



计算机工程与应用  
Computer Engineering and Applications  
ISSN 1002-8331, CN 11-2127/TP

## 《计算机工程与应用》网络首发论文

题目: 旋转目标检测算法在卫星影像中的应用  
作者: 李巍, 戴朝霞, 张向东, 张亮, 沈沛意  
网络首发日期: 2020-09-22  
引用格式: 李巍, 戴朝霞, 张向东, 张亮, 沈沛意. 旋转目标检测算法在卫星影像中的应用. 计算机工程与应用.  
<https://kns.cnki.net/kcms/detail/11.2127.TP.20200922.1531.016.html>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

# 旋转目标检测算法在卫星影像中的应用

李 巍<sup>1</sup>, 戴朝霞<sup>2</sup>, 张向东<sup>1</sup>, 张 亮<sup>3</sup>, 沈沛意<sup>3</sup>

1. 西安电子科技大学 通信工程学院, 西安 710126
2. 中国电子科技网络信息安全有限公司, 成都 610041
3. 西安电子科技大学 计算机科学与技术学院, 西安 710126

**摘 要:** 近年来, 深度学习在卫星影像目标检测领域中得到了快速的发展, 如何精准高效定位目标物体是卫星影像目标检测研究中的主要难点。提出了一种基于旋转矩形空间的 YOLOv3 改进算法去精准定位卫星影像目标, 对原有网络进行改进, 增加角度变换的数据预处理过程, 防止实例角度变化对网络训练造成影响。使用双旋转坐标进行回归训练, 增加了角度锚点, 提高了网络对卫星目标的检测有效性。提出了基于旋转矩形空间的非极大值抑制改进算法, 可以有效去除多余的旋转预测框。实验结果表明, 改进 YOLOv3 算法相较于原始 YOLOv3 算法拥有更好的可视化效果, 可以有效准确地定位卫星影像的目标物体, 有效避免了密集场景下预测框的遮挡问题, 在保证实时性的前提下, 将均值平均精度提高了 0.8%。

**关键词:** 卫星影像; 目标检测; 深度学习; 旋转矩形框

**文献标志码:** A    **中图分类号:** TP75    **doi:** 10.3778/j.issn.1002-8331.2005-0008

李巍, 戴朝霞, 张向东, 等. 旋转目标检测算法在卫星影像中的应用. 计算机工程与应用

LI Wei, DAI Zhaoxia, ZHANG Xiangdong, et al. Application of rotating target detection algorithm in satellite image. Computer Engineering and Applications

## Application of rotating target detection algorithm in satellite image

LI Wei<sup>1</sup>, DAI Zhaoxia<sup>2</sup>, ZHANG Xiangdong<sup>1</sup>, ZHANG Liang<sup>3</sup>, SHEN Peiyi<sup>3</sup>

1. School of Communication Engineering, Xidian University, Xi'an 710126, China
2. China Electronic Technology Cyber Security Co., Ltd., Chengdu 610041, China
3. School of Computer Science and Technology, Xidian University, Xi'an 710126, China

**Abstract:** Recently, deep learning has developed rapidly in the field of satellite image target detection, how to accurately and effectively locate the target object is the main challenge in the research of satellite image target detection. An improved YOLOv3 algorithm based on rotating rectangular space is proposed, which can accurately locate satellite image targets. This paper adds angle conversion as a preprocessing step, which can eliminate the distortion of instance angle during network training. Two kinds of rotating bounding box coordinates are proposed for regressing network and angle anchors are added to priori boxes, which can improve the effectiveness of network de-

**基金项目:** 国家重点研发计划 (No.2019YFB1311600); 宁波 2025 重点研发计划 (No.2018B10071)。

**作者简介:** 李巍(1994-),男,硕士研究生,研究领域为深度学习与计算机视觉;戴朝霞(1972-),女,硕士,高级工程师,研究领域为网络安全;张向东(1970-),男,博士,教授,硕导,研究领域为智能交通系统(ITS),图像处理和计算机视觉;张亮(1981-),男,通信作者,博士,教授,硕导,CCF 会员,研究领域为多核嵌入式系统,计算机视觉,深度学习,同时定位和地图绘制(SLAM),人机交互,图像处理等领域,E-mail:liangzhang@xidian.edu.cn;沈沛意(1969-),男,博士,教授,硕导,CCF 会员,研究领域为计算机视觉,DSP/FPGA 理论及应用,数字图像处理以及计算机网络。

tection. A novel rotated non-maximum suppression is proposed to handle rotated bounding boxes, which can effectively remove redundant rotating prediction box. The experimental results shows that the improved YOLOv3 algorithm has a significant improvement over the original effect, which can effectively and accurately locate the target object of the satellite image. Improved YOLOv3 has a better visualization effect and effectively avoids the occlusion problem of prediction frames in dense scenes and the mean Average Precision of detection is improved by 0.8% under the premise of ensuring real-time performance.

**Key words:** satellite imagery; target detection; deep learning; rotating bounding box

## 1 引言

卫星遥感目标检测在军事侦察、船舶管理、灾害监测、无人驾驶等领域有着十分重要的应用。随着人工智能技术的蓬勃发展,深度学习在目标检测领域不断突破现有瓶颈,相比于传统的方法,其在图像的目标检测、分类分割上更加具有性能优势。深度学习技术所具备的高效通用、灵活易学性,使其成为卫星数据处理领域一个热门研究方向。深度学习在目标检测领域可以分为“two-stage detection”和“one-stage detection”两种检测方法,两者的不同之处在于,前者是一个从粗到细的两阶段处理过程,后者则是一步完成<sup>[1]</sup>。两阶段检测模型主要包括 R-CNN<sup>[2]</sup>、FastR-CNN<sup>[3]</sup>、FasterR-CNN<sup>[4]</sup>、R-FCN<sup>[5]</sup>等,检测过程的第一阶段提供大量候选区域,第二阶段则进行具体判定并筛选分数最高的区域。虽然两阶段方法的检测准确率较高,但是不能应用于实时检测。单阶段检测模型主要包括 YOLOv1<sup>[6]</sup>、YOLOv2<sup>[7]</sup>、YOLOv3<sup>[8]</sup>、SSD<sup>[9]</sup>等系列方法。单阶段检测模型拥有更快的速度,可以直接完成端到端的预测,抛弃了两阶段粗检测和精检测的过程,但一定程度上降低了物体检测的精度。

在卫星影像目标检测领域中,检测算法需要维持检测速度和精度之间的平衡。Benjdira 等<sup>[10]</sup>对一阶段检测和两阶段检测模型在卫星影像数据集上进行了细致对比,在精度相当的情况下,一阶段检

测算法的灵敏度和处理时间均优于两阶段检测算法。Li 等<sup>[11]</sup>设计了一个端到端的网络用于检测像素级标注的卫星船舶数据,可以将不同级别的特征图进行融合并提取多尺度特征,虽然可以准确定位,但是标注成本过高且检测速度慢。Adam<sup>[12]</sup>在 YOLOv1 上进行改进,解决卫星影像的高图像分辨率问题和小目标物体问题,但无法解决多尺度问题。戴伟聪等<sup>[13]</sup>将密集相连模块加入到 YOLOv3 中去检测卫星遥感图像的目标物体,提高了目标的检测精度和召回率。魏伟等<sup>[14]</sup>对 YOLOv3 的网络结构和锚点进行改进,提高了 YOLOv3 在航拍目标检测的准确性,但是在物体密集的情况下,会存在预测框遮挡问题,视觉效果较差。

针对于如何精确定位,本文设计了基于旋转矩形空间的 YOLOv3 改进网络用于卫星影像的实时检测,使用旋转矩形框对带有角度的目标物体进行精确定位。主要有以下几点改进:针对于实例角度会随图像缩放而改变的现象,提出了角度转换算法,计算图像缩放后实例的角度;提出了两种旋转框坐标用于回归网络的设计,可以更好地定位带有角度的物体;添加角度锚点到先验框中,帮助网络稳定角度的预测;使用逐边裁剪算法和 Green-Riemann 定理计算旋转矩形的 IOU 值;提出了基于旋转矩形空间的非极大值抑制算法,可以有效去除多余的旋转预测框,使算法的检测精度得到较大提升。

## 2 模型总体设计

网络结构如图 1，主要包括预处理部分和三个

主体子模块：53 层卷积层构成基础骨干网络、特征金字塔网络、用于提取旋转矩形框的预测模块。

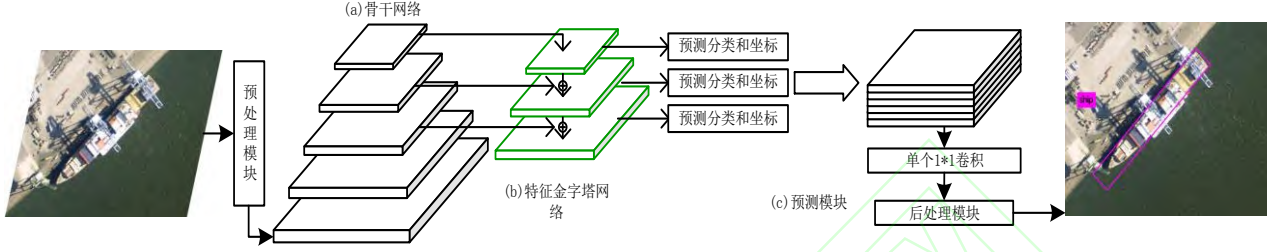


图 1 模型网络结构图

Fig.1 Model network structure diagram

### 2.1 预处理模块

预处理模块对输入的数据进行调整及增强，使数据能够适应当前网络，得到更为精良的模型。主要具有两个功能：（1）图像缩放：因为本文网络结构的输入大小固定，所以需要将图像缩放到合适的大小，对应于网络结构输入层的大小；（2）数据调整：图像中的物体经过图像缩放后，目标物体的角度信息相较于整个图像会发生变化，需要对角度进行计算变换。图 2 给出了目标物体角度变化的示意图，以及角度坐标变换公式要用到的参数。坐标  $(x_1, y_1)$  和  $(x_2, y_2)$  为图（a）目标物体上的两点，物体的角度  $\theta_0$  为两点直线与  $x$  轴正向所形成的夹角， $(w_0, h_0)$  为图（a）的宽高。将图（a）缩放为图（b），坐标  $(x_3, y_3)$  和  $(x_4, y_4)$  为图（b）目标物体上的两点，对应于图（a）的  $(x_1, y_1)$  和  $(x_2, y_2)$ ，物体的角度  $\theta_1$  为两点直线与  $x$  轴正向所形成的夹角， $(w_1, h_1)$  为图（b）的宽高。图（b）目标物体的角度与图（a）目标物体的角度之间的换算公式为式（1）。

$$\begin{aligned}
 x_3 &= x_1 \times \frac{w_1}{w_0}, y_3 = y_1 \times \frac{h_1}{h_0} \\
 x_4 &= x_2 \times \frac{w_1}{w_0}, y_4 = y_2 \times \frac{h_1}{h_0} \\
 \tan \theta_0 &= \frac{y_2 - y_1}{x_2 - x_1} \\
 \tan \theta_1 &= \frac{y_4 - y_3}{x_4 - x_3} = \frac{y_2 \times \frac{h_1}{h_0} - y_1 \times \frac{h_1}{h_0}}{x_2 \times \frac{w_1}{w_0} - x_1 \times \frac{w_1}{w_0}} \\
 &= \frac{(y_2 - y_1) \times \frac{h_1}{h_0}}{(x_2 - x_1) \times \frac{w_1}{w_0}} = \tan \theta_0 \times \frac{h_1 \times w_0}{h_0 \times w_1} \\
 \theta_1 &= \arctan \left( \tan \theta_0 \times \frac{h_1 \times w_0}{h_0 \times w_1} \right)
 \end{aligned} \quad (1)$$

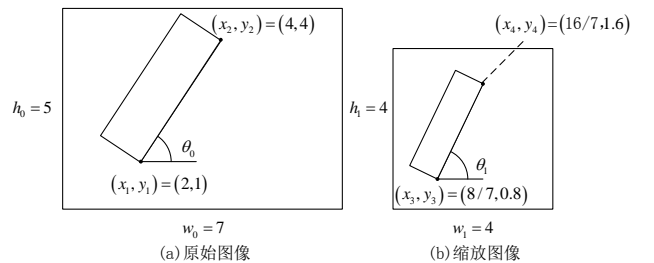


图 2 物体角度变化的示意图

Fig.2 Schematic diagram of object angle change

### 2.2 特征提取

目标检测任务首先需要对输入的图形进行基础特征提取工作，针对如何将图形分解成基础特征，



再将分解的基础特征逐渐抽象成更为高级的语义进行概念上表达这一问题，本文采用残差网络 DarkNet-53 和特征金字塔网络相结合的方式提取特征。残差网络 ResNet 在 2015 年的 ImageNet 大规模视觉识别竞赛中取得了图像分类和物体识别的优胜，跳跃式连接这一概念首次被应用在深度学习中<sup>[15]</sup>。残差结构主要解决的是随网络层数增加带来的网络退化问题。Facebook 在 2017 年 CVPR 论文中提出了特征金字塔网络思想，用于提取特征<sup>[16]</sup>。卫星影像上的目标物体比自然图像具有更大的尺寸比，通过特征金字塔结构可以在速度上和准确性上兼顾，并获得更加鲁棒的多尺度特征。

### 2.3 预测模块

本文预测模块将每层特征金字塔网络的输出作为输入，两个  $1 \times 1$  的卷积层分别对输入进行特征提取，得到两种不同的旋转框坐标预测以及它们各自的置信度和分类预测，如图 3 所示。本文使用双坐标参与网络训练可以改善网络性能，优化目标检测的结果，其原理是通过网络结构上的改变将两种坐标信息加入到网络回归中，优化最终的目标函数。其中坐标表示 1（图 4（a））只在网络训练的过程中用于回归，改善网络性能。坐标表示 2（图 4（b））用于网络训练时的回归，并在测试阶段输出最终的结果。

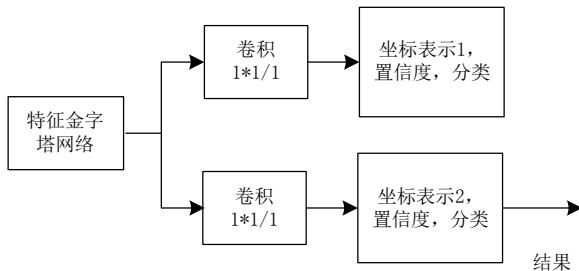


图 3 预测模块示意图

Fig.3 Schematic diagram of prediction module

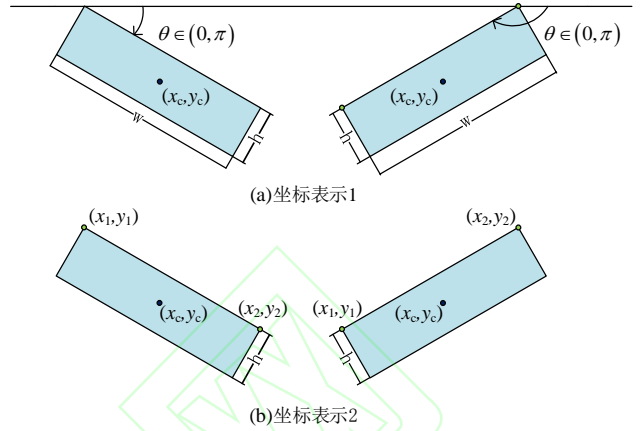


图 4 两种旋转坐标示意图

Fig.4 Schematic diagram of two rotating coordinates

### 2.4 角度锚点提出与设计

本文将角度锚点加入到了 RPN 网络中，不仅在每个格子中设置长宽比例不同的 anchor box，而且还在每个长宽比上设置不同的旋转角度，具有不同旋转角度和长宽比的 anchor box 可以尽可能地覆盖图像上不同角度、不同位置、不同尺度的物体。一些大物体可能会占据图像的中心，本文将最后一层特征图的大小设置成奇数，这样会有一个物体位置预测占据图像中心。因此在本文网络结构中，特征金字塔有三层，其特征层大小对应不同大小的 RPN 网络格子，分别为  $13 \times 13$ 、 $26 \times 26$ 、 $52 \times 52$ ，每层的每个格子有三种不同长宽比的 anchor box 以及每个长宽比有 3 种角度的预设，这样每个网格将产生 9 个 anchor box，这些 anchor box 将作为神经网络的候选区域并预测其中是否有物体，同时调整预测框的位置。图 5 为本文加入角度锚点后 RPN 的示意图。

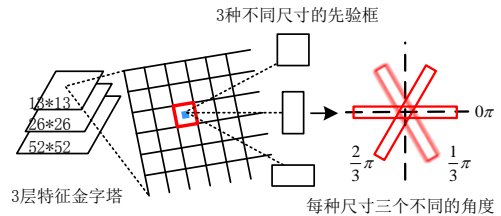


图 5 角度锚点示意图

Fig.5 Angle anchor point diagram

### 2.5 旋转矩形的交并比计算

目标之间的交并比（IOU）计算是现在主流目

标检测算法中的关键步骤，主要应用于后处理阶段，是目标检测算法中测量指标 mAP 计算的重要参数。本文设计了一个简单且能够处理旋转矩形框的 IOU 算法。将要计算的两个旋转矩形的坐标按顺时针排列，然后使用由 Sutherland I E 和 Hodgman G W 提出的逐边裁剪算法找到两个旋转矩形相交的点<sup>[17]</sup>。其基本思想是将两个多边形中的一个当作窗口，将多边形关于窗口的裁剪分解为多边形关于窗口各边所在直线的裁剪。裁剪算法中的裁剪线由窗口的每条边及其延长线构成，裁剪线将平面分为可见与不可见两部分。另一个多边形每条边有两个端点  $s$ 、 $p$ ，按顺时针顺序排列，经过裁剪线裁剪后有四种输出情况，如图 6 所示。(a) 中  $s$ 、 $p$  端点同位于外侧则没有输出；(b) 中  $s$  端点位于内侧， $p$  端点位于外侧，输出  $sp$  与边界交点  $I$ ；(c) 中  $s$ 、 $p$  端点同位于内侧，输出  $p$ ；(d) 中  $s$  端点位于外侧， $p$  端点位于内侧，输出  $sp$  与边界交点  $I$  以及  $p$ 。

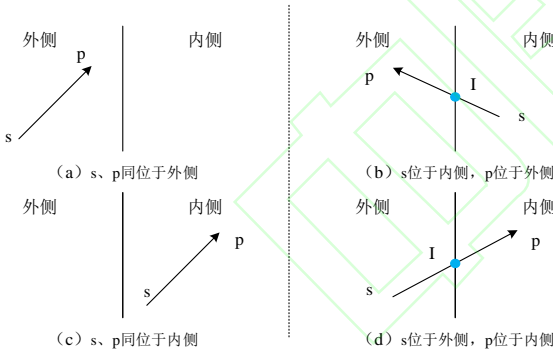


图 6 多边形与裁剪线的位置关系

Fig.6 Schematic diagram of the Sutherland-Hodgman cutting process

图 7 为两个旋转矩形使用逐边裁剪算法进行裁剪过程的示意图，按照从左到右、从上到下的顺序进行裁剪。通过逐边裁剪算法得到两个旋转矩形按顺时针旋转的交点，然后对交点使用 Green-Riemann 定理简单高效地计算两个旋转矩形的交叉面积<sup>[18]</sup>，如公式 (2) 所示：

$$A = \frac{1}{2} \times \sum_{i=0, j=(i+1)\%n}^n (x_i \times y_j - x_j \times y_i) \quad (2)$$

公式 (2) 中  $A$  为两个旋转矩形的交叉面积，

$(x_i, y_i)$  为两个旋转矩形的交点， $n$  为两个旋转矩形交点的数量。(注：两个旋转矩形的交点按照顺时针排序)。

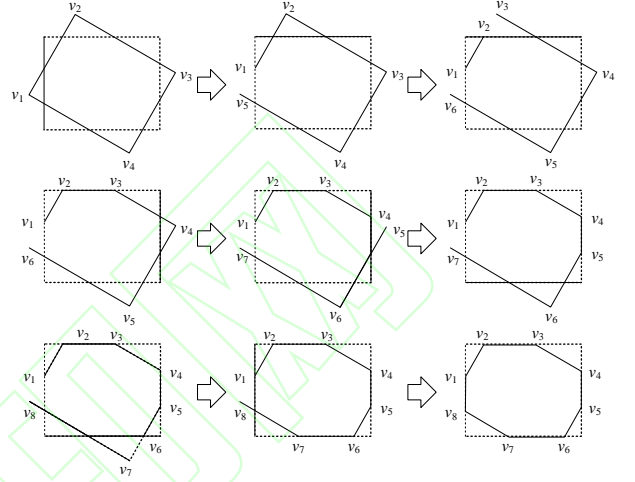


图 7 逐边裁剪算法裁剪过程的示意图

Fig.7 Schematic diagram of the Sutherland-Hodgman cutting process

## 2.6 基于旋转矩形的非极大值抑制

本文的后处理阶段主要为非极大值抑制处理，主要针对卫星影像这种带有角度的目标物体。假设有两个同心且形状相同的旋转矩形框，保持两个旋转矩形框的角度差不变，逐渐加大两者的长宽比，两者 IOU 的计算值会随着长宽比的变化急剧减小；保持两个旋转矩形框的长宽比不变，逐渐加大两者的角度差，两者 IOU 的计算值会随着角度差的变化急剧减小。图 8 展示了上述两种情况。

图 8 (a)、(b)、(c) 中两个旋转矩形框之间都具有相同的角度差 (角度为 45 度)，但是其长宽比不同，分别为 1:3、1:6、1:9，三个 IOU 的计算值相差巨大，分别为 0.308、0.134、0.085。图 8 (d)、(e)、(f) 中两个旋转矩形框长宽比相同 (长宽比为 1:6)，但是其角度差不同，分别为 15 度、30 度、45 度，三个 IOU 的计算值同样相差巨大，分别为 0.444、0.200、0.134。如果按照非极大值抑制算法中设定一个阈值  $I$ ，两个预测框之间的交并比计算超过这个阈值  $I$ ，则过滤掉预测分数小的预测框，那么遇到图 8 中类似两个预测框中心离得较近且角度差极大或者长宽比极高的情况就会变得难以处

理,因为两个预测框的 IOU 通常值很小。将阈值下调,一般需要设置一个非常小的值才可以将上述情况去除掉,但是阈值设置小数值时又会遇到新的问题,因为卫星影像检测中往往会遇到如图 9 所示的情况。在图 9 中,红色实线框为真实物体,黑色虚线框为神经网络预测框,浅蓝的颜色区域为两个黑色虚线框交叉的面积。在阈值较小的情况下图中一些有效的检测框会被抑制掉。

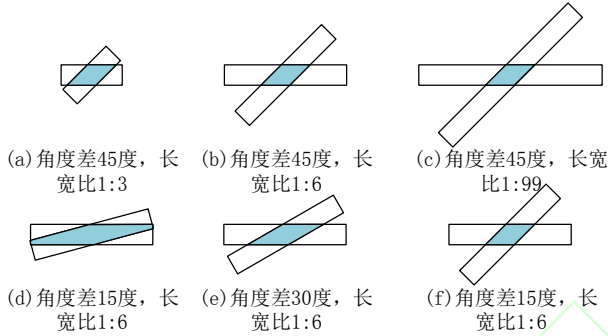


图 8 物体的角度和长宽比与 IOU 关系示意图

Fig.8 Schematic diagram of the relationship between the angle and aspect ratio of the object and the IOU

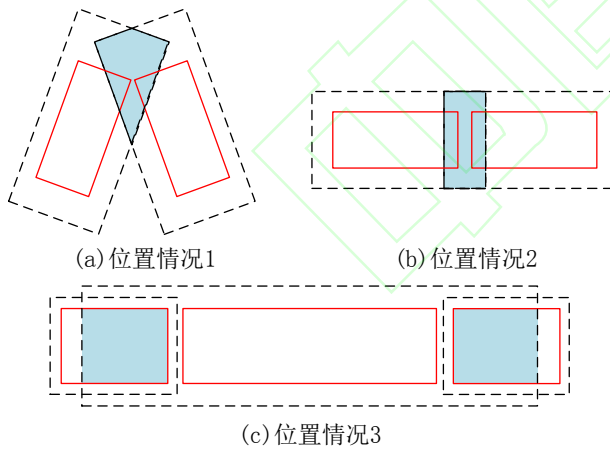


图 9 卫星影像检测目标预测之间的位置关系

Fig.9 Position relationship between satellite image detection target prediction

本文对传统的非极大值抑制算法进行改进,通过在不同的角度差和长宽比下设置不同阈值来抑制多余的预测框,保留有效的预测框。在两个旋转框同心的情况下,保持角度差和长宽比其中一个变量,调整另一个变量,参考两个旋转矩形的交并比值大小来设置不同的阈值,并根据检测的效果不断对每个阈值进行调整。

### 3 实验与分析

#### 3.1 实验数据集

本文使用三个公开的高分辨率可见光卫星影像数据集来评估模型的性能,分别是中国科学院自动化研究团队建立的 High Resolution Ship Collection 2016 (HRSC2016) 数据集<sup>[19]</sup>、武大遥感国重实验室夏桂松和华科电信学院白翔一起建立的 Dota 数据集<sup>[20]</sup>和中科院大学高清航拍目标数据集 UCAS-AOD。

HRSC2016 数据集主要由各类舰船图像组成,舰船所在场景有海面 and 近岸两种。数据集主要来源于 Google Earth。数据集的图像从 Mumansk、Everett、Newport Rhode Island、Mayport Naval Base、Norfolk Naval Base、San Diego Naval Base 六个著名港口收集。数据集图像一般分辨率为 0.4m~2-m。图像大小在 300×300 至 1500×900 之间,大部分图像分辨率超过了 1000×600。数据使用 xml 格式保存图像的分辨率层级、港口信息、地理坐标、数据源、标尺信息、图像采集日期等信息。HRSC2016 数据集总共有 1061 张图像,通过科学的方法将其分为包含 436 幅图像和 1207 个实例样本的训练集、181 幅图像和 541 个实例样本的验证集、444 幅图像和 1228 个实例样本的测试集。

HRSC2016 数据集主要数据类别为舰船,并将其分为 3 个等级的识别任务:(1) L1 级任务:舰船检测任务,种类只有船舶一种,任务是将目标物体从背景中分离。(2) L2 级任务:根据舰船用途划分的识别任务,将船舶分为 4 大类,包括:军舰、潜艇、航母以及商船。(3) L3 级任务:根据舰船型号划分的识别任务,在 L2 任务的基础上,将军舰和航母按照大小型号进行细致的分类,比如:军舰具体细分为阿利伯克驱逐舰 (Arleigh Burke)、佩里级护卫舰 (Perry)、奥斯汀级船坞登陆舰 (Austin) 等,航母具体细分为企业级航母 (Enterprise)、小鹰级航母 (Kitty Hawk) 以及尼米兹级航母 (Nimitz) 等。商船则按照用途划分为运输汽车船、气垫船以及游艇等。其位置标注信息包括矩形框:左上点  $(x_1, y_1)$ 、右下点  $(x_2, y_2)$ ;带有角度的矩形框:中心点  $(x_c, y_c)$ ,宽  $w$ ,高  $h$ ,角度  $\theta$  以及像素级别的分割掩膜。

Dota 数据集总共有 2806 张卫星图像,分为 16 个类别,包括飞机、集装箱起重机、船只、游泳池、储蓄罐、环形路线、棒球内场、网球场、英式足球场、直升飞机、大型车辆、桥、篮球场、小型车辆、田径场、海港,共计 188282 个实例。大部分图像分辨率约为 4000×4000,需要进行裁剪才能使用神



经网络训练。其数据主要来源为卫星 JL-1、卫星 GF-2、中国资源卫星数据和应用中心以及 Google Earth。Dota 数据集中每个实例的位置采用标注四个顶点坐标的方式，可以定位不同方向和形状的四边形，四个点按顺时针方向依次排序，标注形式为  $(x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4)$ 。同时每个实例还被分配类别和是否难以检测的标签。数据集使用 txt 格式进行保存。在数据划分上，随机将原始数据集划分为 1/2、1/6、1/3 作为训练集、验证集和测试集。Dota 数据集的实例在各种角度方向上表现得非常平衡，通过数据集去观察实例物体时可以进一步接近真实场景，这对神经网络模型的学习训练有着至关重要的作用。在最新的 Dota 数据集标注还包含 10 像素以下的小目标实例。

UCAS 数据集总共有 1510 张卫星图像，包含飞机和汽车 2 个类别，共计 14596 个实例。所有图像分辨率都为  $1280 \times 659$ 、 $1372 \times 941$ 、 $1280 \times 685$  这三种。数据集中每个实例的标注形式为  $(x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4, \theta, x, y, width, height)$ ，带角标的  $(x_i, y_i)$  为旋转矩形的四个顶点坐标， $\theta$  为目标的旋转角度， $(x, y)$  为水平边界框的左上点， $(width, height)$  为水平边界框宽高。数据主要来源为 Google Earth。数据集使用 txt 格式进行保存。因为原数据没有将数据集进行划分，本文将按照两个类别的实例数量 1/2、1/6、1/3 作为训练集、验证集和测试集。同样数据集实例的角度分布均衡。

### 3.2 实验结果及分析

#### (1) 检测结果定量分析

本文通过在 HRSC2016 数据集、Dota 数据集和 UCAS\_AOD 数据集使用均值平均精度 (mean Average Precision, mAP) 来评估本文算法。表 1 展示了模型在 HRSC2016 数据集三个等级任务下每个类别的召回率、精确度和平均精度等指标，gts 表示每个类别下测试数据中真实目标的数量，dets 表示每个类别下模型预测此类目标的数量。模型从 L1 任务到 L3 任务 mAP 不断下降。先从模型预测目标的数量分析，模型在三个任务下的预测总数量为 2948、3027、5942。由于 HRSC2016 数据集上 L1、L2、L3 是不同等级的分类任务，三者在训练集训练的实例总数并没有改变，只是单个种类实例的数量却随着种类的细分而下降，模型预测数量却急剧上升，保持了高召回率的同时却导致精确度快速下降，导致在 L3 任务上 mAP 大幅下降的结果。

Dota 数据集的测试结果如表 2 所示。结果相比于 HRSC2016 数据集测试的结果，其检测准确率下降，模型的 mAP 为 38.2%。模型在 Dota 数据集上性能整体下降主要有三点原因：1、Dota 数据集具有很多小目标物体，10-50 像素的目标占比 57%，还包含 10 像素以下的小目标。2、Dota 数据集包含多种实例密度不同的图像，特别是有很多小目标高度聚集的图像。3、Dota 训练集各个类别的实例数量极端不平衡。

表 1 HRSC2016 数据集测试实验的明细结果

Table 1 Detailed results of the HRSC2016 data set test experiment

HRSC2016 数据集						
	类别	gts	dets	recall	precision	ap
L1	ship	1188	2948	91.2%	36.7%	85.8%
					map	85.8%
L2	ship	290	1004	69.3%	20.0%	43.2%
	air	110	221	98.2%	48.9%	90.5%
	war	562	1305	95.4%	41.1%	90.4%
	mer	226	497	78.3%	35.6%	65.2%
			3027		map	72.3%
L3	ship	290	1221	69.7%	16.5%	42.8%
	air.	13	76	92.3%	15.8%	72.8%
	war.	26	518	73.1%	3.7%	9.2%
	mer.	57	292	40.4%	7.9%	30.1%
	Nim.	35	105	94.3%	31.4%	78.0%
	Ent.	12	96	91.7%	11.5%	72.2%
	Arl.	158	593	94.9%	25.3%	80.4%
	Whi.	40	287	87.5%	12.2%	43.7%
	Per.	128	619	90.6%	18.7%	64.2%
	San.	22	219	86.4%	8.7%	45.6%
	Tic.	3	453	94.2%	21.4%	78.4%
	Aus.	35	195	91.4%	16.4%	60.9%
	Tar.	50	174	94.0%	27.0%	86.8%
	Com.A	40	177	87.5%	19.8%	46.1%
	Con.	18	164	77.8%	8.5%	21.1%
	Car.A	16	53	87.5%	26.4%	73.7%
	Con.A	119	385	88.2%	27.3%	63.8%
	Med.	10	191	90.0%	4.7%	56.8%
	Car.B	16	124	87.5%	11.3%	54.0%
					map	56.9%



为了验证单阶段网络的准确率会受到类别不平衡的影响, 本文还使用类别实例数量较为平衡的 UCAS\_AOD 数据集进行对比验证, 表 3 展示了其详细的实验结果, 模型的 mAP 为 92.1%。结合表 4 三个训练集每个类别实例数量的统计, 发现 UCAS\_AOD 训练集两个类别的实例数量分布平衡, 模型检测效果最好, HRSC2016 训练集 L1 任务到 L3 任务, 每个类别实例数量差异变大, 检测效果从高到底, Dota 训练集中实例数量差异最大, 最高类别实例数量与最低类别实例数量之比约 1000:1, 模型检测效果最差。

## (2) 可视化检测结果

图 10 和图 11 展示了算法改进前后使用非旋转矩形候选框和旋转矩形候选框的检测效果对比图, 图 10 为 HRSC2016 数据集的三个识别任务的检测结果, 图 11 为 Dota 数据集和 UCAS\_AOD 数据集的检测效果。图 10 中从左到右三列分别为 L1 舰船检测任务、L2 舰船种类识别任务、L3 舰船型号识别任务。

表 2 Dota 数据集测试实验的明细结果

Table 2 Detailed results of the Dota data set test experiment

Dota 数据集					
类别	gts	dets	recall	precision	ap
Plane	2543	38533	96.4%	6.2%	75.7%
BD	91	3153	74.1%	2.0%	10.3%
Bridge	86	7057	39.3%	0.5%	5.5%
GTF	39	3361	42.1%	0.5%	2.0%
SV	15584	392354	93.4%	3.7%	48.8%
LV	7522	144892	83.0%	4.3%	51.1%
Ship	15772	148435	88.8%	9.5%	72.6%
TC	1028	19560	93.1%	4.8%	71.8%
BC	119	7094	75.4%	1.3%	37.8%
ST	2412	29731	93.5%	7.6%	70.7%
SBF	26	4802	69.6%	0.3%	17.1%
RA	36	2417	68.6%	1.0%	15.5%
Harbor	2147	81833	80.8%	2.1%	45.4%
SP	561	10054	84.0%	4.7%	59.7%
HC	69	3588	52.3%	0.9%	26.9%
CC	14	863	21.4%	0.3%	0.1%
map					38.2%

表 3 UCAS\_AOD 数据集测试实验的明细结果

Table 3 Detailed results of the UCAS\_AOD data set test experiment

UCAS_AOD 数据集					
--------------	--	--	--	--	--

类别	gts	dets	recall	precision	ap
Plane	2417	3903	98.3%	60.9%	97.2%
Car	2014	3579	91.5%	51.5%	86.9%
map					92.1%

图 10 (a)、(b) 和 (c) 分为两行, 第一行为使用非旋转矩形候选框的检测结果, 第二行为使用旋转矩形候选框的检测结果。图 11 中左列为使用非旋转矩形候选框的检测结果, 右列为使用旋转矩形候选框的检测结果。在目标背景简单以及目标为单个情况下, 旋转矩形框和非旋转矩形框在可视化效果方面较为相同, 如图 10 (a) 和 (b) 所示。当物体相邻较近或者相互挨着的情况下, 如图 10 (c) 和图 11 所示, 旋转矩形框标定的图像更为清晰, 可以帮助工作人员快速定位目标并分析处理数据。非旋转矩形框标定的图像会存在几点问题: 大物体预测框囊括相邻小物体预测框的问题、相邻物体间预测框相互覆盖和重叠的问题以及一个预测框覆盖多个实例的问题, 这对快速发现目标物体的位置信息影响较大, 特别是当较小物体排布特别紧密时, 大量重叠的预测矩形框会遮挡小目标物体。综上所述, 在可视化效果方面, 基于旋转矩形框检测识别的效果更好, 因为旋转矩形框更贴合卫星图像目标, 而非旋转矩形框则在预测旋转目标的同时还会覆盖多余的区域, 如果遇到目标物体密集度非常高且目标物体处于多角度的场景时, 非旋转矩形框标定的图像显得特别杂乱。

表 4 训练集每个类别的实例数量统计表

Table 4 Statistics table of the number of instances of each category in the training set

HRSC2016 训练集 L1 舰船检测任务							
ship	1788						
HRSC2016 训练集 L2 舰船类别识别任务							
war	841	ship	438	mer	314	air	195
HRSC2016 训练集 L3 舰船型号识别任务							
ship	438	Tar.	99	Com.A	50	air.	22
Arl.	272	mer.	70	war.	49	Ent.	21
Tic.	267	Whi.	68	Con.	34	Car.A	15
Con.A	170	Nim.	53	San.	30	Med.	15
Per.	137	Aus.	53	Car.B	25		
Dota 训练集							
SV	209659	harbor	11437	HC	1014	BD	427

ship	60645	ST	7974	BC	897	GTF	363
LV	43383	TC	5031	bridge	665	CC	245
plane	11819	SP	3083	SBF	586	RA	219
UCAS_AOD 训练集							
plane	5065	car	5100				

### (3) 与其他算法的比较

使用 Faster R-CNN<sup>[4]</sup>、Cascade R-CNN<sup>[21]</sup>、SSD<sup>[9]</sup>、YOLOv3<sup>[8]</sup>以及本文改进方法进行对比,结果如表 5 所示。

由表 5、图 10 和图 11 可见,改进的 YOLOv3 算法与原 YOLOv3 相比,时间上由于引入旋转检测框导致平均检测时间稍长 0.026s,但是在三个数据集的 5 个任务上平均 mAP 提高了 0.8%,可视化效果更好。两阶段方法 Faster R-CNN 和 Cascade R-CNN 检测准确率更高,但是本文改进的方法检测速度是两者检测速度的 10 倍、25 倍。而 SSD 算法无论在检测精度还是检测速度上都较差。因此,本文改进的 YOLOv3 算法在实时性及检测精度上达到了较好的平衡,可视化效果更佳。

表 5 不同算法的检测结果对比

Table 5 Comparison of detection results of different algorithms

	HRSC2016mAP/%			Dota mAP/%	UCAS mAP/%	平均检测时间 /s
	c1	c4	c19			
Faster R-CNN	87.6	72.7	46.9	<b>58.1</b>	89.2	0.173
Cascade R-CNN	<b>88.5</b>	<b>78.6</b>	53.9	43.3	89.5	0.4380
SSD	87.4	71.1	48.6	35.2	88.3	0.0308
YOLOv3	87.1	70.3	50.9	43.1	89.9	<b>0.0146</b>
本文改进方法	85.8	72.3	<b>56.9</b>	38.2	<b>92.1</b>	0.0172

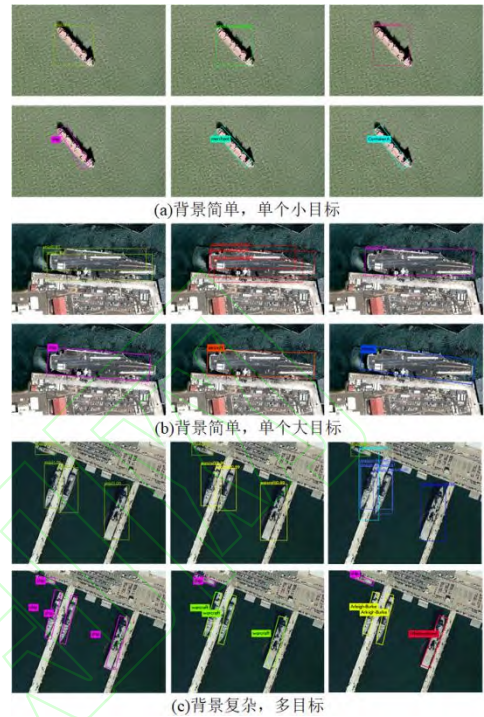


图 10 HRSC2016 数据集检测结果对比图

Fig.10 Comparison chart of HRSC2016 data set test results



图 11 Dota 和 UCAS\_AOD 数据集检测结果对比图

Fig.11 Comparison chart of Dota and UCAS\_AOD data set test results

## 4 结束语

本文设计了基于旋转矩形空间的单阶段卫星影像目标检测模型,使用多种方法对其进行改进,是本文的主要工作内容。使用残差网络结构和特征金字塔网络结构相结合的方式提取特征。针对实例角度会随着图像缩放而改变的现象,对角度标签在输入网络时进行了预处理工作,使角度符合缩放后实例的实际情况。提出了两种旋转框坐标用于网络

回归的设计, 添加角度锚点到先验框中, 优化了目标函数。使用逐边裁剪算法和 Green-Riemann 定理计算旋转矩形的 IOU 值, 并针对于卫星影像提出了基于旋转矩形空间的非极大值抑制算法。

但本文工作还存在不足, 单阶段检测模型如果遇到种类实例数量不平衡的数据集, 检测准确率会降低, 接下来还需要克服这方面的影响。本文基于旋转矩形的非极大值抑制算法, 虽然效果比传统的非极大值抑制算法可视化效果好, 但是需要人工按照不同的情况去精细设置多个不同角度和不同长宽比下的 NMS 阈值, 朝着自适应化的方向改进算法同样是值得研究的重点。

## 参考文献:

- [1] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features[C]//Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001. IEEE, 2001, 1: 1-1.
- [2] 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.
- [3] Girshick R. Fast R-CNN[C]. international conference on computer vision, 2015: 1440-1448.
- [4] Ren, Shaoqing, He, Kaiming, Girshick, Ross, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 39(6):1137-1149.
- [5] Dai J, Li Y, He K, et al. R-fcn: Object detection via region-based fully convolutional networks[C]//Advances in neural information processing systems. 2016: 379-387.
- [6] 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.
- [7] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.
- [8] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.
- [9] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//European conference on computer vision. Springer, Cham, 2016: 21-37.
- [10] Benjdira B, Khursheed T, Koubaa A, et al. Car detection using unmanned aerial vehicles: Comparison between Faster R-CNN and YOLOv3[C]//Proceedings of 2019 1st International Conference on Unmanned Vehicle Systems-Oman (UVS), 2019: 1-6.
- [11] Li M, Guo W, Zhang Z, et al. Rotated region based fully convolutional network for ship detection[C]//IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2018: 673-676.
- [12] Van Etten A. You only look twice: Rapid multi-scale object detection in satellite imagery[J]. arXiv preprint arXiv:1805.09512, 2018.
- [13] 戴伟聪, 金龙旭, 李国宁, 等. 遥感图像中飞机的改进 YOLOv3 实时检测算法[J]. 光电工程, 2018, 45 (12): 180350.
- [14] Dai W C, Jin L X, Li G N, et al. Real-time airplane detection algorithm in remote-sensing images based on improved YOLOv3[J]. Opto-Electronic Engineering, 2018, 45(12): 180350.
- [14] 魏玮, 蒲玮, 刘依. 改进 YOLOv3 在航拍目标检测中的应用[J]. 计算机工程与应用, 2020, 56(7): 17-23.
- [14] WEI Wei, PU Wei, LIU Yi. Application of Improved YOLOv3 in Aerial Target Detection. CEA, 2020, 56(7): 17-23.
- [15] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]. IEEE Conference on Computer Vision & Pattern Recognition(CVPR). 2016.
- [16] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.
- [17] Sutherland I E, Hodgman G W. Reentrant polygon clipping[J]. Communications of the ACM, 1974, 17(1):32-42.
- [18] Green G. An essay on the application of mathematical analysis to the theories of electricity and magnetism[M]. Wezäta-MelinsAktiebolag, 1828.
- [19] Zikun Liu, Hongzhen Wang, Lubin Weng, et al. Ship Rotated Bounding Box Space for Ship Extraction From High-Resolution Optical Satellite Images With Complex Backgrounds[J]. IEEE Geoscience & Remote Sensing Letters, 2016, 13(8):1-5.
- [20] 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.
- [21] Cai Z, Vasconcelos N. Cascade r-cnn: Delving into high quality object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 6154-6162.