

基于改进 YOLOv3 模型的交通视频目标检测算法研究

梁秦嘉, 刘 怀, 陆 飞

(南京师范大学电气与自动化工程学院, 江苏 南京 210023)

[摘要] 针对现有的基于 YOLOv3 的目标检测算法在多尺度目标检测上存在速度与精度难以平衡的问题, 在已有算法的基础上改进形成新的 YOLOv3 多尺度目标检测算法. 该算法首先通过 k-means++ 聚类为各个尺度选择候选锚框的数量和长宽比维数, 有效降低原始算法在初始聚类点所造成的聚类偏差; 其次将 YOLOv3 的检测尺度从 3 扩展到 4, 以提高对不同尺度下目标检测的精度; 最后为避免梯度衰落, 将检测层前的 6 个卷积层转换为 2 个残差单元. 在 UA-DETRAC 数据集上的实验结果表明, 该方法比原始 YOLOv3 的准确率和召回率分别提高了 7.91% 和 4.57%, 同时此算法的处理速度可实现对交通视频的实时处理.

[关键词] 交通监控, 目标检测, 深度学习, 卷积神经网络

[中图分类号] TP391.4 **[文献标志码]** A **[文章编号]** 1672-1292(2021)02-0047-07

Traffic Video Target Detection Algorithm Based on Improved YOLOv3

Liang Qinjia, Liu Huai, Lu Fei

(School of Electrical and Automation Engineering, Nanjing Normal University, Nanjing 210023, China)

Abstract: Aiming at the problem that the existing target detection algorithm based on YOLOv3 is difficult to balance the speed and accuracy of multi-scale target detection, the paper proposes an improved YOLOv3 multi-scale target detection algorithm. The algorithm firstly selects the number of candidate anchor frames and the aspect ratio dimensions for each scale through k-means++ clustering, which effectively reduces the clustering deviation caused by the original algorithm at the initial clustering points. Secondly, the YOLOv3 detection scale is extended from 3 to 4, in order to improve the accuracy of target detection under different scales. Finally, in order to avoid gradient fading, the six convolutional layers before the detection layer are converted into two residual units. The experimental results on UA-DETRAC dataset show that the accuracy and recall rate of this method are 7.91% and 4.57% higher than the original YOLOv3, respectively. Meanwhile, the processing speed can meet the real-time requirements of traffic video.

Key words: traffic monitoring, target detection, deep learning, convolutional neural network

随着城市建设的不断发展, 公共设施逐渐完善, 城市道路上交通流量也急剧增长. 如何对交通图像或视频等监控数据中的各种目标进行准确的检测已成为城市交通管理所面临的主要挑战^[1]. 近年来, 由于图像处理单元的性能不断提高, 使得深度学习方法在目标检测领域取得了广泛的应用^[2], 如行人检测^[3]、人脸检测^[4]、航拍目标检测^[5]等.

目前, 以神经网络算法为基础的目标检测模型主要有两类. 一类是基于目标候选框^[6]的模型, 如 R-CNN^[7]、Fast-R-CNN^[8]、Faster-R-CNN^[9]等, 这类模型采用选择性搜索方法, 寻找目标候选区域, 然后针对候选区域采用模型进行进一步的处理, 虽然检测精度较高, 但速度较慢^[10], 无法达到实时性要求. 另一类是基于端到端^[11]的模型, 如 SSD^[12]、YOLOv2^[13]、YOLOv3^[14-15]等, 此类算法直接通过卷积神经网络的全局特征判断目标的位置和种类, 因此检测的速度更快, 但精度略低^[16]. YOLOv3 网络模型在目标检测的精度与速度方面均有良好的表现, 因此在视频目标检测领域得到了广泛的应用^[17], 但由于其网络结构的限制, 导致原始算法较难适应视频中目标在运动时产生的尺度差异, 易出现误检和漏检, 以及对小目标检测不准确的问题.

收稿日期: 2020-08-23.

基金项目: 国家自然科学基金项目(61603194).

通讯作者: 刘怀, 博士, 副教授, 研究方向: 数字图像处理、实时控制系统. E-mail: liuhuai@njnu.edu.cn

本文提出一种基于改进 YOLOv3 网络模型的交通视频目标检测算法. 首先, 针对交通视频中的目标在运动时会发生尺度变化的问题, 将原始 YOLOv3 的检测尺度从 3 个扩展到 4 个; 同时, 为了在不同尺度下选择合适的锚框大小, 对数据集采用 k-means++ 聚类算法进行重新聚类; 针对原始 YOLOv3 在检测阶段对视频中目标易出现漏检及误检的问题, 将每个尺度输出检测前的 6 个卷积层转换为 2 个残差单元, 提高对目标的检测准确率.

1 改进的 YOLOv3 算法

1.1 YOLOv3 算法

YOLO 是一种端到端的目标检测网络模型, 可通过神经网络直接判断目标的所属类别并标示具体位置, 故其检测速度较快. YOLOv3 是 YOLO 系列的第三代算法, 其网络结构如图 1 所示.

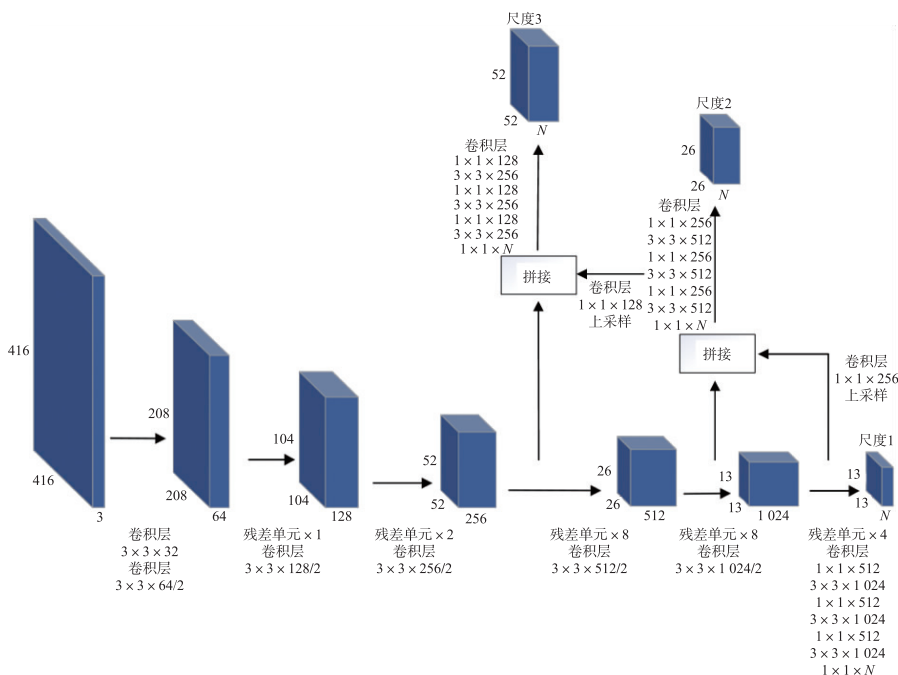


图 1 YOLOv3 网络结构图

Fig. 1 The network of YOLOv3

为了进行特征提取, YOLOv3 使用了连续的 3×3 和 1×1 卷积层, 并借鉴了残差网络的思想, 设置了 5 个残差块, 每个残差块由多个残差单元组成. 残差单元结构^[18]是指在卷积神经网络的卷积层外部加入中间层实现与越层之间的快捷连接形成残差模块, 从而参与后续神经网络的构成. 假设残差单元的输入为变量 x , 其要拟合的最终输出也即为原始映射函数 $L(x)$, 同时定义残差映射函数 $F(x) = L(x) - x$, 则有 $L(x) = F(x) + x$. 这样即可认为原始映射函数等于前馈网络中主路径的残差映射函数 $F(x)$ 与输入 x 两者相加. 与越层的快捷连接除改变函数计算公式外未引入其他参数, 也不会使神经网络算法的计算速度变慢. 残差单元的结构如图 2 所示.

YOLOv3 对输入图像进行了 5 次下采样, 在预测阶段采用最后 3 层下采样的结果来预测 3 个尺度的目

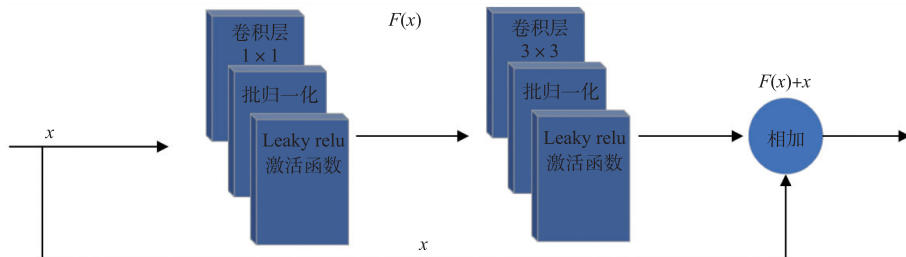


图 2 残差单元结构图

Fig. 2 The structure of the residual unit

标. 在尺度 3 上,使用 8 倍下采样的特征图检测小目标;在尺度 2 上,采用 16 倍下采样的特征图检测中等大小的目标;在尺度 1 上,使用 32 倍下采样的特征图检测大目标. 对于卷积神经网络,网络浅层所提取的目标特征的细节信息较多,所检测到的目标位置会更为准确,因此检测的精度相对较高;而网络深层可以提取目标的语义信息,提取到的特征更全面,但目标检测精度较低,因此采用特征融合来检测目标. 为了将浅层与深层所检测到的特征进行融合,YOLOv3 模型借鉴特征金字塔(feature pyramid networks, FPN)的结构,通过上采样的方式对更深一层的特征图进行大小调整,则不同尺度的特征图将拥有相同的大小. 将来自较早层的特征与来自较深层的特征合并在一起,因而 YOLOv3 对不同目标的检测都有很好的性能.

1.2 YOLOv3 算法的改进

1.2.1 网络结构的改进

为了进一步提高 YOLOv3 的性能,本文提出一种增强的目标检测算法. 原始 YOLOv3 使用 3 种尺度检测不同大小的目标,特征层大小分别设定为 13×13 、 26×26 、 52×52 ,但在交通视频中由于行人及各种交通工具的目标数量较多,运动时目标的尺度变化也较大,原算法不能将这些目标很好地检测出来. 本文将原始 YOLOv3 的 3 个检测结构增加为 4 个,根据网络结构将增加的特征层大小设定为 104×104 . 同时采用特征金字塔的思想,使用 YOLOv3 网络模型中的卷积层对图像进行多次卷积操作. 首先通过上采样的方法,将最底层采集到的 13×13 大小的特征层进行扩展,使其扩大为 26×26 大小的特征层,之后将其与上一个检测尺度的特征层进行融合,将结果再次输入回检测层中. 对上文设计的 4 个尺度进行相同的操作处理,实现 4 个不同的尺度融合. 本文设计了 4 个检测尺度,将浅层获取到的目标细节信息和深层获取到的高级语义信息进行特征的拼接融合,使得网络可以获得更为准确的目标信息,同时保持网络的复杂度,以更好地适用于视频场景中复杂环境下的多目标多尺度检测.

对 N 层的卷积神经网络,其训练过程中的网络梯度变化为:

$$\frac{\partial L_{\text{loss}}}{\partial X_1} = \frac{\partial F_M(X_{N_M}, \mathbf{W}_{N_M}, b_{N_M})}{\partial X_N} \times \dots \times \frac{\partial F_2(X_{N_2}, \mathbf{W}_{N_2}, b_{N_2})}{\partial X_1}, \quad (1)$$

式中, L_{loss} 为计算损失; X_N 为任意层的输入; \mathbf{W}_N 为该层卷积矩阵; b_N 为该层偏置项; $F_M(X_{N_M}, \mathbf{W}_{N_M}, b_{N_M})$ 表示第 M 层的计算损失; X_{N_M} 为第 M 层的输入; \mathbf{W}_{N_M} 为第 M 层卷积矩阵; b_{N_M} 为第 M 层偏置项. 根据式(1)可知,当网络逐渐加深时,梯度会逐渐衰落变小甚至消失.

在 YOLOv3 网络的目标检测层之前,有 6 个卷积层. 当卷积层数较多时,易出现梯度衰落甚至消失的情况,导致检测精度降低. 由于残差结构可有效避免梯度衰落情况的发生,将卷积层转化为残差结构后,网络的梯度变化如式(2)所示:

$$\frac{\partial L_{\text{loss}}}{\partial X_1} = \left[1 + \frac{\partial F_M(X_{N_M}, \mathbf{W}_{N_M}, b_{N_M})}{\partial X_N} \right] \times \dots \times \left[1 + \frac{\partial F_2(X_{N_2}, \mathbf{W}_{N_2}, b_{N_2})}{\partial X_1} \right]. \quad (2)$$

受残差网络结构的启发,同时考虑网络层数,本文将 6 个卷积层转换为 2 个残差单元,如图 3 所示. 最终改进后的网络结构如图 4 所示.

1.2.2 初始锚框的选择

借鉴 Faster R-CNN, YOLOv3 中引入了锚框机制. 锚框即初始候选框,其宽度与高度是固定的,选择合适的初始候选框对目标检测的精度和速度有很大的影响. 原始 YOLOv3 模型是通过 k-means 聚类算法在数据集上自动找到好的先验框,但由于在选择初始的聚类点时, k-means 聚类算法有比较大的随机性,可能会导致较大的聚类偏差. 针对此问题,本文采用 k-means++ 聚类算法,对本文所使用的数据集标注框进行重新聚类分析,以减小因随机性而带来的聚类偏差.

为了减小锚框大小对检测的影响,本文采用平均交并比 AvgIOU 作为目标聚类分析的度量指标. AvgIOU 的目标函数如下:

$$I_{\text{IOU}} = \frac{\text{area}(\text{box}_{\text{Tra}} \cap \text{box}_{\text{Pre}})}{\text{area}(\text{box}_{\text{Tra}} \cup \text{box}_{\text{Pre}})}, \quad (3)$$

$$\text{AvgIOU} = \arg\max \frac{\sum_{i=1}^k \sum_{j=1}^{n_k} I_{\text{IOU}}(\text{box}, \text{centroid})}{n}, \quad (4)$$

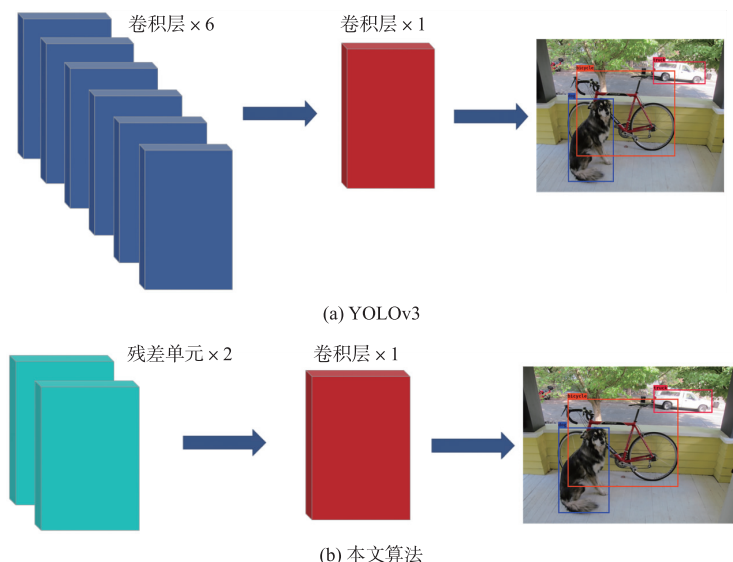


图 3 卷积层到残差结构的转换

Fig. 3 The conversion from convolution layer to residual structure

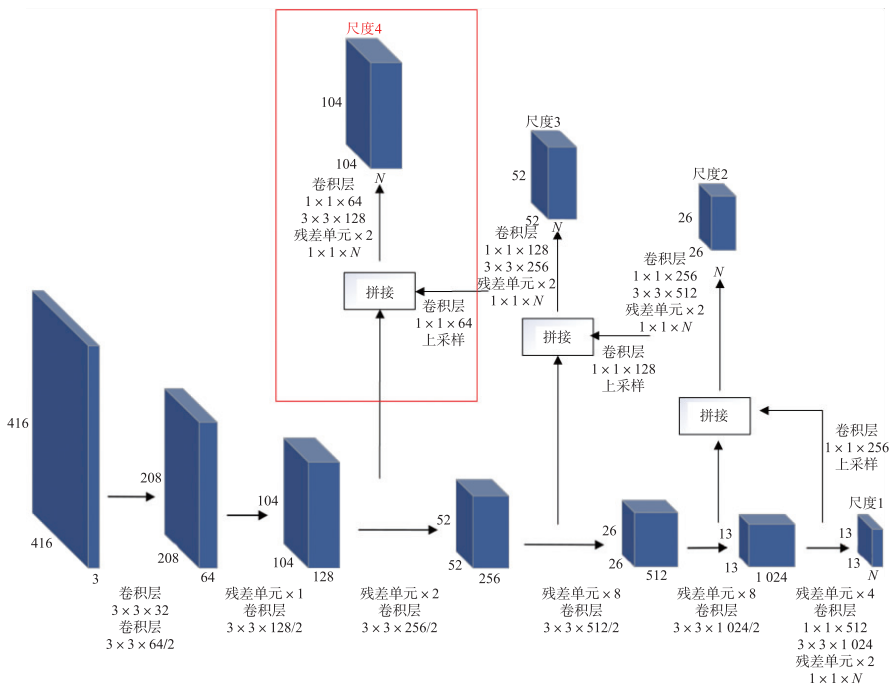


图 4 改进 YOLOv3 网络结构图

Fig. 4 The network of improved YOLOv3

式中, I_{iou} 表示真实框和预测框交集面积和并集面积的比值; box 为样本中目标框的尺寸; $centroid$ 为聚类的簇中心; n_k 是第 k 个聚类中心的样本数量; k 是聚类中心个数; n 为样本的总数量。

在 UA-DETRAC 数据集上应用 k-means++ 聚类。图 5 为不同 k 值下得到的 AvgIOU, 可以看出, 随着 k 的增加, 目标函数的变化越来越稳定。结合图 5, 并考虑 AvgIOU 和检测层的数量, 本文选择 12 个锚框。对数据集进行 k-means++ 聚类生成聚类中心后, YOLOv3 将这些聚类按比例平均划分。考虑到锚框个数给神经网络的复杂度和计算量带来的影响, 对不同尺度均选择 3 个锚框。通过 k-means++ 聚类, 得到各个锚框的宽和高维度的尺寸分别为 (21, 22), (32, 24), (26, 40), (45, 40), (63, 32), (62, 56),

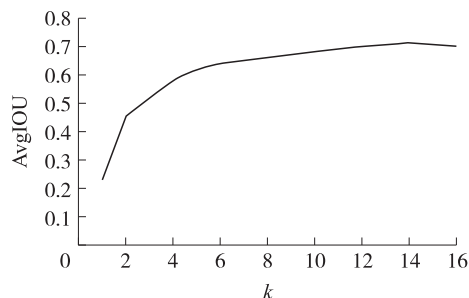


图 5 不同 k 值下的聚类分析结果

Fig. 5 Clustering analysis results under different k values

(73,60),(106,57),(131,118),(92,116),(200,130),(316,181).

为了检验锚框尺寸对检测速度和精度的影响,本文将原本 YOLOv3 的 9 个锚框重新聚类,命名为 YOLOv3-9anchors;同时,为了验证不同尺度个数对目标检测精度与速度的影响,在本文改进算法 4 个尺度的基础上增加第 5 个尺度,该检测尺度大小为 208×208 ,其他处理方式与本文相同,与 YOLOv3 中同大小的特征图相连,记为 YOLOv3-5scales,对该算法选择 15 个锚框.

2 实验结果及分析

2.1 实验数据集

为了验证改进算法模型的具体性能,本文以车辆目标为例,采用 UA-DETRAC 数据集对算法进行测试^[19].同时,为了验证在雾霾及阴雨等恶劣天气情况下算法的有效性,部分视频场景来源于自行拍摄.

2.2 评价指标

本文对多目标检测模型的评价指标主要有平均精度、召回率以及平均检测时间.其中,Precision 表示视频中所检测的每一种目标的检测精度;mAP 表示平均精度,即所有目标检测精度的平均值;平均时间为不同网络模型检测完成一幅图像时所花费的平均时间.精度和召回率的计算公式为:

$$\text{Precision} = \frac{TP}{FP+TP}, \quad (5)$$

$$\text{Recall} = \frac{TP}{FP+TN}, \quad (6)$$

假设将公共汽车设定为需要检测的目标,TP 表示正确检测的数量;FP 表示将其他类型如小型汽车等被错误检测为公共汽车的数量;FN 表示公共汽车被错误检测为其他目标类型的数量.

2.3 实验环境与训练

本文算法是在深度学习框架 keras 下进行的.实验的设施配置为:CPU 为 Intel i5-9400,主频 2.90GHz,16GB 内存,GPU 为 NVIDIA 1070,8GB 显存,操作系统为 windows 10.

本文对数据集进行重新标注训练,检测目标包含小型汽车(car)、公共汽车(bus)、大型货车(truck)三类.为了提高训练的效果,使用了不同角度旋转图像和改变图像的饱和度、曝光和色调等数据增强方法.在训练阶段,初始学习率为 0.001,权值衰减为 0.000 5.当训练批次为 60 000 和 70 000 时,学习率分别降至 0.000 1 和 0.000 01,使损失函数进一步收敛.由于本文算法并未使模型的复杂度明显增加,因此训练时间相比原始 YOLOv3 没有显著变化.

2.4 结果分析

利用改进的 YOLOv3 模型对 UA-DETRAC 数据集中的目标进行检测.同时对上文所提到的 YOLOv3-9anchors 与 YOLOv3-5scales 算法进行聚类以选择合适的锚框.针对 YOLOv3-9anchors,聚类后的锚框宽高分别为(19,21),(30,25),(27,39),(43,41),(62,33),(63,55),(71,58),(104,57),(128,119).针对 YOLOv3-5scales,聚类后的锚框宽高分别为(10,18),(12,23),(22,19),(19,31),(37,19),(31,29),(31,44),(44,39),(59,26),(57,53),(92,42),(83,74),(133,59),(142,108),(267,123).

不同目标检测算法的测试结果如表 1 所示.由表可知,改进 YOLOv3 网络模型的检测精度为 84.59%,相比原始 YOLOv3 网络模型,改进后算法的精度提高了 7.91%;召回率为 96.93%,相比 YOLOv3 提高了 4.57%.YOLOv3-tiny 的 mAP 值为 73.25%,召回率为 90.65%,均低于改进 YOLOv3 模型.基于不同主干网络 ResNet101 和 VGG16 的 Faster-R-CNN 算法的平均精度分别为 85.03%和 86.42%,比本文算法分别高了 0.44%、1.83%.在检测速度方面,改进 YOLOv3 对图像进行检测的平均用时为 30.72 ms,较 YOLOv3 稍有不足.在 YOLO 系列检测算法中,属于简化版本的 YOLOv3-tiny 因其主干网络浅,结构简单,检测速度较快,平均用时仅为 6.38 ms.而

表 1 不同算法实验结果对比

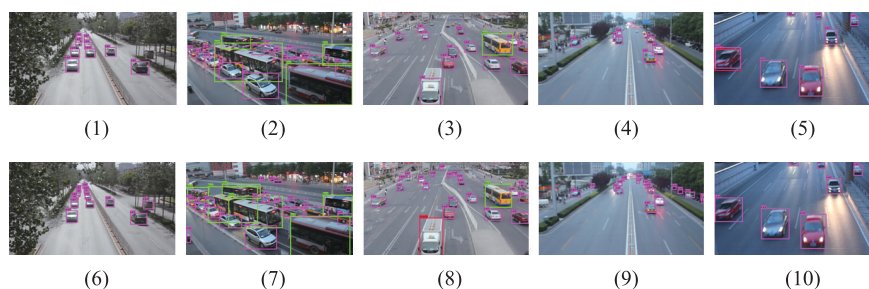
Table 1 Comparison of experimental results of different algorithms

网络	mAP/%	Recall/%	平均时间/ms
Faster-R-CNN(resnet101)	85.03	92.15	250.21
Faster-R-CNN(VGG16)	86.42	92.27	163.52
YOLOv3-tiny	73.25	90.65	6.38
YOLOv3	76.68	92.36	29.67
YOLOv3-9anchors	82.07	94.53	29.65
YOLOv3-5scales	83.56	96.21	41.25
改进 YOLOv3	84.59	96.93	30.72

Faster-R-CNN 系列算法在检测速度方面有明显的缺陷,平均时间这一指标远高于 YOLO 模型,无法满足交通视频的实时性要求. YOLO 模型中大部分用时在 30 ms 左右,检测速度大幅提升,可实现实时监测.

对目标框进行重新聚类后的 YOLOv3-9anchors 相较于原始算法的 mAP 和召回率分别提升了 5.39% 和 2.17%,表明初始锚框大小的选择对目标的检测精度有着较大影响. 而 YOLOv3-5scales 的精度和召回率虽然相较于原始 YOLOv3 模型提高了 6.88% 和 3.85%,与本文改进的 YOLOv3 算法精度相差无几,但检测速度相较于本文算法明显下降.

整体而言,改进 YOLOv3 模型在上述评价指标上均较 YOLOv3 更加出色. 由于采用了 k-means++ 聚类方法,使得目标候选框的选择更适合于图像类数据;将检测层前的卷积层转化为残差单元,避免了出现梯度衰落;同时检测尺度由 3 个增加至 4 个,提高了目标检测精度,可有效检测不同尺度的目标. 改进 YOLOv3 方法对于候选框的分类和选择恰当,并选择了合适的尺度,未使网络模型变得过于复杂,提高精度的同时也保持了较快的速度. YOLOv3 与改进 YOLOv3 模型在测试集上的目标检测对比效果如图 6 所示.



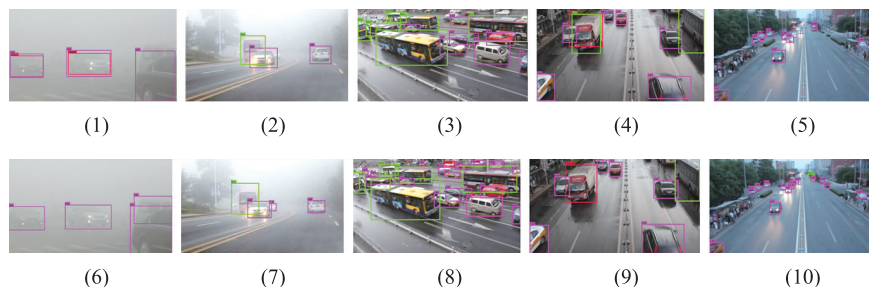
(1)~(5) 为 YOLOv3 检测结果, (6)~(10) 为改进 YOLOv3 检测结果

图 6 白天及夜间的情况下 YOLOv3 模型与改进 YOLOv3 模型算法检测效果对比

Fig. 6 Comparison of detection effect between YOLOv3 and the improved YOLOv3 algorithm in daytime and nighttime conditions

由图 6 可以看出, YOLOv3 对很多目标出现漏检及误检情况. 在场景 1 中, 虽然主车道上车辆均被检出, 但 YOLOv3 未检测出右侧树后的车辆, 改进 YOLOv3 准确地得出了检测结果. 在场景 2 中, YOLOv3 仅检测出主车道上的车辆, 改进 YOLOv3 对上方的小目标车辆及左侧出现的车辆局部均能正确检测. 在场景 3 中, YOLOv3 将视频下方的卡车错检为小型汽车, 且后方的的小目标均未检测, 改进 YOLOv3 正确检测出了卡车及远处的小目标. 在场景 4 中, YOLOv3 仅检测出近距离的几个目标, 改进 YOLOv3 对远处小目标及两侧有遮挡存在的目标均进行了正确检测. 在场景 5 中, YOLOv3 对最左侧车辆检测出 car 和 truck 两个结果, 改进 YOLOv3 可准确检测出每一个目标.

图 7 所示为恶劣天气情况下两种模型的目标检测对比图. 可以看出, YOLOv3 的检测精度依然较低. 在场景 1 中, YOLOv3 对左侧车辆检测出 car 与 truck 两个标签, 改进 YOLOv3 检测正确. 在场景 2 中, YOLOv3 仅检测到前方 3 个目标, 改进 YOLOv3 检测到了后方的车辆. 在场景 3 中, YOLOv3 仅能检测到距离较近的目标, 改进 YOLOv3 可以准确检测远处小目标. 在场景 4 中, YOLOv3 对画面中的卡车进行误检. 在场景 5 中, 由于阴雨导致画面较为模糊, YOLOv3 仅能检测到少数目标, 且将公共汽车误检为小型汽



(1)~(5) 为 YOLOv3 检测结果, (6)~(10) 为改进 YOLOv3 检测结果

图 7 恶劣天气情况下 YOLOv3 模型与改进 YOLOv3 模型算法检测效果对比

Fig. 7 Comparison of the detection effect of YOLOv3 and the improved YOLOv3 algorithm under severe weather conditions

车,改进 YOLOv3 依然可以检测到画面中的大部分车辆。

由此可见,改进 YOLOv3 对小目标的检测及存在遮挡情况和受恶劣天气影响的环境下,都比原始算法有更高的检测精度。

3 结论

针对当前交通视频中检测目标多、尺度不断变化等问题,提出了基于改进 YOLOv3 模型的交通视频目标检测算法。通过对 YOLOv3 增加检测尺度提高对多尺度目标的检测精度,同时采用 k-means++ 聚类算法选择更适合数据集的候选框,最后为了避免网络层数较多导致的梯度衰落,将输出检测层的 6 个卷积层转化为 2 个残差单元。通过在 UA-DETRAC 数据集上与原始 YOLOv3 进行比较实验,改进后算法的准确率和召回率相比原始 YOLOv3 分别提高了 7.91% 和 4.57%,同时检测时间仅增加了 0.09 ms,可以达到针对交通视频的实时监控要求。

[参考文献] (References)

- [1] 李明熹,林正奎,曲毅. 计算机视觉下的车辆目标检测算法综述[J]. 计算机工程与应用,2019,55(24):20-28.
- [2] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553):436-444.
- [3] WU L, WANG Y, GAO J, et al. Where-and-when to look: deep siamese attention networks for video-based person re-identification[J]. IEEE Transactions on Multimedia, 2019, 21(6):1412-1424.
- [4] YOLCU G, OZTEL I, KAZAN S, et al. Deep learning-based face analysis system for monitoring customer interest[J]. Journal of Ambient Intelligence and Humanized Computing, 2020, 11(1):237-248.
- [5] 崔艳鹏,王元皓,胡建伟. 一种改进 YOLOv3 的动态小目标检测方法[J]. 西安电子科技大学学报,2020,47(3):1-7.
- [6] 黄继鹏,史颖欢,高阳. 面向小目标的多尺度 Faster-RCNN 检测算法[J]. 计算机研究与发展,2019,56(2):319-327.
- [7] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA:IEEE, 2014:580-587.
- [8] GIRSHICK R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile:IEEE, 2015:1440-1448.
- [9] KOIRALA A, WALSH K B, WANG Z, et al. Deep learning for real-time fruit detection and orchard fruit load estimation: benchmarking of 《MangoYOLO》[J]. Precision Agriculture, 2019, 20(6):1107-1135.
- [10] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6):1137-1149.
- [11] 戴立伟,黄山. 交通场景目标检测指标优化研究[J]. 激光与光电子学进展,2020,57(14):344-352.
- [12] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multiBox detector[C]//European Conference on Computer Vision 2016. Glasgow, UK, 2020.
- [13] SANG J, WU Z, GUO P, et al. An improved YOLOv2 for vehicle detection[J]. Sensors, 2018, 18(12):4272.
- [14] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL]. [2020-07-27]. <https://arxiv.org/abs/1804.02767>.
- [15] 孔方方,宋蓓蓓. 改进 YOLOv3 的全景交通监控目标检测[J]. 计算机工程与应用,2020,56(8):20-25.
- [16] ZHAO X, PU F, WANG Z, et al. Detection, tracking, and geolocation of moving vehicle from UAV using monocular camera[J]. IEEE Access, 2019, 7:101160-101170.
- [17] XU Z F, JIA R S, SUN H M, et al. Light-YOLOv3: fast method for detecting green mangoes in complex scenes using picking robots[J]. Applied Intelligence, 2020, 50(12):4670-4687.
- [18] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA:IEEE, 2016:770-778.
- [19] WEN L, DU D, CAI Z, et al. UA-DETRAC: a new benchmark and protocol for multi-object detection and tracking[J]. Computer Science, 2020, 193:102907.

[责任编辑:严海琳]