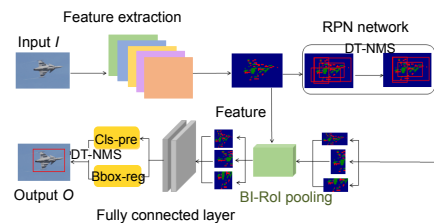




DOI: 10.12086/oe.2019.190159

基于双阈值-非极大值抑制的 Faster R-CNN 改进算法

侯志强^{1,2}, 刘晓义^{1,2*}, 余旺盛³, 马素刚^{1,2}¹西安邮电大学计算机学院, 陕西 西安 710121;²西安邮电大学陕西省网络数据分析与智能处理重点实验室,
陕西 西安 710121;³空军工程大学信息与导航学院, 陕西 西安 710077

摘要: 根据目标检测算法中出现的目标漏检和重复检测问题, 本文提出了一种基于双阈值-非极大值抑制的 Faster R-CNN 改进算法。算法首先利用深层卷积网络架构提取目标的多层卷积特征, 然后通过提出的双阈值-非极大值抑制 (DT-NMS) 算法在 RPN 阶段提取目标候选区域的深层信息, 最后使用了双线性插值方法来改进原 Roi pooling 层中的最近邻插值法, 使算法在检测数据集上对目标的定位更加准确。实验结果表明, DT-NMS 算法既有效地平衡了单阈值算法对目标漏检问题和目标误检问题的关系, 又针对性地减小了同一目标被多次检测的概率。与 soft-NMS 算法相比, 本文算法在 PASCAL VOC2007 上的重复检测率降低了 2.4%, 多次检测的目标错分率降低了 2%。与 Faster R-CNN 算法相比, 本文算法在 PASCAL VOC2007 上检测精度达到 74.7%, 性能提升了 1.5%。在 MSCOCO 数据集上性能提升了 1.4%。同时本文算法具有较快的检测速度, 达到 16 FPS。

关键词: 计算机视觉; 目标检测; 非极大值抑制; 卷积神经网络

中图分类号: TP391.4

文献标志码: A

引用格式: 侯志强, 刘晓义, 余旺盛, 等. 基于双阈值-非极大值抑制的 Faster R-CNN 改进算法[J]. 光电工程, 2019, 46(12): 190159

Improved algorithm of Faster R-CNN based on double threshold-non-maximum suppression

Hou Zhiqiang^{1,2}, Liu Xiaoyi^{1,2*}, Yu Wangsheng³, Ma Sugang^{1,2}¹College of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an, Shaanxi 710121, China;²Shaanxi Key Laboratory of Network Data Analysis and Intelligent Processing, Xi'an University of Posts and
Telecommunications, Xi'an, Shaanxi 710121, China;³Information and Navigation Institute of Air Force Engineering University, Xi'an, Shaanxi 710077, China

Abstract: According to the problems of target missed detection and repeated detection in the object detection algorithm, this paper proposes an improved Faster R-CNN algorithm based on dual threshold-non-maximum suppression. The algorithm first uses the deep convolutional network architecture to extract the multi-layer convolution features of the targets, and then proposes the dual threshold-non-maximum suppression (DT-NMS) algorithm in the

收稿日期: 2019-04-08; 收到修改稿日期: 2019-05-30

基金项目: 国家自然科学基金资助项目(61703423, 61473309); 西安邮电大学研究生创新基金资助项目(CXJJ2017019)

作者简介: 侯志强(1973-), 男, 博士, 教授, 博士生导师, 主要从事图像处理、计算机视觉和信息融合的研究。E-mail: hzq@xupt.edu.cn

通信作者: 刘晓义(1995-), 女, 硕士研究生, 主要从事计算机视觉, 目标检测, 图像处理的研究。E-mail: 18829290763@163.com

RPN(region proposal network). The phase extracts the deep information of the target candidate regions, and finally uses the bilinear interpolation method to improve the nearest neighbor interpolation method in the original RoI pooling layer, so that the algorithm can more accurately locate the target on the detection dataset. The experimental results show that the DT-NMS algorithm effectively balances the relationship between the single-threshold algorithm and the target missed detection problem, and reduces the probability of repeated detection. Compared with the soft-NMS algorithm, the repeated detection rate of the DT-NMS algorithm in PASCAL VOC2007 is reduced by 2.4%, and the target error rate of multiple detection is reduced by 2%. Compared with the Faster R-CNN algorithm, the detection accuracy of this algorithm on the PASCAL VOC2007 is 74.7%, the performance is improved by 1.5%, and the performance on the MSCOCO dataset is improved by 1.4%. At the same time, the algorithm has a fast detection speed, reaching 16 FPS.

Keywords: computer vision; object detection; non-maximum suppression; convolutional neural network

Citation: Hou Z Q, Liu X Y, Yu W S, et al. Improved algorithm of Faster R-CNN based on double threshold-non-maximum suppression[J]. *Opto-Electronic Engineering*, 2019, 46(12): 190159

1 引言

目标检测是计算机视觉领域重要的研究方向和研究热点,可应用于自动驾驶、机器人、视频监控、行人检测等多个领域^[1-6]。目标定位的精度问题一直受到人们极大的关注^[7-9]。

近几年来,基于深度学习的目标检测方法得到了很大的突破,其主要的可以分为两类。一类是一阶段检测方法,这类方法检测速度快,但检测精度略低。目前主要方法有 YOLO^[10]和 SSD^[11]等。另一类是两阶段检测方法,首先使用区域推荐产生候选目标,随后使用卷积神经网络进行处理。此类方法有 R-CNN^[12], SPP-NET^[13]和 Fast R-CNN^[14]等。由于 Fast R-CNN 检测处理一幅图像需要很长时间,因此,Ren 等又提出 Faster R-CNN^[15]算法。

传统的 Faster R-CNN 算法使用非极大值抑制算法 NMS 进行候选框滤除。它采用“非一即零”的思想,只保留分类目标得分最高的候选框,阈值设定较大,这在很大程度上加大了目标在高度重叠时被漏检的风险。因此,soft-NMS 算法^[16]选择一个小的阈值,并采用“权值惩罚”策略去解决这个问题,在一定程度上减少了目标漏检。但是测试发现,soft-NMS 算法的使用会大大增加候选框的数量,从而出现同一目标被重复检测和多次检测出现错分目标的新问题,尤其是当图像中存在多个目标且目标的重叠度很高的时候。这是因为,NMS 算法和 soft-NMS 算法都属于单阈值-非极大值抑制算法。阈值通过实验结果进行设定,阈值的选择非常重要,阈值设置过大,容易出现目标漏检问题,阈值设置过小,又容易出现目标误检问题。单阈值的选择很难平衡好这两者之间的关系。

针对以上问题,本文的主要工作有:1) 为了解决单阈值 NMS 算法中出现的目标漏检和目标重复检测且出现目标错分的问题,提出了基于双阈值的非极大值抑制 DT-NMS(dual threshold-non-maximum suppression, DT-NMS)算法。为了突出本文算法在目标重复检测问题上的表现,提出了重复检测率和多次检测的目标错分率作为衡量指标。最终,通过对阈值的合理设定,既有效地平衡了单阈值算法对目标漏检问题和目标误检问题的关系,又针对性地减小了同一目标被多次检测的概率。2) 针对目标定位不够准确的问题,采用双线性插值法去替换原 RoI pooling 层中的最邻近插值法,使算法无论是在单个目标还是多个目标的定位性能上都有所改善。3) 改进后的 Faster R-CNN 算法在 VGG-Net-16 的网络结构上重新进行了网络训练和参数的调整。并在 PASCAL VOC2007 数据集和 MSCOCO 数据集上进行了大量实验验证,通过与多种近年来性能优越的主流算法进行了对比分析,实验结果表明,本文算法在检测精度和速度上都具有很好的表现。

2 基本原理

2.1 Faster R-CNN 算法

CNN 和 Fast R-CNN 是两阶段目标检测方法中的典型算法,但检测速度慢。主要原因是对所有区域进行特征提取时存在大量的重复计算。并且,选择性搜索(selective search, SS)等传统候选框生成方法在 CPU 上占用了大量时间。为了解决这个问题,Ren 等提出了 Faster R-CNN 算法。Faster R-CNN 算法将 VGG-Net-16 作为基础网络提取深度特征。针对 Fast

R-CNN, 算法作出的改进是:

- 1) 使用深度网络 RPN 代替传统的选择性搜索等方法生成候选区域;
- 2) 实现了 RPN 网络和检测网络的特征共享。

2.2 RPN 网络

RPN 是 Faster R-CNN 算法中的一个重要改进,它输入的是任意一幅图像,输出的是一组经过打分的候选框。RPN 网络首先利用“滑动窗口”的思想代替传统的选择性搜索等方法去生成候选区域,在每个滑动窗口处生成 $K(K=9)$ 个大小不同的锚边框(anchor boxes),以满足不同尺寸目标的需要。其次将这些锚边框送入两个全卷积层分类得分层和边框回归层中,分别预测出 $2 \times K$ 个得分和 $4 \times K$ 个坐标。最后经过打分,挑选出得分排在前 300 的窗口作为建议窗口进行下一步的运算。

3 基于 Faster R-CNN 算法的改进

3.1 改进的非极大值抑制算法

目前,大部分的目标检测方法都使用式(1)对候选框进行非极大值抑制算法(non-maximum suppression, NMS)处理。

$$s_f = \begin{cases} s_i, & U_{IoU}(M, b_i) < N_t \\ 0, & U_{IoU}(M, b_i) \geq N_t \end{cases}, \quad (1)$$

其中: s_i 为该候选框的原始得分, s_f 为该候选框的最后得分, M 为得分最高的候选框, b_i 为待检测框, $U_{IoU}(M, b_i)$ 为待检测框 b_i 和 M 的交并比(intersection over union, IoU), N_t 为预设阈值。

但是,当图像中出现两个距离十分相近的目标时,其中分数较低的框很有可能因为与当前得分最高的框的交并比大于预设阈值而被直接抑制,从而出现目标漏检问题。如图 1(a)所示,当图中的两匹马距离十分接近时,NMS 方法只保留得分最高的候选框,将两匹马看作是同一目标,从而出现了目标的漏检问题。

soft-NMS 算法采用“权值惩罚”策略很好地解决了这个问题,具体方法:

$$s_f = \begin{cases} s_i, & U_{IoU}(M, b_i) < N_t \\ s_i \times (1 - U_{IoU}(M, b_i)), & U_{IoU}(M, b_i) \geq N_t \end{cases}. \quad (2)$$

式中符号同式(1)。如图 1(b)所示,soft-NMS 算法使用不同的候选框分别框住了两匹马,且类别预测得分高于 NMS 算法。说明 soft-NMS 算法在一定程度上解决了 NMS 算法的漏检问题。但随着候选框的数量增加,同一目标被重复检测和多次检测出现错分目标的概率也随之增加。如图 2(a)所示,soft-NMS 算法在汽车上出现了重复检测,且在汽车上两次检测的类别不同,出现了类别分错的情况。

综上,本文提出了基于 DT-NMS 的解决方法:首先将检测框按得分排序,以阈值 N_t 的范围采用



图 1 NMS 算法和 soft-NMS 算法的结果对比。(a) NMS 算法; (b) soft-NMS 算法

Fig. 1 Comparison of results obtained by the NMS algorithm and soft-NMS algorithm. (a) NMS algorithm; (b) soft-NMS algorithm

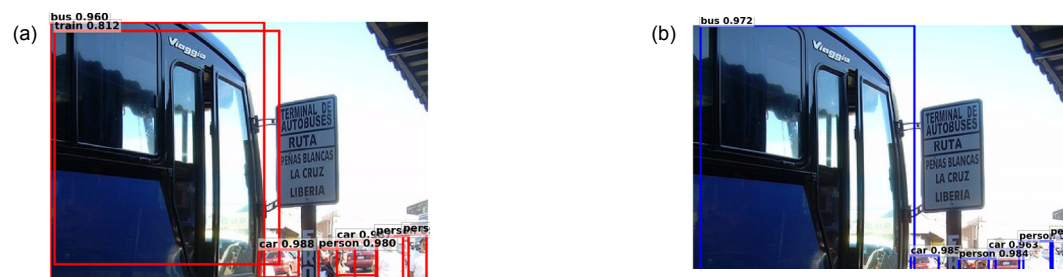


图 2 soft-NMS 算法和本文算法的结果对比。(a) soft-NMS 算法; (b) 本文算法

Fig. 2 Comparison of the results obtained by the soft-NMS algorithm and our proposed algorithm. (a) soft-NMS algorithm; (b) Ours algorithm

权值惩罚策略,对于当前得分最高的框 M ,给每一个待检测框 b_i 一个相应的权值,该权值的大小与 b_i 和 M 的 $U_{IoU}(M, b_i)$ 成反比。每一个待检测框的最后得分由权值和原始得分的乘积决定。这样做,大大降低了目标在高度重叠时被漏检的概率。

其次,根据另一个预设阈值 N_i ,

$$s_f = \begin{cases} s_i & U_{IoU}(M, b_i) < N_t \\ s_i \times (1 - U_{IoU}(M, b_i)) & N_t \leq U_{IoU}(M, b_i) \leq N_i \\ 0 & U_{IoU}(M, b_i) > N_i \end{cases}, (3)$$

式中: s_i 为该检测框的原始得分, s_f 为该检测框的最后得分, M 为得分最高的候选框, b_i 为待检测框, $U_{IoU}(M, b_i)$ 为当前检测框 b_i 和 M 的交并比, N_t 和 N_i 为阈值。

由式(3),删除部分与 M 的 $U_{IoU}(M, b_i)$ 过大的待检测框,这是因为, $U_{IoU}(M, b_i)$ 过大时, M 与 b_i 为同一目标的可能性非常大,这些框会加大目标被重复检测的概率。

最终,我们选取出得分在前 300 的窗口作为建议窗口用于之后检测阶段。也就是说,本文提出的非极大值抑制方法在降低漏检率的同时对候选框进行再一次过滤,通过双阈值的调节,选择了一个合适的 N_t ,既可以很好地平衡 NMS 算法出现的目标漏检和目标误检问题的关系,同时, N_i 的加入使得 $(N_i, 1)$ 区间内 soft-NMS 算法中出现的目标重复检测和目标错分问题很好地被解决。如图 2(b),本文算法很好地解决了汽车的重复检测和错分目标问题,且在目标的分类得分上也高于 soft-NMS 算法。

3.2 改进的 RoI pooling 模块

Faster R-CNN 算法在 RPN 网络中会产生大小不同的候选框,并在 RoI pooling 层中采用最近邻插值法将这些框映射为固定大小。对于缩放后坐标不能刚好为

整数的候选框,最近邻插值法直接选择离目标点最近的整数点坐标去代替原始坐标,然后进行最大池化操作。但是这样做会使目标的回归位置与目标的真实位置之间误差变大,从而影响最终的目标检测率。

为了解决这个问题,同时受到 Mask R-CNN^[17] 算法的启发,本文在 RoI pooling 层使用双线性插值法去替换原始的最近邻插值法。与 Mask R-CNN 算法不同的是:1) 算法所应用的网络结构不同,Mask R-CNN 算法使用 Res-Net 作为基础网络,而本文则使用 VGG-Net-16 的网络结构。因此,重新进行了网络的训练和参数的调整。2) 使用算法的应用目的不同,前者对目标进行分割,本文则对 RoI pooling 层进行改进来提升目标检测的性能。

图 3 显示了原算法与本文改进算法的区别。图 3(a)为 RoI pooling 的量化过程。输入图像经过 RPN 网络会得到特征图上的候选区域信息(proposals 中的一个红框表示一个候选区域),再将所有候选区域信息送入 RoI pooling 层。图中以其中一个候选区域的操作为例,其余的候选区域操作与之类似,图中的绿色矩形表示了该候选区域。在 RoI pooling 层中,首先对候选区域的顶点坐标进行向下取整,然后将候选区域分成 7×7 个单元,每个单元四个顶点坐标再次进行向下取整,图中的粉色格子表示了这些单元。取整之后,每个单元中都有 $n \times m$ 个完整的像素,图中以 2×2 的大小为例。最后, RoI pooling 层在每个单元中选择像素的最大值,进行最大池化。这样做使候选区域中有一部分像素值被丢弃,只利用这些像素去做目标位置的回归运算,即使最后得到的目标预测位置为整数,但是依然不准确,它和目标的真实位置存在很大偏差,而这个偏差最终会影响检测的准确度。

图 3(b)为本文使用的 BI-RoI pooling 方法。在得到

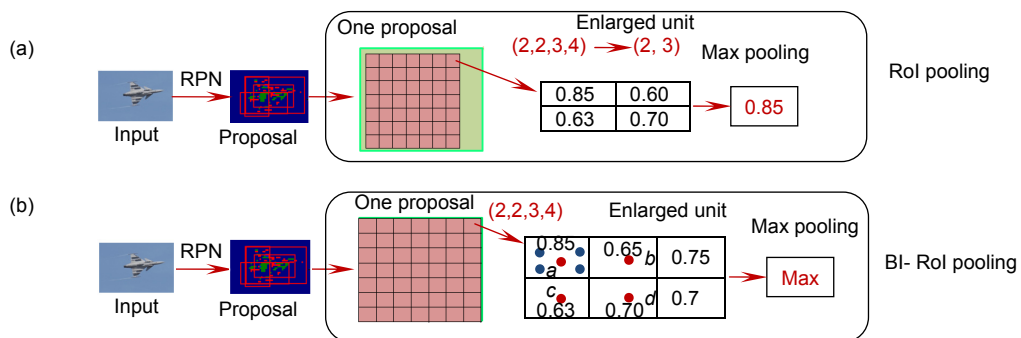


图 3 RoI pooling 和 BI-RoI pooling 方法对比。(a) RoI pooling 的量化过程; (b) BI-RoI pooling 的量化过程
Fig. 3 Comparison of RoI pooling and BI-RoI pooling methods. (a) Quantization process of RoI pooling; (b) Quantization process of BI-RoI pooling

特征图的候选区域信息后,与 RoI pooling 方法不同,本文取消了 RoI pooling 中的取整操作,首先将候选区域分成 7×7 个单元,使候选区域和每一个单元的浮点坐标保持不变,每个单元中像素点不再恰好是 $n \times m$ 个。然后将每个单元平均分成 4 等份,取每一等份的中心点作为采样点,假设采样点分别为 a, b, c, d 点。

$$\begin{aligned}
 f(i+u, j+v) = & (1-u) \times (1-v) \times f(i, j) \\
 & + (1-u) \times v \times f(i, j+1) \\
 & + u \times (1-v) \times f(i+1, j) \\
 & + u \times v \times f(i+1, j+1), \quad (4)
 \end{aligned}$$

根据式(4)计算这四个点处的像素值,最后取其最大值,进行池化操作。以图 3(b)中 a 点为例,假设 a 点坐标为 $(i+u, j+v)$,其中 i, j 均为非负整数, u, v 为 $[0, 1]$ 区间的浮点数, a 点附近的 4 个蓝色圆点分别表示离 a 点最近且坐标为整数的点,它们的坐标分别为 $(i, j), (i+1, j), (i, j+1), (i+1, j+1)$ 。 $f(i, j)$ 表示坐标为 (i, j) 的点的像素值。其他三个点 b, c, d 处的像素值计算类似。

改进后的算法通过在候选区域中选择采样点的方

式来计算图像对应位置的像素值,充分地使用了整个候选区域中的全部像素信息。在此基础上进行目标位置回归更加准确,目标框的预测位置也就更加接近其真实位置。

3.3 本文算法的检测框架

综合上述对本文算法关键部分的描述,本文的具体检测步骤如算法 1 所示,算法流程图如图 4 所示。

4 分析与讨论

本文实验使用的数据集为 PASCAL VOC2007, PASCAL VOC2012 和 MSCOCO,并分别在对应的测试集上进行了实验结果测试,使用平均准确度(mean average precision, mAP)作为性能评测标准。本文算法在处理器为 Intel(R) Core i5 8400CPU 和 GTX1080GPU 的计算机上运行,实验平台为基于 Faster R-CNN 的 Caffe 版本。

4.1 定性分析

为验证本文算法的检测性能,将本文算法的检测

算法 1 一种基于双阈值-非极大值抑制的 Faster R-CNN 改进算法

Input: Image I

Output: Image O with detected object

Step one: Extracting features through a deep convolutional network to obtain features map of the input image

Step two: Use the DT-NMS algorithm to obtain proposals information in the RPN network

Step three: Map the deep information of the proposals information back to the features map

Step four: After using the BI-RoI pooling method, the proposals information of different sizes are fixed to a uniform size

Step five: Detected image obtained by two fully connected layers

Algorithm over

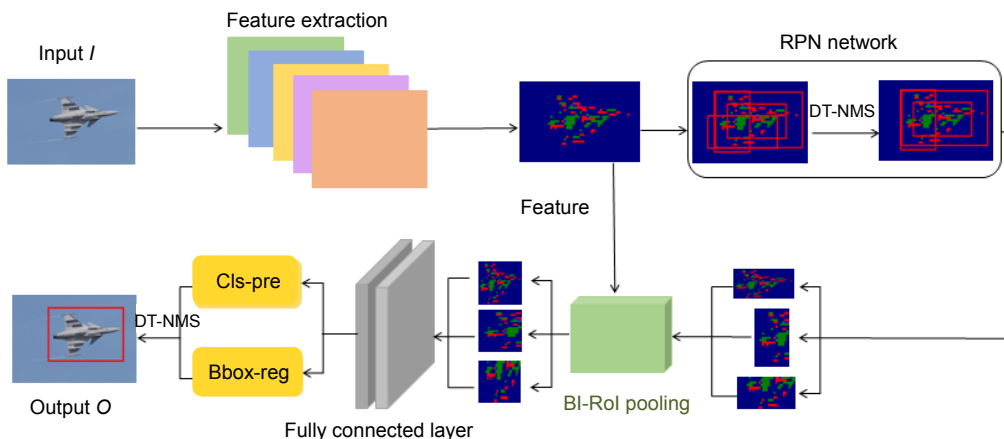


图 4 本文算法框架

Fig. 4 Algorithm framework of this paper

结果与 Faster R-CNN 算法, soft-NMS 算法的检测结果进行比较。从中选取了部分代表性的结果如下。

图 5 为 PASCAL VOC 数据集上的检测结果。由图 5 可以看出, 不论是单目标图像还是多目标图像, 本文提出的算法都比 Faster R-CNN 算法好。具体表现为

- 1) 能够检测出更多的目标;
- 2) 对同样识别出的目标, 其得分更高;
- 3) 对高度重叠的目标, 其准确度更高, 修正了 Faster R-CNN 算法检测时出现的漏检问题。

同时, 对比 soft-NMS 算法而言, 本文算法在目标重复检测问题上也表现得更好, 从而使得目标的定位也更加准确。比如图 5 中第 4 行重复检测到的车辆, 第 5 行重复检测到的人, 本文算法都给予了很好的修正。

本文算法在 MSCOCO 数据集上的检测结果也很突出。由图 6 可以看出, 本文算法在目标框的定位上明显优于其他对比算法, 且得分也很高。图 7 的结果

显示, 通过使用 DT-NMS 算法, 很好地解决了 soft-NMS 中重复检测和错分目标的问题。在多目标重叠时效果尤其明显。

本文算法在以上两个数据集上都有明显的提升, 说明其具有普适性。

4.2 定量分析

4.2.1 改进的非极大值抑制模块

双阈值的选择可以很容易地通过实验结果确定出来。在本文中, 我们采用联合调参法来选择最终阈值。阈值改变过程中, 保持网络的其他结构不变, 以算法的平均准确度作为判断条件, 实验如表 1 所示。

本文选择 $N_1=0.35$, $N_2=0.95$ 作为最终的阈值, 此时算法的平均准确度最高为 74.72%。

为了突出本文算法在目标重复检测问题上的表现, 提出了重复检测率和多次检测的目标错分率作为衡量指标, 其计算式分别为

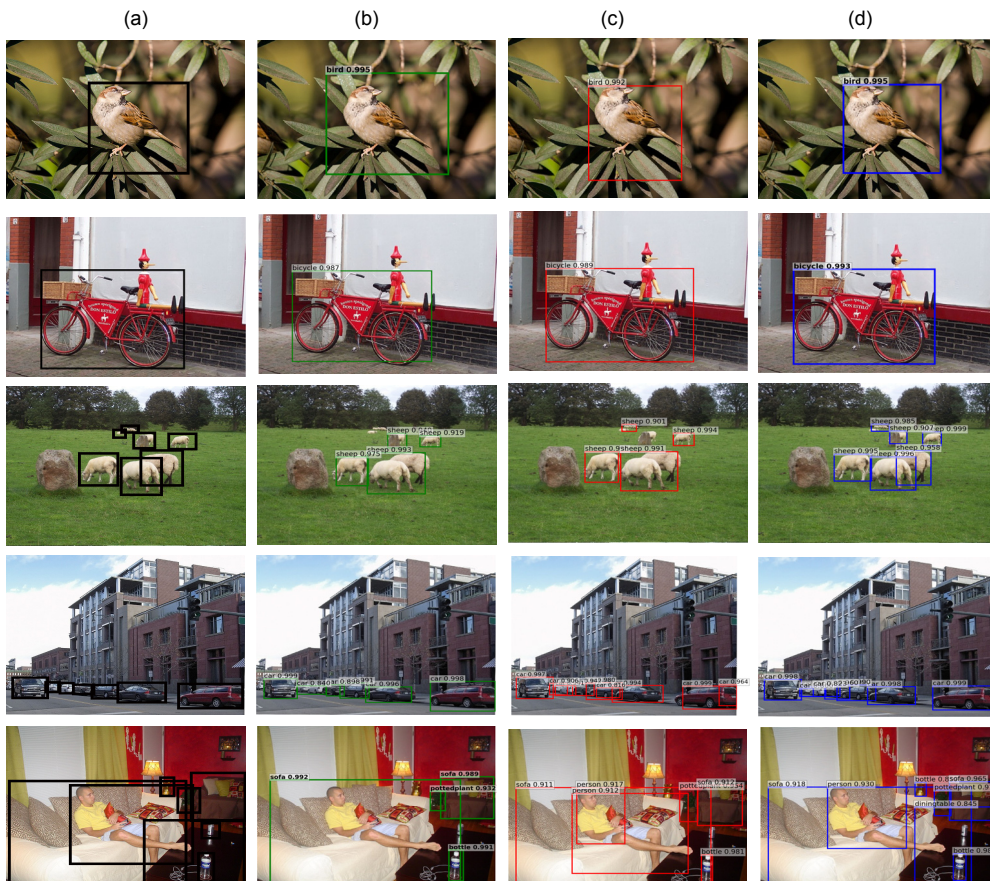


图 5 本文算法与其他算法在 PASCAL VOC 上的实验结果对比。(a) Ground-truth; (b) Faster R-CNN; (c) soft-NMS; (d) 本文算法

Fig. 5 Comparison of experimental results obtained by the proposed algorithm and other algorithms on PASCAL VOC. (a) Ground-truth; (b) Faster R-CNN; (c) soft-NMS; (d) Our algorithm

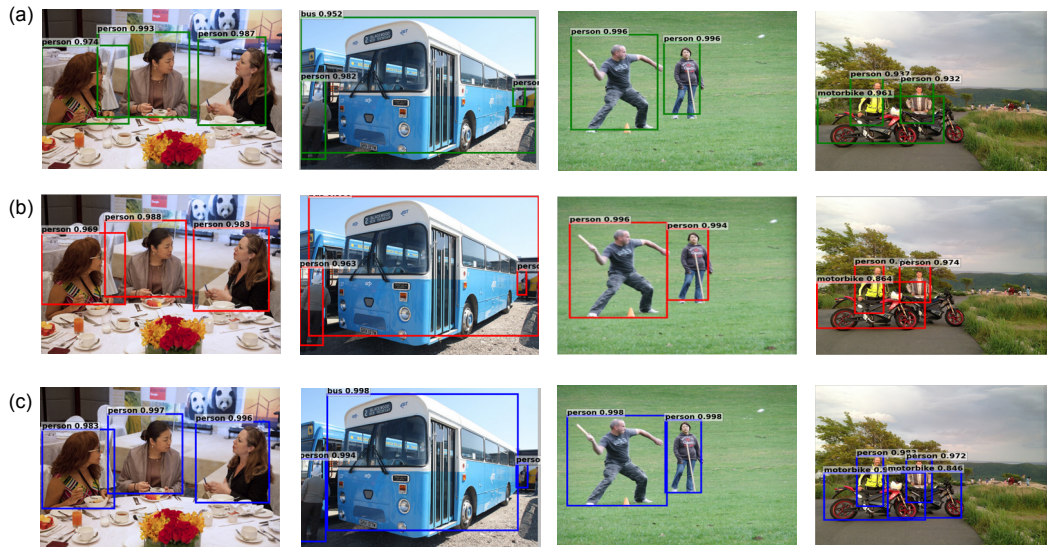


图 6 本文算法与其他算法在 MSCOCO 数据集上的目标定位实验结果对比。(a) Faster R-CNN; (b) soft-NMS; (c) 本文算法

Fig. 6 Comparison of target positioning experiments obtained by the proposed algorithm and other algorithms on the MSCOCO dataset. (a) Faster R-CNN; (b) soft-NMS; (c) Our algorithm



图 7 本文算法与其他算法在 MSCOCO 数据集上的重复检测实验结果对比。(a) Faster R-CNN; (b) soft-NMS; (c) 本文算法

Fig. 7 Comparison of repeated detection experiments among the proposed algorithm and other algorithms on the MSCOCO dataset. (a) Faster R-CNN; (b) soft-NMS; (c) Our algorithm

表 1 不同阈值组合下的算法的准确度对比

Table 1 Comparison of algorithm precision under different threshold combinations

N_i	N_t				
	0.3	0.35	0.4	0.45	0.5
0.8	73.95	74.05	74.01	74.10	74.19
0.85	74.15	74.20	74.18	74.17	74.10
0.9	74.43	74.50	74.05	74.30	74.36
0.95	74.62	74.72	74.66	74.58	74.44

$$R_{rate} = \frac{R_{total}}{N_{pictotal}} \times 100\% , \quad (5)$$

$$C_{error-rate} = \frac{C_{error}}{R_{total}} \times 100\% , \quad (6)$$

其中： R_{total} 表示出现重复检测的总数， R_2 和 R_3 分别为同一目标出现重复检测 2 次和 3 次的次数， C_{error} 表示重复检测且每次分类不同的次数， R_{rate} 和 $C_{error-rate}$ 为重复检测率和多次检测的目标错分率， $N_{pictotal}$ 表示测试图片的总数。

我们在 PASCAL VOC2007 数据集上对 4952 张图片进行了逐一测试，如表 2 所示，相比于 soft-NMS 算法，本文的重复检测率降低 2.4%，多次检测的错分率降低了 2.2%，整体结果好于 soft-NMS 算法。

4.2.2 改进的 RoI pooling 模块

改进后的 Faster R-CNN 算法分别在 VGG-Net-16 和 AlexNet 的网络结构上重新进行了网络训练和参数的调整。

比较表 3 中的第 1~3 行可以发现，使用本文所提出的 DT-NMS 算法，比 Faster R-CNN 算法的 mAP 提升 1.1%，速度也比 Faster R-CNN 算法稍快一些。同时，能够在平均准确度 mAP 上与 soft-NMS 算法基本相同。结合表 2 可以看出，本文算法的重复检测率远低于 soft-NMS 算法，说明 DT-NMS 算法在减小同一目标的重复检测这一问题上表现得更好。

比较表 3 中的第 3、4 行可以发现，在使用 DT-NMS

算法的基础上使用双线性插值法对 RoI pooling 层进行改进，mAP 仍有 0.4% 的提升，且对算法的检测速度并没有影响。说明对 RoI pooling 层的改进起到了一定的效果。

改进算法在 AlexNet 网络上的性能提升更为明显。表 3 第 5~8 行可以看出，相比于 Faster R-CNN 算法，本文算法的 mAP 达到了 63.7%，性能提升了 3.8%。比 soft-NMS 算法提升了 0.5%。且速度几乎没有影响。

综上所述，本文算法能够在降低目标的重复检测率的同时，使目标框的定位也更加准确。有效地提升了 Faster R-CNN 算法的检测性能。

4.2.3 算法综合性能的定量分析

为了进一步分析本文算法的检测性能，表 4 和表 5 显示了本文算法和其他算法模型的检测结果对比。由表 4 可以看到，在 PASCAL VOC 数据集的 20 类检测目标上，本文算法在鸟(bird)，瓶子(bottle)，猫(cat)等多个类别的检测精度优于 Fast RCNN，Faster R-CNN，算法的平均检测精度为 74.7%，比同一系列的 A-Fast-RCNN^[18]算法高 3.3%，比 RON320^[19]算法还高 0.5%。同时，由表 5 可以看到，与相同网络结构下的一阶段检测算法相比，本文算法的检测精度也有优势，高于 SSD 算法 0.4%，高于 Yolov2 算法 1%。

改进算法在 MSCOCO 数据集上的检测结果也很明显。由表 6 可以看到，相比于 Faster R-CNN 算法，

表 2 本文算法和 soft-NMS 在重复检测问题上的结果对比

Table 2 Comparison of the results obtained by the proposed algorithm and soft-NMS on the repeated detection problem

Algorithm	R_{total}	R_2	R_3	C_{error}	$R_{rate}/\%$	$C_{error-rate}/\%$
soft-NMS	592	543	49	64	11.9	10.8
Ours	471	441	30	38	9.5	8.6

表 3 本文算法中各模块的检测精度和速度

Table 3 Detection accuracy and speed of each module in the algorithm

Algorithm	mAP	Detection speed/(FPS)	Threshold	Network
Faster R-CNN	73.2	15.3	$N_t=0.6$	VGG16
Faster+soft-NMS	74.3	15.1	$N_t=0.3$	VGG16
Faster+DT-NMS	74.3	16.0	$(N_t, N_i)=(0.35,0.95)$	VGG16
Faster+DT-NMS+BI-RoI pooling	74.7	16.0	$(N_t, N_i)=(0.35,0.95)$	VGG16
Faster R-CNN	59.9	15.8	$N_t=0.6$	AlexNet
Faster +soft-NMS	63.2	15.6	$N_t=0.3$	AlexNet
Faster+DT-NMS	63.4	15.6	$(N_t, N_i)=(0.35,0.95)$	AlexNet
Faster+DT-NMS+BI-RoI pooling	63.7	15.6	$(N_t, N_i)=(0.35,0.95)$	AlexNet

表 4 本文算法和其他算法在 PASAL VOC2007 数据集上的检测结果

Table 4 Test results of the proposed algorithm and other algorithms on the PASAL VOC2007 dataset

Algorithm	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table
Fast R-CNN	77.0	78.1	69.3	59.4	38.3	81.6	78.6	86.7	42.8	78.8	68.9
A-Fast-RCNN	75.7	83.6	68.4	58.0	44.7	81.9	80.4	86.3	53.7	76.1	72.5
Faster R-CNN	76.5	79.0	70.9	65.5	52.1	83.1	84.7	86.4	52.0	81.9	65.7
RON320	75.7	79.4	74.8	66.1	53.2	83.7	83.6	85.8	55.8	79.5	69.5
soft-NMS	76.9	81.3	74.8	64.9	60.2	81.8	86.2	85.6	55.8	80.0	67.5
Ours	76.0	81.0	74.8	64.6	62.5	81.6	85.7	87.5	56.8	81.2	68.2

Algorithm	dog	horse	motor	person	plant	sheep	sofa	train	tv	mAP
Fast R-CNN	84.7	82.0	76.6	69.9	31.8	70.1	74.8	80.4	70.4	70.0
A-Fast-RCNN	82.6	83.9	77.1	73.1	38.1	70.0	69.7	78.8	73.1	71.4
Faster R-CNN	84.8	84.6	77.5	76.7	38.8	73.6	73.9	83.0	72.6	73.2
RON320	84.5	81.7	83.1	76.1	49.2	73.8	75.2	80.3	72.5	74.2
soft-NMS	82.8	82.3	79.6	81.0	43.4	77.1	71.7	79.8	73.7	74.3
Ours	84.3	82.9	78.9	81.0	43.8	78.3	70.0	80.9	73.9	74.7

表 5 本文算法和一阶段检测算法在 PASAL VOC2007 数据集上的检测结果

Table 5 Test results of the proposed algorithm and one-stage detection algorithm on the PASAL VOC2007 dataset

Algorithm	Network	mAP
YoLo	VGG16	63.4
Faster R-CNN	VGG16	73.2
YoLov2	VGG16	73.7
SSD	VGG16	74.3
Faster+soft-NMS	VGG16	74.3
Ours	VGG16	74.7

表 6 本文算法和其他算法在 MSCOCO 数据集上的检测结果

Table 6 Test results of the proposed algorithm and other algorithms on the MSCOCO dataset

Algorithm	Network	AP(0.5:0.95)	AP@0.5
YoLov2	VGG16	21.6	44.0
SSD	VGG16	23.2	41.2
Faster R-CNN	VGG16	24.4	45.7
Faster+soft-NMS	VGG16	25.5	46.7
Ours	VGG16	25.8	47.1
YoLov3	Darknet	33.0	57.9
RetinaNet	ResNet101	34.4	53.1

本文算法的检测精度提升了 1.4%。同时，与 soft-NMS 的算法相比，在相同网络结构中本文算法仍然具有一定的优势，比 SSD 算法的检测精度高 2.6%，比 YoLov2^[20] 算法高 4.2%。由于 RetinaNet^[21] 算法和 YoLov3^[22] 算法分别使用更深的 ResNet-101 网络和 DarkNet 网络进行检测，故检测结果略好于本文算法。

5 结论

本文提出了一种改进的基于双阈值-非极大值抑制的 Faster R-CNN 算法，能够很好地平衡单阈值-非

极大值抑制算法所带来的漏检和误检问题。且在解决目标被重复检测这一问题上表现得更好。相比于 soft-NMS 算法，本文的重复检测率降低了 2.4%，多次检测的目标错分率降低了 2%。同时，通过对 RoI pooling 层的改善，使算法在目标定位上更加准确，在 PASCAL VOC2007 数据集和 MSCOCO 数据集上检测精度都较 Faster R-CNN 算法有所提升。人为设定阈值的非极大值抑制算法存在一定的局限，下一步工作中，我们将考虑把自适应阈值的非极大值抑制算法引入新的基础网络，如 ResNet^[23] 和 DenseNet^[24] 以获得更好的

目标检测性能。

参考文献

- [1] Borji A, Cheng M M, Jiang H Z, *et al.* Saliient object detection: a benchmark[J]. *IEEE Transactions on Image Processing*, 2015, **24**(12): 5706–5722.
- [2] Luo H B, Xu L Y, Hui B, *et al.* Status and prospect of target tracking based on deep learning[J]. *Infrared and Laser Engineering*, 2017, **46**(5): 0502002.
罗海波, 许凌云, 惠斌, 等. 基于深度学习的目标跟踪方法研究现状与展望[J]. *红外与激光工程*, 2017, **46**(5): 0502002.
- [3] Hou Z Q, Han C Z. A survey of visual tracking[J]. *Acta Automatica Sinica*, 2006, **32**(4): 603–617.
侯志强, 韩崇昭. 视觉跟踪技术综述[J]. *自动化学报*, 2006, **32**(4): 603–617.
- [4] Xin P, Xu Y L, Tang H, *et al.* Fast airplane detection based on multi-layer feature fusion of fully convolutional networks[J]. *Acta Optica Sinica*, 2018, **38**(3): 0315003.
辛鹏, 许悦雷, 唐红, 等. 全卷积网络多层特征融合的飞机快速检测[J]. *光学学报*, 2018, **38**(3): 0315003.
- [5] 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.
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.
- [6] Wang S M, Han L L. Moving object detection under complex dynamic background[J]. *Opto-Electronic Engineering*, 2018, **45**(10): 180008.
王思明, 韩乐乐. 复杂动态背景下的运动目标检测[J]. *光电工程*, 2018, **45**(10): 180008.
- [7] Zhou X Y, Liu J, Lu X, *et al.* A method for pedestrian detection by combining textual and visual information[J]. *Acta Electronica Sinica*, 2017, **45**(1): 140–146.
周炫余, 刘娟, 卢笑, 等. 一种联合文本和图像信息的行人检测方法[J]. *电子学报*, 2017, **45**(1): 140–146.
- [8] Cao M W, Yu Y. Moving object detection based on multi-layer background model[J]. *Acta Electronica Sinica*, 2016, **44**(9): 2126–2133.
曹明伟, 余焯. 基于多层背景模型的运动目标检测[J]. *电子学报*, 2016, **44**(9): 2126–2133.
- [9] Zhang Z S, Qiao S Y, Xie C H, *et al.* Single-shot object detection with enriched semantics[C]//*Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 2018: 5813–5821.
- [10] Redmon J, Divvala S, Girshick R, *et al.* You only look once: unified, real-time object detection[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 2016: 779–788.
- [11] Liu W, Anguelov D, Erhan D, *et al.* SSD: single shot MultiBox detector[EB/OL]. (2016-12-29) [2019-05-28]. arXiv: 1512.02325v1. <https://arxiv.org/abs/1512.02325v1>.
- [12] Girshick R, Donahue J, Darrell T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, USA, 2014: 580–587.
- [13] He K M, Zhang X Y, Ren S Q, *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.
- [14] Girshick R. Fast R-CNN[C]//*Proceedings of IEEE International Conference on Computer Vision*, Santiago, Chile, 2015: 1440–1448.
- [15] Ren S, He K M, Girshick R, *et al.* Faster R-CNN: towards real-time object detection with region proposal networks[EB/OL]. (2015-06-04)[2019-05-28]. arXiv: 1506.01497. https://arxiv.org/abs/1506.01497?source=post_page.
- [16] Bodla N, Singh B, Chellappa R, *et al.* Soft-NMS – improving object detection with one line of code[C]//*Proceedings of IEEE International Conference on Computer Vision*, Venice, Italy, 2017: 5562–5570.
- [17] He K M, Gkioxari G, Dollár P, *et al.* Mask R-CNN[C]// *Proceedings of IEEE International Conference on Computer Vision*, Venice, Italy, 2017: 2980–2988.
- [18] Wang X L, Shrivastava A, Gupta A. A-Fast-RCNN: hard positive generation via adversary for object detection[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, USA, 2017: 3039–3048.
- [19] Kong T, Sun F C, Yao A B, *et al.* RON: reverse connection with objectness prior networks for object detection[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, USA, 2017: 5244–5252.
- [20] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, USA, 2017: 6517–6525.
- [21] Lin T Y, Goyal P, Girshick R, *et al.* Focal loss for dense object detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, doi: 10.1109/TPAMI.2018.2858826.
- [22] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2019-05-28]. arXiv: 1804.02767. <https://arxiv.org/abs/1804.02767>.
- [23] He K M, Zhang X Y, Ren S Q, *et al.* Deep residual learning for image recognition[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, USA, 2016: 770–778.
- [24] Huang G, Liu Z, van der Maaten L, *et al.* Densely connected convolutional networks[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, USA, 2017: 2261–2269.

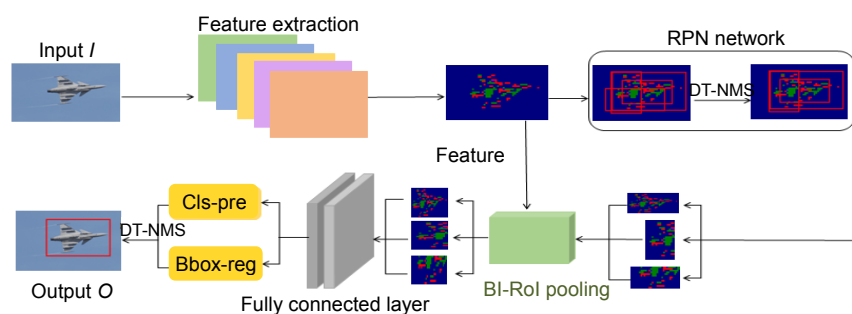
Improved algorithm of Faster R-CNN based on double threshold-non-maximum suppression

Hou Zhiqiang^{1,2}, Liu Xiaoyi^{1,2*}, Yu Wangsheng³, Ma Sugang^{1,2}

¹College of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an, Shaanxi 710121, China;

²Shaanxi Key Laboratory of Network Data Analysis and Intelligent Processing, Xi'an University of Posts and Telecommunications, Xi'an, Shaanxi 710121, China;

³Information and Navigation Institute of Air Force Engineering University, Xi'an, Shaanxi 710077, China



Algorithm framework of this paper

Overview: The Faster R-CNN algorithm uses the non-maximum suppression algorithm for proposals filtering. It adopts the idea of “non-one or zero”, leaving only the candidate box with the highest score of the classification targets, which greatly increases the risk that the target will be missed when it is highly overlapping. Therefore, the “weight penalty” strategy is employed by the soft-NMS algorithm to solve this problem, which reduces the target missed detection to a certain extent. However, the test found that the use of the soft-NMS algorithm will greatly increase the number of proposals, resulting in a new problem that the same target is repeatedly detected and multiple detections have mis-targeted the targets, especially when there are multiple targets in the image and the degree of overlap of the targets is high. According to the problems of target missed detection and repeated detection in the object detection algorithm, this paper proposes an improved Faster R-CNN algorithm based on double threshold-non-maximum suppression. The algorithm first uses the VGG-Net-16 deep convolutional network architecture to extract the multi-layer convolution features of the targets, and then proposes the dual threshold-non-maximum suppression (DT-NMS) algorithm in the RPN (region proposal network). The stage extracts the deep information of the target candidate regions, and finally uses the bilinear interpolation method to improve the nearest neighbor interpolation method in the original RoI pooling layer, so that the algorithm can locate the targets more accurately on the detection dataset. In order to highlight the performance of the DT-NMS algorithm on the target repetitive detection problem, this paper first proposed the repeated detection rate and the object mis-distribution rate of multiple detections as the measurement index. By simply setting the threshold in the DT-NMS algorithm, the relationship between the single-threshold algorithm and the target misdetection problem is effectively balanced, and the probability that the same target is detected multiple times is reduced. The improved Faster R-CNN algorithm re-adjusts network training and parameters on the VGG-Net-16 network structure, and a lot of experimental verification on the PASCAL VOC data set has been implemented. The experimental results show that compared with the soft-NMS algorithm, the repeated detection rate of the proposed algorithm in PASCAL VOC2007 is reduced by 2.4%, and the target error rate of multiple detections is reduced by 2%, indicating that the improved algorithm solves the problem of target missed detection and repeated detection in the traditional algorithms. Compared with the Faster R-CNN algorithm, the detection accuracy of this algorithm on the PASCAL VOC2007 is 74.7%, and the performance is improved by 1.5%. At the same time, the algorithm has a fast detection speed, reaching 16 FPS.

Citation: Hou Z Q, Liu X Y, Yu W S, *et al.* Improved algorithm of Faster R-CNN based on double threshold-non-maximum suppression[J]. *Opto-Electronic Engineering*, 2019, 46(12): 190159

Supported by National Natural Science Foundation of China (61703423, 61473309) and Xi'an University of Posts and Telecommunications Graduate Innovation Fund (CXJJ2017019)

* E-mail: 18829290763@163.com