

基于特征重加权的小样本遥感图像目标检测算法

周 博, 葛洪武, 李 珩, 李 旭

(中国电子科技集团公司 第 54 研究所, 石家庄 050081)

摘要: 针对遥感图像具有目标尺度多变、目标模糊、背景复杂的特点, 提出了一种基于特征重新加权的遥感小样本目标检测算法 RE-FSOD; 该模型包括 3 部分: 元特征提取器、特征重加权提取器、预测模块, 其中元特征提取器由 CSPDarknet-53、FPN 以及 PAN 构成, 负责提取数据的元特征; 特征重加权提取器用于生成特征重加权向量, 用于调整元特征来强化对于检测新类有帮助的特征; 预测模块由 YOLOv3 的预测模块构成, 在此基础上将定位损失函数替换为 CIOU 损失函数, 提升模型的定位精度; 最后在 NWPU VHR-10 遥感数据集上进行了训练和测试, 实验结果表明, 该方法相较于基线方法 FSODM 的在 3-shot、5-shot、10-shot 情况下分别提升了约 19%、11%、8%。

关键词: 小样本目标检测; YOLO; 迁移学习; 特征重加权; 注意力机制

Few-shot Object Detection on Remote Sensing Images Based on Feature Reweighting

ZHOU Bo, GE Hongwu, LI Heng, LI Xu

(The 54 th Research Institute, China Electronics Technology Group Corporation, Shijiazhuang 050081, China)

Abstract: Aimed at the characteristics of variable target size, fuzzy target, and complex background, a few-shot object detection model based on feature reweighting is proposed. The model consists of three parts of element feature extractor, feature reweighting extractor and prediction module. The element feature extractor is composed of CSPDarknet-53, feature pyramid network (FPN) and path aggregation network (PAN), which is responsible for extracting the element features of data. The feature reweighting extractor is used to generate the feature reweighting vectors, and adjust the element features to enhance the helpful features for detecting new classes. The prediction module is composed of the prediction module of YOLOv3. On this basis, the positioning loss function is replaced by CIOU to improve the positioning accuracy of the model. Finally, the training and testing are carried out on the NWPU VHR-10 remote sensing data set. The experimental results show that compared with the baseline method FSODM, the mean average precision (mPA₅₀) improves by 19%, 11% and 8% at the conditions of 3-shot, 5-shot and 10-shot respectively.

Keywords: fewshot object detection; YOLO; transfer learning; feature reweighting; attention mechanism

0 引言

目标检测的任务是找出图像或视频中人们感兴趣的物体, 并且同时检测出它们的位置和大小, 目标检测不仅要解决分类问题, 还要解决定位问题, 作为计算机视觉的基本问题之一, 目标检测构成了许多其他视觉任务的基础, 例如实例分割、图像标注和目标跟踪等。目标检测算法可以分为一阶段目标检测算法以及二阶段目标检测算法, 其中一阶段目标检测算法主要包含 YOLO^[1-3]、SSD^[4]等算法; 二阶段目标检测算法主要包括 RCNN^[5]、Fast RCNN^[6]、Faster RCNN^[7]等算法。一阶段目标检测算法减少了空间和时间的占用, 速度有较大提升, 但精度低于二阶段目标检测算法。上述算法的性能都依赖于大量的数据集, 但是在某些场景下, 往往难以收集到大量的数据集或者获取数据集的代价非常高。

受限于大规模的数据集, 小样本目标检测算法在计算机视觉领域得到了广泛的研究; 小样本目标检测算法旨在

学习可转移的知识, 将这些知识泛化到新类中, 从而对只有少量样本的新类进行检测。Wang 等人基于 Faster RCNN 提出了两阶段微调方法^[8] (TFA, two-stage fine-tuning approach), 在保持整个特征提取器不变的情况下, 只微调分类器和回归器, 首次证明了简单的迁移学习方法在小样本目标检测任务上效果好于元学习方法。Sun 等人提出了基于对比候选框编码的小样本目标检测算法^[9] (FSCE, few-shot object detection via contrastive proposal encoding), 首次将对对比学习引入到小样本目标检测方法。Li 等人提出了基于类别特征均衡的小样本目标检测算法^[10] (CME, class margin equilibrium) 方法来优化特征空间划分和新类表示, 让模型更好地学习小样本的特征; Qiao 等人提出了基于解耦 Faster R-CNN 的小样本目标检测算法^[11] (DeFRCN, decoupled faster r-CNN for few-shot object detection) 方法, 缓解了 Faster RCNN 中前景和背景的识别冲突问题, 还提出了一个离线的分数校准模块, 以缓解定位和回归之间的冲突, 但是该模块增加了模型的计算量。Kang 等人以

收稿日期: 2023-10-12; 修回日期: 2023-11-05。

基金项目: 国防基础科研计划资助(JCKY2020210B021)。

作者简介: 周 博(1998-), 男, 硕士研究生。

引用格式: 周 博, 葛洪武, 李 珩, 等. 基于特征重加权的小样本遥感图像目标检测算法[J]. 计算机测量与控制, 2024, 32(2): 283-290.

YOLOv2 为基础网络提出了元学习小样本目标检测算法 (MetaYolo^[12], few-shot object detection via feature reweighting), 在此基础上引入了特征重加权提取器, 用于强化对于检测新类有帮助的元特征, Li 等人以 YOLOv3 为基础网络, 提出了遥感小样本目标检测算法^[13] (FSODM, few-shot object detection model), 通过引入多尺度的特征重加权提取器来提升模型对于不同尺度目标的检测性能。

遥感图像具有目标尺度差异大、目标模糊、背景复杂度高特点, 上述小样本目标检测算法大多基于日常生活中的常见物体开发, 在遥感图像目标检测任务上精度仍然较低, 并且大多基于二阶段的 Faster RCNN 算法, 计算复杂度高, 检测速度较慢。其中 FSODM 以一阶段的 YOLOv3 为基础, 针对遥感图像提出了多尺度的特征重加权提取器, 提高了不同尺度目标的检测精度, 并且有较快的检测速度, 但是 FSODM 提取到的图像的深层语义信息以及浅层定位信息仍然不够丰富。针对上述问题, 以 FSODM 为基础, 提出了一种新的特征重加权小样本目标检测算法 (RE-FSOD, feature reweighting few-shot object detection), 贡献主要如下:

1) 将元特征提取器的骨干网络更换为 CSPDarknet-53 结构, 在 Neck 部分加入路径聚合网络^[14] (PAN, path aggregation network) 结构, CSPDarknet 相对于初始的 Darknet-53 引入了交叉跨阶段网络^[15] (CSP, cross stage partial network) 结构, 并且将空间金字塔池化^[16] (SPP, spatial pyramid pooling) 更换为快速空间金字塔池化 (SPPF, spatial pyramid pooling fast), 能够减少参数量以及计算复杂度, 并且提供了更强大的特征表示能力。PAN 结构能够将浅层丰富的定位信息传递到深层, 增强对不同尺度检测目标的定位能力, 使得在 3-shot、5-shot、10-shot 情况下分别提升了大约 10%、3%、4%。

2) 提出了结合注意力机制和残差结构的 C2fSE 模块, 由压缩激励注意力机制^[17] (SE, squeeze and excitation) 和带有残差结构^[18]的通道到像素模块 (C2f, channel to pixel) 构成, 将其添加到原始的特征重加权提取器卷积层后, 增加网络的深度和感受野, 提升网络的语义特征提取能力, 使得在 3-shot、5-shot、10-shot 情况下分别提升了大约 4%、3%、4%。

3) 将定位损失函数替换为 CIUO^[19] 损失函数, CIUO 损失函数同时考虑了检测目标的大小、长宽比、纵横比等因素, 加速模型的收敛, 且提升了模型的定位性能, 使得在 3-shot、5-shot、10-shot 情况下分别提升了大约 6%、9%、3%。

1 RE-FSOD 算法

1.1 小样本目标检测问题设置

小样本目标检测算法旨在从源数据集 D_s 中学习通用的元知识, 利用少量的目标数据集 D_t 将元知识迁移到目标任务中, 使得小样本目标检测算法在目标数据集上能够快速收敛。假设目标数据集中有 N 类样本, 每类样本具有 K 个

标签, 称为 N -way- K -shot^[20] 任务。

小样本目标检测算法的数据集由支持集与查询集构成, 给定一个 k -shot 任务, 支持集 $S_i = \{(I_k, M_k)\}$, 其中 I_k 代表输入图像并且 $I_k \in R^{h \times w \times 3}$, M_k 代表对应目标的掩膜, $k = 1, 2, 3, \dots, K$ 。查询集 Q_i 包含 N_q 张图片, N_q 为训练集或者测试集中所有图片的数目, 同样包含有 k 类目标。RE-FSOD 算法的一次迭代训练的输入为 $T_i = \{Q_i, S_i\}$ 。

1.2 FSODM 介绍

FSODM 算法是 Li 等人于 2021 年提出的第一个小样本遥感目标检测算法, FSODM 算法基于 YOLOv3 目标检测算法进行开发, 在此基础上引入了多尺度的特征重加权模块, 用来调整特征图的特征, 强化那些对于检测新类有帮助的特征, 使得算法能够在少量样本下取得较好的检测精度。相较于其他流行的小样本目标检测算法 TFA、FSCE 等算法, FSODM 结合了多尺度的特征重加权模块, 能够更好地处理尺度多变的遥感目标, 并且 FSODM 算法基于一阶段的 YOLOv3 算法开发, 在推理阶段可以移除掉新增的特征重加权模块, 具有较少的计算量以及更快的推理速度, 但 FSODM 的检测精度仍然较低, 为了能够更准确地识别遥感目标, 仍需要对算法进一步改进, FSODM 结构如图 1 所示。

1.3 RE-FSOD 整体结构

RE-FSOD 算法以 FSODM 为基础, 与 FSODM 有相同的架构, 同样由元特征提取器、特征重加权提取器、预测模块 3 部分构成, 整体架构如图 2 所示, 详细结构如图 3 所示。

元特征提取器的输入为查询集中的图片, 能够提取 3 个不同尺度的元特征图, 特征重加权提取器为数据集的所有目标类别生成 3 个尺度的特定类特征重加权向量, 将对应尺度的特征重加权向量与元特征图进行 1×1 的通道卷积来调整元特征图的权重, 强化对于检测新类有帮助的特征, 将特征图输入到预测模块中生成 (x, y, w, h, o, c) , (x, y, w, h) 代表边界框坐标, o 代表置信度, c 代表类别分数。

2 RE-FSOD 关键技术

2.1 元特征提取器

元特征提取器旨在从查询集中提取到鲁棒的元特征, I 为输入到元特征提取器的查询图片, $I \in R^{h \times w \times c}$, 提取得到的元特征为 $F_i = e(I) \in R^{h_i \times w_i \times m_i}$, 用 i 代表 3 个尺度的序号, $i = 1, 2, 3$ 。其中 h_i, w_i, m_i 表示尺度 i 的特征图的大小。依照 FSODM 的设置, 将特征提取器的图片输入设置为 512×512 , 3 个特征图的尺寸大小分别为 $16 \times 16 \times 1024$ 、 $32 \times 32 \times 512$ 、 $64 \times 64 \times 256$ 。

FSODM 中元特征提取器由 Darknet-53 以及特征金字塔网络 (FPN, feature pyramid network) 构成。RE-FSOD 将元特征提取器的骨干网络替换为 CSPDarknet-53, 并且在 FPN 的基础上添加了 PAN 结构。Darknet-53 中存在大量的残差块, 并且使用步长为 2, 卷积核大小为 3×3 的卷积层 Conv2D 代替池化层进行下采样, 残差块可以增加网络的深度, 使得网络能够提取更高级的语义特征, 可以避免梯度的

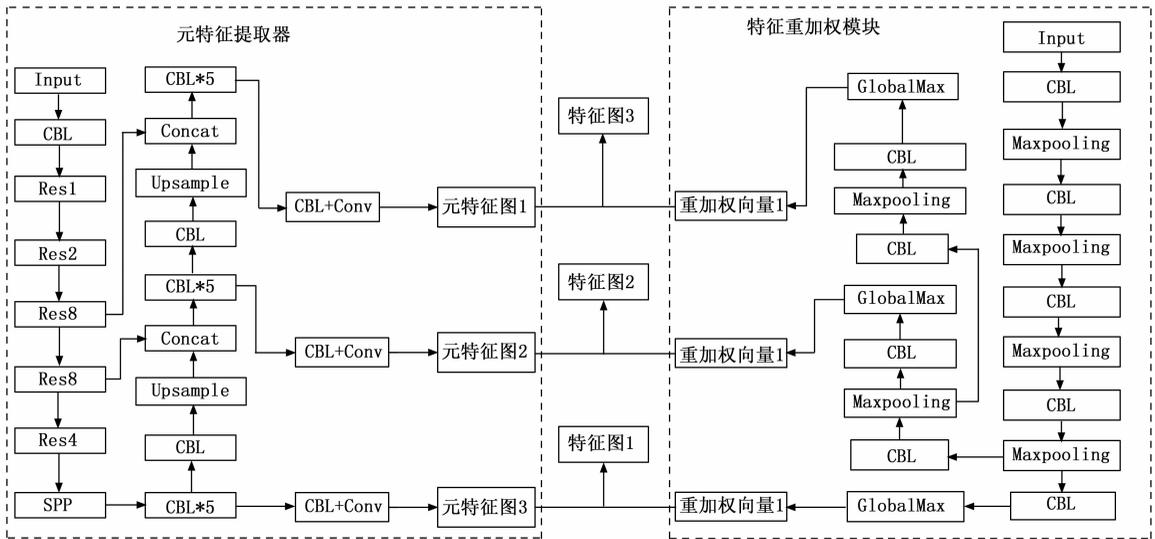


图 1 FSODM 网络结构图

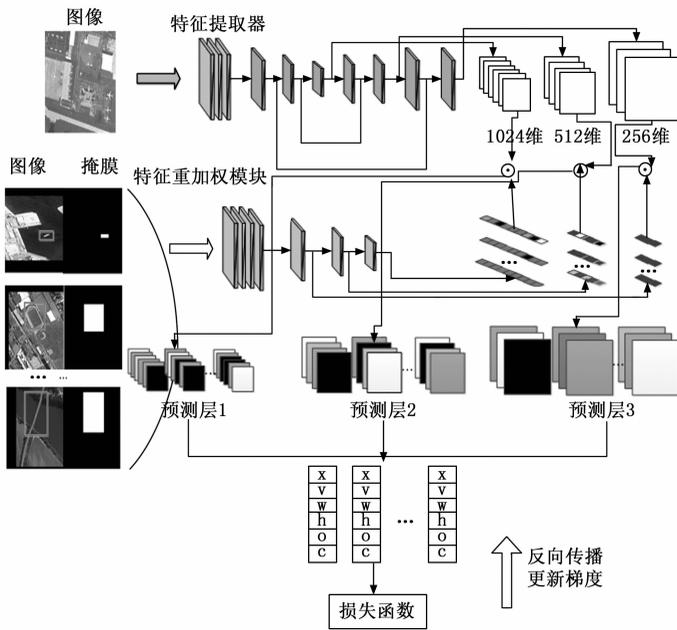


图 2 RE-FSOD 架构图

到不同大小的目标特征, SPPF 在 SPP 结构上进行了改进, 复用了已有的特征图, 并且将 3 个最大池化层的卷积核设置为 5, 大大提升了推理速度。

一般浅层的特征图具有更多定位信息和较少的语义信息, 深层的特征图尺寸变得更小, 维度变得更大, 因此深层的特征图具有更多的语义信息和较少的定位信息。在 FSODM 元特征提取器使用了 FPN 结构, FPN 是一个自顶向下的特征金字塔, 把浅层的语义特征传递下来, 对整个金字塔进行增强, 它只增强了语义信息, 但是没有对定位信息进行传递, PAN 结构针对 FPN 的缺点, 在 FPN 的基础上增加了一个自底向上的金字塔, 对 FPN 进行补充, 将定位特征传递上去, 这样生成的特征图同时拥有了丰富的定位和语义信息, 能够提升目标检测算法的精度。

2.2 特征重加权提取器

特征重加权提取器的输入是检测目标的感兴趣区域 (ROI), 如图 1 所示, 包含各个类的图片以及对应类别目标的掩膜, 一般输入图像会包含多个检测目标, 为了使特征重加权提取能够识别特定类别的目标, 因此只选取对应类别的一个目标, 根据此目标位置将边界框内的区域像素设置为 1 并且将其余位置的像素点设置为 0 得到掩膜。将输入图像与掩膜拼接成 4 维向量输入到特征重加权提取器中, 生成对应类别多个尺度的特征重加权向量, 特征重加权向量与元特征图进行 1×1 的卷积来调整元特征图, 强化对于检测新类有帮助的特征。

FSODM 中使用卷积层与最大池化层构成特征重加权网络, 为了使得特征重加权能够提取到更丰富的语义信息, 提出了融合注意力机制的 C2fSE 模块, 在此基础上将 C2fSE 模块融入到特征重加权提取器中, C2fSE 模块由 C2f 模块以及 SE 模块构成, C2f 模块的卷积层通过增加卷积核的大小和步幅来扩大感受野, 使得模型能够在更大的区域内获取特征, 同时 C2f 中的残差结构能够加深网络的层数,

消失或者爆炸。CSPDarknet-53 相比于 Darknet-53 网络在残差结构中引入了 CSP 结构, CSP 将原输入分为两个分支, 分别使用 1×1 的卷积进行特征变换, 并且使特征图的通道数减半, 其中一个分支的特征图经过 N 个 BottleNeck 模块, 最后在通道维度上进行连接, CSP 分流的结构可以减少计算量, 有效地传递信息, 使得网络能够更好地学习特征。

CSPDarknet-53 使用了 SPPF (最大池化特征金字塔) 模块, 原方法中在骨干网络中使用的方法为 SPP, 两者的作用都是实现局部特征与全局特征的融合, SPP 存在 4 个分支, 其中一个分支为卷积层, 其余 3 个分支分别为卷积核为 5, 9, 13 的最大池化层, 之后将 4 个分支的输出在通道维度上进行连接, 将局部特征与全局特征融合进而提取

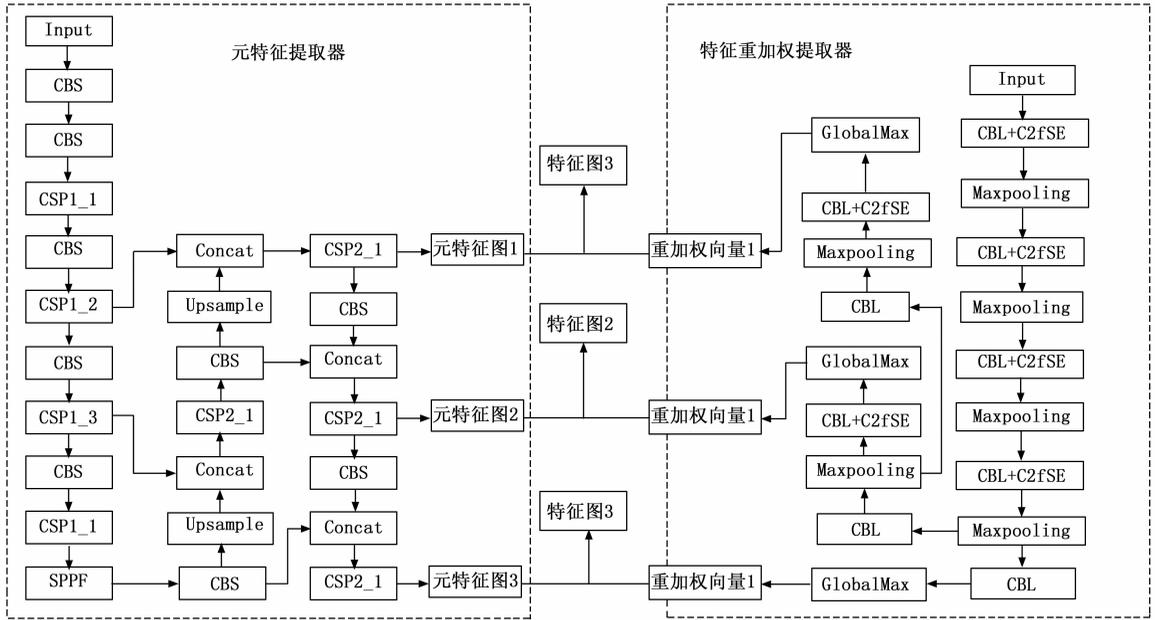


图 3 RE-FSOD 网络结构图

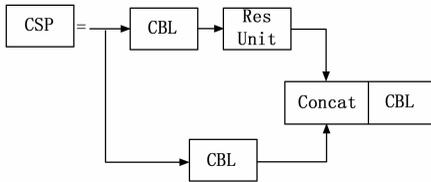


图 4 CSP 结构示意图

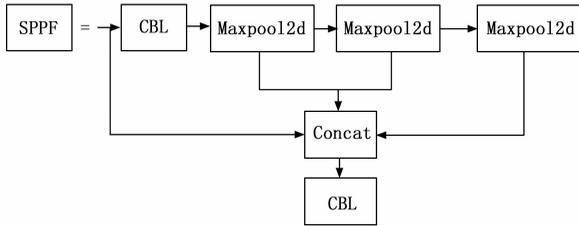


图 5 SPPF 示意图

增强网络的特征提取能力。SE (Squeeze-and-Excitation) 注意力机制首先通过压缩操作，将每个通道的二维特征压缩为 1 个实数，得到通道维度的全局特征，之后通过激励操作为每个通道生成一个权重，最后将得到的权重调整初始的特征图，SE 通过挤压与激励的操作使模型更加关注信息量大的特征，从而抑制那些不重要的特征，进而提升特征重加权网络的特征提取能力。

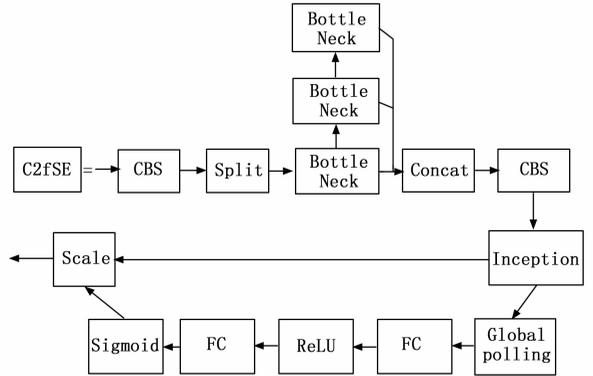


图 7 C2fSE 结构示意图

2.3 损失函数

FSODM 损失函数由定位损失、类别损失及置信度损失构成，相较于 FSODM，RE-FSOD 将定位损失由均方误差损失更换为 CIUO 损失。

FSODM 均方误差损失公式如式 (1) 所示：

$$L_{loc} = \frac{1}{N_{pos}} \sum_{pos} \sum_i (coord'_i - coord_p)^2 \quad (1)$$

式中， $l \in \{x, y, w, h\}$ ， x 和 y 为边界框中心点位置， w 和 h 分别为预测框的宽和高，如果 anchor 与真实框的 IOU 大于 0.5 或者 anchor 在所有 anchor 中和某个真实框具有最

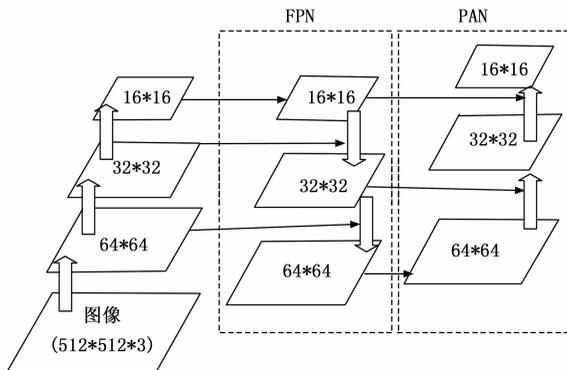


图 6 FPN+PAN 结构示意图

大的 IOU, 则认为此 anchor 为正样本, 否则认为此 anchor 为负样本。pos 代表所有的正样本, $coord_p^i$ 是预测的边界框, $coord_p^r$ 为真实框。当大物体与小物体有相同的 IOU 时, 根据均方误差的公式, 大物体的损失必定大于小物体的损失, 导致不同尺度检测目标定位损失的不一致。因此将损失函数更换为 CIOU 损失函数, CIOU 损失函数以 IOU 作为损失值, 统一了不同尺度目标的损失, 同时 CIOU 损失函数考虑了预测框与真实框中心点的距离以及预测框与真实框之间的宽高比, 加速了模型的收敛并且提高了模型的定位精度。IOU 是预测框与真实框之间的重叠程度, 其计算公式如 (2) 所示:

$$IOU = \frac{A \cap B}{A \cup B} \quad (2)$$

式中, A 为预测框的面积, B 为真实框的面积。

CIOU 损失函数宽高比惩罚项如式 (3) 所示:

$$V = \frac{4}{\pi^2} \left(\arctan \frac{\omega_{gr}}{h_{gr}} - \arctan \frac{\omega}{h} \right)^2 \quad (3)$$

式中, ω_{gr} 与 h_{gr} 是真实框的宽和高, ω 和 h 是预测框的宽和高。

CIOU 中心点距离惩罚项由式 (4)、式 (5) 定义:

$$D = \frac{d}{c^2} \quad (4)$$

$$d = \rho^2(A_{ctr}, B_{ctr}) \quad (5)$$

式中, A_{ctr} , B_{ctr} 分别是真实框与预测框的中心点, $\rho^2(\cdot)$ 代表两点间的欧式距离, c 是真实框 A 与预测框 B 的最大外接矩形的对角线, d 为 A 与 B 的中心点的距离。

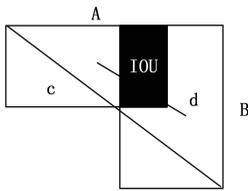


图 8 中心点距离示意图

其中: α 为权重系数, 如式 (6) 所示:

$$\alpha = \frac{V}{(1 - IOU) + V} \quad (6)$$

CIOU 损失函数最终定义如式 (7) 所示:

$$L_{ciou} = 1 - IOU(A, B) + \frac{\rho^2(A_{ctr}, B_{ctr})}{c^2} + \alpha V \quad (7)$$

正样本置信度损失如式 (8) 所示:

$$L_{obj} = \frac{1}{N_{pos}} \sum_{pos} - [P_i \cdot \log P_o + (1 - P_i) \cdot \log(1 - P_o)] \quad (8)$$

式中, p_o 为预测的置信度 o , o 为预测框中存在物体的可能性, p_i 为真实框的值, 有目标为 1, 没有目标时为 0。

负样本置信度损失如式 (9) 所示:

$$L_{noobj} = \frac{1}{N_{neg}} \sum_{neg} - [P_i \cdot \log P_o + (1 - P_i) \cdot \log(1 - P_o)] \quad (9)$$

式中, neg 代表所有负样本。

总置信度损失如式 (10) 所示:

$$L_o = L_{obj} \cdot \omega_{obj} + L_{noobj} \cdot \omega_{noobj} \quad (10)$$

式中, ω_{obj} 和 ω_{noobj} 是正样本和负样本损失的权重, 用来平衡正负样本损失。类别损失如式 (11) 所示:

$$L_c = \frac{1}{N_{pos}} \sum_{pos} - \log \left(\frac{e^{c_{pr}}}{\sum_{i=1}^N e^{c_{pi}}} \right) \quad (11)$$

式中, c_{pr} 是真实类别分数, c_{pi} 是预测类别分数, 已经使用置信度来判断预测框中是否存在目标, 因此分类损失中忽略了负样本。

RE-FSOD 总的损失函数表示为:

$$L = L_{ciou} + L_{obj} + L_c \quad (12)$$

2.4 预测模块

预测模块与 YOLOv3 类似, 依据 YOLOv3 的设置, 采用基于 anchor (锚框) 的方法, anchor 对应大物体尺寸设置为 (116×90)、(156×198)、(373×326), 对中等物体尺寸设置为 (30×61)、(62×45)、(59×119), 对小物体尺寸设置为 (10×13)、(16×30)、(33×23)。将 3 个尺度的特征图输入到预测模块中, 生成 (x, y, ω, h, o, c) , (x, y, ω, h) 代表边界框坐标, o 代表置信度, c 代表类别分数。与 YOLOv3 预测模块相比, 只有类别预测不同, YOLOv3 每个预测框生成 k 个类别分数, 由于 RE-FSOD 对每个类都有一个类特征图, 每个预测框只预测一个类别分数 c , 对应于输入图像的同位置的一组预测框称为 c_{pi} ($i = 1, 2, \dots, K$), 每个预测框类别的最终概率为:

$$P_{ci} = \frac{e^{c_{pi}}}{\sum_{j=1}^K e^{c_{pj}}} \quad (13)$$

其中: $\sum_{i=1}^K P_{ci} = 1$ 。

2.5 训练和推理流程

数据集使用公开的 NWPU VHR-10 十分类遥感数据集, 选择船舶、储罐、篮球场、棒球场、汽车、地面轨道、港口为基类, 飞机、棒球场、网球场为新类。训练策略依据 FSODM 中的设置, 训练过程分为两阶段: 第一阶段在具有充足标签的基类上进行训练, 称为基类训练; 第二阶段使用第一阶段生成的模型在小样本数据集上进行微调。基类训练的数据集 D_{train} 中包括 640 张图片, 共包括 7 类目标, 其中包含船舶目标 240 个、储罐目标 524 个、篮球场目标 127 个、地面轨道目标 130 目标个、港口目标 180 个、桥梁目标 99 个、汽车目标 472 个。在 k -shot 任务下, 微调阶段的小样本数据集 D_{test} 包括 10 类目标, 从基类数据集的图片随机选取, 直至选取的图片包含所有类的 k 个的目标。依据 1.1 中的设置, 每次迭代的输入为 $T_i = \{Q_i, S_i\}$ 。一张查询图像对应一组支持图像, 支持图像由图像及其对应目标类别的掩膜构成如图 9 所示, 一轮训练中, 查询图像就是训练集的所有图像, 其所对应的支持图像由训练集中的图像随机组合生成。

训练以及推理流程如下:

1) 生成基类训练集 D_{train} 以及推理所用的小样本数据集 D_{test} 。

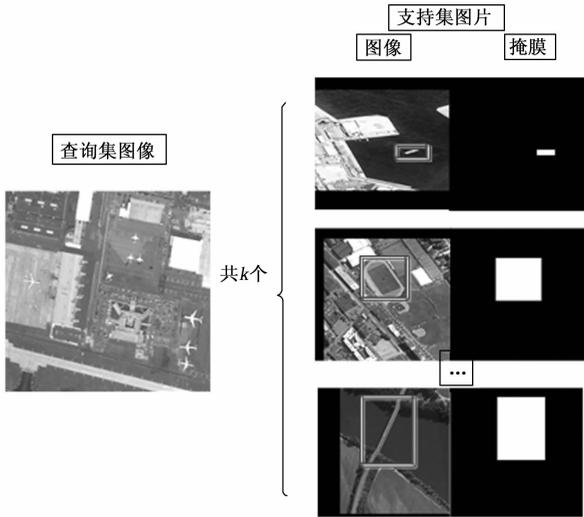


图 9 查询图像与支持图像示意图

2) 初始化特征提取器、特征重加权网络、预测模块的网络权重。

3) 将 D_{train} 中数据输入到网络中训练 900 轮，保存模型。

4) 加载步骤 3) 中的模型，利用 D_{test} 训练 10 轮进行微调，生成最后的模型。

5) 利用 D_{test} 中的数据集，将支持集输入到特征重加权网络中生成重加权向量。

6) 将查询图像输入到网络中生成元特征图，利用重加权向量与元特征图进行 1×1 的通道卷积生成调整后的元特征图，将其输入到预测模块进行解码生成最后的预测结果。解码过程只有类别预测与 YOLOv3 不同，其余均相同。

3 实验与结果分析

3.1 实验环境

CPU 为 intel Xeon[®] Gold 6310, GPU 为 V100-32 GB, 24 G 运行内存, Ubuntu20.04 操作系统, Pytorch 版本为 0.3.1, 计算架构为 CUDA8。

超参数说明：训练图像的分辨率设置为 512×512 , 批次大小设置为 16, 使用 SGD 优化器, 基类训练阶段学习率为 0.01, 微调阶段学习率设置为 0.001, 优化器权重衰减设置为 0.0005, 动量因子设置为 0.937, 在基类训练阶段训练 900 轮, 在微调阶段训练 10 轮, 训练集与测试集以 8:2 的比例进行划分, 其中训练集 640 张图片, 测试集 160 张图片。

3.2 实验评价指标

RE-FSOD 通过 mAP (mean Average Precision) 平均精度均值以及 AP (Average Precision) 平均精度来评价小样本目标检测算法的性能。

$$P = \frac{TP}{TP + FP} \quad (14)$$

$$R = \frac{TP}{TP + FN} \quad (15)$$

$$mAP = \frac{1}{n} \sum_{n=1}^N P(k)R(k) \quad (16)$$

式中, P 代表预测出真正正例占所有预测为正确的比例, R 代表预测出的真正正例占所有预测为正确的比例, 用来反映漏检情况, 在式 (16) 中, n 是代表类别数, 根据 P 和 R 利用式 (16) 计算得出 mAP 。实验采用 mAP_{50} 作为评价指标。

3.3 数据集介绍及数据增强

数据增强流程:

1) 在图像输入到特征提取网络之前, 首先将图像向左右、上下两个方向随机平移, 平移的距离不过图像尺度的 20%。

2) 裁剪平移后的图像, 使用 HSV 进行数据增强, 以 0.5 的概率将图像随机进行翻转, 最后使用多尺度策略将图像放缩, 输入到网络当中。

将图像进行翻转、平移使检测目标在图像中的位置发生变化, 进行 HSV 数据增强调整图像的色彩, 使得输入到重加权模块的每张图片都是不同的, 两种方法大大丰富了样本的多样性, 可以使模型提取到更加鲁棒的特征表示, 提升模型的泛化能力。

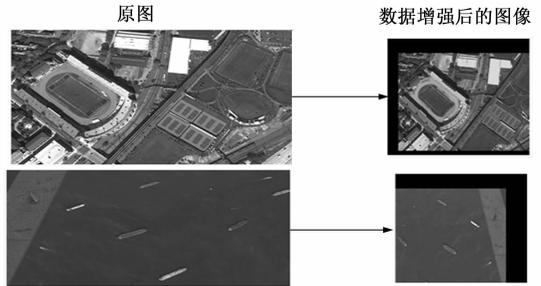


图 10 数据增强示意图

3.4 消融实验

为了评估 RE-FSOD 模型在小样本场景下的检测性能, 将数据集分为两部分, 其中以飞机、棒球场、网球场作为新类, 其余类别作为基类。并且进行了 3-shot、5-shot、10-shot 实验。

改进 1 为使用 CSPDarknet-53 骨干网络, 并且添加 PAN 结构以及 SPPF 结构, 改进 2 为将定位损失函数替换为 CIOU 损失函数, 改进 3 为在特征重加权提取器中添加 C2fSE 结构。同时为了验证改进的有效性, 进行消融实验, 实验结果如表 1 所示, 使用改进 1 之后, 即使用 CSPDarknet-53 作为骨干网络, 并且添加 PAN 结构后, 在 3-shot、5-shot、10-shot 情况下 mAP 分别提升了 10%、2.8%、3.9%, 证明改进 1 能够有效地提升的目标检测算法精度, 并且在 3-shot 实验中, mAP 有大幅提升, 说明改进 1 在样本极少的情景性能更加出色。在改进 1 的基础上添加改进 2, 将定位损失函数替换为 CIOU 损失, 在 3-shot、5-shot、10-shot 的情况下分别提升了 6.3%、9.6%、3%, 验证了改进 2 的有效性, 说明 CIOU 损失函数相比于均方误差损失函数能够提升模型的定位精度。在此基础上引入 C2fSE 模块, 在 3-shot、10-shot 情况下分别提升了 3%、1.1%, 5-shot 情况下基本持平, 证明

了改进 3 的有效性。并且在不同样本条件下进行了 3 个改进模块两两组合的实验, 相较于只改进 1 个模块, 均有不同程度的提高, 说明改进模块具有较好的独立性。

表 1 消融实验结果 mAP_{50}

样本数	方法	飞机	棒球场	网球场	平均
3	FSODM	0.15	0.57	0.25	0.32
	改进 1	0.199	0.796	0.266	0.42
	改进 2	0.224	0.64	0.282	0.382
	改进 3	0.182	0.664	0.246	0.364
	改进 1,2	0.312	0.805	0.334	0.483
	改进 1,3	0.286	0.756	0.293	0.445
	改进 2,3	0.268	0.793	0.37	0.477
	改进 1,2,3	0.305	0.859	0.373	0.513
5	FSODM	0.58	0.84	0.16	0.53
	改进 1	0.534	0.87	0.269	0.558
	改进 2	0.614	0.862	0.4	0.625
	改进 3	0.543	0.848	0.295	0.562
	改进 1,2	0.635	0.886	0.443	0.654
	改进 2,3	0.587	0.871	0.396	0.618
	改进 1,3	0.604	0.852	0.449	0.635
	改进 1,2,3	0.628	0.868	0.456	0.651
10	FSODM	0.60	0.88	0.47	0.65
	改进 1	0.774	0.86	0.426	0.689
	改进 2	0.727	0.873	0.449	0.683
	改进 3	0.756	0.88	0.446	0.694
	改进 1,2	0.732	0.882	0.54	0.719
	改进 2,3	0.728	0.843	0.541	0.704
	改进 1,3	0.766	0.865	0.457	0.696
	改进 1,2,3	0.775	0.878	0.532	0.728

RE-FSOD 算法相较于 FSODM 算法在 3-shot、5-shot、10-shot 情况下的 mAP 分别提升了 19%、11%、8%。并且样本越少提升效果越明显, 说明 RE-FSOD 相比于 FSODM 在小样本条件下检测遥感图像目标具有明显优势。

3.5 不同算法对比实验结果

为验证 RE-FSOD 算法的性能, 与当前流行的小样本目标检测算法 FSODM、TFA、FSCE、PAMS-Det 进行了对比实验, 结果如表 2 所示, RE-FSOD 相较于性能最好的 PAMS-Det 算法在 3-shot、5-shot、10-shot 的情况下 mAP 提升了 14%、10%、7%。

RE-FSOD 以 FSODM 算法为基础进行改进, 为验证 RE-FSOD 改进模块的性能, 进行如图 11 所示的实验, 在 10-shot 场景下, RE-FSOD 相较于 FSODM 算法能够减少错分漏分的情况, 验证了改进模块的优越性。

3.6 特征重加权向量评估

为了验证特征重加权模块的效果, 高维向量难以可视化, 使用 T-SNE 模块对其进行降维, T-SNE 算法中, 高维向量的分布与降维后向量的分布是比较接近的, 但是如果高维向量中本来有一段距离, 降维后的向量距离就会被拉大, 直观地说, 如果原来在高维空间中距离很近, 降维后

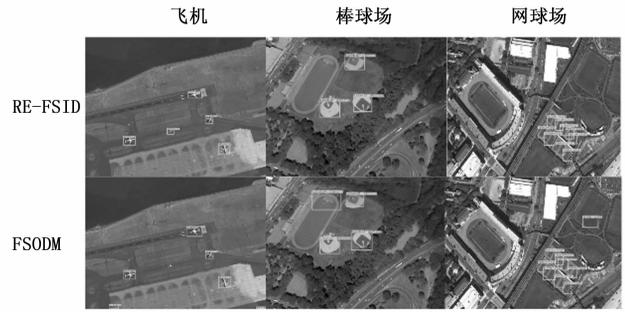


图 11 实验结果对比图

表 2 不同算法实验结果 mAP_{50}

方法	样本数	飞机	棒球场	网球场	平均
FSODM	3	0.15	0.57	0.25	0.32
	5	0.58	0.84	0.16	0.53
	10	0.60	0.88	0.48	0.65
TFA	3	0.12	0.61	0.13	0.29
	5	0.51	0.78	0.19	0.49
	10	0.60	0.85	0.49	0.65
FSCE	3	0.37	0.52	0.24	0.40
	5	0.48	0.67	0.28	0.48
	10	0.57	0.83	0.37	0.59
PAMS-Det ^[21]	3	0.21	0.76	0.16	0.37
	5	0.55	0.88	0.20	0.55
	10	0.61	0.88	0.50	0.66
RE-FSOD	3	0.30	0.86	0.37	0.51
	5	0.63	0.87	0.46	0.65
	10	0.78	0.88	0.53	0.73

距离仍然很近, 但是如果高维空间有距离, 降维后距离就会被拉大, 这样的特性有利于对 1 024 维、512 维、256 维的高维特征重加权进行降维, 并进行可视化。

从基类训练集的支持集中随机选取了共 140 张图片输入到特征重加权模块生成重加权向量, 每类 20 张, 使用 T-SNE 将高维数据降到 2 维, 进行可视化。得到结果如图 12 所示。



图 12 不同维度重加权向量对比

不同的点代表不同类别的特征重加权向量, 可以见到相同类别的重加权向量聚集在一起, 这表明重加权模块通过梯度下降后能够成功地表示来自原始输入图片的类别信息。并且使用样本点到中心点的欧式距离对聚类效果进行了评估, 公式如下:

$$D = \sqrt{(x_1 - x_{\text{mean}})^2 + (y_1 - y_{\text{mean}})^2} \quad (17)$$

式中, $(x_{\text{mean}}, y_{\text{mean}})$ 是类中心点的坐标, (x_1, y_1) 是样本点的坐标。

计算所有样本点到其类中心点的欧式距离的平均值, 得到 256 维的平均欧式距离为 785, 512 维平均欧式距离为 689, 1 024 维平均欧式距离为 627, 说明 1 024 维的特征重加权向量聚类结果比 512 维以及 256 维的特征重加权向量聚类结果更好, 说明维度更高的特征重加权向量承载了更多信息, 因此具有更高维度的特征重加权向量能够更好地表示支持样本中的类别信息。

4 结束语

RE-FSOD 以 FSODM 方法为基础, 在元特征提取器中使用 CSPDarknet-53 作为骨干网络, 并且添加了 PAN 结构以及 SPPF 结构, 能够提取更加鲁棒的元特征, 在特征重加权提取器中引入结合注意力机制与残差结构的 C2fSE 模块, 能够提取到更丰富的语义信息, 并且使用 CIOU 作为定位损失函数, 提升定位精度并加速算法的收敛。经过实验证明, 特定重加权向量的维度越高, 提取到特定类别重加权向量的语义信息越丰富, RE-FSOD 相较于 FSODM 提出的 3 个改进方法均能有效地提升算法的检测精度, 并且相较于其他先进小样本目标检测方法具有明显的优势, 说明该方法能够更好的处理尺度多变、背景复杂、目标模糊的遥感图像。目前 RE-FSOD 在推理速度、计算量、占用内存等方面仍然有待于进一步提升, 后续计划通过利用剪枝、量化等方法对模型进行轻量化, 在保证模型精度的同时降低模型的参数量并且提升模型的推理速度。

参考文献:

- [1] 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.
- [2] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.
- [3] REDMON J, FARHADI A. Yolov3: an incremental improvement [J]. Arxiv Preprint Arxiv: 1804.02767, 2018.
- [4] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: single shot multibox detector [C] //Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, the Netherlands, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.
- [5] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [6] GIRSHICK R. Fast r-CNN [C] //Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [7] 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.
- [8] WANG X, HUANG T E, DARRELL T, et al. Frustratingly simple few-shot object detection [J]. Arxiv Preprint Arxiv: 2003.06957, 2020.
- [9] SUN B, LI B, CAI S, et al. Fscf: few-shot object detection via contrastive proposal encoding [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 7352-7362.
- [10] LI B, YANG B, LIU C, et al. Beyond max-margin: class margin equilibrium for few-shot object detection [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 7363-7372.
- [11] QIAO L, ZHAO Y, LI Z, et al. Defrcn: Decoupled faster r-CNN for few-shot object detection [C] //Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 8681-8690.
- [12] KANG B, LIU Z, WANG X, et al. Few-shot object detection via feature reweighting [C] //Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 8420-8429.
- [13] LI X, DENG J, FANG Y. Few-shot object detection on remote sensing images [J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1-14.
- [14] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.
- [15] WANG C Y, LIAO H Y M, WU Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020: 390-391.
- [16] HE K, ZHANG X, REN S, 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.
- [17] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [18] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [19] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression [C] //Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34 (7): 12993-13000.
- [20] 张振伟, 郝建国, 黄健, 等. 小样本图像目标检测研究综述 [J]. 计算机工程与应用, 2022, 58 (5): 1-11
- [21] ZHAO Z, TANG P, ZHAO L, et al. Few-shot object detection of remote sensing images via two-stage fine-tuning [J]. IEEE Geoscience and Remote Sensing Letters, 2021, 19: 1-5.