

## YOLO9000 模型的车辆多目标视频检测系统研究

李鹏飞, 刘瑶, 李珣, 张宏伟

(西安工程大学 电子信息学院, 西安 710048)

**摘要:** 提出了一种基于 Darknet 框架下 YOLO9000 算法的车辆多目标检测方法; 该方法在 YOLO9000 算法基础上, 根据训练结果和车辆目标特征对 YOLO9000 网络模型进行改进, 并对其算法参数进行调整, 获得更为适合于当前道路视频检测的 YOLO9000-md 网络模型下车辆多目标检测方法; 为验证检测方法的有效性和完备性, 对其模型进行了验证对比分析, 同时对视频车辆进行了检测实验, 实验结果表明, 基于 YOLO9000-md 网络结构的车辆多目标检测方法, 验证检测正确率达到了 96.15%, 具有一定的鲁棒性和通用性, 为今后进行基于视频的更加智能化和自动化的数据分析提供了可靠的理论依据。

**关键词:** 智能交通; 目标检测; 网络模型; 正确率

## A Detection Method of Multi-target for Vehicles Based on YOLO9000 Model

Li Pengfei, Liu Yao, Li Xun, Zhang Hongwei

(School of Electronic Information, Xi'an Polytechnic University, Xi'an 710048, China)

**Abstract:** A multi-vehicle detection method was proposed, which consists of YOLO9000 under Darknet framework. The YOLO9000 structure was improved according to the training results and vehicle target characteristics, the algorithm parameters are adjusted. Finally, The YOLO9000-md network model was obtain, which is found more suitable for road vehicles detection. In order to verify the validity and completeness of this method, the model was verified and contrasted. At the same time, the vehicles under video was tested. The experimental results show that, the accuracy rate reaches 96.15% based on the YOLO9000-md model, which has certain robustness and versatility. It provides a reliable theoretical basis for more intelligent and automated data analysis based on video in the future.

**Keywords:** Intelligent traffic; target detection; network model; correct rate

## 0 引言

针对现如今高交通流的信息发展时代下, 道路交通拥堵问题及车辆信息识别的需求, 智能交通系统<sup>[1]</sup>成为解决此类问题的有效手段。在此背景下, 我们必须对车辆的微观信息进行有效采集。车辆检测是智能交通系统数据前端采集的一个重要环节<sup>[2]</sup>, 同时作为目标检测技术的一个具体应用方向, 基于机器学习的车辆检测更是近年研究热点。而深度学习<sup>[3]</sup>作为机器学习中重要的一部分, 在未来车路协同环境和微观交通对象信息提取的工程应用中, 在进一步提高检测率、鲁棒性的车辆多目标实时检测系统中提供了良好的技术基础。

文献 [4] 中, Szegedy 等人采用深度学习的方法在 VOC 2007 数据集测试结果中使 mAP 达到了 30%; R-CNN<sup>[5]</sup>则将传统机器学习和深度学习结合起来, 有效的将 mAP 提升至 48%; 不久 Uijlings 等人通过修改网络模型的

结构将 mAP 提高到 66%。随后通过进一步优化目标检测网络模型结构, 出现了 SPP-Net<sup>[6]</sup>、Fast R-CNN<sup>[7]</sup>、Faster R-CNN<sup>[8]</sup>、YOLO<sup>[9]</sup>等, mAP 变化如表 1 所示。YOLO 与传统依赖先验知识的特征提取算法不同, 它是采用深度卷积神经网络的全新实时目标检测方法, 对几何变换、形变和光照具有一定程度的不变性, 进一步的将 mAP 提升到了 78.6%, 改善了车辆外观多变带来的影响, 使检测系统具有更高的灵活性和泛化能力。

表 1 mAP 变化

Detection Frameworks	Train	mAP
R-CNN	2007+2012	48%~66%
Fast R-CNN	2007+2012	70%
Faster R-CNN	2007+2012	73.2%~76.4%
YOLO v2	2007+2012	78.6%

本文针对车辆多目标检测系统中, 车辆外形、车体结构以及道路场景复杂等变化导致的检测率低的问题<sup>[10]</sup>, 利用 Darknet 深度学习框架的 YOLO9000 算法, 对本文构建的多目标车辆 VOC 数据集进行特征训练学习, 进一步对网络模型进行改进和参数的调整, 最终得到本文改进的车辆多目标实时检测模型。

## 1 YOLO9000 算法

YOLO9000 是采用 YOLO v2 + join training 算法, 在

收稿日期: 2018-12-28; 修回日期: 2019-03-01。

基金项目: 陕西省自然科学基金基础研究计划项目(2016JQ5106); 陕西省教育厅专项科研项目(16JK1342); 西安工程大学控制科学与工程建设经费资助项目(107090811); 西安工程大学研究生创新基金项目(chx201816)。

作者简介: 李鹏飞(1962-), 男, 西安市人, 博士, 教授, 主要从事机器视觉和工业检测方向的研究。



Ubuntu16.04, 深度神经网络参数配置平台为 Darknet 框架, 整个开发环境为 Ubuntu16.04+Opencv 3.1.0+CUDA 8.0+CUDNN5.0, 配合应用基于深度卷积神经网络的 GPU 加速库 CUDNN, 对数据进行快速训练及实时性验证。

基于 CNN<sup>[13]</sup> 的车辆检测方法需要从大量样本中学习车辆特征, 若样本集不具有代表性, 很难选择出好的特征。为保证数据的多样性, 本文分时段对同一目标路段 (西安市碑林区金花南路 19 号西安工程大学立交桥) 进行图像采集。同时, 根据光照和车流密度的不同, 文章采集了早上 6:00-7:00 的自由流 (车辆数 < 300 辆/小时)、9:00-11:00 的同步流 (300 ≤ 车辆数 ≤ 900 辆/小时)、7:30-8:30 的阻塞流 (900 < 车辆数 < 1300 辆/小时) 等<sup>[14]</sup> 车流密度条件下的三组数据组合成一个混合样本 (由训练集和验证集组成), 获取样本如图 4 所示 (部分)。将训练样本按 80% 和 20% 的比例随机分开, 用于模型的训练和模型性能的验证 (见 2.2.1 节)。

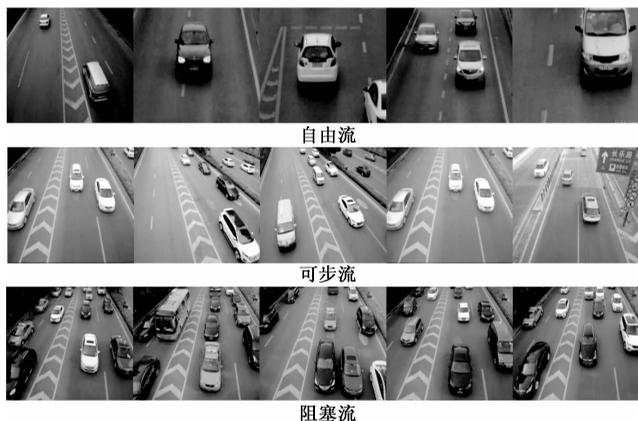


图 4 混合数据样本

## 2.2 实验结果分析

### 2.2.1 训练结果验证分析

本文使用在 ImageNet 上预训练得到的网络模型对 YOLO9000 和 YOLO9000\_md 模型进行参数调整, 将我们制作的车辆 VOC 数据集导入模型, 在 GPU 加速情况下训练 1 小时后, 获取了迭代 20000 次之后的检测模型。我们将初始学习率设为 0.001 和 0.0001, 在迭代 2000 次、6000 次和 15000 次时以之前的 0.1 倍来改变学习率, 保存迭代 10000 次和 20000 次的不同权重模型, 将训练结果进行实验对比分析。

#### 1) Loss 曲线分析:

图 5 为 YOLO9000 模型和 YOLO9000\_md 模型训练时的 Loss 曲线图。子图 (a) 是 YOLO 9000 模型以 0.0001 的初始学习率在迭代 1000 次时的 Loss 曲线图, 其训练开始由于学习率过小发散比较严重, 在迭代接近 400 次逐渐趋于收敛; 子图 (b) 是将学习率调整为 0.001 后的 Loss 曲线图, 对比图 (a) 明显训练开始抖动降低, 并且在迭代 200 次时就迅速收敛, 但是两条曲线开始摆动都处于发散状态,

且 Loss 最大值都达到了 2000 多; 子图 (c) 是 YOLO9000\_md 模型以 0.001 的初始学习率训练时的 Loss 值, 训练时一直处于收敛状态比较平稳, Loss 值一直在 300 以内, 并且在迭代 400 次后逐渐无限趋于 0。虽然图 (b) YOLO9000 在收敛速度上具有优势, 但在检测实验过程汇总中发现, 前期收敛速度的微小差别对后期检测效果影响较小。因此做了进一步的验证。

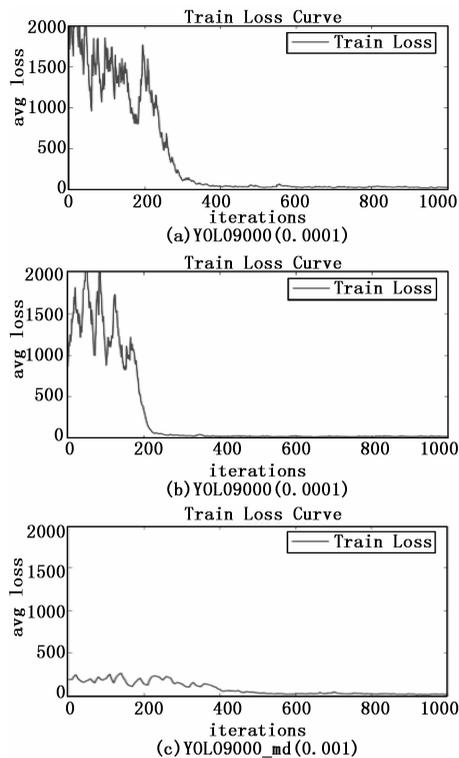


图 5 Loss 曲线图

#### 2) 准确性分析:

我们进一步对验证集进行测试的结果如表 1 所示。其中, Total 表示实际有多少个目标包围框, 即待检测实际目标个数; Correct 表示正确的识别出了多少个包围框, 就是我们在检测一张图片时, 网络模型会检测出很多目标包围框, 每个目标包围框都有其置信概率, 概率大于阈值的包围框与实际的目标包围框计算 IOU, 找出 IOU 最大的包围框, 如果这个最大值大于预设的 IOU 的阈值, 那么 Correct 就加一; Proposal 表示所有检测出来的包围框中, 大于阈值的包围框的数量; Precision 表示精确度如式 (3) 所示; Recall 表示召回率, 是检测出车辆的个数与验证集中所有车辆个数的比值, 如式 (4) 所示:

$$Precision = \frac{Correct}{Proposal} \quad (3)$$

$$Recall = \frac{Correct}{Total} \quad (4)$$

由表 2 可知: 在验证 364 个目标时, 以 0.0001 的初始学习率得到的 YOLO9000 模型可准确检测出 314 个车辆目标, 其 Precision 只有 76.77%, Recall 值是 86.26%; 以

0.001 的初始学习率得到的模型其 Precision 值提升到了 89.03%，同时 Recall 提高到了 95.88%；改进模型 YOLO 9000\_md 以 0.001 的初始学习率训练后得到的验证结果显示，其 Precision 和 Recall 值同时提升了 0.2% 以上，达到了很好的折中效果。

表 2 测试结果

模型(学习率) 参数	Total	Correct	Proposal	Precision /%	Recall /%
YOLO 9000 (0.0001)	364	314	409	76.77	86.26
YOLO 9000 (0.001)	364	349	392	89.03	95.88
YOLO 9000_md (0.001)	364	350	392	89.29	96.15

### 3) 检测结果分析:

为进一步验证模型在实际交通视频检测中，训练的不同权重模型对检测结果的影响。本文将 YOLO9000 模型以 0.001 的初始学习率在经过半个小时的训练后得到了迭代 10000 次的权重模型，一小时后得到迭代 20000 次的权重模型，将不同的权重模型进行车辆多目标的检测，对检测结果进行了直观的对比，如图 6 所示。

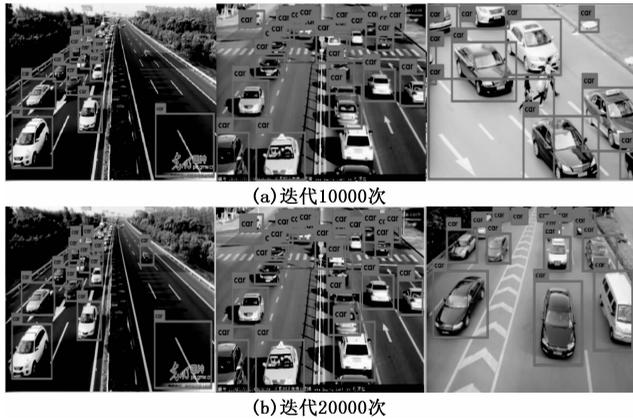


图 6 YOLO9000 模型

对比图 6 实验结果，YOLO9000 迭代 20000 次检测效果优于 10000 次结果。子图 (a) 的第一张右后方、第二张中间下方对比子图 (b) 存在明显的漏检情况；第一张右下方将背景误检为了车辆目标，第三张中右后方将地面杂物也误检为了车辆；这是由于预测框与检测出的标记框即 IOU 数值存在较大差异。分析可知，迭代次数在同等条件下较多的迭代次数，将获得更好的检测效果，但对于资源的耗费将更多。

#### 2.2.2 视频下的多目标实时检测结果分析

因 YOLO 方法受训练样本影响较大，需要样本多样性且具有代表性，在应用本文采集的组合不同车流密度得到的样本训练模型之后，为验证其在交通视频下检测结果都具有准确性，本文根据之前实验结果（详见 2.2.1），以 0.001 的初始学习率用 YOLO9000 和改进 YOLO9000\_md 模型将采集到的视频图像进行实时检测如图 7 所示，最后

得出其测试结果的混淆矩阵。混淆矩阵主要用于比较分类结果和实际检测结果，可以把分类结果的比例显示在一个混淆矩阵里面，进而分析图像检测分类的准确程度。结果如图 8、图 9 所示。



图 7 视频检测结果

		Predicted	
		object	background
Actual	object	2738	142
	background	303	—

图 8 YOLO9000 模型

		Predicted	
		object	background
Actual	object	2743	127
	background	291	—

图 9 YOLO9000\_md 模型

在交通视频的检测中，可以很好地检测出视频中的车辆，如图 8 所示。我们一共提取到了 673 张图像，一共包括 2880 个待检测目标，由图 9 可知，YOLO9000 模型将背景检测为车辆的有 142 个，也就是误检；而将车辆目标没有检测出来 303 辆，也就是漏检；最终正确检测出车辆目标 2738 个。相比图 8，图 9 改进 YOLO9000\_md 模型正确检测出目标 2743 个，比 YOLO9000 模型的正确检测出的目标多了 5 个，同时对做到了很好的处理误检和漏检情况做了很好处理，情况也有所减少。因此，针对视频下的多车辆目标实时检测，本文训练获得的 YOLO9000\_md 模型具有更好的检测效果。

### 3 总结

本文针对城市多车道中多目标视频检测问题，利用深度学习的 Darknet 框架下的 YOLO9000 目标检测算法，通过对 YOLO9000 模型参数的分析，根据车辆目标特征及其运动特征，进行不同网络结构的训练和对比，提出了一种能够获得较准确检测结果的 YOLO9000\_md 网络模型，该模型将目标检测问题转换为目标的二分类问题，通过大量