激光与光电子学进展 先进成像

用于卫星遥感图像的多尺度目标检测算法

项建弘[1,2]，陈振兴[1,2]，王霖郁[1,2*]

[1]哈尔滨工程大学信息与通信工程学院，黑龙江 哈尔滨 150001；

[2]哈尔滨工程大学先进船舶通信与信息技术重点实验室，黑龙江 哈尔滨 150001

摘要 为解决在卫星遥感图像的多尺度目标检测中出现的背景混乱、小目标检测精度低、漏检率高等问题，提出一种用于卫星遥感图像的多尺度目标检测算法。在主干网络中使用通道和空间注意力模块，并重新设计特征融合网络，实现上采样-下采样-上采样的多重融合，并在其中加入通道权重参数，让网络更加关注重要的层次，实现不同层次特征信息的充分利用，使细节特征信息得到增强。在DIOR数据集中的实验结果表明，所提算法不仅显著提升对小目标的检测效果，而且提高对复杂场景中目标的检测精度，与YOLOv5m相比，对部分较小或者复杂的目标检测效果提升明显，精度提升4.5个百分点以上，整体精度提升3.1个百分点。

关键词 遥感；神经网络；多尺度目标检测；注意力机制；通道权重；特征融合

中图分类号 P237 **文献标志码** A **DOI**：10.3788/LOP212670

Multiscale Object Detection Algorithm for Satellite Remote-Sensing Images

Xiang Jianhong[1,2], Chen Zhenxing[1,2], Wang Linyu[1,2*]

[1]*College of Information & Communication Engineering, Harbin Engineering University, Harbin* 150001, *Heilongjiang, China*;

[2]*Key Laboratory of Advanced Ship Communication and Information Technology, Harbin Engineering University, Harbin* 150001, *Heilongjiang, China*

Abstract A multiscale object detection algorithm for satellite remote-sensing images is proposed to solve the problems of background confusion, low precision of small object detection, and high miss rate in multiscale object detection. The channel and spatial attention module is used in the backbone network, and the feature fusion network is redesigned to realize the multiple fusion of up-down-up sampling. The channel weight parameter is added to enable the network to pay more attention to critical channels, fully utilize different feature information levels, and enhance the detailed feature information. In a DIOR dataset, not only the detection effect of small objects but also the detection accuracy of objects in complex scenes is improved. Compared with that using YOLOv5m, the detection effect of some small or complex objects is improved significantly, the accuracy is improved by more than 4.5 percentage points, and the overall accuracy is improved by 3.1 percentage points.

Key words remote sensing; neural network; multiscale object detection; attention mechanism; channel weight; feature fusion

1 引 言

在目标检测中往往会出现对大小尺度不同的目标进行检测的情况，且在实际场景中小目标具有重要的应用[1]，比如在卫星图像分析中，需要检测地面的汽车、飞机和船舶等目标。目前在目标检测领域，小目标存在尺度较小和分辨率不高等问题，导致其检测精度远远低于大目标和中等目标，甚至在卫星图像中小目标还存在着模糊不清的状况。因此当前目标检测领域的重点研究方向之一就是研究对小目标检测的有效方法和提高对小目标的检测性能。

基于卷积神经网络的深度学习目标检测算法大致分为多阶段(two-stage)目标检测算法和单阶段(one-stage)目标检测算法两类。two-stage算法首先生成候

收稿日期：2021-10-08；**修回日期**：2021-11-03；**录用日期**：2021-11-16；**网络首发日期**：2021-11-29

基金项目：国防通信抗干扰重点实验室(9140C020201120C02002)

通信作者：*wanglinyu@hrbeu.edu.cn

选区域，再对候选区域进行预测，如R-CNN系列算法[2-4]。以SSD[5]、YOLO[6-9]为代表one-stage检测算法直接对不同区域的候选框尺度和长宽比进行计算，然后分类与回归，拥有较快的速度。RetinaNet[10]代表另一类one-stage算法，在输出的feature map上分别使用两个全卷积神经网络(FCN)完成分类和回归任务。

在多尺度目标检测中，常使用图像金字塔的方法来进行预测，针对小目标检测精度不高的问题，常见的解决方法是特征金字塔网络(FPN)[11]，FPN将深层feature map融合浅层，增强对小目标的检测效果。path aggregation network (PAN)[12]通过增加下采样特征融合，对FPN输出后的feature map中的位置等信息进行再利用，对小目标的检测有大幅度提升。针对卫星遥感图像小目标检测困难的问题，喻钧等[13]利用K-means++聚类算法确定出合适锚点(anchor)，再利用FPN的思想引入更深层尺度进行特征融合，并使用generalized intersection over union(GIoU)作为损失函数，提高检测的精确度。汪亚妮等[14]在SSD中加入注意力，并进行特征融合，最后用非极大值抑制(Soft-NMS)进行后处理，虽然检测效果提升较多，但整体效果仍不理想。李竺强等[15]提出CLRNet，CLRNet使用连续学习的方式精调检测模型，并改进锚点缩放尺度的层次性来提高检测精度，对遮挡目标的提升较大，但对于遥感图像中出现的大量小目标的效果不佳。农元君等[16]在YOLOv3的基础上精简网络，同时对多尺度预测网络进行改进与优化，引入空间注意力模块来增强遥感目标的特征，达到提高检测速度和精度的目的，但是由于其网络注重实时速度而降低卷积层深度，检测精度提升后仍然不太理想。田婷婷等[17]采用多尺度空洞卷积特征融合检测器来解决遥感图像目标尺度多样和背景复杂等问题，用空洞卷积替代普通卷积，用跳跃连接融合不同尺度信息并使用多维注意力模块，有效提高检测精度，但是使用的数据集种类与样本数量较少，无法充分展现模型泛化能力。

与一般图像相比，遥感影像背景复杂、分辨率不高，从而导致漏检、错检情况频繁发生。为解决上述问题，本文将通道和空间注意力模块(CBAM)[18]添加到YOLOv5的backbone中，在为FPN+PAN增加上采样特征融合的同时设置各融合通道权重[19]，最终对小目标检测效果提高明显并提升整体的检测精度。

2　YOLOv5-WMFPN算法

2.1　YOLOv5算法原理

YOLOv5改进YOLOv3部分网络结构，有效提高整体精度(mAP)，其具有深度和宽度两个参数，可用来得到不同大小网络结构的YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x等4种模型，它们只是使用的具体某一模块的数量与对应的通道数不同，但总体结构一致。YOLOv5网络结构如图1所示。

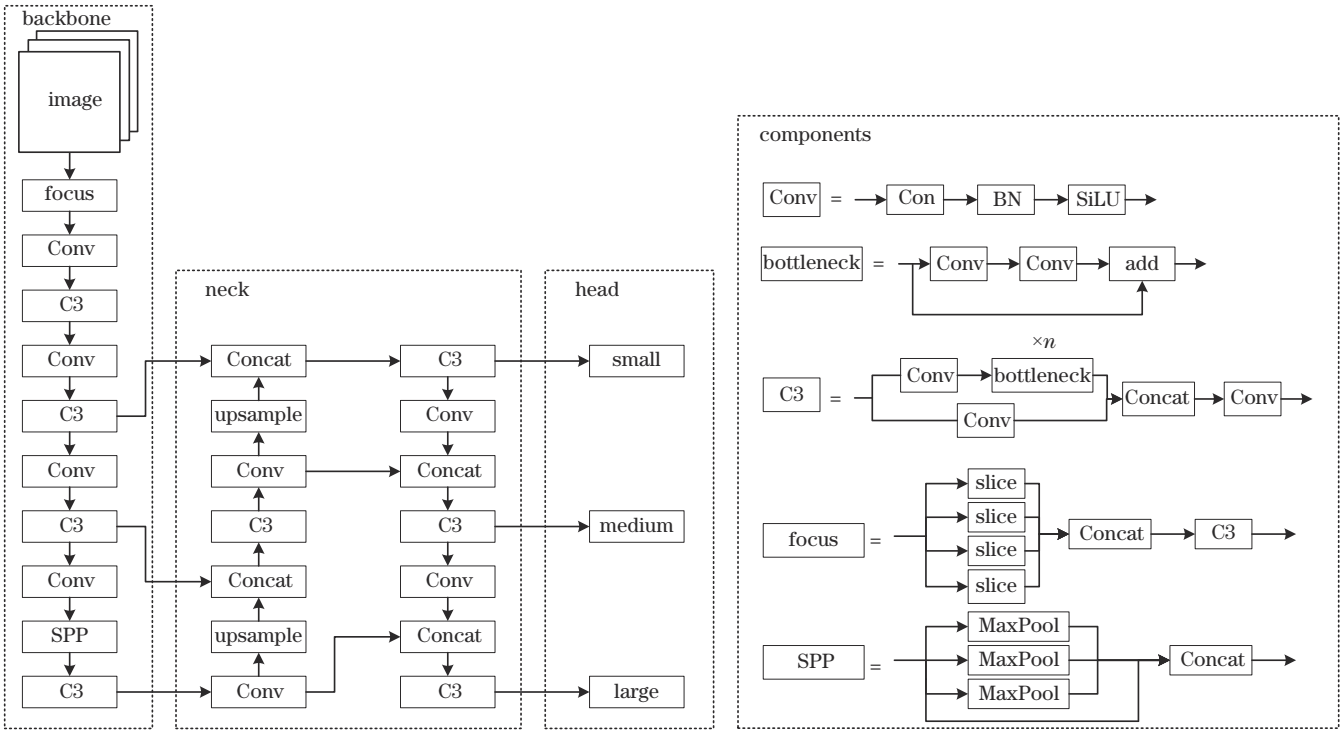

图1　YOLOv5网络结构图

Fig. 1　YOLOv5 network structure diagram

backbone为特征提取网络，其主要结构是包含跨阶段局部网络(CSPNet)[20]的Darknet53、简化多个bottleneck得到的C3模块，该模块在backbone里激活bottleneck中的残差短路，而在neck中不激活。网络中

每个Conv模块由3×3卷积层、批归一化(BN)、SiLU激活函数[21]组成。C3在backbone中以3×3、步长为2的卷积核进行下采样，使feature map的边长缩短为原来的1/2，通道数增大两倍。图片通过fcous模块，使用切片操作将输入通道扩充4倍，在没有信息丢失情况下得到的2倍下采样的feature map。经过多组Conv和C3模块的卷积计算后，feature map的长宽缩短为输入图像的1/32，经过一个空间金字塔池化(SPP)[22]模块，采用5/9/13的最大池化，再进行Concat融合，提高感受野。

neck层的主要结构是FPN+PAN，特征由深层向浅层上采样融合，再由浅层下采样与深层融合，选取3个不同尺度的feature map，用于检测大、中、小等3个尺度的目标。

图像训练前使用Mosaic增强，其思想来自CutMix[23]，如图2所示。

这种数据增强方式简单来说就是把4张图片，通过随机缩放、裁减、排布的方式进行拼接。该方法的优点是缩放拼接后丰富了检测物体的背景和小目标，裁剪后增加许多残缺目标，随机排布又增加了数据集的随机性。BN计算时会将由4张图片组合的新图片算作1张图片进行训练，大幅度提高了训练的速度，并丰富了数据集中样本的数量，提升了鲁棒性。

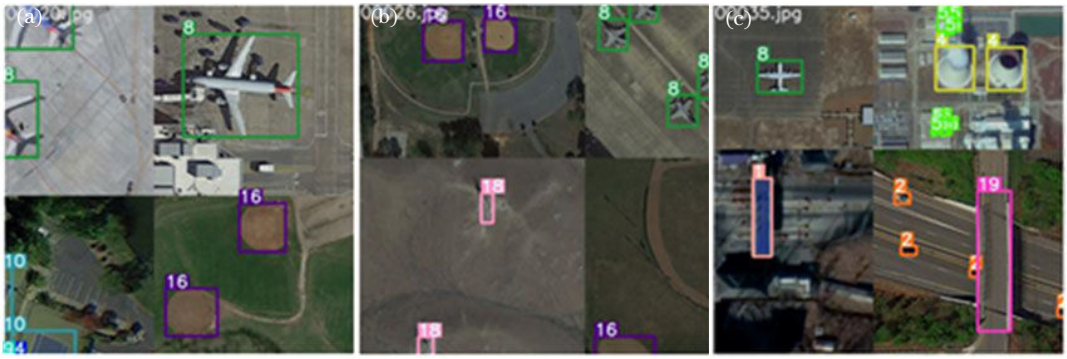

图2　Mosaic增强。(a)(b)(c)使用后效果

Fig. 2　Mosaic data augmentation. (a) (b) (c) Effect after use

2.2　损失函数选用

在检测目标画出回归框(bounding box)时，选用CIoU[24]作为损失函数：

$$\begin{cases} L_{\mathrm{CIoU}} = 1 - R_{\mathrm{IoU}} + \dfrac{\rho^2(B,G)}{c^2} + \alpha v \\ R_{\mathrm{IoU}} = \dfrac{|G \cap B|}{|G \cup B|} \end{cases}, \tag{1}$$

$$\begin{cases} v = \dfrac{4}{\pi^2}\left(\arctan\dfrac{w_{\mathrm{gt}}}{h_{\mathrm{gt}}} - \arctan\dfrac{w}{h}\right)^2 \\ \alpha = \dfrac{v}{(1 - R_{\mathrm{IoU}}) + v} \end{cases}, \tag{2}$$

式中：R_{IoU}为bounding box(B)与ground truth(G)的交并比；$\rho(B,G)$为B与G中心点之间的欧氏距离；c为B与G最小外接矩形的对角线的长度；v就是B与G长宽比的距离；w、h和w_{gt}、h_{gt}分别为B和G的宽高。选用CIoU后，目标框的位置经过计算后会增加，且模型对遮挡物体的识别置信度更高，检测效果更好。

选用交叉熵损失函数作为分类的损失：

$$\begin{cases} L = \ell(\boldsymbol{x},\boldsymbol{y}) = \{l_1, \cdots, l_n, \cdots l_N\}^{\mathrm{T}} \\ l_n = -w_n\left\{y_n \cdot \log\sigma(x_n) + (1 - y_n) \cdot \log\left[1 - \sigma(x_n)\right]\right\} \end{cases}, \tag{3}$$

式中：$\sigma(x) = \dfrac{1}{1 + \exp(-x)}$。输入经过Sigmoid函数处理后再进行交叉熵损失计算，实现数值稳定性。

2.3　增加注意力的backbone网络

在主干网络中引入CBAM卷积注意力模块来提升对目标的特征提取能力，CBAM由通道注意力模块(CAM)和空间注意力模块(SAM)两部分组成。

由于特征的每一个通道都有特定的卷积核，通道注意力的关键是关注什么样的特征更有意义，CAM模块使用全局平均池化和最大池化两种方式来分别利用不同的信息。CAM模块结构如图3所示。

CAM的计算过程可描述为

$$\boldsymbol{y}_{\mathrm{C}} = \sigma\left\{f_{\mathrm{MLP}}\left[\mathrm{AvgPool}(\boldsymbol{x})\right] + f_{\mathrm{MLP}}\left[\mathrm{MaxPool}(\boldsymbol{x})\right]\right\}, \tag{4}$$

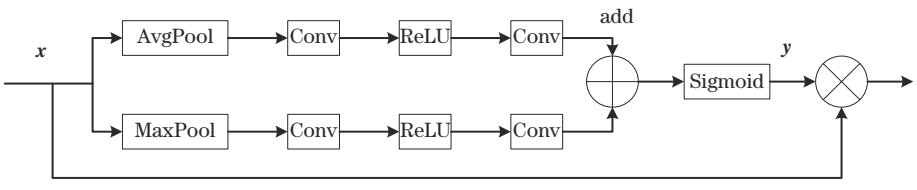

图3　CAM结构图

Fig. 3　CAM structure diagram

式中：输入 $\boldsymbol{x}$ 是一个 $H\times W\times C$ 的 feature map；f_{MLP} 表示 MLP 感知器；σ 表示 Sigmoid 激活函数。使用全局平均池化和最大池化得到两个 1×1×C 的 feature map。在 MLP 感知器中，用卷积将卷积核个数为 C/r 的卷积层压缩，使用激活函数 ReLU 处理后再送入卷积核个数为 C 的卷积层；将两个池化并卷积处理后的特征相加并使用 Sigmoid 归一化后得到权重 $\boldsymbol{y}_{\mathrm{C}}$。输入特征 $\boldsymbol{x}$ 与 $\boldsymbol{y}_{\mathrm{C}}$ 和相乘即可得到经权重缩放后 feature map。

空间注意力模块 SAM 与通道注意力模块 CAM 不同，它关注哪里的特征更有意义，其结构如图 4 所示。

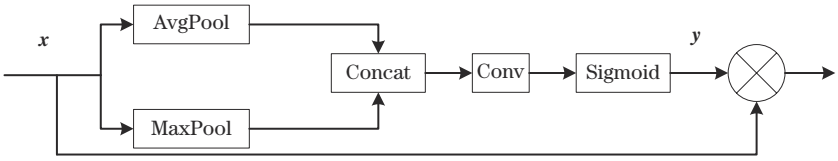

图 4　SAM 结构图

Fig. 4　SAM structure diagram

SAM 的计算过程可描述为

$$\boldsymbol{y}_{\mathrm{S}}=\sigma\left\{f^{7\times7}\left[\mathrm{AvgPool}(\boldsymbol{x});\ \mathrm{MaxPool}(\boldsymbol{x})\right]\right\},\quad(5)$$

式中：$f^{7\times7}$ 表示 7×7 卷积层。SAM 模块接在 CAM 模块后，首先在通道上进行最大池化和平均池化处理并在通道上拼接，然后经过 7×7 卷积层，使通道数降为 1，经过 Sigmoid 后得到权重 $\boldsymbol{y}_{\mathrm{S}}$。最后将输入 feature map 与 $\boldsymbol{y}_{\mathrm{S}}$ 相乘获得最终的 feature map。

使用 CBAM 后，backbone 网络对图像的特征提取更具有针对性，虽少量增加计算参数，但提高了特征提取的针对性。

2.4　带权重的多重特征金字塔融合网络

为了增强对小目标的检测效果，再次对底层信息进行上采样并融合到顶层中，提出一种新的融合结构，为减少运算，删掉部分节点并对 FPN+PAN 输出的 P5 层进行上采样，与 P4、P3 层再次融合，形成 FPN+PAN+FPN 结构（MFPN）。经过又一次上采样融合后，对小目标的检测效果得到了明显提升，但检测中等大小目标时会出现部分检测效果略微下降的问题，这种方法平等对待各输入特征，但是不同的输入特征具有不同的分辨率，它们通常对输出特征的贡献是不同的。为了解决这个问题，在输出前的融合计算时，给每个输入增加一个额外的权重，让网络更加关注重要的层次。带权重的多重特征金字塔融合网络（WMFPN）与原结构对比如图 5 所示。

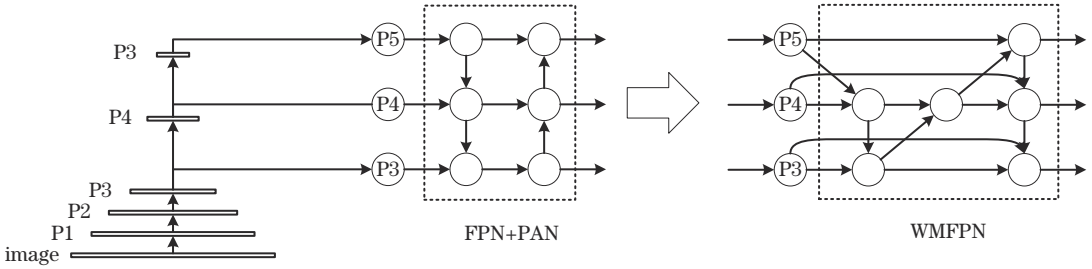

图 5　FPN+PAN 与 WMFPN 结构对比

Fig. 5　Structure comparison between FPN+PAN and WMFPN

特征图输出给 head 检测结构前进行带权重的通道融合，P5 层没有增加额外输入，结构保持不变，P4、P3 层的输入增加一条来自于原 backbone 的对应层，3 条通路分别对应 3 个可学习的权重。P4 层融合网络结构如图 6 所示。

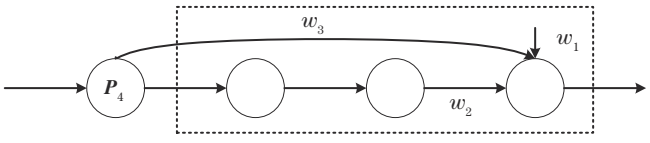

图 6　P4 层融合网络结构

Fig. 6　Fusion network structure of P4 layer

权值 w_1 对应的通路来自于 P5 层上采样 $\left[\mathrm{upsample}(\boldsymbol{P}_{5\mathrm{out}})\right]$，权值 w_2 对应的通路代表上次特征融合的 P4 层（$\boldsymbol{P}_{4\mathrm{last}}$），权值 w_3 对应的通路来自 backbone 网络中的 P4 层（$\boldsymbol{P}_{4\mathrm{backbone}}$）。

则最后 P4 的输出为

$$\boldsymbol{P}_{4\mathrm{out}}=\mathrm{Conv}\left[\frac{w_1\cdot\mathrm{upsample}(\boldsymbol{P}_{5\mathrm{out}})+w_2\cdot\boldsymbol{P}_{4\mathrm{last}}+w_3\cdot\boldsymbol{P}_{4\mathrm{backbone}}}{w_1+w_2+w_3+\varepsilon}\right]。\quad(6)$$

引入权重后，该网络类似于增加了通道选择的注意力机制，可确定融合时应该更关注哪一条通路。将同一张图像使用原网络与所提网络分别进行检测，在输出 P4 层中选 32 个通道的 feature map 进行对比，结果如图 7 所示。

从图 7 可以看出，所提网络输出的 feature map 中特征更加明显，有目标存在的区域与背景之间的差异更

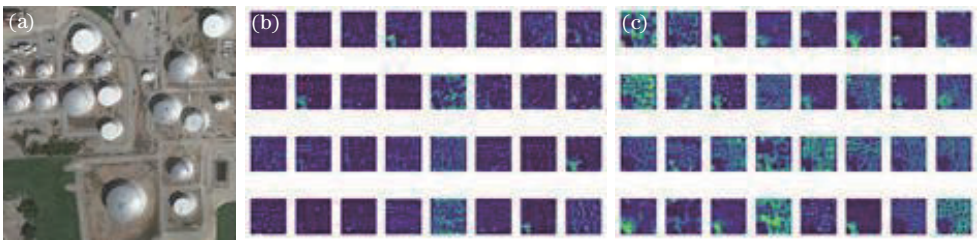

图 7　所提网络与原网络输出的 feature map 对比。(a)原图;(b)原网络 feature map;(c)所提网络 feature map

Fig. 7　Output feature map comparison between proposed network and original network. (a) Original picture; (b) original network feature map; (c) proposed network feature map

明显。对比原图可知,所提网络由于在特征融合时更加关注一些重要的通道,使目标区域的特征更突出。

P5 层 feature map 大小为原图像的 1/32,将 FPN+PAN 结构下采样后的 feature map 直接输出,用于检测大物体。然后 2 倍上采样与 backbone 中对应 P4 层的 feature map 融合,引入该层原特征,通过权重参数调整使其更注重所需的通道,将 P4 层最后的输出用于检测中等物体。然后再将 P4 层的 feature map 上采样后与 P3 层融合,同理融合在 backbone 中对应层的 feature map,最后的输出用于检测小物体。

将 CBAM 注意力模块与 WMFPN 网络整合,构成新的 YOLOv5-WMFPN,如图 8 所示。

将原图像缩放为 640×640 的图像,输入 backbone,使用 focus 和多个 C3 与 Conv 后得到 1/32 原

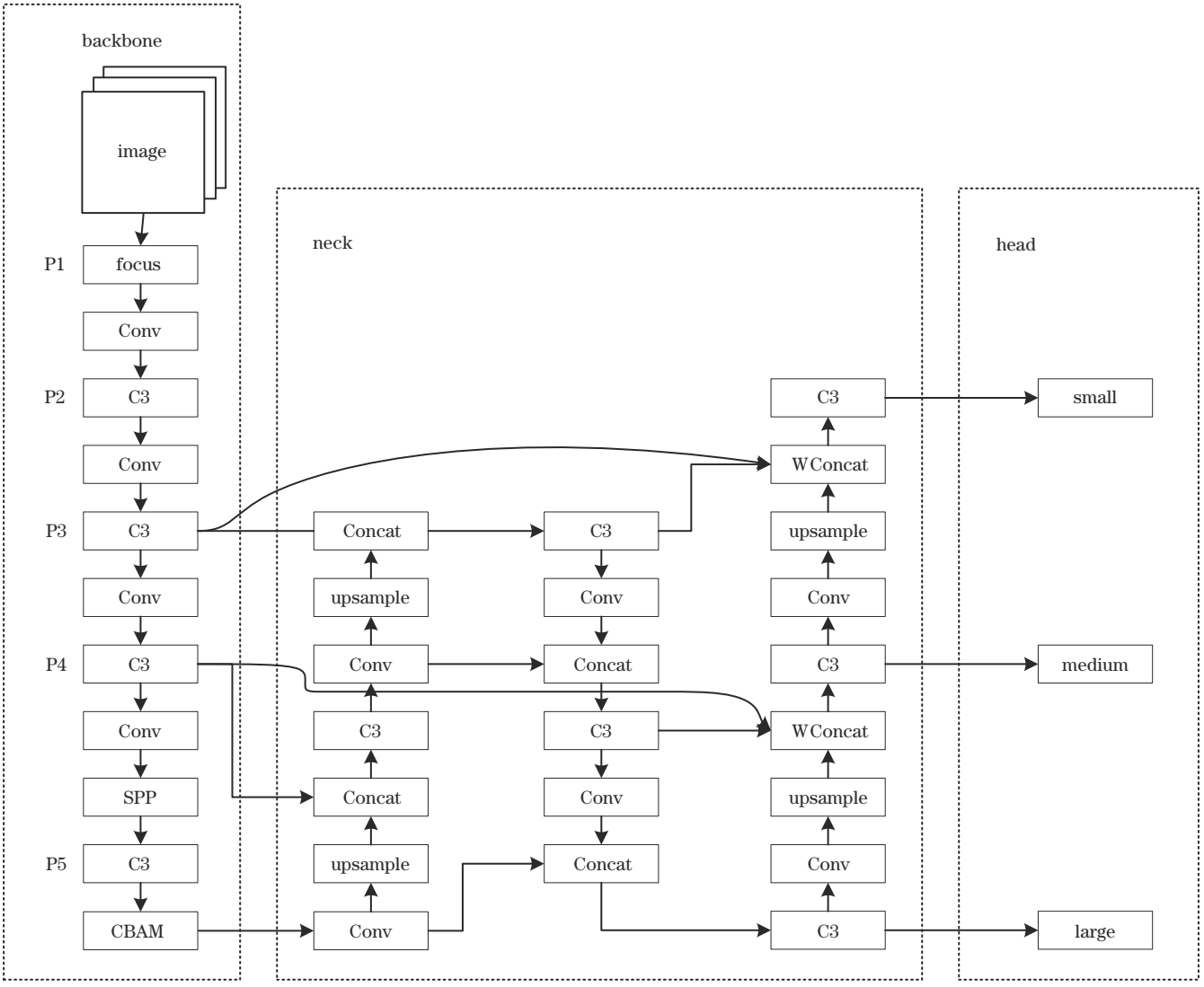

图 8　YOLOv5-WMFPN 结构图

Fig. 8　YOLOv5-WMFPN structure diagram

图宽高的 feature map，经 SPP 层可以大幅增加感受野，将上下文特征分离而获得更加多局部特征的信息。通过 C3 层并使用 CBAM 注意力模块，使主干网络的特征提取更加具有针对性。在 neck 网络进行多尺度的特征融合。经过上采样-下采样-上采样的结构，并在 P4、P3 层输出之前使用带权重的通道融合，使融合时更关注所需要的通道，最后将不同尺度的 P5、P4、P3 层送入检测层分别检测大目标、中目标、小目标。

3 分析与讨论

3.1 数据集选用及评价指标

选用 DIOR 光学遥感图像数据集[25]中的 20 个类别进行检测，该数据集由 23463 个遥感图像和 192472 个对象实例组成，图像大小为 800 pixel×800 pixel，空间分辨率为 0.5～30 m，训练集数目为 5862，验证集数目为 5863，测试集数目为 11738。该数据集类别如图 9 所示：

图 9 光学遥感图像数据集类别。(a)飞机；(b)机场；(c)棒球场；(d)篮球场；(e)桥梁；(f)烟囱；(g)大坝；(h)高速公路服务区；(i)高速公路收费站；(j)高尔夫球场；(k)地面田径场；(l)港口；(m)立交桥；(n)船舶；(o)体育场；(p)储罐；(q)网球场；(r)火车站；(s)车辆；(t)风车

Fig. 9 DIOR optical remote sensing image dataset categories. (a) Airplane; (b) airport; (c) baseball field; (d) basketball count; (e) bridge; (f) chimney; (g) dam; (h) expressway service area; (i) expressway toll station; (j) golf course; (k) ground track field; (l) harbor; (m) overpass; (n) ship; (o) stadium; (p) storage tank; (q) tennis court; (r) train station; (s) vehicle; (t) wind mill

数据集中的 20 个对象分别是飞机、机场、棒球场、篮球场、桥梁、烟囱、大坝、高速公路服务区、高速公路收费站、高尔夫球场、地面田径场、港口、立交桥、船舶、体育场、储罐、网球场、火车站、车辆和风车，记为 c1～c20。

使用的准确率评价指标为 mAP@0.5，即预测框与真实框的交并比(IoU)超过 50% 时认为该检测结果是正确的。mAP 的表达式为

$$\begin{cases} P=\dfrac{N_{TP}}{N_{TP}+N_{FP}} \\ R=\dfrac{N_{TP}}{N_{TP}+N_{FN}} \\ P_{AP}=\displaystyle\int_0^1 P(R)\mathrm{d}R \\ P_{mAP}=\dfrac{\displaystyle\sum_{i=1}^{N}P_{APi}}{N} \end{cases}, \tag{7}$$

式中：N_{FN}是正样本被判负的样本数；N_{FP}是负样本被判正的样本数；N_{TP}是正确判定的样本数；AP指平均精准度，是每一个类别的精确率；mAP指各类别AP的平均值。

3.2　所提算法与YOLOv5m比较

使用CPU为i7 9700k、GPU为RTX2080ti、内存为32 GB的设备进行训练、验证及测试，Python版本为3.8，Pytorch版本为1.7.1。batchsize设置为16，训练的epoch为100。输入图像尺寸缩放调整为640×640，初始学习率为0.01，学习率周期衰减设置为0.2，即完成训练100 epoch后最终学习率为0.002。YOLOv5算法有深度和宽度两个参数，按模型大小可分为YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x。选择YOLOv5m模型进行对比实验，所提算法与YOLOv5m的各类AP对比如表1所示。

表1　YOLOv5-WMFPN与YOLOv5m各类AP对比

Table 1　Comparison of various APs between YOLOv5-WMFPN and YOLOv5m algorithm　　unit: %

Algorithm	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10
YOLOv5m	78.4	78.2	73.8	88.1	45	77.5	64.3	62	59.6	77.9
YOLOv5m-WMFPN	89.1	83	74.4	91.9	47	82	64.6	63.6	65	79.2
Algorithm	c11	c12	c13	c14	c15	c16	c17	c18	c19	c20
YOLOv5m	72.6	60.4	58.4	88.9	65.6	76.6	86.7	63	55.1	77
YOLOv5m-WMFPN	77.3	63.1	61.1	89.2	69.3	77.4	89.7	64.8	58.6	81.3

从表1可以看出，所提算法AP在每一类上都有一定提升，尤其在c1、c2、c6、c9、c11、c20等类中提升效果显著，达4.5个百分点以上，对以小目标为主的飞机目标提升尤为明显，达到10.7个百分点。

所提算法与YOLOv5m的整体效果对比如表2所示。

表2　YOLOv5m-WMFPN与YOLOv5m的整体效果对比

Table 2　Overall effect comparison between YOLOv5m-WMFPN and YOLOv5m

Algorithm	P /%	R /%	mAP@0.5 /%	mAP@[.5:.95] /%	Detection speed /(frame·s^{-1})
YOLOv5m	86.1	65.3	70.5	46.6	113.6
YOLOv5m-WMFPN	84.9	68.2	73.6	50.5	102.0

从表2可以看出，所提算法的mAP@0.5提升3.1个百分点，mAP@[.5:.95]提升3.9个百分点，虽然增加了特征融合网络的层数使计算量增多，但牺牲少量的计算效率却换来了明显的性能提升。在数据集20个对象类中，以飞机、风力发电机、船舶、车辆这几类小目标为主，以下主要对这几类目标进行展示与结果分析。

图10是对飞机、风力发电机目标的检测，使用的数据集为卫星遥感图片，飞机目标非常小，但是其特征比较明显，往往停在机场内周围环境不复杂，对检测干扰比较小。YOLOv5m算法在针对很小的飞机目标时往往会出现漏检的问题，而所提算法对于很小的飞机目标也可以检测，虽然也会出现个别漏检情况，但是相较于YOLOv5m算法提升很大。在飞机目标稍大时，两种算法都基本可以全部检测出飞机目标，且精确度都比较高。风力发电机往往在平原山坡上，仅会受到周围环境的干扰，但由于遥感图像往往是俯视拍摄的，风力发电机的特征不算明显，所提算法不仅将YOLOv5m漏检的目标均检测出来，精度也有较大提升。

在卫星遥感图像检测中经常会出现目标过于小导致人眼也无法准确识别的情况，也会出现目标处于复杂背景下因为分辨率不足而模糊的情况。图11中，车辆目标十分微小，在图中车辆的特征不明显，而且多分布于城区中，背景十分复杂，对车辆目标的检测存在很大的干扰。船舶目标在水域背景颜色偏深色且分辨率低而模糊时，会导致检测难度增加而无法判断其是否为正确目标，当船舶稍大且清晰时，这样它的特征也比较明显，检测效果也会比较好。相比YOLOv5m，所提算法在这样的场景下降低了漏检的概率，提高了检测的精度。

经过对比发现，所提YOLOv5m-WMFPN算法不仅检测出YOLOv5m中漏检的目标，而且准确度也比较高，整体效果优于YOLOv5m算法。在小目标检测中，这些极小的目标往往存在分辨率低、图像模糊、携带的信息少等问题，有时人眼也很难分辨。所提WMFPN增强了对小目标的检测，改善了分辨率低、图像模糊、携带的信息少等情况下的检测效果，在与YOLOv5m的对比实验中，降低了漏检率，能够检测出更多的目标，在目标极小时有良好的检测效果。

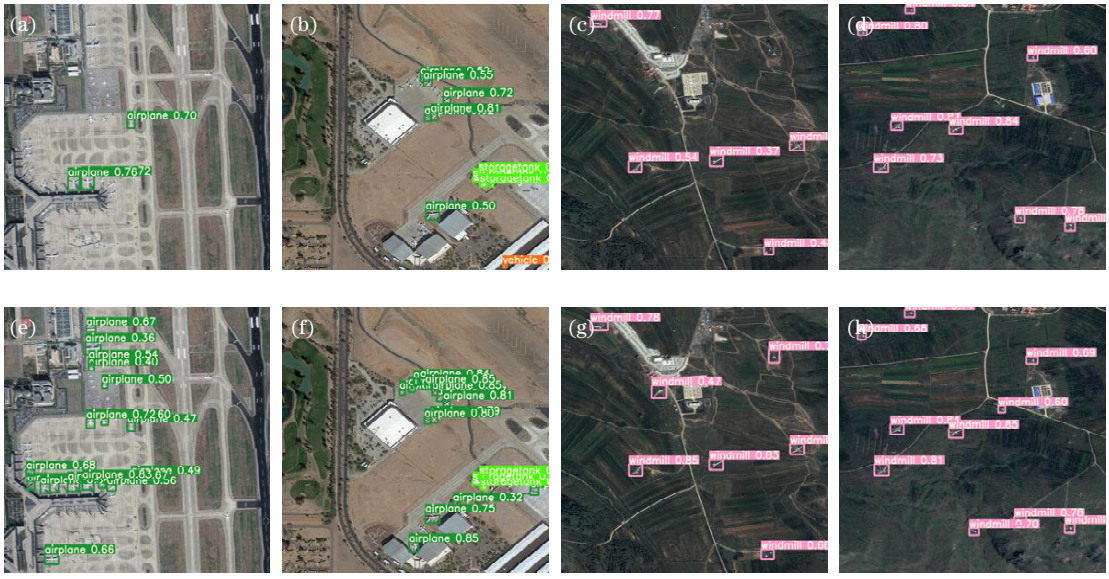

图 10　飞机、风力发电机目标检测。(a)(b) YOLOv5m 飞机；(c)(d) YOLOv5m 风力发电机；(e)(f) YOLOv5m-WMFPN 飞机；(g)(h) YOLOv5m-WMFPN 风力发电机

Fig. 10　Airplane and wind mill object detection. (a) (b) YOLOv5m airplane; (c) (d) YOLOv5m wind mill; (e) (f) YOLOv5m-WMFPN airplane; (g) (h) YOLOv5m-WMFPN wind mill

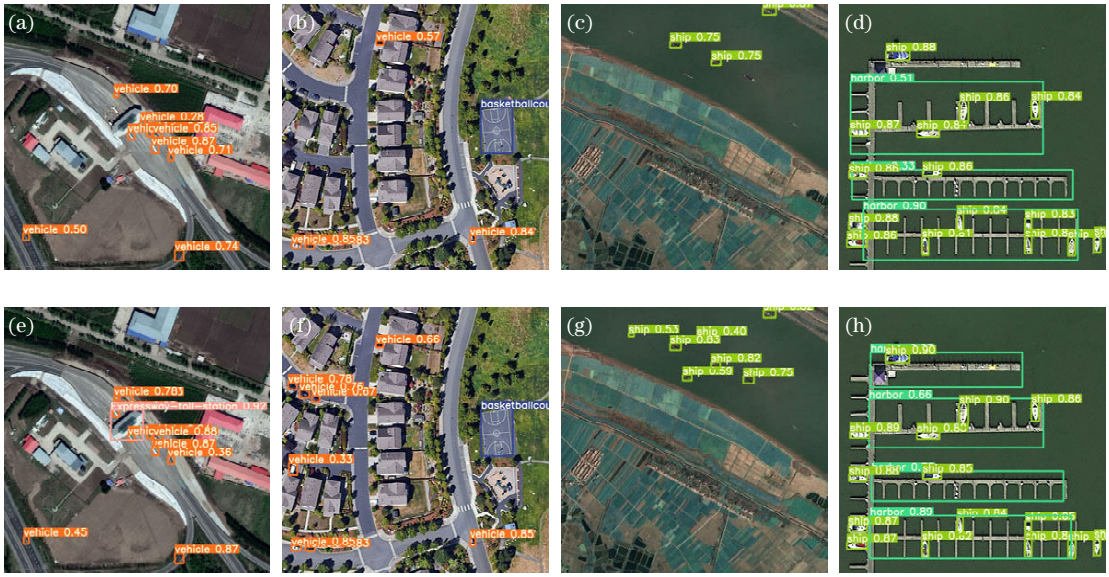

图 11　车辆、船舶检测。(a)(b) YOLOv5m 车辆；(c)(d) YOLOv5m 船舶；(e)(f) YOLOv5m-WMFPN 车辆；(g)(h) YOLOv5m-WMFPN 船舶

Fig. 11　Vehicle and ship object detection. (a) (b) YOLOv5m vehicle; (c) (d) YOLOv5m ship; (e) (f) YOLOv5m-WMFPN vehicle; (g) (h) YOLOv5m-WMFPN ship

3.3　消融实验

为对比某一个模块对检测效果的影响，进行了消融实验。实验 1 为 YOLOv5m 算法原结构的性能，实验 2、3、4、5 则为启用对应模块后的性能。消融实验结果如表 3 所示。

从表 3 可以看出，任一模块对检测精度都有提升，但随着网络更加复杂，检测速度也有相应的下降。在 MFPN 结构之上增加通道权重后，在特征融合时可以更加关注需要的通道特征，使 mAP 获得较大的提升。模型的性能并不随着模块的增加而线性提升，在逐步改进 YOLOv5m 的主干网络和特征融合网络的过程中，模型的检测能力也随之逐步提升。

3.4　YOLOv5-WMFPN 与主流算法性能对比

为了进一步验证所提算法的效果，将其与其他主流算法的性能进行比较，结果如表 4 所示。

从这几种算法性能对比可以看出：SSD 算法作为早期单阶段检测算法，整体检测效果垫底；Faster RCNN 算法作为早期性能较好的算法，有着庞大的计算参数量，但是由于出现较早受限于 VGG16 的性能，总体效果不佳；但使用 ResNet50 和 FPN 的 Faster

表 3 消融实验结果

Table 3 Ablation experiment results

Experiment	MFPN	WMFPN	CBAM	mAP /%	Detection speed /(frame·s^{-1})
1				70.5	113.6
2			√	71.2	109.4
3	√			71.5	107.3
4	√	√		73.3	106.1
5	√	√	√	73.6	102.0

表 4 各算法性能比较

Table 4 Performance comparison of each algorithm

Algorithm	Backbone	Neck	mAP /%
Faster RCNN	VGG16		57.6
Faster RCNN-FPN	ResNet50	FPN	67.4
SSD	VGG16		54.3
Yolov3	Darknet53	FPN	64.2
Yolov3-SPP	Darknet53	FPN	67.5
Efficientdet-d2	Efficientnet	BiFPN	55..5
Efficientdet-d4	Efficientnet	BiFPN	63.9
YOLOv5s	C3CSP	FPN+PAN	67.2
YOLOv5m	C3CSP	FPN+PAN	70.5
YOLOv5m-WMFPN	C3CSP-CBAM	WMFPN	73.6

RCNN 相较于旧版性能提升 9.8 个百分点，得益于特征金字塔结构实现多尺度目标检测，使整体检测效果提升很大；YOLOv3 作为使用最为广泛的 YOLO 系列算法，检测效果比较一般，增加 SPP 网络后扩大了感受野，使 YOLOv3 的检测效果有了较大的改善；Efficientdet 算法是较新且效果较好的算法，其 d4 版本比 d2 拥有更多的模块数量和更深的网络，因此获得了更好的效果，但整体效果不太理想；YOLOv5 算法在 YOLOv3 的基础上使用了精简的 backbone，使用 FPN+PAN 的特征融合网络结构，并在训练前对图片进行 Mosaic 数据增强，模型大小不到 YOLOv3 四分之一的 YOLOv5s 就能达到与 YOLOv3-SPP 相同的检测效果；而网络深度和模块数量相比较 YOLOv5s 略微增多 YOLOv5m 效果提升也很明显；所提 YOLOv5-WMFPN 算法效果在各对比的算法中最佳，并在小目标的检测中提升显著，整体 mAP 达到了 73.6%。

4 结 论

为解决卫星遥感图像的目标检测中目标过小或者模糊导致检测效果不佳的问题，提出 YOLOv5-WMFPN 算法，在 backbone 中使用通道注意力与空间注意力模块 CBAM，提高了对特征提取的针对性。提取的 feature map 经过上采样-下采样-上采样结构的特征融合，不仅使细节特征信息得到增强，而且也使 feature map 富含位置信息，并使用通道融合权重，使通道融合时让网络关注更加重要的通道，实现不同层次特征信息的充分利用，在稍微增加计算量的同时将精度大幅提高。实验结果表明，该算法可以有效提升对卫星遥感图片的多尺度检测精度，对部分小目标和复杂目标检测效果提升明显，与 YOLOv5m 相比，在 DIOR 数据集上 mAP 提升 3.1 个百分点，达 73.6%。

参 考 文 献

[1] Huang J, Rathod V, Sun C, et al. Speed/accuracy trade-offs for modern convolutional object detectors[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 3296-3297.

[2] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 580-587.

[3] Girshick R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision, December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1440-1448.

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

[5] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.

[6] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.

[7] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6517-6525.

[8] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2021-09-18]. https://arxiv.org/abs/1804.02767.

[9] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2021-09-18]. https://arxiv.org/abs/2004.10934.

[10] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//2017 IEEE International

Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2999-3007.

[11] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 936-944.

[12] Liu S, Qi L, Qin H F, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 8759-8768.

[13] 喻钧, 康秦瑀, 陈中伟, 等. 基于全卷积神经网络的遥感图像海面目标检测[J]. 弹箭与制导学报, 2020, 40(5): 15-19, 23.
Yu J, Kang Q Y, Chen Z W, et al. Sea surface target detection in remote sensing images based on full convolution neural network[J]. Journal of Projectiles, Rockets, Missiles and Guidance, 2020, 40(5): 15-19, 23.

[14] 汪亚妮, 汪西莉. 基于注意力和特征融合的遥感图像目标检测模型[J]. 激光与光电子学进展, 2021, 58(2): 0228003.
Wang Y N, Wang X L. Remote sensing image target detection model based on attention and feature fusion[J]. Laser & Optoelectronics Progress, 2021, 58(2): 0228003.

[15] 李竺强, 朱瑞飞, 马经宇, 等. 联合连续学习的残差网络遥感影像机场目标检测方法[J]. 光学学报, 2020, 40(16): 1628005.
Li Z Q, Zhu R F, Ma J Y, et al. Airport detection method combined with continuous learning of residual-based network on remote sensing image[J]. Acta Optica Sinica, 2020, 40(16): 1628005.

[16] 农元君, 王俊杰. 基于嵌入式的遥感目标实时检测方法[J]. 光学学报, 2021, 41(10): 1028001.
Nong Y J, Wang J J. Real-time object detection in remote sensing images based on embedded system[J]. Acta Optica Sinica, 2021, 41(10): 1028001.

[17] 田婷婷, 杨军. 基于多尺度特征融合网络的遥感影像目标检测[J]. 激光与光电子学进展, 2022, 59(16): 1628002.
Tian T T, Yang J. Object detection for remote sensing image using multi-scale feature fusion network[J]. Laser & Optoelectronics Progress, 2022, 59(16): 1628002.

[18] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 3-19.

[19] Tan M X, Pang R M, Le Q V. EfficientDet: scalable and efficient object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 10778-10787.

[20] Wang C Y, Mark Liao H Y, Wu Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 14-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 1571-1580.

[21] Elfwing S, Uchibe E, Doya K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning[J]. Neural Networks, 2018, 107: 3-11.

[22] He K M, Zhang X Y, Ren S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.

[23] Yun S, Han D, Chun S, et al. CutMix: regularization strategy to train strong classifiers with localizable features [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 6022-6031.

[24] Zheng Z, Wang P, Ren D, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation[EB/OL]. (2020-05-07)[2021-09-18]. https://arxiv.org/abs/2005.03572v3.

[25] Li K, Wan G, Cheng G, et al. Object detection in optical remote sensing images: a survey and a new benchmark[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 159: 296-307.