

基于感知延伸与锚框最适匹配的 遥感图像目标检测算法

洪伟¹, 赵祥模², 王鹏³, 李晓艳², 邸若海², 吕志刚², 王储²

(1.西安工业大学 兵器科学与技术学院, 陕西 西安 710021;
2.西安工业大学 电子信息工程学院, 陕西 西安 710021;
3.西安工业大学 发展规划处, 陕西 西安 710021)

摘 要:针对遥感图像目标成像小、背景复杂、分布拥挤的问题,将旋转目标检测方法 S²ANet 作为基线网络,提出一种基于感知延伸与锚框最适匹配的遥感图像目标检测算法(HQ-S²ANet)。构建协同注意力模块(SEA),捕获特征像素间关系的同时扩展模型感知区域,实现目标与全局的关系建模;针对遥感图像背景复杂问题,改进特征金字塔(FPN)特征融合过程,在特征融合下采样过程中将感知延伸卷积模块与常规卷积交替堆叠形成感知延伸特征金字塔模块(HQFPN),保证低层细节位置信息的同时,延伸感知范围以增强模型信息捕捉能力;为解决遥感目标图像分布拥挤的问题,利用高质量锚框匹配方法(MaxIoUAssigner_HQ),通过常数因子控制锚框真值分配,在保证召回率的同时,防止低质量锚框匹配产生。实验结果表明,在 DOTA 数据集下,与 S²ANet 算法相比,HQ-S²ANet 平均精度(mAP)提高 3.1%,召回率(Recall)均值提高 1.6%,而参数量仅增加 2.61M,所提算法有效增强了遥感图像目标检测能力。

关 键 词:遥感图像;特征融合;锚框匹配;旋转检测

中图分类号:TP7 **文献标志码:**A **文章编号:**1000-2758(2023)04-0820-11

遥感(remote sensing),指一切非接触的超视距探测,一般指运用传感器对物体辐射及反射特性的探测。伴随着遥感技术不断更新迭代,遥感图像分辨率不断提升,遥感数据中常见的桥梁、飞机等目标也可在图像中清晰可见。由此可见遥感图像可以从宏观的角度反馈信息。因此准确地对遥感图像进行图像分析在军事、民生保障、气象测绘等领域十分重要^[1]。

在深度卷积网络的框架下,遥感图像检测中大部分现有方法依赖于复杂的 R-CNN 框架,如 SCR-Det^[2]、CAD-Net^[3]、ReDet^[4],该框架由两部分组成:区域建议网络(RPN)及 R-CNN 检测头^[5]。RPN 用于从水平锚框生成高质量的感兴趣区域(ROI),并

利用 ROI 池化算子从 ROI 中提取准确的特征,最后利用 R-CNN 进行目标分类以及边界框回归。但水平 ROI 通常会导致定向边界框和目标之间严重错位^[6],遥感图像中目标分布较为密集,水平 ROI 通常会包含多个实例。因此,需要具有不同角度、长宽比的符合需求的锚框,但生成锚框的同时会导致计算开销大幅增加。近期,ROI 转换器的提出^[7],利用水平 ROI 替换旋转 ROI 的方法,解决了锚框重复生成的问题,但仍需复杂的 ROI 操作。与基于 R-CNN 的检测器相比,单阶段检测器,如 YOLO-ACN^[8]、S²ANet^[9]、SSD^[10]、RetinaNet^[11],通过回归边界框,并使用规则且密集的采样锚点直接对目标进行分类。这种体系结构具有较高的计算效率,但在精度

上往往低于双阶段检测器。现阶段,遥感图像目标检测算法大多由传统双阶段目标检测算法针对遥感图像特点改进得到。传统算法虽能在常规场景的数据集上表现出良好的检测性能,但是由于遥感场景中存在目标成像小、变化方向任意、变化尺度大、分布密集、背景复杂等问题,改进后的传统算法仍难以有效提取遥感场景中目标的细节位置特征及充足的语义信息,导致检测效果无法达到预期^[12]。

为提升遥感场景下目标识别能力,考虑到遥感图像目标变化方向的不确定性,本文针对遥感图像目标成像小、背景复杂、分布拥挤的问题提出基于感知延伸与锚框最适配的遥感图像目标检测算法,主要工作如下:①在特征融合前利用协同注意力模块(SEA),捕获特征像素间关系的同时扩展模型感知区域,实现目标与全局的关系建模;②通过分块裁剪输入图像以扩大目标在图像中所占比例,利用感知延伸特征金字塔(HQFPN)完成特征融合,HQFPN下采样过程中将感知增强卷积模块与常规卷积交替构建感知延伸下采样模块(EN_DilatedConv),减少下采样过程中低层细节位置特征的丢失,解决传统特征融合金字塔模块^[13]在特征融合时信息丢失的问题;③在为真实边界框分配锚框时,采用高质量锚框匹配算法(MaxIoUAssigner_HQ),在保证召回率的同时,减少低质量锚框匹配的产生。

1 S²ANet 目标检测算法

S²ANet 模型的总体框架如图 1 所示。该模型使用 ResNet50^[14] 模块从输入图像提取特征,并使用包含特征对齐模块和定向检测模块组成的目标检测头网络生成旋转预测边界框及类别置信度,最后使用非极大值抑制(NMS)筛选结果。FAM 模块由锚优化网络(ARN)及对齐卷积层(ACL)组成。ARN 是一个轻型网络,由锚分类分支及锚框回归分支组成。锚分类分支将锚细分为不同类别,锚框回归分支将水平锚细化为高质量旋转锚框。为了加快模型推

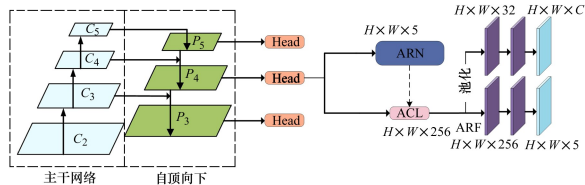


图 1 S²ANet 结构图

理速度,推理阶段不执行锚分类分支,转而通过锚框回归调整对齐卷积(AlignConv)中采样位置。

首先通过规则网络 $R_n = \{(r_x, r_y)\}$ 对输入特征图 X 采样,然后将加权的采样值相加。例如 $R = \{(-1,1), (-1,0), \dots, (0,1), (1,1)\}$ 表示核大小为 3×3 ,膨胀为 1,对于输出特征图 Y 上的每个 $p \in R$,采样加权计算公式为

$$Y(p) = \sum_{r \in R} W(r) \cdot X(p+r) \quad (1)$$

相对于普通卷积,对齐卷积增加偏置项 O 。

$$Y(p) = \sum_{r \in R} W(r) \cdot X(p+r+O) \quad (2)$$

对于位置 p ,偏置项 O 的计算为基于锚的采样位置及常规采样位置 $(p+r)$,位置 p 处的边界框表示为 $(x+\omega+h+\theta)$ 。对于 $r \in R$,基于锚框的采样位置 L'_p 可表示为

$$L'_p = \frac{1}{S} \left(x + \frac{1}{k} (\omega, h) \cdot r \cdot R^T(\theta) \right) \quad (3)$$

式中: k 表示为卷积核大小; S 表示采样步长,其中 $R(\theta) = (\cos\theta, -\sin\theta; \sin\theta, \cos\theta)^T$ 为旋转矩阵; p 位置的偏置项 O 可表示为

$$O = \{L'_p - p - r\}_{r \in R} \quad (4)$$

通过该方式可以将给定位置 p 的轴对齐卷积特征 $X(p)$ 转换为基于对应锚框的任意方向卷积特征。

图 2a) 为普通 2D 卷积基于锚框生成的采样点,图 2b) 为经过 AlignConv 生成的旋转锚框,箭头表示采样点偏移。



a) 普通卷积采样点 b) 对齐卷积采样点

图 2 普通卷积与对齐卷积采样点

ACL 将 ARN 生成的旋转锚解码为绝对锚框 (x, w, h, θ) 后与计算所得偏移量及图像提取特征输入对齐卷积中,以提取对齐特征。ODM 模块利用主动旋转滤波器(ARF)缓解分类置信度和位置回归精度的不一致性,使得目标检测更加精准。ARF 是一个 $K \times K \times N$ 大小的滤波器,利用 ARF 旋转在卷积操作中生成的特征图具有 N 个方向,以实现对方向信息的编码。将 ARF 用于卷积层,可以获得方向敏

感特征。ODM 通过选择响应最强的方向通道汇集方向敏感特征,达到提取方向不变特征的目的。最后将方向敏感特征和方向不变特征分成 2 个子网络,分别进行边界框回归及目标分类。FAM 与 ODM 组成的目标检测头,可以对齐具有不同方向的对象特征,从而提升目标分类的鲁棒性。

2 HQ-S²ANet 目标检测算法

为解决遥感场景中目标成像小、背景复杂、分布拥挤的问题,本文提出一种改进的 S²ANet 遥感图像目标检测算法 HQ-S²ANet,使用感知延伸特征融合模块对骨干网络提取特征进行融合;为锚框匹配真实标注框时采用 MaxIoUAssigner_HQ 算法,在保证召回率的同时,抑制低质量匹配的发生,以增强模型检测能力。

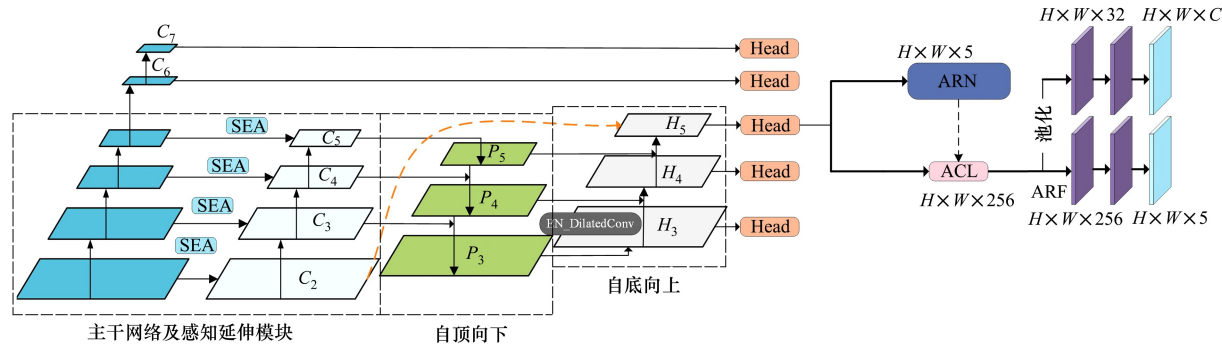


图 3 HQ-S²ANet 结构图

2.1 感知延伸特征融合模块

2.1.1 协同注意力模块(SEA)

自注意力机制(self-attention)作为注意力机制的重要组成部分,依赖特征个体相关性,通过接受的输入生成像素间关联性权重,避免外部信息对其的影响。通道注意力 ECA-Net^[15]在不增加网络计算量的同时使网络增强有用信息去除冗余特征。为缓解自注意力获取通道间关系能力较弱以及 ECA 缺乏特征局部特征关系的问题,将自注意力对上下文信息的捕捉能力以及 ECA 能够加强通道间

依赖关系的效果进行结合。输入特征经过 1×1 卷积后得到 3 个维度的向量 Q (query), K (key), V (value)。将 Q 与 K 向量的转置矩阵相乘,经过 softmax 后得到权重系数 A 。当矩阵 A 与矩阵 V 相乘后得到经过自注意力处理的输入向量权重 S 后,利用 ECA 对其进行通道间关系处理。此时得到的矩阵较通过原始输入特征得到的矩阵,加强通道间依赖关系的同时避免了传统通道注意力因降维跨通道交互所带来的影响,由此提出了协同注意力模块 SEA。如图 4 所示,利用协同注意力模块处理主干网络输出特征,以增强特征间相关性。

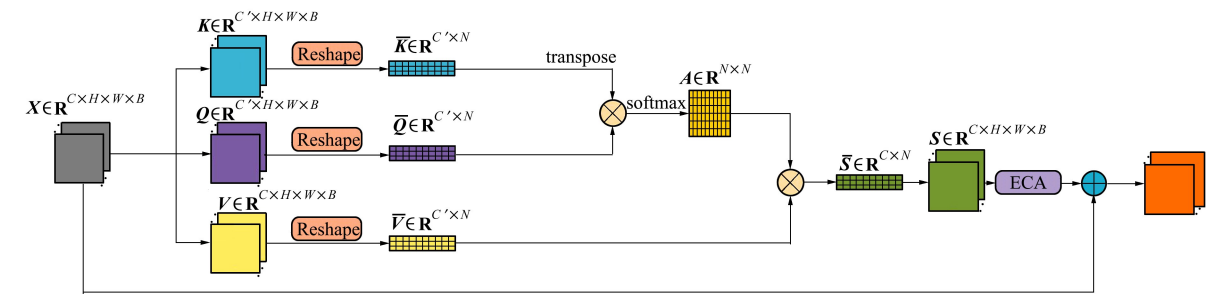


图 4 协同注意力结构图

$$\mathbf{A} = \text{softmax}\left(\frac{\mathbf{K}^T \mathbf{Q}}{\sqrt{C'}}\right) \quad (5)$$

将权重系数 $\mathbf{A}$ 乘以 $\mathbf{V}$ 向量,得到加权后的输入特征评分 $\mathbf{S}$ 。

$$\bar{\mathbf{S}} = \bar{\mathbf{V}}\mathbf{A} \quad (6)$$

将 $\mathbf{S}$ 向量 reshape 为三维特征向量后,经过 ECA 强化通道间关系,与输入特征向量通过 add 操作得到 SEA 模块的输出特征 $\mathbf{Z}$ 。

$$\mathbf{Z} = \text{ECA}(\text{reshape}(\bar{\mathbf{V}}\mathbf{A}) + \mathbf{X}) \quad (7)$$

2.1.2 感知延伸特征金字塔模块(HQFPN)

为缩短信息路径,利用低层存在的精确位置定位信息增强特征金字塔融合效果,构建自底向上路径。随着特征传递层数的增加,特征信息丢失逐渐增多、局部化响应能力不足。因此,在自底向上的路径中建立从低层到高层的捷径分支,以弥补信息丢失,同时通过传播低层细节的强响应,进一步增强整个特征层次的局部化响应能力。

给定主干网络上的特征图 $C_i^{C \times H \times W}$,则自底向上特征图 $H_i^{256 \times H \times W}$ 可表示为

$$H_3 = P_3 \quad (8)$$

$$H_4 = \text{Conv}(P_4) + \text{EN_DilatedConv}(H_3) \quad (9)$$

$$H_5 = \text{Conv}(P_5) + \text{Downsample}(H_4) + \text{Downsample}(C_2) \quad (10)$$

式中:Conv 表示通道为 256,核为 3×3 的卷积操作;Downsample() 表示下采样操作;EN_DilatedConv 表示感知延伸下采样模块。

传统特征融合模块中,低层特征图细节定位信息较为丰富,高层特征图语义信息较为丰富,语义信息经过自上而下路径传递,中间经过卷积操作及信息交互过程较少,对语义信息影响较小;浅层特征需沿骨干网络传递,细节特征传递至顶层需要经过骨干网络中多次卷积、池化等操作,导致特征细节定位信息丢失较为严重,不利于浅层特征中小目标的定位。因此在原有特征融合模块基础上构建自下而上路径,避免底层特征经过骨干网络,使底层细节位置特征传递至深层时保留充足的有效信息。

图 5 中红色连接为传统特征金字塔低层特征传递路径,蓝色连接为 HQFPN 传递路径,构建捷径分支路径后,低层特征传递所经过的网络层结构大幅减少,有效地保留了低层特征中的细节位置信息,提升了小目标识别能力。

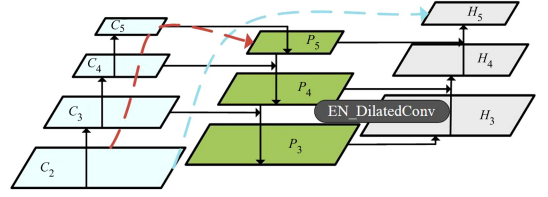


图5 低层特征传递示意图

遥感图像存在一定随机性,部分目标类别排列稀疏,数据量小。该类遥感图像包含的实际信息十分有限,图像中目标区域较为局限,其余区域多为无关背景,大尺寸卷积操作虽能较好地提取目标语义特征,但是对于遥感图像,扩大卷积核不仅会增加算法计算开销,也会影响算法对密集分布目标的识别能力。空洞卷积在普通卷积操作的基础上通过膨胀系数 d 控制卷积核稀疏采样尺度,并将卷积核稀疏采样间隔区域填充^[16]。

$$S_{w,h} = K_{w,h} + (K_{w,h} - 1) \cdot (d - 1) \quad (11)$$

式中: $S_{w,h}$ 为空洞卷积核尺寸; $K_{w,h}$ 为原卷积核尺寸。

故在 HQFPN 自底向上的下采样过程中使用空洞卷积,在不增加有效卷积单元数量的同时延伸感知范围。

图 6 为标准卷积与空洞卷积操作示意图,空洞卷积根据空洞率将特征间隔部分填充 0,导致特征图出现网格隔影^[17],一定程度弱化了特征,影响下采样后特征的显著性。为解决该问题,本文构建 EN_DilatedConv 模块,将空洞卷积与均值池化组合使用。对于低层特征,一部分采用空洞率为 2 的 3×3 空洞卷积与输入特征卷积后的特征组成,另一部分则由 3×3 大小的全局均值池化作用于输入特征图后得到的特征组成,两部分特征进行相加后,作为下采样最终的输出结果

$$\mathbf{O} = \mathbf{F}_1 + \mathbf{F}_2, \mathbf{O} \in \mathbf{R}^{H \times W \times C} \quad (12)$$

$$\mathbf{F}_1 = \text{Conv}_D(\mathbf{X}), \mathbf{F}_2 = \text{AvgPool}(\mathbf{X}) \quad (13)$$

式中: $\mathbf{X}$ 为输入特征图; $\text{Conv}_D()$ 表示核大小为 3,步长为 2,空洞系数为 2,填充为 2 的空洞卷积; AvgPool 表示全局平均池化操作。

在自下而上特征传递路径中,自最底层开始,对特征图采取下采样操作,逐层传递特征信息,与对应自上而下路径上的特征图相加融合。此外将 C_2 的特征图经过 8 倍池化后,与自下而上传递路径中最顶层特征图融合,使最顶层输出特征图既包含自身

丰富的高阶语义信息,又融合原始低层特征图池化得到的细节信息,达到增强特征图信息表征的能力。

最终将 H_3, H_4, H_5, C_6, C_7 作为预测特征图,输出特征图既延伸了感知范围,又包含了低层丰富的细节定位信息及高层充足的语义信息。

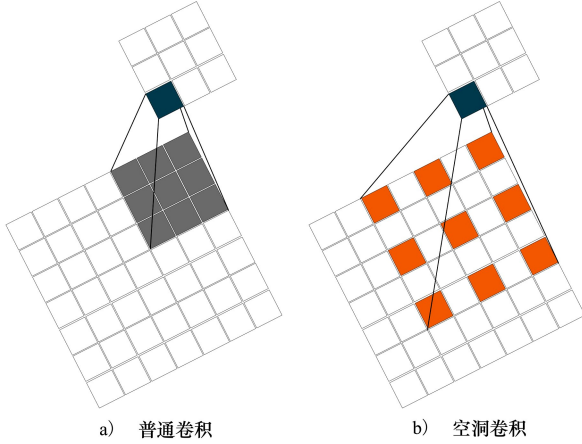


图 6 标准与空洞卷积操作示意图

2.2 锚框最适匹配方法 (MaxIoUAssigner_HQ)

经过感知增强后的特征图传递到检测头中, FAM 及 ODM 会得到生成的高质量旋转锚框与真实标注框的交并比 (IOU), 根据设定好的正样本及负样本阈值为所有锚框分配属性。在原网络中, S^2ANet 将 IOU 值低于负样本阈值的锚框判定为负样本, 不参与真实标注框的匹配。由于航空遥感图像目标尺度小、分布拥挤, 网络训练前期产生较多的低质量匹配均被忽略, 导致召回率过低, 不利于网络的损失计算, 进而导致精度过低。改进后的 S^2ANet 实现过程中, 令低质量锚框也参与真实标注框的匹配, 虽保证了召回率, 但一定程度上导致低质量匹配的发生。

$$I = \begin{pmatrix} a_{11} & \cdots & a_{1B} \\ \vdots & & \vdots \\ a_{G1} & \cdots & a_{GB} \end{pmatrix} \quad (14)$$

$$G_A = [\max(a_{1j}), \max(a_{2j}), \cdots, \max(a_{Gj})] \quad (15)$$

$$G_i = [j_1, j_2, \cdots, j_G] \quad (16)$$

$$A = [\max(a_{i1}), \max(a_{i2}), \cdots, \max(a_{iB})] \quad (17)$$

$$A_i = [i_1, i_2, \cdots, i_B] \quad (18)$$

B 为生成锚框 (Anchor) 个数, G 为真实标注框个数。矩阵 I 记录网络生成锚框与对应真实标注框的 IOU 值, G_A 记录真实标注框与 Anchor 的最大 IOU 值, G_i 记录对应 Anchor 的索引。 A 记录 Anchor 与真实标注框的最大 IOU 值, A_i 记录对应真实标注框的索引。网络生成锚框与真实标注框对比时会对 G_A 中的值做筛选, 若小于负样本阈值, 则被划分为背景, 判定该标注框无符合要求锚框。该操作会使部分真实标注框未被分配锚框, 导致召回率过低。为解决这一问题, MaxIoUAssigner 设定正样本阈值下限, 若 IOU 小于负样本阈值但是大于正样本阈值下限, 则不将其划分为背景, 并将该锚框与对应真实标注框做匹配。

由于遥感图像部分目标分布密集, 单个锚框会与多个真实目标标注框发生重叠, 且由于一个锚框只允许与一个真实标注框完成匹配, 高质量匹配可能会被低质量锚框覆盖。若当前锚框已与最大 IOU 值对应的真实标注框完成匹配。 i, j 为真实标注框索引, m 为 Anchor 索引。

$$R_A[i] = G_i[m] \quad (19)$$

$$A_i[G_i[m]] = i \quad (20)$$

此时若另一真实标注框的最大 IOU 也为与该锚框交并得到的

$$R_A[j] = G_i[n] \quad (21)$$

$$G_i[n] = G_i[m] \quad (22)$$

此时

$$A_i[G_i[m]] = j \quad (23)$$

匹配 IOU 大于该样本阈值, 导致与该锚框 IOU 最大的真实标注框 i 被重写为 j , 该锚框被分配了 IOU 值较低的真实标注框, 甚至可能为负样本, 导致锚框匹配质量过低, 影响边界框回归以及分类损失, 进而影响模型精度。针对以上问题, 本文提出 MaxIoUAssigner_HQ 高质量锚框匹配方法。

表 1 为 MaxIoUAssigner_HQ 算法伪代码, 在为真实标注框分配其 IOU 最大锚框时通过常数因子控制, 若该 IOU 值为当前锚框与真实标注框交并最大值, 则允许匹配。若不为最大值, 则与最大值做差值, 当差值小于常数因子时, 允许匹配发生, 当大于常数因子时则拒绝匹配。通过常数因子限定匹配的发生, 在保证召回率的同时, 防止低质量匹配发生, 避免高质量锚框被重写。

表 1 MaxIoUAssigner_HQ 算法伪代码

MaxIoUAssigner_HQ 算法	
输入: i // 真实标注框索引	
$G_A[]$ // 真实标注框与锚框的 IoU_{max}	
$G_i[]$ // 对应锚框索引	
$A[]$ // 锚框与真实标注框的 IoU_{max}	
$A_i[]$ // 真实标注框索引	
n // 真实标注框个数	
输出: $R_A[]$ // 锚框与真实标注框匹配结果	
1: while $i < n$	
2:	if $G_A[i] > min_pos$
3:	if $G_A[i] == A[G_i[i]]$ // 与该锚框最大 IOU 值作比较
4:	$R_A[i] = A_i(G_A[i])$ // 相同允许匹配
5:	else if $A[G_i[i]] - G_A[i]$
	$< = value$ // 差值小于常数因子允许匹配
6:	$R_A[i] = A_i(G_A[i])$
7:	end if
8:	$i++$
9:	end if
10:	end while
11:	return $R_A[]$

3 实验结果与分析

3.1 数据集与评价指标

为验证改进算法的有效性,设计实验在 DOTAV1.0 数据集上进行验证。DOTA 数据集来源包含不同传感器和平台,包括 Google Earth、JL-1 卫星拍摄,以及中国资源卫星数据和应用中心的 GF-2 卫星拍摄。数据集共计 2 806 幅航拍图,每张图像的像素尺寸在 800×800 到 4 000×4 000 的范围内,包含不同尺度、方向和形状的物体。类别包括: plane, baseball-diamond, bridge, ground-t-field, small-vehicle, large-vehicle, ship, tennis-court, basketball-court, storage-tank, soccer-ball-field, roundabout, harbor, swimming-pool, helicopter。由于 DOTA 测试集未公开相应类别标签文件,本次实验采用官方给定训练集进行训练,不设置验证集,划分官方验证集作为测试集,验证集与测试集无相同图片。采用

mAP 作为评价指标。

3.2 实验环境及参数设置

实验环境为 Ubuntu18.04,Pytorch 版本为 1.8.1,服务器硬件配置为 NVIDIA RTX3070(8GB 显存)显卡,Inter core i5 处理器。

本文采用的 DOTA 数据集中图像分辨率集中在 800×800 到 4 000×4 000 的范围内,若直接将大分辨率图像输入网络或导致参数量过大,图形计算单元无法负荷。经过实验,将输入分辨率改为 2 048×2 048后,网络无法在当前硬件环境下运行。若分辨率过小,将导致可学习特征减少,网络无法获取足够的空间定位信息及语义信息,进而导致网络检测性能下降。故本文对数据集中图片进行裁剪操作,将大分辨图片裁剪为若干尺寸为 1 024×1 024 的子图。实验时输入图像尺寸为 1 024×1 024,由于硬件环境制约,设置 batch_size 为 2,初始学习率设为 0.002 5,动量设为 0.9,使用随机梯度下降(SGD)优化器。考虑到数据集规模及模型复杂度,为防止模型出现过拟合现象,将训练 Epochs 设为 20,训练过程中,第一个 Epoch 使用 Warmup 算法进行预热;输出预测结果后,使用 NMS 算法进行处理。

3.3 结果分析

图 7 展示了 S²ANet 与 HQ-S²ANet 在 DOTA 遥感目标验证集的部分检测结果图,图中上半部分为原始图片,下半部分为框选区域局部放大图。图中可看出目标排列密集、尺寸小,变化范围大、变化方向任意的特点。S²ANet 会出现漏检或者误检的问题。S²ANet 对图 7a)左侧尺寸较小的卡车不够敏感,在图 7c)中会忽略右侧汽车目标且出现了误检及重复检测的问题,而 HQ-S²ANet 避免了上述问题的发生。

为验证所提算法的性能,在划分测试集上进行实验,并对比了文中模型以及其他先进模型。结果如表 2 所示。类别分别为 PL: plane, BD: baseball-diamond, BR: bridge, GTF: ground-t-field, SV: small-vehicle, LV: large-vehicle, SH: ship, TC: tennis-court, BC: basketball-court, ST: storage-tank, SBF: soccer-ball-field, RA: roundabout, HA: harbor, SP: swimming-pool, HC: helicopter。

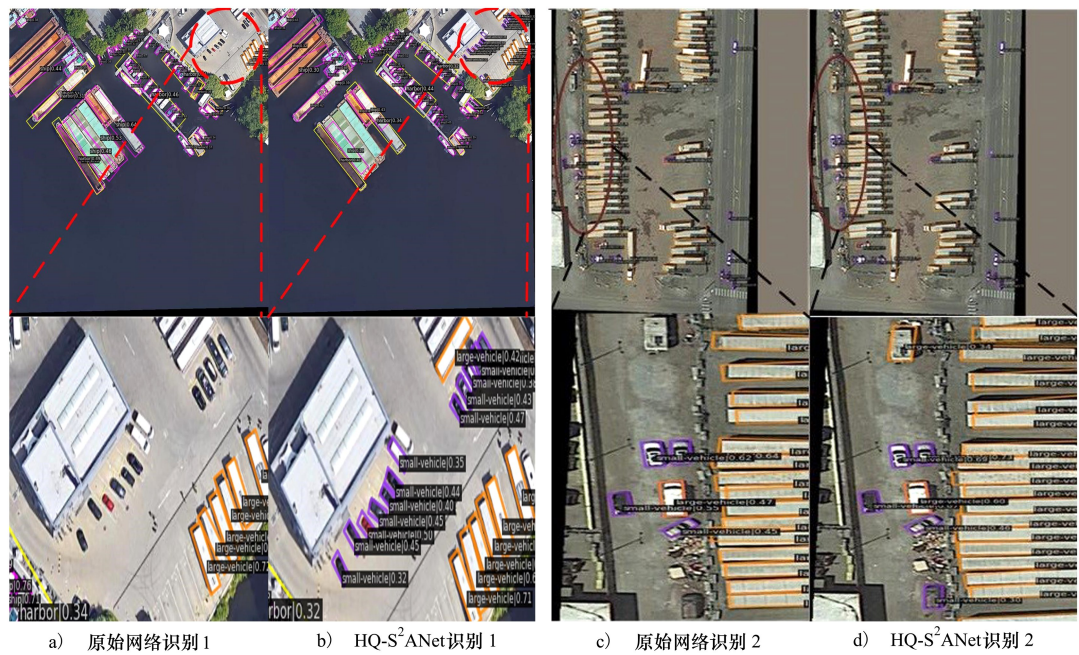


图 7 DOTA 数据集上部分检测结果

表 2 HQ-S²ANet 与其他先进模型在 DOTA 数据集上的性能比较

模型	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	mAP
SASM ^[18]	89.00	72.10	47.30	67.30	59.00	75.10	77.70	90.00	59.60	68.20	40.80	62.00	71.40	56.20	31.20	64.50
CSL ^[19]	88.30	75.90	40.70	67.60	66.60	71.40	78.60	90.50	57.70	60.80	49.70	66.70	59.80	53.40	44.70	64.80
Rotated_repoint ^[20]	89.30	73.50	37.10	50.40	60.70	58.60	75.80	87.80	52.30	66.30	47.60	63.20	55.00	54.00	23.00	59.60
Rotated_CFA ^[21]	89.20	75.20	47.90	70.90	71.40	78.00	88.40	90.10	56.20	68.90	46.90	60.80	68.20	54.70	43.40	67.30
Rotated_retinanet	88.70	76.40	39.70	69.20	65.30	71.20	77.70	90.60	68.40	60.60	51.60	63.10	61.10	51.30	49.30	65.70
Rotated_atss ^[22]	89.50	76.20	46.50	72.20	64.00	75.50	83.10	90.70	65.50	67.40	52.30	63.80	63.60	58.80	51.00	68.00
S ² ANet	89.60	75.10	48.10	69.70	67.70	78.60	86.20	90.70	66.50	69.20	54.90	65.30	65.20	56.80	39.60	68.20
HQ-S ² ANet(本文)	89.40	77.40	48.00	72.00	73.20	81.70	88.40	90.60	69.20	70.20	57.00	66.10	71.60	58.00	56.90	71.30

在 DOTA 数据集上, HQ-S²ANet 的 mAP 为 71.3%, 与原算法 S²ANet 精度相比提高 3.1%, 与 Rotated_atss 算法相比提高 3.3%。Rotated_retinanet 精度为 65.70%。Rotated_CFA 精度为 67.30%, 该算法虽能较好地检测 PL, BR 这类分布较为稀疏且目标大小适中的目标, 但对于密集目标, 如 SV, 其精度较低。改进算法对 SV, HC 这类分布密集小目标的 AP 明显较高, 表明其具备解决遥感图像检测难点的能力, 能从复杂背景中提取特征信息较少且语义信息较弱的目标。

图 8 为 HQ-S²ANet 与 S²ANet 特征融合后的热力图结果, 图 8a)~8c) 为 S²ANet 输出结果, 图 8d)

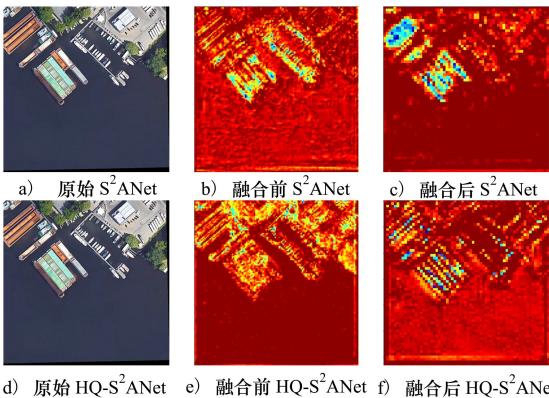


图 8 S²ANet 与 HQ-S²ANet 特征可视化

~8f) 为 HQ-S²ANet 输出结果。图 8b), 8e) 为 FPN 与 HQFPN 最底层特征图特征可视化, 包含丰富细节定位信息。图 8e) 中经过 SEA 加权后的特征对小目标给予更高关注度, 与图 8b) 相比, HQ-S²ANet 关注区域更广、对目标关注程度更高。图 8f) 为图 8e) 经过 HQFPN 下采样模块细化特征后的热力图。较图 8c) 经过传统下采样处理, 图 8f) 扩大感受野的同时捕获邻近目标间关系, 目标边缘特征更加明确。

3.4 消融实验

将 HQFPN, MaxIoUAssigner_HQ、协同注意力模块依次嵌入到 S²ANet 模型, 为验证模块有效性, 结果如表 3 所示。

表 3 DOTA 数据集上的消融实验研究

算法	HQFPN	Max_HQ	SEA	mAP/%
S ² ANet				68.20
	✓			68.90
	✓	✓		69.70
	✓	✓	✓	71.30

嵌入各个模块后, 均能获得比原始网络更高的精度。其中嵌入 HQFPN 模块后, 由于特征图经过增强, 通过大感受野的方式融合了低层细节位置信息以及高层语义信息, mAP 提升 0.7%。再进一步优化锚框匹配策略, 为抑制锚框低质量匹配的发生, 嵌入 MaxIoUAssigner_HQ 模块后, mAP 提升 0.8%, 证明该模块在保证召回率的同时提升了网络识别精度。为捕获特征中的上下文信息并获取通道间依赖关系, 将 SEA 模块加入网络。特征图经过注意力加权后, 更加注重有效信息的提取, mAP 提升 1.6%, 证明网络完成局部间信息交互的同时捕获了通道间信息依赖, 能够显著提升目标检测精度。

3.4.1 注意力模块对比实验

该实验在 DOTA 数据集上对 2 种注意力融合的先后顺序进行了测试, 第一种方式为利用通道注意力对自注意力中 Q, K, V 向量进行加权后计算关联度矩阵, 第二种方式为利用通道注意力对自注意力机制得到的向量评分矩阵进行加权处理后与向量 S 相加。由于目标种类过多, 表中只取尺寸较小及分布较为密集的 2 类目标, 如表 4 所示。

若过早地使用通道注意力对自注意力机制中 Q, K, V 矩阵进行加权, 由于通道间权重经过 softmax 处理, 值域为 0~1 之间, 原始矩阵在与其相乘后, 一

定程度上弱化了特征。自注意力机制在后续处理过程中, 多次运用到 Q, K, V 进行计算, 导致特征的弱化影响被不同程度放大了。故使用 Self-ECA 构造协同注意力结构, 对网络的预测精度影响较小。

表 4 不同注意力融合方式的 HQ-S²ANet 性能比较 %

方式	AP(GTF)	AP(SH)	mAP
ECA	70.60	87.80	70.30
Self	65.80	87.30	70.20
ECA-Self	70.70	87.80	71.10
Self-ECA	72.00	88.20	71.30

3.4.2 下采样方式对比实验

表 5 对比了在 DOTA 数据集上 3 种下采样方式, 方式 1 在低层使用空洞系数为 2 的空洞卷积, 高层使用普通卷积; 方式 2 下采样全过程使用空洞系数为 2 的空洞卷积; 方式 3 在低层使用空洞系数为 2 的空洞卷积及均值池化对特征进行提取后将 2 种特征图相加传入下一层, 高层则使用普通卷积。

表 5 不同下采样方式的 HQ-S²ANet 性能比较 %

方式	AP(SV)	AP(HC)	mAP
1	72.60	50.20	70.70
2	70.90	53.50	70.20
3	73.20	56.90	71.30

根据结果分析, 下采样方式 1 与方式 2 均因空洞卷积的使用, 导致了网格隔影的产生, 在一定程度上弱化了特征, 这对以小目标为主的遥感图像检测影响较大。方式 3 利用均值池化与空洞卷积相加的方式, 在获得大感受野的同时弥补了空洞卷积对特征的弱化, 从而提升了网络性能。

3.4.3 锚框阈值对比实验

表 6 为 HQ-S²ANet 在控制变量的基础上, 在 DOTA 数据集中对比不同阈值下锚框匹配的检测结果。实验采用的模型为 HQ-S²ANet, 将阈值设定为 0.4, 0.6, 0.8 进行实验。从实验结果可以看出, 当阈值为 0.8 时, 部分高质量匹配仍会被负样本匹配重写, 由于未能有效抑制锚框低质量匹配发生, 导致召回率虽优于阈值为 0.4 的对照组, 但模型精度仍不理想。阈值为 0.4 时, 部分正样本锚框被忽略, 导致召回率过低影响网络精度。根据结果判断, 最优阈值处于 0.4~0.8 区间, 故设定阈值为 0.5, 0.7, 进一

步实验判断。阈值设定为 0.5,0.7 时,模型精度高于 0.4,0.8 且低于 0.6,故判断最优阈值为 0.6。基于模型精度的考虑,设定阈值为 0.6。

表 6 不同锚框阈值的 HQ-S ² ANet 性能比较 %			
阈值	召回率(GTF)	召回率(SH)	mAP
0.4	87.70	91.80	69.70
0.5	86.30	91.90	70.90
0.6	90.10	91.90	71.30
0.7	85.80	91.50	70.90
0.8	89.20	91.60	70.40

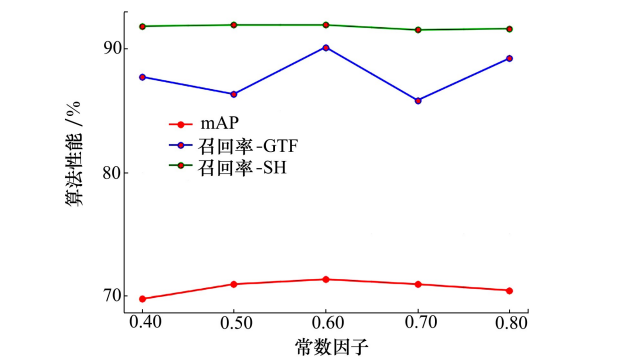


图 9 锚框阈值与 mAP 及部分类别召回率关系

3.4.4 算法复杂度分析

表 7 表明,基于 S²ANet 与改进的 HQ-S²ANet 算法相比,浮点型运算量仅增加 6.08G,参数量增加 2.61M,识别精度更高,较 S²ANet 提升 3.1%。与当前遥感图像目标检测领域主流算法 Rotated_retinanet 及 Rotated_reppoints 相比,虽参数量有所提升,但精度提升明显。由此证明:本文所提改进算法

HQ-S²ANet 在计算开销增加较少的前提下,大幅提升了算法识别能力。

表 7 模型算法复杂度与 mAP 统计			
算法	Flops/G	Params/M	mAP/%
Rotated_reppoints	194.32	36.85	59.60
Rotated_retinanet	221.96	39.02	65.70
S ² ANet	197.62	38.60	68.20
HQ-S ² ANet(本文)	203.70	41.21	71.30

4 结 语

遥感图像的目标检测在多个领域具有重要研究意义,为解决遥感图像目标成像小、分布拥挤、背景复杂的问题,提出一种基于感知延伸与锚框最适匹配的遥感图像目标检测算法 HQ-S²ANet。在特征融合下采样过程中将感知延伸卷积模块与常规卷积交替堆叠构建 HQFPN 模块,减少下采样过程中低层细节位置信息的丢失以增强网络对小目标的识别能力;为解决遥感目标图像分布拥挤的问题,在为真实边界框分配锚框时,提出高质量锚框匹配算法,在保证召回率的同时,减少低质量锚框匹配的产生;同时设计了协同注意力模块,以较低的计算开销增强了模型对目标与全局特征的建模能力,进而提升模型识别能力。通过在 DOTA 数据集上的实验,以及与先进模型的对比表明,所提算法可以提升在遥感图像中的目标检测能力,验证了所提模型的有效性。锚框匹配算法中,限定阈值对不同数据可适应性不足,下一步考虑网络通过学习得到最适阈值,在不影响精度的情况下提升方法的鲁棒性。

参考文献:

[1] 胡凡. 遥感图像中密集小目标检测算法设计与实现[D]. 南京:东南大学, 2021
HU Fan. Design and implementation of dense small target detection algorithm in remote sensing images[D]. Nanjing: Southeast University, 2021 (in Chinese)

[2] YANG X, YANG J, YAN J, et al. SCRDet: towards more robust detection for small, cluttered and rotated objects[C]//2019 IEEE/CVF International Conference on Computer Vision, Seoul, 2019: 8231-8240

[3] ZHANG G, LU S, ZHANG W. CAD-Net: a context-aware detection network for objects in remote sensing imagery[J]. IEEE Trans on Geoscience and Remote Sensing, 2019(99): 1-10

[4] 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

[5] HAN J, DING J, XUE N, et al. Redet: a rotation-equivariant detector for aerial object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 2786-2795

- [6] XIA G S, BAI X, DING J, et al. DOTA: a large-scale dataset for object detection in aerial images[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 3974-3983
- [7] DING J, XUE N, LONG Y, et al. Learning roi transformer for oriented object detection in aerial images[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 2849-2858
- [8] LI Y, LI S, DU H, et al. YOLO-ACN: focusing on small target and occluded object detection[J]. IEEE Access, 2020(99): 1
- [9] HAN J, DING J, LI J, et al. Align deep features for oriented object detection[J]. IEEE Trans on Geoscience and Remote Sensing, 2020, 60: 1-11
- [10] WEI L, DRAGOMIR A, DUMITRU E, et al. SSD: single shot multiBox detector[C] // European Conference on Computer Vision, 2016: 21-37
- [11] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C] // Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980-2988
- [12] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C] // IEEE Conference on Computer Vision and Pattern Recognition, 2017
- [13] LI W, CHEN Y, HU K, et al. Oriented reppoints for aerial object detection[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 1829-1838
- [14] 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
- [15] ANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11534-11542
- [16] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[J/OL]. (2016-04-16) [2022-06-15]. <https://arxiv.org/abs/1511.07122>
- [17] MEHTA S, RASTEGARI M, CASPI A, et al. Espnet: Efficient spatial pyramid of dilated convolutions for semantic segmentation [C] // Proceedings of the European Conference on Computer Vision, 2018: 552-568
- [18] HOU Liping, LU Ke, XUE Jian, et al. Shape-adaptive selection and measurement for oriented object detection[C] // the 36th AAAI Conference on Artificial Intelligence, 2022
- [19] YANG X, YAN J. Arbitrary-oriented object detection with circular smooth label[C] // European Conference on Computer Vision, 2020: 677-694
- [20] YANG Z, LIU S, HU H, et al. Reppoints: point set representation for object detection[C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 9657-9666
- [21] GUO Z, LIU C, ZHANG X, et al. Beyond bounding-box: convex-hull feature adaptation for oriented and densely packed object detection[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 8792-8801
- [22] ZHANG S, CHI C, YAO Y, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 9759-9768

Remote sensing target detection algorithm based on perceptual extension and anchor frame best-fit matching

HONG Wei¹, ZHAO Xiangmo², WANG Peng³, LI Xiaoyan²,
DI Ruohai², LYU Zhigang², WANG Chu²

(1.School of Ordnance Science and Technological, Xi'an Technological University, Xi'an 710021, China;
2.School of Electronic Information Engineering, Xi'an Technological University, Xi'an 710021, China;
3.Development Planning Service, Xi'an Technological University, Xi'an 710021, China)

Abstract: Aiming at the small imaging, complex background and crowded distribution of remote sensing image targets, a remote sensing image target detection algorithm (HQ-S²ANet) based on perceptual extension and anchor frame optimal matching is proposed by using the rotating target detection method S²ANet as a baseline network. Firstly, a cooperative attention(SEA) module is built to capture the relationship among the feature pixels when extending the model perception area to realize the relationship modeling between the target and the global. Secondly, the feature pyramid (FPN) feature fusion process is improved to form a perceptual extension feature pyramid module (HQFPN), which guarantees the low-level detail position information in the down sampling process when extending the perception area to enhance the model information capturing capability. Finally, a high-quality anchor frame is used to detect the target by using the high quality anchor frame as the baseline network. The high-quality anchor frame matching method (MaxIoUAssigner_HQ) is used to control the anchor frame truth value assignment by using a constant factor to ensure the recall rate while preventing the generation of low-quality anchor frame matching. The experimental results show that, under the DOTA dataset, the average accuracy(mAP) of HQ-S²ANet is improved by 3.1%, the parameters number increased by only 2.61M and the average recall(recall) is improved by 1.6% compared with the S²ANet algorithm, and the present algorithm effectively enhances the detection capability of the remote sensing image target.

Keywords: remote sensing image; feature fusion; anchor frame; rotation detection

引用格式:洪伟, 赵祥模, 王鹏, 等. 基于感知延伸与锚框最适匹配的遥感图像目标检测算法[J]. 西北工业大学学报, 2023, 41(4): 820-830

HONG Wei, ZHAO Xiangmo, WANG Peng, et al. Remote sensing target detection algorithm based on perceptual extension and anchor frame best-fit matching[J]. Journal of Northwestern Polytechnical University, 2023, 41(4): 820-830 (in Chinese)