公路视频实时车辆检测分类与车流量统计算法*

查伟伟,白　天
(中国科学技术大学 软件学院,安徽 合肥 230000)

摘　要:公路视频实时车辆检测分类与车流量统计是计算机视觉领域的一个经典问题。传统设置检测带法,易漏检复检,自动化性不好。基于深度网络的 one-stage 算法实时性好,但是经常会把变化的背景、运动的非车辆物体纳入其中,同时对光照变化敏感,夜间分类效果不好。因此,提出采用 one-stage 做目标检测,并不直接获取分类结果,而是根据标注框将物体切割出来,去除背景,提升抗背景扰动性能和分类效果;再送入一个经过迁移学习的浅层神经网络;将分类输出和目标检测网络的位置输出合并送入一个全图匹配算法,进行车流量统计。该方法在保障实时性的同时降低了漏检和复检率。

关键词:卷积神经网络;目标检测与分类;实时车流量统计;YOLOv3 网络
中图分类号:TP181　　**文献标识码**:A　　DOI: 10.19358/j.issn.2096-5133.2020.03.012
引用格式:查伟伟,白天.公路视频实时车辆检测分类与车流量统计算法[J].信息技术与网络安全,2020,39(3):62-67,72.

Highway video real-time vehicle detection classification and traffic flow statistics algorithm

Zha Weiwei, Bai Tian
(Department of Software Engineering, University of Science and Technology of China, Hefei 230000, China)

Abstract: Real-time vehicle detection, classification and traffic statistics based on road video are classic problems in the field of computer vision. The traditional method of setting the detection belt is prone to missed inspection and re-inspection, so the automation performance is not good. The real-time performance of one-stage algorithm based on deep network can be guaranteed, but the changing background, moving non-vehicle objects are often included, and the change of illumination is sensitive at the same time, so the classification at night is not good. Therefore, an algorithm is proposed to perform target detection by one-stage, and the classification result is not directly obtained. Instead, it cuts out the object according to the bounding box, removes the background, and improves resistance to background disturbances and classification accuracy. Then it is sent to a transfer learning shallow neural network. The classified output and the position output of the target detection network are combined and sent to a full map matching algorithm for traffic flow statistics. While ensuring real-time performance, the rate of missed inspections and re-inspections is reduced.

Key words: convolutional neural network; target detection and classification; real-time traffic statistics; YOLOv3 network

0　引言

公路视频的车辆分类与车流量统计是运动物体目标检测识别与跟踪问题,可以通过传统图像方法和现代深度网络实现。传统图像方法由于计算量较小,因此实时性相对较高。现代深度网络在背景分割、目标分类的准确度上有着压倒性的优势。

传统方法进行运动物体检测一般是在图像连续的帧之间做差分,进行背景去除或者对每个像素进行前景背景建模。经典的方法如帧差法[1]、光流法[2]、背景减除法[3]、高斯背景建模[4],这些方法实时性高,但弊端也显而易见:对高速和低速运动物体检测效果差,抗背景扰动性差,对光照度变化敏感,切割出的物体有较大的背景边缘。这些弊端不利于后续的分类和统计。文献[5]给出了背景建模的难点并构建了合成数据库用以评价其他方法。

深度卷积神经网络[6](Deep Convolutional Neu-

* 基金项目:福建省交通运输科技发展项目(201431)

ral Network,DCNN)应用之后,给类似于图像的网格计算带来了福音。从 RCNN[7]、Fast RCNN[8]到 Faster RCNN[9]以及后续的 Mask RCNN[10]等一系列深度网络在目标检测和分类准确率上有了极大的提升。其中 Mask RCNN 甚至可以将目标沿着边缘直接切割出来。特征金字塔 FPN[11]的提出更是对细小物体轮廓检测的精确度产生了极大影响,可以在低层次特征图上获取小物体轮廓。RetinaNet 提出用 FocalLoss[12]代替原来的交叉熵损失,缓解了 two-stage 网络的类别不平衡问题,进一步提高了检测速度。与此同时,像 YOLO[13]、SSD[14]、对 SSD 的改进 RefineDet[15],以及对特征图融合的改进 M2Det[16]这样的 one-stage 深度网络在提升准确度的同时,实时性上也有所提升。在经过 YOLO 网络和 YOLO9000[17]算法的奠基之后,YOLOv3[18]已经可以对视频进行端到端的实时目标检测和分类,但如果将其直接用在公路视频上,研究发现由于车辆的由远及近,在远处捕捉的目标误差很大;夜间光照度弱的时候,背景常常被误判捕捉;甚至是摄像头视频左上角的时间变化都会被捕捉成运动物体。这不便于车流量统计。

基于以上问题和需要,本文提出一个基于修改后的 YOLOv3 和浅层网络的重构网络和后续的去重统计车流量算法。实验结果表明相比传统检测带法,该算法明显降低了复检和漏检率,相比于目前的目标跟踪方法降低了时间复杂度。

1 相关网络

1.1 YOLOv3 检测网络

YOLOv3 是一个端到端的 one-stage 目标检测网络。其端到端的特性是指可以直接从视频输入到视频输出。这依赖其 one-stage 的优势——实时性好,依据其官网提供的数据,YOLOv3-320 可以达到 45 f/s,YOLOv3-tiny 可以达到 220 f/s,可以实现在 mAP 和速率之间权衡。公路视频一般在 25 f/s,可以满足实时性需要。但相对于 two-stage 网络,YOLOv3 也有其明显的劣势,由于其候选 anchor 数目很多,而真正进行分类和坐标回归预测的只有少数真正含有目标物体的 anchor。YOLOv3 的损失包括了是否含有目标物体的置信度损失、分类损失以及坐标预测损失。大量的 anchor 预测的都是背景导致负例比例严重大于正例比例,而负例损失虽然没有计入分类损失和坐标预测损失,但是却会被计入是否含有目标物体的置信度损失中。通过实际预测较少的 anchor 提高了 YOLOv3 的速率但也降低了预测的准确度。

1.2 浅层分类网络

自 Alexnet 之后,人们自主设计的分类网络层出不穷。经典的如 VGG、GoogleNet、Inception Network、ResNet。与这些网络相比,浅层网络层数少,时效高,准确率也不会丢失几个百分点。较之后动则上百层的分类网络,浅层网络可以仅有 5 层卷积层、3 层池化层以及 4 个全连接层。外加一些 ReLU、dropout 等提高网络性能的手段。考虑到实时性的要求,本文采用这样一个层数较浅的分类网络去提高抗背景的扰动性,去除目标检测网络中错误切割出的运动目标以及变化的背景。

2 公路视频实时车辆检测分类与车流量统计

本文提出一种基于 YOLOv3 与浅层神经网络整合后的实时车辆检测与分类网络,在满足实时性的同时提高网络对公路视频的车辆检测和分类的准确性。之后根据两个网络的输出提出一种基于传统图像方法的车流量统计算法。

2.1 检测分类网络

检测网络采用 YOLOv3,模型采用 YOLOv3-416,根据官网提供数据,YOLOv3-416 可达 35 f/s。利用 Darknet 平台提供的 detector demo API 测试,可以实现公路视频的实时端到端输出。实测直接用 YOLOv3 训练自己的车辆数据集进行检测分类时,网络由于数据集的大小、公路视频的噪声干扰、环境光照度变化等因素导致分类效果不理想,甚至会错误分类一些类似模糊的背景以及摄像头上方变化的时间信息。实时性可以保证,准确性达不到实际需求。

针对以上问题,本文对 YOLOv3 网络的主体部分进行改造。改造后的网络如图 1 所示。

为保证分类的准确性不受公路视频噪声、光线变化等因素的干扰,应当先对图片数据进行一次筛选,不应当直接分类,所以改造 YOLOv3 主干网络输出层前的 1×1 网络通道数为 3×7,3 为每个 grid 预测 3 个尺寸的 anchor;7 包含 x、y、w、h 四个坐标,物体置信度 prob 以及 iscar、notcar 两个分类。两个分类中的 notcar 用来对采集的视频图片数据进行清洗,筛除行人、摄像头上角变化的数字等一系列运动的物体。考虑到车辆在刚进入摄像头画面时是

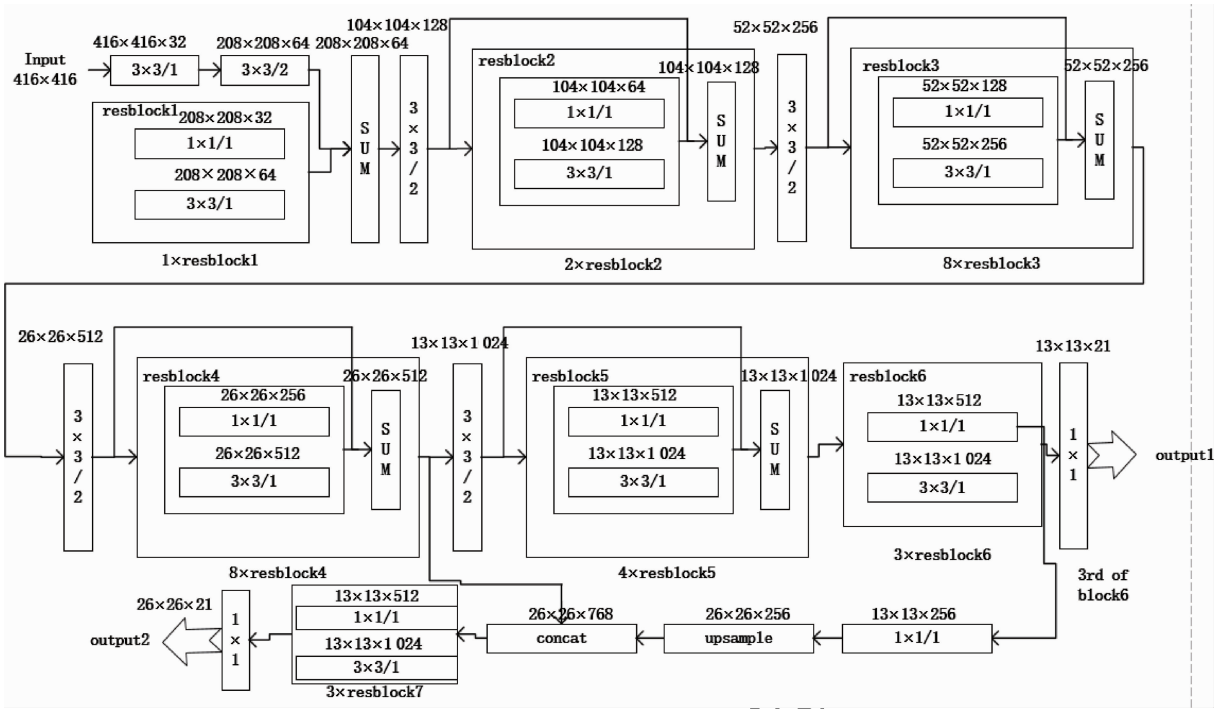

图 1　改造后的 darknet53

不清晰的,不利于后续分类,所以希望捕捉到尽量大一些清晰些的车辆,同时进一步提高实时性。基于以上两点,在 FPN(特征金字塔)中,只向上采样一次,在原有 13 × 13 的特征图中再得到 26 × 26 的融合特征输出,分别对应于图 1 中的 output1 和 output2。对主干网络的输出进行坐标回归以及非极大值抑制后得到的输出为准车辆的 boundingbox 坐标(小目标物体由于多次特征提取导致坐标回归不精确,很大概率会在后续的初步根据置信度筛选目标框中被过滤)。根据坐标对原图进行裁剪,将目标裁剪出来,进一步排除背景变化对分类的干扰。为保证实时性,选择一个层数不深的浅层神经网络作为后继,浅层分类可并行。将裁剪后的目标图片 resize 成固定大小的图片作为浅层网络的输入。浅层网络输出为 car、bus、motor、truck、none 五个分类,其中 none 用来筛除误检测的背景等图片。考虑到自身公路视频车辆数据集的大小问题,为了防止过拟合,并不对车辆数据集进行去噪和手动筛选,仍然采用含有噪声扰动以及运动模糊的原公路视频车辆数据集进行训练。为了防止高偏差,浅层网络的特征提取层采用在 ImageNet 上的预训练参数,仅对 4 个全连接层用自身车辆数据集迁移学习。

2.2　融合网络的训练

融合后的网络采取分开训练的方式,改造后的 YOLOv3 网络使用 ImageNet 进行初始权重的训练,初始学习率设置为 0.1,weight decay 设置为 0.000 5,momentum 设置为 0.9,batchsize 设为 128。训练 30 个 epochs。使用厦门公路局提供的车辆数据集制作成 coco 数据集在初始权重上再训练 10 个 epochs。浅层分类网络的训练采取 finetune 的方式,初始权重在 ImageNet 上预训练得到,保存在 npy 文件中,将该文件中网络权重转化成 tensorflow 的 ckpt 格式。加载过后冻结最后的四个全连接层,将厦门公路局提供的车辆数据集通过改造的 YOLOv3 网络进行目标检测,裁剪出车辆目标以及一些 none 分类目标图片作为训练集,训练集比例 car:bus:motor:truck:none 为 1:1:1:1:1。将该裁剪后的图片训练集对浅层网络进行 finetune,得到最终的权重文件。

2.3　车流量统计算法

将浅层网络输出的 4 个车辆分类(none 分类直接过滤)以及 YOLOv3 网络输出的对应目标坐标作为车流量统计算法的输入。因为 YOLOv3 网络在目标检测阶段对每一帧图片都实时地输出了目标的中心点坐标和长宽,可以充分利用帧与帧之间的大

量重复信息,类似于帧差法求运动前景的轮廓,当物体运动过快或者选择不连续的帧进行差分的时候就会出现拖长的轮廓。同理,可以对每隔几帧的图片的 anchor box 中心点坐标以及长宽进行匹配。目标中心点坐标相差小于阈值,并且长宽变化小于阈值时判定为同一辆车。反之判定为新的车辆进入视野,车辆计数器加一。

算法流程如下:

(1)初始化车辆计数器 Count = 0。

(2)使用 YOLOv3 检测算法检测交通视频中的第一帧画面里的所有车辆(静态图片检测),得到其检测区域中心坐标信息和长宽信息组成向量 $\boldsymbol{P}_w^1$, $\boldsymbol{P}_w^2,\cdots,\boldsymbol{P}_w^n$(前一帧每辆车检测区域中心坐标和长宽信息用一个向量 $\boldsymbol{P}_w^i$ 表示,其中包括中心点坐标 x、y,区域长宽 w、h 共四个分量),并保存到本地中,更新车辆计数器 Count = n。

(3)对于交通视频中的第 k 帧画面,使用 YOLOv3 算法框架,得到检测区域中心坐标信息和长宽信息向量 $\boldsymbol{P}_k^1,\boldsymbol{P}_k^2,\cdots,\boldsymbol{P}_k^m$(后一帧每辆车检测区域中心坐标和长宽信息用一个向量 $\boldsymbol{P}_w^i$ 表示,其中包括中心点坐标 x、y,区域长宽 w、h 共四个分量)。

(4)计算坐标 $\boldsymbol{P}_k^1,\boldsymbol{P}_k^2,\cdots,\boldsymbol{P}_k^m$ 与坐标 $\boldsymbol{P}_w^1,\boldsymbol{P}_w^2,\cdots,\boldsymbol{P}_w^n$ 之间的欧式距离,得到距离矩阵如式(1)所示。

$$\begin{bmatrix} L_{11} & L_{12} & \cdots & L_{1n} \\ L_{21} & L_{22} & \cdots & L_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ L_{m1} & L_{m2} & \cdots & L_{mn} \end{bmatrix} \tag{1}$$

(5)计算矩阵每一行 L_{i1}, L_{i2},$\cdots$,L_{in}($i = 1,2,\cdots,m$),若存在 $L_{ij} < \varepsilon (j = 1,2,\cdots,m)$,则 Count 不变;否则,Count ++。

(6)提取交通视频剩余帧画面,重复步骤(3)~步骤(5),直至结束。

算法流程图如图 2 所示。

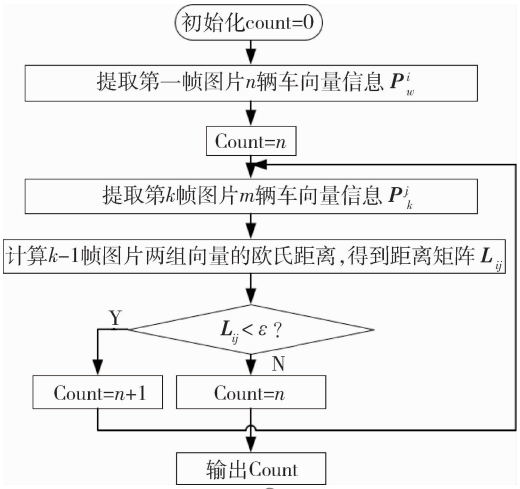

图 2 车流量统计算法流程图

3 实验与分析

3.1 实验配置

实验用到两个数据集:ImageNet 数据集和从厦门公路局提供的公路视频提取到的车辆数据集。其中公路视频数据集包括 45 012 张训练图片和 13 587 张测试图片。训练集和测试集均划分为 5 个类:car, truck, motor, bus, none。其中 none 代表非车辆分类的运动物体或背景。并且对该数据集只进行了人工打标签操作,并未进行任何数据集清洗,图片包含噪声,目标物体存在运动模糊,与摄像头存在距离变化。因此训练出的模型泛化能力好。

实验环境包括硬件设备和软件配置。硬件参数为:Intel(R) Xeon(R) Silver 4116 CPU @ 2.10 GHz; 6 条 32 GB 内存;Ubuntu16.04;Tesla V100 16 130 MB 显存。软件参数为:TensorFlow 1.13.1;Keras 2.2.4; Python3.7.3;Jupyter Notebook。

实验模型参数分为两部分,第一部分 YOLOv3 网络采用 YOLOv3-416 改造后的模型,权重参数使用在 ImageNet 数据集和自身车辆数据集上训练得到的初始参数。该权重是在 ImageNet 数据集上分了几个阶段训练得到的,比直接使用公路视频提取的车辆数据集训练得到的权重在目标检测上效果好。第二部分浅层网络初始训练采用 ImageNet 数据集。初始训练过后使用车辆测试数据集进行测试,top1 准确率为 83.31%。初始训练和测试完毕后使用车辆训练数据集再次训练,冻结之前的所有特征提取层。训练 10 个 epochs 过后再次使用车辆测试数据集进行测试,top1 准确率达到 96.30%。

3.2 结果与分析

3.2.1 检测分类效果

由于公路视频车辆检测和分类难以找到具体的对照指标。为了体现本文重构模型的分类效果,将本文模型与直接用 YOLOv3 训练车辆数据集得到的模型的分类进行对比。将车辆数据集制作成 YOLOv3 官方要求的 coco 数据集进行训练 YOLOv3

网络,初始权重采用 ImageNet 上的预训练权重。设置上述 5 个分类,模型采用 YOLOv3-416,Batchsize 设置为与浅层网络同样的大小 128,训练 10 个 epoch。随机抽取测试集 5 个分类中车辆的分类结果和置信度大小,本文重构网络分类效果如图 3 所示。

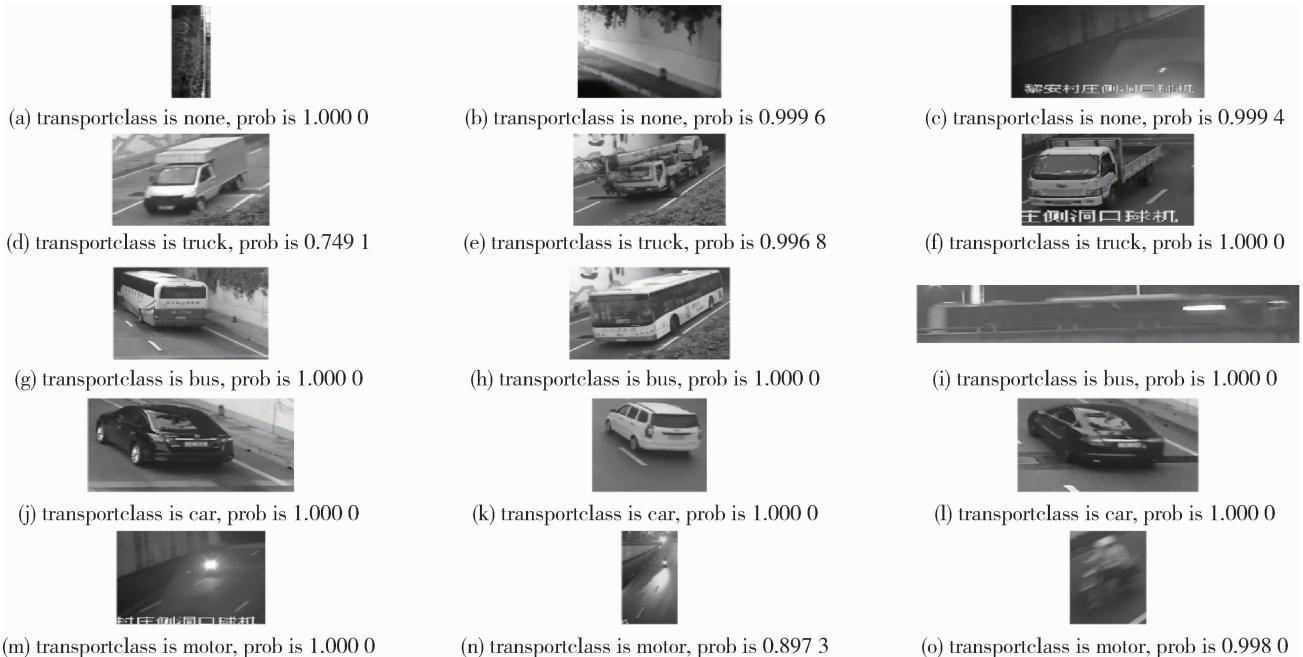

图 3　本文重构网络识别效果图

由图 3 第一行子图可见,本文重构网络对于夜间背景和光线昏暗的隧道等可见度低的背景筛除效果较好;由第二行第一个子图可见,对于特殊的箱式货车,模型可以正确给出分类结果,但分类置信度不高,后期可以通过对这种特殊的货车进行数据增强以提高分类置信度;由第三行和第四行可见,模型对于白天车型分类准确率和置信度很高,完全可以投入到实际应用中;由第五行前两张图片可见,模型对于夜间微小车辆分类具有鲁棒性,由最后一张图片可见运动模糊的车辆在切片过后分类效果较好。

由图 4 很容易看出夜间直接用 YOLOv3 网络训练自己的数据集已经不能检测公路视频中的车辆类型,还有噪声被当成目标错误地捕捉。由图 3 可以看出本文的重构网络夜间可以很好地去除 YOLOv3 错误捕捉的噪声构成的目标,去除大量噪声的错检。但对车辆检测的置信度也存在一定的下降,会存在漏检的情况,这时可以通过降低设置的置信度阈值来进行一定的补偿。但仍需后期进一步改进夜间的效果。

综合上述对比图可以看出,基于 one-stage 的

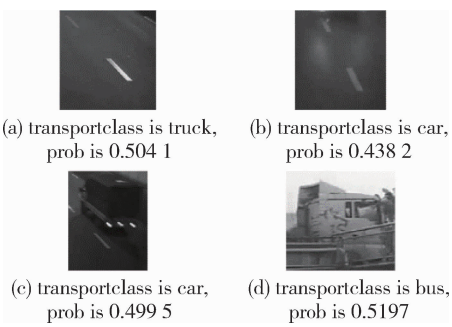

图 4　YOLOv3 网络识别效果图

YOLOv3 网络训练得到的模型直接用在公路视频车辆测试集上,分类存在一定的错误率,尤其是在存在运动模糊的小型物体如 motor 和较大背景周期噪声的 truck 上,错误率较高,而这些情况在公路视频上很常见。置信度的差距尤为明显。这体现出直接在车辆数据集上训练的泛化能力不足。后期可以考虑在 YOLOv3 上冻结某些层进行训练。

3.2.2　车流量统计效果

为验证本文车流量统计算法在掺杂光照强度和背景变化干扰下的效果,随机抽取 3 条道路,每条道路抽取 2 个视频,包含白天、夜晚。人工输出一定视频时长中车辆数目,具体如表 1 所示。

表 1　车流量实际情况统计表

道路	道路 1		道路 2		道路 3	
	视频 1	视频 2	视频 1	视频 2	视频 1	视频 2
时间段	白天	夜晚	白天	夜晚	白天	夜晚
时长/min	30	30	30	30	30	30
车辆总数	186	166	142	153	127	148

实验中采用匹配准确率作为指标：

$$\text{precision} = \frac{\text{right}}{\text{right} + \text{error}} \tag{2}$$

其中，right 代表正确检测并匹配对的机动车数量，error 代表未检测到或者匹配出错导致误检漏检的机动车数量。为验证本文提出的算法效果，本文使用基于虚拟检测带的方法对实验视频数据进行处理并做对比，如表 2 所示。

表 2　车流量统计效果对照表

道路/视频		本文方法		虚拟检测带法	
		准确率/%	误检数	准确率/%	误检数
道路 1	视频 1	95.7	8	87.1	24
	视频 2	83.8	27	81.9	30
道路 2	视频 1	95.8	6	87.3	18
	视频 2	85.6	22	81.7	28
道路 3	视频 1	96.1	5	87.4	16
	视频 2	87.2	19	83.1	25

4　结束语

本文提出一种传统图像方法和深度学习方法相结合的基于公路视频的实时车辆检测分类和车流量统计算法，在保证实时性的同时提高了公路视频车辆分类的准确率。建立了基于 YOLOv3 网络和浅层网络的重构网络，当输入为有噪声污染的图像和公路实时视频时，分类置信度和准确率相较于原 YOLOv3 网络有明显的提高，对于夜间视频有一定的改善效果。基于逐帧匹配的车流量统计算法，相比于传统的检测带法，本文方法误检漏检率低、人工干预少。该方法还有很多可以提升的空间，其中检测网络还是会有一定的小目标物体被检测，导致后续的分类网络会分类一些在图像边缘就捕捉到的轮廓模糊的车辆，从而降低分类准确率；其次目前由于数据集的限制，没有对车辆具体类别再继续细化下去。后续研究会考虑对公路的具体每个车道进行车流量统计。

参考文献

[1] TESAURO G. Temporal difference learning and TD-Gammon[J]. Communications of the ACM, 1995, 38(3): 58-68.

[2] LUCAS B D, KANADE T. An iterative image registration technique with an application to stereo vision[C]. Proceedings of the 7th International Joint Conference on Artificial Intelligence, 1981, 2: 674-679.

[3] MA X X, GRIMSON W E L. Edge-based rich representation for vehicle classification[C]. Tenth IEEE International Conference on Computer Vision (ICCV'05), 2005, 2: 1185-1192.

[4] STAUFFER C, GRIMSON W E L. Adaptive background mixture models for real-time tracking[C]. Computer Vision and Pattern Recognition. DOI: 10.1109/CVPR.1999.784637.

[5] BRUTZER S, HÖFERLIN B, HEIDEMANN G. Evaluation of background subtraction techniques for video surveillance[C]. 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2011: 1937-1944.

[6] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C]. Advances in Neural Information Processing Systems, 2012: 1097-1105.

[7] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[J]. arXiv preprint arXiv: 1311, 2524, 2013.

[8] GIRSHICK R. Fast r-cnn[C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.

[9] REN S, HE K, GIRSHICK R, et al. Faster r-cnn: towards real-time object detection with region proposal networks[C]. Advances in Neural Information Processing Systems, 2015: 91-99.

[10] HE K, GKIOXARI G, DOLLÁR P, et al. Mask r-cnn[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961-2969.

[11] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.

[12] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980-2988.

（下转第 72 页）

[6] 科普中国. 可扩展标记语言[EB/OL]. [2019-11-10]. https://baike.baidu.com/item/可扩展标记语言.
[7] 科普中国. JSON[EB/OL]. [2019-09-26]. https://baike.baidu.com/item/JSON/2462549.

(收稿日期:2020-01-05)

作者简介:

王磊(1995 -),男,硕士研究生,主要研究方向:计算机科学与技术。

王皓(1973 -),男,硕士,高级工程师,主要研究方向:计算机科学与技术。

(上接第 67 页)

[13] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
[14] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]. European Conference on Computer Vision. Springer, Cham, 2016: 21-37.
[15] ZHANG S, WEN L, BIAN X, et al. Single-shot refinement neural network for object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4203-4212.
[16] ZHAO Q, SHENG T, WANG Y, et al. M2det: a single-shot object detector based on multi-level feature pyramid network[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33: 9259-9266.
[17] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.
[18] REDMON J, FARHADI A. YOLOv3: an incremental improvement[J]. arXiv preprint arXiv: 1804.02767, 2018.

(收稿日期:2020-01-05)

作者简介:

查伟伟(1992 -),男,硕士研究生,主要研究方向:机器视觉。

版权声明