



(12)发明专利申请

(10)申请公布号 CN 111179217 A

(43)申请公布日 2020.05.19

(21)申请号 201911229726.1

(22)申请日 2019.12.04

(71)申请人 天津大学

地址 300072 天津市南开区卫津路92号

(72)发明人 于瑞国 汪婧 王臣汉 李雪威

喻梅 姜汉 高洁 刘志强 应翔

(74)专利代理机构 天津市北洋有限责任专利代

理事务所 12201

代理人 刘子文

(51) Int. Cl.

G06T 7/00(2017.01)

G06N 3/04(2006.01)

G06N 3/08(2006.01)

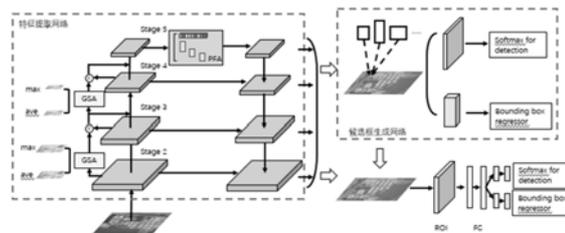
权利要求书1页 说明书6页 附图3页

(54)发明名称

一种基于注意力机制的遥感图像多尺度目标检测方法

(57)摘要

本发明公开一种基于注意力机制的遥感图像多尺度目标检测方法,针对遥感图像背景复杂、小目标众多,目标尺寸多样化的特点,采用特征金字塔网络进行不同尺度目标的预测。为了更加准确地关注到目标所在区域,在网络中引入注意力机制,有效地提高目标检测精度。其中,包括全局空间注意力模块和像素特征注意力模块。全局空间注意力模块通过从浅层特征中提取空间位置相关信息,并与深层特征融合,强化深层特征的位置表达能力。像素特征注意力模块以多尺度卷积核产生与输入同尺寸的特征图,并以通道注意力的方式为每层特征图赋予权值,得到细节信息表现良好的像素级注意力图在不增加计算量的同时增加了感受野。



1. 一种基于注意力机制的遥感图像多尺度目标检测方法,其特征在于,包括以下步骤:

步骤一,将遥感图像经过特征金字塔网络,提取不同尺度的特征,从底层网络中提取浅层特征,高层网络中提取深层特征;

步骤二,引入全局空间注意力模块(GSA,global spatial attention),从底层网络中提取浅层特征的注意力信息,对深层特征进行优化;

步骤三,将最后一层提取出的特征经过像素特征注意力模块(PFA,pixel feature attention),增大感受野,并减少细节信息损失;

步骤四,将步骤一产生的不同尺度的特征图,送到候选框生成网络(RPN,region proposal network)生成候选框;

步骤五,将候选框中的目标映射到不同尺度的特征图上,通过ROI pooling将每个特征图上的候选框变形到固定尺寸,并提取相关目标的特征;

步骤六,对候选框进行分类和边框回归。

2. 根据权利要求1所述一种基于注意力机制的遥感图像多尺度目标检测方法,其特征在于,步骤一具体包括以下步骤:采用特征金字塔网络,包括自底向上和自顶向下两个部分,自底向上为卷积和池化的过程,分辨率不断减小;将每一个块作为一个stage,共包括五个stage;自顶向下过程从stage 5的特征图开始进行上采样,然后与自底向上生成的相同大小的特征图进行横向连接;再采用3x3的卷积核对每个融合结果进行卷积,消除上采样混叠效应。

3. 根据权利要求2所述一种基于注意力机制的遥感图像多尺度目标检测方法,其特征在于,步骤二是在步骤一中进行自底向上提取特征的过程中进行的操作,首先对浅层的特征进行全局平均池化和全局最大值池化,然后通过逐像素相加将两张注意力图融合产生全局注意力图;最后将全局注意力图与深层特征图相乘。

4. 根据权利要求2所述一种基于注意力机制的遥感图像多尺度目标检测方法,其特征在于,步骤三是在步骤一自底向上的stage 5后引入像素特征注意力模块,通过通道注意力方式,为特征图的每一层赋予权值;将全局信息和细节信息进行了融合,提高网络对目标区域的定位能力。

5. 根据权利要求1所述一种基于注意力机制的遥感图像多尺度目标检测方法,其特征在于,步骤四是候选框生成网络通过采用anchors对提取的不同尺度的卷积特征图进行处理,提取不同大小的候选框,寻找可能包含目标的预定义数量的区域。

一种基于注意力机制的遥感图像多尺度目标检测方法

技术领域

[0001] 本发明属于人工智能、深度学习和计算机视觉领域,涉及目标检测和显著特征提取技术,尤其是一种基于注意力机制的遥感图像目标检测技术。

背景技术

[0002] 目前在相关技术中,目标检测技术主要是针对自然图像,方法主要分为两类:一种是two-stage的目标检测方法,应用较为广泛。Two-stage的方法以Faster R-CNN、FPN等方法为代表,将目标检测分为两个阶段:区域生成和目标分类。先利用预训练的CNN网络进行特征提取,然后通过区域生成网络RPN生成一定数量的候选区域,再对候选区域进行细粒度的物体分类和边框回归。Two-stage方法可以获得较为精确地检测结果,但是缺点是由于需要先生成大量不同尺度的候选框,所以不能达到实时的检测速度。

[0003] 另一类是one-stage的目标检测方法。One-stage方法可以直接在网络中提取特征来预测物体的分类和位置,虽然损失了一些准确率但是极大地提高了检测速度。例如以YOLO为代表,将待检测图像缩放到统一尺寸,为了检测不同位置的目标,将图像等分成网格,如果某个目标的中心落在一个网格单元中,此网格单元就负责预测该目标。YOLO的检测速度较FasterR-CNN有了近10倍的提升。但是one-stage方法的检测准确率不如two-stage类的方法,网格划分的大小对检测结果影响较大。

[0004] 在遥感图像目标检测方面,由于遥感图像具有背景复杂、角度多样化、尺寸大小不一,并且存在大量的小目标等特点,难度相对较大。但是现有遥感图像的目标检测还是采用跟自然图像一样的方法,没有针对遥感图像中存在的背景繁杂导致网络难以准确注意到关键的图像区域等问题提出解决方案。

发明内容

[0005] 本发明的目的是为了克服现有技术中的不足,提供一种基于注意力机制的遥感图像多尺度目标检测方法,能够有效的优化传统的针对自然图像的检测方法在面对遥感图像时难以在繁杂的背景中准确定位目标的问题。

[0006] 本发明的目的是通过以下技术方案实现的:

[0007] 一种基于注意力机制的遥感图像多尺度目标检测方法,包括以下步骤:

[0008] 步骤一,将遥感图像经过特征金字塔网络,提取不同尺度的特征,低层网络中提取浅层特征,高层网络中提取深层特征;

[0009] 步骤二,引入全局空间注意力模块(GSA,global spatial attention),从低层网络中提取浅层特征,用浅层特征的注意力信息去优化深层特征;

[0010] 步骤三,将最后一层提取出的特征经过像素特征注意力模块(PFA,pixel feature attention),增大感受野,并减少细节信息损失;

[0011] 步骤四,将步骤一产生的不同尺度的特征图,送到候选框生成网络(RPN,region proposal network)生成候选框;

[0012] 步骤五,将候选框中的目标映射到不同尺度的特征图上,通过ROI pooling将每个特征图上的候选框变形到固定尺寸,并提取相关目标的特征;

[0013] 步骤六,对候选框进行分类和边框回归。

[0014] 进一步的,步骤一具体包括以下步骤:采用特征金字塔网络,包括自底向上和自顶向下两个部分,自底向上为卷积和池化的过程,分辨率不断减小;将每一个块作为一个stage,共包括五个stage;自顶向下过程从stage 5的特征图开始进行上采样,然后与自底向上生成的相同大小的特征图进行横向连接;再采用3x3的卷积核对每个融合结果进行卷积,消除上采样混叠效应。

[0015] 进一步的,步骤二是在步骤一中进行自底向上提取特征的过程中进行的操作,首先对浅层的特征进行全局平均池化和全局最大值池化,然后通过逐像素相加将两张注意力图融合产生全局注意力图;最后将全局注意力图与深层特征图相乘。

[0016] 进一步的,步骤三是在步骤一自底向上的stage 5后引入像素特征注意力模块,通过通道注意力方式,为特征图的每一层赋予权值;将全局信息和细节信息进行了融合,提高网络对目标区域的定位能力。

[0017] 进一步的,步骤四是候选框生成网络通过采用anchors对提取的不同尺度的卷积特征图进行处理,提取不同大小的候选框,寻找可能包含目标的预定义数量的区域。

[0018] 与现有技术相比,本发明的技术方案所带来的有益效果是:

[0019] 1.本发明所提出的基于注意力机制的遥感图像多尺度目标检测方法,为遥感图像的目标检测提供了一种新思路,注意力机制的引入可以让网络更加准确的定位到目标所在位置,并且该方法具有较好的泛化性,对于自然图像的目标检测也具有较好的效果。

[0020] 2.GSA模块有效地减少了小目标的信息损失,提升了训练速度,加快了网络的收敛,对不同尺寸的目标均能达到较好的检测效果。

[0021] 3.PFA模块增加了感受野,在不增加计算量的前提下,减少了网络的误检,有效地提高了目标检测的准确率。

附图说明

[0022] 图1为本发明整体结构示意图。

[0023] 图2是全局空间注意力模块示意图。

[0024] 图3是像素特征注意力模块示意图。

[0025] 图4a和图4b分别是FPN方法和本发明方法(AMOD)的目标检测结果图。

具体实施方式

[0026] 以下结合附图和具体实施例对本发明作进一步详细说明。应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。

[0027] 本发明提出了一种基于注意力机制的遥感图像多尺度目标检测模型,如图1所示,为本发明的整体结构示意图,其中包括的两个模块如图2、图3所示,整体处理步骤如下:

[0028] 步骤S1.1:采用ResNet101作为特征提取网络,通过卷积和池化操作提取浅层特征和深层特征,浅层特征分辨率高,位于提取网络的低层,深层特征分辨率低,位于提取网络的高层。

[0029] 步骤S2.1:通过提取浅层的全局空间注意力,与深层特征融合,优化深层特征因分辨率低而导致的小物体信息丢失的问题。对于 $F \in \mathbb{R}^{H \times W \times C}$ 的浅层特征图,在维度C上分别进行全局平均池化和全局最大值池化得到 $F' \in \mathbb{R}^{H \times W \times 1}$ 的特征图,然后通过逐像素相加的方式融合,如公式(1)所示:

$$[0030] \quad F_{gsa} = \text{sigmoid}(\text{AvgPool}(F) \oplus \text{MaxPool}(F)) \quad (1)$$

[0031] 其中F为浅层特征图, F_{gsa} 为全局空间注意力的输出结果,sigmoid为sigmoid激活函数,AvgPool为全局平均池化,MaxPool为全局最大值池化, \oplus 为逐像素相加。

[0032] 步骤S2.2:将浅层特征进行下采样变成与深层特征一样的大小,再与其相乘,即将更准确的位置信息传递给深层特征,如公式(2)所示:

$$[0033] \quad F_{next} = \text{down}(F_{gsa}) \otimes F' \quad (2)$$

[0034] 其中down代表下采样, F' 为深层特征图, F_{gsa} 为浅层特征的全局注意力图, \otimes 为逐像素相乘。

[0035] 步骤S3.1:计算像素特征注意力的通道注意力分支。为获得特征图各通道的全局信息,先对特征图进行全局平均池化,特征图的大小变为 $1 \times 1 \times C$,再进行 1×1 的卷积来整合特征图,计算公式如(3)所示:

$$[0036] \quad M_G(F) = \sigma(f^{1 \times 1}(\text{AvgPool}(F))) \quad (3)$$

[0037] 其中 $f^{1 \times 1}$ 代表 1×1 的卷积,AvgPool为全局平均池化, σ 为relu激活函数, $M_G(F)$ 为通道注意力分支的输出。

[0038] 步骤S3.2:计算像素特征注意力的卷积分支。将特征图通过 1×1 的卷积,如公式(4)所示:

$$[0039] \quad M_C(F) = f^{1 \times 1}(F) \quad (4)$$

[0040] 其中 $f^{1 \times 1}$ 代表采用核为1的卷积, $M_C(F)$ 为卷积分支的输出。

[0041] 步骤S3.3:计算像素特征注意力的多尺度卷积分支。先采用 5×5 ,步长为2的卷积核作用于输入,得到尺寸减半的特征图。然后,再使用 5×5 ,步长为1的卷积核和尺寸为3,步长分别为2和1的卷积核,分别得到为输入尺寸 $1/2$ 和 $1/4$ 的特征图。最后对较小的特征图进行上采样,将二者融合到一起,再次上采样恢复到原特征图大小,公式如(5)所示:

$$[0042] \quad M_m(F) = \text{up}(\text{up}(f^{5 \times 5}(f^{5 \times 5}(F)) \oplus \text{up}(f^{3 \times 3}(f^{3 \times 3}(f^{5 \times 5}(F))))) \quad (5)$$

[0043] 其中up代表上采样, $f^{3 \times 3}$ 和 $f^{5 \times 5}$ 分别表示为卷积核为3和5的卷积, \oplus 为逐像素相加, $M_m(F)$ 为多尺度卷积分支的输出。

[0044] 步骤S3.4:将特征图F输入像素特征注意力模块的不同分支得到的特征图进行融合,如公式(6)、(7)所示:

$$[0045] \quad M(F) = \sigma(M_G(F) \oplus \sigma((M_C(F) \otimes M_m(F)))) \quad (6)$$

$$[0046] \quad F_{out} = M(F) \quad (7)$$

[0047] 其中 σ 为relu激活函数, \oplus 为逐像素相加, \otimes 为逐像素相乘, $M_G(F)$ 、 $M_C(F)$ 、 $M_m(F)$ 分别为特征图F通过全局通道注意力分支、卷积分支、多尺度卷积分支后得到的结果, F_{out} 为像素特征注意力模块的输出结果。

[0048] 步骤S3.5:自顶向下过程从stage 5的特征图开始进行上采样,然后与自底向上生

成的相同大小的特征图进行横向连接。再采用3x3的卷积核对每个融合结果进行卷积,产生不同尺度的特征图。

[0049] 步骤S4.1:采用3×3的卷积核对不同尺度的特征图进行卷积,同时以特征图的每个区域的中心点为中心学习多个不同尺度的不同长宽比的区域,称之为anchor。将每个anchor与真实值计算交并比IOU,如公式(8)所示,大于0.5的称为positive anchor,否则称为negative anchor。本发明所提出的方法anchor设定的面积尺寸为 $\{32^2, 64^2, 128^2, 224^2, 256^2\}$,三种不同的长宽比例为 $\{1:1, 1:2, 2:1\}$ 。

$$[0050] \quad IOU = \frac{box \cap gt}{box \cup gt} \quad (8)$$

[0051] 其中,box表示提取出的候选框,gt为每个目标真实的框。

[0052] 步骤S4.2:利用步骤S4.1中RPN网络生成的候选框和步骤S3生成的不同尺度的特征图,将候选框映射到特征图上。

[0053] 步骤S5:将步骤S4.2中的特征图中的候选框变形到固定大小,然后进行全连接操作,进行预测。对每个候选框,依据其大小选择一个stage进行预测,如公式(9)所示:

$$[0054] \quad k = \left\lfloor k_0 + \log_2(\sqrt{wh}/128) \right\rfloor \quad (9)$$

[0055] 其中,w和h分别表示候选框的宽和高。参数 k_0 为一正整数,表示 $w \times h = 128^2$ 的情况下候选框应该映射到的目标stage,在实验中,一般取固定值4。128为基准面积,是由数据集的尺寸来决定的。 k 为拟用来检测目标的stage。

[0056] 步骤S5.2:利用softmax进行具体类别的分类,如公式(10)所示:

$$[0057] \quad L_{cls}(p_i, p_i^*) = -\log[p_i^* p_i + (1-p_i^*)(1-p_i)] \quad (10)$$

[0058] 其中 p_i 为anchor预测为目标的概率, p_i^* 为真实值的标签, $L_{cls}(p_i, p_i^*)$ 为分类损失函数。

[0059] 步骤S5.3:利用L1 loss进行边框回归获得物体的精确位置,如公式(11)所示:

$$[0060] \quad L_{reg}(t_i, t_i^*) = \begin{cases} 0.5(t_i - t_i^*)^2, & \text{if } |t_i - t_i^*| < 1 \\ |t_i - t_i^*| - 0.5, & \text{otherwise} \end{cases} \quad (11)$$

[0061] 其中 $t_i = \{t_x, t_y, t_w, t_h\}$ 是一个向量,表示预测的框的4个坐标, t_i^* 是positive anchor对应的真实框的坐标向量, p_i 为anchor预测为目标的概率, p_i^* 为真实值的标签, $L_{reg}(t_i, t_i^*)$ 为边框回归损失函数。

[0062] 步骤S5.4:进行分类和边框回归的联合训练,总的损失函数如公式(12)所示:

$$[0063] \quad L(\{p_i\}, \{u_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (12)$$

[0064] 其中 p_i 为anchor预测为目标的概率, p_i^* 为真实值的标签, $t_i = \{t_x, t_y, t_w, t_h\}$ 是一个向量,表示预测的框的4个坐标, t_i^* 是positive anchor对应的真实框的坐标向量, $L_{cls}(p_i, p_i^*)$ 为分类损失函数, $L_{reg}(t_i, t_i^*)$ 为边框回归损失函数, λ 是平衡权重, N_{cls} 为mini-batch的大小, N_{reg} 为anchor的数量。

[0065] 为了验证方法具有较好的泛化性,本发明在遥感数据集NWPU(含10种目标类别)和RSOD(含4种目标类别),以及自然数据集PASCAL VOC 07(含20种目标类别)上进行了实验。

遥感数据集按照8:2的比例随机划分训练集和测试集。

[0066] 将每个类别的平均精度 (AP) 以及所有类别的平均精度均值 (mAP) 作为评价标准可定量评估检测的效果,值越大越好。如表1所示,实验结果表明,无论是对于遥感图像还是自然图像,本发明提出的模型均取得了最佳的检测效果。在NWPU数据集上,对于storage tank、ship等小目标以及bridge、harbor这种形状较长的目标检测效果提升较大,相比FPN方法提升了3.75%;在RSOD数据集上各类别检测效果均有不同程度的提升;在PASCAL VOC 07上,SSD300和YOLO3采用了PASCAL VOC 07和12作为训练集和测试集,而其余各个模型均只采用了PASCAL VOC 07作为训练集,可以看出本发明的AMOD模型在更少的训练集下却能达到最好的效果。本发明提出的模型比YOLO3高出0.8%,比FPN高出了1.3%。

[0067] 表1 NWPU-VHR 10、RSOD、PASCAL VOC 07数据集上不同方法检测结果对比

Class\ Method	Faster R-CNN	SSD300	RFCN	FPN	YOLO3	AMOD(Our)
	ResNet101	VGG16	ResNet101	ResNet101	Darknet53	ResNet101
Ship	94.39	83.74	95.83	95.26	91.07	96.11
Vehicle	82.02	38.43	83.97	91.37	73.56	95.34
Airplane	99.35	90.61	99.94	100	99.99	100
Bridge	74.42	98.18	75.89	83.35	92.86	93.40
Harbor	87.94	88.21	89.82	92.55	92.88	96.54
GTF	97.37	99.99	98.72	96.69	99.89	96.82
Tennis court	92.03	87.61	93.55	97.20	94.92	99.90
storage tank	66.65	77.37	66.91	84.86	93.40	95.84
BC	85.25	69.28	90.65	91.36	75.26	93.01
BD	93.37	97.44	97.67	94.72	99.40	97.88
mAP	87.28	83.09	89.30	92.73	91.28	96.48
Aircraft	83.54	71.89	83.69	91.93	88.38	94.28
Playground	97.81	98.58	99.54	98.16	99.65	99.66
Oiltank	98.11	90.72	98.44	99.18	98.91	99.58
Overpass	88.62	90.21	94.42	96.90	96.94	97.20
mAP	92.02	87.85	94.03	96.54	95.97	97.67
mAP	77.02	77.66	72.94	78.70	79.26	80.00

[0069] 表1中,通过比较目前最先进方法的目标检测的mAP可以得知,无论是对于遥感图像还是自然图像,本发明所提出的基于注意力机制的遥感图像多尺度目标检测方法均可以有效地提高目标检测的准确。尤其是在NWPU数据集上,比Faster R-CNN提升了9.2%,比YOLO3提升了5.2%。

[0070] 在训练速度上,基于注意力机制的遥感图像多尺度目标检测方法可以更快的定位到目标所在区域,提升了训练速度,加快了收敛。如表2所示,相比较于基准FPN方法,改进的AMOD方法在速度上有了明显的提升。在NWPU数据集上提升了37.5%的训练速度,在RSOD上提升了20%,在PASCAL VOC 07上提升了35%。

[0071] 表2 NWPU、RSOD、PASCAL VOC 07数据集上不同方法训练速度对比

	FPN迭代次数	AMOD迭代次数
NWPU	40000	25000

RSOD	50000	40000
PASCAL VOC 07	200000	130000

[0073] 部分遥感图像检测结果如图4a和图4b所示,可以看出FPN方法没有检测出所有的桥和网球场,而本发明改进的AMOD方法由于添加了像素特征注意力模块,增加了感受野,减少了漏检,有效的提升了目标检测精度。

[0074] 本发明并不限于上文描述的实施方式。以上对具体实施方式的描述旨在描述和说明本发明的技术方案,上述的具体实施方式仅仅是示意性的,并不是限制性的。在不脱离本发明宗旨和权利要求所保护的范围情况下,本领域的普通技术人员在本发明的启示下还可做出很多形式的具体变换,这些均属于本发明的保护范围之内。

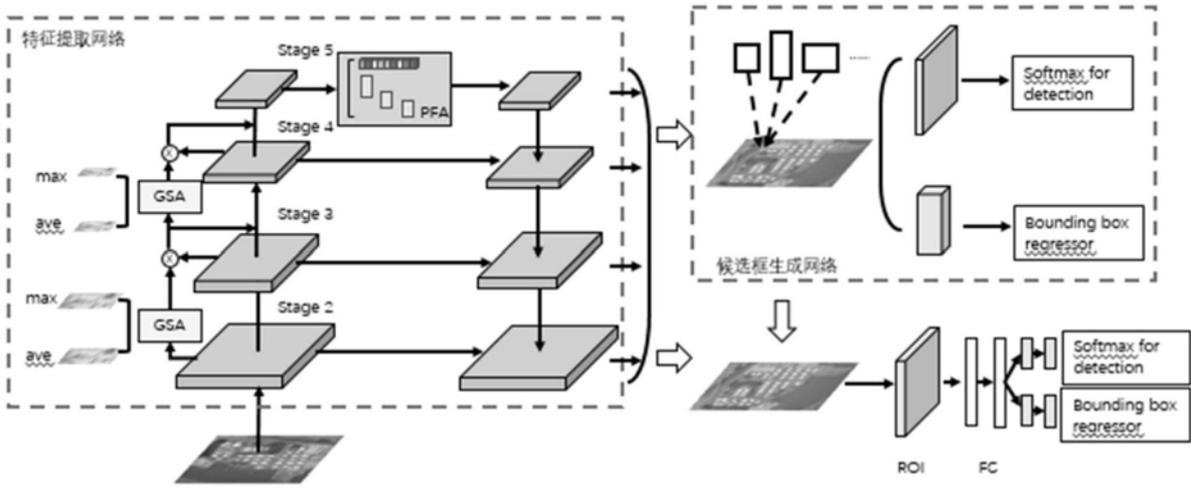


图1

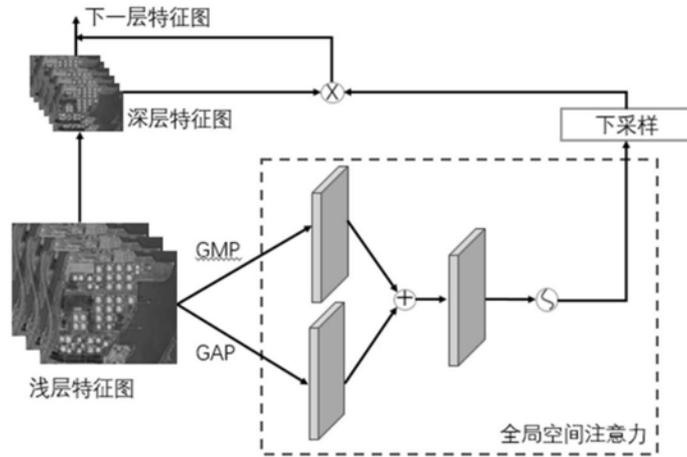


图2

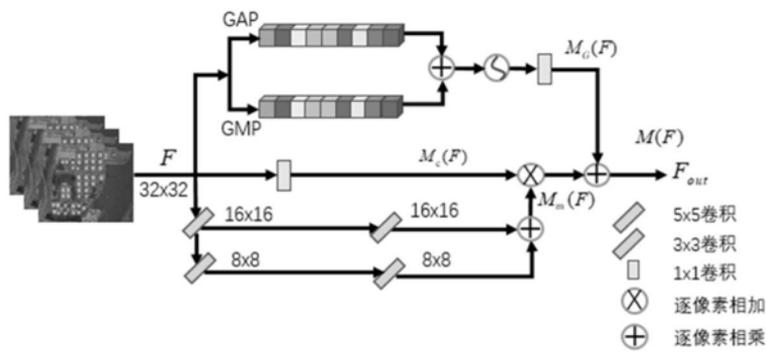


图3

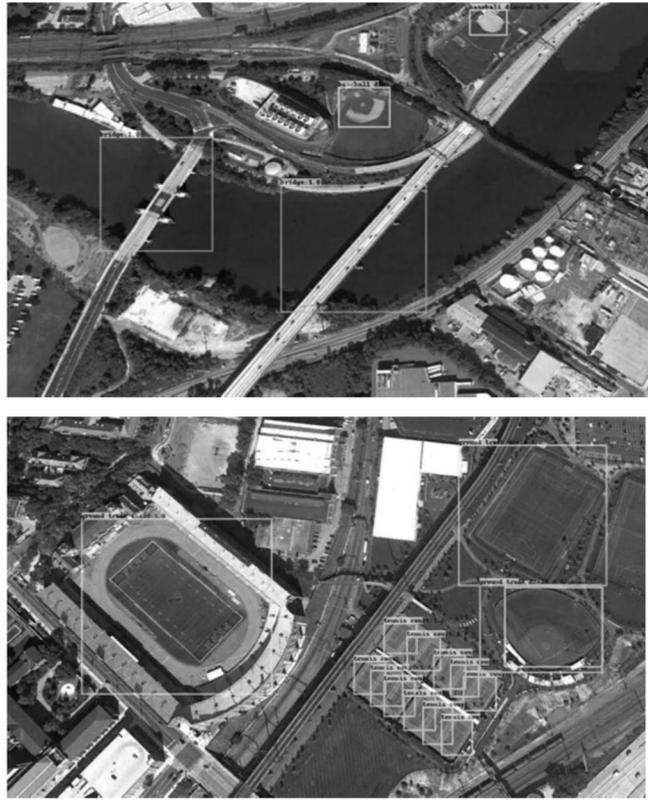


图4a

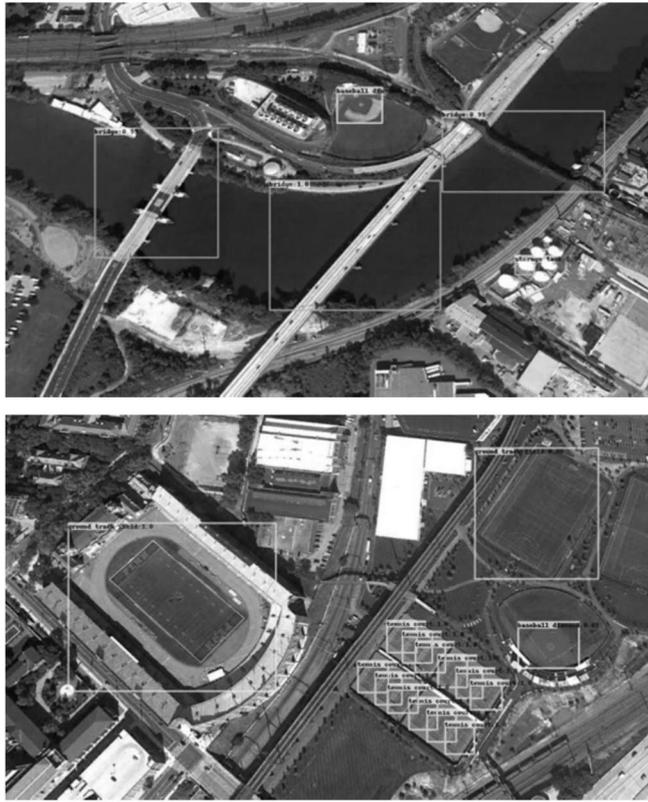


图4b