**一种基于复杂背景下的航天目标检测方法**

**一、发明或者实用新型的名称**

一种基于复杂背景下的航天目标检测方法

**二、所属技术领域**

本发明属于机器学习的计算机视觉领域，通过对复杂背景下的航天目标（飞机，轮船）图像特征进行充分学习，解决小目标、飞机(或轮船)相互遮挡等复杂背景下难以检测的问题。

**三、背景技术**

航天目标检测在军用方面具有重要的应用价值。这类图像相比于轮廓完整、姿态单一图像，存在着飞机(轮船)相互遮挡、拍摄距离过远造成的飞机(轮船)的尺寸较小、环境复杂等难以检测的问题。传统的目标检测算法，如：梯度直方图（HOG）+支持向量机（SVM）、可变形的组件模型（DPM）、非极大值抑制（NMS）等，检测过程主要分为三个阶段，首先是生成目标建议框，随后是提取每个建议框中的特征，最后是根据特征进行分类，然而传统的目标检测算法在预测的时间和精度方面都很不理想。

自从深度神经网络算法首次在ImageNet数据集上取得了举世瞩目的成就后，目标检测领域就逐渐开始利用深度学习来做研究。第一代R-CNN首次使用深度网络模型来提取特征，最终以49.6%的准确率开创了目标检测算法的新时代，随后，Fast R-CNN、Faster R-CNN也相继而出，这类算法是基于候选区域的，虽然具有很高的检测精度，但是实时性很差，不适用于飞机(轮船)检测。随着深度学习的不断发展，基于回归算法的检测模型再次将目标检测领域带到了一个新高度，其中以SSD和YOLO为代表的检测算法做到了真正意义上的实时检测效果。这类算法不需要生成候选区域阶段，所以也称之为One-Stage检测算法，这类算法现已广泛应用于工业检测、机器人导航、智能视频监控、航空航天等诸多领域。

YOLOV3是目前在实时性和检测精度方面都能达到很好性能的检测算法之一，国内外诸多研究者将其应用于航空航天领域，并提出了不同的改进方案。如，Kharchenko将YOLOV3算法应用到无人机拍摄的地面小物体检测，该方法具有极高的检测能力、定位精度和实时处理；Dai W C将YOLOV3算法应用到遥感图像中飞机检测，并将密集相连模块用于骨干网络，实现了多尺度遥感图像中的飞机检测。这些方法通过减少骨干网络层数、改变卷积层等方式对算法进行改进，但是在小目标及相互遮挡等复杂环境背景下，检测的效果并不理想。

**四、目的**

小目标、飞机或轮船相互遮挡等复杂背景下难以检测的问题，对飞机和轮船检测的准确性及实时性提出了很大的挑战。将实时性较高的YOLOV3算法应用到复杂背景下的航天目标检测上不能达到让人满意的程度。

YOLOV3是目前在实时性和检测精度方面都能达到很好性能的检测算法之一，国内外诸多研究者将其应用于航空航天领域，并提出了不同的改进方案。本文在YOLOV3算法的基础上进行了四点改进，大大提升算法在小目标及相互遮挡等复杂环境背景的检测的效果。

**五、技术方案**

1、总览

本文提出了一个改进的YOLOV3网络结构。所提方法的简介如图1所示。

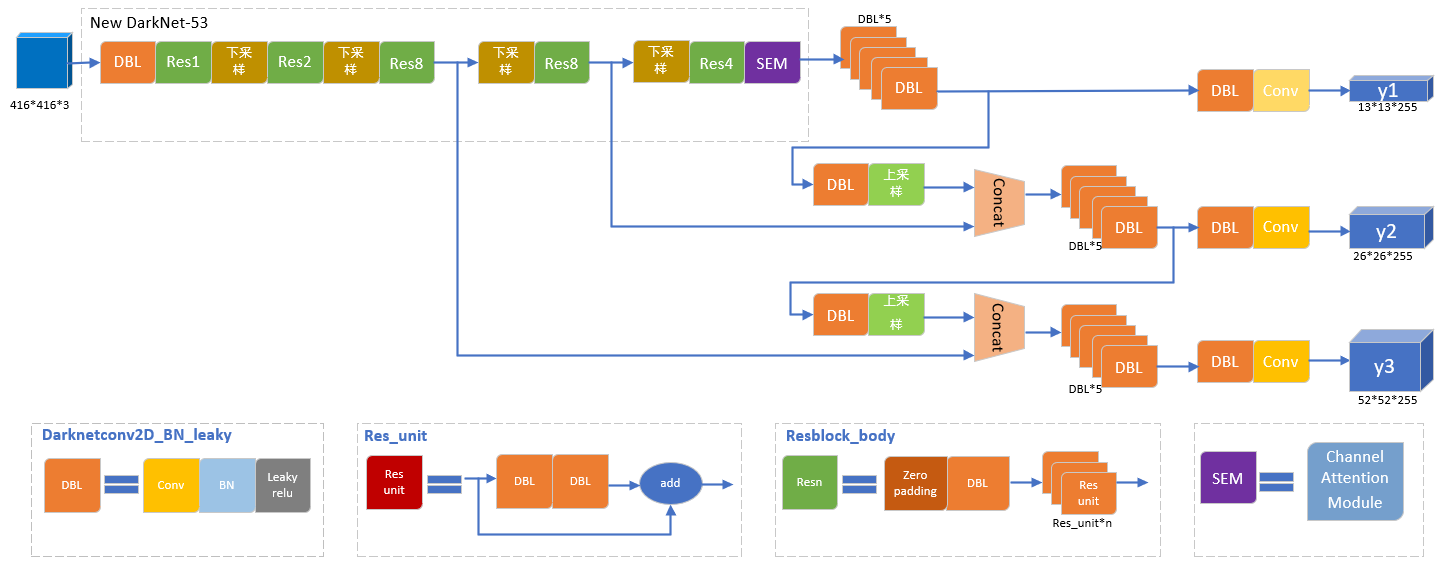


图1 改进的YOLOV3整体结构

首先，输入是416\*416\*3的图片，送入特征提取的骨干网络DarkNet-53，该网络一共53层卷积，除去最后一个全连接层总共52个卷积用于做主体网络。这52个卷积层是这样组成的：首先是1个通道数32的卷积核，然后是5组重复的残差单元Resblock\_body，每个单元由1个单独的卷积层与一组重复执行的卷积层构成，重复执行的卷积层分别重复1次、2次、8次、8次、4次；在每个重复执行的卷积层中，先执行1x1的卷积操作，再执行3x3的卷积操作，然后是下采样，通道数减半，再恢复。因为池化的下采样操作会带来信息损失，那么就把池化层去掉。但是池化层去掉随之带来的是网络各层的感受野变小，这样会降低整个模型的预测精度。我们将池化下采样替换成扩张率为2的空洞卷积瓶颈，空洞卷积的主要贡献就是，如何在去掉池化下采样操作的同时，而不降低网络的感受野。由于YoloV3感受野区域广，牺牲了空间分辨率，致使深层网络的小目标信息损失，缺乏细节信息。为了提取更多的语义信息，我们在骨干网络的最后加入了一个通道注意力模块，每个通道的特征图被认为是一个特征探测器，通道的注意力主要关注在输入图像中什么是有意义的。

在YOLOV3中采用类似FPN的上采样和融合做法，最后融合了3个尺度，大小分别是13\*13，26\*26和52\*52，在多个尺度的特征图上做检测，对于小目标的检测效果提升还是比较明显的。虽然在YOLO v3中每个网格预测3个边界框，看起来比YOLO v2中每个网格单元预测5个边界框要少，但因为YOLO v3采用了多个尺度的特征融合，所以边界框的数量要比之前多很多。

本文的主要贡献概括如下。

（1）提出了一个包含21766张图片的复杂背景下的航天对象数据集，并将改进后的YOLOV3算法应用到复杂背景下的飞机(轮船)检测。

（2）同时引入空间域和通道域注意力机制，使其在复杂背景下可以更好的忽略无关信息而关注重点信息，提高检测精度。

（3）通过线性衰减置信度的方式，对非极大值抑制NMS算法进行优化，以提升模型对遮挡飞机和轮船的检测能力。

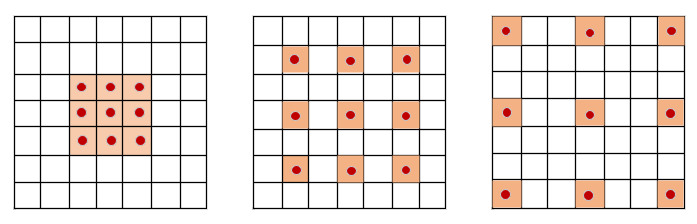
（4）对IOU算法进行了优化，解决了当把IOU作为LOSS时，因预测框和真实框不相交带来的Loss=0，从而无法进行优化的问题。

（5）将骨干网络中的卷积层替换为空洞卷积，使其保持较高分辨率及较大感受野，提高模型对小目标检测的准确性。

2、高分辨率、大感受野的骨干网络

对于小目标的检测，一方面需要高分辨率的特征图，才能检测较小区域的目标信息;另一方面需要较开阔的感受野或者更全局的信息，才能准确判断目标的位置和语义特征。为了改善复杂背景图像中小目标的检测，结合空洞卷积和FPN结构，提出一种高分辨率、大感受野的骨干网络用于特征提取。

空洞卷积是通过改变卷积核的内部间隔来扩大卷积核的感受野，图２为３种不同间隔的空洞卷积核，其中rate表示卷积核内部空洞的间隔，图2(a)为３×３、rate＝１的空洞卷积，卷积核的感受野范围是３×３。图2(b)为３×３、rate＝２的空洞卷积，卷积核的感受野增大到7×7。图2(c)为３×３、rate＝３的空洞卷积，卷积核的感受野增大到15×15，保证卷积网络能够提取到较大视野内的特征信息[２３]。



(a) (b) (c)

图2 空洞卷积 （a）rate=1；(b)rate=2; (c)rate=3

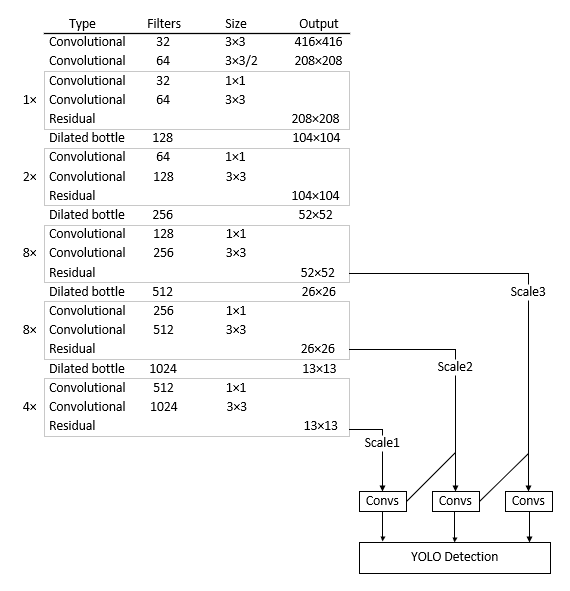


图3 改进后YOLOV3的骨干网络

空洞卷积的好处在于不做池化损失信息的情况下，加大了感受野，让每个卷积输出都包含较大范围的信息。该结构减少下采样的数量后进行特征提取，势必将带来深层网络的感受野减小的问题[２３]。故改进后的YOLOV3沿用FPN 结构，带１×１卷积结构的空洞卷积瓶颈层，如图4所示。这种复杂度较低的空洞卷积残差结构中，Conv为卷积层，Add为相加运算，ReLU 为激活函数。因使用大小为３×３，rate＝２的空洞卷积，从而整体上增大了骨干网络的感受野和特征表达能力[２４]。同时，空洞卷积残差结构仍具有残差单元的网络参数较少、计算复杂度较低的优点，使用１×１的卷积实现跨通道的特征融合，较好地整合了信息。

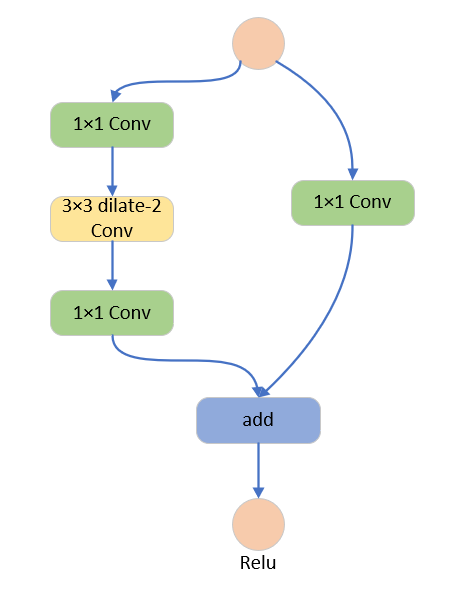


图4 带1×1卷积结构的空洞卷积瓶颈层

最终，改进的YOLOV3能够在深层网络中保持较高分辨率和较大的特征图的感受野，增强了YOLOV3算法对小目标的感受野和检测能力。

3、通道注意力机制

在改进的YOLOV3骨干网络中，虽然使用空洞卷积瓶颈层能够在深层网络中保持较高分辨率和较大的特征图的感受野，但是由于感受野区域广，牺牲了空间分辨率，致使深层网络的小目标信息损失，缺乏细节信息。虽然YOLOV3采用FPN，将高层特征与浅层特征信息融合，提高了小目标的检测效果，但是步长较大的高层导致小目标的语义特征丢失严重，该特征融合方法对精确定位飞机和检测较小的场面飞机有不利的影响。为了定位到感兴趣的信息，抑制无用信息，在骨干网络的最后，我们引入通道注意力模块，以挖掘更丰富的语义特征，提高检测精度。

我们利用特征的通道间关系生成通道注意图。由于每个通道的特征图被认为是一个特征探测器，通道的注意力主要关注在输入图像中什么是有意义的。为了有效地计算通道注意力，我们压缩了输入特征图的空间维数。平均池化是目前常用的一种信息聚合方法，它能够有效地学习目标对象的范围，在很多注意力模块中采用它来计算空间数据信息。此外，我们认为最大池化收集了另一个关于不同的物体判别性特征的重要线索，以推断出更精细的通道注意力。因此，我们同时使用平均池化和最大池化在空间维度压缩特征图。实验证实，相较于单独使用一种池化方法，同时利用这两种特性大大提高了网络的表示能力，这表明我们的设计选择是有效的。图5为通道注意力模快，下面描述了具体的操作。

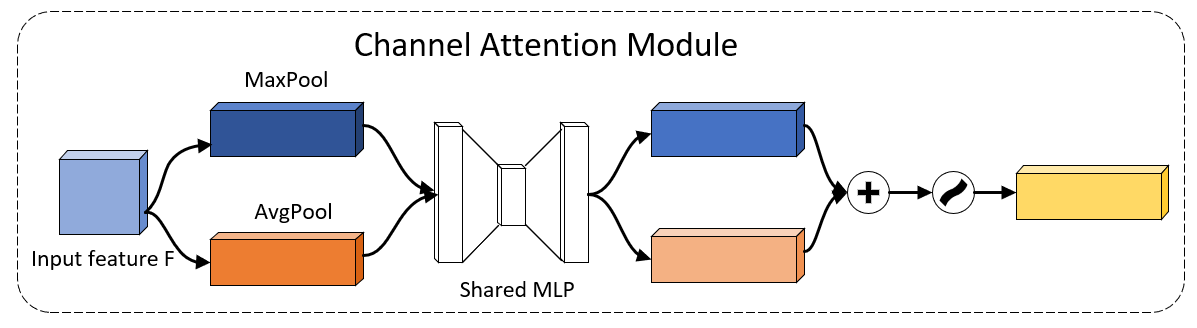


图5 通道注意力模块

我们首先使用平均池化和最大池化操作聚合特征图的空间信息，得到两个不同的空间背景描述符和，它们分别表示平均池化特征和最大池化特征。然后将两个描述符送到一个共享网络，以生成我们的通道注意图 。共享网络由多层感知器(MLP)和一个隐含层组成。为了减少参数开销，将隐藏激活大小设置为，其中r为约简比率。在将共享网络应用于每个描述符之后，我们使用元素求和来合并输出特征向量。简而言之，通道注意力计算公式为：

(1)

其中，指的是sigmoid函数，，。

4、线性衰减的NMS算法

绝大部分目标检测方法，最后都要用到NMS进行后处理，使用NMS去除重复框，降低误检率。NMS算法是将检测框按照得分进行排序，保留得分最高的检测框，同时删除与该框重叠面积大于阈值的其它框。NMS可用公式表示为

 （2）

其中，M为置信度得分较大的候选框，bi为被比较的物体预测框，I(M，bi)为M与bi的交并比，Nt为抑制阈值。这种贪心算法对物体的检测存在问题。如图6所示，A、B两架飞机，A飞机的预测框的置信得分为0.9，B飞机预测框的置信得分为0.6，B飞机严重遮挡A飞机，且A、B两架飞机预测框的IOU大于0.5。一方面：使用NMS算法处理多余预测框，当A、B飞机预测框的IOU大于0.5（YOLOV3设定阈值为0.5）时，保留置信度得分较高的B飞机，将A飞机的预测框的置信度得分置为0，从而造成无法检测出B飞机的结构。另一方面：NMS阈值也不太容易确定，设置过小会出现A飞机无法成功检测，设置过高会增大误检率。本文使用线性衰减的NMS算法，解决了YOLOV3中NMS造成遮挡飞机(或轮船)无法精确检测的问题。当交并比I(M, bi)大于抑制阈值Nt时，对公式(1)中的置信度得分Sconfi进行线性平滑处理，优化后的NMS算法表示为：

 （3）

线性衰减的NMS算法的具体步骤为：a.将回归产生的N个预测框按照置信度得分进行排序；b.选择得分最大的预测框，计算其与其它预测框的IOU；c.比较由步骤b计算得到的各个IOU与阈值之间的大小，如果大于阈值，使用公式(2)重新计算预测框得分，否则保持原值不变；d.将得分最大的预测框存入集合L中，执行N-1，剩余的预测框重复步骤a-d,直到N=1，即处理完所有的预测框；e.删除集合L中置信度得分小于阈值的预测框，显示得分大于阈值的预测框，即为最终检测框。

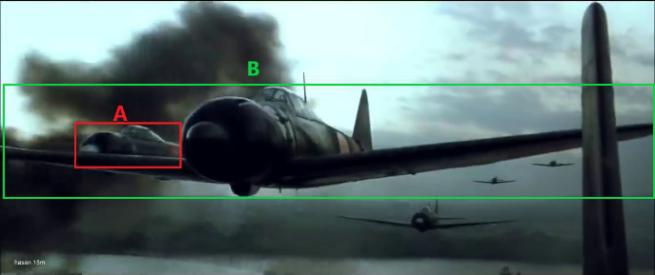
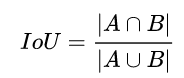


图6 相互遮挡的飞机

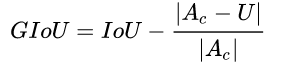
5、 优化IOU算法

IOU是交并比，是目标检测中最常用的指标，在基于anchor的方法中，它的作用不仅用来确定正样本和负样本，还可以用来评价预测框和真实框的距离。IOU可用公式表示为：

 （4）

其中，IOU为置信度，A和B为两个候选框，为A、B取交集，为A、B取并集。

IOU具有2个优点：一是直接反应预测框和真实框的检测效果，二是尺度不变性。然而检测任务中的BBox的回归损失(MSE loss, l1-smooth loss等）优化和IoU优化不是完全等价的。并且Ln范数对物体的scale也比较敏感。所以解决方法是直接把IOU作为Loss。但是当IOU作为Loss时，它无法直接优化A、B没有重叠的部分，此时Loss=0,没有梯度回传，无法进行学习训练。本文使用IOU优化算法GIOU，解决了YOLOV3中IOU造成的无法优化问题。GIOU的计算公式为：

 （5）

其中，Ac为两个框的最小闭包区域面积；U为A、B的并集。GIOU的特性如下：（1）与IOU相似，GIOU也是一种距离度量，作为Loss时，，满足损失函数的基本要求。（2）GIOU是IOU的下界，即GIOU(A,B)<=IOU(A,B)，在两个框无重合时，IOU=GIOU（3）IOU的取值是[0,1],但GIOU有对称区间，取值是[-1,1]。在两者重合时取最大值1，在两者无交集且无限远时取最小值-1，因此GIOU是一个非常好的距离度量指标（4）与IOU只关注重叠区域不同，GIOU不仅关注重叠区域，还关注其它的非重合区域，能更好的反应两者的重合度。

1. **有益效果**

1、实验设置

本次实验使用的数据集为自制数据集，采用人工进行标注，分为飞机和轮船两种类别。该数据集包含了各种型号的单架飞机(轮船)、相互遮挡的飞机(轮船)、小目标以及其它目标等复杂图像，共计21766张图片以及相应的XML文件。如图7所示。



图7 数据集示例

2、实验设置补充

实验使用Pytorch深度学习框架，在NVIDIA 1660显卡上完成训练及测试。

在本次实验中，我们将训练集、验证集和测试集的比例设置为8：1：1。使用YOLOV3的9个Anchor，分别为10x13、16x30、33x23、30x61、62x45、59x119、156x198、373x326。实验中使用Adam优化器，设置momentum=0.9，decay=0.0005，最大迭代次数为5000。在开始训练时learning rate = 10-3 以达到稳定整个网络的目的，迭代2000次之后调整为10-2，达到3000次之后调整为10-3。

3、实验性能比较

为了验证本方法能够较好的检测复杂场景下的飞机(轮船)，分别在不同的优化方法上进行实验。

(1)基于线性衰减的NMS算法检测

针对飞机(轮船)相互遮挡问题，本文采用了线性衰减的NMS算法。从测试集中选取了遮挡比例不同的图片进行测试。然后将提出的方法与现有成熟方法YOLOV3和Faster-RCNN进行了对比，由表1所示。由表1结果可知，遮挡比例在50%以下，三种方法的检测精度接近；当遮挡比例大于50%时，本文的方法比YOLOV3和Faster-RCNN的检测精度要高。具体可视化图如图8所示，我们可以观察到针对不同比例的遮挡，本方法的检测能力是极好的。

表1 不同检测方法在遮挡数据集下的实验效果

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Overlapped（%） | Recall（%） | Map（%） |
| YOLOV3 | <=50% | 87.11 | 86.02 |
| >50% | 74.01 | 73.21 |
| Faster-RCNN | <=50% | 85.21 | 84.56 |
| >50% | 70.26 | 70.34 |
| Our | <=50% | 91.24 | 88.04 |
| >50% | 90.08 | 87.56 |



图8 线性衰减的NMS可视化数据

（2）基于GIOU算法检测

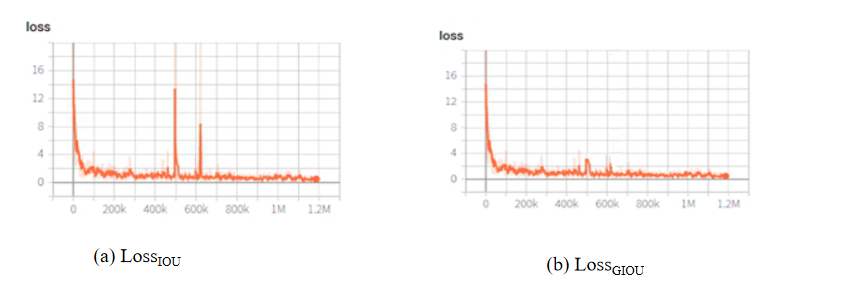
实验中采用GIOU作为损失函数，GIOU可以避免由于不重叠带来的LOSS=0而无法优化的问题，由图8所示，当epoch=200时，可以很明显的观测到GIOU的loss曲线比IOU的loss曲线更平滑。

图9 loss可视化对比

（3）基于注意力机制和空洞卷积检测

我们在YOLOV3的骨干网络中引入通道注意力模块，根据引入的位置和的个数不同，检测精度也有不同。通过对比多种策略，我们最终采用在每隔2个卷积操作加入一个注意力模块。除此之外，原YOLOV3是在基于stride=2的卷积操作进行下采样以次来增大感受野，而这种方法虽然增大了感受野，但是也会产生空间层级化信息丢失、小物体信息无法重建等问题，在本方法中我们使用带有rate的空洞卷积来增大感受野，经实验表明，rate=2时，效果更佳。得到的部分测试结果如下所示。其中，图10表示在多尺度目标存在的情况下，算法能够精确识别定位目标物体；图11表示算法对极小目标的检测能力；图12是在复杂背景下的检测效果。相较于原YOLOV3，算法不仅对小目标的检测精度明显提升，而且能够在复杂背景下更好地提取到目标信息，准确定位目标物体。数据指标如表2所示。



图11 小目标检测

表2 不同方法的检测精度对比

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | F1(%) | Precision(%) | Recall(%) | MAP(%) |
| YOLOV3 | 80.23 | 70.11 | 85.11 | 74.94 |
| Faster-RCNN | 80.58 | 70.02 | 83.25 | 74.26 |
| SSD | 79.23 | 69.82 | 80.21 | 72.67 |
| R-FCN | 80.14 | 70.03 | 83.78 | 73.92 |
| Ours | 92.01 | 88.24 | 93.89 | 89.04 |

针对小目标及遮挡飞机和轮船等难以检测的问题，将实时性较高的YOLOV3算法应用到复杂背景下的飞机和轮船检测上。我们在YOLOV3模型的基础上，对骨干网络、IOU算法和NMS算法分别进行了改进。实验表明本方法较大的改善了小目标、飞机(轮船)遮挡等复杂环境下的检测效果，将准确度从74.04%提高到89.04%。