

文章编号: 2095-2163(2022)12-0192-05

中图分类号: TP391.41

文献标志码: A

改进 YOLOv4 的车辆图像检测算法研究

刘瑞峰, 孟利清

(西南林业大学 机械与交通学院, 昆明 650224)

摘要: 针对自动驾驶车辆真实行驶场景下因环境复杂、车辆间目标遮挡、环境背景遮挡等导致的车辆检测误检、漏检和定位不准的问题,本文提出了一个改进 YOLOv4 模型的车辆检测算法。该算法在 YOLOv4 网络的 Backbone 与 Neck 的通道处以及 Neck 的上采样与下采样处分别添加 7 处 CBAM 注意力机制,以提升网络提取有效特征的能力。并利用 k-means 聚类算法生成适合数据集的锚框。为检验模型的有效性,对数据集进行重新整理与划分,将与车辆无关的种类删去,将 Car、Bus、Truck 三类合并为 Vehicle 一类,随后进行实验,并与当前主流的其他目标检测模型进行对比。实验证明,改进的 YOLOv4 算法比原算法 AP 提升了 4.8%,准确率提升了 4.54%,召回率提高了 0.9%,优于大部分主流算法。提出的模型为复杂环境下自动驾驶领域的车辆识别提供了有效方法。

关键词: YOLOv4; 车辆检测; 注意力机制; 聚类算法; 目标检测

Research on vehicles image detection algorithm for improved YOLOv4

LIU Ruifeng, MENG Liqing

(College of Machinery and Transportation, Southwest Forestry University, Kunming 650224, China)

[Abstract] For the problem of vehicles misdetection, missed detection and inaccurate localization caused by complex environment, inter-vehicle target occlusion and environmental background occlusion in real driving scenarios of autonomous vehicles, a vehicle detection algorithm with improved YOLOv4 model is proposed in this paper. The algorithm adds seven CBAM attention mechanisms at the channels of Backbone, Neck of the YOLOv4 network and at the upsampling and downsampling of Neck, respectively, to improve the ability of the network to extract effective features. And the k-means clustering algorithm is used to generate the anchor frames suitable for the dataset. In order to test the effectiveness of the model, the dataset is rearranged and divided, the categories unrelated to vehicles are deleted, and the three categories of Car, Bus and Truck are combined into one category of Vehicle, then experiments are conducted and compared with other current mainstream target detection models. The experiments demonstrate that the improved YOLOv4 algorithm improves AP by 4.8%, accuracy by 4.54%, and recall by 0.9% over the original algorithm, which is better than most mainstream algorithms. The proposed model provides an effective method for vehicle recognition in the field of autonomous driving in complex environments.

[Key words] YOLOv4; vehicles detection; attention mechanism; clustering algorithm; target detection

0 引言

近年来,随着经济和社会的发展,机动车作为一种方便的交通工具得到大规模的使用,截至 2022 年 3 月底,全国机动车保有量达 4.02 亿辆。然而众所周知的是,机动车辆事故也可能会给社会及个人带来严重危害。基于此,无人驾驶汽车的开发和推广对保障行车安全、提高通行效率发挥着重要作用^[1]。自动驾驶日益受到广泛关注。目标检测作为自动驾驶环境感知中图像方面的关键技术,也吸引了众多学者参与研究^[2]。丁乐乐^[3]提出了一种深度学习与强化学习相结合的方法来对大巴车、轿

车和货车进行分类。然而,由于缺乏有效训练数据集,即需进一步证明模型的适应性。余胜等人^[4]提出了复杂场景下车辆识别方法,对漏标注、误标注的车辆进行重新标注,但是却没有关注到自动驾驶场景下的遮挡问题。在自动驾驶视角下,车辆目标往往会存在如目标间遮挡、环境背景遮挡等现实问题^[5],会直接影响目标检测的准确度。在真实环境下对车辆检测的错误,会对自动驾驶造成不小的干扰,甚至引发后续安全问题。

因此,对于真实场景下因车辆间遮挡、环境背景遮挡而导致的车辆检测中产生的误检、漏检等问题,亟需研发能够在复杂场景中对车辆进行较为准确检

基金项目: 云南省教育厅科学研究基金项目(2022Y572)。

作者简介: 刘瑞峰(1998-),男,硕士研究生,主要研究方向:图像识别;孟利清(1963-),男,博士,教授,主要研究方向:图像识别。

通讯作者: 孟利清 Email: 1341288094@qq.com

收稿日期: 2022-10-28

测的车辆检测算法。研究可知, YOLOv4 算法有着检测准确率高和检测速度快的优点, 能满足车辆检测对实时性和准确度的要求^[6]。但是 YOLOv4 算法应用于车辆检测时, 也存在因复杂场景下车辆重叠遮挡和环境背景遮挡问题导致的误检和漏检问题^[7]。

因此, 本文提出了一个改进的 YOLOv4 算法, 在 YOLOv4 网络的 Neck 部分添加了 CBAM 通道和空间注意力机制。并采用 k-means 均值算法, 生成合适的锚框应用于该网络, 提升算法对真实环境下不同尺寸车辆的识别率。以期为自动驾驶车辆识别提供一个有效的方法。

1 改进的 YOLOv4 算法

1.1 YOLOv4 算法

YOLOv4 的输入为固定大小。主干网络 (Backbone) 与 YOLOv3 算法相比, 使用了 CSPDarkNet 网络。CSPDarkNet 网络的整体 CSP 结构是将原来的网络分为 2 部分。一部分是原来的残差结构, 另一部分通过运算后直接与前一部分的运算结果结合, 有效提升了网络的学习能力。YOLOv4 网络在保持一定轻量化的同时, 拥有较高的识别精度。主干网络与颈部网络通过 3 个尺寸的通道头进行连接。3 种尺度用以检测大、中、小型的物体。网络的颈部 (Neck) 采用了 PANet 结构, 用来进行特征融合。为 YOLOv4 网络还提出了 moasic 数据增强方法, 使网络能够在有限的数据集上得到更好的效果。

1.2 注意力机制

CBAM 注意力机制包括通道注意力机制 (Channel Attention Module, CAM) 和空间注意力机制 (Spatial Attention Module, SAM)^[8]。CBAM 模块中, 通道注意力模块的输出作为空间注意力模块的输入。CAM 模块中, 输入特征图先经过一个全局平均池化和全局最大池化层, 再经过 Sigmoid 激活操作, 生成通道注意力特征 M_c , 再与该模块的输入相乘, 得到空间注意力模块需要的特征图。通道注意力特征 M_c 计算公式可表示为:

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPooling(F))) \quad (1)$$

其中, “ σ ” 为 Sigmoid 操作。

SAM 空间注意力模块中, 输入特征图同时进入全局平均池化和全局最大池化, 通过 2 个池化层进

行通道拼接, 经过 $7 * 7$ 卷积层卷积操作后输入 Sigmoid 层, 由此得到的空间注意力特征 M_s 与该子模块的输入做乘法, 得到最终生成的特征。空间注意力特征计算 M_s 可由式 (2) 进行描述:

$$M_s(F) = \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) \quad (2)$$

1.3 k-means 算法

聚类算法是能够通过数据的特点, 将具有相似特征的数据分为同类的算法^[9]。k-means 算法是通过选定 K 个中心点, 再计算数据中每个例子到中心点的欧氏距离, 将距离相近的数据划分到一类。分类后, 重新计算同类数据的中心点, 得到 K 个新的中心点后重复上述过程。当各类别中的数据划分不再变化时, 最终确定 K 个中心点。k-means 聚类算法的损失函数具体如下:

$$J(c, \mu) = \min \sum_{i=1}^M \|x_i - \mu_{c_i}\|^2 \quad (3)$$

将 K-means 聚类算法引入到锚框的选定上, 可以提高检测的准确率^[10]。通过 k-means 算法计算锚框时, 需要引入新的损失函数来评价锚框。采用 IoU 的损失函数见式 (4):

$$D(i, c) = 1 - IoU(i, c) \quad (4)$$

其中, i 为目标真实框的长宽; c 为中心点的长宽; IoU 是计算聚类中心与真实框交并比的过程。交并比是 2 个框在图像上的交集与并集之比, 交并比越接近 1, 说明 2 个框的重合程度越高。因此, 用 1 减去交并比的差来作为本文聚类算法的损失函数, 能够较为合理地评价锚框对于指定数据集的优劣。本文由 BBD100K 数据集生成的 anchor box 尺寸大小为 $[6, 2]$, $[7, 5]$, $[10, 4]$, $[11, 7]$, $[17, 6]$, $[23, 11]$, $[42, 17]$, $[78, 31]$, $[165, 72]$ 。

1.4 改进的 YOLOv4 算法

本文主要对 YOLOv4 算法的网络结构、锚框生成进行改进。网络结构上, 本文将 CBAM 注意力模块分别添加在 PANet 的 3 个输入后面, 再分别添加 4 个 CBAM 注意力模块到 2 个上采样和 2 个下采样模块上的输出上。注意力模块对输入的通道和空间位置进行加权后输出, 不改变特征图的维度。改进后的 YOLOv4 模块如图 1 所示。同时, 网络采用了由 k-means 聚类算法在数据集上计算出的锚框, 可以使锚框尺寸适合数据集中的物体。

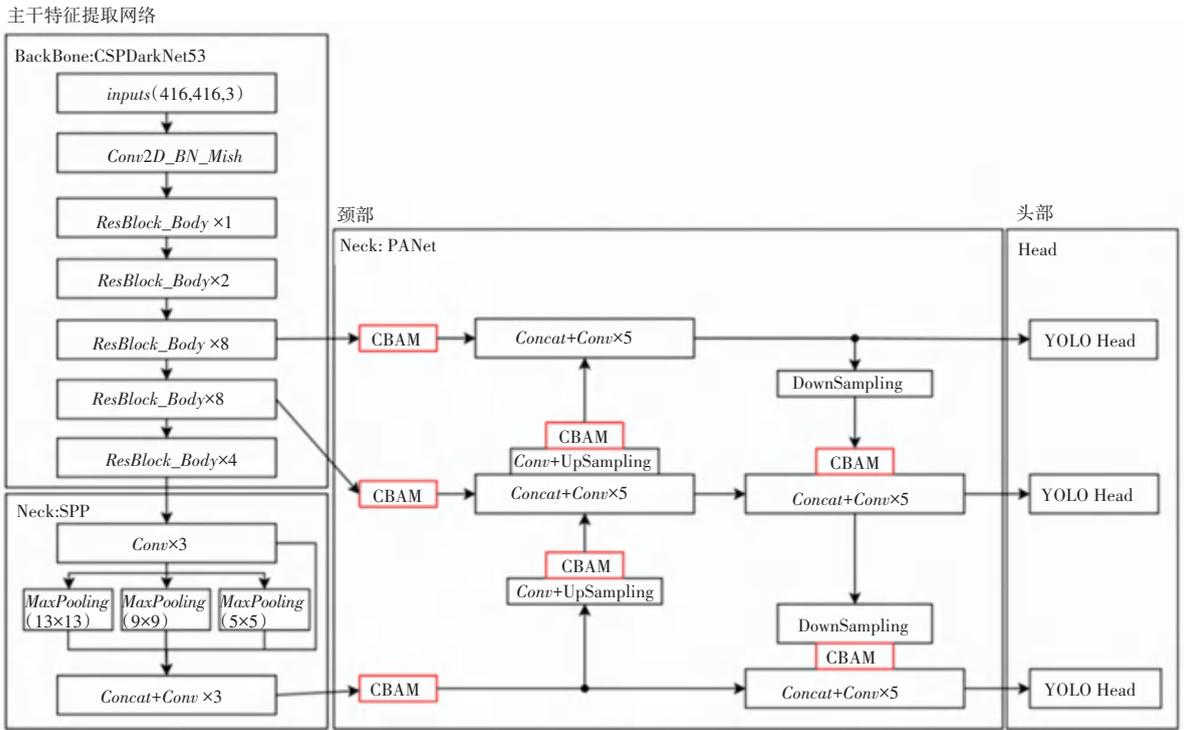


图1 改进的YOLOv4目标检测网络

Fig. 1 Improved YOLOv4 target detection network

2 实验

2.1 数据集和实验平台

数据集选用由加州大学伯克利分校发布的BBD100k数据集,这是自动驾驶领域规模最大,场景最多样的驾驶视频数据集之一。数据集中包含晴天、阴雨天、夜晚等各种车辆行驶中采集的图像数据。本文从中选取了10 000张图片作为数据集,其中2 000张为测试集,1 600张为验证集,6 400张为训练集。本文删去原数据集的pedestrian、rider、train、motorcycle、bicycle、traffic light、traffic sign、trailer类,将car、truck、bus合并为vehicle类。

实验平台为Windows10, GPU为RTX3070,显存为8 G; CPU为R9 5900hx,内存为16 G。

2.2 参数设置

消融实验中,优化器种类为Adam,初始学习率设为0.001,最小学习率为0.000 01, momentum设为0.937,学习率下降方式设为cos。epoch设为100,使用主干网络预训练权重,前50个epoch冻结Backbone训练,冻结时batch_size设为8,解冻阶段batch_size设为4。输入图片尺寸设为416×416,使用mosaic数据增强方法。

对比实验中,优化器为Adam,学习率下降方式

设为cos, epoch设为100,冻结训练50个epoch后进行解冻训练。其他参数因各算法网络结构不同,选择各模型推荐参数进行训练。

2.3 消融实验

为验证改进网络的效果,进行消融实验。改进前后网络在测试集上的表现见表1。由表1可知,添加了注意力机制后的网络在数据集上的准确率Precision达到了82.23%,比基线网络提高了4%。添加了k-means算法生成的锚框后,准确率达到了83.24,比基线网络提高了5.3%。添加了注意力机制并采用k-means算法生成锚框的模型与基线网络相比,准确率达到了82.62%,提升了4.54%;召回率达到了61.07%,提升了0.9%;AP值达到了70.52,提高了4.8%。

表1 消融实验

Tab. 1 Ablation experiments

网络名称	Precision/%	Recall/%	AP/%
YOLOv4	79.03	60.50	67.26
YOLOv4, and CBAM	82.23	58.51	66.96
YOLOv4, and k-means	83.24	57.64	68.30
YOLOv4, CBAM and k-means	82.62	61.07	70.52

2.4 多种检测方法对比

为检验方法的有效性,本文选取了YOLOv3、

Faster R-CNN、GhostNet-YOLOv4、YOLOv5s_v6.1 进行对比实验。首先使用上述网络分别在数据集中训练,然后使用测试集对上述训练过的网络进行测试,测试结果见表 2。更直观的实验结果如图 2 所示。具体来说,图 2(a)为车辆行驶数据原图,图 2(b)为 Faster R-CNN 算法检测结果、图 2(c)为 GhostNet-YOLOv4 算法检测结果,图 2(d)为 YOLOv5s_v6.1 算法检测结果,图 2(e)为 YOLOv3 算法的检测结果,图 2(f)为 YOLOv4 算法的检测结果,图 2(g)为改进 YOLOv4 算法的检测结果。

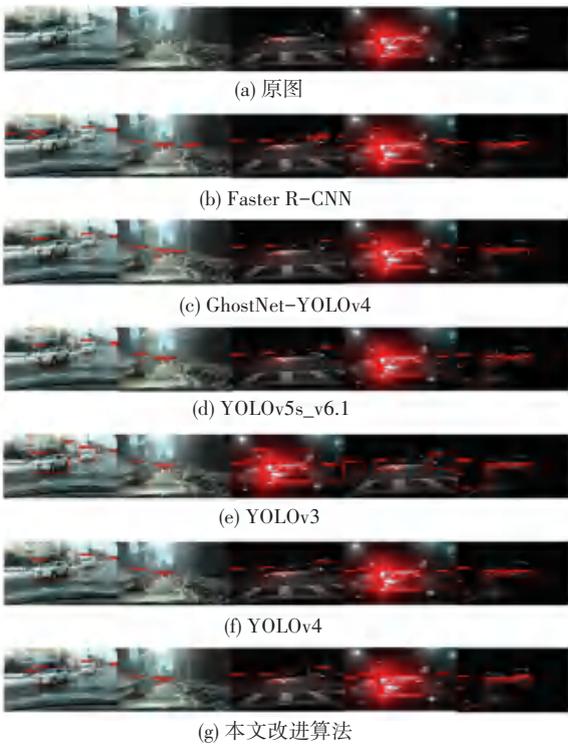


图 2 不同目标检测算法与改进算法实验结果

Fig. 2 Experimental results of different target detection algorithms and improved algorithms

表 2 多种模型对比评价

Tab. 2 Comparative evaluation of multiple models %

模型名称	Precision	Recall	AP
Faster R-CNN	39.30	66.84	59.83
GhostNet-YOLOv4	77.05	50.67	58.86
YOLOv5s_v6.1	92.54	48.41	65.50
YOLOv3	81.09	58.27	66.88
本文改进模型	82.62	61.07	70.52

从表 2 中可以看到, YOLOv5s 和本文的模型准确率最高, 分别达到了 92.54% 和 82.62%; Faster R-CNN 和本文模型召回率最高, 分别达到了 66.84% 和 61.07%; 本文算法的 AP 高于其他算法, 达到了 70.52%。

图 2 中, 图 2(g) 的第一张图片表明, 只有本文模型检测到了被树木遮挡的车辆, 其他模型均出现了漏检的情况; 对于图 2 的第 4 张夜晚的情景, 图 2(e) 和图 2(g) 中均检测到了左方的车辆, 而图 2(c) 和图 2(d) 均未能检测出左方车辆, 图 2(b) 虽然检测到了左方车辆, 但是却出现了误检情况。从图 2 中可以看出, 改进算法能够适应不同光照条件, 与文中的其他参照算法相比, 不易出现漏检和误检情况。由图 2(f) 和图 2(g) 可看到, YOLOv4 算法对被环境遮挡的车辆无法进行准确识别, 而改进算法在车辆被严重遮挡的情况下, 依然能够准确识别出车辆。

3 结束语

复杂外部环境下的车辆检测算法对于自动驾驶车辆的安全和实时控制至关重要。虽然已经取得了一些研究成果, 然而这些研究对于车辆拍摄角度下的车辆遮挡、环境遮挡问题还未能提供有效解决方案。针对这种情况, 本文对自动驾驶汽车行驶过程中的车辆图像检测问题展开研究, 选用了对于主流目标检测算法具有很大的挑战性的数据集 BBD100k, 数据集是由车辆在实际路面中采集而来, 其中包含了晴天、夜晚等各种行驶场景下的图像。从该数据集中选取 10 000 张图片作为实验的数据集, 并对原数据集中的类别进行合并和调整, 删去了和车辆无关的类别。本文选择 YOLOv4 网络作为基线网络进行改进, 在网络的 7 个位置添加了 CBAM 注意力机制, 添加通道注意力机制使网络忽略不重要的通道, 添加空间注意力机制使得网络忽略遮挡部分的特征, 提升网络对被遮挡车辆识别准确率。利用 k-means 聚类算法生成适合该数据集的锚框, 提升对不同尺寸大小车辆的检测能力, 并通过实验进行验证。实验证明, 改进的 YOLOv4 网络在该数据集上的 AP 值达到了 70.64%, 比基线网络提高了 5%, 证明改进的 YOLOv4 算法在自动驾驶中对车辆检测具有更高的精度, 该方法适用于复杂条件下的自动驾驶场景。

参考文献

- [1] 李柏. 复杂约束下自动驾驶车辆运动规划的计算最优控制方法研究[D]. 杭州: 浙江大学, 2018.
- [2] WANG Hongbo, HOU Jiaying, CHEN Na. A survey of vehicle re-identification based on deep learning[J]. IEEE Access, 2019, 7: 172443 - 172469.
- [3] 丁乐乐. 基于深度学习和强化学习的车辆定位与识别[D]. 成都: 电子科技大学, 2016.