

基于改进 SSD 的车辆小目标检测算法研究

郭健忠¹, 余腾飞¹, 崔玉定¹, 周兴林²

(1. 武汉科技大学 汽车与交通工程学院, 湖北 武汉 430065;

2. 武汉科技大学 机械自动化学院, 湖北 武汉 430065)

摘要:针对 SSD (single shot multibox detector, 单步多盒检测) 算法在车辆的自动紧急制动 (AEB) 中对远方目标检测效果差、检测速度慢、对硬件资源需求高的问题, 提出了一种基于 SSD 的改进算法。首先用 MobileNetv2 替换 SSD 中的 AGG-16 作为检测网络, 以减少参数数量和计算量, 降低网络对硬件性能的需求; 其次, 提出了特征增强和融合的方法, 反复挖掘目标信息, 并把不同特征层的信息进行融合, 以提高对小目标检测的能力; 最后, 对先验框解码过程进行改进, 减少网络需要解码的先验框数量, 再次减少计算量, 提高网络检测速度, 并调整先验框的尺寸, 进一步增强小目标检测的能力。把改进后的网络和 SSD300、YOLO、MobileNetv2-SSD 等网络在 KITTI 数据集上进行检测和对比分析, 实验结果表明, 改进后的网络对小目标识别的速度有所加快, 鲁棒性更好, 准确率更高, 同时也降低了对硬件配置资源的需求。

关键词:卷积神经网络; 目标检测; 小目标; 特征增强; 特征融合

中图分类号: TP301.6

文献标识码: A

文章编号: 1673-629X(2022)03-0001-07

doi: 10.3969/j.issn.1673-629X.2022.03.001

Research on Vehicle Small Target Detection Algorithm Based on Improved SSD

GUO Jian-zhong¹, YU Teng-fei¹, CUI Yu-ding¹, ZHOU Xing-lin²

(1. School of Automobile and Traffic Engineering, Wuhan University of Science and Technology, Wuhan 430065, China;

2. School of Mechanical Automation, Wuhan University of Science and Technology, Wuhan 430065, China)

Abstract: In order to solve the problems of poor detection effect, slow detection speed and high demand for hardware resources, an improved algorithm based on SSD is proposed in automatic emergency braking (AEB) of vehicles. Firstly, AGG-16 in SSD is replaced by MobileNetv2 as the detection network to reduce the number of parameters and the amount of computation, and reduce the demand of network hardware performance. Secondly, the method of feature enhancement and fusion is proposed, which excavates target information repeatedly and integrates the information of different feature layers to improve the ability of small target detection. Finally, the decoding process of the priori box is improved to reduce the number of priori boxes that need to be decoded, reduce the amount of computation again, improve the speed of network detection, and adjust the size of the priori box to further enhance the ability of small target detection. The improved network is detected and compared with SSD300, YOLO, MobileNetV2-SSD and other networks on KITTI dataset. The experimental results show that the improved network has higher speed, better robustness, higher accuracy, and reduces the demand for hardware configuration resources.

Key words: convolutional neural network; object detection; small target; feature enhancement; feature fusion

0 引言

近几年来,随着计算机视觉技术的快速发展,基于神经网络的目标检测技术不断增强,其在汽车行业的运用受到了广泛的关注和研究。目前,百度、滴滴、博世等公司已把神经网络算法运用到汽车无人驾驶上,并取得了良好效果。可以预见的是,神经网络算法必

将在无人驾驶领域大放光彩。

但是在实际应用当中,有两个难点需要解决,一是车载设备的计算能力和存储空间往往受到诸多限制,故算法必须对硬件资源的需求少;二是算法要对远距离上的小目标有较强的检测能力,检测速度快,以便汽车有充足的时间应对。采用轻量型神经网络如

收稿日期: 2021-02-23

修回日期: 2021-06-24

基金项目: 2018 年国家重大科研项目 (51827812)

作者简介: 郭健忠 (1969-), 男, 副教授, 研究方向为汽车电子与信息技术; 通讯作者: 崔玉定 (1976-), 男, 博士研究生, 讲师, 研究方向为车辆系统动力学控制。

MobileNet 可以有效降低模型对硬件的要求。2016 年, Iandola 等人^[1]提出 Squeeze Net, 采用 1×1 卷积核对上层的特征图进行卷积, 主要目的是减少特征图的维数。后来, Sifre 等人^[2]提出了深度可分离卷积的结构, 进一步减少了网络的参数量和计算量。而深度可分离卷积的本质是使用深度卷积和逐点卷积将卷积操作分解为冗余信息更少的稀疏化表达。2017 年, 谷歌公司的 Howard 等人^[3]发布了一种可以在手机等移动终端运作的神经网络结构: MobileNet, 采用了深度可分离卷积的结构。之后, Howard 等人^[4]尝试对 MobileNet 进行改进, 提出了 MobileNetv2, 主要利用了线性瓶颈层和倒残差对网络进行改进, 明显地提升了网络的预测精度。即使如此, 在车载设备的限制和要求下, MobileNetv2 的检测精度仍不令人满意。为提高其检测精度和对小目标的检测能力, 可考虑结合其他算法, 取长补短, 以期提高整体算法的性能。

卷积神经网络中因对目标检测的方式不同, 可把算法大致分为基于候选区域和基于回归的目标检测算法两类^[5]。基于回归的目标检测算法省略了候选区域生成步骤, 直接将特征提取、目标分类以及目标回归在同一个卷积神经网络中实现, 将目标检测过程简化为一种端到端的回归问题, 使基于深度学习的目标检测算法在速度上有了大幅度提升, 这无疑比基于候选区域的算法更适合。这类算法的代表有: YOLO 系列^[6-7]、SSD^[8]、FPN^[9]以及 Retina Net^[10]。其中, SSD 网络是当中检测精确度相对更高的网络结构。为提升 SSD 模型在中小目标检测方面的能力, 很多学者对此进行了研究。例如, Tang 等^[11]在原有 SSD 模型的基础之上采用多视窗的方法进行检测从而提升准确率, 但是这种方法的区域划分不固定, 对于目标检测的准确性和鲁棒性产生了一定的影响。Jeong 等^[12]提出了 RSSD 模型, 改善了 SSD 的特征融合方式, 利用分类网络增加获取的 Feature Map 之间的联系, 减少了重复框的出现, 同时增加了 Feature Map 的个数使得模型能够检测到更多尺度的目标; Fu 等^[13]提出了 DSSD, 使用 Resnet-101 替代原来的 VGG Net, 在分类回归之前使用残差模块, 在辅助的 SSD 额外特征卷积层后面添加了几层反卷积层, 整个网络呈现出宽窄宽的结构, 提高了小目标的检测精度。陈幻杰等^[14]对 SSD 模型的改进, 通过小目标特征区域进行放大提取, 额外提取多个高层特征层的信息, 利用更深的网络结构改善中等目标的检测结果等方法, 较大地提高了对中小目标的准确率, 检测速度相比原有的 SSD 模型下降了将近一半, 在满足准确性的基础上实时性受到了较大的影响。

该文结合 MobileNetv2 和 SSD 模型构成 MobileNetv2-SSD 网络, 并对网络结构进行改进, 增强

对特征的提取效果, 使网络对小目标有更强的检测能力, 同时降低了对硬件资源的需求。实验表明, 改进后的网络有更强的小目标检测能力。

1 MobileNetv2-SSD 网络

在车辆目标检测阶段, 因需要迅速、准确地识别目标, 所以基于回归的目标检测算法更加适合。考虑到 SSD 在准确度和检测速度上比其他基于回归的目标检测算法更有优势, 故该文采用 SSD 网络进行目标检测。但 SSD 网络对小目标的检测效果较差, 而且整个模型相对较大, 在实际应用时对硬件要求高, 考虑到硬件成本问题, 应结合使用规模较小的网络, 降低对硬件配置资源的需求, 而 MobileNetv2 作为轻量化的较为符合需求。基于以上分析, 考虑结合 MobileNetv2 和 SSD 网络。结合后的网络兼具 SSD 网络检测速度快、准确度高和 MobileNetv2 网络参数少、模型小的特点, 较为符合要求。

SSD 模型是当中检测精确度相对较高的网络结构, 其不同大小的特征图上预测目标。SSD 网络检测目标时, 以每一个网格为基础, 生成多个先验框, 之后每一个先验框都会检测框内是否有目标, 若判断出框内有目标, 就根据目标的大小和位置, 计算出每一个先验框的宽高偏差和中心点的偏移量, 通过解码得到预测框的中心和宽高。利用框中物体的置信度对框进一步筛选, 经过非极大值抑制处理, 筛选出得分最高的框, 标记在图片上。

MobileNetv2 是一种轻量化的卷积神经网络, 它对输入图像先升维、深度分解卷积、再降维, 这种被称为倒残差的结构可以减少参数量和计算量。在 ImageNet 数据集上做实验, MobileNet 与 VGG16 的准确率几乎一样, 但参数量和计算量只有其 $1/9$ 。

MobileNetv2-SSD 网络结合两种网络的优势, 检测速度快, 检测准确度高, 然而它对小目标的检测效果不尽如人意。在目标检测物的在图像中占比较小时, 也会存在如下问题:

(1) 小目标呈现的信息量较小, 导致检测效果不佳, 训练时需要反复挖掘信息。

(2) 在训练阶段, 对于小物体的标注容易存在偏差, 在目标物体较小的情况下, 标记的误差会对检测效果产生较大的影响。

(3) 实际应用场景中, 小目标数量较少, 使得网络训练时负样本过多, 导致模型的退化, 严重制约了检测网络的性能。

2 改进型 MobileNetv2-SSD 网络

为解决以上问题, 该文提出特征增强和特征融合

的方法使网络反复挖掘小目标的信息,以加强网络对小目标的检测能力。同时,为了进一步减少参数量和计算量,并使得网络更符合检测车辆特征和实际应用的需要,对先验框进行改进。

2.1 网络结构设计

改进后的模型结构如图1所示。输入为 $300 \times 300 \times 3$ 的图片,经过一次标准卷积和两次深度可分离卷积后进入图1所示结构。用 MobileNetv2 网络取代原本 SSD 中的 VGG 网络,用于多尺度特征提取,之后用特征增强模块挖掘提取到的信息。考虑到大特征图更易提取小目标,故对大特征图进行特征融合,以加强对小目标的提取能力。然后,对增强或者融合后的特征图进行归一化和使用 leaky ReLU 等操作,防止梯度消失

和梯度爆炸。最后,用非极大值抑制进行目标分类。

模型主要分为两个部分,第一部分是用于提取目标特征的 MobileNetv2 基础模块和 SSD 基础模块,一步步缩小尺度,得到不同尺度下的特征图;第二部分是多尺度特征检测网络,其作用是进行不同尺度条件下的特征提取,并把提取到的特征进行增强和融合,然后进行归一化处理 and 线性修正,最终得到一维向量进行目标检测。从图1可以看出,检测模块(detections)得到6个特征图中目标的置信度和坐标,通过非极大值抑制(non-maximum suppression, NMS)模块过滤掉重复预测的目标。通过该模型可以有效提取待识别目标的信息,且具有准确度高和检测速度快等优点,对变化目标具有较好的鲁棒性。

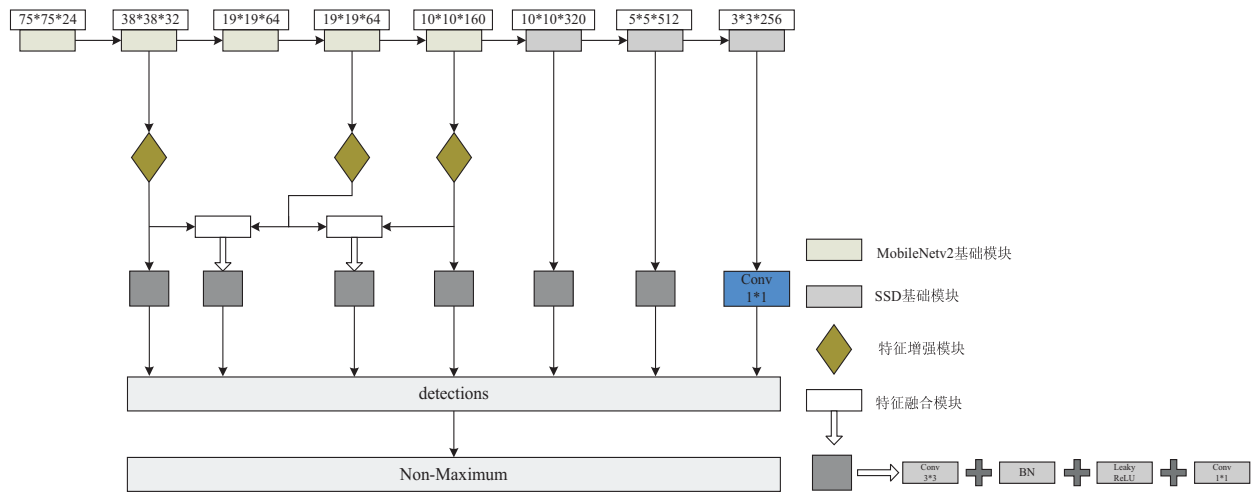


图1 M2-SSD 网络结构

网络的基本配置如表1所示。表中 t 是输入通道的倍增系数(即瓶颈层中的通道数是输入通道数的多少倍); n 是该模块重复的次数; c 是输出通道数; s 是该模块第一次重复时的 stride(后面重复都是 stride = 1)。

表1 M2-SSD 网络基本配置

input	operator	t	c	n	s
$300^2 \times 3$	conv2d		32	1	2
$150^2 \times 32$	bottleneck1	1	16	1	1
$150^2 \times 16$	bottleneck2	6	24	2	2
$75^2 \times 24$	bottleneck3	6	32	3	2
$38^2 \times 32$	bottleneck4	6	64	4	1
$19^2 \times 64$	bottleneck5	6	64	3	1
$19^2 \times 64$	bottleneck6	6	160	3	2
$10^2 \times 160$	bottleneck7	6	320	1	2
$10^2 \times 320$	conv2d	—	512	1	1
$5^2 \times 512$	conv7_2	—	256	1	—
$3^2 \times 256$	conv8_2	—	256	1	—
$1^2 \times 256$	conv9_2	—	—	1	—

瓶颈残块(bottleneck residual block)是网络中很重要的结构。它把数据先升维,使用轻量级深度卷积做过滤;再应用深度可分离卷积进行处理。为便于实际应用,如果维度越低,卷积层的乘法计算量就越小,想要减少运算量,就需要尽可能将网络设计成低维度的。如果整个网络都是低维的,那么整体计算速度就会很快。但是如果维度太低,激活变换 ReLU 函数可能会滤除很多有用信息,造成信息丢失或破坏。针对这个问题,使用 linear bottleneck 来代替原本的非线性激活变换,而在提取特征数据阶段,高维张量更能提取到足够多的信息,故先对数据进行升维,再降维。

2.2 特征增强和特征融合

为了提高大特征图对小目标的检测能力,丰富不同尺度特征图的特征信息,提出了特征增强模块。如图2所示,模块整体为一个残差网络结构,相比于学习原始的信号,残差网络学习到了信号的差值,还增加了网络的宽度。残差网络可以通过跳层连接,增强了梯度的流动,从而使得信息前后向传播更加顺畅,不对称的结构使得反向传播时,错误信号可以不经任何中间权重矩阵变换直接传播到低层,一定程度上可以缓

解梯度弥散问题。多重 1×1 和 3×3 的卷积可以充分挖掘特征图信息,而批量标准化可以解决梯度消失的问题,由于小目标的文本信息量较少,故这样设计特征增强模块可以有效增强对小目标的检测能力。

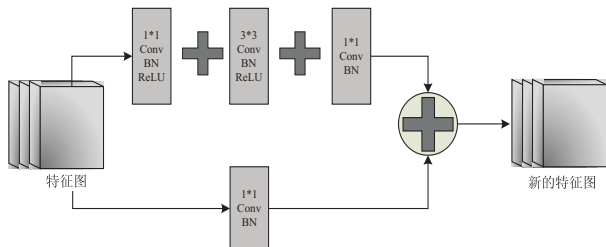


图2 特征增强模块结构

为了加强各个尺度特征图之间的联系,增强检测能力,提出特征融合模块,结构如图3所示。首先,用 1×1 的卷积核对两个特征图进行卷积,使其通道统一,然后采用双线性插值使特征图的大小统一,之后用 `concat` 函数进行合并。特征融合能使得特征图的语义更丰富,对大尺度特征图进行特征融合可以提高对小目标的检测能力。

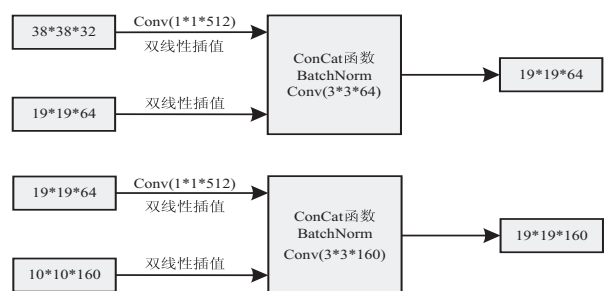


图3 特征融合模块结构

SSD 模型在目标检测部分与图像金字塔有着相似的结构,对图像下采样得到不同大小的特征图。由于大尺寸特征图的细节特征信息丰富,能检测出细微轮廓,而小尺寸的特征图丢失较多细节信息,但对较大轮廓检测效果好,因此对不同特征层进行融合能够提高网络语义信息,增强细节与整体之间的联系,提高网络对小目标的检测能力。

特征融合具体步骤如下:该文采用特征融合的思想,在以上特征增强的基础框架上,首先对不同大小特征层,用 `padding=0` 的 1×1 卷积核在这些特征图上做卷积操作,保证了融合时的通道统一。由于这里把不同大小的特征图进行融合,所以需采用双线性插值操作使特征图大小统一。改进后的特征层有更丰富的语义信息,能更好表征目标的特性,对车辆小目标检测的准确率有较大幅度改善。由于 SSD 中的检测模块的特征层尺寸较小,且包含语义信息较强,不再进行特征增强。

2.3 先验框的改进

传统的 SSD 网络检测目标时,会对所有先验框解

码,然后进行置信度排序,置信度低的被删除。该文提出的改进是在解码前先对先验框进行置信度排序,在每个同目标类别中去除置信度较低的先验框,保留一定数量的高置信度先验框,再进行解码得到真实框,如图4所示。这样做的目的是减少需要解码的先验框个数,减少了计算量。

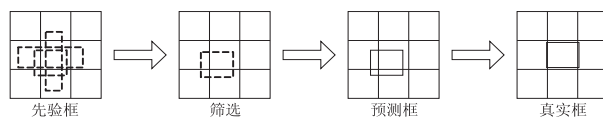


图4 先验框改进后解码过程

考虑到本实验中目标种类较少,且要求模型轻量化,故只保留 50 个同类别的先验框。即对先验框进行置信度排序后,同一目标类别中,只保留 50 个置信度最高的先验框。之后对每一类别的 50 个先验框进行解码,把大小和位置修正,再从中选出 IOU 最大的预测框作为真实框输出。

为了提高对小目标的检测能力,调整先验框的尺寸。特征图先验框的个数预测依次为 4、6、6、6、4、4。先验框长宽比依次为 $1:1, 2:2, 1:2, 2:1, 1:3, 3:1$, 先验框尺度计算公式如式(1)所示。

$$S_k = S_{\min} + \frac{S_{\max} - S_{\min}}{m - 1}(k - 1), k \in [1, m] \quad (1)$$

式中, S_k 表示先验框的尺寸相对于原图的比例, S_{\min} 和 S_{\max} 分别表示比例的最小值与最大值。为了更好地检测小目标,将 S_{\min} 和 S_{\max} 的值调整为 0.16 和 0.88,这样能够有效避免由于目标过小导致在训练阶段真实标签无法找到相应的先验框与之匹配的问题。尺寸调整后的先验框大小如表2所示。

表2 先验框尺寸

特征图	k	S_k	Min_size	Max_size
BN4	—	0.08	24	48
BN4-6	1	0.16	48	102
BN6-7	2	0.34	102	156
BN7	3	0.52	156	210
conv7_2	4	0.7	210	264
conv9_2	5	0.88	264	318

SSD 默认框长宽比采用 $t \in \{1, 2, 3, 1/2, 1/3\}$ 以增强先验框长宽比对目标形状的鲁棒性。针对先验框长宽比 $t=1$ 的先验框,添加 $S_k^* = \sqrt{S_k S_{k+1}}$ 形成大正方形先验框。对特征图 BN4、Conv7_2、Conv9_2 不使用长宽比为 3、1/3 的先验框,而在特征图 BN4-6、BN6-7、BN7 都使用 6 个默认框,使用上述所有长宽比。每个先验框长度、宽度的计算公式如下:

$$h = S_k \sqrt{t}, w = \frac{S_k}{\sqrt{t}}, t \in \{1, 2, 3, 1/2, 1/3\} \quad (2)$$

先验框中心坐标为 $(\frac{i+0.5}{|h_k|}, \frac{j+0.5}{|h_k|})$, 其中 h_k 是第 k 个特征图尺寸大小。 $i, j \in \{0, 1, \dots, |h_k| - 1\}$, 这里的 i, j 表示网格单元的序号 (第几行第几列), 因为 SSD 采用了不同大小的网格, 所以先验框坐标要归一化到 $[0, 1]$, 使它们独立于网格大小。

2.4 损失函数的计算

由于针对小目标检测, 正负样本特征的不平衡, 往往造成简单的负样本主导网络的训练并导致模型的退化, 这严重制约了检测网络的性能。为解决此问题, 该文引入 Focal Loss 损失函数。具体可以表示为:

$$FL(p_i) = -\alpha_i (1 - p_i)^\gamma \log(p_i) \quad (3)$$

其中, α_i 用于控制正负样本的权重, 取比较小的值来降低负样本的权重; $(1 - p_i)^\gamma$ 用于控制难易样本的权重, 目的是通过减少易分样本的权重, 从而使得模型在训练的时候更加专注难分样本的学习; p_i 表示某个目标的类别概率。

损失函数由位置损失和分类损失加权组成, 公式如下:

$$L(t^*, t, b^*, b) = \frac{1}{N} (L_{\text{class}}(t^*, t) + a L_{\text{loc}}(b^*, b)) \quad (4)$$

其中, t^* 为真实的目标类别, t 为预测目标类别, b^* 为真实的标签框位置坐标, b 为预测目标坐标, N 表示与真实框相匹配的先验框的个数, $L_{\text{class}}(t^*, t)$ 为分类损失函数, a 为权重, $L_{\text{loc}}(b^*, b)$ 为位置损失函数。

位置损失可表示为:

$$L_{\text{loc}}(b^*, b) = \sum_{i \in \text{Pos}} \sum_m \text{smooth}_{L_i}(b_i^* - b_i) \quad (5)$$

其中, Pos 表示正样本, m 为先验框的位置 $m \in \{x, y, w, h\}$, x, y 表示先验框的中心点坐标, w, h 分别表示先验框的宽和高。Smooth $_{L_i}$ 损失函数在 Fast R-CNN 中被提出, 主要目的是为了防止梯度爆炸。

分类损失可表示为:

$$L_{\text{class}}(t^*, t) = - \sum_{c=1}^l \sum_{i \in \text{Pos}} \alpha_i (1 - t_{i,c})^\gamma \log(t_{i,c}) - \sum_{c=0}^l \sum_{j \in \text{Neg}} (1 - \alpha_j) t_{j,c}^* \log(1 - t_{j,c}^*) \quad (6)$$

其中, $t_{i,c}$ 为预测目标为目标的类别, $t_{j,c}^*$ 表示真实类别为背景的目标类别, Neg 表示负样本, c 为目标类别, l 为先验框个数, α_i 和 γ 为超参数, 且 $\alpha_i \in [0, 1]$, $\gamma \in [0, 5]$ 。

3 实验结果与分析

本实验用的计算机配置为 windows10 操作系统, Intel Core(TM) i9-10900K 处理器, 内存为 16 GB, 显卡为 NVIDIA GTX 1080, 开发环境使用的是

Tensorflow1.8。

3.1 数据集介绍

实验使用的是 KITTI 数据集, 并进行了调整。KITTI 数据集是目前国际上最大的自动驾驶场景下的计算机视觉算法评测数据集, 该数据集包含市区、乡村和高速公路等场景采集的真实图像数据, 每张图像中最多达 15 辆车和 30 个行人, 还有各种程度的遮挡与截断。KITTI 数据集有 8 个类别, 将标注文件中 “Van”、“Truck”、“Tram” 和 “Car” 类别合并为 “Car” 类, 其他类别作为背景。

为了增加训练的数据量, 提高模型的泛化能力, 同时增加噪声数据, 提升模型的鲁棒性, 对数据集进行数据增强。比如使用随机缩放和旋转对样本做一点处理, 随机地缩放小目标的 size (20% 以内) 还有随机旋转 (15° 左右), 或者对样本进行随机反转、裁剪、平移、添加噪声。

3.2 实验参数设置

经过数据增强后, 数据集总共有 7 600 张图片, 选用其中 6 200 幅图像作为训练集, 剩下 1 400 幅图像作为验证集。实验中设定冲量常数为 0.9, 权重衰减系数为 0.000 5, 学习速率为 0.000 8。在训练集中每次迭代处理 8 个样本, 进行 240 个迭代周期, 训练迭代次数为 54 000 次。

3.3 模型评价

以模型训练时的损失函数变化和测试时的各类别 AP 的平均值作对比分析, 评价模型性能的优劣。其中, 损失函数由位置损失和分类损失加权组成, 计算总损失值与迭代次数的关系, 各类别 AP 的平均值即 mAP (mean average precision) 由召回率和精确率得出。

召回率的计算方法为:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

精确率的计算方法为:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

式中, TP (true positive) 指正样本中预测为正的数, FN (false negative) 指正样本中预测为负的数量, FP (false positive) 指负样本中预测为正的数。

3.4 实验结果

模型对数据增强后的 KITTI 数据集进行检测, 得到的实验结果如图 5 所示。图 5 (a) 为损失函数的损失值与迭代次数之间的关系图, 训练完成后的损失值为 0.03, 图 5 (b) 为训练时改进后的先验框的置信度得分和训练次数的关系图。

实验结果表明, 改进网络的损失值较小, 置信度得分较高, 证明了网络的有效性。

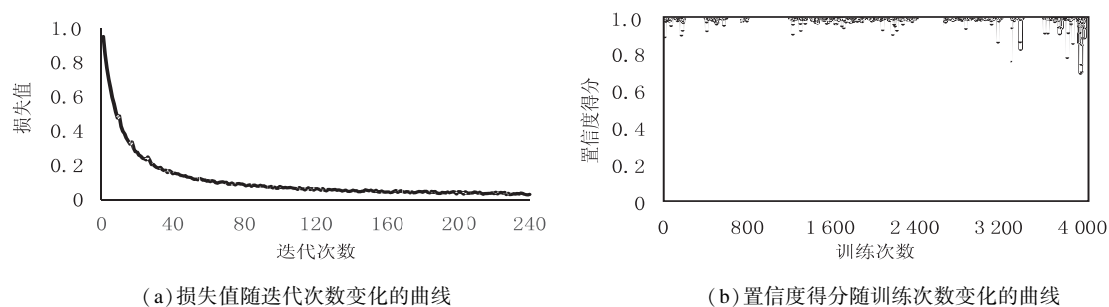


图 5 损失值和置信度得分的变化曲线

图 6 为改进型 MobileNetv2-SSD 网络的检测效果。其中,(a)为算法对远处的小目标的检测效果;(b)为在强光照条件下,算法对正面目标和背面目标的检测效果;(c)为算法对侧面目标和被截断目标的

检测效果;(d)为目标处在被遮挡的阴影背景下算法的检测效果。可以看到,在复杂多变的应用场景下改进后的算法仍有较好的检测效果。但为了突出其先进性,需与其他算法进行对比分析。

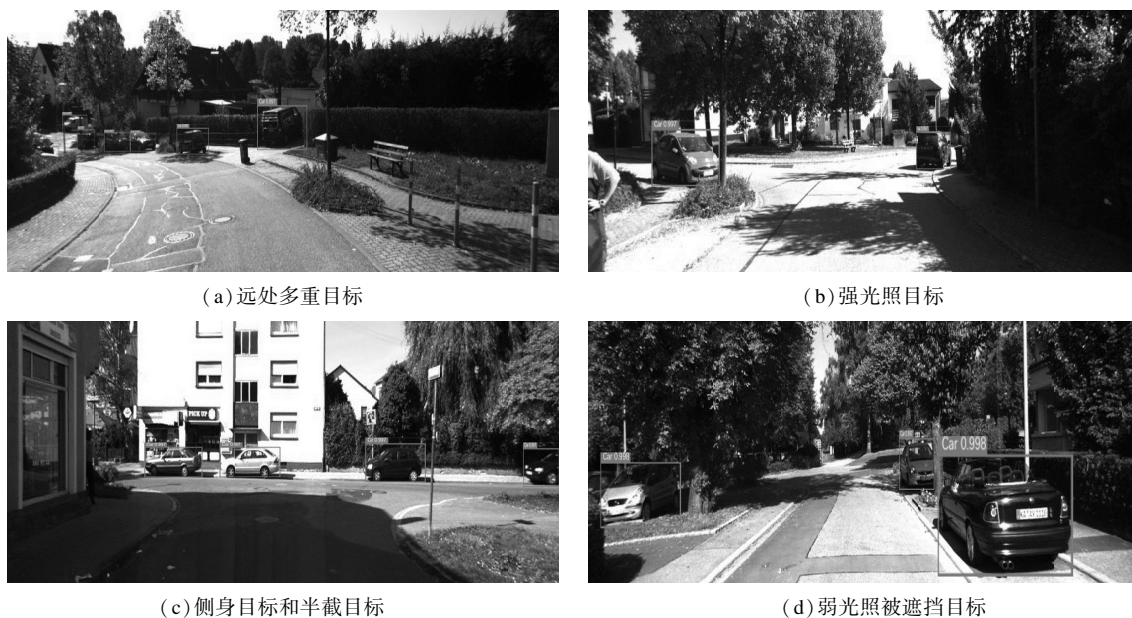


图 6 改进型 MobileNetv2-SSD 网络的检测效果

3.5 对比分析

首先与未进行改进的 MobileNetv2-SSD 网络作对比。图 7(a)为原本的 MobileNetv2-SSD 网络的检测效果,(b)为改进型 MobileNetv2-SSD 网络的检测效果。可以看出,原本的 MobileNetv2-SSD 网络对远处

的小目标检测有遗漏,对被截断的目标无法识别,对给较少信息的目标检测能力较差,而改进型 MobileNetv2-SSD 网络有针对性地提高了对小目标的检测能力,对目标信息反复挖掘,所以检测效果较好。



图 7 两种网络的检测效果对比

之后是与其他算法作对比分析。表3是不同检测算法在KITTI数据集上的检测结果对比,Faster R-CNN由于把目标识别和定位分两步完成,导致模型较大,而且检测速度较慢。其他的算法都是单步识别网络,同时进行目标的识别和定位,所以检测速度较快。其中,YOLOv2有着最快的检测速度,是SSD300检测速度的两倍,但检测准确度比不上SSD300。YOLOv3有着最高的检测准确度,达到了91.58%,但网络仍比较大。而文中的检测网络在检测速度上仅次于YOLOv2,检测准确度上仅次于YOLOv3,而且基本持平,网络大小上远小于其他网络。由此证明,文中的检测网络在综合性能上有较强的优越性。

表3 不同检测网络的检测结果

检测网络	输入	检测速度/ (帧/s)	mAP/ %	网络大小/MB
Faster R-CNN ^[15]	600X-	11.63	76.90	-
SSD300 ^[12]	300X300	58.32	81.00	-
YOLO ^[9]	448X448	68.37	58.54	753
YOLOv2 ^[7]	416X416	114.26	69.48	193
YOLOv3 ^[16]	416X416	48.60	91.58	236
文中网络	300X300	95.21	90.17	50.41

4 结束语

基于MobileNetV2-SSD网络进行改进,针对SSD网络的小目标检测能力不足的问题增强特征的提取和运用,为减小计算量,对先验框的使用过程进行改进。实验表明,改进后的网络相比原网络有更高的准确度,同时检测速度更快,这在车辆小目标检测中有较高的应用价值。该实验的进一步研究是扩大检测目标的种类,增加对行人、电动车、摩托车等目标的检测,以增强应用价值。

参考文献:

- [1] IANDOLA F N, HAN S, MOSKEWICZ M W, et al. SqueezeNet: AlexNet-level accuracy with 50X fewer parameters and 0.5 MB model size[J]. arXiv:1602.07360, 2016.
- [2] SIFRE L. Rigid-motion scattering for image classification[D]. Paris: Ecole Polytechnique, 2014.
- [3] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[C]//IEEE conference on computer vision and pattern recognition. Hawaii, USA: IEEE, 2017: 1-9.
- [4] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: inverted residuals and linear bottle necks[C]//IEEE conference on computer vision and pattern recognition. Salt Lake City: IEEE, 2018: 4510-4520.
- [5] 许必霄, 宫婧, 孙知信. 基于卷积神经网络的目标检测模型综述[J]. 计算机技术与发展, 2019, 29(12): 87-92.
- [6] 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. Las Vegas, Las Vegas, USA: IEEE, 2016: 429-442.
- [7] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//2017 IEEE conference on computer vision and pattern recognition. Honolulu, USA: IEEE, 2017: 6517-6525.
- [8] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Computer vision - ECCV 2016. Amsterdam, The Netherlands: Springer, 2016: 21-37.
- [9] LIN T Y, DOLL R P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Hawaii, USA: IEEE, 2017: 2117-2125.
- [10] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318-327.
- [11] TANG C, LING Y S, ZHENG K D, et al. Object detection method of multi-view SSD based on deep learning[J]. Infrared and Laser Engineering, 2018, 47(1): 302-310.
- [12] JEONG J, PARK H, KWAK N. Enhancement of SSD by concatenating feature maps for object detection[C]//2017 IEEE conference on computer vision and pattern recognition. Hawaii, USA: IEEE, 2017: 236-246.
- [13] FU C Y, LIU W, RANGA A, et al. DSSD: deconvolutional single shot detector[C]//2017 IEEE conference on computer vision and pattern recognition. Hawaii, USA: IEEE, 2017: 2372-2376.
- [14] CHEN H J, WANG Q Q, YANG G W, et al. SSD object detection algorithm with multi-scale convolution feature fusion[J]. Journal of Frontiers of Computer Science and Technology, 2019, 13(6): 1049-1061.
- [15] 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 & Machine Intelligence, 2017, 39(6): 1137-1149.
- [16] LI Y D, HAN Z Q, XU H Y, et al. YOLOv3-Lite: a lightweight crack detection network for aircraft structure based on depthwise separable convolutions[J]. Applied Sciences, 2019, 9(18): 3781.