

基于优化 YOLOv3 算法的交通灯检测

孙迎春, 潘树国*, 赵涛, 高旺, 魏建胜

东南大学仪器科学与工程学院, 江苏 南京 210096

摘要 为解决 YOLOv3 算法在检测道路交通灯时存在的漏检率高、召回率低等问题, 提出一种基于优化 YOLOv3 算法的交通灯检测方法。首先, 采用 K-means 算法对数据进行聚类分析, 结合聚类结果和交通灯标签的统计结果, 确定先验框的宽高比及其数量。然后, 根据交通灯尺寸特点, 精简网络结构, 分别将 8 倍降采样信息、16 倍降采样信息与高层语义信息进行融合, 在两个尺度上建立目标特征检测层。同时, 为了避免交通灯特征随着网络的加深而消失的问题, 分别减少两个目标检测层前的两组卷积层, 简化特征提取步骤。最后, 在损失函数中, 利用高斯分布特性评估边界框的准确性, 以提升对交通灯检测的精度。实验结果显示, 优化 YOLOv3 算法的检测速度可达 30 frame/s, 平均精度较原网络提升 9 个百分点, 可以有效完成对交通灯的检测。

关键词 机器视觉; YOLOv3; 交通灯检测; BDD100K 数据集; K-means 算法; 高斯分布

中图分类号 TP391.4

文献标志码 A

doi: 10.3788/AOS202040.1215001

Traffic Light Detection Based on Optimized YOLOv3 Algorithm

Sun Yingchun, Pan Shuguo*, Zhao Tao, Gao Wang, Wei Jiansheng

School of Instrument Science and Engineering, Southeast University, Nanjing, Jiangsu 210096, China

Abstract To solve the problems of high missed-detection rate and low recall rate existed in the YOLOv3 algorithm for detecting traffic lights, a traffic light detection method based on the optimized YOLOv3 algorithm is proposed. First, the K-means algorithm is used to cluster the data. By combining the clustering results with the statistical results of traffic light labels, the number and the width-height ratios of the prior boxes are determined. Then, the network structure is simplified according to the size characteristics of traffic lights. The $8 \times$ downsampling information and the $16 \times$ downsampling information are fused with high-level semantic information, and the object feature detection layer is established on two scales. Meanwhile, to avoid the disappearance problem of traffic light features with the deepening of the network, two sets of convolution layers are reduced before two object-detection layers, and thus the feature extraction steps are simplified. Finally, in the loss function, Gaussian distribution characteristics are used to evaluate the accuracy of the boundary box to improve the precision of traffic light detection. The experimental results reveal that the detection speed of the optimized YOLOv3 algorithm can reach 30 frames/s and the average precision is 9 percent higher than that of the original network, which effectively completes the detection of traffic lights.

Key words machine vision; YOLOv3; traffic light detection; BDD100k dataset; K-means algorithm; Gaussian distribution

OCIS codes 150.0155; 150.1135; 100.4996

1 引 言

随着计算机视觉相关技术的不断发展, 目标检测算法在自动驾驶领域取得了重大进展, 对行人、车辆、道路线等各类道路目标进行检测的算法应运而生。交通灯是一种室外场景语义视觉定位的重要标识物, 对辅助自动驾驶及构建高精度地图有着重要

意义。因此, 基于计算机视觉的道路交通灯检测成为智能车辆和自动驾驶领域的重要研究课题之一^[1]。

传统基于颜色的交通灯检测方法易受到光照、车辆尾灯等因素影响^[2-3]。深度卷积神经网络(CNN)由于能够自主完成对目标特征的学习, 提取到关键信息, 因而具有较强的鲁棒性^[4-5]。近年来,

收稿日期: 2019-12-31; 修回日期: 2020-02-13; 录用日期: 2020-03-16

基金项目: 国家自然科学基金(41774027, 41774022)、国家重点研发计划(2016YFB0502101)

* E-mail: psg@seu.edu.cn

主要利用目标候选框思想和回归思想,基于 CNN 构成目标检测算法^[6-7]。R-CNN^[8]、Fast R-CNN^[9]、Faster R-CNN^[10]等基于目标候选框思想的双步检测算法首先提取目标候选框,再利用检测网络完成模型训练。SSD(single shot multibox detector)^[11]、YOLO(you only look once)^[12]、YOLOv3^[13]等基于回归思想的单步检测算法摒弃了提取目标候选框的步骤,直接利用检测网络产生目标的类别和位置信息,具有更高的检测速度。

为了提高交通灯检测的速度和精度,本文将 BDD100K 数据集^[14]中的交通灯作为训练和检测的目标,在结构上对单步检测算法中的深度卷积神经网络 YOLOv3 进行优化。首先根据聚类结果及标签统计结果精简网络结构,分别将 8 倍降采样信息、16 倍降采样信息与高层语义信息融合,在两个尺度上建立目标特征检测层。为充分利用低层特征图提取的细节信息,避免特征随网络的加深而消失的问题,分别减少两个目标检测层前的两组卷积层,简化特征提取步骤,以获得表征目标信息的不变特征。最后,改进损失函数,利用高斯分布特性增加对边界框可靠性的判断。实验结果表明,YOLOv3 优化网络可以有效提升交通灯的检测速度和精度。

2 YOLOv3 算法

2.1 YOLOv3 算法原理

YOLOv3 将图像划分成 $S \times S$ 的网格,目标中心所在的网格负责完成对目标的预测。为了完成对 C 类目标的检测,每个网格需要预测 B 个边界框及 P 个条件类别概率($P=C$),并输出边界框中包含目标及边界框准确度的置信度信息。每个边界框对应的置信度 $D_{\text{confidence}}$ 的计算方法为

$$D_{\text{confidence}} = P(o) \times I_{\text{pred}}^{\text{truth}}, \quad (1)$$

式中: o 为被检测目标; $P(\cdot)$ 为网格中含有被检测目标的概率; $I_{\text{pred}}^{\text{truth}}$ 为预测边界框与真实边界框的交并比(IOU)。如果网格包含目标,即目标的中心落在网格内,则 $P(o)$ 为 1,否则为 0;每个边界框对应的类别置信度由每个边界框的置信度与条件类别概率的乘积构成,计算方法为

$$E_{\text{confidence}} = P(c_l | o) \times P(o) \times I_{\text{pred}}^{\text{truth}}, \quad (2)$$

式中: c_l 为被检测目标的类别; l 为类别序号, $l=1, 2, \dots, C$ 。

预测到的边界框中心坐标为 (x_0, y_0) ,边界框的宽为 w_0 、高为 h_0 。为了增加网络的收敛速度,对

边界框的中心坐标、宽和高参数进行归一化处理,计算方法为

$$\begin{cases} w = \frac{w_0}{w_{\text{img}}} \\ h = \frac{h_0}{h_{\text{img}}} \\ x = x_0 \cdot \frac{S}{w_{\text{img}}} - x_c \\ y = y_0 \cdot \frac{S}{h_{\text{img}}} - y_c \end{cases}, \quad (3)$$

式中: (x_c, y_c) 为中心点所在的网格坐标; (x, y) 为归一化后中心点所在的网格坐标; w_{img} 和 h_{img} 分别为图像的宽度和高度; w 和 h 分别为归一化后图像的宽度和高度。根据每个网格的输出,计算得到网络的输出大小为 $S \times S \times (5 \times B + C)$ 。

2.2 Darknet-53 网络

Darknet-53 网络是 YOLOv3 算法的重要组成部分,用来提取目标的特征。Darknet-53 由 53 个卷积层组成,每个卷积层包含卷积运算、批归一化处理及 Leaky ReLU 激活函数,网络相关参数如图 1 所示。Darknet-53 采用步长为 2 的卷积核完成降采样,共包含 5 次降采样,其中分别在 8 倍,16 倍,32 倍降采样中完成对目标的预测。同时,为增加网络的深度并减少模型训练的计算量,Darknet-53 引入残差结构。由于融合了低维特征的细节信息和高维特征的语义信息,Darknet-53 网络在保持精度不变的情况下提高了检测的速度。

3 YOLOv3 优化算法

3.1 交通灯数据集标签聚类分析

anchor 具有固定的宽高值,是用来预测边界框偏移的参考框。Anchor 选择的优劣对最终的检测结果有直接的影响。YOLOv3 采用 K-means 算法对训练集标签进行维度聚类分析,以获得 anchor 的参数,减少训练过程中的匹配误差。

欧氏距离作为评价标签相似性的度量指标,会使大尺寸标签产生更大的误差,影响聚类结果,不利于目标检测的速度和精度。YOLOv3 采用平均交并比(Avg IOU, I_{Avg})作为度量标签相似性的指标,能避免标签尺寸大小在聚类过程中引入的误差。距离公式表示为

$$d(b, c) = 1 - I_{\text{Avg}}(b, c), \quad (4)$$

式中: b 为标签; c 为聚类中心。标签与聚类中心矩形框的尺寸越相似,聚类距离就越小。

	type	number of filters	size	output size
	convolutional	32	3×3	256×256
	convolutional	64	3×3/2	128×128
1×	convolutional	32	1×1	
	convolutional	64	3×3	
	residual			128×128
	convolutional	128	3×3/2	64×64
2×	convolutional	64	1×1	
	convolutional	128	3×3	
	residual			64×64
	convolutional	256	3×3/2	32×32
8×	convolutional	128	1×1	
	convolutional	256	3×3	
	residual			32×32
	convolutional	512	3×3/2	16×16
8×	convolutional	256	1×1	
	convolutional	512	3×3	
	residual			16×16
	convolutional	1024	3×3/2	8×8
4×	convolutional	512	1×1	
	convolutional	1024	3×3	
	residual			8×8

图 1 Darknet-53 相关参数

Fig. 1 Relative parameters of Darknet-53

为了设置与交通灯尺寸及比例相匹配的 anchor, 首先对数据集中的标签分布情况进行统计, 如图 2 所示。

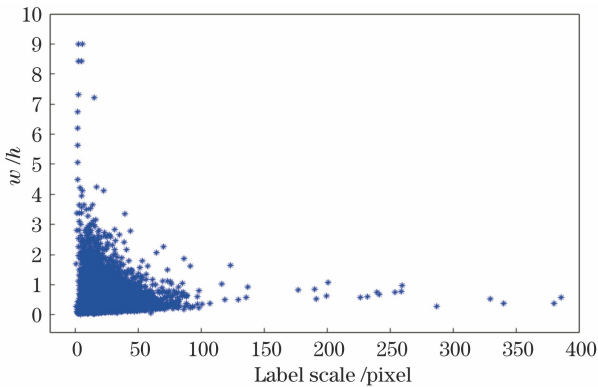


图 2 交通灯标签分布情况

Fig. 2 Distribution of traffic light label

图 2 中, 横轴代表标签的尺度变化范围, 通过将原始标签面积调整为网络输入分辨率面积, 再对其开根号得到; 纵轴代表标签中宽与高的比例。由图 2 可知, 数据集中交通灯目标的尺寸集中在 0~100 pixel×100 pixel 内, 宽高比集中在 0~1 内。

然后对数据集中的交通灯标签进行聚类分析, 从 1 开始不断增加聚类中心的个数, 得到聚类中心数量 k 与 I_{Avg} 之间的关系, 如图 3 所示。当 k 大于 3 时, 曲线逐渐变得平缓, 聚类出现相似的结果, 为了

避免冗余, 取聚类中心的数量为 3 时对应的 anchor 作为检测交通灯的预测框, 分别为 (5, 13), (8, 23), (14, 44)。使用 K-means 算法得到的聚类结果尺寸都较小, 如表 1 所示。考虑到尺寸大于 50 pixel×50 pixel 的标签在数据集中占有一定比例, 在聚类分析得到的 3 个 anchor 的基础上添加一组尺寸较大的 anchor 以匹配数据集中像素大于 50 pixel×50 pixel 的标签, 最终选取的 anchor 参数为 (8, 23), (14, 44), (5, 13), (30, 61), (62, 45), (59, 119)。

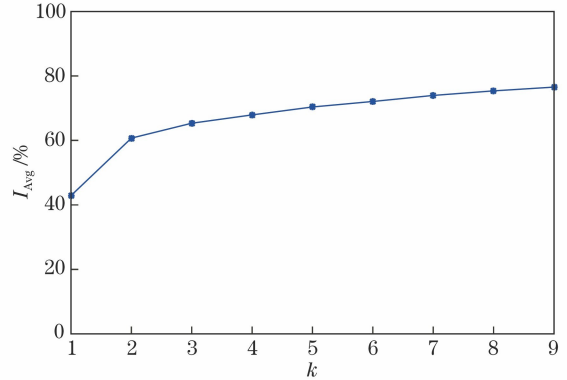


图 3 聚类分析结果

Fig. 3 Result of clustering analysis

表 1 不同 k 值下的聚类结果

Table 1 Clustering results under different k values

$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	$k=9$
(5, 12)	(4, 12)	(4, 12)	(3, 12)	(4, 10)	(3, 11)
(5, 21)	(5, 20)	(6, 18)	(5, 18)	(5, 16)	(5, 13)
(8, 18)	(9, 27)	(6, 11)	(6, 10)	(7, 12)	(5, 21)
(13, 39)	(7, 14)	(7, 30)	(7, 17)	(7, 22)	(7, 12)
	(14, 47)	(12, 21)	(8, 33)	(8, 33)	(8, 18)
		(14, 47)	(12, 21)	(12, 46)	(8, 31)
			(15, 49)	(13, 22)	(12, 45)
				(19, 56)	(14, 23)
					(19, 57)

3.2 面向小尺寸交通灯的网络结构优化

加深网络的层数虽然可以使得模型提取到更丰富的语义信息, 但在对低像素目标进行检测时, 易产生特征消失的问题。根据数据集中目标维度聚类分析的结果, 简化网络的特征提取过程, 以提升最终的检测精度。

交通灯在数据集图像中的尺寸占比较小, 像素较低, Darknet-53 网络不易提取到目标的特征。为避免学习能力随网络层数加深而退化的问题, 精简网络结构, 分别将 8 倍降采样层、16 倍降采样层与高层语义信息层进行融合, 在两个尺度上建

立目标特征检测层。同时在两个目标检测层前分别减少两组卷积层,以充分利用低层特征图提取

的细节信息。优化后的 YOLOv3 网络结构如图 4 所示。

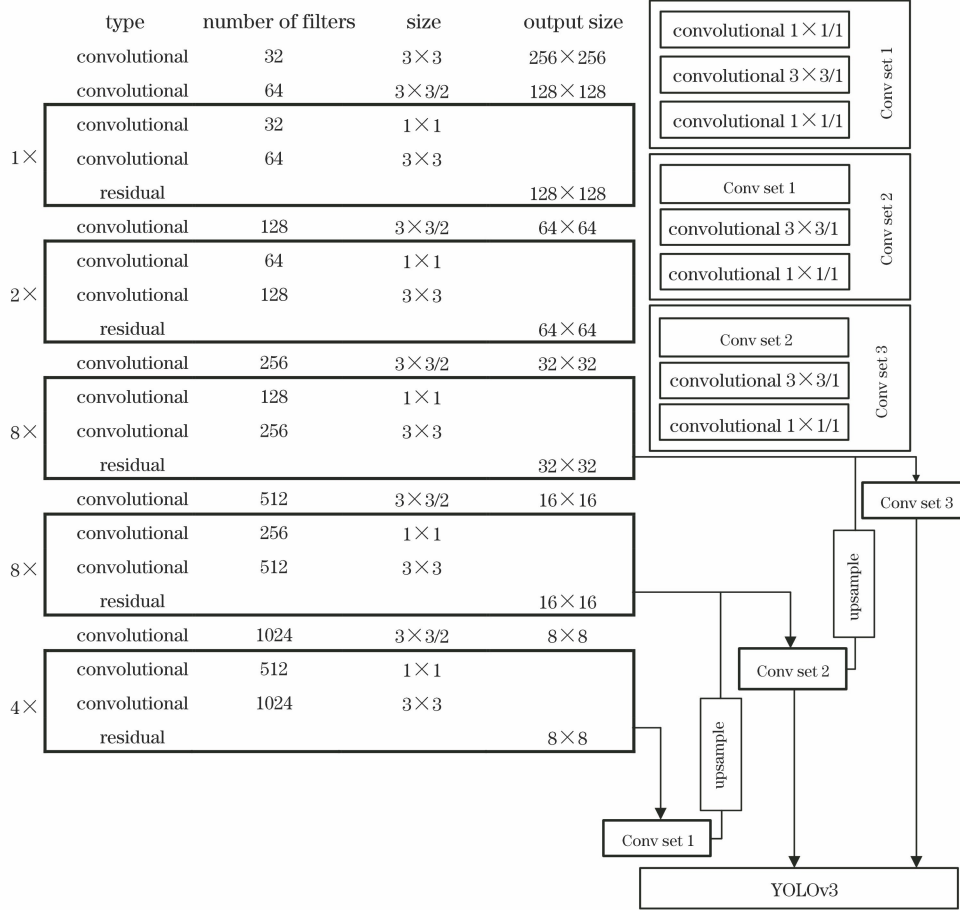


图 4 优化后的 YOLOv3 网络结构

Fig. 4 Optimized YOLOv3 network structure

在对交通灯目标进行检测时,考虑到网络的输入分辨率对模型检测精度和速度的影响,首先将数据集图像统一缩放成 3 个通道分辨率为 672 pixel×672 pixel 的形式;再利用 Darknet-53 提取交通灯目标的特征,交替使用 3×3 和 1×1 大小的卷积核进行运算。为得到更适合检测交通灯目标的深度卷积神经网络,精简原始 YOLOv3 网络,并融合大特征图的细节信息和小特征图的语义信息,在两个尺度上训练出最终的交通灯目标检测模型。

3.3 基于高斯分布特性的损失函数改进

损失函数用来度量神经网络预测值与真实值之间的误差,是决定模型检测效果的重要参数之一。在 YOLOv3 损失函数中,利用高斯分布特性评估每个边界框的不确定性,以提升网络的精度。

YOLOv3 算法的损失函数设计主要从边界框坐标预测误差、边界框的置信度误差、分类预测误差这三个方面进行考虑。YOLOv3 损失函数 $L_{\text{loss}}(o)$ 计算公式为

$$\begin{aligned}
 L_{\text{loss}}(o) = & \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] + \\
 & \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} (2 - w_i \times h_i) [(w_i - \hat{w}_i)^2 + (h_i - \hat{h}_i)^2] - \\
 & \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} [\hat{C}_i \ln C_i + (1 - \hat{C}_i) \ln(1 - C_i)] - \\
 & \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{noobj}} [\hat{C}_i \ln C_i + (1 - \hat{C}_i) \ln(1 - C_i)] - \\
 & \sum_{i=0}^{S^2} I_i^{\text{obj}} \sum_{l \in C_{\text{classes}}} \{ \hat{P}_i(l) \ln [P_i(l)] + \\
 & [1 - \hat{P}_i(l)] \ln [1 - P_i(l)] \}, \quad (5)
 \end{aligned}$$

式中:网格序号 $i=0, 1, \dots, S^2$; 边界框序号 $j=0, 1, \dots, B$; x_i 为第 i 个网格的边界框中心点横坐标的预测值, y_i 为第 i 个网格的边界框中心点纵坐标的预测值; \hat{x}_i 为第 i 个网格的边界框中心点横坐标的真实值, \hat{y}_i 为第 i 个网格的边界框中心点纵坐标的真实值; w_i 为第 i 个网格的边界框宽度的预测值, h_i 为

第 i 个网格的边界框高度的预测值; \hat{w}_i 为第 i 个网格的边界框宽度的真值, \hat{h}_i 为第 i 个网格的边界框高度的真值; C_i 为第 i 个网格的置信度预测值; \hat{C}_i 为第 i 个网格的置信度真值; $P_i(l)$ 为第 i 个网格包含的目标类别预测值; $\hat{P}_i(l)$ 为第 i 个网格包含的目标类别真值; $C_{classes}$ 为被检测目标的类别; λ_{coord} 为权重系数; λ_{noobj} 为惩罚权重系数; I_{ij}^{obj} 表征第 i 个网格第 j 个边界框是否负责预测目标, 当负责预测目标时取值为 1, 否则为 0; I_{ij}^{noobj} 表征第 i 个网格第 j 个边界框是否负责预测目标, 当负责预测目标时取值为 0, 否则为 1。

受 Gaussian YOLOv3 的启发^[15], 为了提高交通灯检测精度, 利用高斯分布特性改进损失函数, 进一步判断交通灯边界框的可靠性, 以边界框中心点 x 方向坐标为例, 修改后的边界框 x' 坐标预测误差计算方法为

$$L_x = - \sum_{m=1}^W \sum_{n=1}^H \sum_{k=1}^K \gamma_{mnk} \times \ln\{N[x_{mnk}^G | \mu_{t_x}(x_{mnk}), \sum_{t_x} x_{mnk}] + \epsilon\}, \quad (6)$$

式中: t_x 为边界框的中心坐标相对于网格左上角 x 坐标的偏移量; W 和 H 分别对应每个宽和高的网格数; K 对应 anchor 的数目; $\mu_{t_x}(x_{mnk})$ 为输出层第 (m, n) 个网格中第 k 个 anchor 的 t_x 的均值; $\sum_{t_x} x_{mnk}$ 为对应的 t_x 的不确定性; x_{mnk}^G 为 t_x 的真值; γ_{mnk} 为权重参数; ϵ 为 10^{-9} 。

4 实验与分析

4.1 网络的训练与实验环境配置

使用 Intel® Xeon (R) CPU E5-2690 v4 处理器, TITAN V 显卡, 128 GB CPU 内存, 12G 显卡内存, 基于 Darknet 深度学习框架, 在 Ubuntu16.04 操作系统下对 YOLOv3 网络和优化的 YOLOv3 网络分别进行训练。训练开始阶段, 衰减系数设置为 0.0005, 学习率设置为 0.001, 并且选择 step 模式更新学习率, 在训练迭代次数达到 15000 和 20000 时, 将学习率分别降低至初始学习率的 10% 和 1%。图 5 和图 6 分别表示训练过程中 YOLOv3 优化算法平均损失值及平均交并比的变化趋势。

随着训练次数的增加, 损失值不断下降, 最终稳定在 1.5 左右, 平均交并比不断上升, 最终在 0.55 上下浮动。由曲线的收敛情况分析, 25000 次训练后可以停止迭代。

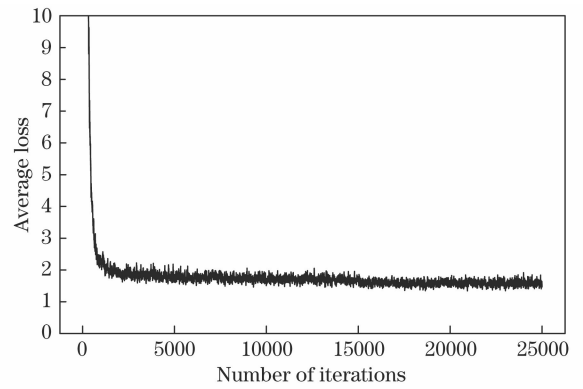


图 5 YOLOv3 优化算法的平均损失函数曲线
Fig. 5 Average loss function curve of optimized YOLOv3 algorithm

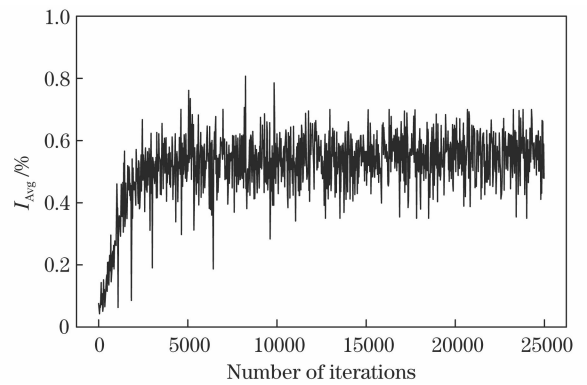


图 6 YOLOv3 优化算法的平均交并比曲线
Fig. 6 Average IOU curve of optimized YOLOv3 algorithm

4.2 网络的测试

分别使用 YOLOv3 网络和优化的 YOLOv3 网络对测试集中 5646 张图片进行测试, 网络的检测结果如图 7 所示。

对比图 7(a)、(b) 可以发现, YOLOv3 对交通灯的检测存在错检问题, 误将卡车的边框当作目标, 而优化的 YOLOv3 网络可以避免对交通灯的错检。对比图 7(c)、(d) 可以发现, YOLOv3 对交通灯的检测存在漏检问题, 没有检测到道路正上方的 1 个交通灯, 而优化的 YOLOv3 网络可以检测出原 YOLOv3 算法漏检的目标。

为了准确评估模型检测效果, 计算模型检测目标时的精确率 (P) 和召回率 (R)。

$$P = \frac{Q_{TP}}{Q_{TP} + Q_{FP}}, \quad (7)$$

$$R = \frac{Q_{TP}}{Q_{TP} + Q_{FN}}, \quad (8)$$

式中: Q_{TP} 为模型正确检测出来的目标数量; Q_{FP} 为模型误检的目标数量; Q_{FN} 为模型漏检的目标数量。平均精准度 (AP)^[16] 综合了精确率和召回率

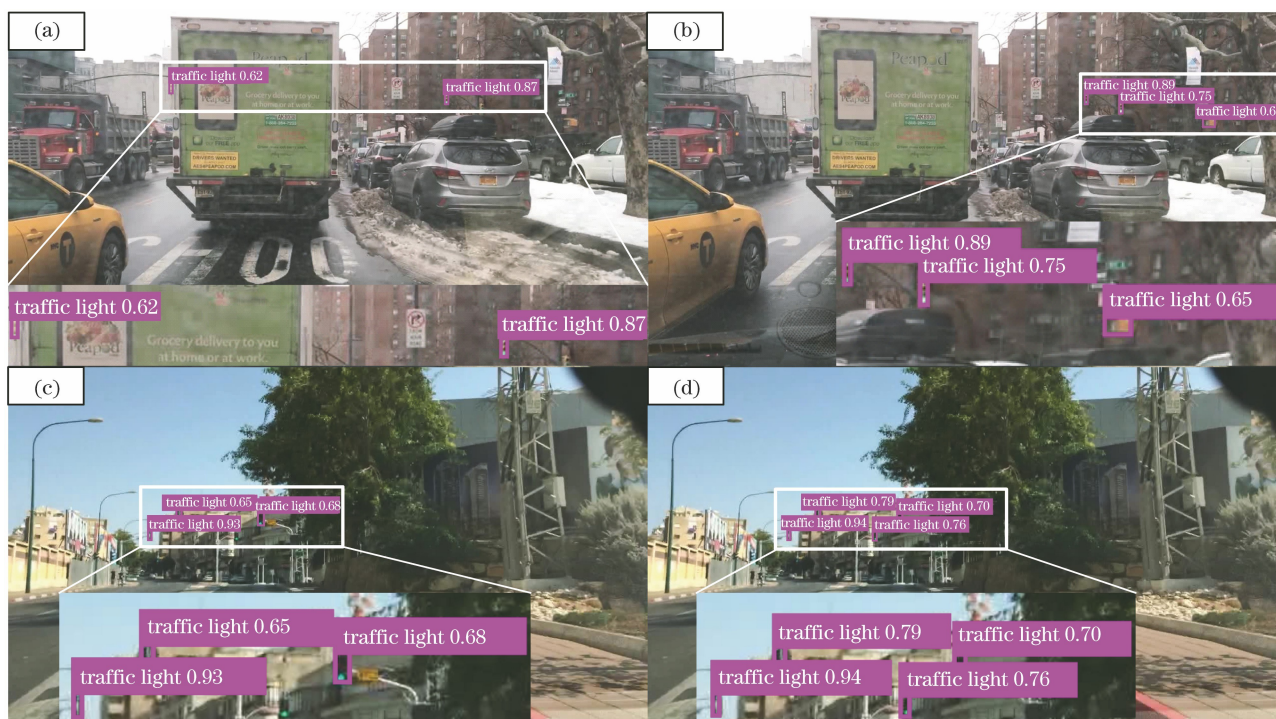


图 7 各网络对交通灯的检测效果。(a)(c) YOLOv3 网络的检测效果；
(b)(d) 优化的 YOLOv3 网络的检测效果

Fig. 7 Detection results of traffic lights by each network. (a)(c) Detection results of YOLOv3 network;
(b)(d) detection results of optimized YOLOv3 network

的结果,用于评估模型在单个检测类别上的表现。YOLOv3 网络和优化的 YOLOv3 网络对 BDD100K 数据集中交通灯的检测 AP 曲线如图 8、9 所示。

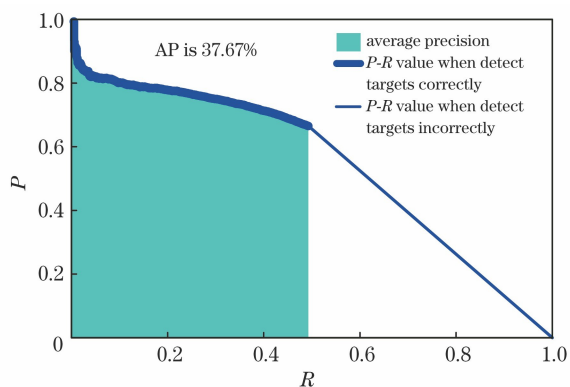


图 8 YOLOv3 网络的 AP 曲线

Fig. 8 AP curve of YOLOv3 network

优化的 YOLOv3 网络检测交通灯的平均精准度达到 46.78%, 召回率达到 57.58%, 准确率达到 69.70%, 测试单张图片的时间为 33 ms, 检测图片的速度为 30 frame/s, 与 YOLOv3 相比, 平均精准度提升 9.11 个百分点, 召回率提升 8.3 个百分点, 精确率提升 3.19 个百分点, 测试单张图片的时间减少

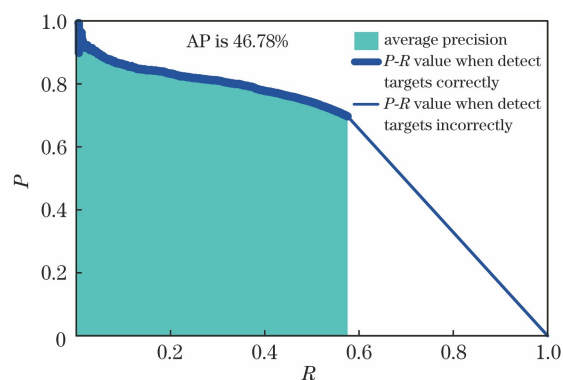


图 9 优化 YOLOv3 网络的 AP 曲线

Fig. 9 AP curve of optimized YOLOv3 network

4 ms, 检测速度提升 3 frame/s。不同网络下的总体目标检测结果如表 2 所示。

表 2 总体目标检测结果对比

Table 2 Comparison of total object detection results

Algorithm	Time / ms	Speed / (frame·s ⁻¹)	Precision / %	Recall / %	AP / %
YOLOv3	37	27	66.51	49.28	37.67
Optimized YOLOv3 algorithm	33	30	69.70	57.58	46.78

根据光线变化、天气变化和目标尺寸的大小,将测试集分别划分为白天和夜晚两种子测试集、雨天和非雨天两种子测试集、大目标和小目标两种子测试集。在3种不同的划分标准下,使用YOLOv3网络和优化的YOLOv3网络对目标进行检测,得到不同场景和目标尺寸下的平均精准度,如图10所示。

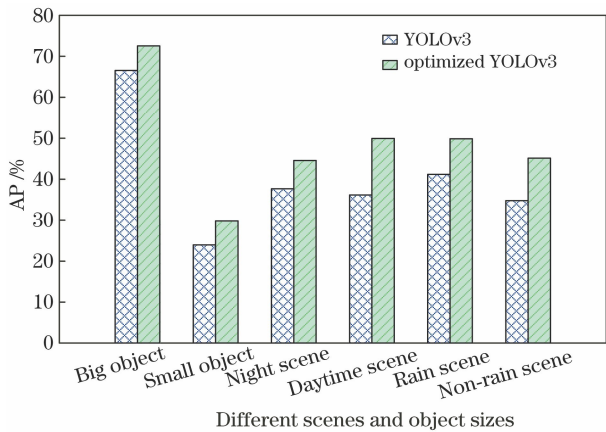


图10 不同场景和目标尺寸下的平均精准度对比
Fig. 10 Comparison of average precision under different scenes and object sizes

图10中,优化的YOLOv3网络在白天和夜晚、

雨天和非雨天、大目标和小目标共6种情况下的平均精准度均高于原YOLOv3网络。其中,对于白天和夜晚、雨天和非雨天4种场景,与原YOLOv3网络相比,优化的YOLOv3网络在日间场景下的平均精准度的提升最高,约14个百分点;对于大目标和小目标2种不同的目标尺寸,YOLOv3网络和优化的YOLOv3网络检测大目标时的平均精准度高,均高于60%,检测小目标时的平均精准度低,均低于30%。YOLOv3网络和优化的YOLOv3网络在不同目标尺寸和不同场景下的检测精确率、召回率和平均精准度如表3、4所示。

表3 不同目标尺寸下的检测精确率、召回率和平均精准度

Table 3 Detection precision, recall, and average precision under different object sizes

Network	Big object			Small object		
	P / %	R / %	AP / %	P / %	R / %	AP / %
YOLOv3	83.95	69.32	66.53	54.59	37.93	24.00
Optimized YOLOv3	86.71	75.68	72.54	56.58	44.48	29.84

表4 不同场景下的检测精确率、召回率和平均精准度

Table 4 Detection precision, recall, and average precision under different scenes

Network	Night scene			Daytime scene			Rain scene			Non-rain scene		
	P / %	R / %	AP / %	P / %	R / %	AP / %	P / %	R / %	AP / %	P / %	R / %	AP / %
YOLOv3	65.36	49.98	37.67	70.45	44.04	36.15	70.28	51.00	41.18	63.71	47.29	34.77
Optimized YOLOv3	68.32	57.24	44.58	72.06	59.11	49.96	73.79	58.86	49.89	70.01	57.10	45.15

YOLOv3优化算法通过精简网络,极大地避免了特征消失的问题,增强了网络对目标特征细节信息的提取能力,使不同场景和不同目标尺寸下的召回率都得到7个百分点到15个百分点的提升。同时,YOLOv3优化算法利用高斯分布特性,增强网络对每个边界框不确定性的判断,使不同场景和不同目标尺寸下的精确率都得到1.5个百分点到6.5个百分点的提升。在总体平均精准度提升9个百分点的情况下,网络结构的精简与高斯分布特性的运用还缩减了模型的检测时间,检测速度从27 frame/s提升到30 frame/s。不同场景和目标尺寸下的检测结果对比如图11所示。

5 结 论

提出了一种YOLOv3优化算法的交通灯检测方法。首先对交通灯数据集标签进行聚类分析,结合聚类分析结果及数据集标签统计情况对anchor参数进行优化;为避免随网络加深而产生梯度消散的问题,精简网络结构,同时融合低层细节信息和高层语义信息,在两个尺度上建立目标特征检测层;最后利用高斯分布特性增强对交通灯边界框可靠性的判断,提高交通灯的检测精度。实验结果表明,YOLOv3优化算法可以有效提升交通灯的检测精度和速度。由于先验框的设置依赖于聚类算法,存



图 11 检测结果对比。(a) YOLOv3 算法检测结果;(b) YOLOv3 优化算法检测结果
 Fig. 11 Comparison of detection results. (a) Detection results of YOLOv3 algorithm;
 (b) detection results of optimized YOLOv3 algorithm

在一定的随机性,下一步将结合 Guided Anchoring 机制^[17]及模型高效压缩方法开展研究。

参 考 文 献

[1] Wang W H, Sun S L, Jiang M X, et al. Traffic

- lights detection and recognition based on multi-feature fusion[J]. *Multimedia Tools and Applications*, 2017, 76(13): 14829-14846.
- [2] Shi X M, Zhao N, Xia Y J. Detection and classification of traffic lights for automated setup of road surveillance systems[J]. *Multimedia Tools and Applications*, 2016, 75(20): 12547-12562.
- [3] Kim H K, Yoo K Y, Park J H, et al. Traffic light recognition based on binary semantic segmentation network[J]. *Sensors*, 2019, 19(7): 1700.
- [4] Zhang R, Li W P, Mo T. Review of deep learning [J]. *Information and Control*, 2018, 47(4): 385-397, 410.
张荣, 李伟平, 莫同. 深度学习研究综述[J]. *信息与控制*, 2018, 47(4): 385-397, 410.
- [5] Hua X, Wang X Q, Wang D, et al. Multi-objective detection of traffic scenes based on improved SSD[J]. *Acta Optica Sinica*, 2018, 38(12): 1215003.
华夏, 王新晴, 王东, 等. 基于改进 SSD 的交通大场景多目标检测 [J]. *光学学报*, 2018, 38(12): 1215003.
- [6] Ma Q, Zhu B, Cheng Z D, et al. Detection and recognition method of fast low-altitude unmanned aerial vehicle based on dual channel[J]. *Acta Optica Sinica*, 2019, 39(12): 1210002.
马旗, 朱斌, 程正东, 等. 基于双通道的快速低空无人机检测识别方法 [J]. *光学学报*, 2019, 39(12): 1210002.
- [7] Ju M R, Luo H B, Wang Z B, et al. Improved YOLO V3 algorithm and its application in small target detection[J]. *Acta Optica Sinica*, 2019, 39(7): 0715004.
鞠默然, 罗海波, 王仲博, 等. 改进的 YOLO V3 算法及其在小目标检测中的应用[J]. *光学学报*, 2019, 39(7): 0715004.
- [8] 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, June 23-28, 2014, Columbus, OH, USA. New York: IEEE, 2014: 580-587.
- [9] Girshick R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE, 2015: 1440-1448.
- [10] Ren S Q, He K M, 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] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector [M]//Leibe B, Matas J, Sebe N, et al. *Computer vision-ECCV 2016. Lecture notes in computer science*. Cham: Springer, 2016, 9905: 21-37.
- [12] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 779-788.
- [13] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2019-12-05]. <https://arxiv.org/abs/1804.02767>.
- [14] Yu F, Xian W Q, Chen Y Y, et al. BDD100k: a diverse driving video database with scalable annotation tooling[EB/OL]. (2018-05-12)[2019-12-05]. <https://arxiv.org/abs/1805.04687>.
- [15] Choi J, Chun D, Kim H, et al. Gaussian YOLOv3: an accurate and fast object detector using localization uncertainty[EB/OL]. (2019-04-09)[2019-12-05]. <https://arxiv.org/abs/1804.02767>.
- [16] Everingham M, van Gool L, Williams C K I, et al. The pascal visual object classes (VOC) challenge[J]. *International Journal of Computer Vision*, 2010, 88(2): 303-338.
- [17] Wang J Q, Chen K, Yang S, et al. Region proposal by guided anchoring[C]//2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE, 2019: 2965-2974.