针对遥感图像细长旋转目标检测的YOLOv5模型研究

仲波，杨璐源
（重庆邮电大学计算机科学与技术学院，重庆 400065）

摘要：针对遥感图像中细长旋转目标，（例如桥梁，港口）检测时精度较低（桥梁检测精度不超过50%，港口精度不超过70%），角度预测困难等问题，本文提出了一种改进YOLOv5模型来提升遥感图像中的细长旋转目标的检测精度。本文对YOLOv5进行三个方面进行改进：首先，针对细长旋转目标长度较大，经常跨图像造成漏检的问题，本文在输入端增大输入图像的尺寸从而保持了目标的完整性；其次，在特征提取时将主干网的常规卷积改用可形变卷积（DNC）,减少非目标区域的干扰从而保证提取到的细长目标特征与实际情况更加吻合；最后，改进的模型还额外预测了目标长宽比，并使用预测的长宽比对预测的长和宽的结果进行调整，进一步提高预测的精度。将本文改进后的YOLOv5模型用于DOTA数据集，结果表明，该方法对遥感图像中的细长旋转目标有较好的效果，将桥梁和港口的检测精度分别提高到了52.1%和75.16%.

关键词：遥感图像；细长旋转目标；YOLOv5；DCN
中图分类号：TP391.4

Improved YOLOv5 in Remote Sensing Slender and Rotating Target Detection

Zhong bo, Yang Luyuan
(Chongqing University of Posts and Telecommunications, College of Computer Science and Technology, City Chong Qing, 400065)

Abstract: Although the deep-learning based object detecting methods for remote sensing imagery have achieved very high accuracy on most of the targets, the slender and rotating targets in remote sensing images, such as bridges and harbors, are exceptional. The problem is that the predefined anchor boxes usually cannot well cover these targets. Thus, in this study an improved YOLOv5 model is proposed to better detect the slender rotating targets in remote sensing imagery. Firstly, the size of input images is doubled to maintain the integrity of the target. Secondly, the deformable convolution (DCN) is applied in the backbone net-work during feature extraction to increase the coverage of the feature and decrease the interruption by the background while extracting the slender rotating targets. Finally, besides length and width of a bounding box, the aspect ratio of the bounding box is also added to the loss function to emphasize the im-portance of the uniqueness of the slender rotating targets and to improve the prediction accuracy subse-quently. The three modifications have been applied to the YOLOv5 model respectively and together on DOTA dataset. The results show that the proposed method can better detect slender and rotating targets in remote sensing images and the accuracy for bridge and harbor has been improved to 52.1% and 75.16% respectively.

Key words: remote sensing image; slender rotating target; YOLOv5; DCN

0 引言

目标检测是计算机视觉中一项基本而又具有挑战性的任务。它要求算法为图像中每个感兴趣的实例预测一个带有类别标签的边界框。近年来，随着卷积神经网络的发展，目标检测

基金项目：中国科学院战略性先导科技专项(XDA20100101)
作者简介：仲波(1978－), 男, 副研究员、硕导，主要研究方向：遥感大数据、遥感数据处理等. E-mail: zhongbo@radi.ac.cn

在自然图像中取得了巨大的成功。由于针对遥感图像的目标检测方法能够在城市规划、军事侦察等方面得到了广泛的发展和应用[1]。使用主流目标检测模型（例如 Faster-RCNN[2]，YOLOv3[3]，YOLOv4[4]，SSD[5]等），在遥感图像目标检测中也得到了广泛的应用，并取得了一定的进展。然而遥感图像自身具有背景复杂、规模大、目标小且方向任意、尺度差异较大等特点，使得主流目标检测模型在一些目标上的检测精度很难提高。尤其是拥有极大的长宽比的细长目标难度最大，其中港口的检测精度仅接近 70%，桥梁的检测精度低于 50%，这类目标的大小差距极大，有些桥梁仅占检测图像的很小的区域，而且有些则跨越了很多个检测图像；另外，细长旋转目标长和宽的比值差异巨大，正是这一差异造成预测边界框与目标之间的交并比（Intersection over Union, IoU）变化十分敏感，微小的角度变化，会导致 IoU 出现较大的变化，从而造成预测框难以匹配目标，使得这类目标的检测精度会低于其他类型目标的精度。细长旋转目标很少被单独研究，但是在遥感图像中却很常见；因此，有必要专门针对遥感图像中的细长旋转目标检测进行研究。

伴随着 DOTA[6]数据集的出现以及基于深度学习目标检测算法的快速发展，第一个遥感图像旋转目标检测模型 FR-O 被提出，该模型基于 Faster R-CNN 并在最终的结果中加入了角度值的预测，最终预测的结果精度并不高，特别是桥梁远低于其他类别的精度。Jiang等人提出了 R^{2}CNN[7]，该算法使用了 3×11 和 11×3 两种卷积核拟提高细长旋转目标的检测精度；但是仅仅增加以上两种卷积核其实依然很难提取到细长旋转目标的有效特征。Dai 等人在 2017 年提出了可形变卷积[8]（Deformable Convolutional Networks, DCN），可形变卷积在原始的常规卷积基础之上加入了卷积位置偏移量，该偏移量有 x 和 y 两个方向，并且都是通过对原始特征层卷积得到，使得可形变卷积可以自适应地学习感受野。2018 年 Zhu 等人针对于 DCN 可能出现的感受野对应位置超出目标范围的问题，提出了 DCNv2[9]；其在 DCN 的基础之上给每个采样点增加了一个惩罚项 Δm_k，影响该特征对输出的重要性。

本文鉴于以上方法并且在 YOLOv5 的基础之上，针对细长旋转目标长宽比大，特征难以提取的问题，对 YOLOv5 进行的改进。在 YOLOv5 模型中加入了可形变卷积，加强了对细长旋转目标的特征提取。在最后的预测值中相较于 YOLOv5 加入了长宽比的预测，使得能够使用预测的长宽比对预测边界框的长宽比进行计算。在不严重影响模型效率的同时，也提高了模型对细长旋转目标的检测精度。

1 YOLOv5

1.1 YOLOv5 结构

现有的目标检测算法分为两类。第一类为两阶段模型，例如 Faster-RCNN。两阶段模型的第一个阶段为提出候选框，第二个阶段是对候选框进行判断并且预测出最后的结果。两阶段模型精度较高，但是模型的效率较低。第二类是单阶段模型，其中 YOLO 系列模型是单阶段模型中最具有代表性的模型之一，单阶段模型是一个端到端的模型，它将原始图像经过卷积操作之后获得预测的位置和类别信息，因此效率比两阶段模型高。

YOLO 系列模型经过几代的发展形成了现在 YOLOv5。YOLOv5 是现有单阶段模型中少有兼具效率和精度的模型。图 1 展示了一个 YOLOv5 的结构示例。 YOLOv5 主干网由 1 个 Focus 层，3 个 CSP1 层，4 个 CBL 层和 1 个 SPP 层组成。经过主干网输出原始图像 1/8，1/16 和 1/32 三种尺度的特征图。YOLOv5 的特征增强部分使用了 PANet[10]和 FPN[11]结构，将主

干网输出的三种尺度的特征图输入到特征增强部分进行特征融合。经过特征融合后使得输出的三个尺度的特征图都拥有丰富的语义信息和位置信息。最终模型根据预测值并结合锚框计算出预测的边界框。由于本文是对旋转目标进行检测，需要在最终结果中预测出角度信息，因此将预测旋转目标 YOLOv5 模型命名为 YOLOv5-O，并将在后文使用。

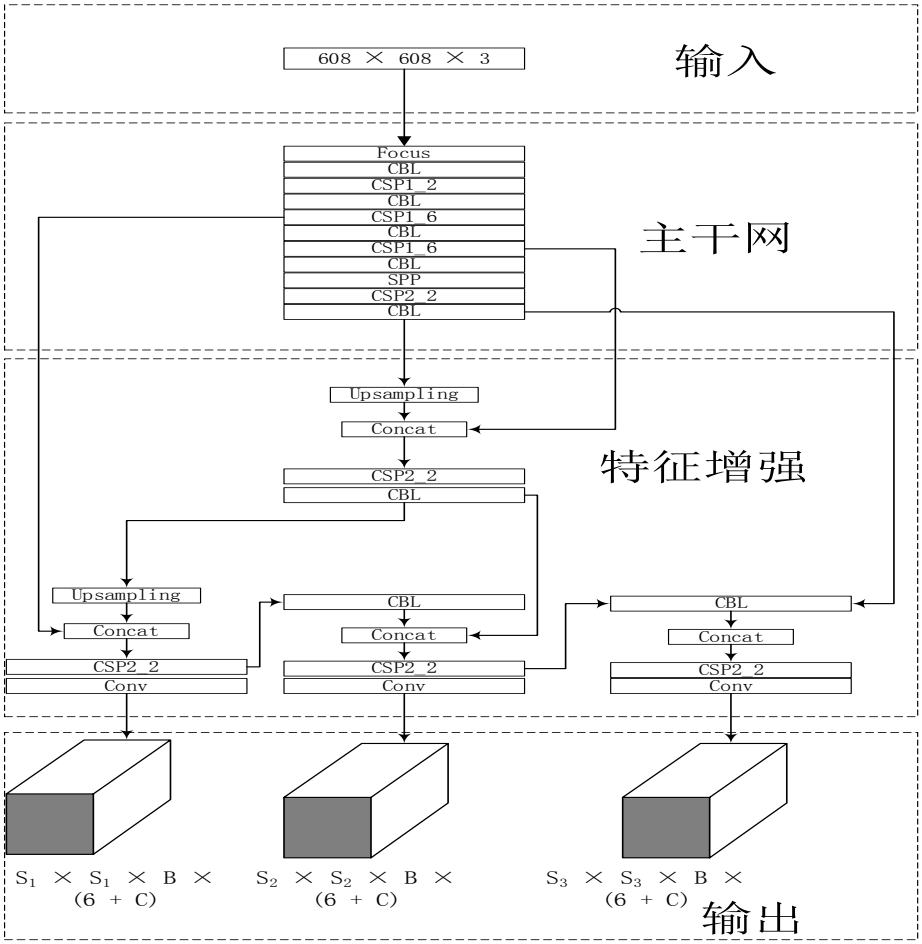


图 1　YOLOv5-O 模型结构

Fig. 1　YOLOv5-O Model Structure

1.2　YOLOv5-O 边界框预测

YOLOv5-O 先将输入的原始图像划分为 S×S 个网格，每个网格单元预测 B 个边界框的坐标信息，包含 B 个边界框的得分以及 C 个类别的得分。

1.2.1　YOLOv5-O 边界框分类信息预测

每个网格单元 B 个边界框得分反映了模型对所预测的边界框包含目标的可能性。形式上，将置信度（confidence）定义为公式（1）：

$$confidence = \Pr(Object) \times IoU_{pred}^{truth} \tag{1}$$

其中 $\Pr(Object)$ 为当前预测边界框中出现目标的可能性，如果边界框中包含物体，则 $\Pr(Object)$ 为 1，使得置信度为当前预测边界框与真实标注的 IoU，否则 $\Pr(Object)$ 为 0，使得置信度也为 0。

每个网格单元还预测了 C 个条件类别概率，即在一个栅格包含一个目标的前提下，它

属于某个类的概率。YOLOv5-O 会为每个边界框预测一组类概率，即 $\Pr(Class_i|Object)$。结合公式（1），预测边界框的类别概率如公式（2）所示：

$$\Pr(Class_i|Object)\times\Pr(Object)\times IoU_{pred}^{truth}=\Pr(Class_i)\times IoU_{pred}^{truth} \quad (2)$$

1.2.2 YOLOv5-O 边界框位置信息预测

每个网络单元为每个边界框预测 4 个坐标和 1 个角度，分别为 t_x，t_y，t_w，t_h，t_θ。如图 2 所示，红色的框为预先设定的锚框，蓝色框为预测的边界框，c_x 和 c_y 分别表示当前预测的边界框的中心坐标相对于图像左上角的偏移量，利用预测值 t_x，t_y 计算出当前中心点相对于当前网格左上角的偏移量，结合 c_x 和 c_y 可计算出当前预测边界框的中心坐标在整张图像上的坐标 b_x 和 b_y。p_w 和 p_h 是当前锚框的宽和长，可利用预测值 t_w 和 t_h 并结合 p_w 和 p_h 获得当前预测边界框的宽（b_w）和长（b_h），YOLOv5-O 对角度 b_θ 直接进行回归，且角度的回归范围 0°~180°。预测边界框的中心坐标，长，宽和角度计算公式（3）所示：

$$\begin{cases} b_x=\sigma(t_x)\times 2-0.5+c_x \\ b_y=\sigma(t_y)\times 2-0.5+c_y \\ b_w=p_w\times(\sigma(t_w)\times 2)^2 \\ b_h=p_h\times(\sigma(t_h)\times 2)^2 \\ b_\theta=t_\theta \end{cases} \quad (3)$$

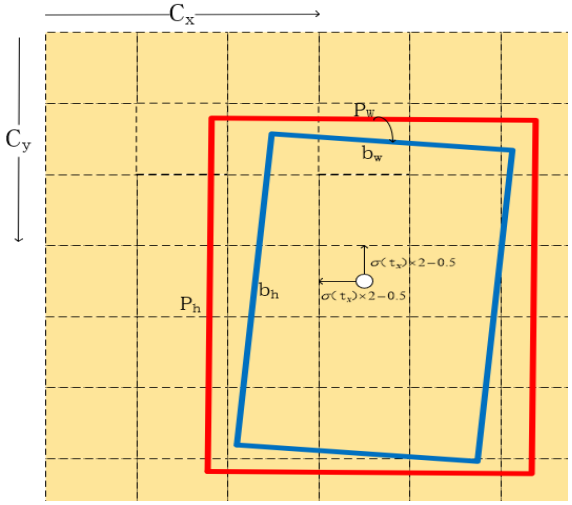


图 2 YOLOv5-O 边界框预测

Fig. 2 YOLOv5-O Bounding Box Prediction

2 YOLOv5-O 网络改进

YOLOv5 包含 YOLOv5s，YOLOv5m，YOLOvl，YOLOv5x，分别表示模型不同的大小（内存），其中 YOLOv5s 最小，YOLOv5x 最大。本文后续的改进会增加模型的大小并且导致模型的训练时间和推理时间加长，因此综合考虑了模型的精度和效率，将 YOLOv5m 作为本文的基础模型，由于是旋转框目标检测模型，因此命名为 YOLOv5m-O。

2.1 对 YOLOv5m-O 输入端的改进

原始 YOLOv5m-O 的输入图像大小为 608×608×3。遥感图像与普通图像相比，遥感图像

中存在着较多的细长目标，图像的大小会影响细长目标的完整性，对模型的训练造成影响。考虑以上遥感图像的特殊性，有必要对 YOLOv5m-O 的输入端和数据增强做针对性的改进。

原始 YOLOv5m-O 的输图像是使用的 608×608×3，如果以 608×608×3 的大小对图像进行裁剪，会裁剪掉较多的细长目标，对训练造成较大的影响。因此，本文将 YOLOv5m-O 模型的输入图像改为了 1024×1024×3，尽量保证了更多图像中目标的完整性，如图 3 所示：

(a) 608×608×3 图像　　(b) 1024×1024×3 图像

图 3　输入图像

Fig. 3　Input Image

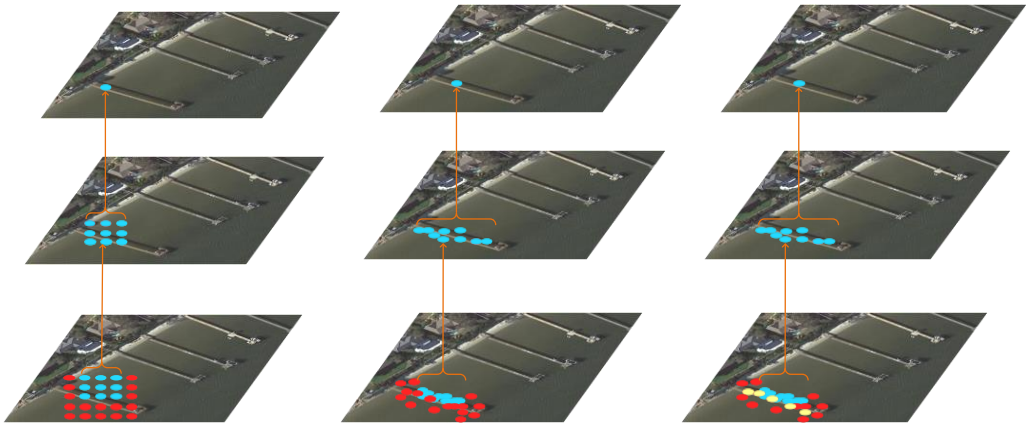
2.2　在 YOLOv5m-O 中加入 DCNv2

在遥感图像中存在很多细长的旋转目标（例如：桥梁，港口等），使用常规卷积时形状始终为矩形，其中大部分信息来自于非目标区域，因此很难对细长旋转目标进行有效的特征提取，如图 5 所示，仅仅使用常规的卷积可能会出现对目标的特征提取不完全并且会提取到很多的背景信息，对于正样本，采样的特征应该更多关注于感兴趣区域（Regions of Interest, RoI）内，如果特征中包含了过多超出 RoI 的特征，那么一定会对结果造成影响和干扰。Zhu 等人提出了一种能够更加充分的提取目标有效特征的卷积方式 DCNv2。

对于网络的每个采样点的响应，不是图像上的所有区域都对输出有影响，去除掉一些不相关的区域，采样点的响应可以保持不变。因此，如果将每个采样点的支持区域限制到了最小，可以和整幅图产生相同响应的区域，这一区域称之为错误边界显著性区域。在目标检测中将错误边界显著性区域限制在感兴趣区域，去除掉不相关的背景信息不仅不会对检测结果造成影响反而会提升检测的精度。图 4（a）为常规卷积，图 4（b）为 DCN，可以看出无论是常规卷积还是 DCN 其错误边界显著性区域都会超过 RoI。DCNv2 在 DCN 的基础之上给每个采样点的采样特征乘以一个惩罚项，如图 4（c）所示，蓝色的采样点，红色的采样点和黄色的采样点分别拥有不同的权重，以判定这一部分区域的特征是否对输出有影响，如公式（4）所示：

$$y(p) = \sum_{k=1}^{K} \omega_k \cdot x\left(p + p_k + \Delta p_k\right) \cdot \Delta m_k \tag{4}$$

DCNv2 加入了 Δm_k 惩罚项，如果 Δm_k 为 0，则表示这部分区域的特征对输出没有影响，因此可以调整超过 RoI 的采样点对输出的影响。



(a) 常规卷积 (b)DCN (c) DCNv2

图 4 不同的卷积方式

Fig. 4 Different Convolution

在 YOLOv5m-O 中的主干网中包含 3 个 CSP1 和 1 个 CSP2，其中最基本的模块单元 CBL 由卷积、批归一化和 Leaky Relu 激活函数构成。CBL 基本结构如图 5 所示。CBL 使用的是常规卷积，对于在复杂背景下的遥感图像中细长旋转目标的特征并不能进行有效的提取。

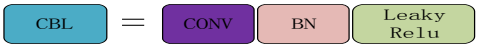


图 5 CBL 模块

Fig. 5 CBL Module

本文将 YOLOv5m-O 中 CBL 模块的常规卷积全部替换为 DCNv2，将该模块命名为 DBL。如图 6 所示 DBL 模块由 DCNv2，批归一化和 Leaky Relu 激活函数构成。该模块在保持输出特征图的大小和通道数的同时，由于计算了每个采样点的偏移量和权重的原因，能够提取到更多的前景的特征，将有效特征集中在 RoI。

图 6 DBL 模块

Fig. 6 DBL Module

2.3 YOLOv5m-O 预测框长和宽回归方式改进

YOLOv5m-O 是一种基于先验框的目标检测模型。YOLOv5m-O 包含三个尺度的检测头，每个尺度检测头的输出特征图中的每个网格点又分别包含 3 个先验框；因此，YOLOv5m-O 模型包含 9 个锚框。以上锚的长和宽是 YOLOv5m-O 模型中的自适应锚框计算生成算法通过 *K*-means 计算获得。图 7 所示，红色为锚框，绿框为目标的标注边界框，由于细长旋转目标长宽比较大，即使锚框的大小是经过 *K*-means 算法计算获得，也很难很好的覆盖目标，造成边界框的长和宽难以回归。原始 YOLOv5m-O 边界框长和宽计算公式如公式（5）：

$$\begin{cases} b_w = p_w \times (\sigma(t_w) \times 2)^2 \\ b_h = p_h \times (\sigma(t_h) \times 2)^2 \end{cases} \tag{5}$$

图 7　锚框和边界框

Fig. 7　Anchor Box and Bounding Box

p_w 和 p_h 是当前锚框的宽和长，利用模型输出预测值 t_w 和 t_h 并结合 p_w 和 p_h 可获得当前预测边界框的宽（b_w）和长（b_h）。根据公式（5）计算出预测框的宽（b_w）和长（b_h），在训练时 YOLOv5m-O 的边界框宽和长的损失计算公式如下，

$$l_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B}I_{ij}^{obj}\left(2-w_i^j\times h_i^j\right)\times\left[\left(w_i^j-\hat{w}_i^j\right)^2+\left(h_i^j-\hat{h}_i^j\right)^2\right] \tag{6}$$

其中，I_{ij}^{obj} 表示第 i 个网格的第 j 个锚框是否负责预测目标边界框的位置，若负责预测则为 1，否则为 0。w_i^j 和 h_i^j 为当前预测边界框的宽和长，$\hat{w}_i^j$ 和 $\hat{h}_i^j$ 为目标边界框的宽和长。由于本文主要的研究目标为遥感图像中的细长目标，用此方法难以对边界框长宽进行回归。因此，针对以上问题，提出了一种新的边界框长和宽的回归方式，方法是在 YOLOv5m-O 的预测值的基础之上额外预测一个长宽比 r。相较于原始的 YOLOv5m-O 模型的损失函数，改进后的 YOLOv5m-O 模型在边界框损失中额外计算了长宽比的损失，公式如下，

$$\begin{aligned}&l_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B}I_{ij}^{obj}\left(2-w_i^j\times h_i^j\right)\times\left[\left(w_i^j-\hat{w}_i^j\right)^2+\left(h_i^j-\hat{h}_i^j\right)^2\right]\\&+l_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B}I_{ij}^{obj}\left[\left(r_i^j-\hat{r}_i^j\right)^2\right]\end{aligned} \tag{7}$$

其中，r_i^j 是预测边界框的长宽比，$\hat{r}_i^j$ 是目标边界框的长宽比。YOLOv5m-O 模型的损失函数包含边界框损失函数，置信度损失函数和分类损失函数，因此，改进后的 YOLOv5m-O 模型的损失函数如公式（8）所示，

$$\begin{aligned}&Loss=bbox\_loss+conf\_loss+prob\_loss=l_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B}I_{ij}^{obj}\left[\left(x_i-\hat{x}_i^j\right)^2+\left(y_i-\hat{y}_i^j\right)^2\right]+\\&l_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B}I_{ij}^{obj}\left(2-w_i^j\times h_i^j\right)\times\left[\left(w_i^j-\hat{w}_i^j\right)^2+\left(h_i^j-\hat{h}_i^j\right)^2\right]+l_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B}I_{ij}^{obj}\left[\left(r_i^j-\hat{r}_i^j\right)^2\right]\\&+l_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B}I_{ij}^{obj}\left[\left(\theta_i^j-\hat{\theta}_i^j\right)^2\right]+\sum_{i=0}^{S^2}\sum_{j=0}^{B}I_{ij}^{obj}\left[\hat{C}_i^j\ln C_i^j+\left(1-\hat{C}_i^j\right)\ln\left(1-C_i^j\right)\right]-\\&l_{noobj}\sum_{i=0}^{S^2}\sum_{j=0}^{B}I_{ij}^{noobj}\left[\hat{C}_i^j\ln C_i^j+\left(1-\hat{C}_i^j\right)\ln\left(1-C_i^j\right)\right]-\sum_{i=0}^{S^2}I_{ij}^{obj}\sum_{c\in classes}\left[\hat{P}_i^j\ln P_i^j+\left(1-\hat{P}_i^j\right)\ln\left(1-P_i^j\right)\right]\end{aligned} \tag{8}$$

在推理阶段将预测的长宽比加入到预测边界框长和宽的计算之中，其公式如下，

$$\begin{cases} b_w = p_h \times (\sigma(t_h) \times 2)^2 \times r \\ b_h = p_w \times (\sigma(t_w) \times 2)^2 / r \end{cases} \quad (9)$$

其中 r 为预测的长宽比，其值为 w/h。因此，使用预测的长乘以长宽比获得最终的边界框的宽，再使用预测的宽除以长宽比获得最终的边界框的长。

3 实验与结果分析

3.1 数据集分析

目前遥感图像目标检测数据集中带有旋转标注框最常用的是武汉大学遥感国家重点实验室夏桂松团队和华科电信学院白翔团队合作在 2019 年发布的 DOTA 数据集[6]。目前经过几版本的发展，DOTA 数据集 DOTA-v1.0、DOTA-v1.5 和 DOTA-v2.0 三个版本。本文选择了目前使用最广泛的 DOTA-v1.5，其中包含 16 个类别，旋转框标注方式是四个顶点八个坐标 $x_1,y_1,x_2,y_2,x_3,y_3,x_4,y_4$ 表示的不规则四边形，在本文实验中，需要将四个顶点八个坐标 $x_1,y_1,x_2,y_2,x_3,y_3,x_4,y_4$ 的表示方法转换为中心坐标，长宽和角度的 x,y,w,h,θ 的表示方法。

由于本文只针对于细长旋转目标进行检测，因此只选取了 DOTA-v1.5 数据集中包含了的港口和桥梁的图像进行训练和测试。训练集和测试集数据的数量如表 1 所示：

表 1　数据集信息

Tab. 1　Dataset Information

类别	*训练图片*	*训练目标*	*测试图片*	*测试目标*
Bridge	1528	3687	117	235
Harbor	2030	13176	578	4557

3.2 评价指标

目前在目标检测领域一般使用平均准确率（mean Average Precision，mAP）作为评价指标。mAP 由精确率（*precision*）和召回率（*Recall*）计算获得，能够有效的避免单一指标的缺陷，其中精确率和召回率计算公式分别为公式（10）和公式（11）所示：

$$precision = \frac{TP}{FP+FN} \quad (10)$$

$$recall = \frac{TP}{TP+FN} \quad (11)$$

在以上公式中，FN（False Negative）为假的负样本，FP（False Positive）为假的正样本，TN（True Negative）为真的负样板，TP（True Positive）为真的正样本。精确率是针对预测结果而言的，它表示的是预测为正例中有多少是真正的正例。召回率是针对原来的样本而言的，它表示的是样本中的正例有多少被预测正确了。mAP 通过精确率和召回率计算获得，能够反映出精确率和召回率两项评价指标，因此，本文采用 mAP 作为实验的评价指标。

3.3 实验平台与训练策略

3.3.1 实验平台

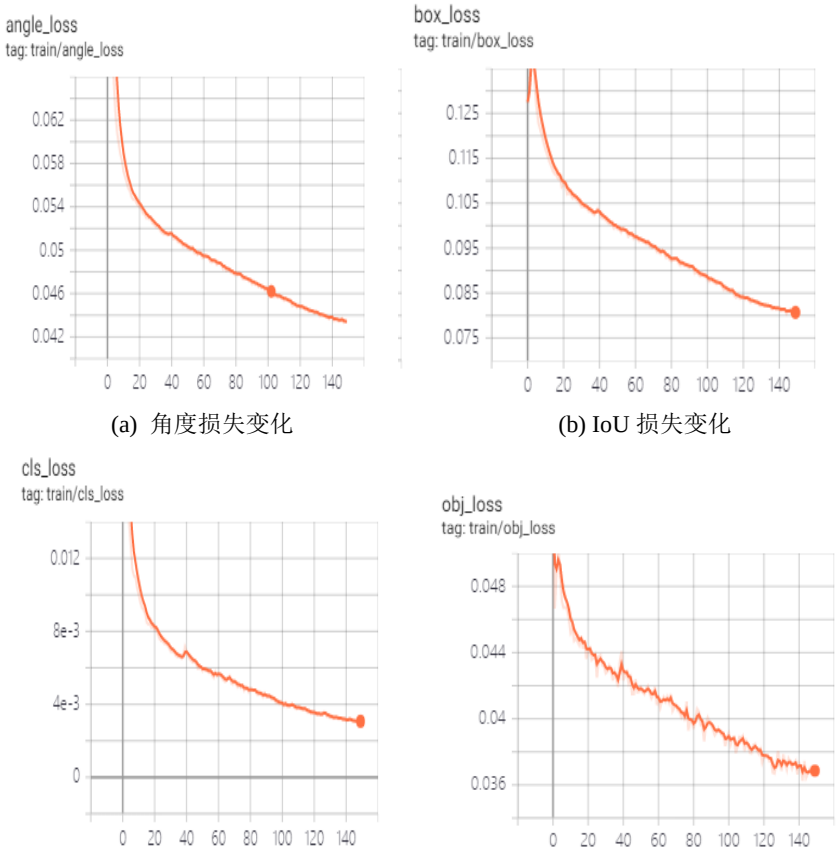
本文实验平台配置 CPU 为 Intel(R) Xeon(R) Silver 4114 CPU @ 2.20GHz，内存 64G，显卡为 TITAN RTX ，显存 24G，操作系统为 Ubuntu 18.04.1。开发环境使用了 Python 3.7.10，

Pytorch 1.6.0，torchvision 0.7.0 和 CUDA 11.0.207。

3.3.2 训练策略

在 YOLOv5 训练过程中，在不同的层使用了不同的学习率调整策略。此外，YOLOv5 训练分为了两阶段，分为 warmup 阶段和 warmup 以后的阶段。在 warmup 阶段，采用了一维线性插值的方法对每次迭代的学习率进行更新，在 warmup 以后的阶段使用了余弦退火算法对学习率进行更新。通过以上的学习率调整策略能够更好的进行模型训练。

初始学习率为 0.01，warmup 阶段动量为 0.8，warmup 阶段之后动量为 0.937，训练 epoch 为 150，batch size 为 2。下图分别为，训练阶段角度损失的变化，边界框损失的变化，类别概率的损失变化和置信度损失变化。



(a) 角度损失变化　　(b) IoU 损失变化

(c) 分类损失变化　　(d) 置信度损失变化

图 8　损失变化曲线

Fig. 8　Loss Curve

3.4　实验结果与结果分析

为了方便实验对比，本文使用的基础模型为 YOLOv5m-O，将加入了白噪声的模型命名为 YOLOv5m-O-A，加入了 DCNv2 的模型命名为 YOLOv5m-O-B，使用了改进长宽回归方式的模型命名为 YOLOv5m-O-C，使用本文提出的三种改进方式的模型命名为 YOLOv5m-O-D。并且与现有的单阶段旋转框检测模型进行了对比。本文实验和对比实验均采用了相同的 DOTA 数据集作为训练集和测试集。

表 2 实验精度

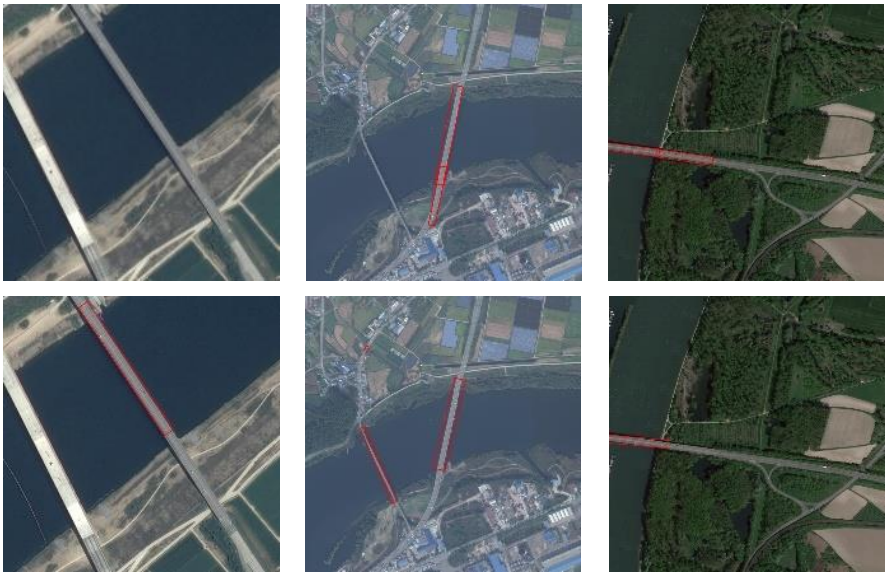
Tab. 2 Experimental Accuracy

Model	*Backbone*	*Bridge*	*harbor*
RetinaNet-O	R-50-FPN	41.81%	62.57%
DRN[12]	H-104	43.52%	67.62%
R^3Det[13]	R-101-FPN	50.91%	66.91%
PIoU[14]	DLA-34	24.10%	57.10%
S^2ANet[15]	R-50-FPN	48.37%	65.26%
YOLOv5m-O	Darknet	46.17%	67.73%
YOLOv5m-O-A	Darknet	48.30%	68.25%
YOLOv5m-O-B	Darknet	48.55%	70.96%
YOLOv5m-O-C	Darknet-DCN	50.97%	71.52%
YOLOv5m-O-D	Darknet-DCN	**52.10%**	**75.16%**

如上表所示，基础的 YOLOv5m 旋转框检测模型 YOLOv5m-O 对于处桥梁和港口的检测 mAP 分别为 46.17%和 67.73%，增大输入图像的 YOLOv5m-O-A 桥梁和港口的 mAP 分别提升了 2.13%和 0.52%，使用了 DCNv2 的 YOLOv5m-O-B 桥梁和港口的 mAP 分别提升了 2.38%和 3.23%，改进了预测框长宽回归方式的模型 YOLOv5m-O-C 在桥梁和港口 mAP 分别提升了 4.8%和 3.79%，最终集成了本文改进的所有方式的模型 YOLOv5m-O-D 的 mAP 在桥梁和港口分别提升了 5.93%和 7.43%。与当前已经提出一些遥感图像旋转目标检测单阶段模型比较，如上表所示桥梁和港口最高 mAP 分别为 R3Det 的 50.91%和 DRN 的 67.62%，本文提出的方法相较于最高精度提高了 1.19%和 7.54%。

图 9 图像分别是 YOLOv5m-O 的目标检测结果图和经过本文方法改进 YOLOv5m-O-D 的目标检测结果图，图 9（a）和（b）图像为港口的目标检测，可以看出改进后的模型能够更好的拟合目标的轮廓，由于改进后的模型预测了边界框的长宽比，并且在推理的时候将长宽比加入到模型长宽的计算，使得模型能够得出更加准确的边界框的长和宽。图 9 的后四组图像是对桥梁的预测，在改进后的模型相较于改进之前能够预测出更多的桥梁目标，减少了漏检（图 9（c）和图 9（d）图像）能够预测出更多的桥梁目标，并且同时也减少了误检（图 9（e）和图 9（f）图像），有效的区分出了桥梁和道路，这是由于在本文的模型中加入了 DCNv2 可以使得模型能够提取出有效的特征，从而提升了目标预测的成功率。同时本文增加了新的数据增强方式，提高了模型的鲁棒性。综上，改进后的 YOLOv5m-O-D 模型相较于改进前和其他模型针对于遥感图像细长旋转目标有较好效果。

(a) 对照组 a (b) 对照组 b (c) 对照组 c



(d) 对照组 d (e) 对照组 e (f) 对照组 f

图 9 实验对比

Fig. 9 Experimental Comparison

4 结束语

本文主要研究遥感图像中的细长旋转目标检测，针对遥感图像细长旋转目标长宽难以回归和特征难以提取的问题提出了一种基于 YOLOv5 模型的改进方法。主要改进三个方面，首先，在输入端增大了输入图像的大小，以保证训练目标的完整性；其次，在主干网使用了 DCNv2 保证了对目标特征提取的有效性；最后，使用预测的边界框长宽比参与到预测的长和宽的计算。将改进后的模型应用于 DOTA 数据集上，实验结果表明了本文改进后的 YOLOv5m-O 模型提升了细长旋转目标的 mAP，解决了一部分本文所提的问题。但是对于桥梁目标精度提升较少，后续工作将对再针对于桥梁的检测继续进行改进。

[参考文献] (References)

[1] 聂光涛, 黄华.光学遥感图像目标检测算法综述[J].自动化学报, 2021, 47(08):1749-1768.
[2] REN S Q, HE K M, 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.
[3] REDMON J, FARHADI A.YOLOV3：an incremental improvment[J].arXiv：1804.02767,2018.
[4] BOCHKOVSKIY A, WANG C Y, LIAO H Y M.YOLOv4：optimal speed and accuracy of object de-tection[C]//IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2020.
[5] LIU W, ANGUELOV D, ERHAN D, et al.SSD：single shot multibox detector[C]//European Conference on Computer Vision.Cham：Springer, 2016.
[6] G.-S. Xia, X. Bai, J. Ding, Z. Zhu, S. Belongie, J. Luo, M. Datcu, M. Pelillo, and L. Zhang. Dota: A Large-scale Dataset for Object Detection in Aerial Images[J]//IEEE Conference on Computer Vision and Pattern Recogni-tion, June 2018.
[7] Yingying Jiang, Xiangyu Zhu, Xiaobing Wang, Shuli Yang, Wei Li, Hua Wang, Pei Fu, Zhenbo Luo. R2CNN: Rotational Region CNN for Orientation Robust Scene Text Detection[J].arXiv：arXiv:1706.09579, 2017.
[8] Dai J, Qi H, Xiong Y, Li Y, Zhang G, Hu H, Wei Y. De-formable convolutional networks.[C]//In Proceedings of the IEEE International Conference on Computer Vision, Venice, October 2017.
[9] Xizhou Zhu, Han Hu, Stephen Lin, Jifeng Dai. Deforma-ble ConvNets v2: More Deformable, Better Re-sults[J].arXiv：arXiv: 1811.11168, 2018.
[10] LIU S, QI L, QIN H, et al.Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2018.
[11] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, Serge Belongie. Feature Pyramid Networks for Object Detection[J].arXiv：arXiv: 1612.03144, 2016.
[12] Xingjia Pan, Yuqiang Ren, Kekai Sheng, Weiming Dong, Haolei Yuan, Xiaowei Guo, Chongyang Ma, and Chang-sheng Xu. Dynamic refinement network for oriented and densely packed object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2020.
[13] Xue Yang, Qingqing Liu, Junchi Yan, Ang Li, Zhiqiang Zhang, and Gang Yu. R3Det: Refined single-stage detec-tor with feature refinement for rotating object[C]//In Pro-ceedings of the AAAI Conference on Artificial Intelli-gence, 2021.
[14] Zhiming Chen, Kean Chen, Weiyao Lin, John See, Hui Yu, Yan Ke, and Cong Yang. PIoU Loss: Towards accurate orented object detection in complex environments[C]//In Proceedings of the European Conference on Computer Vision, 2020.
[15] J. Han, J. Ding, J. Li, and G. S. Xia. Align deep features for oriented object detection[J].IEEE Transactions on Geo-science and Remote Sensing, 2021.