

基于显著性特征和角度信息的遥感图像目标检测



袁星星 吴 秦

江南大学人工智能与计算机学院 江苏 无锡 214122

江南大学江苏省模式识别与计算智能工程实验室 江苏 无锡 214122

(6171914011@stu.jiangnan.edu.cn)

摘要 遥感图像中的目标具有密集性、多尺度和多角度等特性,这使得遥感图像多类别目标检测成为一项具有挑战性的课题。因此,文中提出了一种新的端到端的遥感图像目标检测框架。该框架通过提取显著性特征和不同卷积通道之间的相互关系来增强目标信息,抑制非目标信息,从而提高特征的代表能力。同时,在不增加模型参数的情况下,在卷积模块中添加多尺度特征模块来捕获更多的上下文信息。为了解决遥感图像中目标角度多变这一问题,该框架在区域建议网络中加入了角度信息,得到有角度的矩形候选框,并在训练过程中添加注意力损失函数来引导网络学习显著性特征。该框架在公开的遥感图像数据集上进行了相关验证,在水平任务框和方向任务框上的实验结果证明了所提方法的有效性。

关键词: 遥感图像;目标检测;多尺度特征;通道自学习;显著性学习

中图分类号 TP391

Object Detection in Remote Sensing Images Based on Saliency Feature and Angle Information

YUAN Xing-xing and WU Qin

School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, Jiangsu 214122, China

Jiangsu Provincial Engineering Laboratory for Pattern Recognition and Computational Intelligence, Wuxi, Jiangsu 214122, China

Abstract Multi-class object detection of remote sensing images is a challenging subject since objects in remote sensing images are dense, multi-oriented and multi-scale. This paper presents a new framework for object detection in remote sensing images. The framework enhances object information, suppresses non-object information, and improves the ability of feature representation by extracting salient features and the relationship between different convolutional channels. Meanwhile, multi-scale features are added to the convolution module to capture more context information without adding extra parameters to the model. To solve the problem of variable object angle in remote sensing image, we add angle information to the Region Proposal Network (RPN) to get oriented rectangular object proposals. In the training stage, the attention loss function is added to guide the significance of network learning. The proposed framework is validated on the public remote sensing image data set, and the experimental results on the horizontal boxes task and the oriented boxes task prove the effectiveness of the proposed method.

Keywords Remote sensing image, Object detection, Multi-scale feature, Channel self-learning, Saliency learning

1 引言

随着传感器技术和航空航天遥感技术的发展,研究人员可以从高质量的遥感图像中便利地获取图像内容和丰富的语义信息。基于遥感影像的地理空间目标检测是分析目标相关信息的重要方法^[1-2]。目前,研究人员提出了很多算法来检测遥感和卫星图像中不同类型的对象^[3],如城市区域检测^[4]、建筑物检测^[5]、油桶检测^[6]、车辆检测^[7]、飞机检测^[8]。

与相机在地面上从水平视角获得的自然图像不同,遥感图像是航拍俯视获得的,因此很容易受到天气和照明等条件

的影响。与自然图像相比,高分辨率遥感图像中的物体往往排列密集且角度是随机的,如图 1(a)和图 1(b)所示。同时,高分辨率遥感图像背景复杂,图像中的物体尺寸小且尺度多变。综上所述,遥感图像目标检测成为了一项具有挑战性的任务。随着近年来深度学习的快速发展,卷积神经网络(Convolutional Neural Networks, CNN)已成为一种新的特征表示方法,大大提升了目标检测的性能^[9]。目前基于 CNN 的目标检测算法大致可以分为两种:基于区域的 CNN 方法(如 R-CNN^[10], Fast R-CNN^[11]和 Faster R-CNN^[12])和无区域方法(如 YOLO^[13]和 SSD^[14])。R-CNN, Fast R-CNN 和 Faster

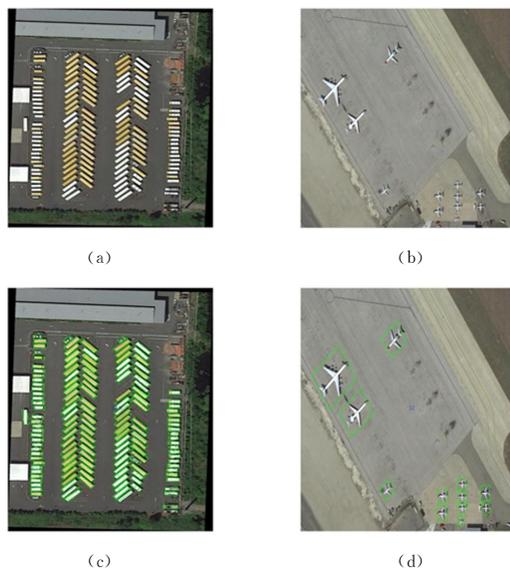
到稿日期:2019-12-02 返修日期:2020-05-15 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(61972180);江苏省自然科学基金(BK20181341)

This work was supported by the National Natural Science Foundation of China(61972180) and Natural Science Foundation of Jiangsu Province of China(BK20181341).

通信作者:吴秦(qinwu@jiangnan.edu.cn)

R-CNN 是典型的基于深度学习的目标检测算法,它们通过提取候选区域对相应区域进行分类和回归。其中,R-CNN 首先用选择策略(Selective Search,SS)方法生成区域建议,然后对每个区域建议进行特征提取,最后将提取好的特征放到 SVM 分类器和回归器中。R-CNN 虽然提高了目标检测的准确率,但是它的时间开销非常大。Fast R-CNN 则是基于 R-CNN 的改进,它通过同时从 CNN 分类器输出边界框和相应的标签来改进目标检测过程。Fast R-CNN 在进一步提高目标检测的准确率的同时也节省了时间开销。然而,因为 Fast R-CNN 仍然使用选择策略的方法生成区域建议,所以检测过程受到了时间限制。为了避免在生成区域建议时采用耗时的选择策略方法,Faster R-CNN 目标检测算法提出了区域建议网络(Region Proposal Network,RPN)。区域建议网络是全卷积网络(Fully Convolutional Networks,FCN),利用“锚”框来有效地预测候选框。Faster R-CNN 在区域建议生成过程和检测过程之间共享特征,大大减少了时间消耗。与基于区域的 CNN 方法相比,类似 YOLO 的无区域方法将目标检测作为回归问题,单个网络可以直接预测边界框和对应类别,因此检测速度很快。虽然 SSD 也将目标检测视为回归问题,但是其将小卷积滤波器应用于像素图以预测类别得分和物体偏移,而不是应用于全连接的层。此外,不同尺度的特征图用于不同尺度的目标检测,因此,与 YOLO 相比,SSD 的检测精度得到大大提高,但检测效果却不如 Faster R-CNN。



注:图1(a)和图1(b)分别为遥感卫星图像中的大型车辆和飞机,图1(c)和图1(d)为其对应的检测结果,绿色框中的内容为检测到的对象

图1 部分遥感卫星图像样本及本文方法的检测结果
(电子版为彩色)

Fig.1 Some samples of remote sensing satellite images and their corresponding detection results with proposed method

近年来,Xia 等^[15]将 Faster R-CNN^[11]目标检测框架应用到遥感图像的检测中,使用区域建议网络生成“锚”框来有效地预测候选框,但是此方法存在候选框冗余和漏检的问题。为了解决高分辨率图像人工标注样本有限的问题,Han 等^[16]提出了 R-P-Faster R-CNN 框架,该框架在 ImageNet 预训练

好的模型上进行训练,在卫星图像上取得了较好的检测效果。由于卷积神经网络中卷积核的几何形状是固定的,因此其模拟几何变换是有限的。为了打破这一局限,Xu 等^[9]将可变形神经网络应用于遥感图像目标检测,但候选框仍然存在冗余问题。Guo 等^[17]运用多尺度卷积神经网络来解决遥感图像中的多尺度问题,该方法对小目标的检测效果有一定提升,但对多类别的检测结果并不理想。在特征提取阶段,为了更有效地提取目标的有效信息并抑制无效信息,Hu 等^[18]提出了挤压激励模块,将其用于强调有用信息并抑制无用信息。但是,对于稠密且多尺度的卫星遥感图像,这些框架的检测结果均不理想。

针对遥感图像目标检测中物体旋转角度多、尺度变化大、物体排列密集、背景复杂等问题,本文提出了一种基于显著性特征和角度信息的遥感图像检测框架。该框架的主要创新点如下:

- (1)通过旋转区域建议网络,缓解了候选框冗余和定位不准确的问题。
- (2)使用具有不同内核大小、不同扩张率的网络来捕获多尺度信息,获得更多的上下文信息,解决了遥感图像中物体多尺度的问题。
- (3)提出显著性学习模块和通道自学习模块来加强目标信息,削弱非目标信息,提高特征的代表能力。该框架在训练和检测时,将图片重构成不同大小以增强模型的鲁棒性。

2 显著性特征和角度信息网络

本文方法的总体步骤如图 2 所示,检测结果样本如图 1(c)和图 1(d)所示。首先,利用基础网络 resnet-101 对图像进行特征提取后得到 C_4 。一方面,本文通过通道自学习模块处理得到各通道的权重以反映各通道的重要程度;另一方面,通过通道显著性学习模块处理得到空间权重以反映目标的位置信息。然后,根据两者的权重得到 C_4 的权重,即得到最终特征。最后,将得到的最终特征输入旋转 RPN,生成旋转候选框。

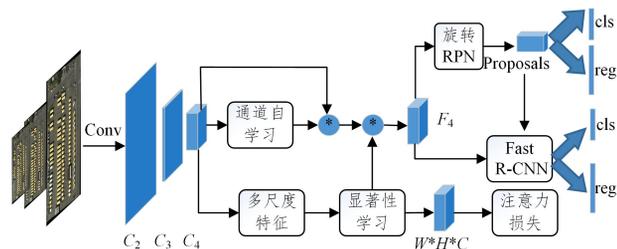


图2 基于显著性特征和角度信息的网络结构图

Fig.2 Network structure based on saliency features and angle information

2.1 多尺度特征

虽然图像可以通过卷积、池化操作来增大感受野,获得更丰富的上下文信息,但是这同时会导致图像信息损失。空洞卷积向卷积层引入了一个称为“扩张率”的新参数,通过填充卷积核,能够在扩大感受野的同时不丢失信息,使用不同扩张率的空洞卷积可以生成不同的感受野,具体如图 3 所示。

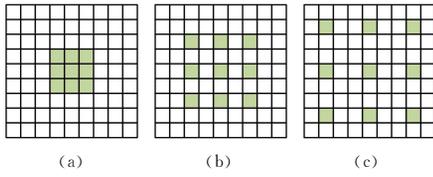


图3 不同扩张率的空洞卷积示意图

Fig.3. Diagram of dilated convolution with different expansion rates

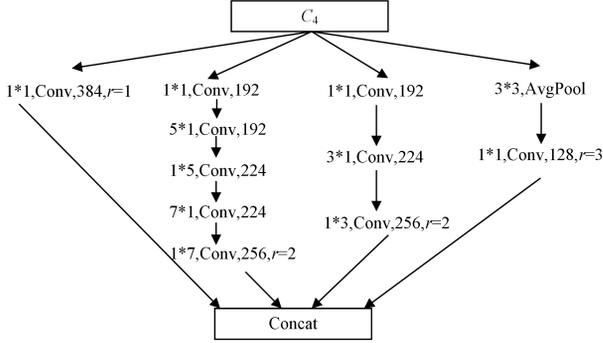


图4 多尺度特征的结构图

Fig. 4 Structure of multi-scale feature

假设原始特征为 F , 图 3(a)–图 3(c) 分别对应 1, 2, 3 的扩张率。当卷积核为 3×3 , 使用扩张率为 1 的空洞卷积时, 生成特征 F_1 , 如图 3(a) 所示, F_1 上 1 点相对 F 的感受野为 3×3 ; 当使用扩张率为 2 的空洞卷积时, 生成特征 F_2 , 如图 3(b) 所示, F_2 上 1 点相对 F 的感受野为 5×5 ; 当使用扩张率为 3 的空洞卷积时, 生成特征 F_3 , 如图 3(c) 所示, F_3 上 1 点相对 F 的感受野为 7×7 。为了在更大的区域中捕获更多的上下文信息, 在不增加参数数量的情况下, 本文方法在卷积操作中添加空洞卷积, 如图 4 所示, 多尺度特征由 4 个分支组成。由于遥感图像中目标的占比相对较小, 经过对比不同扩张率的实验结果, 最后将在第一个分支的卷积层中的扩张率设置为 1, 第二个和第三个分支的最后一层卷积中的扩张率设置为 2, 第四个分支的最后一层卷积中的扩张率设置为 3。

2.2 通道自学习

在遥感图像中, 需要检测的物体在整个图像中的占比非常小。卷积核作为神经网络的核心, 通常在局部感受野中聚集空间信息和特征维度信息, 最终得到全局信息, 但是却忽略了通道之间的信息。因此, 我们可以通过自动学习来获取每个特征通道的重要程度。通道自学习模块由两个部分组成, 如图 5 中上虚线框中的内容所示。

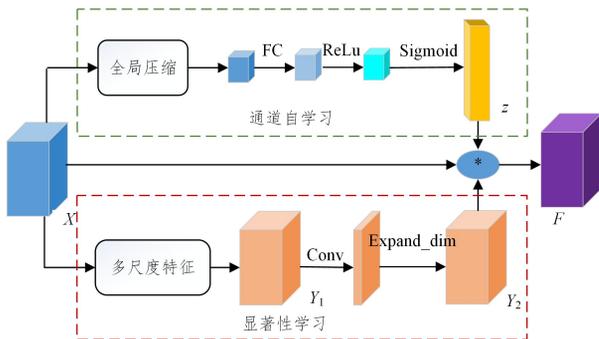


图5 特征处理的结构图

Fig. 5 Structure of feature processing

首先, 顺着空间维度, 对基础网络 resnet-101 所提取的特征 C_4 (记为 X) 进行全局压缩, 对每个通道的特征求均值, 获得每个通道的全局信息, 输出的维度和输入的特征通道数相同。其次, 通过参数 w 为每个特征通道生成权重, 有效的特征权重较大, 无效或效果小的特征权重小。最后, 将学习得到的权重在通道维度上逐项加权到先前的特征上, 对特征进行自适应调整。假设特征 X 的大小为 $W \times H \times C$, 则全局压缩的过程用式 (1) 和式 (2) 表示, 自适应调整的过程用式 (3) 表示。

$$y^k = \frac{1}{W * H} \sum_{i=1}^W \sum_{j=1}^H u^k(i, j) \quad (1)$$

$$z = \xi(W_2 \delta(W_1 y)) \quad (2)$$

$$X1 = X * z \quad (3)$$

其中, $u^k(i, j)$ 是特征 X 的第 k 个通道上第 i 行第 j 列的元素, $y = [y^1, y^2, \dots, y^k, \dots, y^C]^T$ 。 ξ 和 δ 分别表示 ReLu 激活函数和 Sigmoid 激活函数。其中, W_1 和 W_2 是权重向量, $W_1 \in R^{c/r}$, $W_2 \in R^c$, c 是通道数, r 是通道缩小比例, 其值设为 $16^{[18]}$ 。

2.3 显著性学习

过多的噪声会干扰目标信息, 模糊目标之间的边界, 导致漏检和虚警增加。Wang 等^[19] 通过设计单通道注意遮罩, 解决了面部遮挡问题, 增强了目标信息, 弱化了非目标信息。受该方法的启发, 我们在模型中也加入了显著性学习模块, 其过程如图 5 中下虚线框中的内容所示。原始特征经过多尺度特征后得到特征 Y_1 , 对 Y_1 进行卷积操作, 得到两通道的显著特征图, 它们分别代表前景和背景的得分。然后, 对显著特征图进行 softmax 操作, 经 softmax 操作后, 显著性特征值在 $[0, 1]$ 之间。选择一个通道与特征 Y_1 相乘, 得到新的特征, 特征 Y_1 上对应目标区域的特征值将变大, 非目标区域的特征值将变小, 使目标信息得到加强, 非目标信息受到抑制。

2.4 旋转框

带有水平框的 RPN^[12] 被广泛用于生成对象的候选位置。然而, 由于视角的关系, 遥感图像中的目标大多是倾斜的。如果在遥感图像中仍然使用水平候选框, 那么在后续的处理中会丢失许多包含目标的候选框, 导致定位不准确。因此, 我们在遥感影像中使用旋转候选框作为目标框。本文将提案/对象的边界框表示为定向矩形 (Oriented Rectangle)。我们用 (x, y, w, h, θ) 来表示一个以坐标 (x, y) 为中心的长方形, 其宽度为 w , 高度为 h , 角度为 θ 。因为边界框具有一定的角度, 所以在轴对齐边界框上计算交并比 (IoU) 可能会导致倾斜边界框的 IoU 不准确, 进而影响识别和定位。为了解决这一问题, Tang 等^[20] 提出了三角测量倾斜 IoU 的计算方法。我们使用旋转非最大抑制 (Non-Maximum Suppression, NMS) 进行后续处理, 并针对不同的对象类设置不同的旋转 NMS 阈值。边界框回归步骤中使用的参数定义如式 (4) 和式 (5) 所示:

$$\begin{cases} t_x = (x - x_a) / w_a, & t_y = (y - y_a) / h \\ t_w = \log(w / w_a), & t_h = \log(h / h_a) \\ t_\theta = \theta - \theta_a \end{cases} \quad (4)$$

$$\begin{cases} t_x^* = (x^* - x_a) / w_a, & t_y^* = (y^* - y_a) / h_a \\ t_w^* = \log(w^* / w_a), & t_h^* = \log(h^* / h_a) \\ t_\theta^* = \theta^* - \theta_a \end{cases} \quad (5)$$

其中, (x, y) 是预测框的中心, w 和 h 分别表示预测框的宽度和高度, θ 是预测框的旋转角度; (x_a, y_a) 是“锚”的中心, w_a , h_a 和 θ_a 表示“锚”的宽度、高度和旋转角度, (x^*, y^*) 是真实框的中心; $t_x, t_y, t_w, t_h, t_\theta$ 表示预测框与“锚”的偏移量, $t_x^*, t_y^*, t_w^*, t_h^*, t_\theta^*$ 表示真实框与预测框的偏移量。

2.5 损失函数

本文使用多任务损失函数进行端对端的训练。在分类阶段采用交叉熵损失函数, 在回归损失时使用 $L1$ 平滑损失函数。在显著性学习模块中, 为了引导网络学习显著性特征, 我们采用了监督学习的方法。首先, 根据真实标注框得到一个二值图作为标签, 然后将二值图和显著性特征图的交叉熵损失作为注意力损失。损失函数的定义如式(6)所示:

$$L = \lambda_1 \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda_2 \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) + \lambda_3 \frac{1}{i * j} \sum_i \sum_j L_{att}(s_{ij}, s_{ij}^*) \quad (6)$$

其中, p_i^* 是真实框的标签, 如果“锚”包含物体, 则 p_i^* 值为 1, 否则为 0, p_i 表示第 i 个“锚”预测为物体的可能性的值, i 是批大小中“锚”的索引。 $t_i = [t_x, t_y, t_w, t_h, t_\theta]$, $t_i^* = [t_x^*, t_y^*, t_w^*, t_h^*, t_\theta^*]$ 。 $L_{cls}(p_i, p_i^*)$ 是分类结果的对数损失, L_{reg} 是回归损失。 s_{ij}, s_{ij}^* 分别表示显著特征图像素的预测和标签。分类损失 L_{cls} 、回归损失 L_{reg} 和注意力损失 L_{att} 由平衡参数 λ_1, λ_2 和 λ_3 加权, 实验中 λ_1, λ_2 和 λ_3 设为 2, 4, 1。

3 实验及结果分析

3.1 数据集

为了验证本文提出的框架的有效性, 我们在 DOTA 数据集和 RSOD 数据集上进行了实验。DOTA 是用于航空图像目标检测的大规模数据集^[15], 共有 2806 张图片, 包含 188282 个实例和 15 个类别。每张图像的大小不一样, 在 $800 * 800$ 到 $4000 * 4000$ 像素之间, 包含各种不同比例、方向和形状的对象, 它在标记各个目标时使用 4 点坐标 $(x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4)$ 。训练集、验证集和测试集各占 1/2, 1/6 和 1/3。由于每张图像的大小不一样, 且长度和宽度不固定, 我们将每个图像划分为固定大小为 $800 * 800$ 的子图像, 以便从图像中检测目标。

RSOD 数据集共有 4 个类别, 分别是油桶、飞机、桥和操场, 其中油桶有 165 张图片, 包含 1586 个实例; 飞机有 446 张图片, 包含 4993 个实例; 桥有 176 张图片, 包含 180 个实例; 操场有 189 张图片, 包含 191 个实例。

3.2 实验设置

本文所有的实验都是基于 Faster R-CNN 这一深度学习框架而设计的。为了解决遥感卫星图像标记样本数量有限的问题, 我们使用了在 ImageNet 数据集上预训练的 ResNet-101 初始化底层网络。在 DOTA 数据集上共训练了 36 万次, 前 15 万次的学习率为 $3 * 10^4$; 15 万次到 25 万次的学习率为 $3 * 10^6$; 剩余的学习率为 $3 * 10^8$ 。在旋转 RPN 阶段, “锚”的长宽比分别设置为 1:1, 1:2, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 1:9。在训练阶段, 每幅图像的大小被随机重构为 $800 * 800$, $600 * 600$, $1000 * 1000$ 和 $1100 * 1100$ 4 个尺度; 在测试阶

段, 每张图片也在这 4 个尺度的基础上分别进行测试。因为 DOTA 数据集有两个任务边界框, 分别是水平边界框和方向边界框, 所以对于 DOTA 数据集中的图像, 我们分别评价两种边界框(水平边界框和方向边界框)的检测效果。

3.3 评价指标

我们在 DOTA 数据集上采用平均精度(AP)和类平均精度(mAP)来评估模型的性能; 在 RSOD 数据集上则采用 F1 来度量。首先定义精度(precision)、召回率(recall)和 F1, 如式(7)所示:

$$\begin{aligned} precision &= \frac{TP}{TP + FP} \\ recall &= \frac{TP}{TP + FN} \\ F1 &= \frac{2 * precision * recall}{precision + recall} * 100\% \end{aligned} \quad (7)$$

设置步长为 0.1, 取 $recall = [0, 0.1, 0.2, \dots, 1]$ 时对应的 precision 值, 这些 precision 值的平均值即为 AP。mAP 则是计算所有类的 AP 的平均值。

当“锚”框与真实框区域之间的面积交并比大于 0.7 时, 则认为“锚”框为正确检测的目标框(TP), 否则认为其是错误检测的目标框(FP)。此外, 如果有多个“锚”框与真实框区域之间的面积交并比大于 0.7, 则仅将具有最大面积交并比的矩形框视为正确检测的目标框(TP), 其他的矩形框视为误检框(FP)。FN 表示正确检测框被误判为错误检测框的数目。

这里显示的测试结果是将我们的预测结果提交到 DOTA 数据集的官方网站而获得的¹⁾。

3.4 实验结果

为了探讨本文所提框架的合理性, 我们对第 3 节所提的框架结构进行了不同的设置。首先, 我们构建了遥感图像的基本目标检测框架 Faster R-CNN, 该框架的实验结果在表 1 中用“O”表示, 在原框架的基础上使用旋转 RPN 的检测结果, 在表 1 中用“OO”表示。其次, 在旋转 RPN 框架中加入了通道自学习模块(简称“OOC”)。然后, 显著性学习模块被添加到 OOC 框架中(简称“OOCs”), 并通过多尺度图片进行训练。最后, 将多尺度特征模块添加到 OOCs 框架中(简称“OOCsM”)。

对于这 5 种不同的设置, 我们在 DOTA 数据集上分别进行了实验。在 DOTA 数据集上使用方向边界框作为真实框的结果, 如表 1 所列。Faster R-CNN 框架的平均精度为 52.9%, 在加入旋转 RPN 后的结果提高了 8.2%; 加入通道自学习处理后, 平均精度比仅基于旋转 RPN 的结果提高了 1.1%; 与仅基于 OOC 的结果相比, 增加了显著性学习模块的框架的平均精度提高了 5.9%。为了验证多尺度特征模块的有效性, 我们设置了两种扩张率, 对应结果如表 1 中的 OOC-SM123 和 OOCsM135。其中, OOCsM123 代表在第一个分支的卷积层中的扩张率为 1, 第二个和第三个分支的最后一层卷积中的扩张率为 2, 第四个分支的最后一层卷积中的扩张率为 3。OOCsM135 代表在第一个分支的卷积层中的扩张率为 1, 第二个和第三个分支的最后一层卷积中的扩张率为 3, 第四个分支的最后一层卷积中的扩张率为 5。实验结果显

¹⁾ <http://captain.whu.edu.cn/DOTAweb/evaluation.html>

示,扩张率为 1,3,5 的结果略高于扩张率为 1,2,3 的结果,其平均精度比仅基于 OOCs 的结果提高了 0.8%。

在 DOTA 数据集上使用水平边界框作为真实框的结果如表 2 所列。Faster R-CNN 框架的平均精度为 60.5%,在加入旋转 RPN 后的结果提高了 2.1%。加入通道自学习处理模块后,平均精度比仅基于旋转 PRN 的结果提高了 2.5%。

与仅基于 OOC 的结果相比,增加了显著性学习模块的平均精度提高了 8.2%。最后在加入多尺度特征模块时,我们设置了 1,2,3 和 1,3,5 这两种扩张率,实验结果如表 2 中的 OOC-SM123 和 OOC-SM135,扩张率为 1,2,3 的结果略高于扩张率为 1,3,5 的结果,其平均精度比仅基于 OOCs 的结果提高了 1.2%。

表 1 DOTA 数据集上的方向边界框任务

Table 1 Oriented bounding boxes tasks on DOTA data set

(单位:%)

Method	mAP	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC
O	52.9	79.1	69.1	17.2	63.5	34.2	37.2	36.2	89.2	69.6	59.0	49.4	52.5	46.7	44.8	46.3
OO	61.1	80.7	72.9	35.6	59.3	55.0	44.4	53.6	89.4	72.6	73.7	54.7	58.1	55.2	63.7	47.8
OOC	62.2	81.1	74.6	37.4	67.4	55.8	47.3	53.4	90.0	74.3	74.3	53.3	60.7	59.1	62.8	40.9
OOCs	68.1	89.4	81.0	43.4	71.8	67.8	51.3	63.7	89.1	80.0	84.4	62.2	60.7	64.7	66.1	45.0
OOC-SM123	68.9	89.2	81.6	44.0	70.5	65.6	53.3	64.4	89.8	79.8	84.9	62.2	63.9	65.8	65.4	52.5
OOC-SM135	69.1	89.1	79.7	45.0	72.4	67.3	51.9	63.9	88.4	78.3	85.2	64.9	61.9	63.5	66.4	59.6
NLoss	68.3	88.3	80.3	43.1	72.6	65.4	49.2	63.2	88.3	80.0	84.2	61.3	63.3	64.2	66.6	53.7

表 2 DOTA 数据集上的水平边界框任务

Table 2 Horizontal bounding boxes tasks on DOTA data set

(单位:%)

Method	mAP	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC
O	60.5	80.3	77.6	32.9	68.1	53.7	52.5	50.0	90.4	75.1	59.6	57.0	49.8	61.7	56.5	41.9
OO	62.6	79.2	72.8	43.1	56.3	59.6	54.3	59.9	81.6	68.4	73.7	50.2	57.9	65.9	69.5	46.1
OOC	65.1	80.7	75.0	45.8	65.8	60.4	56.5	60.6	81.6	70.5	74.8	55.2	60.6	67.0	74.5	46.3
OOCs	73.3	90.0	80.1	54.3	69.0	71.7	74.1	78.1	90.4	80.7	84.9	61.8	60.7	75.8	74.7	53.3
OOC-SM123	74.5	90.0	82.5	53.7	68.7	71.9	74.1	77.9	90.6	85.5	84.9	63.4	63.0	76.1	75.5	60.0
OOC-SM135	74.1	89.8	83.1	53.5	69.4	72.2	74.6	78.0	89.5	80.0	85.3	63.7	62.0	75.3	75.6	60.6
NLoss	73.8	89.5	81.3	53.8	69.9	71.8	74.1	77.7	90.4	82.0	84.2	63.3	63.4	75.7	76.5	53.3

另外,为了验证注意力损失函数的有效性,我们对损失函数中不添加注意力损失函数的情况进行了实验,即损失函数中 $\lambda_3=0$,其实验结果如表 1、表 2 中的“NLoss”项所列。对比表 1 中的最后两行可知,在方向边界框任务中添加注意力损失函数比未添加注意力损失函数时的精度提高了 0.8%。对比表 2 中的最后两行可知,在水平边界框任务中,添加注意力损失函数比未添加注意力损失函数时的精度提高了 0.3%。由此可见,注意力损失函数对于提升目标检测的精度有一定的效果。

3.5 与其他算法的比较

本小节将本文所提的框架与 YOLOv2^[13]、R-FCN^[21]、

SSD^[14]、F-H(带水平框的 Faster R-CNN)和 F-O(带方向框的 Faster RCNN)^[15]进行比较。不同方法在 DOTA 数据集上方向边界框的比较结果如表 3 所列。

本文方法在 DOTA 数据集上旋转边界框任务的平均精度为 68.9%,相比其他算法其平均精度最高,比目前最好的结果高出了 15.1%,特别是小型车的平均精度比目前最高的平均精度高出了 33.2%,储罐的平均精度比目前最高的平均精度高出了 19.3%。

不同方法在 DOTA 数据集上水平边界框任务的比较结果如表 4 所列。

表 3 不同方法在 DOTA 数据集上方向边界框任务的比较

Table 3 Comparison of various methods for oriented bounding box tasks on DOTA dataset

(单位:%)

Method	mAP	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC
FR-H	36.3	47.2	61.0	9.8	51.7	14.9	12.8	6.9	56.3	60.0	57.3	47.8	48.7	8.23	37.3	23.1
FR-O	52.9	79.1	69.1	17.2	63.5	34.2	37.2	36.2	89.2	69.6	59.0	49.4	52.5	46.7	44.8	46.3
YOLOv2	21.4	39.6	20.3	36.6	23.4	8.9	2.1	4.8	44.3	38.3	34.7	16.0	37.6	47.2	25.5	7.45
R-FCN	26.8	37.8	38.2	3.6	37.3	6.7	2.60	5.6	22.9	46.9	66.0	33.4	47.2	10.6	25.2	18.0
SSD	10.6	39.8	9.1	0.64	13.2	0.3	0.4	1.1	16.2	27.6	9.2	27.2	9.1	3.0	1.1	1.0
Ours	68.9	89.2	81.6	44.0	70.5	65.6	53.3	64.4	89.8	79.8	84.9	62.2	63.9	65.8	65.4	52.5

表 4 不同方法在 DOTA 数据集上水平边界框任务的比较

Table 4 Comparison of various methods for horizontal bounding box tasks on DOTA dataset

(单位:%)

Method	mAP	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC
FR-H	60.6	80.3	77.6	32.9	68.1	53.7	52.5	50.0	90.4	75.1	59.6	57	49.8	61.7	56.5	41.9
YOLOv2	39.2	76.9	33.9	22.7	34.9	38.7	32.0	52.4	61.6	48.5	33.9	29.3	36.8	36.4	38.3	11.6
R-FCN	47.2	79.3	44.3	36.6	53.5	39.4	34.2	47.3	45.7	47.7	65.8	37.9	44.2	47.2	50.6	34.9
SSD	10.9	44.7	11.2	6.2	6.9	2.0	10.2	11.3	15.6	12.6	17.9	14.7	4.55	4.55	0.53	1.01
Ours	74.5	90.0	82.5	53.7	68.7	71.9	74.1	77.9	90.6	85.5	84.9	63.4	63.0	76.1	75.5	60.0

本文的框架不仅在 DOTA 数据集上取得了最好的结果,在 RSOD 数据集上也取得了较好的结果。由于 RSOD 数据集只有水平边界框作为真实框,因此我们只比较水平边界框的结果,如表 5 所列。

表 5 不同方法在 RSOD 数据集上的比较

Table 5 Comparison of various methods on RSOD dataset
(单位:%)

Method	Oiltank	Aircraft	Overpass	Playground
AlexNet	93.56	87.95	80.18	80.47
GoolegNet	92.83	85.90	62.34	91.62
SENet	97.70	86.49	78.95	87.64
Ours	98.32	84.88	85.60	92.42

结束语 为了解决遥感图像中物体密集、角度多、尺度种类多、背景复杂等问题,本文提出了基于显著性特征和角度信息的遥感图像检测框架。该框架通过构造旋转区域建议网络来解决物体多角度的问题。同时,在框架中加入通道自学习和显著性学习模块,使得目标信息遭到加强,非目标信息遭到削弱。此外,在保持参数量的同时,在网络中加入多尺度特征模块,以获得更丰富的上下文信息。在训练过程中,添加注意力损失函数来引导网络学习显著性特征。在两个公开的数据集上的结果显示,本文方法的检测精度与其他方法相比有所提升,因此验证了本文方法的有效性。但是,有些类型的检测结果仍不太理想,如桥和直升机。因此,在下一步的工作中,我们计划提升候选框的质量,进一步避免冗余和漏检,从而提升检测效果。

参 考 文 献

- [1] BLASCHKE T. Object based image analysis for remote sensing [J]. *Remote Sens*, 2010, 65(1): 2-16.
- [2] BONTEMPS S, BOGAERT P, TITEUX N, et al. An object-based change detection method accounting for temporal dependencies in time series with medium to coarse spatial resolution [J]. *Remote Sensing of Environment*, 2008, 112(6): 3181-3191.
- [3] CHEN S Z, WANG H, XU F, et al. Target classification using the deep convolutional networks for SAR images [J]. *IEEE Transaction on Geoscience and Remote Sensing*, 2016, 54(8): 4806-4817.
- [4] ZHONG P, WANG R A. A multiple conditional random field's ensemble frame work for urban area detection in remote sensing optical images [J]. *IEEE Transaction on Geoscience and Remote Sensing*, 2007, 45(12): 3978-3988.
- [5] STANKOV K, HE D C. Detection of buildings in multispectral very high spatial resolution images using the percentage occupancy hit-or-miss transform [J]. *Remote Sens*, 2014, 7(10): 4069-4080.
- [6] WEN X, SHAO L, FANG W, et al. Efficient feature selection and classification for vehicle detection [J]. *IEEE Trans. Circuits Syst. Video Technol*, 2015, 25(3): 508-517.
- [7] LINE E, LARS A, HANS K. Classification-based vehicle detection in high resolution satellite images [J]. *ISPRS J. Photogrammetry. Remote Sens*, 2009, 64(5): 65-72.
- [8] LIU L, SHI Z. Airplane detection based on rotation invariant and sparse coding in remote sensing images [J]. *Optik-International Journal for Light and Electron Optics*, 2014, 125(18): 5327-5333.
- [9] XU Z, XU X, WANG L, et al. Deformable convnet with aspect ratio constrained NMS for object detection in remote sensing imagery [J]. *Remote Sensing*, 2017, 9(12): 1312.
- [10] GIRSHICK R, DONAHUE J, DARRELL T, et al. Region-based convolutional networks for accurate object detection and segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(1): 142-158.
- [11] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2017, 39(6): 1137-1149.
- [12] REN S, HE K, 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.
- [13] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified real-time object detection [C] // *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2016: 779-788.
- [14] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot Multi Box Detector [C] // *Proceeding of the 14th European Conference on Computer Vision*. New York, 2016: 21-37.
- [15] 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*. 2018.
- [16] HAN X, ZHONG Y, ZHANG L. An efficient and robust integrated geospatial object detection framework for high spatial resolution remote sensing imagery [J]. *Remote Sensing*, 2017, 9(7): 666.
- [17] GUO W, YANG W, ZHANG H J, et al. Geospatial object detection in high resolution satellite images based on multi-scale convolutional neural network [J]. *Remote Sensing*, 2018, 10(1): 131.
- [18] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 42(8): 99.
- [19] WANG J, YUAN Y, YU G. Face attention network: An effective face detector for the occluded faces [J]. *arXiv:1711.07246*, 2017.
- [20] TANG T, ZHOU S, DENG Z, et al. Arbitrary oriented vehicle detection in aerial imagery with single convolutional neural networks [J]. *Remote Sensing*, 2017, 9(11): 1170.
- [21] DAI J, LI Y, HE K, SUN J. R-FCN: object detection via region-based fully convolutional networks [C] // *Advances in Neural Information Processing Systems 29*. Cambridge: The MIT Press, 2016: 379-387.



YUAN Xing-xing, born in 1994, post-graduate. Her main research interests include image processing and artificial intelligence.



WU Qin, born in 1978, Ph.D, associate professor. Her main research interests include computer vision and pattern recognition.