyramid networks for small object detection [J]. ACM Transactions on Multimedia Computing, Communications and Applications, 2023, 19(2s): 1-21.
- [19] XU X, LIU Y, LYU L, et al. MAD-YOLO: A quantitative detection algorithm for dense small-scale marine benthos [J]. Ecological Informatics, 2023, 75: 102022.
- [20] WANG M, YANG W, WANG L, et al. FE-YOLOv5: Feature enhancement network based on YOLOv5 for small object detection [J]. Journal of Visual Communication and Image Representation, 2023, 90: 103752.
- [21] MA N, ZHANG X, ZHENG H T, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design [C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 116-131.
- [22] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [J]. arXiv preprint arXiv:2207.02696, 2022.
- [23] LEE Y, HWANG J, LEE S, et al. An energy and GPU-computation efficient backbone network for real-time object detection [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019.
- [24] TIAN C, XU Y, ZUO W, et al. Asymmetric CNN for image super resolution [J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2021, 52(6): 3718-3730.
- [25] TAN M, PANG R, LE Q V. EfficientDet: Scalable and efficient object detection [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781-10790.
- [26] JIANG Y, TAN Z, WANG J, et al. GiraffeDet: A heavy-neck paradigm for object detection [J]. arXiv preprint arXiv:2202.04256, 2022.
- [27] 程晓悦, 赵龙章, 胡穹, 等. 基于膨胀卷积平滑及轻型上采样的实时语义分割 [J]. 激光与光电子学进展,

- 2020, 57(2): 185-192.
- CHENG X Y, ZHAO L ZH, HU Q, et al. Real-time semantic segmentation based on inflated convolutional smoothing and lightweight upsampling [J]. Advances in Lasers and Optoelectronics, 2020, 57(2):185-192.
- [28] LIU Y, SHAO Z, HOFFMANN N. Global attention mechanism: Retain information to enhance channel-spatial interactions [J]. arXiv preprint arXiv: 2112.05561, 2021.
- [29] 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, 2016: 1874-1883.
- [30] XIA G S, BAI X, DING J, et al. DOTA: A large-scale dataset for object detection in aerial images [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 3974-3983.
- [31] YANG X, YANG J, YAN J, et al. ScrDet: Towards more robust detection for small, cluttered and rotated objects [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 8232-8241.
- [32] WANG P, SUN X, DIAO W, et al. FMSSD: Feature-merged single-shot detection for multiscale objects in large-scale remote sensing imagery [J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 58(5): 3377-3390.
- [33] XU T, SUN X, DIAO W, et al. ASSD: Feature aligned single-shot detection for multiscale objects in aerial imagery [J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1-17.
- [34] HAN J, DING J, LI J, et al. Align deep features for oriented object detection [J]. 2020, DOI: 10.48550/arXiv.2008.09397.
- [35] REDMON J, FARHAD A. YOLOv3: An incremental improvement [J]. arXiv preprint arXiv: 1804.02767, 2018.
- [36] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection [J]. arXiv preprint arXiv:2004.10934, 2020.

作者简介



单慧琳, 2003 年于武汉大学获得学士学位, 2009 年于南京信息工程大学获得硕士学位, 现为南京信息工程大学和无锡学院副教授、硕导, 主要研究方向为深度学习、图像处理等。

E-mail: shanhuilin@nuist.edu.cn

Shan Huilin received her B. Sc. degree from WuHan University in 2003, M. Sc. degree from Nanjing University of Information Science & Technology in 2009. Now she is an associate professor and master's supervisor in Nanjing University of Information Science & Technology and Wuxi University. Her main research interests include deep learning and image processing.



吕宗奎, 南京信息工程大学硕士研究生, 主要研究方向为深度学习、图像处理等。

E-mail: 1372805287@qq.com

Lyu Zongkui is a M. Sc. candidate at Nanjing University of Information Science and Technology. His main research interests include deep learning and image processing.