〈图像处理与仿真〉

YOLOv5-LR：一种遥感影像旋转目标检测模型

高明明[1]，李沅洲[1]，马 雷[2]，南敬昌[1]，周芊邑[3]
（1. 辽宁工程技术大学 电子与信息工程学院，辽宁 葫芦岛 125105；2. 中国科学院自动化研究所，北京 100190；
3. 辽宁工程技术大学 电气与控制工程学院，辽宁 葫芦岛 125105）

摘要：真实遥感图像中，目标呈现任意方向分布的特点，原始 YOLOv5 网络存在难以准确表达目标的位置和范围、以及检测速度一般的问题。针对上述问题，提出一种遥感影像旋转目标检测模型 YOLOv5-Left-Rotation，首先利用 Transformer 自注意力机制，让模型更加注意感兴趣的目标，并且在图像预处理过程中采用 Mosaic 数据增强，对后处理过程使用改进后的非极大值抑制算法 Non-Maximum Suppression。其次，引入角度损失函数，增加网络的输出维度，得到旋转矩形的预测框。最后，在网络模型的浅层阶段，增加滑动窗口分支，来提高大尺寸遥感稀疏目标的检测效率。实验数据集为自制飞机数据集 CASIA-plane78 和公开的舰船数据集 HRSC2016，结果表明，改进旋转目标检测算法相比于原始 YOLOv5 网络的平均精度提升了 3.175%，在吉林一号某星推扫出的大尺寸多光谱影像中推理速度提升了 13.6%，能够尽可能地减少冗余背景信息，更加准确检测出光学遥感图像中排列密集、分布无规律的感兴趣目标的区域。

关键词：遥感图像；滑动窗口；注意力机制；旋转目标检测；YOLOv5
中图分类号：TP391 **文献标识码**：A **文章编号**：1001-8891(2024)01-0043-09

YOLOv5-LR: A Rotating Object Detection Model for Remote Sensing Images

GAO Mingming[1]，LI Yuanzhou[1]，MA Lei[2]，NAN Jingchang[2]，ZHOU Qianyi[3]
(1. *College of Electronic and Information Engineering*, *Liaoning Technical University*, *Huludao* 125105, *China*;
2. *Institute of Automation*, *Chinese Academy of Sciences*, *Beijing* 100190, *China*;
3. *College of Electrical and Control Engineering*, *Liaoning Technical University*, *Huludao* 125105, *China*)

Abstract: In a real remote sensing image, the target is distributed in any direction and it is difficult for the original YOLOv5 network to accurately express the location and range of the target and the detection speed is moderate. To solve these problems, a remote sensing image rotating target detection model, YOLOv5-Left-Rotation, was proposed. First, the transformer self-attention mechanism was used to make the model pay more attention to the targets of interest. In addition, Mosaic data were enhanced in the image preprocessing, and the improved Non-Maximum Suppression algorithm was used in post-processing. Second, an angle loss function was introduced to increase the output dimensions of the network, and the prediction box of the rotating rectangle was obtained. Finally, in the shallow stage of the network model, a sliding window branch was added to improve the detection efficiency of large-sized remote sensing sparse targets. The experimental datasets were the self-made aircraft dataset CASIA-plane78 and the public ship dataset HRSC2016. The results show that the average accuracy of the improved rotating target detection algorithm is improved by 3.175% compared with that of the original model, and the reasoning speed is improved by 13.6% in a large multispectral image swept by a Jilin-1 satellite. It can optimally reduce the redundant background information and more accurately

收稿日期：2022-11-18；**修订日期**：2022-12-30.
作者简介：高明明（1980-），女，博士，副教授，博士生导师，中国电子学会会员，主要研究方向为：人工智能、智能感知、微波毫米波技术等。
通信作者：李沅洲（1999-），男，硕士，硕士研究生，主要研究方向为遥感图像处理和目标识别。E-mail：yuanzhou.li@foxmail.com
基金项目：国家自然科学基金青年科学基金（61701211）；辽宁省应用基础研究计划项目（2022JH2/101300275）；辽宁省应用基础研究计划项目（22-1083）；北京市科技计划项目（Z201100005820010）。

detect the densely arranged and irregularly distributed areas of objects of interest in optical remote sensing images.

Key words: remote sensing images, sliding window, attention mechanism, rotating object detection, YOLOv5

0 引言

目标检测任务一直是计算机视觉的核心问题，它的目的是对图像中的目标进行定位。就遥感而言，目标检测是一项与各种应用密切相关的关键任务，例如，农作物收获分析、资源地理制图、军事航海、情报侦察、灾害管理、交通规划等[1-2]。由于其应用的广泛性，遥感图像有着各种感兴趣的对象，这些对象中自身存在任意方向旋转、多尺度、小目标、密集分布等情形，因此，研究目标检测算法如何适应当前大量多样化的遥感数据，具有重大的研究意义和应用价值[3-4]。

近年来，深度学习的进展为目标检测和定位提供了突破。在文献[5]中，卷积神经网络（CNNs）已经取代了传统的目标检测算法，增加网络深度使得它可以检测具有复杂背景的遥感图像中的小物体[6]。基于 CNN 的目标检测算法主要分为两类，第一类是两阶段目标检测算法，如 SPP-Net[7]、Faster R-CNN[8-9]等。这种类型的方法非常准确，但是检测速度非常慢。第二类算法直接将图像输入到网络中来预测物体的类别和位置信息，省去了区域检测过程，转化为回归问题，从而大幅提高了算法的检测速度，这种类型常见算法包括 SSD[10]、YOLO[11-14]系列等[15]。然而，上述适用于自然场景图像进行多分类的目标检测算法，当遇到当前大量多样化的遥感数据时，检测性能难以满足实际需求。

面对自然场景图像，原始的 YOLOv5 网络在众多优秀的检测方法中表现突出，目标检测边框为水平矩形框，但由于在真实的遥感图像中，目标呈现任意方向分布的特点，采用与坐标轴平行的一般矩形框，难以准确表达检测目标的位置和范围[16]。遥感图像具有超大的图像尺寸[17]，整幅遥感影像尺寸往往可达 20000×20000 以上，对于现阶段的计算硬件资源，针对常规图像的深度学习目标检测算法无法直接处理大尺寸的遥感图像，因此普遍采用带有重叠区域的滑动窗口切割法，将图片切割成尺寸合适的子图[18]，这样无疑加大了网络的计算量，导致检测速度明显下降。

针对上述问题，本文在 YOLOv5 的基础上进行模型改进，首先采用旋转矩形框的表示方式，应用 Transformer 自注意力模块[19]来增强网络对目标与图像背景的区分能力[20]，引入角度损失函数，增加旋转参数，使结果更紧凑地框选出目标。再对数据进行预处理通过 Mosaic 方法进行数据增强，随机缩放增加小目标，让网络的鲁棒性更好。采用 K-Means 方法进行聚类得到预设锚框的长度和宽度。其次，本文在网络浅层阶段，增加滑动窗口，来提高大尺寸遥感稀疏目标的检测效率，在保证检测精度的前提下，进一步来提高检测的速度。最后，提出旋转检测网络 YOLOv5-LR，以飞机数据集 CASIA-plane78 和舰船数据集 HRSC2016[21]作为实验数据，测试网络性能[22]。

1 一种遥感影像旋转目标检测模型

1.1 改进 YOLOv5 的网络结构

旋转检测网络 YOLOv5-LR 结构如图 1 所示。首先，对 Backbone 网络 CSPDarkNet53 添加 Transformer 自注意力模块，只增加了少量参数，就可以增强网络对目标与图像背景的区分能力。其次，由于卷积神经网络具有强大的特征表示能力和回归能力，在进行多层的卷积操作之后，可以进行更深层次的特征提取，并且卷积神经网络具有旋转不变性，所以本文认为经过多层网络后能够提取到目标的角度信息。因此在 Detect 层，引入一个旋转角度参数，通过构造新的角度损失函数对模型进行约束，使得 YOLOv5-LR 能够完成较为精准的旋转目标检测。最后，本文在 YOLOv5 网络特征提取与特征融合阶段之间，增加 CBS 卷积模块、全局平均汇聚层以及单层 Conv 卷积，构成判断滑动窗口分支，可以有效进行提前判读、加快检测速度。

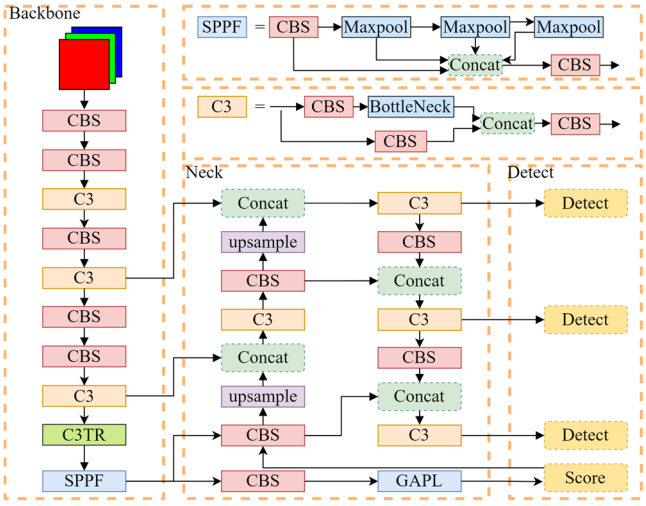

图 1 YOLOv5-LR 的网络结构

Fig.1 Network structure of YOLOV5-LR

1.2 Transformer 自注意力机制

当在人工处理一张光学遥感图像时，通常会自然地关注到你感兴趣目标（舰船、飞机等），并且会忽略掉一些背景信息（完全是海面、市区房屋建筑的部分等），例如，当人们想要在一个白色书架里寻找一些书本，目光主要会停留在书本上面，而后面的白色书架是带有什么花纹信息则往往不会被注意到。因此，人类利用视觉系统观察世界时，对每一个位置的注意力分布是不相同的。由于人类的注意力是有限的，为了节省这种资源，人们在面对自己感兴趣的目标时，会投入更多的注意力，而对自己想要寻找的目标以外的事物会投入较少的注意力。研究者将人类这种现象模拟到机器学习中，“注意力机制”因此诞生。

Transformer 模型最初是应用在文本数据上的，现在已经推广到各种深度学习的任务中。它完全基于注意力机制，没有任何卷积层或循环神经网络层，与先前循环神经网络实现输入表示的自注意力机制（self-attention）相比，更加轻量化。本文将第 9 层 C3 模块中的 Bottleneck 替换为 Transformer 块，相比于传统的 Transformer 结构，仅仅使用了 Encoding 部分。这样做的好处是告知模型哪里更需要关注，并赋予其较大权重，而不存在目标的背景区域就赋予一个更小的权重，让模型像人类视觉一样注意到感兴趣的目标，同时在接下来引入判断滑动窗口分支时，也会起到一个结合全局信息的作用。

1.3 旋转矩形框

原始的 YOLOv5 面对自然场景下的目标表现良好，检测结果呈水平矩形框。但由于光学遥感图像是从空对地进行拍摄，目标的形状、方向有任意分布的特点，使用水平矩形框难以精准地对目标的位置信息进行描述。例如，以 *xy*-平面作为图像边框的坐标轴，则水平矩形框与坐标轴平行，若目标边缘与坐标轴之间的夹角较大，则水平矩形框的冗余信息也较大，此时如果采用旋转矩形框，冗余信息则明显减少，如图 2 所示。因此，本文引入角度损失函数，增加一个旋转角度参数，使得检测结果更加紧凑地框选出目标。

水平矩形框的参数包括检测框的中心点的横、纵坐标，以及高度、宽度 4 个参数，通常表示为（x, y, w, h）。旋转检测方法即使用带有角度的旋转包围盒，框选感兴趣对象进行目标检测，相比一般水平矩形框，新增一个方向角度参数 θ，使用五参数表示法（x, y, w, h, θ）。但由于旋转角度定义的方式非唯一，五参数表示法主要分为 OpenCV 定义法和长边定义法两类，如图 3 所示。

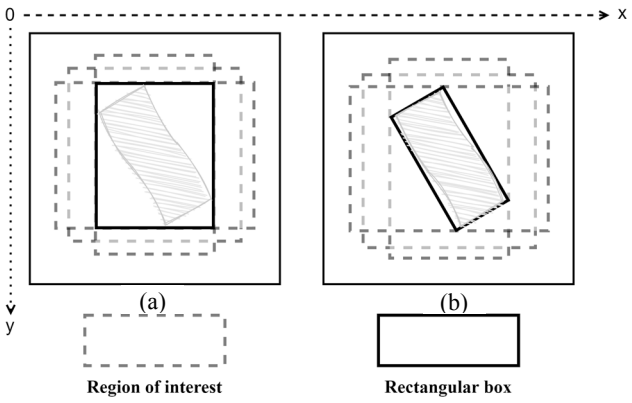

图 2 水平矩形框(a)与旋转矩形框(b)示意图

Fig.2 Schematic diagram of horizontal rectangular box (a) and rotating rectangular box (b)

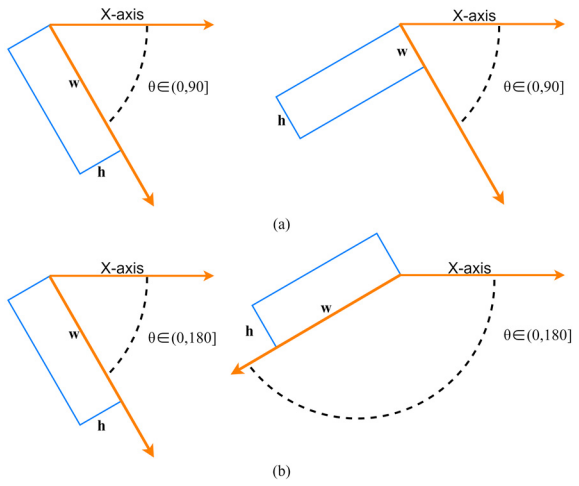

图 3 旋转框定义法。(a) OpenCV 定义法；(b)长边定义法

Fig.3 Definition method of rotation box. (a) OpenCV definition; (b) Long edge definition

1.4 滑动窗口分支

遥感图像具有超大的图像尺寸，而检测器所感兴趣区域的分布相对于整幅图像极其稀疏，当前主流目标检测器想要对整幅遥感图像进行检测，普遍采用对图像进行切片操作或使用滑动窗口扫描整幅图像。为了避免绝大部分窗口仅包含简单背景，却浪费大量计算资源，导致检测速度降低，本文尝试在网络模型的浅层阶段，增加滑动窗口分支。例如，窗口内全部是海面、云雾，本文认为这类滑动窗口是易于和含有目标的图像区分的，即易于被分类网络所区分的。因此通过滑动窗口分支，图像可以在进入模型深层阶段之前被判断是否含有目标或是否属于简单背景，如图 4 所示，如果当存在感兴趣区域的概率大于给定阈值，则图像顺利进入模型深层阶段，当其概率小于给定阈值，则跳出模型网络。

本文引入的滑动窗口分支主要由 3 个子模块组成，其结构如图 5 所示，首先加入 YOLOv5 网络中特

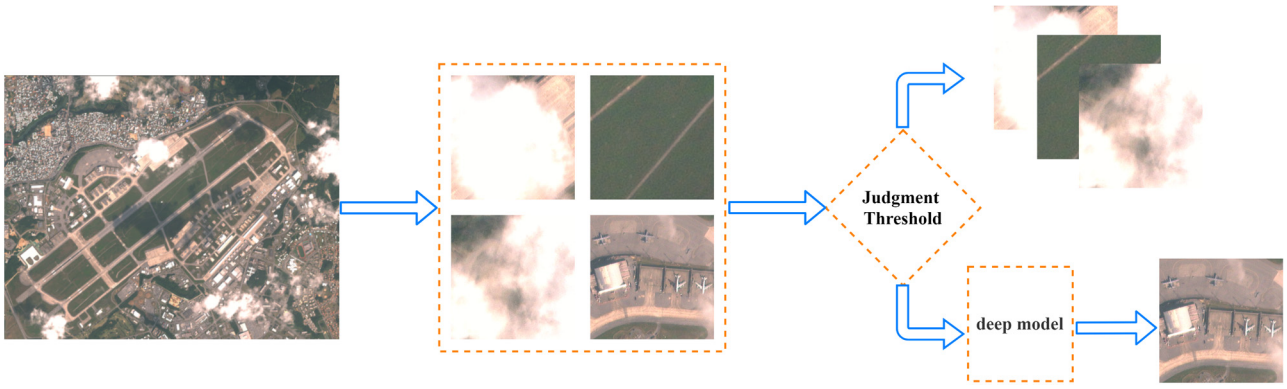

图 4 模型滑动窗口分支流程

Fig.4 Model sliding window branching process

征提取的基础组件，由卷积层（convolution）、批量规范化层（Batch normalization）以及激活函数（Sigmoid weighted Liner Unit），其主要作用为对模型 Backbone 提取的特征进行通道降维。其次加入全局平均池化层（Global Average Pooling Layer），对上一个模块提取的特征信息进行空间降维。最后，添加一个普通卷积层，它相当于一个全连接层，对池化操作后的特征信息进行提取，Sigmoid 整理后，得到最终的输出结果，即代表滑动窗口内存在目标的概率大小。

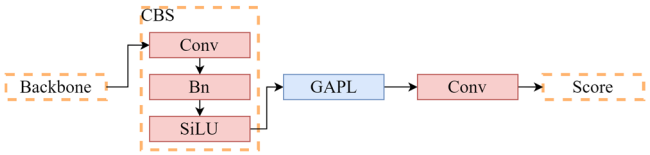

图 5 滑动窗口分支结构

Fig.5 Sliding window branch structure

1.5 损失函数

在深度学习中，需要可视化一个模型的优劣程度，因此通常会定义出一个函数，并希望通过各种优化策略把它下降到最低点，即损失函数。

YOLOv5 的损失主要由 3 部分组成：首先是分类损失，采用的是二值交叉熵损失（Binary Cross Entropy Loss）。分类损失函数公式如式(1)所示：

$$L_{\text{cla}}=-\sum_{i}^{N}\sum_{j}^{C}I_i^{\text{obj}}[O_{ij}\ln\left(\hat{C}_{ij}\right)+\left(1-O_{ij}\right)\ln\left(1-\hat{C}_{ij}\right)] \quad (1)$$

式中：N 为正负样本个数；C 为数据集真实目标的类别总数；当第 i 个矩形框存在感兴趣目标时，I_i^{obj} 等于 1，若不存在则等于 0。O_{ij} 表示预测目标的第 i 个矩形框的真实值是否为第 j 个类别，如果是第 j 个类别，则取 1，否则取 0。$\hat{C}_{ij}$ 表示为模型最后的输出通过 Sigmoid 函数得到目标为 j 的概率。

其次是置信度损失，仍使用二值交叉熵损失，公式如式(2)所示：

$$L_{\text{conf}}=-\sum_{i}^{N}[O_i\ln\left(\hat{C}_i\right)+\left(1-O_i\right)\ln\left(1-\hat{C}_i\right)] \quad (2)$$

式中：$O_i\in[0,1]$，表示真实置信度 IoU（Intersection over Union），它其实就是用来衡量预测目标矩形框和真实目标矩形框的重合程度，用这两个区域的重叠部分除以两个区域的并集部分即可得到 IoU 的计算值，通过设定的阈值，与这个计算值进行比较，若 IoU 大于这个阈值，则将预测出来的矩形框认为较为准确，反之，认为预测不准确。$\hat{C}_i$ 为网络最终得到的预测置信度。

最后是定位损失，采用 CIoU 损失函数。CIoU 公式本文不再赘述，定位损失函数如式(3)所示。

$$L_{\text{loc}}=\sum_{i}^{N}I_i^{\text{obj}}\left(1-\text{CIoU}\right) \quad (3)$$

为了得到旋转矩形框来更加紧凑地框选出目标，本文增加一个旋转角度参数，引入角度损失函数 L_{ang}，使得总体损失值对处理真实目标矩形框的位置、大小和旋转角度变得敏感起来。由于光学遥感数据集大多数采用四点式（x,y,w,h）标注方式，或八点式（$x_1,y_1,x_2,y_2,x_3,y_3,x_4,y_4$）标注方式，没有直接给出角度信息，本文可利用点位坐标来自动计算真实矩形框的角度。先回归，再预测目标矩形框的宽高与中心点坐标，再将矩形框进行一定角度的旋转，来逼近真实目标矩形框的位置。因此本文设计角度损失函数采用 L2 范数损失函数，如式(4)所示：

$$L_{\text{ang}}=\sum_{i}^{N}I_i^{\text{obj}}\left(\hat{C}_{\theta i}-O_{\theta i}\right)^2 \quad (4)$$

式中：$\hat{C}_{\theta i}$ 表示经过网络模型最终预测输出角度值；$O_{\theta i}$ 表示根据点位自动计算得出的真实目标矩形框的旋转角度。由于旋转框的定义方式是与框的边界相关的，并且在上文提到的两种定义方式的角度变化都是

具有周期的特性。因此这两种定义方式均会出现预测矩形框相对于真实矩形框的 IoU 较大，但损失也较大的情况。但由于如果使用 OpenCV（Open Source Computer Vision Library）定义法，角度的取值在其定义边界时，不仅会造成角度损失偏大，还会造成边长产生较大损失，对模型的回归有一定影响。因此本文使用长边定义法的方式来避免损失过大的情况，对于长宽比较大的物体，更加有效消除了不必要的损失。

本文针对预先判断感兴趣区域的滑动窗口分支，定义滑动窗口损失函数如式(5)所示：

$$L_{\text{slid}} = -\sum_{i}^{M}[S_i \ln\left(\hat{S}_i\right) + \left(1-S_i\right)\ln\left(1-\hat{S}_i\right)] \tag{5}$$

式中：M 为整幅遥感图像作为窗口依次送入模型的数量；S_i 为第 i 个滑动窗口内是否包含感兴趣区域的真实值，即包含目标则为 1，反之为 0。$\hat{S}_i$ 为滑动窗口分支的输出，作为对滑动窗口内含有目标的概率。

所以在训练过程中，总损失函数由分类损失、置信度损失、定位损失、角度损失以及滑动窗口损失组成，总损失函数如式(6)所示：

$$\begin{aligned}\text{Loss} = {} & \lambda_{\text{cla}}L_{\text{cla}} + \lambda_{\text{conf}}L_{\text{conf}} + \lambda_{\text{loc}}L_{\text{loc}} \\ & + \lambda_{\text{ang}}L_{\text{ang}} + \lambda_{\text{slid}}L_{\text{slid}}\end{aligned} \tag{6}$$

式中：λ_{cla}、λ_{conf}、λ_{loc}、λ_{ang}、λ_{slid} 均为可调节的权重参数，用于平衡不同 loss 之间的重要度。

2 实验与结果分析

2.1 数据集

本文实验使用的光学遥感数据集为实验室自制飞机数据集 CASIA-plane78[23]和公开的舰船数据集 HRSC2016。CASIA-plane78 数据由国产自主产权系列卫星拍摄，从多景超高分辨率图像中切取机场部分，最终制作尺寸为 4096 pixel×4096 pixel，包含多种成像条件的机场图像，共计 534 幅。数据集细分为 28 类，带有具体型号的飞机目标。由于图像分辨率较高，不加入滑动窗口分支训练时，无法直接输入到目标检测器进行训练预测，所以将数据集切片处理，保留 500 的重叠区域，以 1024 pixel×1024 pixel 进行滑动窗口切割，最终制作训练集包含 10495 幅含有目标的图像，1350 幅简单背景图像，制作验证集 4437 幅图像，测试集共计 3820 幅图像。

HRSC2016 数据集是公开的光学遥感舰船检测数据集，包含有 1061 幅从 Google 地球收集的遥感图像。由于其像素大小分布差异较大，本文在构建数据集时使用数据增强来扩充样本的数量，并统一将图像的尺寸调整为 1024 pixel×1024 pixel，并最终生成 1125 幅图像组成的训练集以及 780 幅图像组成的测试集。

2.2 评价指标

目标检测中常见的性能评估指标有：精确率（Precision）、召回率（Recall）、平均精度均值（mAP）、速度评估指标（FPS）等。

精确率又称查准率，指模型预测的所有目标中预测正确的比例，可评估预测的准不准。召回率又称查全率，指所有真实目标中，模型预测正确的目标比例，可评估预测的全不全。精确率与召回率的计算如式(7)：

$$\begin{aligned}\text{Precision} &= \frac{\text{TP}}{\text{TP+FP}} \\ \text{Recall} &= \frac{\text{TP}}{\text{TP+FN}}\end{aligned} \tag{7}$$

式中：TP 为预测正确的矩形框数量，即真样本。FP 作为非目标但预测为目标的数量，即假样本。FN 为真实目标漏检的数量。而 AP 为精确率-召回率曲线 p(r)与坐标轴所围成的面积，衡量的是学习出来的模型在每个类别上的好坏，其值越高则代表算法在该数据集上的表现越好。平均精度均值（mAP）就是取所有类别上 AP 的平均值，表示的是学习出的模型在所有类别上的好坏。AP 公式如式(8)所示：

$$\text{AP} = \int_0^1 p(r)\,\mathrm{d}r \tag{8}$$

速度评估指标 FPS 常用来衡量目标检测模型的运行速度，它的数值大小表示的是在每秒钟能够推理的图像的数量。通常视频播放速度为 24 FPS，因此目标检测算法想要满足高速实时性的要求，运行速度就要达到至少 24 FPS。

2.3 实验平台

本文所有实验在 Ubuntu20.04.2LTS 系统下进行，使用深度学习框架 pytorch1.10.1，CUDA 版本为 11.4，显卡为四块 NVIDIA GeForce RTX 3090。模型训练的迭代次数设置为 500，Batch size 设置为 64。初始学习率为 0.01，在训练过程中采用余弦退火策略动态调整，动量设置为 0.937，权重衰减率设为 0.0005。

2.4 实验结果

本文算法在 YOLOv5 网络的基础上进行改进。在 CASIA-plane78 数据集上对 Faster R-CNN、YOLOv4、YOLOv5 以及本文改进后的算法 YOLOv5-LR 对 FPS 以及 mAP 进行对比，数据如表 1 所示。可以看出，相较于其他网络，本文模型在检测速度、以及检测精度有比较优秀的表现。原始 YOLOv5 网络虽然检测速度最快，但是检测精度相较于本文模型较低。本文提出改进算法 YOLOv5-LR 在保证了检测速度相差不大，满足实时检测的需求，mAP 值提升了 3.42%。

表 1 CASIA-plane78 数据集不同检测方法对比

Table 1 Comparison of different detection methods in CasIA-Plane78 dataset

Methods	mAP/%	FPS
Faster R-CNN	87.9	16.3
YOLOv4	89.1	59.8
YOLOv5	93.5	117.6
Ours	96.7	106.4

除此之外，通过对 CASIA-plane78 数据集的大量测试结果进行观察，发现在用滑动窗口扫描整幅图像，且滑动窗口附有重叠区域的情况下，当感兴趣目标恰好落在重叠区域时候，感兴趣目标会被检测器多次检测到，并误将一个物体当成了两个物体。这样会使同一个感兴趣目标被多个矩形框重叠框选。经过分析，本文认为检测器在检测的后处理部分的非极大值抑制算法 NMS 不够准确，仅仅通过设置阈值来保留置信度最高的矩形框，剔除其他与该框重叠的矩形框。通常情况下，阈值被定义为某一个数值，它常与置信度最高的矩形框与其他各个矩形框的重叠面积进行比较，即两个框之间的 IoU 大于阈值则后者为冗余框。

图 6 为改进前后部分检测结果对比图，其中，第一行为原始 YOLOv5 检测结果；第二行为改进算法 YOLOv5-LR 检测结果。可以看出原始网络在排列密集、分布无规律的小型飞机目标，会出现漏检，一个目标被检测为多个目标的情况，如图 6(a1)、(a2)和(a3)；使用改进后的 YOLOv5-LR 没有发生多检、漏检的情况如图 6(b1)、(b2)和(b3)。YOLOv5 对于排列密集的大型目标很不友好，检测得到的矩形框由于相互之间重叠情况较多、检测矩形框的内部冗余背景较大，使得观察结果较为困难，且会出现误检的现象如图 6(a4)。而改进后的 YOLOv5-LR 算法引入角度损失函数，增加一个旋转角度参数，在这种情况下，能够更加紧凑地框选出密集大型目标如图 6(b4)，检测结果更为简单明了。

以上方法存在以下问题，例如，对重叠区域进行合并检测结果时，两个大小不同的矩形框框选同一个感兴趣区域，如图 7(a)所示，且矩形框的位置关系为大框包含小框，然而 IoU 正好小于阈值，检测器不会自动删除某一个矩形框，而是都保留了下来，但显然其中一个是冗余框。如果仅仅将阈值调小，检测器会过滤掉其中一个矩形框，那个矩形框一定是两框之中置信度较小的矩形框。重要的是不能确定置信度较小的矩形框也是在包含关系中较小的那一个矩形框。这样意味着仅仅过滤置信度较小的矩形框也许会将大的、包含整个目标的矩形框给删除掉，留下一个小的、仅仅包含目标部分区域的矩形框，如图 7(b)所示。

因此，本文对后处理部分的非极大值抑制算法 NMS 进行改进，增加一项过滤冗余框的规则，即在滑动窗口的重叠区域里，如果出现上述大框包含小框的情况，优先去删除较小的矩形框，保留包含范围较大的矩形框。由于这种做法是在检测过程的后处理部分，不会影响网络对本身就存在重叠的目标进行检测的精度，并且在整幅图像对应的滑动窗口的重叠区域中，能够保留更完整的目标，同时会解决“多检”的情况。如图 7(c)所示。改进后的非极大值抑制算法 NMS 已应用到 YOLOv5-LR，效果如图 7(d)所示。

为了体现本文改进 YOLOv5 的有效性，针对长宽比更大的舰船数据集 HRSC2016 进行了目标检测算法性能比较，与目前现有旋转目标检测算法以及原始 YOLOv5 进行了对比，如表 2 所示，本文改进算法 YOLOv5-LR 相比于原始 YOLOv5 的检测精度提升了 2.93%。

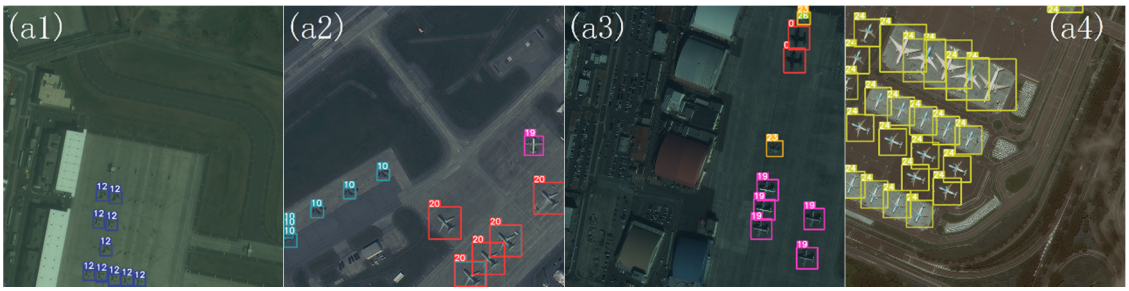

图 6 数据集 CASIA-plane78 不同方法的检测结果。(a) YOLOv5 算法；(b)改进后 YOLOv5-LR 算法

Fig.6 Detection results of different methods in dataset CasIA-Plane78. (a) YOLOv5 algorithm; (b) Improved YOLOV5-LR algorithm

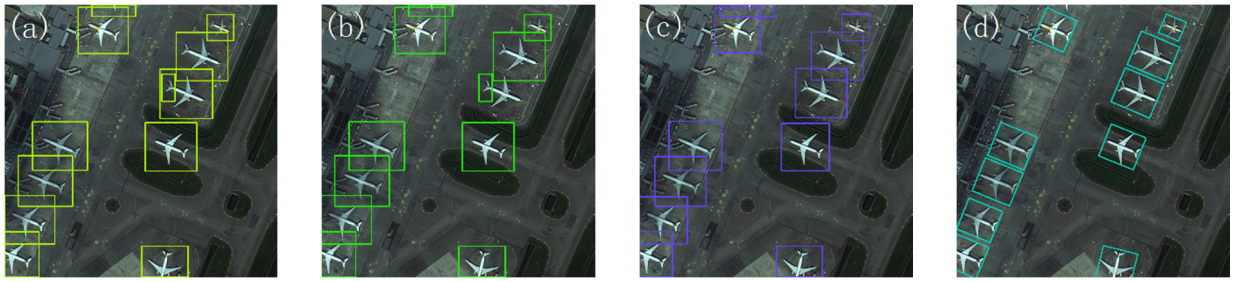

图 7 不同 NMS 结果对比。(a)原检测结果；(b)调整 NMS 阈值结果；(c)改进 NMS；(d)YOLOv5-LR 结果

Fig.7 Compare the results of different NMS. (a) Original detection result; (b)Result of adjusting the NMS threshold; (c) Improve the NMS; (d) The result of YOLOv5-LR

表 2 HRSC2016 数据集不同检测方法对比

Table 2 Comparison of different detection methods in HRSC2016 dataset

Methods	mAP/%	FPS
RR-CNN	79.6	5.06
R3Det	89.2	12.13
YOLOv5	95.5	126.58
Ours	98.3	104.17

图 8 展示了原始 YOLOv5 及改进算法 YOLOv5-LR 的检测效果。第一行为原始 YOLOv5 检测结果；第二行为改进算法 YOLOv5-LR 检测结果。如图 8(a1)所示，当舰船目标的边界长、宽，相对于光学图像的边界较为平行的时候，原始网络检测结果良好，冗余背景信息不多但依旧存在。在改进网络检测结果图 8(b1)中，在矩形框范围内尽可能剔除不必要的背景信息，针对这种情况检测效果提升不大，但在光学遥感图像的拍摄中，舰船目标不可能总是平行于整幅图像的边界，由于这种随机性的存在，原始网络仅仅在这种情况下表现良好，并不能体现算法的适用性。

当舰船目标分布密集且舰船较大，原始 YOLOv5 会出现定位不准确、漏检的情况，如图 8(a2)、(a3)和(a4)。本文发现，造成漏检严重的情况是因为舰船目标相对于其他目标，长宽比较大，因此使用正矩形框选目标后，目标常位于矩形框的对角线处，而对角线两侧包含大量冗余背景信息，这样会使得整个矩形框的范围比较大，在面对密集排列目标时，一个已经包含整个目标的矩形框会覆盖另一个目标的部分区域，这样使得本来另一个目标的矩形框与该矩形框的 IoU 值比较大，导致非极大值抑制算法 NMS 的错杀。由于目标密集，多个目标的多尺度检测框重叠区域较大，因此仅调小 NMS 的阈值可能造成更多的漏检，调大阈值可能造成更多的误检、多检。改进算法 YOLOv5-LR，利用自注意力机制增强了网络对目标与背景的辨别能力，增加角度损失函数，来使得矩形框具备角度参数，让检测结果更加紧凑地框选出目标、剔除冗余信息，有效解决遥感图像因目标排列密集、尺度变化多样导致的漏检、定位不准确的问题。

本文引入的滑动窗口分支，在大尺寸光学遥感图像稀疏目标场景中，对输入图像使用提前判读操作。当滑动窗口是不存在感兴趣目标（舰船、飞机等）的简单背景滑窗，不去进行复杂的深层网络检测任务。实验表明，与原始 YOLOv5 网络相比，YOLOv5-LR 引入的滑动窗口分支，对大尺寸光学遥感图像的推理速度明显优于对先对图像进行切割、小尺寸图像逐个进行检测、合并检测结果的大尺寸光学遥感图像的传统检测步骤。

常见的光学遥感图像数据集，如本文使用的 HRSC2016 是对原始遥感图像进行了切割，只保留了部分感兴趣区域制作为数据集样本。而整幅遥感影像

尺寸往往可达 20000×20000 以上，当前公开的遥感图像数据集较少，本文引入滑动窗口分支的实验基于吉林一号某星推扫出的大尺寸多光谱影像。本文对 12 景光学遥感图像进行测试，在 YOLOv5-LR 未加入滑动窗口分支前对整幅图像的推理时间为 24705 ms，在加入滑动窗口分支后推理时间缩短至 21336 ms，在推理精度不降低的情况下，推理速度提升了 13.6%，如图 9 展示了大尺寸光学遥感图像的旋转目标检测的部分结果。

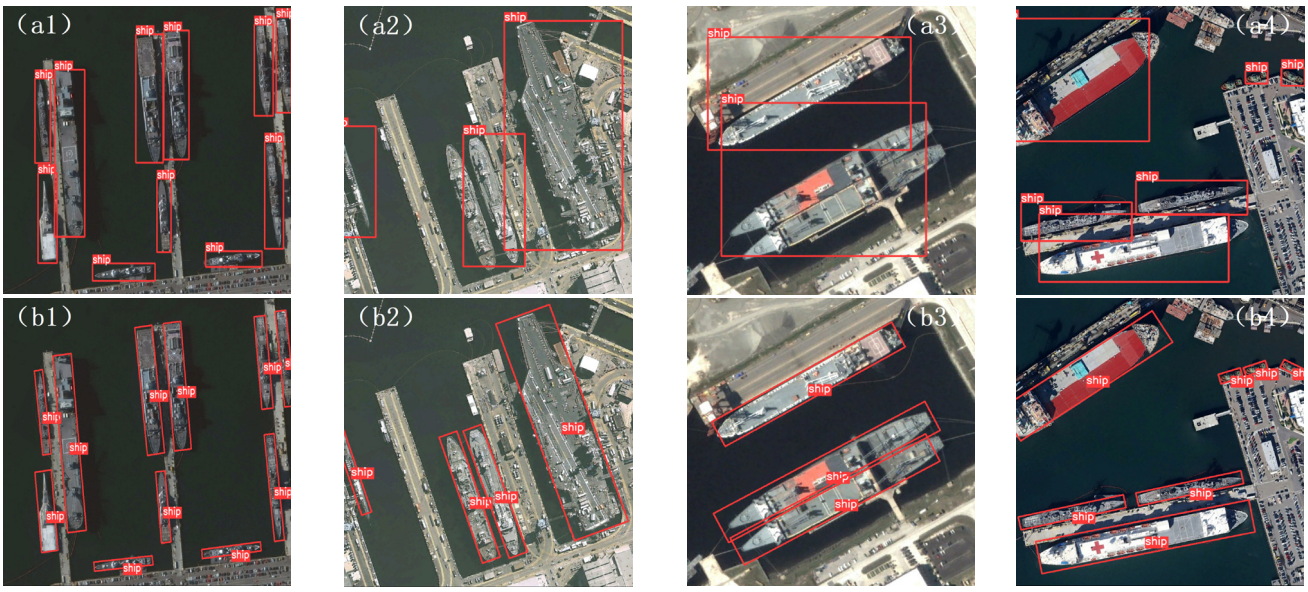

图 8 大尺寸光学遥感图像的旋转目标检测的部分结果

Fig.8 Partial results of rotating object detection in large size optical remote sensing images

图 9 数据集 HRSC2016 不同方法的检测结果。(a) YOLOv5 算法；(b)改进后 YOLOv5-LR 算法

Fig.9 Detection results of different methods in HRSC2016 dataset. (a) YOLOv5 algorithm; (b) improved YOLOV5-LR algorithm

3 结束语

本文提出了基于改进 YOLOv5 的遥感旋转目标检测算法 YOLOv5-LR。首先利用 Transformer 自注意力机制增强网络对目标与图像背景的区分能力，引入旋转角度参数，使得原检测网络选取目标的正矩形，变为包含角度参数的旋转矩形框，并对图像后处理部分的非极大值抑制算法 NMS 进行改进。其次，本文尝试在网络模型的浅层阶段，增加滑动窗口分支，来提高大尺寸遥感稀疏目标的检测效率，在保证检测精度的同时，进一步提高检测速度。

实验表明，本算法在 CASIA-plane78 和 HRSC2016

数据集上取得良好的表现，与原始 YOLOv5 网络相比在精度上有所提升，增强了冗余背景信息的剔除能力，有效解决遥感图像因目标排列密集、尺度变化多样导致的多检、漏检、定位不准确的问题。在遥感图像方面，YOLOv5-LR 的旋转检测框更适合拟合目标形态，不仅保持了先进检测算法的精度，同时实现了更快的检测速度。引入滑动窗口分支之后，对于大尺寸光学遥感图像的推理速度有所提升，但是对于不包含简单背景的大尺寸光学遥感图像来说，引入滑动窗口分支可能会导致推理速度更慢。因此，研究如何更高效地对大尺寸光学遥感图像进行检测，并基于旋转框来自动区分目标首尾将是本文后续研究的一个方向。

参考文献：

[1] ZHANG X, CHEN G, LI X, et al. Multi-oriented rotation-equivariant network for object detection on remote sensing images[J]. *IEEE Geoscience and Remote Sensing Letters*, 2022, **19**: 1-5.

[2] WANG Yi, Syed M A B, Mahrukh K, et al. Remote sensing image super-resolution and object detection: Benchmark and state of the art[J]. *Expert Systems with Applications*, 2022, **197**: 116793.

[3] XI Y Y, JI L Y, YANG W T, et al. Multitarget detection algorithms for multitemporal remote sensing data[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, **60**: 1-15,

[4] WANG Y Q, MA L, TIAN Y. State-of-the-art of ship detection and recognition in optical remotely sensed imagery[J]. *Acta Automatica Sinica*, 2011, **37**(9): 1029-1039.

[5] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [J]. *Communications of the ACM*, 2017, **60**(6): 84-90.

[6] WANG W, FU Y, DONG F, et al. Semantic segmentation of remote sensing ship image via a convolutional neural networks model[J]. *IET Image Processing*, 2019, **13**(6): 1016-1022.

[7] 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.

[8] 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.

[9] FANG F, LI L, ZHU H, et al. Combining faster r-cnn and model-driven clustering for elongated object detection[J]. *IEEE Transactions on Image Processing*, 2020, **29**: 2052-2065.

[10] LIU W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector[C]//*Proc of the European Conference on Computer Vision*, 2016: 21-37.

[11] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: unified, real time object detection[C]//*Computer Vision and Pattern Recognition*, 2017: 6517-6525.

[12] Redmon J, Farhadi A. YOLO9000: Better, faster, stronger[C]//*IEEE conference on Computer Vision and Pattern Recognition*, 2017: 6517-6525.

[13] Redmon J, Farhadi A. Yolov3: An incremental improvement[C]//*IEEE conference on Computer Vision and Pattern Recognition*, 2018, arXiv:1804.0276.

[14] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4:Optimal Speed and Accuracy of Object Detection [C]//*IEEE conference on Computer Vision and Pattern Recognition*, 2020. arXiv: 2004.10934.

[15] ZHU Wentao, LAN Xianchao, LUO Huanlin, et al. Remote sensing aircraft target detection based on improved faster R-CNN[J]. *Computer Science*, 2022, **49**(6A): 378-383.

[16] LI D, ZHANG J. Rotating target detection for tarpaulin rope based on improved YOLOv5[C]// 5th *International Conference on Artificial Intelligence and Big Data* (ICAIBD), 2022: 299-303.

[17] YANG X, YANG J R, YAN J C, et al. SCRDet: Towards more robust detection for small, cluttered and rotated objects[C]//2019 *IEEE/CVF International Conference on Computer Vision* (ICCV), 2019: 8232-8241.

[18] WANG B R, LI M. A structure to effectively prepare the data for sliding window in deep learning[C]// *IEEE* 6th *International Conference on Signal and Image Processing* (ICSIP), 2021: 1025-2018.

[19] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: transformers for image recognition at scale[J/OL]. *Computer Science*, 2010, https://arxiv.org/abs/2010.11929.

[20] LAN Lingxiang, CHI Mingmin. Remote sensing change detection based on feature fusion and attention network[J]. *Computer Science*, 2022, **49**(6): 193-198.

[21] LIU Z, WANG H, WENG L, 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*, 2017, **13**(8): 1074.

[22] LI Y, LI M, LI S, et al. Improved YOLOv5 for remote sensing rotating object detection[C]//6th *International Conference on Communication, Image and Signal Processing* (CCISP), 2021: 64-68.

[23] Institute of Automation. Chinese Academy of Sciences Remote sensing artificial intelligence algorithm competition platform[EB/OL]. https://www.rsaicp.com/portal/dataDetail?id=34.