联合双重注意力机制和双向特征金字塔的遥感影像小目标检测

李科文[1]，朱光磊[2]，王辉[1]，祝锐[1]，狄兮尧[3,4]，张天健[3,4]，薛朝辉[3,4]

1. 国家能源集团谏壁发电厂, 镇江 212006;
2. 黄河勘测规划设计研究院有限公司, 郑州 450003;
3. 河海大学 地球科学与工程学院, 南京 211100;
4. 河海大学 江苏省水资源环境遥感监测评估工程研究中心, 南京 211100

摘 要：遥感影像由于成像特点和空间分辨率的制约，导致小目标特征提取难度增加，因此遥感影像小目标检测具有较大挑战。现有深度学习目标检测网络架构多基于自然图像而构建，在顾及遥感影像小目标特殊性方面的研究和探索存在不足。为此，本文提出了联合双重注意力机制和双向特征金字塔的遥感影像小目标检测算法。主要创新工作体现在：（1）针对遥感影像中待检测物体所占比例较小导致其特征信息不易提取以及网络参数量大等问题，基于YOLOv3网络模型，引入LKGNet主干网络和GIoU损失函数，提出LKGNet-YOLO网络；（2）针对网络特征提取过程中容易受噪声信息干扰以及网络多层信息融合能力不足的问题，基于YOLOv3网络模型，引入DA-LKGNet主干网络和加权双向特征金字塔结构，提出DA-LKGNet-YOLO网络。为了验证所提方法的在遥感影像小目标检测领域中的有效性，采用2014年中国科学院大学发布的UCAS-AOD v1.0（UA）数据集和2021年武汉大学发布的AI-TOD数据集进行实验。结果表明：本文方法在UA数据集上0.5阈值的均值平均精度$mAP_{0.5}$为96.21%，在AI-TOD上0.5—0.95阈值的均值平均精度$mAP_{0.5—0.95}$为9.51%，显著优于YOLOv3、RFBNet、SSD、FSSD、RetinaNet、RefineDet等模型，mAP精度分别高出约3.45%—7.52%、1.36%—4.84%。同时，在小尺度目标上的检测水平优于Faster-RCNN和YOLOv7算法。与原始YOLOv3网络相比，本方法在浮点运算次数上降低了48%，参数量减少了42%。上述实验结果证实了本文提出方法的有效性。

关键词：遥感影像，小目标检测，深度学习，YOLOv3，注意力，特征金字塔

中图分类号：TP701/P2

引用格式：**李科文,朱光磊,王辉,祝锐,狄兮尧,张天健,薛朝辉.2024.联合双重注意力机制和双向特征金字塔的遥感影像小目标检测.遥感学报,28(12):3231-3248**

Li K W, Zhu G L, Wang H, Zhu R, Di X Y, Zhang T J and Xue Z H. 2024. Joint dual attention mechanism and bidirectional feature pyramid for remote sensing small targets detection. National Remote Sensing Bulletin, 28(12):3231-3248[DOI:10.11834/jrs.20243043]

1 引 言

随着遥感成像技术的发展，影像中包含的目标信息越来越丰富，随之发展的遥感影像目标检测技术已经成功应用于城市交通规划、军事目标跟踪和地物目标分类等领域。由于遥感目标本身存在多尺度、小目标、分布不均匀等特点，加之不同成像条件及影像空间分辨率的影响（戴伟聪 等，2018），导致针对遥感影像小目标的特征提取与检测具有很大挑战。

近年来，卷积神经网络CNN（Convolutional Neural Network）在图像处理方面大放异彩，并且在大规模目标检测方面已经远远超过传统方法。因此，越来越多的学者将目光聚焦在基于深度学习的遥感影像小目标检测研究。根据是否生成目标候选区域，深度学习遥感影像小目标检测研究可分

收稿日期：2023-03-02；**预印本**：2023-08-16

基金项目：国家自然科学基金(编号：41971279,42271324)；江苏省自然科学基金(编号：BK20221506)

第一作者简介：李科文,研究方向为基于无人机遥感的新能源科技创新。E-mail：12030141@ceic.com

通信作者简介：薛朝辉,研究方向为深度学习等先进机器学习模型在高光谱遥感影像小样本分类、多源遥感数据迁移学习、地表参数遥感估算等方面的理论与应用研究。E-mail：zhaohui.xue@hhu.edu.cn

为：两阶段目标检测和单阶段目标检测（Li等，2020）。

两阶段目标检测算法使用区域候选网络生成候选区域，同时对待检测目标的位置进行初步预测，然后对物体目标的位置、类别进行进一步修正和判断，经过两个阶段得到最终结果。两阶段目标检测算法检测精度较高，但计算效率较低，不能满足实时检测的需求。代表性的两阶段目标检测算法有R-CNN（Girshick等，2014），SPPNet（He等，2014），Fast R-CNN（Girshick，2015），Faster R-CNN（Ren等，2017）等。其中，Faster R-CNN作为R-CNN系列的集大成者，突破性地使用了区域候选网络（RPN）直接提取出候选区域，并将其融入整体网络中，实现端到端的检测，使得综合性能有较大提高，解决了R-CNN和Fast R-CNN因候选区生成对网络目标检测速度的阻碍。但是Faster R-CNN仅使用单个输出层的特征进行分类回归预测，深层特征经历过多的下采样造成小目标特征损失，从而导致检测精度偏低。

单阶段目标检测方法是利用回归的思想来解决目标框定位问题，它并不需要预先生成候选区域，其目标位置和类别信息可以通过网络直接回归得到，因此该类方法也被称为回归型目标检测算法。YOLO算法（Redmon等，2016）是一种具有代表性的单阶段目标检测算法，它采用单个主干网络进行特征提取，在一次评估中回归出图像中待检测目标的边界框和类别概率，但YOLO在小尺寸物体上的检测效果不理想。随后出现的SSD网络（Liu等，2016）、YOLOv2（Redmon和Farhadi，2017）、YOLOv3（Redmon和Farhadi，2018）不断改进了YOLO原有的主干网络，提出计算复杂度更低、准确率更高的网络，通过引入Anchor机制与多层特征融合的思想，提高了模型召回率，降低了漏检率，在一定程度上提升了小目标的检测性能。

现有深度学习目标检测算法大多基于自然图像构建，而遥感影像目标多为小目标。随着卷积神经网络的多重下采样，小目标很容易在深层特征中消失，且遥感影像中背景特征复杂，极易造成漏检。

小目标检测一直是遥感目标检测算法中的研究难点，目前已开展了许多具有借鉴意义的研究工作（聂光涛和黄华，2021）。在网络结构的设计与优化改进方面，Li等（2018）采取反卷积层融合不同层级的特征，进一步增强小目标的检测能力；Wang等（2020）在结合浅层信息后，改进损失函数来增加小目标的训练权重；Zhang等（2018b）在Faster RCNN的基础上，基于RPN网络提议的候选区域，通过反卷积进行上采样以放大特征图尺寸。然而，只有在深层特征中依然存在小目标的前提下，上采样操作才有意义。为此，Dong等（2019）提出一种改进的非最大抑制区域预测模块，从而减少小目标丢失问题。

在多尺度特征融合方面，Wang等（2021a）通过多特征级联决策来逐步消除遥感小目标检测的虚警率和漏检率，但随着特征图尺度的增加，不可避免地增加了计算的复杂度与检测过程的时间消耗。为了提高遥感图像中多尺度飞机目标的检测精度，沙苗苗等（2022）提出一种基于改进Faster R-CNN的遥感图像飞机目标检测方法。为了解决地物样本的尺度不均衡问题，秦登达等（2022）设计了数据集内影像加权融合与地物多尺度特征选择的目标检测策略。

在引入注意力机制方面，周秦汉和王振（2022）使用注意力机制来融合多尺度特征，解决传统卷积神经网络无法获取小尺度目标特征的问题。考虑到场景信息和地物目标之间的关联关系，张菁等（2022）提出全局关系注意力引导场景约束的高分辨率遥感影像目标检测方法。针对主干网络特征提取有效信息量少、特征图信息表征能力弱等问题，张寅等（2022）构造特征增强模块，并融合注意力机制更精确地捕获小目标特征。

小目标检测在遥感影像中面临着一系列的挑战，包括目标场景的差异化、背景特征的复杂性，以及大量漏检、误检等问题。虽然上述算法能够在一定程度上提高小目标检测水平，但同时也引入了计算复杂度增加、检测速度降低、语义信息偏差等新的问题。

首先，小目标检测算法在深度信息挖掘方面仍然存在问题。由于小目标通常具有较小的尺寸和较低的像素值，其特征信息较为有限。因此，如何充分挖掘小目标的深度信息，提高其特征表达能力，仍然是一个亟待解决的问题。

其次，背景干扰是小目标检测中的另一个挑

战。遥感影像中背景复杂多样，包含丰富的地物信息，可能会对小目标的检测造成干扰。目前的方法虽然尝试通过多尺度特征融合等策略来解决这一问题，但仍然存在语义信息偏差的情况，即将背景信息错误地作为目标信息，导致误检或漏检。

此外，虽然多尺度特征融合等策略可以提高目标检测的精度，但也会增加计算复杂度，导致检测速度降低，不利于实时应用和大规模数据处理。

本文针对目前深度学习算法对遥感影像小目标检测存在特征提取能力弱、特征信息融合不强、小目标信息提取难，模型计算复杂度高等问题，从主干特征提取网络构建、注意力机制的引入以及特征信息融合等角度出发，提出了联合双重注意力机制和双向特征金字塔的遥感影像小目标检测方法：基于Ghost module和跨阶段网络，提出了LKGNet主干特征提取网络，进一步融入双重注意力和加权双向特征金字塔模块，构建了轻量化的DA-LKGNet-YOLO目标检测模型，在浮点运算次数和参数量上远低于基础模型。

2 数据资料

本文使用中国科学院大学在2014年发布的UCAS-AOD v1.0（UA）数据集（Zhu等，2015）和武汉大学在2021年发布的AI-TOD数据集（Wang等，2021b）检验提出算法的性能。UA数据集包含汽车和飞机两类目标，共910张图像，其中汽车310张，共计4475个目标物体。该数据集中目标的空间分辨率低，且由于目标所处的环境不同，可以分为复杂背景环境和简单背景环境，适合用于验证提出模型的小目标检测能力。部分实例影像如图1所示。从图1可以看出，在简单背景下，影像中基本只存在检测目标与地面，且检测目标与周围环境有较大的区别；而在复杂背景下，影像中存在的地物类型较多，检测目标自身特征不明显，检测难度较大。

图1 UA目标实例

Fig. 1 Image targets of UA

AI-TOD数据集包含28036张影像，700621个目标，且分为8类目标，分别是飞机、桥、存储罐、船只、游泳池、车辆、人、风车。该数据集中最大目标的尺寸不超过64像素，远小于其余目标检测数据集中的物体尺寸，适合进行小目标检测研究。该数据集的部分实例影像如图2所示。从图2可以看出，相对背景而言，检测目标的尺寸极小，特征信息易丢失，具有一定的检测挑战性。

3 联合双重注意力机制和双向特征金字塔的遥感影像小目标检测研究

3.1 基于GhostNet和CSPNet的轻量化主干网络设计

YOLOv3使用了残差连接构建深层的特征提取网络，并结合特征金字塔网络（FPN）来实现多尺度检测，以在检测精度和速度之间取得平衡。如

图3所示，YOLOv3的特征提取器采用了基于1×1和3×3卷积的Darknet53残差神经网络。Darknet53将卷积层（Conv）、批归一化层（BN）和激活函数（Leaky relu）组成初始采样模块（CBL），并通过特征融合手段（add）形成残差单元（Res unit）。多组残差单元与一组采样模块的线性组合形成了特征采样模块（ResX）。这种方式使得网络可以通过残差连接跨越较大的深度，从而充分利用深层特征进行更准确的目标检测。通过采用残差连接和特征金字塔网络（FPN），YOLOv3可以同时处理不同尺度的特征图，从而能够检测不同大小的目标。然而，深层的特征提取器会导致算法的参数过多，增加了网络模型对数据量的需求，也拖慢了检测速度；其次，遥感影像中存在大量几十个像素的小目标，常规特征提取网络的感受野较小，对于遥感小目标特征和其相似背景无法做出有效区分，导致遥感目标检测算法漏检率和误检率偏高。

图2 AI-TOD目标实例

Fig. 2 Image targets of AI-TOD

为了实现对遥感影像小目标的检测，本文借鉴GhostNet（Han等，2020）网络的轻量化设计，以降低漏检率并减少参数量为出发点，提出一种大感受野、较少参数量、较低运算复杂度的特征提取主干网络，称为LKGNet。如图4所示，该主干网络由一系列基于大卷积核的跨阶段网络模块（LKGCSP）和过渡卷积层组成。该设计方法借鉴了CSPNet的梯度组合思想，增大Ghost Bottleneck中的卷积核，再将其嵌入到CSP结构中，构建LKGCSP结构。在减少由大卷积核引起的网络参数量的同时，进一步提升了模型的学习能力。

3.1.1 LKG Bottleneck模块

LKG Bottleneck模块通过增大感受野，提升小目标检测模型性能。在第一层使用LKG module对特征进行提取并将低维特征映射到高维空间。

LKG module将Ghost module中1×1卷积和3×3卷积替换成两个7×7卷积，大幅增加整个模块的感受野；之后利用1×1卷积再次进行特征提取与通道映射，旨在使用少量的参数生成特征。

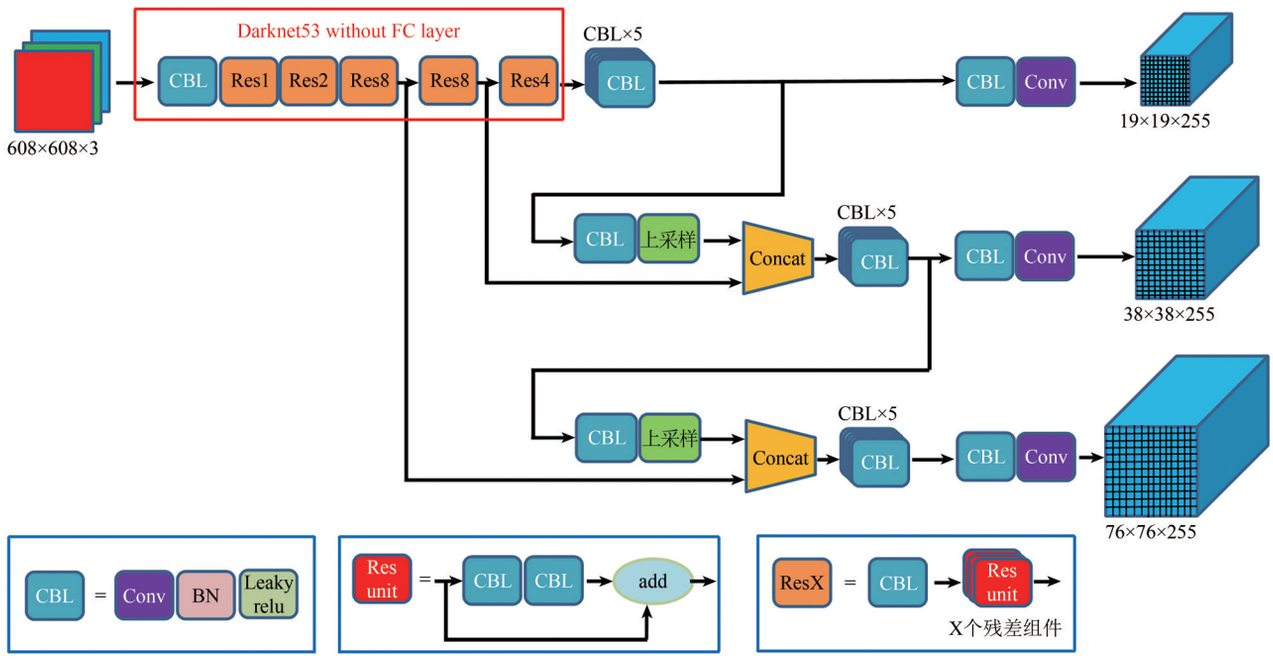


图3 YOLOv3网络结构

Fig. 3 The structure of YOLOv3

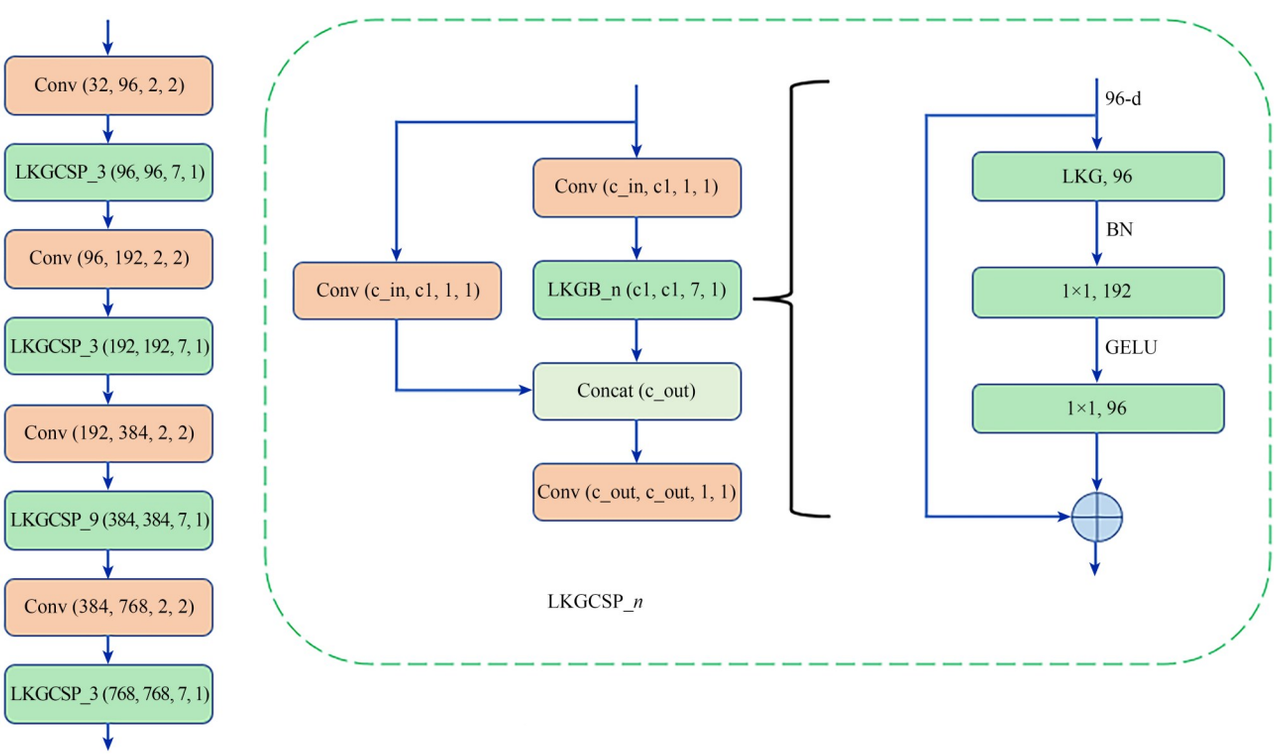


图4 LKGNet网络结构

Fig. 4 The structure of LKGNet

对于给定输入数据 $\boldsymbol{X} \in \mathbb{R}^{c \times h \times w}$，$\tilde{\boldsymbol{X}} \in \mathbb{R}^{c/2 \times h \times w}$，LKG Bottleneck结构的运算过程如下：

$$C_1 = B(conv_1^2(conv_1^1(\tilde{X}, W_1^1), W_1^2) + \widetilde{X}) \quad (1)$$

$$C_2 = g(Conv_1(C_1, W_2^1)) \quad (2)$$

$$C_3 = Conv_2(C_2, W_3^1) \quad (3)$$

$$out = C_3 + X \quad (4)$$

式中，$conv_i^j$为第i层第j次卷积，W_i^j为第i层第j次卷积权重，C_i为第i层卷积后的输出结果，B为Batch Normalization操作，g为gelu激活函数。

3.1.2 跨阶段局部网络

为解决传统网络结构计算量大的问题，跨阶段局部网络（CSPNet）从网络结构设计的角度做出优化，设计了两条梯度信息传递路径。当一个特征进入到CSP结构中会在通道维度上一分为二，在主干通道采用多个卷积层串行连接的方式对输入信息进行前向传播，然后经过转化层进行进一步处理；另一部分信息则是流经跨阶段结构直接与转化层处理后的特征进行梯度信息合并。

CSP结构增加了信息流通路径，不仅减少了网络中的参数量，而且使梯度信息差异化更加明显，丰富了网络中的梯度组合。CSP结构有着很好的适用范围，将LKG Bottleneck嵌入到CSP结构中，构建LKGCSP结构。将重新设计的LKGCSP结构结合过度卷积层作为网络的特征提取模块，最终构成本文目标检测模块的主干网络LKGNet。通常特征提取网络中的空间下采样由步长为2的卷积来实现，在本研究中采用了卷积核为2×2、stride为2的卷积层来实现下采样。在LKGNet中不同阶段的CSP结构中，设置不同的LKG Bottleneck数量。

3.2 双重注意力模块

由于遥感影像目标的背景复杂，所提取的特征在不同通道所包含的特征信息量差异较大，同时在同一通道内不同的位置的特征重要程度也不同。为了更有效的在复杂背景中捕捉遥感小目标的语义信息，本文采用了一种针对通道空间特征的双重注意力模块DA（Dual Attention）（Fu等，2019），该模块包含两个部分：通道注意力模块和空间注意力模块。

通过通道注意力模块可以建立不同通道之间的关系，捕获全局通道信息。如式（5）所示，首先对特征空间维度进行全局池化挤压，将输入特征转化为通道向量。

$$z_c = \boldsymbol{F}_{\mathrm{sq}}(u_c) = \frac{1}{W \times H}\sum_{i=1}^{W}\sum_{j=1}^{H} u_c(i,j) \tag{5}$$

式中，通道向量z_c表示输出的通道向量结果，u_c表示输入特征图的第c个通道，W、H表示输入特征图的宽、高，i、j表示通道中第i行、第j列的像素值。$\frac{1}{W \times H}$是一个标准化项，用于将每个通道的全局信息缩放到相同的尺度。

然后，对特征进行激励，如式（6）所示。其中，F_{ex}代表的是通道激励函数，其作用是计算每个特征通道的注意力权重，以便将注意力集中在对当前任务最有效的通道上。z_c表示式（5）得到的全局池化向量，s表示激励操作后学习到的通道权重输出。W表示全连接操作，W_1和W_2表示两个全连接层的权重矩阵，能够捕获非线性的跨通道交互作用，降低中间层维度以避免过高的模型复杂度。最终将融合全部输入特征信息并恢复输入通道数，保证输入与输出的维度相同。δ表示relu激活函数，σ表示Sigmoid函数，得到每个特征通道的权重。

$$s = F_{\mathrm{ex}}(z_c, W) = \sigma\left(g(z_c, W)\right) = \sigma\left(W_2\delta(W_1 z_c)\right) \tag{6}$$

最后，将得到的通道重要程度权重赋予输入特征。如式（7）所示，F_{scale}代表的是通道缩放函数，它的作用是将计算得到的通道注意力权重s赋予输入特征图u_c，通过动态加权的方式增强重要特征通道中的信息，并抑制一些对当前任务无用的特征信息。

$$X = F_{\mathrm{scale}}(u_c, s) = s \cdot u_c \tag{7}$$

空间注意力模块能够计算输入特征的空间特性，利用空间注意权重关注重要信息同时弱化次要信息，它与通道注意互补。空间注意力如式（8）所示，将通道注意力的输出特征X作为输入，在通道层面做全局最大池化和全局平均池化，得到两个$H \times W \times 1$的特征图，然后将得到的特征图拼接后输入卷积操作使其通道数降为1，即$H \times W \times 1$，再经过Sigmoid生成空间注意力权重$\boldsymbol{V}_s$。

$$\begin{aligned}\boldsymbol{V}_s(X) &= \sigma\left(f^{7\times7}\left((\mathrm{Avg\,Pool}(X);\ \mathrm{MaxPool}(X))\right)\right) = \\ &\sigma\left(f^{7\times7}\left(\left[X^s_{\mathrm{avg}};\ X^s_{\mathrm{max}}\right]\right)\right)\end{aligned} \tag{8}$$

根据上述操作，将DA注意力模块嵌入本文主干网络提出的特征提取LKGCSP模块，构建DALKGCSP（Dual Attention LKGCSP）模块，其结构如图5所示，其中c_in是输入特征的通道数，$c1$是输出特征的通道数。

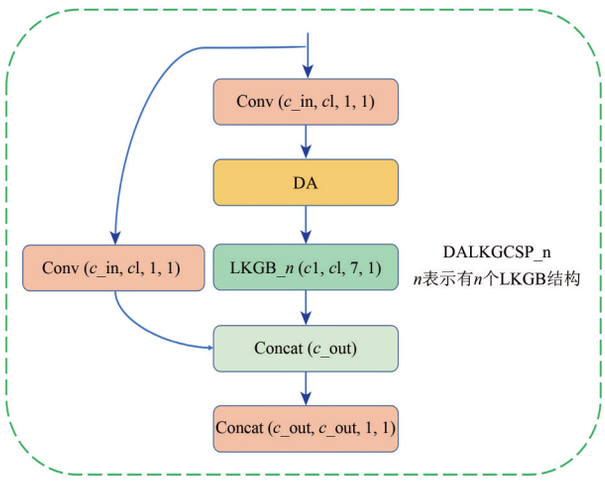


图5 本文提出的DALKGCSP网络结构

Fig. 5 The structure of the proposed DALKGCSP

3.3 融合加权双向特征金字塔网络的检测模型

在YOLOv3算法的特征融合过程中引入特征图金字塔结构FPN（Lin等，2017），构建了一条自上而下的信息流通路径，实现多层特征的融合。但FPN缺少更多方向的信息流动，同时特征在流通路径的传播过程中因为卷积的存在而被稀释。

为解决YOLOv3使用FPN对多层次特征融合能力不够强的问题，本文引入加权双向特征金字塔网络（BiFPN）（Tan等，2020）。BiFPN借鉴了FPN和PANet（Liu等，2018a）的思想，它设计了一个可以放缩的特征融合结构，不仅有向上和向下的流通路径，同时增加了同层次间的信息流通路径以加强特征信息的融合。

本文采用BiFPN改进YOLOv3原有的FPN结构，构建遥感影像小目标检测模型DA-LKGNet-YOLO，其具体结构如图6所示，主要包括3个子模块：本文提出的DA-LKGNet骨干网络，进行多路径特征信息融合的加权双向特征金字塔网络以及用于提取目标边界框的检测器模块。

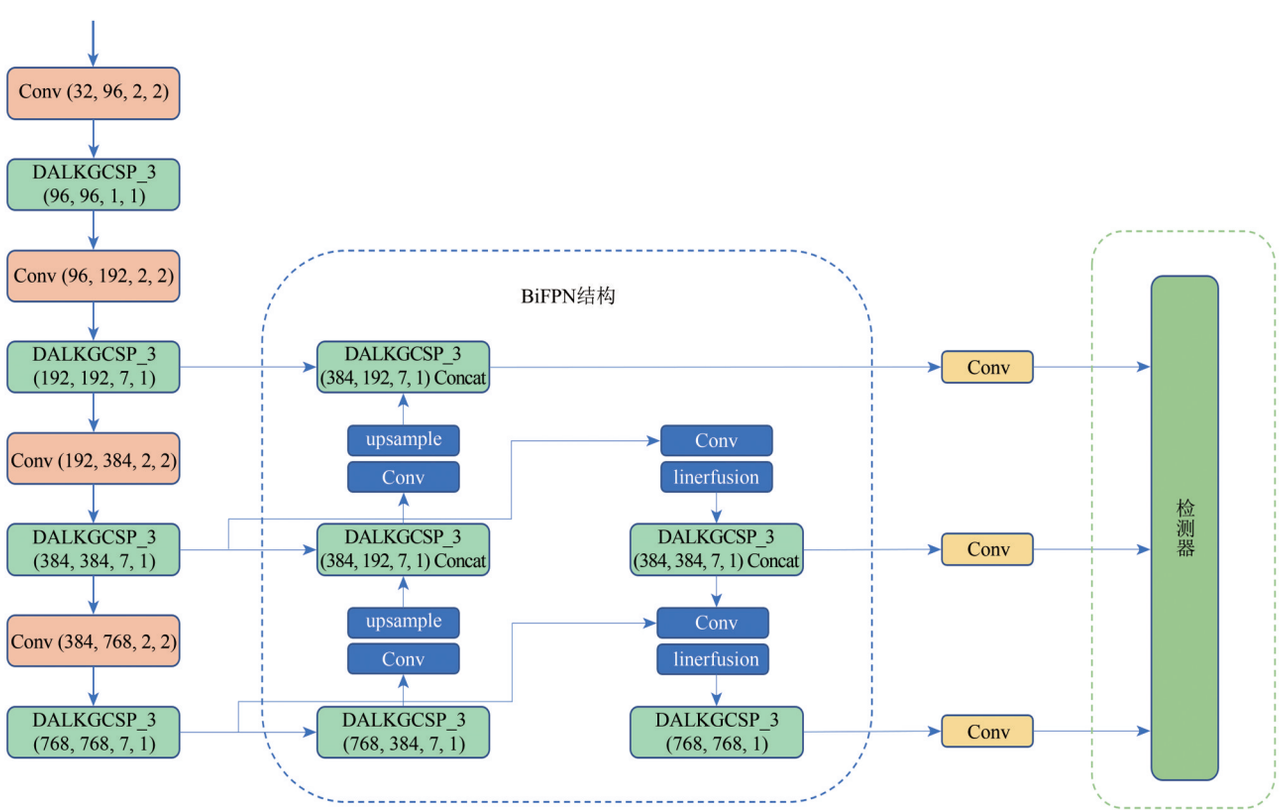


图6 本文提出的DA-LKGNet-YOLO网络结构

Fig. 6 The structure of the proposed DA-LKGNet-YOLO

3.4 广义交并比损失函数

损失函数用来衡量模型预测值和真实值之间的偏差，对于评判模型的性能至关重要。YOLOv3的损失函数由定位损失、置信度损失和分类损失3部分组成，如式（9）所示。其中，定位损失所使用的方法是差值平方，置信度损失和分类损失

使用的是二值交叉熵函数。

$$\begin{aligned}\text{Loss} = & \lambda_1 \sum_{i=0}^{s^2}\sum_{j=0}^{B} I_{ij}^{\text{obj}}\left(\left(\left(x_i-\hat{x}_i\right)^2+\left(y_i-\hat{y}_i\right)^2\right)+\right.\\ & \left.\left(\left(\sqrt{w_i}-\sqrt{\hat{w}_i}\right)^2-\left(\sqrt{h_i}-\sqrt{\hat{h}_i}\right)^2\right)\right)-\\ & \sum_{i=0}^{s^2}\sum_{j=0}^{B} I_{ij}^{\text{obj}}\left(\hat{C}_i^j\log\left(C_i^j\right)+\hat{D}_i^j\log\hat{D}_i^j\right)^2-\\ & \lambda_2\sum_{i=0}^{s^2}\sum_{j=0}^{B} I_{ij}^{2}\left(\hat{C}_i^j\log\left(C_i^j\right)+\hat{D}_i^j\log\hat{D}_i^j\right)-\\ & \sum_{i=0}^{s^2} I_i^{\text{obj}}\sum_{c\in\text{class}}\left(\hat{P}_i^j\log\left(P_i^j\right)+\hat{Q}_i^j\log\hat{Q}_i^j\right)\end{aligned} \tag{9}$$

式中，把一张图片分成$s\times s$个网格，每个网格产生B个锚框；x_i、y_i、w_i、h_i为预测坐标和宽高；$\hat{x}_i$、$\hat{y}_i$、$\hat{w}_i$、$\hat{h}_i$为真实坐标和宽高；λ_1和λ_2表示权重值，I_{ij}^{obj}表示第i个网格中第j个anchor box是否预测到目标，如果是则该值为1，否则为0；λ_2表示第i个网格中的第j个anchor box；$\hat{C}_i^j$表示anchor box是否预测到某个目标，如果是则为1，否则为0，且设$\hat{D}_i^j=1-\hat{C}_i^j$；$P_i^j$是预测每个目标的概率，且设$\hat{Q}_i^j=1-P_i^j$。在YOLOV3中，衡量检测的好坏以及边框损失使用的是交并比IoU（Intersection over Union），其定义如式（10）所示，其中A表示的是预测框，B表示的是真实框。

$$\text{IoU}=\frac{|A\cap B|}{|A\cup B|} \tag{10}$$

IoU损失函数可表示为LIoU=1−IoU，相对于L2损失有两个优点。首先，它能够更好地反应重合程度；其次，它具有尺度不变性，也就是说无论两个重叠的矩形框是大还是小，它的重合程度与矩形框的尺度是无关的。当然这种IoU损失也有一个缺点，即当两个矩形框不相交的时候，此时损失值为0，就无法对参数进行优化。因此，基于以上缺点，将IoU替换为GIoU。GIoU的原理如图7所示，计算公式如式（11）所示。

$$\text{GIoU}=\frac{|A\cap B|}{|A\cup B|}-\frac{|C-(A\cup B)|}{|C|} \tag{11}$$

式中，A和B为两组边框，C为A框和B框的最小外接矩形，$|C-(A\cup B)|$表示在C中不包括A和B的区域面积。GIoU表示原始的IoU减去C中不包括A和B区域的面积再比上C的面积。IoU的取值范围是［0，1］，如果在目标检测中使用IoU，当IoU的值为0时，检测过程中的梯度值也会为0；而GIoU的取值范围是[−1，0)U(0，1]，不会出现梯度值为0的情况；同时，C的引入使非重合区域也得到了网络的关注，使得预测框相较于真实框的位置偏差更加敏感。

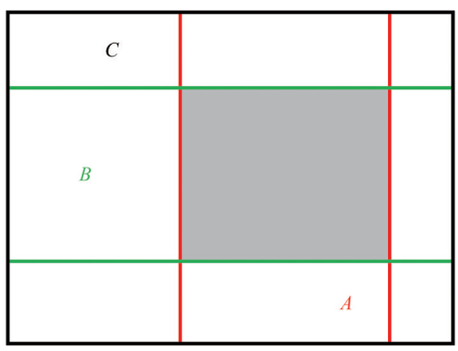


图7 GIoU示意图

Fig. 7 The schematic diagram of GIoU

4 结果与分析

4.1 实验环境与设置

开发平台的CPU为Intel（R） Core（TM）i7−8700，GPU为NVIDIA GeForce RTX3080Ti，操作系统为Ubuntu 16.04LTS，深度学习框架为Pytorch 1.10。采用SGD优化器，批大小设置为18，训练轮次设置为200，初始学习率设置为0.01，权值衰减设置为0.001并在100、120、140、160、180轮时进行衰减。而原始图像缩放至416×416像素后输入进网络中。同时，本实验采用了几何变换、颜色变换与Mosaic数据增强方式。在几何变换中使用缩放（系数为0.5）和左右翻转（概率为0.5）；在颜色变换中使用随机色相、饱和度、色明度，增强系数分别设置为0.015，0.7，0.4；Mosaic增强中以概率0.5随机选取4张图像进行自适应地拼接。

4.2 评价指标

本文采用平均精度均值MAP（Mean Average Precision），准确率（Precision）和召回率（Recall）作为评价指标。其中，根据由准确率和召回率绘制的PR曲线计算AP值，进而mAP由所有类的AP取均值计算。准确率和召回率的定义如式（12）、式（13）所示：

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}} \tag{12}$$

$$\text{Recall}=\frac{\text{TP}}{\text{TP}+\text{FN}} \tag{13}$$

式中，TP代表真正例，FN代表假反例，FP代表假

正例。对于UA数据集，我们采用IoU阈值为0.5的mAP来评估所提出方法的性能。同时，为了更全面地评估本文提出的方法在小目标上的检测能力，我们在AI-TOD数据集上特别采用IoU阈值0.5—0.95的mAP进行性能评判。

4.3 参数敏感性分析

为了测试不同的卷积核超参数对本文所提出主干特征提取网络的影响，使网络获得合适的感受野，在YOLOv3融合LKGNet模块的网络上分别设置卷积核大小为{3，5，7，9}进行实验，结果如表1、表2所示。最终将卷积核大小确定为7，使网络的最终效果达到最优。

表1 卷积核大小的影响(UA)

Table 1 The impact of convolution kernel sizes (UA)

/%

方法	mAP	汽车	飞机
Ghost module	92.60	87.0	98.20
3×3	92.65	87.75	97.55
5×5	93.15	88.22	98.08
7×7	93.75	89.12	98.38
9×9	93.34	88.77	97.91

表2 卷积核大小的影响(AI-TOD)

Table 2 The impact of convolution kernel sizes (AI-TOD)

/%

方法	mAP	AI	BR	ST	SH	SP	VE	PE	WM
Ghost module	7.13	13.71	3.18	4.77	12.01	6.69	12.31	3.15	1.21
3×3	6.95	13.98	3.15	4.58	12.29	5.62	11.52	3.19	1.32
5×5	7.25	13.50	3.55	4.80	12.54	5.96	12.73	3.50	1.45
7×7	7.42	14.01	3.63	4.81	12.84	6.33	12.56	3.75	1.49
9×9	7.35	14.23	3.51	4.77	12.39	6.62	12.02	3.77	1.51

4.4 不同方法的对比分析

进一步，将本文提出的DA-LKGNet-YOLO算法与Faster-RCNN、RetinaNet（Lin等，2020）、RFBNet（Deng等，2017）、RefineDet（Zhang等，2018a）、SSD、FSSD（Li等，2017）、YOLOv3以及YOLOv7（Wang等，2023）进行了对比。我们将上述对比方法的实验设置与相关论文发表的参数保持一致。

各方法的检测结果如表3、表4所示。本文提出的DA-LKGNet-YOLO算法在UA数据集上相较YOLOv3提升了5.31%。与其他先进的目标检测算法相比取得了次优的检测结果。其中对于汽车的目标检测，DA-LKGNet-YOLO与YOLOv7算法仅有0.37%的差距。这是由于该方法使用了更加关注多层特征融合的BiFPN，同时在主干网络中加入了双重注意力模块，增强了有用信息，抑制了目标周围背景噪声，从而使得汽车类别小目标能够更好的被得到检测。

如图8所示，YOLOv3算法存在明显的误检情况，同时被正确检测的目标的置信度明显低于本文提出的方法。例如图8（a）图片中，被误检的汽车目标与周边背景颜色相近。同时图8（a）的飞机类图像中，yolov3错将与飞机纹理相似的地表识别为飞机。这是由于目标在复杂环境下出现颜色粘连，叠加或遮挡等问题时，其自身的纹理特征丧失导致。在主干网络加入注意力机制可以将网络学习特征的空间位置进行评估，抑制周围背景噪声特征对网络的影响；而通道注意力能够增强网络对含有重要特征的通道的有效利用，能有效提升模型的检测能力。如图8（b）的汽车类和飞机类图像检测结果所示，DA-LKGNet-YOLO能够有效避免误检的问题。

表3 不同模型的精度对比(UA)

Table 3 A comparison of different models (UA)

/%

方法	Backbone	mAP	汽车	飞机
Faster-RCNN	ResNet50	91.03	87.15	94.92
RetinaNet	ResNet50	92.76	90.34	95.19
RFBNet	VGG	91.59	85.81	97.37
RefineDet	VGG	88.80	78.92	98.68
SSD	VGG	88.69	84.99	92.39
FSSD	VGG	92.95	87.7	98.2
YOLOv3	Darknet53	90.90	83.00	98.81
YOLOv7	—	96.77	93.88	99.67
DA-LKGNet-YOLO(本文)	DA-LKGNet	96.21	93.51	98.91

表4 不同模型的精度对比(AI-TOD)

Table 4 A comparison of different models (AI-TOD)

/%

方法	Backbone	mAP	AI	BR	ST	SH	SP	VE	PE	WM
Faster-RCNN	ResNet50	11.15	20.66	3.85	20.18	19.00	8.9	11.86	4.49	0.30
RetinaNet	ResNet50	4.67	0.01	6.60	1.83	20.85	0.06	5.66	1.77	0.53
RFBNet	VGG	7.65	14.35	0.92	18.21	13.42	0.02	10.94	3.38	0.01
RefineDet	VGG	7.48	10.77	1.76	12.28	15.65	3.20	11.32	3.98	0.94
SSD	VGG	6.98	14.50	3.11	10.89	13.05	1.90	7.84	3.12	1.44
FSSD	VGG	8.15	15.21	3.25	7.79	18.11	3.09	12.42	4.32	1.01
YOLOv3	Darknet53	6.01	10.58	3.80	4.88	11.66	3.92	9.53	3.31	0.40
YOLOv7	—	11.7	22.11	8.21	18.37	20.64	6.23	12.94	4.98	0.17
DA-LKGNet-YOLO (本文)	DA-LKGNet	9.51	16.11	6.42	7.97	15.32	7.01	14.17	6.81	2.27

(a) YOLOv3 (b) DA-LKGNet-YOLO

图8 本文方法与基准方法的检测结果对比(UA)

Fig. 8 A comparison of detection results between the proposed method and the base method (UA)

表4统计了各方法在AI-TOD数据集上的检测精度。本文提出的方法相较YOLOv3提升3.5%，但精度低于Faster-RCNN和YOLOv7。虽然Faster-RCNN和YOLOv7在飞机（AI）、桥梁（BR）、存储罐（ST）、船（SH）、游泳池（SP）这些相对的大目标类别上检测效果最好，但是本文提出的方法在车辆（VE）、人（PE）、风车（WM）这3个尺寸相对较小的目标类别上的检测效果最好，这验证了本文的方法在小目标识别上的有效性。

如图9所示，在AI-TOD数据集上，YOLOv3算法存在明显的漏检与误检情况，同时被正确检测的目标的置信度明显低于本文提出的方法。例如图9（a）中大量船只被漏检，同时车辆类别的图片中，检测到的车辆的置信度较低，且将路边背景误检为车辆。这是由于过小的目标尺寸导致YOLOv3无法精确抓捕目标物的特征。而在本文提出的方法中，通过融合注意力机制的方式对深层目标特征和背景特征进行针对性分离，着重关注目标物的敏感通道特征，从而优化了YOLOv3中的缺陷。对于图9（a）中的问题，DA-LKGNet-YOLO能够有效解决，并在图9（b）中得到体现。

4.5 不同方法的复杂度分析

为了评估本算法的复杂度，我们将上述对比方法的批大小设置为1，计算各模型处理单张影像的浮点运算次数（GFLOPs）、参数量（Parameters）、预测毫秒时间。从表5中的结果看出，相较于其他方法，DA-LKGNet-YOLO的浮点运算次数处于中游水平，参数量和单张影像预测时间均达到了较高水平。特别地，相较于YOLOv3模型，DA-LKGNet-YOLO的浮点运算次数降低了48%，参数量减少了42%，大大降低了模型的复杂度。

（a）YOLOv3 （b）DA-LKGNet-YOLO

图9 本文方法与基准方法的检测结果对比（AI-TOD）

Fig. 9 A comparison of detection results between the proposed method and the base method (AI-TOD)

表5 复杂度分析

Table 5 Complexity analysis

方法	GFLOPs	参数	时间/ms
Faster-RCNN	46.84	43781193	142
RetinaNet	26.01	36371327	73
RFBNet	74.56	44335262	28
RefineDet	98.31	35685692	40
SSD	67.37	36821726	94
FSSD	118.25	42441822	41
YOLOv3	155.3	61529119	15
YOLOv7	106.5	37622682	12
DA-LKGNet-YOLO（本文）	79.8	35802241	20

5 讨 论

5.1 消融实验

为了验证本文提出方法的性能，在基准网络YOLOv3的基础之上，分别增加广义交并比损失函数、主干网络、注意力模块和加权双向特征金字塔模块进行消融实验。同时为了验证其针对小目标的检测的有效性，我们将两组数据集按不同尺寸大小进行分离，分别统计本文的方法在不同尺度数据上的检测水平。对于UA数据集，我们将所有目标按照小于32^2、32^2—96^2和大于96^2像素检测框像素面积划分为小目标、中目标和大目标，并统计对应的mAP_s，mAP_m，mAP_l。对于AI-TOD数据集，我们将所有目标按照2^2—8^2、8^2—16^2、16^2—32^2和大于32^2像素的检测框像素面积划分为极小目标、小目标、中目标和大目标，并统计对应的mAP_t，mAP_s，mAP_m，mAP_l。

5.1.1 主干网络与广义交并比损失函数

为验证广义交并比损失函数的有效性，首先在基准网络YOLOv3的基础之上使用广义交并比损

失函数，原有网络结构保持不变，两组数据集的整体实验结果如表6、表7所示，其中base代表基准网络YOLOv3。将原始损失函数改进为广义交并比损失函数（GIoU）后，对于UA数据集，平均精度mAP为92.31%，提高了1.41%。对于AI-TOD数据集，mAP为7.04%，提高了1.03%。这证明GIoU损失函数能够更好地感知目标框的位置。

将基准网络YOLOv3的主干网络Darknet-53替换为本文所提出的LKGNet后，检测精度在UA数据集上提升2.85%，在AI-TOD数据集上提升1.41%。这证明较大的感受野有助于提升网络对于小目标的检测性能。

表6 LKGNet消融实验

Table 6 Ablation analysis of LKGNet (UA)

/%

方法	mAP	Car	Airplane
base	90.90	83.00	98.81
base+LKGNet	93.75	89.12	98.38
base+GIoU	92.31	87.80	96.82
base+LKGNet+GIoU	94.17	89.82	98.51

表7 LKGNet消融实验(AI-TOD)

Table 7 Ablation analysis of LKGNet (AI-TOD)

/%

方法	mAP	AI	BR	ST	SH	SP	VE	PE	WM
base	6.01	10.58	3.80	4.88	11.66	3.92	9.53	3.31	0.40
base+LKGNet	7.42	14.01	3.63	4.81	12.84	6.33	12.56	3.75	1.49
base+GIoU	7.04	13.20	4.55	5.58	11.29	4.62	11.02	5.10	0.98
base+LKGNet+GIoU	7.95	14.21	4.92	4.99	13.27	6.63	12.36	5.67	1.61

表8、表9统计了该模块对不同尺寸目标的检测水平。在UA数据集上，3种尺度的目标检测结果最终分别提高了2.4%、2.6%、3%。在AI-TOD数据集上，4种尺度的目标检测结果分别提高了0.7%、2.1%、1.7%和2.9%。这证明了本模块对小目标检测性能有一定提升。

表8 LKGNet多尺度实验(UA)

Table 8 Multiscale analysis of LKGNet (UA)

/%

方法	mAP_s	mAP_m	mAP_l
base	89.7	93.1	94.1
base+LKGNet	90.4	94.6	96.4
base+GIoU	91.5	94.1	95.1
base+LKGNet+GIoU	92.1	95.7	97.1

表9 LKGNet多尺度实验(AI-TOD)

Table 9 Multiscale analysis of LKGNet (AI-TOD)

/%

方法	mAP_t	mAP_s	mAP_m	mAP_l
base	2.2	6.1	8.4	9.7
base+DA	2.6	7.9	8.9	11.5
base+BiFPN	2.4	7.2	9.7	11.9
DA-LKGNet-YOLO	2.9	8.2	10.1	12.6

5.1.2 双重注意力机制与加权双向特征金字塔网络模块

为了验证所改进的双重注意力机制与加权双向特征金字塔网络模块的有效性，本文在改进主干网络的YOLOv3的基础上进行消融实验和小目标实验。其中base代表使用LKGNet主干网络和GIoU的YOLO算法，base+DA为添加双重注意力机制的网络，base+BiFPN表示改进为加权双向特征金字塔结构的网络。

如表10所示，在UA数据集上，DA模块对于数据中汽车类别的提升较为明显，而整体mAP提升0.93%，将模型原有的特征融合模块FPN替换为BiFPN模块时，网络性能指标mAP提升了1.45%。在融合DA模块和BiFPN模块后，整体mAP提高2.04%，各类别也有一定的检测精度提升。表11统计了DA模块和BiFPN模块在AI-TOD数据集上的检测效果。在融合两组模块后，本方法的检测精度得到提升，同时在汽车、人群、风车3种小目标类别上分别提升了1.81%、1.14%和0.66%。上述实验结果表明注意力机制在特征提取过程中能更有效的在复杂背景中捕捉遥感影像小目标的语义信息。

表10 DA与BiFPN消融实验(UA)

Table 10 Ablation analysis of DA and BiFPN (UA)

/%

方法	mAP	Car	Airplane
base	94.17	89.82	98.51
base+DA	95.10	91.92	98.28
base+BiFPN	95.62	92.54	98.70
DA-LKGNet-YOLO	96.21	93.51	98.91

表12、表13统计了该模块对不同尺寸目标的检测水平。在UA数据集上，3种尺度的目标检测结果最终分别提高了2.7%、1.4%、2.1%。在AI-TOD数据集上，4种尺度的目标检测结果分别提高了0.5%、1.7%、2.3%和2.8%。这证明了本模块对小目标检测性能有一定提升。

表11 DA与BiFPN消融实验(AI-TOD)

Table 11 Ablation analysis of DA and BiFPN (AI-TOD)

/%

方法	mAP	AI	BR	ST	SH	SP	VE	PE	WM
base	7.95	14.21	4.92	4.99	13.27	6.63	12.36	5.67	1.61
base+DA	8.95	15.33	6.76	5.56	15.95	5.94	14.01	6.32	1.77
base+BiFPN	9.23	16.09	6.01	7.22	13.92	7.99	13.93	6.76	1.98
DA-LKGNet-YOLO	9.51	16.11	6.42	7.97	15.32	7.01	14.17	6.81	2.27

表12 DA与BiFPN消融实验(UA)

Table 12 Multiscale analysis of DA and BiFPN (UA)

/%

方法	mAP_s	mAP_m	mAP_l
base	92.1	95.7	97.1
base+DA	92.8	96.6	97.1
base+BiFPN	93.9	96.2	98.7
DA-LKGNet-YOLO	94.8	97.1	99.2

表13 DA与BiFPN消融实验(AI-TOD)

Table 13 Multiscale analysis of DA and BiFPN (AI-TOD)

/%

方法	mAP_t	mAP_s	mAP_m	mAP_l
base	2.9	8.2	10.1	12.6
base+DA	3.1	9.5	11.9	14.1
base+BiFPN	3.2	9.8	11.2	14.7
DA-LKGNet-YOLO	3.4	9.9	12.4	15.4

5.2 训练损失与时间

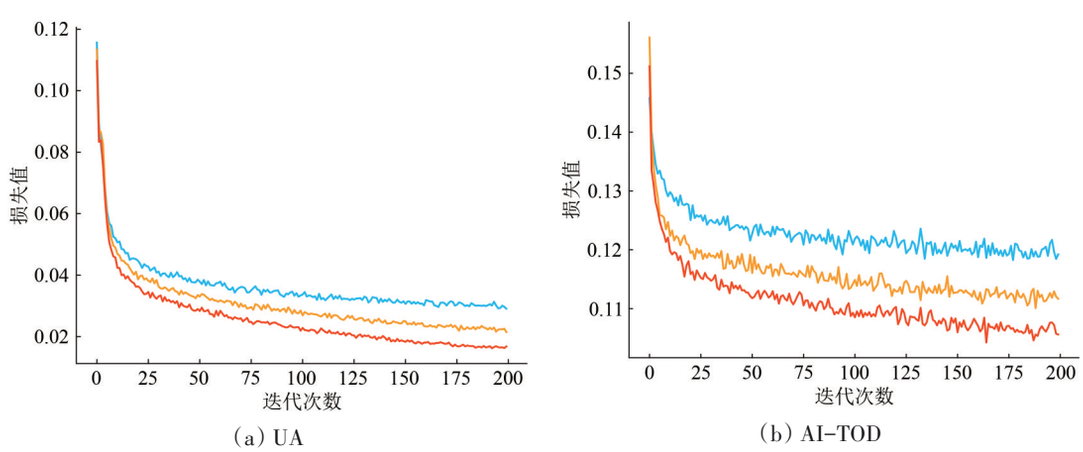
图10展示了消融实验中每一组模块以及最终模型在两组数据集上的训练总损失收敛情况。相较于基础的YOLOv3模型，本文引入的LKGNet模块和GIoU模块均能够在训练前期迅速进行收敛，并保持在相对较小的损失范围中寻找最优结果。在UA数据集上，融合DA模块和BiFPN模块的DA-LKGNet-YOLO模型同样能够更好地拟合目标的检测标签。在AI-TOD数据集上，DA-LKGNet-YOLO模型则保持稳定下降的训练收敛情况。表14展示了每一组模块以及最终模型在两组数据集上的训练时间。在将两组模块组合后，DA-LKGNet-YOLO模型在UA和AI-TOD数据集上分别减少了16.1%和12.5%的训练时间。

表14 训练时间

Table 14 Training time

/s

方法	UA	AI-TOD
base	436	33642
base+LKGNet+GIoU	327	25231
DA-LKGNet-YOLO	381	29436



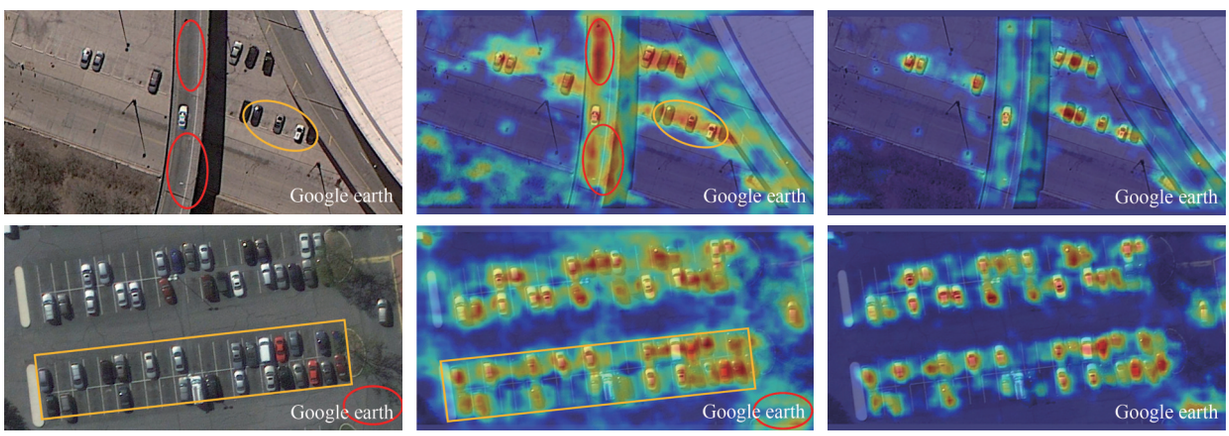
(a) UA (b) AI-TOD

图10 训练损失

Fig. 10 Training loss

5.3 特征图可视化

通过可视化DA模块注意力权重来展示双重注意力模块的工作机制。如图11、图12所示，遥感影像中待检测的小目标（橙色圈部分）所在的区域是期望网络能够重点关注的区域，而剩余的大部分区域（红色圈部分）则包含着容易混淆的背景特征。通过对比引入DA模块前后的可视化特征图可以发现，网络对包含小目标区域的特征进行了重点关注，这体现出网络通过通道注意力机制关注到包含小目标信息丰富的通道，空间注意力机制能有效的理解目标和背景所在的空间位置关系，抑制目标周围的背景信息。

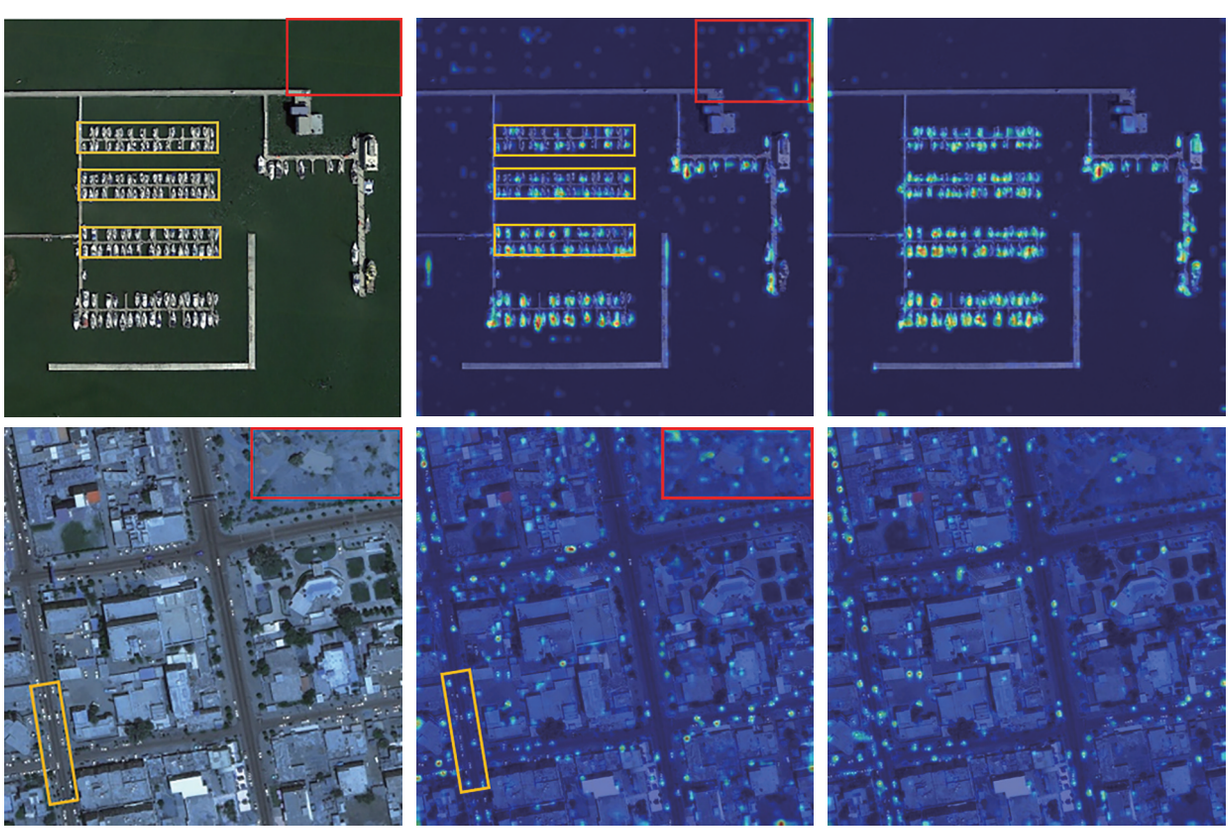


（a）原图 （b）未添加DA （c）添加DA

（a）Original image （b）Without DA （c）With DA

图11 特征图可视化结果（UA）

Fig. 11 Visualization of the learnt feature maps (UA)



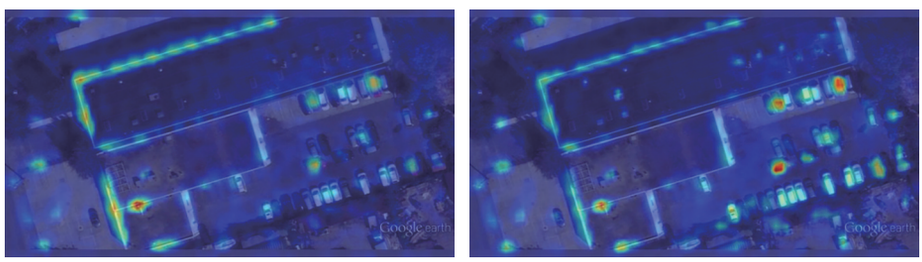
（a）原图 （b）未添加DA （c）添加DA

（a）Original image （b）Without DA （c）With DA

图12 特征图可视化结果（AI-TOD）

Fig. 12 Visualization of the learnt feature maps (AI-TOD)

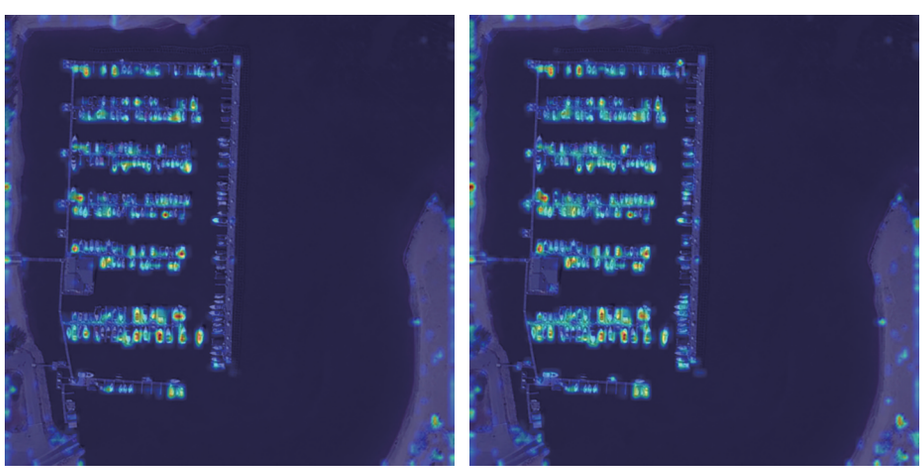
同样，对使用FPN模块以及使用BiFPN模块的网络得到的特征图进行可视化，如图13、图14所示。可以看出，BiFPN结构的特征图对目标有着更好的细节保持能力，同时，经过更多路径的特征信息融合之后物体的特征信息更加清晰。这得益于BiFPN模块不仅同时具备自顶向下与自底向上的特征信息流通路径，而且还增加同层的特征信息流通路径，使网络的各个层都具备了丰富的高级语义特征信息和物体轮廓定位信息，能够使网络提取的特征更加充分的融合。



（a）FPN （b）BiFPN

图13 不同融合结构的特征图可视化（UA）

Fig. 13 Visualization of feature maps learnt by different fusion structures (UA)



（a）FPN （b）BiFPN

图14 不同融合结构的特征图可视化（AI-TOD）

Fig. 14 Visualization of feature maps learnt by different fusion structures (AI-TOD)

6 结 论

本文针对遥感影像小目标检测研究中存在的背景噪声干扰、小目标信息丢失等问题，提出了联合双重注意力机制和双向特征金字塔的遥感影像小目标检测方法。一方面，针对遥感影像中待检测物体所占比例较小导致其特征信息不易提取以及网络参数量大等问题，基于YOLOv3网络模型，引入LKGNet主干网络和GIoU损失函数，提出LKGNet-YOLO网络。通道注意力增大了目标通道的特征响应，减弱了背景区域的特征响应；空间注意力增强了对目标所在区域特征的提取，抑制了小目标周围的背景噪声信息，有益于区分目标与周围环境。另一方面，针对网络特征提取过程中容易受噪声信息干扰以及网络多层信息融合能力不足的问题，基于YOLOv3网络模型，引入DA-LKGNet主干网络和加权双向特征金字塔结构（BiFPN），提出DA-LKGNet-YOLO网络。在UCAS-AOD v1.0数据集和AI-TOD数据集上的实验结果表明：（1）LKGNet主干网络使得整个网络的信息融合能力进一步加强；（2）GIoU损失函数能够更好的感知小目标的位置；（3）双重注意力机制和加权双向特征金字塔在特征提取过程中能更有效的在复杂背景中捕捉遥感小目标的语义信息；（4）实验证明本文提出的方法对小目标具有良好的检测能力。同时与其他先进的目标检测算法相比，该方法取得了优异的检测结果。

志 谢 感谢中国科学院大学提供的UCAS-AOD数据集和武汉大学提供的AI-TOD数据集。

参考文献(References)

Dai W C, Jin L X, Li G N and Zheng Z Q. 2018. Real-time airplane detection algorithm in remote-sensing images based on improved YOLOv3. Opto-Electronic Engineering, 45(12): 180350 (戴伟聪, 金龙旭, 李国宁, 郑志强. 2018. 遥感图像中飞机的改进YOLOv3实时检测算法. 光电工程, 45(12): 180350) [DOI: 10.12086/oee.2018.180350]

Deng L, Yang M, Li T, He Y and Wang C. RFBNet: Deep multimodal Networks with residual fusion blocks for RGB-D semantic segmentation, 2019, p.arXiv: 1907. 00135

Dong R C, Xu D Z, Zhao J, Jiao L C and An J G. 2019. Sig-NMS-based faster R-CNN combining transfer learning for small target detection in VHR optical remote sensing imagery. IEEE Transactions on Geoscience and Remote Sensing, 57(11): 8534-8545 [DOI: 10.1109/TGRS.2019.2921396]

Fu J, Liu J, Tian H J, Li Y, Bao Y J, Fang Z W and Lu H Q. 2019. Dual attention network for scene segmentation//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE: 3141-3149 [DOI: 10.1109/CVPR.2019.00326]

Girshick R. 2015. Fast R-CNN//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago: IEEE: 1440-1448 [DOI: 10.1109/ICCV.2015.169]

Girshick R, Donahue J, Darrell T and Malik J. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE: 580-587 [DOI: 10.1109/CVPR.2014.81]

Han K, Wang Y H, Tian Q, Guo J Y, Xu C J and Xu C. 2020. GhostNet: more features from cheap operations//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE: 1577-1586 [DOI: 10.1109/CVPR42600.2020.00165]

He K, Zhang X, Ren S and Sun J. Spatial pyramid pooling in deep convolutional networks for visual rewgnition, 2014, p. arXiv: 1406.4729

Li K, Wan G, Cheng G, Meng L Q and Han J W. 2020. Object detection in optical remote sensing images: a survey and a new benchmark. ISPRS Journal of Photogrammetry and Remote Sensing, 159: 296-307 [DOI: 10.1016/j.isprsjprs.2019.11.023]

Li M J, Guo W W, Zhang Z H, Yu W X and Zhang T. 2018. Rotated region based fully convolutional network for ship detection//Proceedings of 2018 IEEE International Geoscience and Remote Sensing Symposium. Valencia: IEEE: 673-676 [DOI: 10.1109/IGARSS.2018.8519094]

Li Z X, Yang L and Zhou F Q. 2017. FSSD: feature fusion single shot multibox detector[EB/OL]. https://arxiv.org/abs/1712.00960

Lin T Y, Dollár P, Girshick R, He K M, Hariharan B and Belongie S. 2017. Feature pyramid networks for object detection//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE: 936-944 [DOI: 10.1109/CVPR.2017.106]

Lin T Y, Goyal P, Girshick R, He K M and Dollár P. 2020. Focal loss for dense object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(2): 318-327 [DOI: 10.1109/TPAMI.2018.2858826]

Liu S, Qi L, Qin H F, Shi J P and Jia J Y. 2018a. Path aggregation network for instance segmentation//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE: 8759-8768 [DOI: 10.1109/CVPR.2018.00913]

Liu S T, Huang D and Wang Y H. 2018b. Receptive field block net for accurate and fast object detection//Proceedings of the 15th European Conference. Munich: Springer: 404-419 [DOI: 10.1007/978-3-030-01252-6_24]

Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Y and Berg A C. 2016. SSD: single shot MultiBox detector//Proceedings of the 14th European Conference. Amsterdam: Springer: 21-37 [DOI: 10.1007/978-3-319-46448-0_2]

Nie G T and Huang H. 2021. A survey of object detection in optical remote sensing images. Acta Automatica Sinica, 47(8): 1749-1768 (聂光涛, 黄华. 2021. 光学遥感图像目标检测算法综述. 自动化学报, 47(8): 1749-1768) [DOI: 10.16383/j.aas.c200596]

Qin D D, Wan L, He P E, Zhang Y, Guo Y and Chen J. 2022. Multiscale object detection in remote sensing image by combining data fusion and feature selection. National Remote Sensing Bulletin, 26(8): 1662-1673 (秦登达, 万里, 何佩恩, 张轶, 郭亚, 陈杰. 2022. 结合数据融合与特征选择的遥感影像尺度多样目标检测. 遥感学报, 26(8): 1662-1673) [DOI: 10.11834/jrs.20221249]

Redmon J, Divvala S, Girshick R and Farhadi A. 2016. You only look once: unified, real-time object detection//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE: 779-788 [DOI: 10.1109/CVPR.2016.91]

Redmon J and Farhadi A. 2017. YOLO9000: better, faster, stronger//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE: 6517-6525 [DOI: 10.1109/CVPR.2017.690]

Redmon J and Farhadi A. 2018. YOLOv3: an incremental improvement//Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition. [s.l.]: IEEE: 89-95

Ren S Q, He K M, Girshick R and Sun J. 2017. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6): 1137-1149 [DOI: 10.1109/tpami.2016.2577031]

Sha M M, Li Y and Li A. 2022. Multiscale aircraft detection in optical remote sensing imagery based on advanced Faster R-CNN. National Remote Sensing Bulletin, 26(8): 1624-1635 (沙苗苗, 李宇, 李安. 2022. 改进Faster R-CNN的遥感图像多尺度飞机目标检测. 遥感学报, 26(8): 1624-1635) [DOI: 10.11834/jrs.20219365]

Tan M X, Pang R M and Le Q V. 2020. EfficientDet: scalable and efficient object detection//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE: 10778-10787 [DOI: 10.1109/CVPR42600.2020.01079]

Wang C Y, Bochkovskiy A and Liao H Y M. 2023. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE: 7464-7475 [DOI: 10.1109/CVPR52729.2023.00721]

Wang C Y, Liao H Y M, Wu Y H, Chen P Y, Hsieh J W and Yeh I H. 2020. CSPNet: a new backbone that can enhance learning capability of CNN//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Seattle: IEEE: 1571-1580 [DOI: 10.1109/CVPRW50498.2020.00203]

Wang J W, Yang W, Guo H W, Zhang R X and Xia G S. 2021a. Tiny object detection in aerial images//Proceedings of the 25th International Conference on Pattern Recognition. Milan: IEEE: 3791-3798 [DOI: 10.1109/ICPR48806.2021.9413340]

Wang N, Li B, Wei X X, Wang Y H and Yan H Q. 2021b. Ship detection in spaceborne infrared image based on lightweight CNN and multisource feature cascade decision. IEEE Transactions on Geoscience and Remote Sensing, 59(5): 4324-4339 [DOI: 10.1109/TGRS.2020.3008993]

Zhang S F, Wen L Y, Bian X, Lei Z and Li S Z. 2018a. Single-shot refinement neural network for object detection//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE: 4203-4212 [DOI: 10.1109/CVPR.2018.00442]

Zhang W, Wang S H, Thachan S, Chen J Z and Qian Y T. 2018b. Deconv R-CNN for small object detection on remote sensing images//Proceedings of 2018 IEEE International Geoscience and Remote Sensing Symposium. Valencia: IEEE: 2483-2486 [DOI: 10.1109/IGARSS.2018.8517436]

Zhang Y, Wu X J, Zhao X L, Zhuo L and Zhang J. 2022. Scene constrained object detection method in high-resolution remote sensing images by relation-aware global attention. Journal of Electronics and Information Technology, 44(8): 2924-2931 (张菁, 吴鑫嘉, 赵晓蕾, 卓力, 张洁. 2022. 全局关系注意力引导场景约束的高分辨率遥感影像目标检测. 电子与信息学报, 44(8): 2924-2931) [DOI: 10.11999/JEIT210466]

Zhang Y, Zhu G Y, Shi T J, Zhang K and Yan J H. 2022. Small object detection in remote sensing images based on feature fusion and attention. Acta Optica Sinica, 42(24): 2415001 (张寅, 朱桂熠, 施天俊, 张琨, 闫钧华. 2022. 基于特征融合与注意力的遥感图像小目标检测. 光学学报, 42(24): 2415001) [DOI: 10.3788/AOS202242.2415001]

Zhou Q H and Wang Z. 2022. A remote sensing target detection algorithm based on multi-scale feature enhancement CNNs. Electronics Optics and Control, 29(11): 74-81 (周秦汉, 王振. 2022. 基于多尺度特征增强卷积神经网络遥感目标检测算法. 电光与控制, 29(11): 74-81) [DOI: 10.3969/j.issn.1671-637X.2022.11.013]

Zhu H G, Chen X G, Dai W Q, Fu K, Ye Q X and Jiao J B. 2015. Orientation robust object detection in aerial images using deep convolutional neural network//Proceedings of 2015 IEEE International Conference on Image Processing. Quebec City: IEEE: 3735-3739 [DOI: 10.1109/ICIP.2015.7351502]

Joint dual attention mechanism and bidirectional feature pyramid for remote sensing small targets detection

LI Kewen[1], ZHU Guanglei[2], WANG Hui[1], ZHU Rui[1], DI Xiyao[3,4], ZHANG Tianjian[3,4], XUE Zhaohui[3,4]

1. National Energy Group Jianbi Power Plant, Zhenjiang 212006, China;
2. Yellow River Engineering Consulting Co., Ltd., Zhengzhou 450003, China;
3. School of Earth Sciences and Engineering, Hohai University, Nanjing 211100, China;
4. Jiangsu Province Engineering Research Center of Water Resources and Environment Assessment Using Remote Sensing, Hohai University, Nanjing 211100, China

Abstract: Given imaging characteristics and limitations in spatial resolution, feature extraction for small targets is difficult, which increases the hardship of small-target detection. Existing deep learning target detection network architectures are mostly based on natural images and insufficient for the research on and exploration of small targets in remote sensing images. To overcome these issues, this study proposes a remote sensing image small-target detection algorithm that combines a dual attention mechanism and a bidirectional feature pyramid.

This work provides innovative contributions. (1) To solve the problems of small-target occupation in remote sensing images and the usually huge parameter size, we introduce the LKG bottleneck and generalized intersection over union loss function into YOLOv3 and propose the LKGNet-YOLO network. (2) To resolve the noise disturbance issue and the drawback of feature fusion, we introduce the DA-LKGNet bottleneck and bidirectional feature pyramid network into LKGNet-YOLO and propose the DA-LKGNet-YOLO network.

The remote sensing image dataset (UA) released by the University of Chinese Academy of Sciences in 2014 and the AI-TOD dataset

released by Wuhan University in 2021 is used for experiments to validate the effectiveness of the proposed method in remote sensing image small-target detection. Experimental results demonstrate that the proposed method achieves a mean average precision (mAP) of 96.21% at a threshold of 0.5 on the UA dataset. On the AI-TOD dataset, the $mAP_{0.5—0.95}$ is 9.51% at thresholds ranging from 0.5 to 0.95. Compared with YOLOv3, RFBNet, SSD, FSSD, RetinaNet, and RefineDet, the proposed method has an mAP accuracy that is 3.45%—7.52% and 1.36%—4.84% higher, indicating a considerable performance improvement. Meanwhile, its detection level on small-scale targets is better than that of Faster-RCNN and YOLOv7 algorithms. Compared with the original YOLOv3, our method reduces the number of floating-point operations per second by 48% and the number of parameters by 42%.

A small-target detection method for remote sensing images is proposed in this study. Its contribution lies in the utilization of the YOLOv3 model with LKGNet as the backbone network, along with the design of a dual attention mechanism and a bidirectional feature pyramid network, resulting in a lightweight DA-LKGNet-YOLO model. This model can effectively extract small-sized objects in complex remote sensing images. The experimental results confirm the effectiveness of the proposed method.

Key words: remote sensing image, small target detection, deep learning, YOLOv3, attention mechanism, feature pyramid

Supported by National Natural Science Foundation of China (No. 41971279,42271324); Natural Science Foundation of Jiangsu Province (No. BK20221506)