第 42 卷 第 24 期/2022 年 12 月/光学学报 研究论文

光学学报

基于特征融合与注意力的遥感图像小目标检测

张寅[1,2]，朱桂熠[1,2]，施天俊[3]，张琨[1,2]，闫钧华[1,2*]

[1]南京航空航天大学空间光电探测与感知工业和信息化部重点实验室，江苏 南京 211106；

[2]南京航空航天大学航天学院，江苏 南京 211106；

[3]哈尔滨工业大学空间光电工程中心，黑龙江 哈尔滨 150001

摘要 为解决遥感图像小目标检测中目标特征信息量少、定位困难等难题，提出一种基于特征融合与注意力机制的遥感图像小目标检测算法FFAM-YOLO(Feature Fusion and Attention Mechanism YOLO)。该算法首先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的问题，构造特征增强模块(FEM)以融合较低层级特征图中多重感受野特征，提升算法主干网络的目标特征提取能力；其次，主干网络提取得到高低层级特征图后，建立重构算法的高低层级特征融合结构，利用特征融合模块(FFM)显著增强小目标的特征信息；在增强的有效通道注意力机制(E-ECA)与空间注意力模块(SAM)所组成的级联注意力机制(ESM)作用下，可更精确地捕获小目标特征；最后在输出的两路特征图上进行小目标检测并输出结果。实验结果表明，基于构建的遥感图像小目标数据集USOD(Unicorn Small Object Dataset)，所提算法的查准率达到91.9%，查全率达到83.5%，检测框与真实框之间的交并比阈值(IoU)为0.5时的平均精度(AP)为89%，IoU为0.5∶0.95时的AP达到32.6%，检测速率达到120 frame/s，具有一定的鲁棒性和实时性。

关键词 机器视觉；小目标检测；遥感图像；特征融合；注意力机制；特征增强

中图分类号 TP391.4 **文献标志码** A **DOI**：10.3788/AOS202242.2415001

Small Object Detection in Remote Sensing Images Based on Feature Fusion and Attention

Zhang Yin[1,2], Zhu Guiyi[1,2], Shi Tianjun[3], Zhang Kun[1,2], Yan Junhua[1,2*]

[1]*Space Photoelectric Detection and Sensing of Industry and Information Technology, Nanjing University of Aeronautics and Astronautics, Nanjing* 211106, *Jiangsu, China*;

[2]*College of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing* 211106, *Jiangsu, China*;

[3]*Research Center for Space Optical Engineering, Harbin Institute of Technology, Harbin* 150001, *Heilongjiang, China*

Abstract To deal with issues such as less feature information and difficult positioning raised by small object detection in remote sensing images, this paper proposes a remote sensing image small-target detection algorithm FFAM-YOLO (Feature Fusion and Attention Mechanism YOLO) based on feature fusion and attention mechanism. Firstly, in terms of inadequate effective information in backbone network feature extraction and weak information representation in feature maps, the algorithm constructs a feature enhancement module (FEM) to fuse multiple receptive field features in lower-level feature maps and improve the network′s ability in extracting object features. Secondly, with low-level and high-level feature maps obtained by the backbone network, the algorithm′s low-level and high-level feature fusion structures are rebuilt, and a feature fusion module (FFM) is implemented to enhance the feature information of small targets. Thirdly, small object features are accurately captured by cascade attention mechanism (ESM) consisting of enhanced-efficient channel attention (E-ECA) and spatial attention module (SAM). Finally, the small object is detected in the output dual-branch feature maps, and results are delivered. The experimental results show that with the USOD (Unicorn Small Object Dataset), based on the constructed remote sensing images, the proposed algorithm achieves a precision of 91.9% and a recall of 83.5%, with an average precision AP of 89% for intersection ratio threshold (IoU) between the prediction box and

收稿日期：2022-04-15；**修回日期**：2022-05-16；**录用日期**：2022-06-16

基金项目：国防科技基础加强计划资助(2021-JCJQ-JJ-0834)、国家自然科学基金(61901504,61705104)、中央高校基本科研业务费资助(NJ2020021,NT2020022)

通信作者：*yjh9758@126.com

the ground truth box of 0.5 and an AP of 32.6% for IoU of 0.5∶0.95, respectively, and the detection rate reaches 120 frame/s. The algorithm is with robustness and real-time performance.

Key words machine vision; small object detection; remote sensing image; feature fusion; attention mechanism; feature enhancement

1 引 言

遥感图像小目标检测作为计算机视觉领域研究的重点与难点，在安防、军事、制造等领域均具有很大的应用价值[1]。相较于常规目标，小目标具有特征弱、信息量少等特点，将其从相似背景或毗连目标中区分出来十分困难，当面临光照度低、阴影遮挡等复杂环境，则对遥感图像小目标检测提出更高要求。

深度学习和图形处理器技术的高速发展将目标检测算法性能带向新的高度[2]，基于深度学习的目标检测算法可按双阶段与单阶段检测算法进行区分。双阶段检测算法首先在图像上生成大量候选区域，在此基础上对目标区域位置及范围进行修整并进行目标分类，以 Fast R-CNN[3]、Faster R-CNN[4]、Mask R-CNN[5]等区域卷积神经网络（R-CNN）系列算法为典型代表；单阶段检测算法则跳过候选区域生成步骤，通过回归分析直接产生目标类别概率和预测框坐标信息，该类算法以 YOLO[6]、SSD[7]、YOLO9000[8]、YOLOv3[9]、YOLOv4[10]及 YOLOv5[11]等算法为典型代表。

YOLO 系列目标检测算法因其网络结构简单且同时兼顾检测精度和检测速率而受到广泛关注与应用。在常规目标检测任务中：文献[12]利用注意力机制优化特征融合结构并重新设计损失函数，实现行人、车辆检测；文献[13]通过注意力机制加强特征提取，并通过结合迁移训练实现行人检测；文献[14]将目标局部作为检测重点并引入迁移学习，实现货车检测；文献[15]提出改进的主干网络及锚框机制，实现果实检测；文献[16]将主干网络与 GhostNet 进行融合，以捕获和细化特征并实现舰船检测等。在小目标检测任务中：文献[17]采用 K-means 方法对数据进行聚类分析并优化损失函数，实现尺度大于 50×50 像素的交通灯检测；文献[18]通过替换主干网络并增加特征融合层级，实现尺度大于 32×32 像素的路障检测；文献[19]构建通道注意力模块用于高低层级融合结构，实现尺度大于 32×32 像素的车辆检测；文献[20]通过拓展预测头个数并改进锚框机制与损失函数，实现尺度大于 32×32 像素的车辆检测；文献[21]提出通道注意特征金字塔并剔除检测大目标的模块，实现尺度大于 32×32 像素的车辆检测等。上述研究实现了相对较小尺度目标的检测，但无法有效应对复杂遥感场景图像中更小尺度目标（16×16 像素及以内）的准确检测问题，包括：1）小目标纹理、位置等信息少，算法主干网络对小目标的特征表征能力较弱；2）网络对输入图像进行逐层卷积下采样处理，造成小目标特征大量丢失甚至被忽略；3）网络学习过程易受相近背景噪声及非目标特征主导，导致检测效果不佳。

针对上述不足，本文提出一种基于特征融合与注意力机制的小目标检测算法 FFAM-YOLO（Feature Fusion and Attention Mechanism YOLO）。所提算法首先构造特征增强模块（FEM），并在主干网络低层级特征图引入该模块，以综合提高主干网络对遥感图像小目标特征的表达能力；基于经 FEM 模块所增强的特征，重新构造高低层级特征融合结构，利用特征融合模块（FFM）实现高低层级特征图的优势互补；在级联注意力机制（ESM）作用下，凸显重要的小目标特征；最后，结合输出的双通道特征图与非极大值抑制（NMS）实现小目标检测。

2 FFAM-YOLO

FFAM-YOLO 算法的整体框架如图 1 所示（图中 CBS 为 Conv＋Batch Normalization＋Leaky ReLU，CBL 为卷积＋批量归一化＋Leaky ReLU 激活函数），算法具体步骤为：1）主干网络特征提取，由于网络实现小目标检测主要依靠较低层级特征图，与 YOLOv5 算法相比，FFAM-YOLO 中主干网络仅进行 4 次卷积下采样操作并进行特征提取，选取下采样倍数分别为 4、8、16 的特征图参与高低层级特征融合，可避免仅含有极少小目标特征信息的高层级特征图在融合阶段带入冗余信息，降低模型检测精度；2）对主干网络所提取特征进行增强，受到 YOLOv5 算法通过在最高层级特征图中引入空间金字塔池化[22]（SPP）模块扩展主干网络对特征的映射和接收范围这一操作的启发，构建 FEM，该模块基于感受野模块[23]（RFB）原理，使用残差连接、多个尺度不同的常规卷积及膨胀卷积构成多路结构，通过横向增加网络宽度、提升感受野，增大网络对小目标特征提取的覆盖范围、提高网络对小目标的敏感程度；3）高低层级特征融合，针对小目标及其对应的特征图所呈现的特性，重构高低层级特征融合结构，形成 FFM，以显著增强小目标特征；4）级联注意力机制聚焦重要特征，在增强的有效通道注意力机制（E-ECA）与空间注意力机制（SAM）所组成的 ESM 作用下，抑制背景噪声，给予重要的小目标特征更大权重；5）网络输出端预测结果，基于由高低层级特征融合结构输出的两路特征图进行小目标检测，并经过非极大值抑制处理后得到最终的小目标检测结果。

2.1 特征增强模块

网络主要使用低层级特征图实现小目标预测，低层级特征图中含有丰富的小目标形状、位置等细节信息，但主干网络提取能力有限，且该特征图处于网络的初始阶段，未经过充分处理，使得低层级特征图中语义信息少、感受野受限，不能较好地适应于小目标检测。

本文从 RFB 原理出发构建 FEM，通过在多个支

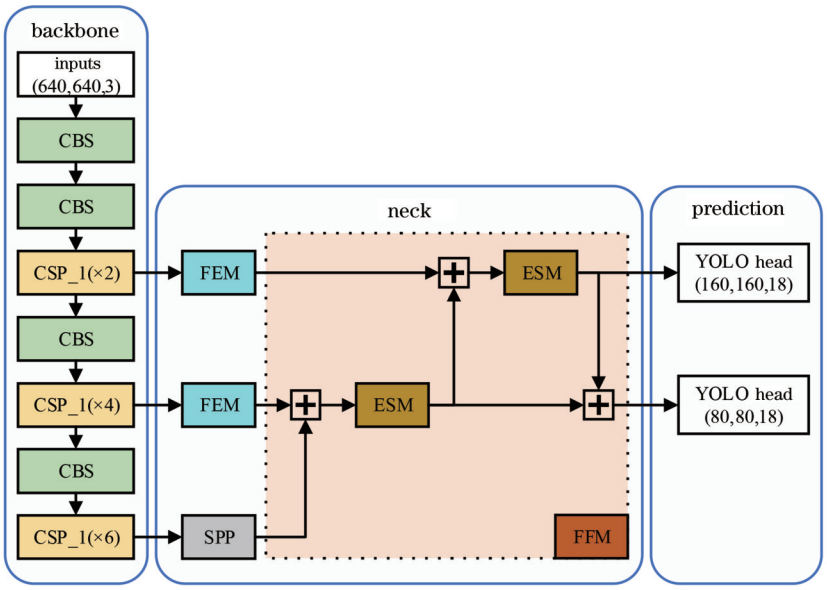

图 1　FFAM-YOLO 算法整体框架

Fig. 1　Overall scheme of FFAM-YOLO algorithm

路上采用由尺度不同、数量不同的常规卷积及膨胀卷积所构成的多分支结构，拼接形成多通道特征图，通过横向拓展网络宽度，同步增大网络的感受野和提升网络对图像中小目标的敏感性与适应性。FEM 结构如图 2 所示，该多分支结构共存在 4 条支路，均先对输入特征图进行 1×1 的卷积操作，初步处理并调整特征图通道数以进行特征图后续处理，其中：第 4 条支路为残差结构，在输出端形成等价映射，保留良好的小目标特征信息；其余 3 条支路均经过如 1×3、3×1 及 3×3 等常规卷积操作并进行级联，以不同尺度卷积操作提取更细腻的小目标特征；中间 2 条支路额外增加空洞卷积层，使提取特征图包含更多上下文信息，增强特征有效性。特征增强模块 FEM 的计算过程可表示为

$$W_1 = f_{\text{conv}}^{3\times3}\left[f_{\text{conv}}^{1\times1}(F)\right], \tag{1}$$

$$W_2 = f_{\text{diconv}}^{3\times3}\left\{f_{\text{conv}}^{3\times1}\left\{f_{\text{conv}}^{1\times3}\left[f_{\text{conv}}^{1\times1}(F)\right]\right\}\right\}, \tag{2}$$

$$W_3 = f_{\text{diconv}}^{3\times3}\left\{f_{\text{conv}}^{1\times3}\left\{f_{\text{conv}}^{3\times1}\left[f_{\text{conv}}^{1\times1}(F)\right]\right\}\right\}, \tag{3}$$

$$Y = \text{Cat}(W_1, W_2, W_3)\oplus f_{\text{conv}}^{1\times1}(F), \tag{4}$$

式中：$f_{\text{conv}}^{1\times1}$、$f_{\text{conv}}^{1\times3}$、$f_{\text{conv}}^{3\times1}$ 和 $f_{\text{conv}}^{3\times3}$ 分别表示卷积核大小为 1×1、1×3、3×1 及 3×3 的常规卷积操作；$f_{\text{diconv}}^{3\times3}$ 表示扩张率为 5 的空洞卷积操作；Cat 表示特征图拼接操作；$\oplus$ 表示特征图按位相加操作；F 表示输入的特征图；W_1、W_2 及 W_3 表示前 3 路经过常规卷积及膨胀卷积后的特征图；Y 表示特征增强后的新特征图。

主干网络低层级特征图经过特征增强模块处理，在不损失分辨率的同时，提升了网络面对遥感图像中小目标遮挡、目标毗连等情况下获取小目标特征的能力。

2.2　特征融合模块

基于高低层级特征图所呈现的不同特性及小目标检测任务需要，本文重构高低层级特征融合结构，形成

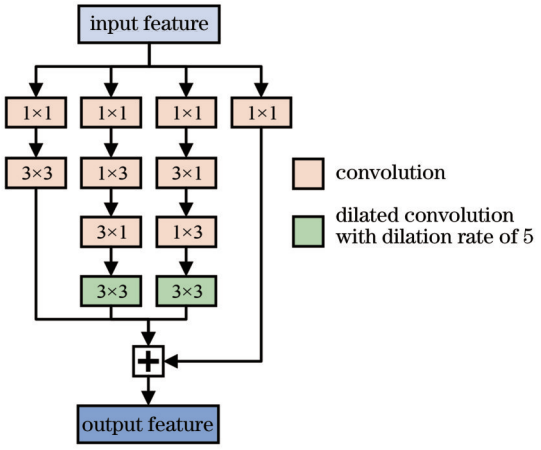

图 2　特征增强模块 FEM

Fig. 2　Feature enhancement module FEM

特征融合模块 FFM，其结构如图 3 所示。使用经特征增强模块 FEM 处理的低层级特征图 X_2(160×160)、X_3(80×80)与 SPP 模块处理的高层级特征图 X_4(40×40)作为特征融合模块 FFM 的输入，首先对特征图 X_4 进行两倍上采样操作，得到尺度与 X_3 相同的特征图并与之融合，对融合后的特征图利用级联注意力机制 ESM 进行处理，分别在通道维度和空间维度上聚合小目标特征，形成特征图 X_3'，随后在特征图 X_3' 基础上重复上述操作并形成特征图 X_2'，特征图 X_2' 与 X_3' 使网络实现小目标语义信息由深层向浅层流动。

网络在训练过程中，随着特征图尺寸不断减小，小目标特征在特征图中所占据的像素数也逐渐减少，如对应于尺寸为 32×32 的小目标，其经过多次卷积下采样后特征图中小目标信息消失殆尽，为避免在网络中引入冗余信息，干扰网络训练及模型收敛，在实现网络小目标信息由浅层向深层流动的过程中，仅对特征图 X_2' 进行卷积下采样处理，得到尺度与特征图 X_3' 相同的

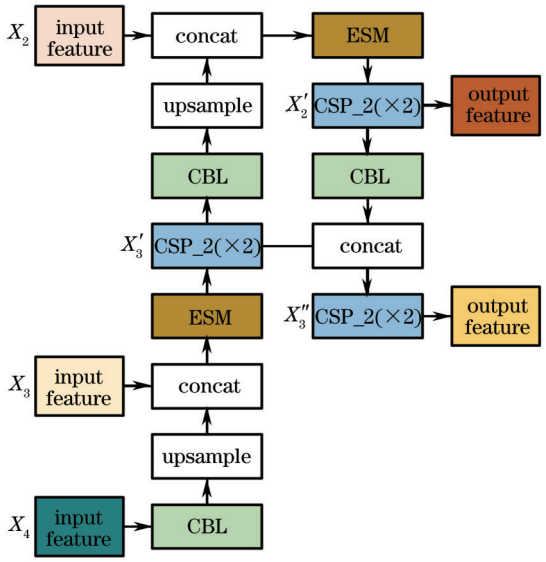

图 3　特征融合模块 FFM

Fig. 3　Feature fusion module FFM

特征图并与之融合形成特征图 X_3''，最终由特征图 X_2'（160×160）与特征图 X_3''（80×80）在网络输出端进行小目标检测。特征融合模块 FFM 计算过程可表示为

$$X_3'=f_{\mathrm{ESM}}\left\{\mathrm{Cat}\left[f_{\mathrm{up}}^{2\uparrow}(X_4),X_3\right]\right\},\tag{5}$$

$$X_2'=f_{\mathrm{ESM}}\left\{\mathrm{Cat}\left[f_{\mathrm{up}}^{2\uparrow}(X_3'),X_2\right]\right\},\tag{6}$$

$$X_3''=\mathrm{Cat}\left[f_{\mathrm{down}}^{2\downarrow}(X_2'),X_3'\right],\tag{7}$$

式中：$f_{\mathrm{up}}^{2\uparrow}$ 表示两倍上采样操作；$f_{\mathrm{down}}^{2\downarrow}$ 表示卷积下采样操作；Cat 表示特征图拼接操作；f_{ESM} 表示级联注意力机制 ESM；X_2、X_3 及 X_4 表示经主干网络选择用于特征融合的高低层级特征图，X_2' 和 X_3' 表示特征信息由高层级流向低层级所生成的特征图，其中 X_2' 也为输出端特征图，X_3'' 表示特征信息由低层级流向高层级所生成的特征图，为输出端特征图。

基于特征融合模块 FFM，网络对关键的小目标特征进行双向融合处理，在精准获取有效信息的同时，避免冗余信息参与网络学习，提升网络对小目标特征的关注能力。

2.3　注意力机制 ESM

注意力机制被广泛用于语音辨识、目标检测等诸多深度学习领域任务，其发展于人类视觉系统选择性、分层次处理信息的感知策略，在深度学习网络中，不仅能够通过网络训练记录并存储信息间的时序及位置关系，还能获取网络特征图不同通道维度及空间维度上的重要性差异，赋予不同权重，凸显重要特征。

本文使用有效通道注意力机制 E-ECA 与空间注意力机制 SAM 所组成的级联注意力机制 ESM，整体结构如图 4 所示。其中有效通道注意力机制 E-ECA 基于高效通道注意力机制[24]（Efficient Channel Network，ECA）进行改进，考虑到对输入特征图进行最大池化处理可以提取特征图上纹理特征等细节信息，进行平均池化处理可以提取特征图上有效背景信息，即与目标相关的上下文特征，故对输入特征图同时进行上述两种方式处理，对处理后的两路特征图进行按位相加操作，并沿用 ECA-Net 中的局部跨通道交互策略和自适应一维卷积结构，通过网络学习得到对应于特征图各通道上的不同权重，获取特征图在通道维度上更为精准的注意力信息。通道注意力机制 E-ECA 的计算过程可表示为

$$X'=\mathrm{AvgPool}(X)\oplus\mathrm{MaxPool}(X),\tag{8}$$

$$\omega_c=\sigma\left[\mathrm{CID}_k(X')\right],\tag{9}$$

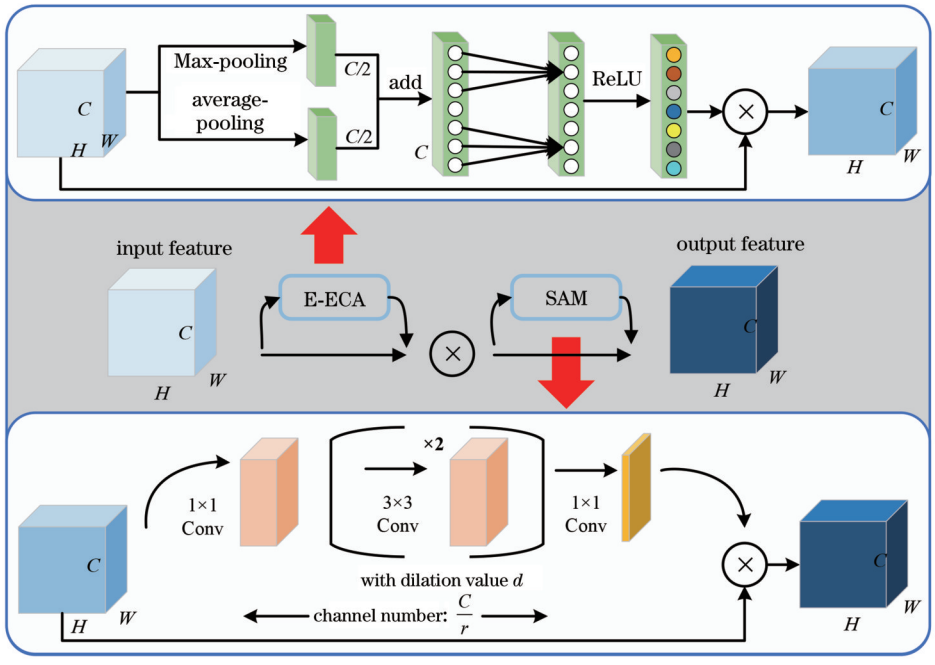

图 4　注意力机制模块 ESM

Fig. 4　Attention module ESM

$$y_c = \omega_c \otimes X, \tag{10}$$

式中：AvgPool表示平均池化操作；MaxPool表示最大池化操作；$\oplus$表示特征图按位相加操作；σ表示ReLU激活函数；CID_k表示核大小为k的一维卷积操作，k值随网络训练自适应变化；$\otimes$表示特征图按位相乘操作；ω_c表示通道注意力机制学习到的权重；X表示输入特征图；X'表示池化操作后的特征图；y_c为经过通道注意力机制后的输出特征图。

空间注意力机制SAM则引入瓶颈注意模块[25](BAM)中相关部分，首先对输入特征图采用1×1卷积降低通道数，在通道维度上进行压缩聚合，处理后的特征图经过多个3×3级联空洞卷积模块，并进一步压缩通道数，精炼特征，最后采用1×1卷积并将特征图通道数压缩为1，获取权重并对输入特征图进行处理，得到输出特征图，其融合多重感受野，小目标特征更加丰富。对于经过通道注意力机制E-ECA处理后的特征图，SAM的计算过程可表示为

$$y' = f_{\mathrm{conv}}^{1\times1}(y_c),\ y'' = f_{\mathrm{diconv}}^{3\times3}\left[f_{\mathrm{diconv}}^{3\times3}(y')\right],\ y''' = f_{\mathrm{conv}}^{1\times1}(y''), \tag{11}$$

$$\omega_s = \mathrm{BatchNorm}(y'''), \tag{12}$$

$$Y = \omega_s \otimes y_c, \tag{13}$$

式中：BatchNorm表示批量归一化操作；y'和y''分别表示进行常规卷积与空洞卷积后的特征图，特征图经过这两步处理后，其通道数大大压缩，通道压缩率设置为16；ω_s表示空间注意力机制学习到的权重；$\otimes$表示特征图按位相乘操作；Y为经过通道注意力与空间注意力级联机制后的输出特征图。

本文将注意力机制嵌入如图3所示的特征融合模块FFM，以增强模型提取小目标特征的能力，提升模型训练效率及鲁棒性。

3 实验结果与分析

3.1 实验平台

本文实验基于Windows 10操作系统，深度学习环境搭载于CUDA11.3及Pytorch1.10.2框架，使用NVIDIA RTX3080Ti GPU加速模型进行训练。根据YOLOv5算法设计增加网络宽度和深度，网络可依次分为YOLOv5n、YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x等5个模型，本文使用YOLOv5m作为实验基础模型，并在其上进行网络改进与优化。

3.2 数据集USOD简介

目前，虽然有一些用于实现遥感图像小目标检测任务的公开数据集，如MS COCO[26]中均包含一部分小尺度目标，但是从图像整体来看，小目标样本数在目标总数量中占比较低，而大、中尺度目标样本数占比较高，导致数据集中正负样本分布不均；如DOTA[27]中小目标样本数占比有所提高，但目标在部分图像上的分布十分密集，甚至有上千个目标，目标在其他部分图像上的分布则较为稀疏，同样存在分布不均的情况。因此，本文基于美国空军实验室(AFRL)所发布的UNICORN 2008[28]，对其进行筛选、图像分割、人工补充标注，形成Unicorn小目标数据集USOD，用于实现遥感图像车辆小目标检测。

UNICORN 2008数据集提供由大型光电传感器以约2 frame/s的速度对于同一区域的成像数据，成像区域大小约为5 km×5 km，空间分辨率约为0.4 m，共包含6471张可见光遥感图像，原始图像大小为8000×12000左右，仅提供图像中部分目标标注信息，集中为车辆小目标。

原始图像尺寸过大则难以进行深度学习模型训练，故选取部分成像质量较好、场景不同的图像进行分割并导入官方提供的标注信息，如图5所示，分割后的小图尺寸为640×640，仅生成少量标注框，故需对分割后图像进行补充标注。

基于图像中已有的标注信息，使用标注软件LabelImg对分割后小图像进行人工补充标注，车辆小目标统一命名为Vehicle，图6所示为人工补充标注示例。对完成人工标注后的图像进行修正与筛选，最终得到包括停车场、公路、临近房屋等不同场景下的图像共3000张，按照7∶3比例进行分配，得到训练集图像2100张，验证集图像900张，以此构建遥感图像小目标数据集USOD。

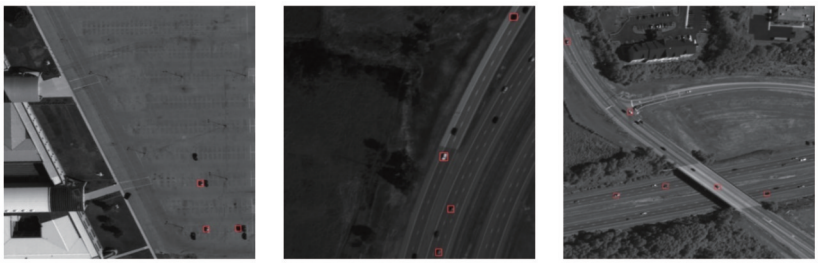

图5 图像官方标注示例

Fig. 5 Examples of image offical annotation

为验证所构建遥感图像小目标数据集USOD的合理性，对数据集中目标x方向和y方向的尺度进行统计，图7为所有目标标注框的尺寸分布图，目标总数量为43378个，尺寸在16×16及以内的目标数量占总数的96.3%，尺寸在32×32及以内的目标数量占总数的99.9%；图8所示为小目标样本数在USOD训练集图像中的分布情况，单张图像小目标样本数在50个及以内的情况占88.2%，小目标样本数在100个及以内的

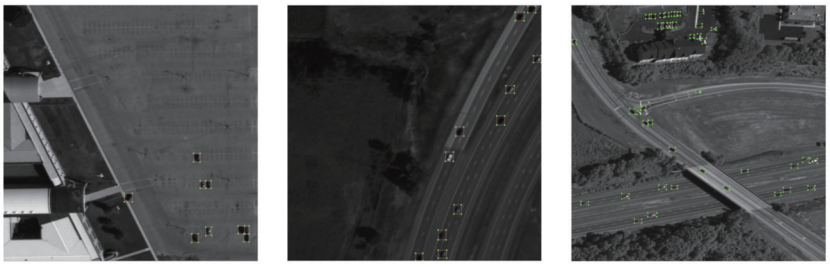

图 6　图像人工补充标注示例

Fig. 6　Examples of manual supplementary annotation

情况占 98.9%，即小目标在该数据集中的分布较均匀，未出现某些图像小目标数量极大的情况。综上，将该构造的数据集 USOD 用于验证小目标检测算法性能较合适。

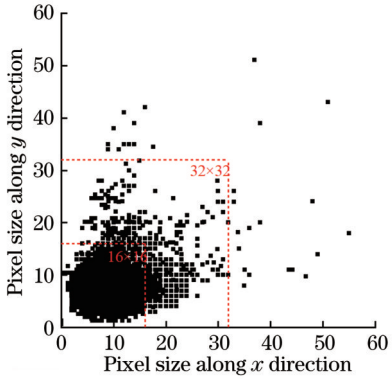

图 7　小目标数据真实框尺寸分布图

Fig. 7　Distribution of small object ground truth box size

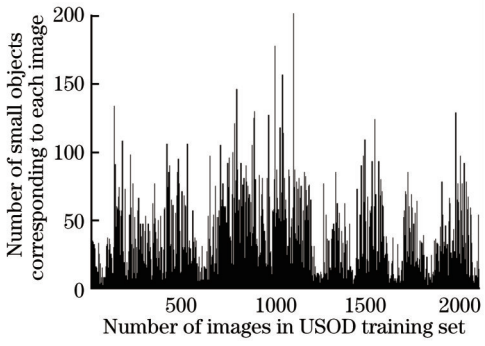

图 8　小目标样本数目分布图

Fig. 8　Distribution of numbers of small objects

3.3　模型训练及评价指标

本文实验将小目标数据集按照 7∶3 的比例划分为训练集和验证集，图像输入网络时采用 Mosaic 增强操作，初始学习率设置为 0.01，采用 SGD 优化器对总损失函数进行优化训练，训练批大小设置为 16，动量参数为 0.937，训练 200 轮次，采用 Early Stopping 方法优化训练过程，防止出现过拟合情况。

本文实验使用深度学习中常用的评价指标，即精确率（P）、召回率（R）和平均精度（AP，可用 A_P 表示）。令 N_{TP} 为正确检出的小目标数目，N_{FP} 为虚警情况下的小目标数目，N_{FN} 为漏检情况下的小目标数目。

精确率 P 的计算公式为

$$P=\frac{N_{TP}}{N_{TP}+N_{FP}}。\tag{14}$$

召回率 R 的计算公式为

$$R=\frac{N_{TP}}{N_{TP}+N_{FN}}。\tag{15}$$

平均精度 A_P 的计算公式为

$$A_P=\int_0^1 P(R)\mathrm{d}R,\tag{16}$$

式中：$P(R)$ 为精确率-召回率曲线。IoU 为检测框与真实框之间的交并比阈值，本实验采用 IoU 为 0.5 时的 AP 和 IoU 为 0.5∶0.95 时的 AP 作为评判指标。

3.4　消融实验结果

为验证本文针对 FFAM-YOLO 算法所提出的各改进模块对遥感图像小目标检测的影响，逐一对各个模块进行评估，评估结果如表 1 所示。表格中“×”表示未添加相应模块，“√”表示添加相应模块，加粗字体表示指标结果最优。可以看出，特征增强与特征融合方式对小目标检测指标提升有较大贡献，这主要是因为原始算法主干网络对小目标的特征提取能力不足，从主干网络流向高低层级特征融合网络的特征有效性降低，使得融合后的小目标特征不明确，造成模型检测性能下降。引入注意力机制使网络对小目标的检测结果更加准确，聚焦小目标显著特征，抑制背景噪声干扰，使得模型最终的检测精度更高。

表 1　算法各模块性能评估结果

Table 1　Evaluation results of each module in proposed algorithm

FEM&FFM	ESM	Algorithm	P /%	R /%	AP for IoU of 0.5 /%	AP for IoU of 0.5∶0.95 /%
×	×	YOLOv5	90.3	81.2	87.1	31.6
√	×	FFAM-YOLO	91.4	83.4	88.2	32.0
√	√	FFAM-YOLO	**91.9**	**83.5**	**89.0**	**32.6**

图 9 展示了 YOLOv5 算法与 FFAM-YOLO 算法在行驶场景、停泊场景的检测结果。行驶场景中，小目标清晰且分布稀疏，三种算法均成功检出小目标；停泊场景中，小目标分布较密集，背景中存在与目标特征接近的地标、屋角等物，YOLOv5 算法存在虚警，FFAM-YOLO 算法经 FEM 和 FFM 削弱背景噪声的影响，消除虚警，经 ESM 提升检测结果置信度。

图 10 展示了 YOLOv5 算法与 FFAM-YOLO 算法在光照度低场景、阴影遮挡场景的检测结果。光照度低场景中，小目标与背景区分不明显，YOLOv5 算法存在漏检，FFAM-YOLO 算法经 FEM 和 FFM 增强并捕获小目标特征，经 ESM 聚焦特征并成功检出目标；阴影遮挡场景中，小目标部分处于树阴，YOLOv5 算法存在漏检，FFAM-YOLO 算法均可将其成功检出。

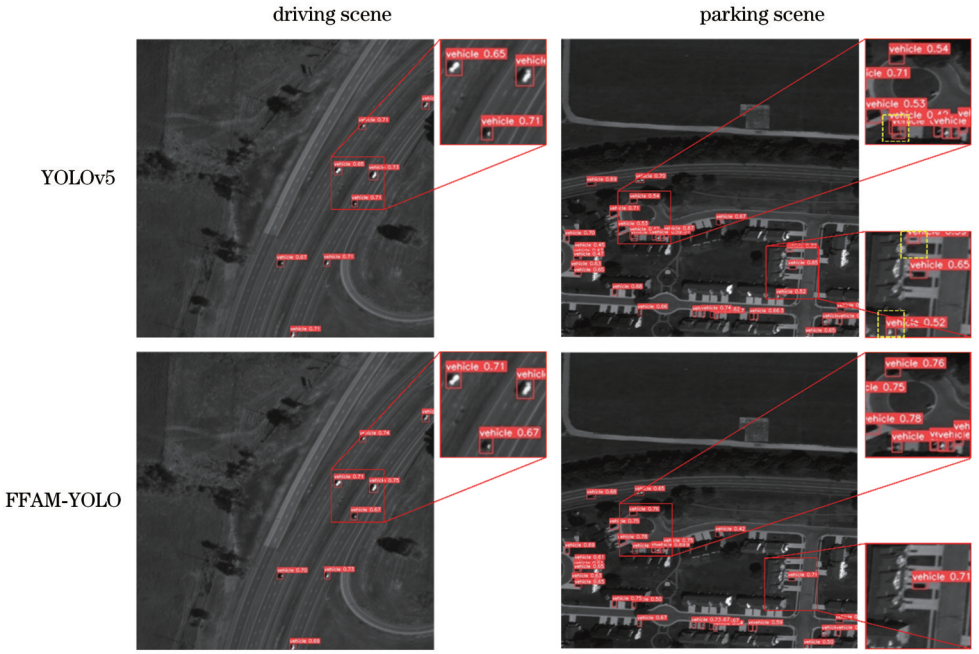

图 9　YOLOv5 与 FFAM-YOLO 对行驶场景及停泊场景的检测结果

Fig. 9　Detection results of YOLOv5 and FFAM-YOLO in driving scene and parking scene

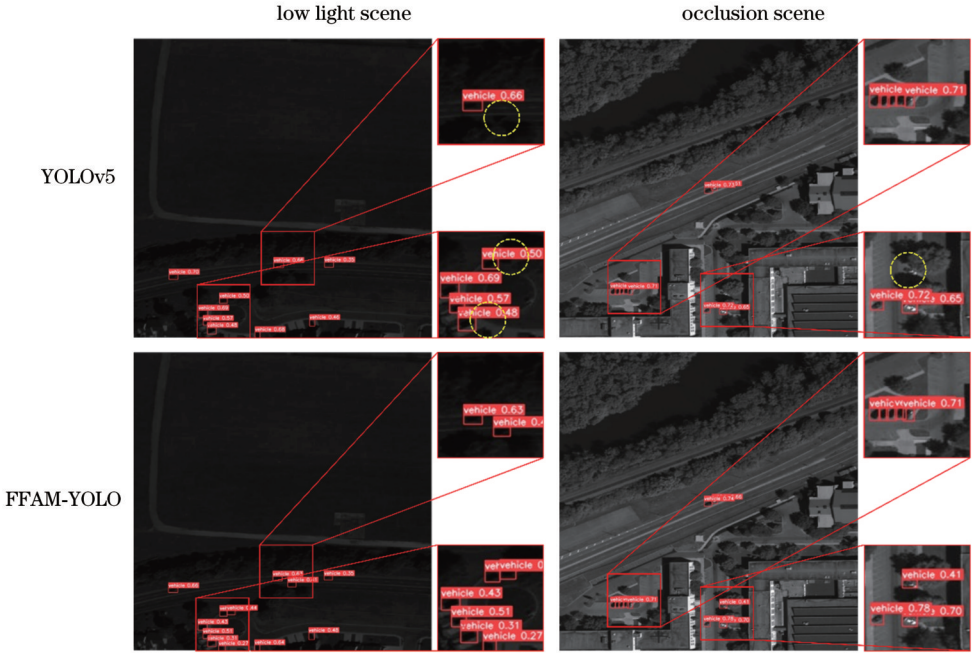

图 10　YOLOv5 与 FFAM-YOLO 对光照度低场景及阴影遮挡场景的检测结果

Fig. 10　Detection results of YOLOv5 and FFAM-YOLO for low light scene and occlusion scene

FFAM-YOLO 算法中引入 FEM 进行主干网络特征增强，针对小目标特性重构高低层级特征融合结构 FFM，采用 ESM 提升小目标特征有效性，因此相比于 YOLOv5 算法能够较好地应对密集排布、光照度低、

阴影遮挡等复杂场景中的小目标检测问题。通过深入分析，FFAM-YOLO算法的误判情况主要集中于如图11所示的复杂场景中，这些场景大多存在光照度低且阴影遮挡的情况，其原因是网络进行特征提取时从目标小范围获取的信息有限，使用FEM模块可提升感受野并增强特征，而该操作同时引入目标与背景或其他目标之间的上下文信息，极端复杂场景下目标与背景过于相似，上下文信息作为信息噪声被引入，干扰算法对小目标的判别。

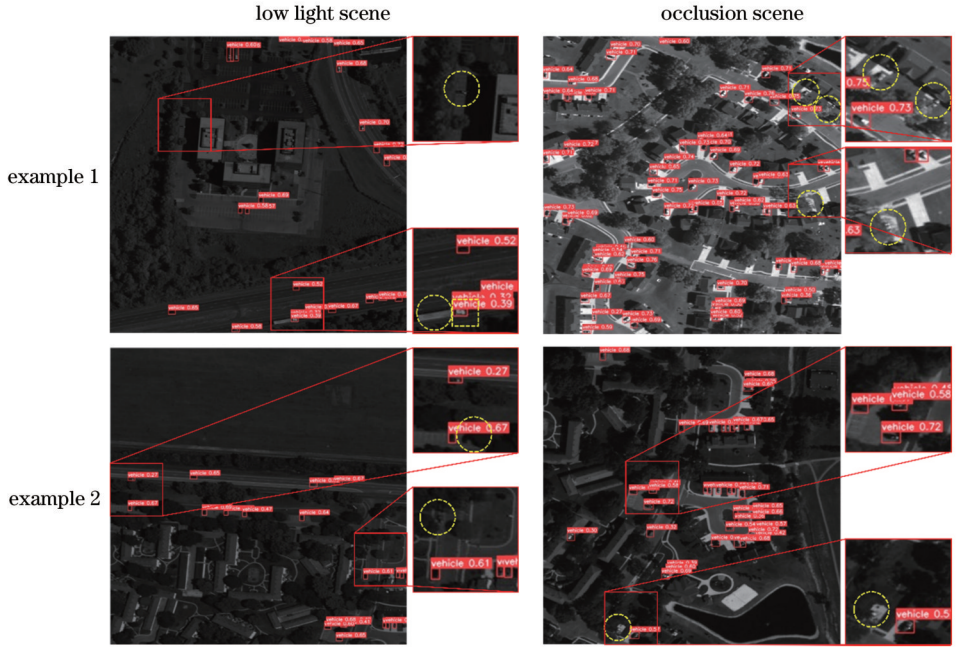

图11　FFAM-YOLO检测误判情况

Fig. 11　FFAM-YOLO detection misjudgment situation

3.5　对比实验结果

为进一步探究FFAM-YOLO算法性能，利用相同的小目标数据集进行验证，将FFAM-YOLO与原始YOLOv5算法进行对比，并将其与其他主流的单阶段目标检测算法如DSSD[29]、RefineDet[30]、YOLOv3及YOLOv4在检测精度和检测速率方面进行比较，结果如表2所示，表格中加粗字体表示指标结果最优。由表2可以得知，FFAM-YOLO算法在遥感图像小目标数据集上的检测精度最高。此外，算法在主干网络中构建特征增强模块，在高低层级特征融合网络中引入注意力机制，因此算法对小目标特征提取的能力得以改善，算法的检测速率有所下降。整体上来看，FFAM-YOLO在检测速率方面仍然具有很大的优势，满足实时检测需求。

表2　本文算法与其他5种算法的性能比较结果

Table 2　Comparison of detetion results of FFAM-YOLO and other five algorithms

Algorithm	Backbone	P /%	R /%	AP for IoU of 0.5 /%	AP for IoU of 0.5:0.95 /%	Detection rate /(frame·s^{-1})
DSSD	ResNet101	64.6	57.4	53.1	16.3	34.3
YOLOv3	DarkNet53	72.0	69.1	57.4	18.5	61.1
YOLOv4	DarkNet53	81.0	82.0	77.8	28.8	61.2
RefineDet	ResNet101	88.1	82.3	79.1	31.4	31.3
YOLOv5	CSPDarkNet53	90.3	81.2	87.1	31.6	**181.9**
FFAM-YOLO	CSPDarkNet53	**91.9**	**83.5**	**89.0**	**32.6**	120.0

如图12所示，选取包括行驶场景、停泊场景、小目标稀疏分布及密集分布情况在内的几个遥感图像样本对DSSD、RefineDet、原始YOLOv5算法与本文改进算法FFAM-YOLO的检测结果进行对比，图中虚线圆圈处为小目标漏检情况，虚线方框处为虚警情况。将各算法的检测结果对标真值标注情况，发现各算法均能有效检测出遥感图像中的车辆小目标，但从细节处看，DSSD算法对于尺寸非常小、特征极难提取情况下的小目标难以实现定位，从而造成漏检，示例①～示例③均存在该情况；RefineDet算法的检测效果略优于

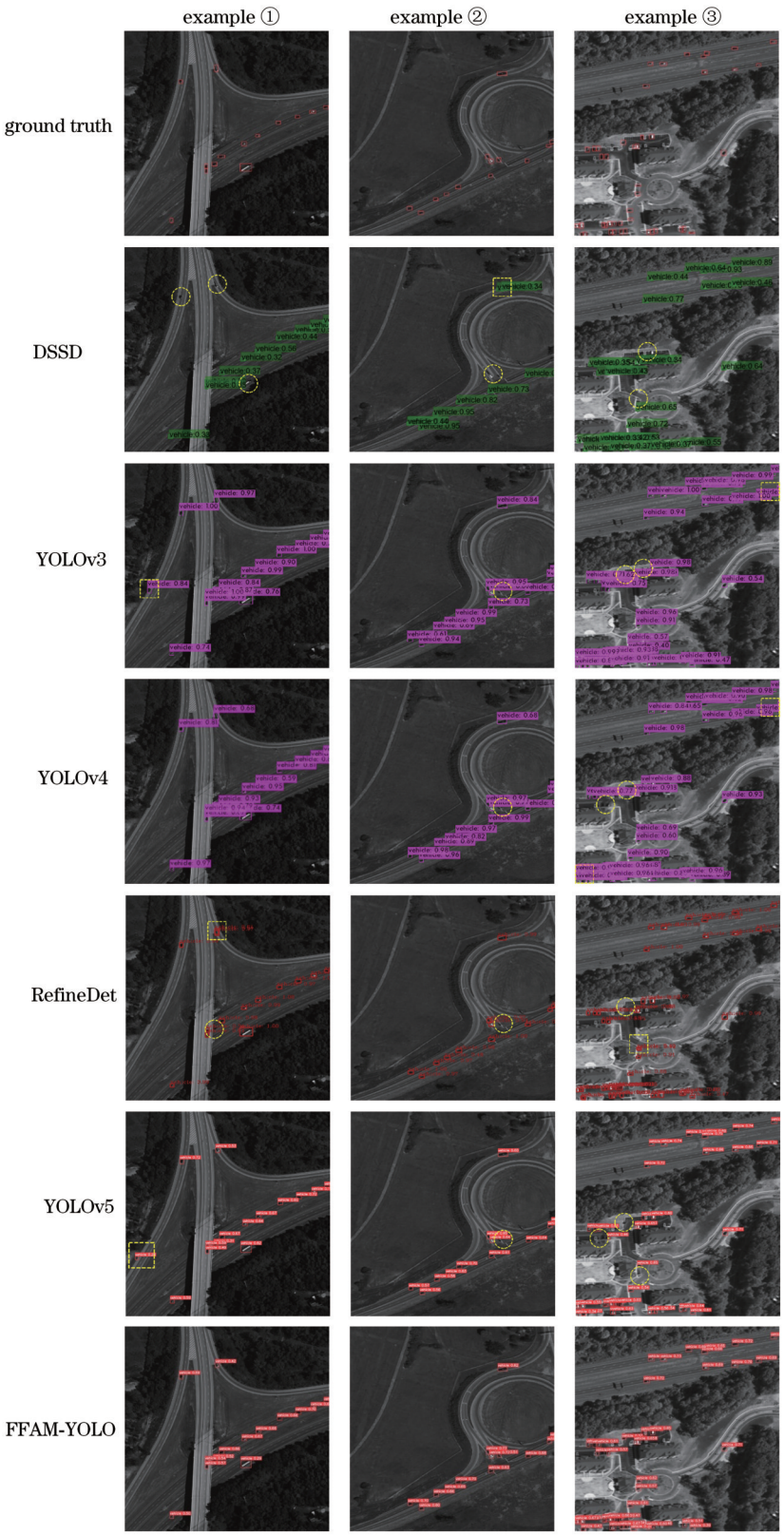

图 12　FFAM-YOLO对比实验的检测结果

Fig. 12　Detection results of contrast experiments using FFAM-YOLO

DSSD，也会出现如示例①中虚警以及示例②和示例③中的漏检情况；原始YOLOv5算法的主干网络对小目标特征的提取不充分，小目标容易受到背景噪声的干扰，出现如示例①中虚警以及示例②和示例③中的漏检情况；本文改进的FFAM-YOLO算法在上述方法的基础上提升了对遥感图像小目标的检测率，尤其是在密集分布场景能有效区分目标、降低漏检率，说明针对小目标检测任务引入构建的特征增强模块和注意力

机制模块、重构特征融合结构，能够使网络更准确定位到遥感图像中的小目标。

4 结 论

针对遥感图像中小目标所占像素数少、特征不明显、与背景区分困难等问题，提出一种基于特征融合与注意力机制的遥感图像小目标检测算法FFAM-YOLO。所提算法通过在主干网络中针对低层级特征图构造特征增强模块FEM，提高模型的特征提取能力，增强特征图的语义信息，并使用重构的高低层级特征融合结构FFM改善网络对小目标的表征能力；在由通道注意力机制E-ECA与空间注意力机制SAM组成的级联注意力机制ESM作用下，凸显对小目标检测有利的融合特征。同时，为解决小目标在数据集中占比小且分布不均的问题，构建遥感图像小目标数据集USOD，用于验证小目标检测算法性能。在小目标数据集上USOD的实验结果表明，本文算法的性能相较于其他算法有一定的提升，性能指标查准率为91.9%，查全率为83.5%，IoU为0.5时的平均精度为89%，IoU为0.5∶0.95时的平均精度为32.6%，算法的准确性和实时性均得到保障。

在小目标检测任务中，单一源图像对小目标的特性表征信息量少，难以提升算法精度，利用多源遥感图像实现融合检测并提升精度是研究的重点方向。此外，小目标检测应用场景趋于多样化，更多需求是在资源及空间受限的设备终端上部署高性能的小目标检测模型，下一步研究将针对小目标的检测网络进行轻量化，推动模型落地。

参 考 文 献

[1] 高新波, 莫梦竟成, 汪海涛, 等. 小目标检测研究进展[J]. 数据采集与处理, 2021, 36(3): 391-417.
Gao X B, Mo M J C, Wang H T, et al. Recent advances in small object detection[J]. Journal of Data Acquisition and Processing, 2021, 36(3): 391-417.

[2] 段仲静, 李少波, 胡建军, 等. 深度学习目标检测方法及其主流框架综述[J]. 激光与光电子学进展, 2020, 57(12): 120005.
Duan Z J, Li S B, Hu J J, et al. Review of deep learning based object detection methods and their mainstream frameworks[J]. Laser & Optoelectronics Progress, 2020, 57(12): 120005.

[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] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2980-2988.

[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] 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.

[8] 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.

[9] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08) [2021-03-02]. https://arxiv.org/abs/1804.02767.

[10] Bochkovskiy A, Wang C, Liao H. Yolov4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2021-02-03]. https://arxiv.org/abs/2004.10934.

[11] Glenn J. YOLOv5[EB/OL]. [2021-02-03]. https://github.com/ultralytics/yolov5.

[12] 鞠默然, 罗江宁, 王仲博, 等. 融合注意力机制的多尺度目标检测算法[J]. 光学学报, 2020, 40(13): 1315002.
Ju M R, Luo J N, Wang Z B, et al. Multi-scale target detection algorithm based on attention mechanism[J]. Acta Optica Sinica, 2020, 40(13): 1315002.

[13] 赵斌, 王春平, 付强, 等. 基于深度注意力机制的多尺度红外行人检测[J]. 光学学报, 2020, 40(5): 0504001.
Zhao B, Wang C P, Fu Q, et al. Multi-scale infrared pedestrian detection based on deep attention mechanism[J]. Acta Optica Sinica, 2020, 40(5): 0504001.

[14] Kasper-Eulaers M, Hahn N, Berger S, et al. Short communication: detecting heavy goods vehicles in rest areas in winter conditions using YOLOv5[J]. Algorithms, 2021, 14(4): 114.

[15] Yan B, Fan P, Lei X Y, et al. A real-time apple targets detection method for picking robot based on improved YOLOv5[J]. Remote Sensing, 2021, 13(9): 1619.

[16] Liu T, Zhou B J, Zhao Y S, et al. Ship detection algorithm based on improved YOLO V5[C]//2021 6th International Conference on Automation, Control and Robotics Engineering (CACRE), July 15-17, 2021, Dalian, China. New York: IEEE Press, 2021: 483-487.

[17] 孙迎春, 潘树国, 赵涛, 等. 基于优化YOLOv3算法的交通灯检测[J]. 光学学报, 2020, 40(12): 1215001.
Sun Y C, Pan S G, Zhao T, et al. Traffic light detection based on optimized YOLOv3 algorithm[J]. Acta Optica Sinica, 2020, 40(12): 1215001.

[18] Benjumea A, Teeti I, Cuzzolin F, et al. YOLO-Z: improving small object detection in YOLOv5 for autonomous vehicles[EB/OL]. (2021-12-22) [2022-01-02]. https://arxiv.org/abs/2112.11798.

[19] Lian J, Yin Y H, Li L H, et al. Small object detection in traffic scenes based on attention feature fusion[J]. Sensors, 2021, 21(9): 3031.

[20] Zhan W, Sun C F, Wang M C, et al. An improved

Yolov5 real-time detection method for small objects captured by UAV[J]. Soft Computing, 2022, 26(1): 361-373.

[21] Kim M, Jeong J, Kim S. ECAP-YOLO: efficient channel attention pyramid YOLO for small object detection in aerial image[J]. Remote Sensing, 2021, 13(23): 4851.

[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] Liu S T, Huang D, Wang Y H. Receptive field block net for accurate and fast object detection[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11215: 404-419.

[24] Wang Q L, Wu B G, Zhu P F, et al. ECA-net: efficient channel attention for deep convolutional neural networks [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11531-11539.

[25] Park J, Woo S, Lee J Y, et al. BAM: bottleneck attention module[EB/OL]. (2018-07-17) [2021-02-05]. https://arxiv.org/abs/1807.06514.

[26] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: common objects in context[M]//Fleet D, Pajdla T, Schiele B, et al. Computer vision-ECCV 2014. Cham: Springer, 2014, 8693: 740-755.

[27] Xia G S, Bai X, Ding J, et al. DOTA: a large-scale dataset for object detection in aerial images[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: 3974-3983.

[28] Leong C, Rovito T, Mendoza-Schrock O, et al. Unified coincident optical and radar for recognition (UNICORN) 2008 dataset[EB/OL]. [2020-04-15]. https://github.com/AFRL-RY/data-unicorn-2008.

[29] Fu C Y, Liu W, Ranga A, et al. DSSD: deconvolutional single shot detector[EB/OL]. (2017-01-23)[2021-02-04]. https://arxiv.org/abs/1701.06659.

[30] Zhang S F, Wen L Y, Bian X, et al. Single-shot refinement neural network for object detection[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: 4203-4212.