

改进的GDT-YOLOV 3目标检测算法

唐悦, 吴戈, 朴燕

引用本文:

唐悦, 吴戈, 朴燕. 改进的GDT-YOLOV 3目标检测算法[J]. 液晶与显示, 2020, 35(8): 852–860.

TANG Yue, WU Ge, PIAO Yan. Improved algorithm of GDT-YOLOV 3 image target detection[J]. *Chinese Journal of Liquid Crystals and Displays*, 2020, 35(8): 852–860.

在线阅读 View online: <https://doi.org/10.37188/YJYXS20203508.0852>

您可能感兴趣的其他文章

Articles you may be interested in

基于深度可分离卷积的交通标志识别算法

Traffic sign recognition algorithm based on depthwise separable convolutions

液晶与显示. 2019, 34(12): 1191–1201 <https://doi.org/10.3788/YJYXS20193412.1191>

基于深度学习的航空对地小目标检测

Detection of small target in aerial photography based on deep learning

液晶与显示. 2018, 33(9): 793–800 <https://doi.org/10.3788/YJYXS20183309.0793>

多尺度特征融合与极限学习机结合的交通标志识别

Traffic sign recognition based on multi-scale feature fusion and extreme learning machine

液晶与显示. 2020, 35(6): 572–582 <https://doi.org/10.3788/YJYXS20203506.0572>

基于LeNet-5的卷积神经网络图像识别算法

Convolutional neural image recognition algorithm based on LeNet-5

液晶与显示. 2020, 35(5): 486–490 <https://doi.org/10.3788/YJYXS20203505.0486>

基于深度学习的乳液泵缺陷检测算法

Defect detection algorithm of lotion pump based on deep learning

液晶与显示. 2019, 34(1): 81–89 <https://doi.org/10.3788/YJYXS20193401.0081>

文章编号:1007-2780(2020)08-0852-09

改进的 GDT-YOLOV3 目标检测算法

唐悦, 吴戈*, 朴燕

(长春理工大学电子信息工程学院, 吉林 长春 130022)

摘要:为了解决视频图像中目标检测准确率低、速度慢等问题,本文提出了一种基于 YOLOV3 改进的目标检测方法。通过引入 GIOU Loss,可解决原 IOU 无法直接优化的非重叠部分问题,在借鉴了密集连接网络的思想之后,将 YOLOV3 中的 3 个残差块更换为 3 个密集块,并结合 Max Pooling 加强密集连接块之间的特征传递,重新替换 IOU 和原网络的连接结构,检测设计出新的网络结构后,减少了参数量,增强了特征的复用与融合,最终实现了优于原方法的效果。实验结果表明:改进的 GDT-YOLOV3 算法与原有的算法相比,无论是在简单还是复杂交通场景中都有较优秀的检测效果,本文所提出的算法平均检出准确率高达 92.77%,速度达到 25.3 f/s,基本满足了实时性。此外在检测精度上,改进的 GDT-YOLOV3 算法要优于 SSD512、YOLOV2 与 YOLOV3 算法。

关键词:目标检测;卷积神经网络;YOLOV3;密集连接网络;K-means

中图分类号:TP394.1;TH691.9 **文献标识码:**A **doi:**10.37188/YJYXS20203508.0852

Improved algorithm of GDT-YOLOV3 image target detection

TANG Yue, WU Ge*, PIAO Yan

(School of Electronics and Information Engineering, Changchun University of
Science and Technology, Changchun 130022, China)

Abstract: In order to solve the problems of low accuracy and slow speed of target detection in video images, an improved YOLOV3-based object detection method is proposed. By introducing GIOU Loss, the non-overlapping part problem that the original IOU cannot directly optimize can be solved. After drawing on the idea of densely connected networks, the three residual blocks in YOLOV3 are replaced with three dense blocks, and the denseness is combined with Max Pooling to strengthen the denseness. After transferring the features between the connected blocks, and replacing the IOU and the original network to detect the connection structure, a new network structure is designed. The number of parameters is reduced, the feature reuse and fusion are enhanced, and the effect is better than the original method. The experimental results show that comparing with the original algorithm, the improved GDT-YOLOV3 algorithm has excellent results in both simple and complex traffic scenarios. The average detection accuracy of the algorithm proposed in this paper is up to 92.77%, and the speed has reached 25.3 f/s, which basically meets the real-time performance. In addition, in terms of detection accuracy, the improved GDT-YOLOV3 performs better than SSD512, YOLOV2, and

收稿日期:2019-11-09; **修订日期:**2020-02-26.

基金项目:国家自然科学基金 (No.60977011, No.20180623039TC, No.20180201091GX)

Supported by National Natural Science Foundation of China (No. 60977011, No. 20180623039TC, No. 20180201091GX)

* 通信联系人, E-mail: 651837271@qq.com

YOLOV3.

Key words: target detection; convolutional neural network; YOLOV3; densenet; K-means

1 引言

随着国家经济飞跃式的增长,人民的生活水平也在不断地提高,在城市道路上,可以看见机动车几乎人人都在使用,而随着互联网科技的发展,大多数人都用饿了么、美团外卖进行订餐,这种足不出户叫外卖的生活方式导致了骑电动摩托车的人也多了起来。机动车和非机动车(电动车、摩托车)的出现改变了人们的生活方式,它取代了原始的人力车和步行的出行方式。在给人们的出行带来极大的方便之时,也给交警部门的工作人员对道路上车流的监管带来了较大的挑战。机动车和非机动车数量的增加,导致了交通事故的频频出现,间接地造成了人员的伤亡以及财产损失。

深度学习在目标检测中被广泛应用,特别是对远小近大和运动速度不同的目标,在背景复杂的交通场景中,进行准确并实时的检测是最具挑战性的任务^[1-7]。深度学习也取得了质的发展,第一类是两阶段基于区域的检测算法,以 Fast RCNN(Fast Region Convolutional Neural Networks)、Faster RCNN(Faster Region Convolutional Neural Networks)、Mask RCNN(Mask Region Convolutional Neural Networks)为主要代表的算法,这类算法对目标的检测准确度很高,但是检测速度稍慢。第二类是单阶段基于回归方法的检测算法,将检测事件当作回归事件进行解决,可以直接将检测目标的位置和类别预测出来。该类算法可以达到实时性检测,但检测的准确度显然不如第一类算法。主要以 YOLO(You Only Look Once)、SSD512(Single Shot MultiBox Detector512)、YOLOV2(You Only Look Oncev2)、YOLOV3(You Only Look Oncev3)为代表性的网络为主。

使用卷积神经网络所提取到的前景目标的特征,确实要比传统方法提取到的特征要好得多,可以避免许多缺点。基于 CNN(Convolutional Neural Networks)的目标检测算法被应用于道路车辆监管、农业果实采摘、辅助驾驶等方面。文献[8]将 VGG(Visual Geometry Group)网络提取

特征,经过初步二分类和回归后得到一系列预选框,将此预选框输入到特征金字塔结构中,并构建注意力掩模模块自适应地学习有效特征,同时融合特征金字塔结构与注意力掩模模块得到更具表征性的特征;文献[9]提出了一种多尺度聚类卷积神经网络算法,来实现对行人的识别与检测,该算法是对 YOLOV3 进行改进,为了得到相应的特征图,首先通过简单聚类对图像特征进行提取,然后通过抽样 K-means 聚类算法结合核函数确定锚点位置,以达到更好的聚类;文献[10]将利用 VGG16 网络作为微调网络,并添加部分深层网络,通过提取目标浅层特征与深层特征进行联合训练,克服检测过程中定位与识别相互矛盾的问题。

本文以基于回归的目标检测方法中的深度卷积网络 YOLOV3 为基础,首先为了解决场景中运动目标尺寸大小不一的问题,将视频图像中的运动物体作为待检测目标。本文基于 YOLOV3 算法,针对视频图像中的运动目标对 YOLOV3 算法做了优化及改进,并设计搭建了一个用道路监控场景中目标检测的 CNN 网络。使用 3 种不同的数据集混合^[11]训练,可以解决单个类别检测效果差的问题。实验结果表明,改进后的 GDT-YOLOV3(GIOU Densenet Transition Module)网络与 SSD512、YOLOV2 与 YOLOV3 比较,检测效果有了明显提升。

本文首先提出了将 GIOU(Generalized Intersection over Union) Loss 引入到 YOLOV3 检测算法中,解决了原 IOU(Intersection over Union)无法直接优化非重叠部分的问题。搭建一个紧密连接的 GDT-YOLOV3 网络,使尺度 1、尺度 2、尺度 3 在进行预测之前,能够收到 Denseblock 输出的多层卷积特征,最大化地实现特征的复用及融合。使用密集连接块提高了 CNN 层间的连接,可用 Max Pooling 前一个密集模块间的输出,可更好地增强 DenseNet 区块间的连接,可解决 CNN 网络的退化问题。针对道路监控中的运动目标,例如骑电动车的行人数据集较少的现象,制作了相关数据集。

2 YOLOV3 网络介绍

YOLO(You Only Look Once)算法最初是 Redmon^[12]等在 2016 年提出的一种基于回归的目标识别方法,到 2018 年发展到了第 3 代,所以被命名 YOLOV3,它的思想是端到端的回归方法,对比先选取候选区域,再特征提取做分类的双阶段算法,它是一种轻快高效的检测方法。

YOLOV3 网络仍然保持着 YOLOV2^[13]网络的优点,不仅在检测速度上有了提升,而且在对目标的检出率也有了提升,尤其在距离视频监控远的小目标的检测方面,检出率有显著的提升。

YOLOV3 网络使用 anchor boxes^[14]的思想,使用了 3 个尺度在 COCO 数据集上进行了预测,每一个尺度有 3 个 anchor boxes,小的先验框被尺度大的特征图使用,可以根据要检测的目标挑选合适的先验框 anchor boxes,最后根据准备预测的尺度对网络结构进行修改。

3 相关工作

3.1 数据集

目前 COCO 数据集的应用较为广泛,尤其在目标检测、语义分割及目标识别等方面,更是被广泛应用,这个数据集包含了自然景象图片,以及生活中随处可见的一些背景复杂和简单的图片,而且 COCO 数据集的数量比较多。

在深度学习中,数据集质量选取的好坏,以及样本数量是否处于一个平衡状态直接影响最终检测效果的高低,大量的样本让网络能够充分学习到需要检测的目标特征。但是本实验中所使用的数据集中骑电动车、摩托车或行人的样本相对于机动车较少,也就会产生样本失衡的问题。因此会有过拟合问题的出现,且最终得到的检测模型在训练后泛化能力比较差。

为了避免这一问题的出现,本文在实验中采用训练整合数据集的方法,首先通过视频图像提取出含有骑车人的图像,然后进行人工标注样本个数;为了保证样本数量的最佳平衡,同时下载 KITTI 数据集,并将这个数据集的格式转换为 COCO 数据集的格式,选择使用这个数据集的原因是,它包含了各种场景下由车载摄像头采集的

道路车辆实际情况,KITTI 数据集中分为 8 个类别:汽车、行人、厢式货车、坐着的行人、骑车的人、杂项、电车。本文对 KITTI 数据集原有的 8 类标签信息进行处理,保留实验需要的 3 个类别标签,即 Car、Pedestrian 和 Cyclist。

本文最终将从 3 个数据集选取共 9 698 张图片,同时用于训练视频监控场景下的运动目标检测模型。测试集采用人工标注的图片约 2 000 张。

3.2 候选边框维度聚类

如果初始候选框选取不好,网络的学习速度也会变得较差。对于本文使用的数据集来说,它包含的种类非常多,每个目标在图片中的姿势、所处位置、大小都略有不同;而对于行人以及骑车的行人数据集来说,它的基本特点都是长宽比明显不同。在 Faster R-CNN 和 SSD 中需要人为事先设置好先验框,主观性比较强。对于人为设置的先验框,在设置先验框时,选择一个维度适合的先验框是非常重要的,如果选择的先验框很好,网络就可以进行相对容易的学习过程,从而得到更好的预测。

本文在 YOLOV3 网络中使用 K-means 算法,将视频监控场景下捕获的运动目标做成数据集,并将数据集中的真实边框进行聚类,可以获得初始候选边框的参数信息。原始的 K-means 算法分别使用了以下几个距离度量方法:曼哈顿距离、欧氏距离以及切比雪夫距离,用其计算两点之间的距离。设置先验框是为了使 predict boxes 与 ground truth 的 GIOU 更好。

本文选取了 $K=1,2,3,\dots,9$,分别对训练集使用 K-means 聚类分析,最终选取的候选边框数量为 9,具体参数为:(10,13),(16,30),(33,23),(30,61),(62,45),(59,119),(116,90),(156,198),(373,326)。YOLOV3 网络在数据集上采用 3 个尺度分别进行预测,具体方式如图 1 所示。

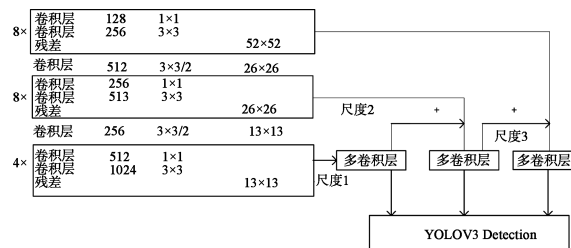


图 1 YOLOV3 网络多尺度预测方式

Fig.1 YOLOV3 network multi-scale prediction method

4 基于 YOLOV3 的改进

本文针对道路监控的运动目标进行检测,对图像中的车和人进行目标检测,对 YOLOV3 提出了 3 点改进。

4.1 改进的 IOU Loss 融合到 YOLOV3

目标检测的中心思想是,如何准确输出图片里物体的类别及所在位置,对于以上两个问题,需要进行位置回归和分类的训练。在位置回归中,使用频率最高的是 Bounding box 和 Ground Truth 框的交并比,通过计算两个框的交集与并集之比,可以预测当前的 predict box 位置与 ground truth 之间的误差,并在训练过程中逐渐回归出一个更好的框。

在位置回归中,通过对 L2 损失函数的分析发现,即使有好的局部最优值,但这个值也未必会是 IOU 的局部最优值,也进一步说明了回归损失与交并比值之间仍有差距。用 IOU 做 loss 的弊端是,当两个目标框无重叠时,交并比 (IOU) 值为 0,如果直接把 IOU=0 作为 loss,梯度不会回传,则对训练来讲是毫无意义的,在实际情况中,这个 0 是非常重要的,它证明了两个框的距离比较远。为了解决 IOU 的缺点,本文提出了 GIOU,该算法的具体过程如下:对于 A、B 两个框,首先要找到最小边界框来涵盖 A、B 区域;其次计算最小框 C 去除了 A、B 外的面积在 C 中所占总面积的比值,最后得到 GIOU 的值就是 IOU 减这个比值, GIOU 为两个框度量准则的算法,如表 1 所示。

表 1 预测框和真实框的度量准则算法

Tab.1 Metric algorithm for prediction box and real box

算法 1:广义的 IOU

输入:两个任意的凹凸形状 $A, B \subseteq S \in R^n$

输出:广义的交并比

对于 A 和 B, 找到一个最小的目标包围框 C, Where $C \subseteq S \in R^n$

$$IOU = \frac{|A \cap B|}{|A \cup B|}$$

$$GIOU = IOU - \frac{|C \setminus (A \cup B)|}{|C|}$$

表 1 中 A 为预测框, B 为真实框, S 是全部框的集合, C 是包含 A 与 B 的最小框, C 也属于 S 集合。

在对目标进行检测中,本文提出了使用广义的交并比 GIOU 取代原始损失,解决了原 IOU 无法直接优化非重叠部分的问题。

4.2 密集相连网络设计

在 YOLOV3 中, Redmon^[7] 等人通过参考残差网络, 设计了 Darknet53 网络。为了实现对道路监控中运动目标的准确检测, 本文使用了 Darknet53, 它通过 ResNet 结构使得训练的复杂度降低, 采用了 1×1 卷积核及 $3 \times 3/2$, stride 为 2 的卷积核取代了 Max Pooling, 降低了参数的数量。

2017 年, Huang 等人提出了新的网络 Densenet^[15], 这个网络借鉴了 Resnet^[16] 思想, 通过对图像特征的最大化利用, 可达到更好的效果和参数。在保证网络中层与层之间最大化的信息传输后, 将每个 Denseblock 中的每一个层连接起来。Densenet 网络结构中有 3 个密集卷积块, 3 个密集卷积块中包含 12 个卷积层, 在每个密集卷积块中, 每一个卷积层可以得到所有之前的卷积层的输出作为输入, 卷积层+池化层连接了毗邻的卷积层, 密集卷积块的结构如图 2 所示。

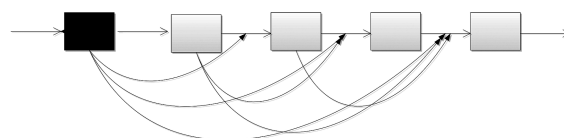


图 2 密集卷积块结构图

Fig.2 Structure of dense convolution block

通过引入这种密集卷积块,可以缓解梯度消失的问题,紧密的连接结构可以提高图片特征的信息传递,在某种程度上降低了参数信息的数目。

本文在基于 Darknet53 网络的基础上使用密集相连增强, Darknet53 的密集连接版本称为 Darknet53-Dense。Darknet53-Dense 由 3 个密集相连模块和 4 个 Transition module 构成, 每个密集相连模块间都有一个 Transition module, 可将特征图的尺寸降低。如图 3 所示, Transition module 中分别对上一个 Dense module 的输出做步长为 2 的 Max Pooling 后与使用 Stride 为 2 的卷积核的输出进行串行连接, 作为下一个 Dense

module 的输入,以这种方式使 Densenet 的模块间连接得到增强,可增强特征的重复使用,减少跨模块间的特征传递损失,使网络可更好地检测到视频图像中的运动目标。图 3 为过渡模块的图解。

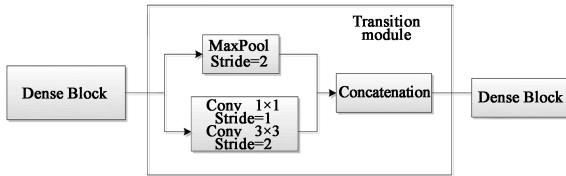


图 3 过渡模块的图解

Fig.3 Diagram of the transition module

4.3 改进的 GDT-YOLOV3 网络设计

在拍摄视频图像过程中会有许多因素的影响,比如相机的内部配件、光照条件及天气因素的影响,尤其在对小目标的检测中,有时人眼都难以准确地看到运动的目标。为了使网络在这些单帧图像上仍然可以有较好的检出率,本文借鉴了 Densenet 网络的思想,将 YOLOV3 网络中尺度 1、尺度 2 及尺度 3 进行预测输入的 3 组 Resnet 模块替换成自定义的 Densenet 模块,搭建一个紧密连接的 GDT-YOLOV3 网络,使尺度 1、尺度 2、尺度 3 在进行预测之前,能够收到 Denseblock 输出的多层卷积特征,最大化地实现特征的复用及融合。改进的 GDT-YOLOV3 网络结构如图 4 所示。

这 3 个 Denseblock 的具体结构如图 5 所示,首先设定 x_0 为模块输入的特征图, x_n 为第 n 个

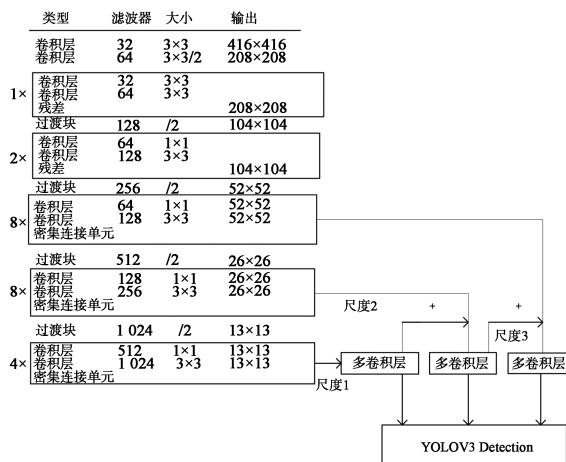
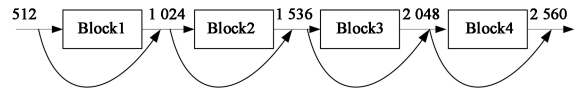


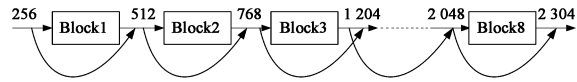
图 4 改进的 GDT-YOLOV3 网络结构

Fig.4 Improved GDT-YOLOV3 network structure



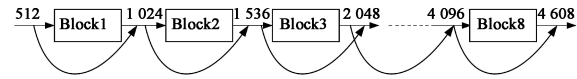
(a) 尺度 1

(a) Scale1



(b) 尺度 2

(b) Scale2



(c) 尺度 3

(c) Scale3

图 5 密集网络块结构图

Fig.5 Dense network block structure

模块的输出特征图, $[x_0, x_1, x_2, \dots, x_{n-1}]$ 表示对 $x_0, x_1, x_2, \dots, x_{n-1}$ 的连接。H()代表 BN-RELU-CONV(3,3)与 BN-RELU-CONV(1,1)的连接, Denseblock 的计算公式为:

$$x_n = H_n([x_0, x_1, \dots, x_{n-1}]), n = 1, 2, \dots, 8. \quad (3)$$

通过推导得出,更改后的 Dense-YOLOV3 网络结构将会在尺度 2 的预测网络中输入 2 304 个特征图,在尺度 3 的预测网络中输入 4 608 个特征图,在尺度 1 的预测网络中输入 2 560 个特征图。

5 实验结果与分析

本实验分别选用另外 4 种网络进行了实验仿真。实验的软硬件平台配置如下:深度学习框架是 darknet53 和 caffe;操作系统是 ubuntu16.04.2, cuda 版本是 8.0.44;电脑配置是 I7-800CPU, 16 G RAM;显卡是 NVIDIA GeForce GTX 1070Ti。

5.1 网络的训练

本文采用了动量项是 0.9 的异步随机梯度下降,设置 subdivisions 为 16,权重的衰减系数为 5×10^{-5} ,权值的初始学习率为 10^{-3} ,由于显卡的限制,在训练阶段设置 batch=32,每次训练的 batch 是 $32/16=2$,输入的图像大小为 416×416 。在 SSD512 网络、YOLOV2、YOLOV3 网络和改进的 GDT-YOLOV3 分别训练 350 000 次,分别

在简单和复杂交通场景下进行试验。测试结果表明,当对网络训练的迭代次数超过一定值时,则会出现过拟合的现象,为了以后测试最佳权重文件更方便,在训练时本文选择每 40 000 次迭代保存一次网络的权重文件。选择的测试集为前 2 000 张图片,使用 4 种网络分别对其所有的权重进行测试,计算召回率的公式为:

$$R = T_p / (T_p + F_N), \quad (4)$$

式中: T_p 表示正确检测出来的目标个数; F_N 代表没有被检测出来的目标个数。测试的部分结果如表 2 所示。

表 2 部分网络召回率与迭代次数关系的对比

Tab.2 Relationship between partial network recall rate and iteration number

迭代次数	YOLOV3/%	GDT-YOLOV3/%
220 000	86.22	89.58
260 000	87.65	95.19
300 000	91.52	93.16
340 000	88.91	85.89

从表 2 可以看出,YOLOV3 网络迭代最好的次数为 300 000 次左右,改进的 GDT-YOLOV3

网络迭代最好也是在 260 000 左右,可以发现改进的 GDT-YOLOV3 网络在检测目标时要比 YOLOV2、YOLOV3 网络和 SSD512 网络的效果好。

5.2 检测效果对比

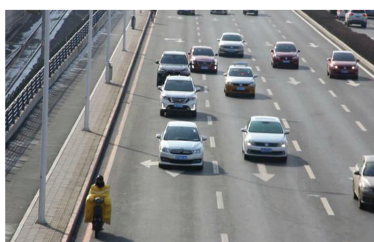
本文选取了 4 种网络的最佳迭代次数对 4 种网络进行对比与测试。表 3 为本文的网络检测模型与其他优秀的目标检测模型的检测精度对比,所有的模型都是采用混合训练集训练的。其中本文的模型检测精度要远高于 SSD512 网络、YOLOV2 和 YOLOV3 网络,虽然检测速度不如 YOLOV3 和 YOLOV2,但也完全满足既实时又准确的要求。

本文对 2 000 张图片进行测试,用这 4 种网络对图像中的目标进行检测。实验结果表明,4 种网络均能准确地检测出道路上的车辆、电动车和行人,虽然原始的 SSD512、YOLOV2、YOLOV3 网络都有较高的检测速度,但是与本文改进的 GDT-YOLOV3 网络的检测精度和速度相比,可发现其他 4 种网络明显不如本文改进的网络模型,而且 GDT-YOLOV3 网络几乎不会发生误检和漏检现象,实验效果如图 6 和图 7 所示。

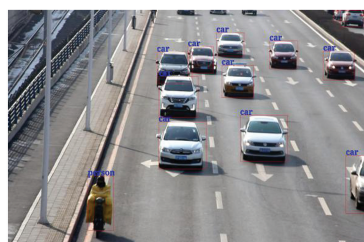
表 3 不同检测模型精度的对比

Tab.3 Comparison of the accuracy of different detection models

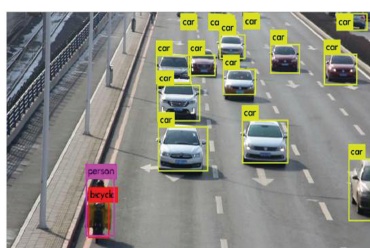
模型	模型精度/%						FPS
	car	people	motorbike	bus	truck	MAP	
SSD512	87.89	85.52	75.22	80.92	78.2	81.55	19.2
YOLOV2	93.65	92.35	79.26	83.78	81.68	86.2	29.3
YOLOV3	95.21	93.15	86.15	86.52	87.31	90.66	28.1
GDT-YOLOV3	96.33	96.98	92.21	89.58	89.72	92.77	25.3



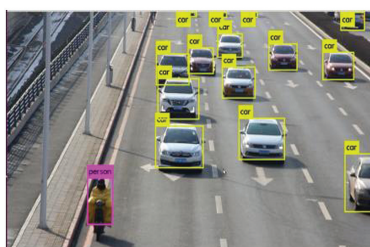
(a) 原图像
(a) original image



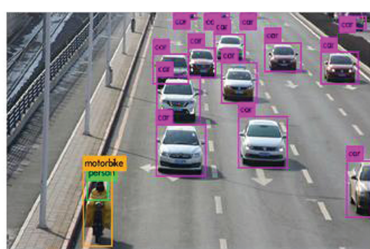
(b) SSD512 网络
(b) SSD512 network



(c)YOLOV2 网络
(c) YOLOV2 network



(d) YOLOV3 网络
(d) YOLOV3 network



(e) 改进的 GDT-YOLOV3 网络
(e) Improved GDT-YOLOV3 network

图 6 简单交通情境下实验检测的对比结果图
Fig.6 Comparison results of experimental tests in a simple traffic situation

这 2 000 张图片包含 51 700 个目标,用这 4 种网络在数据集上进行测试,然后计算检测 5 种目标类别的准确率,准确率的计算公式:

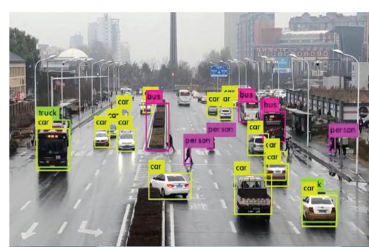
$$P = T_p / (T_p + F_p) . \quad (5)$$



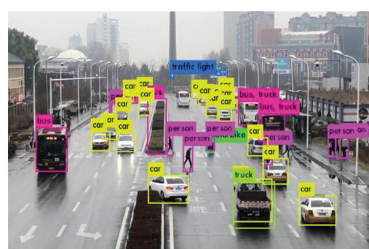
(a)原图像
(a) Original image



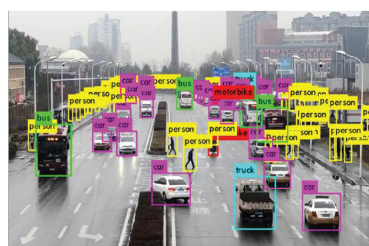
(b)SSD512 网络
(b) SSD512 network



(c)YOLOV2 网络
(c) YOLOV2 network



(d)YOLOV3 网络
(d) YOLOV3 network



(e)改进的 GDT-YOLOV3 网络
(e) Improved GDT-YOLOV3 network

图 7 复杂交通情境下实验检测的对比结果图
Fig.7 Comparison results of experimental tests in complex traffic situations

经过计算得出 SSD512 网络误检车辆个数为 6 260 个,行人误检个数为 7 468 个,电动车误检个数 12 811 个,公共汽车误检个数为 9 864 个,卡车误检个数为 11 270 个;YOLOV2 网络车辆误检车辆个数为 3 282 个,行人误检个数为 3 955 个,电动

车误检个数 10 567 个,公共汽车误检个数为 8 358 个,卡车误检个数为 9 471 个;YOLOV3 网络误检车辆的个数为 2 476,行人的误检个数为 3 541,电动车误检个数为 7 160 个,公共汽车误检个数为 6 969 个,卡车误检个数为 6 653 个;GDT-YOLOv3 网络误检车辆的个数为 1 897,行人的误检个数为 2 078,电动车误检个数为 4 027 个,公共汽车误检个数为 5 378 个,卡车误检个数为 5 315 个。从实验结果可知,本文改进的 GDT-YOLOV3 检测效果更好,与 SSD512 相比分别在车辆、行人、电动车、公共汽车和卡车的检测准确率分别提升了 8.44%、10.46%、16.99%、8.66%、11.52%;与 YOLOV2 相比在车辆、行人、电动车、公共汽车和卡车的检测准确率分别提升了 2.68%、3.63%、6.06%、5.8%、8.06%;与 YOLOV3 相比在车辆、行人、电动车、公共汽车和卡车的检测准确率提升了 1.12%、2.83%、6.06%、3.06%、2.4%,检测速度满足需求。

6 结 论

本文采用 K-means 算法聚类初始目标边框,以 YOLOV3 算法为基础,根据混合数据集的特点,对网络进行了改进。本文设计了一种改进的 GDT-YOLOV3 网络,并分别对 SSD512、YOLOV2 及 YOLOV3 和改进的 GDT-YOLOV3 网络进行训练与测试。实验结果表明,本文方法的平均检出率与 SSD512、YOLOV2、YOLOV3 相比分别提高了 11.22%、6.57%、2.11%,可以对交通场景下不同尺寸的目标有较高的检出率。但是本模型对更复杂的交通场景或重叠的运动目标检测效果不是特别好,在未来会将上述问题作为重点的研究方向。

参 考 文 献:

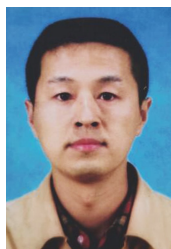
- [1] 刘博艺,程杰仁,唐湘艳,等.复杂动态环境下运动车辆的识别方法[J].计算机科学与探索,2017,11(1):134-143.
LIU B Y, CHENG J R, TANG X Y, *et al.* Moving vehicles recognition in complex dynamic environment [J]. *Journal of Frontiers of Computer Science & Technology*, 2017, 11(1): 134-143. (in Chinese)
- [2] LI H, FU K, YAN M L, *et al.* Vehicle detection in remote sensing images using denoising-based convolutional neural networks [J]. *Remote Sensing Letters*, 2017, 8(3): 262-270.
- [3] 吴言枫,王延杰,孙海江,等.复杂动背景下的“低小慢”目标检测技术[J].中国光学,2019,12(4):853-865.
WU Y F, WANG Y J, SUN H J, *et al.* LSS-target detection in complex sky backgrounds [J]. *Chinese Optics*, 2019, 12(4): 853-865. (in Chinese)
- [4] 高文,朱明,贺柏根,等.目标跟踪技术综述[J].中国光学,2014,7(3):365-375.
GAO W, ZHU M, HE B G, *et al.* Overview of target tracking technology [J]. *Chinese Optics*, 2014, 7(3): 365-375.
- [5] 赵永强,饶元,董世鹏,等.深度学习目标检测方法综述[J].中国图象图形学报,2020,25(4):629-654. (in Chinese)
ZHAO Y Q, RAO Y, DONG S P, *et al.* Survey on deep learning object detection [J]. *Journal of Image and Graphics*, 2020, 25(4): 629-654. (in Chinese)
- [6] 姜文涛,张驰,张晟翀,等.多尺度特征图融合的目标检测[J].中国图象图形学报,2019,24(11):1918-1931.
JIANG W T, ZHANG C, ZHANG S C, *et al.* Multiscale feature map fusion algorithm for target detection [J]. *Journal of Image and Graphics*, 2019, 24(11): 1918-1931. (in Chinese)
- [7] 杨丹,戴芳.运动目标检测的 ViBe 算法改进[J].中国图象图形学报,2018,23(12):1813-1828.
YANG D, DAI F. Improved ViBe algorithm for detection of moving objects [J]. *Journal of Image and Graphics*, 2018, 23(12): 1813-1828. (in Chinese)
- [8] 董潇潇,何小海,吴晓红,等.基于注意力掩模融合的目标检测算法[J].液晶与显示,2019,34(8):825-833.
DONG X X, HE X H, WU X H, *et al.* Object detection algorithm based on attention mask fusion [J]. *Chinese Journal of Liquid Crystals and Displays*, 2019, 34(8): 825-833. (in Chinese)
- [9] 葛雯,史正伟.改进 YOLOV3 算法在行人识别中的应用[J].计算机工程与应用,2019,55(20):128-133.
GE W, SHI Z W. Application of improved YoLOV3 algorithm in pedestrian identification [J]. *Computer Engineering and Applications*, 2019, 55(20): 128-133. (in Chinese)

- [10] 梁华,宋玉龙,钱锋,等.基于深度学习的航空对地小目标检测[J].液晶与显示,2018,33(9):793-800.
LIANG H, SONG Y L, QIAN F, *et al.* Detection of small target in aerial photography based on deep learning [J]. *Chinese Journal of Liquid Crystals and Displays*, 2018, 33(9): 793-800. (in Chinese)
- [11] 李云鹏,侯凌燕,王超.基于 YOLOv3 的自动驾驶中运动目标检测[J].计算机工程与设计,2019,40(4):1139-1144.
LI Y P, HOU L Y, WANG C. Moving objects detection in automatic driving based on YOLOv3 [J]. *Computer Engineering and Design*, 2019, 40(4): 1139-1144. (in Chinese)
- [12] REDMON J, DIVVALA S, GIRSHICK R, *et al.* You only look once: unified, real-time object detection [C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 779-788.
- [13] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 6517-6525.
- [14] DAI J F, LI Y, HE K M, *et al.* R-FCN: object detection via region-based fully convolutional networks [C]//Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona: ACM, 2016: 379-387.
- [15] HUANG G, LIU Z, VAN DER MAATEN L, *et al.* Densely connected convolutional networks [C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 2261-2269.
- [16] HE K M, ZHANG X Y, REN S Q, *et al.* Deep residual learning for image recognition [C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770-778.

作者简介:



唐悦(1994—),女,吉林长春人,硕士,2020年于长春理工大学获得硕士学位,主要从事机器视觉与自动控制方面的研究。E-mail: tangyue0103@163.com



吴戈(1974—),男,吉林长春人,博士,副教授,2011年于长春理工大学获得博士学位,主要从事智能控制与信息处理方面的研究。E-mail: 651837271@qq.com