

doi: 10.11835/j.issn.1000-582X.2020.288

# 一种新的结合卷积神经网络的隧道内停车检测方法

杨祖葭, 丁 洁, 刘晋峰

(太原理工大学 电气与动力工程学院, 太原 030024)

**摘要:**为了更准确地检测高速公路隧道内停车行为,将传统的图像处理技术与深度学习相结合。首先,通过基于混合高斯模型(Gaussian mixture model, GMM)的背景差分法提取出运动目标。接着通过 meanshift 算法跟踪这些运动目标,计算运动目标的速度以及在相邻视频帧中运动目标的相关性,通过与速度阈值和相似度阈值的比较得到静止目标。最后,结合卷积神经网络(convolutional neural network, CNN)分类模型识别静止目标是否为车辆。文中方法在高速公路隧道视频上进行测试,达到至少 84% 的准确率。另外,与没有结合卷积神经网络的传统图像处理方法相比,文中方法至少提高了 63% 的准确率。

**关键词:**人工智能; 计算机视觉; 图像处理; 卷积神经网络

**中图分类号:** TP751

**文献标志码:** A

**文章编号:** 1000-582X(2021)06-049-11

## A new tunnel vehicle stopping detection methodology combined with convolutional neural network

YANG Zuliang, DING Jie, LIU Jinfeng

(College of Electrical and Power Engineering, Taiyuan University of Technology,  
Taiyuan 030024, P. R. China)

**Abstract:** In order to more accurately detect the vehicle stopping in highway tunnels, this paper proposes a new methodology that combines the traditional image processing technology with deep learning. Firstly, the foreground moving targets are extracted using the background difference method based on Gaussian mixture model (GMM). Then the meanshift algorithm is applied to track these foreground moving targets. By calculating the speed of the moving targets and the correlation of the moving targets between the neighboring video frames, and comparing the results with the speed threshold and correlation threshold, the static target is detected. Finally, combined with the convolutional neural network (CNN) classification model, whether the static target is vehicle is identified. The method proposed in this paper is validated using the real highway tunnel vehicle stopping video and achieves an accuracy of at least 84%. Compared with the traditional image processing method without CNN, our method improves at least 63% accuracy.

**Keywords:** artificial intelligence; computer vision; image processing; convolutional neural network

**收稿日期:** 2020-01-07      **网络出版日期:** 2020-04-22

**基金项目:** 国家自然科学基金项目(61604105)。

Supported by National Natural Science Foundation of China(61604105).

**作者简介:** 杨祖葭(1993—), 太原理工大学硕士研究生, 主要研究领域为深度学习、目标检测及图像处理。

**通讯作者:** 丁洁, 女, 副教授, 主要研究领域为图像处理、新型半导体器件建模, (E-mail) dingjie2015@foxmail.com。

高速公路隧道有着空间狭窄密闭、交通流量大、而车辆速度快的特点,因此当隧道内发生停车事件时,极易造成交通事故。尤其是当载有易燃易爆物品的车辆在隧道内行驶时,发生意外停车会造成更为严重的后果。对隧道内进行及时有效的停车检测不仅可以为事故导致的受伤人员节省更多救援时间,还可以通知该路段上的其他驾驶员减少二次事故的发生<sup>[1]</sup>。因此,实时准确地检测高速公路隧道车辆的停车行为并及时反馈是十分必要的。

传统的停车检测是通过埋在地下的传感器来完成的。然而,将传感器埋在地下不仅不易于安装,会损坏路面,而且不易于维修保养。目前,几乎所有的高速公路隧道都安装了视频监控系統。这为使用图像处理方法来检测提供了有利条件。与传统的检测方法相比,图像处理具有安装方便、信息丰富、检测结果直观、维护成本低等优点。

当前,基于图像处理的停车检测已有了一定的研究。Guler 等<sup>[2]</sup>采用 Peripheral 多目标跟踪器从背景减法获得的前景中确定每辆车的位置。然而,该方法会将快速变化的背景检测为车辆。Porikli<sup>[3]</sup>提出了一种基于双背景异常事件检测方法。然而,该方法抗干扰性能差并且如何选择时间常数来建立长、短背景存在很大不确定性。赵敏等<sup>[4]</sup>使用混合高斯模型获得运动目标,通过基于像素时间序列特征的稳态分析法来检测静止目标,根据车辆的区域特征(车辆的颜色、轮廓、长度和宽度)对静止目标的停车车辆进行识别。但是该方法仍然缺乏对车辆目标区域特征识别的鲁棒性。此外,一些研究人员对基于混合高斯模型的方法进行了改进,但仍然没有取得良好的效果<sup>[5-6]</sup>。

本课题检测停车环境为高速公路隧道,为停车检测增加了难度,例如车灯、环境光、车辆目标遮挡都是较为严重的干扰因素。因此,需要新的方法来克服这些问题,自 2012 年 CNN 首次被应用于 ILSVRC 图像分类竞赛并取得显著成绩<sup>[7]</sup>以来,CNN 在图像分类领域得到了广泛应用。GoogleNet 通过引入 Inception 模块获得了 2014 年 ILSVRC 冠军<sup>[8]</sup>。2015 年,另一个更先进的 CNN——ResNet 赢得了冠军<sup>[9]</sup>,ResNet 采用跨层连接的方法,成功缓解了深层神经网络中的梯度消失问题,为数千层网络的训练提供了可能。2017 年 Hu 等<sup>[10]</sup>提出了 SENet,该网络通过学习的方式自动获取每个特征通道的重要程度,在 ILSVRC 2017 竞赛中以 2.251% 的 Top-5 错误率获得了第一名。近年来,由于 CNN 在图像分类领域取得了优异的性能,在本研究中,提出一种新的方法,通过将静止目标检测与本文设计的专用于车辆识别的 CNN 分类模型相结合,对停止车辆进行准确检测。本文的内容安排如下:第一部分构建了停车检测算法的结构,并对算法的细节进行了阐述。第二部分描述了所进行的实验,并对相应结果进行了讨论与分析。第三部分对全文做出了总结。

## 1 结合卷积神经网络的停车检测方法

本文设计的停车检测算法分为三步,如图 1 所示。第一步,在输入视频帧中提取运动目标。第二步,在第一步提取出的运动目标中检测静止的停车目标。需要注意的是,这里的静止目标不同于总是静止的背景对象,而是先动后停的目标,即行驶的车辆发生了停车行为。然而,上述得到的静止目标中往往不仅包含车辆,还包含非车辆物体的干扰,这将导致错误的检测发生。因此,第三步加入车辆识别,利用 CNN 分类模型识别第二步的静止目标是否为车辆,最终完成停车检测。

### 1.1 运动目标提取

运动目标的提取是目标跟踪的前提。因此,要求运动目标提取算法具有较高的处理效率和较强的鲁棒性。常用的运动目标提取算法可以归纳为三类:帧间差分法、光流法和基于 GMM 的背景差分法。

帧间差分法<sup>[11]</sup>通过间隔帧中减去相同点的灰度值来获得灰度的变化值,通过对该值进行阈值化来提取运动区域。但是该方法只能提取边界,不能提取运动目标的整个区域。此外,帧间间隔大小对目标识别的结果也有很大的影响。光流法通过对空间运动场的比较,将物体的运动表示为图像上的光流场<sup>[12]</sup>。然而,此方法计算量大,耗时长,不适用于实时性要求严格的情况。此外,隧道内的车灯可能被误认为光流,从而影响最终的结果。基于 GMM 的背景差分法<sup>[13,14]</sup>利用混合高斯模型建立背景,通过背景差分法提取运动目标。此方法能够及时更新背景,降低误检率,并能够提取目标的完整区域,满足实时性的要求。同时,该方法将重复运动的目标判断为背景,有效地减少了摄像机抖动引起的误判。因此,以基于 GMM 的背景差分法作为提取运动目标的核心算法。

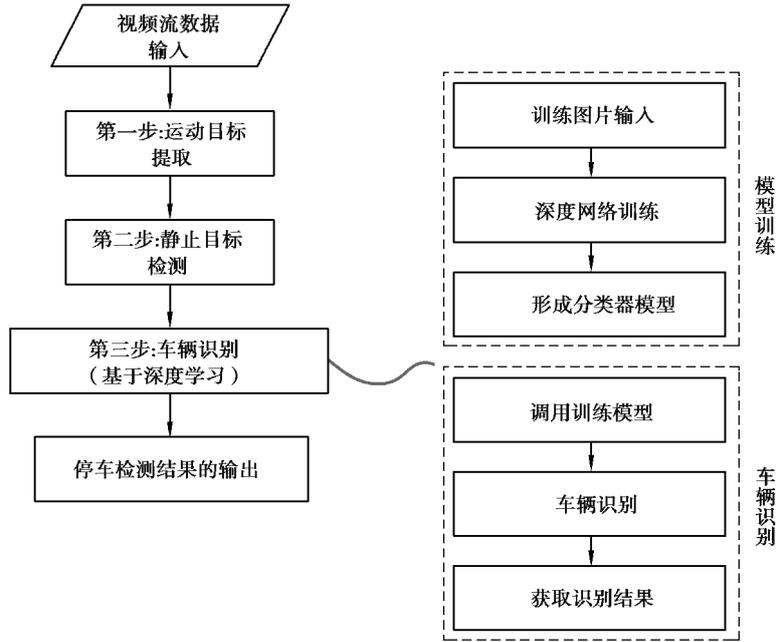


图 1 车辆停车检测方法的流程

Fig. 1 The flow of vehicle stopping detection methodology

基于 GMM 的背景差分法在导入图像后通过背景建模及背景差分的方法分离出前景和背景。具体做法为,设  $t$  时刻图像中每个像素点的像素值为  $X_t$ ,则该时刻的混合高斯模型  $P(X_t)$  为:

$$P(X_t) = \sum_{k=1}^K \omega_{k,t} \eta(X_t, \mu_{k,t}, \Sigma_{k,t}) \quad (1)$$

式中: $K$  是混合高斯模型中高斯概率密度函数的个数, $K$  越大越能描述更复杂的背景,一般取值为 3~5,这里取值为 5; $\omega_{k,t}$  为第  $k$  个高斯概率密度函数在  $t$  时刻的权值,且权值满足条件  $\sum \omega_{k,t} = 1$ ; $\eta(X_t, \mu_{k,t}, \Sigma_{k,t})$  是像素  $X_t$  在  $t$  时刻的第  $k$  个高斯概率密度函数,

$$\eta(X_t, \mu_{k,t}, \Sigma_{k,t}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_{k,t}|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t - \mu_{k,t})^T \Sigma_{k,t}^{-1} (X_t - \mu_{k,t})} \quad (2)$$

式中: $\mu_{k,t}, \Sigma_{k,t}$  分别为第  $k$  个高斯概率密度函数在  $t$  时刻的均值、协方差矩阵; $n$  为所处理图像的通道数,这里处理的是 RGB 三通道彩色图像, $n=3$ ;一般假设像素点的 RGB 色彩空间的各通道相互独立,并具有相同的方差,则其协方差矩阵为  $\Sigma_{k,t} = \sigma_{k,t}^2 \mathbf{I}$ ,式中  $\sigma_{k,t}$  为标准差, $\mathbf{I}$  为三维单位矩阵。以上步骤建立了  $t$  时刻的混合高斯模型。

建立混合高斯模型之后,需要不断进行模型的更新,使模型更加鲁棒地适用于隧道灯光及其他干扰因素导致的背景变化。对  $t$  时刻的混合高斯模型的更新步骤如下:

1) 每个  $t$  时刻的像素值  $X_t$  与当前  $K$  个高斯概率密度函数进行比较,如果满足式(3),即像素值  $X_t$  与该高斯概率密度函数的均值  $\mu_{k,t-1}$  的偏差在  $T_1 \sigma_{k,t-1}$  内,则认为像素值  $X_t$  与第  $k$  个高斯概率密度函数匹配。

$$|X_t - \mu_{k,t-1}| \leq T_1 \sigma_{k,t-1} \quad (3)$$

式中  $T_1$  为匹配阈值,根据文献[4], $T_1 = 2.5$ 。

2) 在提取运动目标时,若像素值  $X_t$  与前  $B_{t-1}$  个高斯概率密度函数中的任何一个匹配,则认为该像素为背景;反之为前景。 $B$  是由第 5)步进行更新的。

3) 各个高斯概率密度函数的权值按式(4)更新,

$$\omega_{k,t} = (1 - \alpha)\omega_{k,t-1} + \alpha \times M_{k,t}, \quad (4)$$

式中  $\alpha$  为学习率, 这里设为 0.002; 若像素点  $X_t$  与第  $k$  个高斯概率密度函数匹配, 则  $M_{k,t}$  为 1, 否则为 0。

4) 与像素值  $X_t$  未匹配的高斯概率密度函数的均值  $\mu_{k,t}$  和标准差  $\sigma_{k,t}$  不更新, 而与像素值  $X_t$  匹配的高斯概率密度函数按公式(5)(6)更新,

$$\mu_{k,t} = \left[ 1 - \alpha \times \eta(X_t, \mu_{k,t}, \Sigma_{k,t}) \right] \mu_{k,t-1} + \alpha \times \eta(X_t, \mu_{k,t}, \Sigma_{k,t}) \times X_t, \quad (5)$$

$$\sigma_{k,t}^2 = \left[ 1 - \alpha \times \eta(X_t, \mu_{k,t}, \Sigma_{k,t}) \right] \sigma_{k,t-1}^2 + \alpha \times \eta(X_t, \mu_{k,t}, \Sigma_{k,t}) (X_t - \mu_{k,t})^T (X_t - \mu_{k,t}). \quad (6)$$

5) 完成参数的更新后, 根据  $\omega_{k,t}/\sigma_{k,t}$  从大到小对高斯概率密度函数进行排序, 将满足式(7)的前  $B_t$  个高斯概率密度函数作为背景:

$$B_t = \arg \min_b \left( \sum_{k=1}^b \omega_{k,t} > T_2 \right). \quad (7)$$

式中阈值  $T_2$  表示背景高斯成分在整个像素的概率分布中所占的比例大小, 根据文献[15],  $T_2$  取经验值 0.7。

6) 循环上述过程, 遍历图像中所有的像素点后即可提取出运动目标。

通过处理连续的输入视频图像, 可以连续地检测监控场景中的运动目标。接下来, 在第 1.2 节中执行静止目标检测。

### 1.2 静止目标检测

该步骤在 1.1 节所获得的运动目标中检测静止目标, 检测过程分为两个阶段, 第一阶段通过估计运动目标的速度, 从运动目标中检测可疑的静止目标。在第二阶段中, 计算相邻间隔帧之间的可疑静止目标区域的相关性, 若相关性系数大于一定阈值, 则认为该区域是真实静止目标区域。

显然, 在第一阶段中, 当运动目标的速度为零时, 运动目标处于停止状态。因此, 可以通过计算运动目标的速度值来检测目标是否停止。具体来说, 通过 meanshift 跟踪算法<sup>[16-18]</sup>可以获得运动目标质心在间隔  $N$  帧中的移动距离, 然后利用运动目标质心移动距离和间隔  $N$  帧对应时间来计算运动目标的速度, 计算公式如式(8)所示:

$$v = \Delta d / \Delta t. \quad (8)$$

式中:  $\Delta d$  是运动目标质心在间隔  $N$  帧中移动的距离;  $\Delta t$  是间隔  $N$  帧相对应的时间。

在理想情况下, 如果  $v$  等于 0, 则认为目标是静止的。但是, 由于运动目标阴影和摄像机抖动等干扰因素对质心位置的影响,  $\Delta t$  间隔时间后静止目标的质心可能不完全重合。因此, 引入了速度阈值  $\lambda$ 。当  $v < \lambda$ , 目标区域则被确定为可疑的静止目标区域。

第二阶段, 计算相邻间隔帧中由第一阶段获取的可疑静止目标区域之间的相关性。相邻间隔帧中需要计算的区域被视为 2 个变量  $X, Y$ , 则相关性系数计算公式如式(9)所示:

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}. \quad (9)$$

式中:  $\text{cov}(X,Y)$  为区域  $X$  和区域  $Y$  的协方差;  $\sigma_X$  和  $\sigma_Y$  分别为区域  $X$  和区域  $Y$  的标准差。如果相关性系数  $\rho_{X,Y}$  大于相关性阈值, 则将可疑的静止目标区域确定为真实静止目标。

### 1.3 车辆识别

在检测到静止目标后, 利用本文设计的 CNN 模型对静止目标进行识别。其中, 1.2 节中获得的静止目标区域作为 CNN 模型的输入。本文设计的 CNN 模型由 5 个卷积层组成, 如图 2 所示。从第一层到第五层使用  $3 \times 3$  大小的卷积核, 通过多层叠加小卷积核, 增强了模型容量和模型复杂度, 并且减少了模型中参数的数量, 加快训练和推理速度。本文所设计的 CNN 分类模型结构基于 Keras 神经网络框架实现。

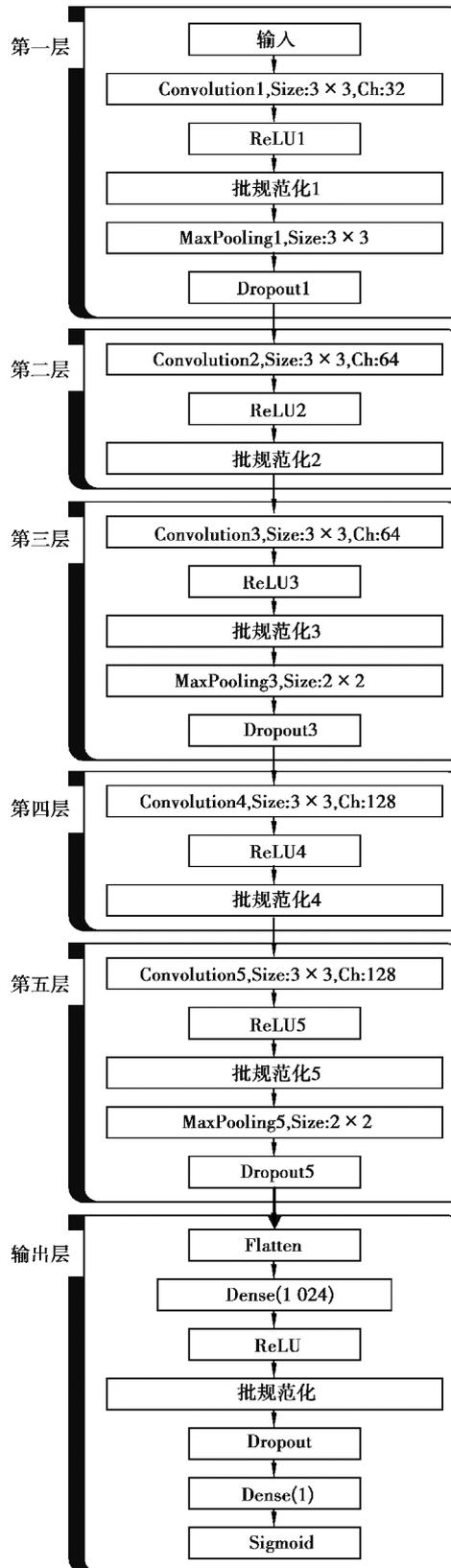


图 2 CNN 分类模型的结构

Fig. 2 The structure of the CNN classification model

在图 2 中,每一个卷积层都以修正线性单元(ReLU)作为卷积后的激活函数。此外,在 ReLU 激活函数之后执行批规范化<sup>[19]</sup>,将其输出规范化到均值为 0、方差为 1 的标准正态分布,缓解网络训练中的梯度消失或爆炸现象,加快模型的训练速度。批规范化分为 4 个步骤。设训练过程中的输入是  $x_i$ ,前两步计算输入数据的均值  $\mu_B$  和方差  $\sigma_B^2$ 。在第三步中,根据前两步中计算出的平均值和方差,使用式(10)进行规范化。

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \quad (10)$$

式中: $\hat{x}_i$  是规范化后的结果, $\epsilon$  是 0.001,用于防止除 0 错误。由于  $\hat{x}_i$  被限于正态分布,会导致网络表达能力下降。为了解决这个问题,第四步引入了 2 个新的参数  $\gamma$  和  $\beta$ ,对  $\hat{x}_i$  的分布进行修正,这 2 个参数值在训练过程中得以更新。批规范化后的最终输出如下:

$$y_i = \gamma x_i + \beta \quad (11)$$

在第一层、第三层、第五层卷积之后采用 Max Pooling 进行池化,将 Feature Map 的尺寸减半。在第一层、第三层、第五层和输出层采用 Dropout 技术<sup>[20]</sup>。Dropout 会随机丢弃一部分神经元及其连接,防止模型训练的过拟合。由于在训练过程中神经元的丢弃是随机的,因此每次训练都相当于处理一个全新的神经网络结构,显著提高本文设计的 CNN 模型的泛化效果。

利用上述设计的 CNN 分类模型,可以排除干扰因素引起的虚假检测,确定静止目标是否为车辆,从而提高检测精度。

#### 1.4 算法步骤

1) 读取一帧图像 frame1,如图 3(a)所示,进行前景像素点的检测,得到的前景像素点图像如图 3(b)所示。对前景像素点图像进行开运算操作,即先腐蚀后膨胀,去除小颗粒噪声和使较大物体的边界更平滑,基本保持目标原有大小不变,如图 3(c)所示。

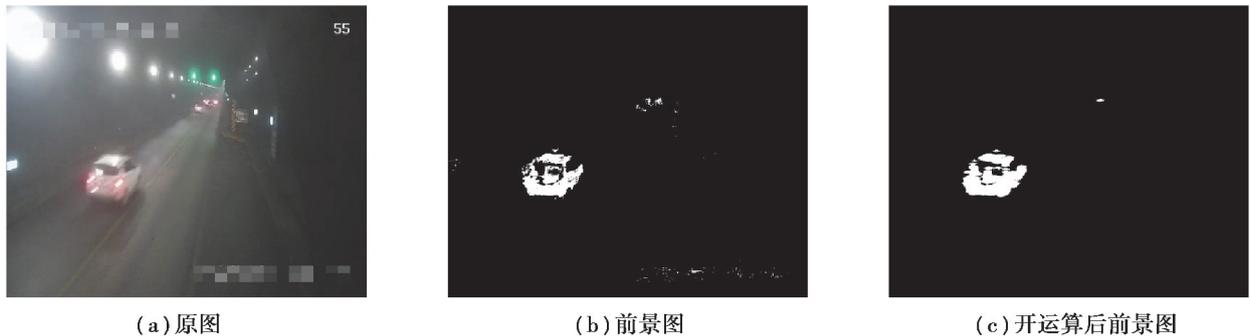


图 3 运动目标检测中的图像  
Fig. 3 Image in moving object detection

2) 检测出前景像素点后,通过连通域分析提取运动目标区域的矩形框(即获得了运动目标区域的形状和位置)。间隔 10 帧后再取一帧 frame2,根据 meanshift 算法跟踪运动目标,即计算 frame1 中的运动目标在 frame2 中相应的位置。

3) 计算运动目标质心的移动距离,再结合间隔帧所对应的耗时来计算车辆的粗略速度:

$$\text{速度} = \frac{\text{质心位移距离}}{\text{间隔帧数} / \text{视频帧率}} \quad (12)$$

根据实验中的调试经验,引入速度阈值 200,当速度小于速度阈值时判定该运动目标为可疑静止目标。

4) 计算 frame1 和 frame2 中该可疑静止目标区域的相关性系数,若大于相关性阈值,则将可疑的静止目标区域确定为静止目标。经过实验,相关性阈值设为 0.8 时获得了最佳效果。

5) 将第 4 步中获取的静止目标区域输入到本文设计的 CNN 隧道车辆识别模型。通过该模型判断静止目标是否为真实车辆(如图 4),如果静止目标被识别为车辆的话则认为发生停车(如图 4 中的(b)(d)),否则没有发生停车行为(如图 4 中的(a)(c))。



图 4 车辆识别中的图像  
Fig. 4 Image in vehicle recognition

6)循环以上步骤,持续对高速公路隧道进行停车检测。

## 2 实验和结果分析

### 2.1 训练 CNN 分类模型

在本研究中,建立了隧道车辆数据集,专门用于隧道车辆分类。隧道车辆数据集包含 1 664 幅图像,其中有车图像 832 幅,无车图像 832 幅。训练集和测试集的数量如表 1 所示。训练集是在样本集中随机抽取组成的,余下的则组成测试集。值得注意的是,本文的数据集中所包含的图片均来自太原市公路管理系统,但是由于保密原因,对图片中敏感位置的标签进行了模糊处理。

表 1 CNN 分类模型的数据集  
Table 1 The data set of the CNN classification model

类别	总数	训练集	测试集
有车图像	832	665	167
无车图像	832	665	167

在 CNN 分类模型的训练过程中,每次从训练集中随机抽取 32 幅图像进行迭代,并计算训练精度和测试精度。图 5 展示了训练过程中评估指标的变化。黑线是训练损失,它反映了模型训练过程中的学习效果,训练损失值越小,学习效果越好;红线表示训练精度;蓝线表示测试精度。模型迭代结束时,训练精度达到 98.81%,测试精度达到 99.70%,此精度符合要求。

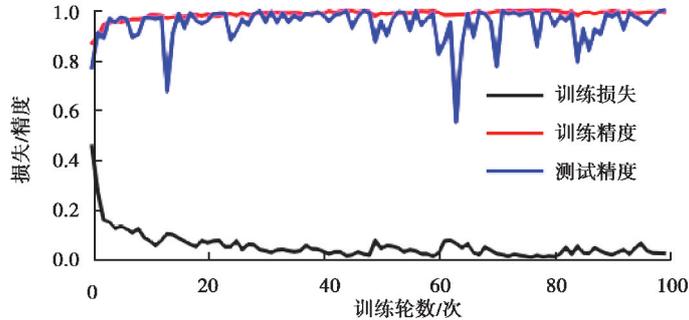


图 5 CNN 分类模型的训练损失、训练精度和测试精度

Fig. 5 The training loss, training accuracy and test accuracy of the CNN classification model

2.2 停车检测的实验结果

首先,通过在高速公路隧道内连接监控摄像服务器进行实时检测,程序运行一天后没有误检。但在此期间,没有发生停车的情况。因此,又从视频历史存储文件中截取了 19 个发生停车事件的监控录像片段,视频的帧率为 25fps,分辨率为 704×576。

为了与传统的图像处理方法进行比较,使用无 CNN 部分的算法进行了实验对比。基于 19 个视频的统计检测结果如图 6 所示,每个视频的详细检测结果如表 2 所示。本文所设计方法的正检率为 84%,误检率为 5%,漏检率为 11%。在没有 CNN 分类模型的情况下,正检率为 21%,误检率为 79%,漏检率为 11%。与没有添加 CNN 的方法相比,本文方法的正检率提高了 63%,误检率降低了 74%。

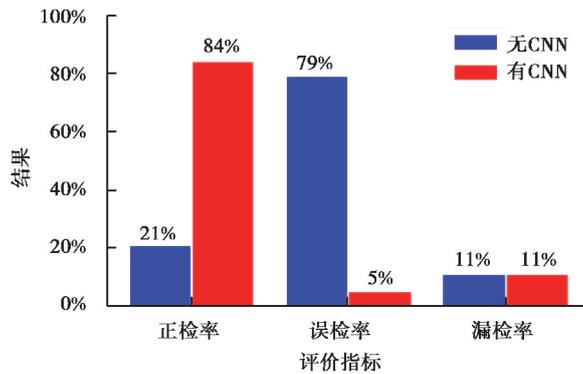


图 6 基于 19 个高速公路隧道视频的车辆停车检测结果

Fig. 6 Detection results of vehicle stopping based on 19 highway tunnel videos

表 2 CNN 分类模型的数据集 19 个高速公路隧道视频停车检测实验结果

Table 2 Experimental results of vehicle stopping detection on 19 highway tunnel videos

视频	正检	误检	漏检	正检	误检	漏检
	无 CNN			有 CNN		
video1	0	1	0	1	0	0
video2	1	0	0	1	0	0

续表 2

视频	正检	误检	漏检	正检	误检	漏检
	无 CNN			有 CNN		
video3	1	0	0	1	0	0
video4	0	1	0	1	0	0
video5	0	1	0	1	0	0
video6	0	1	0	1	0	0
video7	1	0	0	1	0	0
video8	1	0	0	1	0	0
video9	0	1	1	0	0	1
video10	0	1	0	1	0	0
video11	0	1	0	1	0	0
video12	0	1	0	1	0	0
video13	0	1	0	1	0	0
video14	0	1	0	1	0	0
video15	0	1	0	1	0	0
video16	0	1	0	0	1	0
video17	0	1	0	1	0	0
video18	0	1	0	1	0	0
video19	0	1	1	0	0	1

结果表明,如果没有结合 CNN 模型,检测精度会降低。背后原因在于:受光线或其他干扰因素影响的道路区域很容易被误认为是停车区域,因此会降低检测精度。图 7 为误检发生时的情况展示,图中矩形框为因干扰因素引起的错误停车检测。然而,由于 CNN 可以成功地识别非车辆区域,因此这些干扰区域并没有产生误判。CNN 分类模型的引入大大减少此类错误的发生,从而提高停车检测算法的可靠性。



图 7 某些场景中错误的检测结果

Fig. 7 Misidentified detection results in some scenes

### 2.3 相关性阈值的实验结果

值得注意的是,上述实验结果是在相关性阈值为 0.8 时获得的,而不同的相关性阈值会对实验结果产生影响。因此,将阈值设置为不同的值作为对比实验,对 19 个视频进行了检测。实验结果如表 3。

表 3 相关性阈值的实验结果

Table 3 Experimental results of correlation threshold

相关性阈值	无 CNN			有 CNN		
	正检率/%	误检率/%	漏检率/%	正检率/%	误检率/%	漏检率/%
0.0	0	100	16	37	53	16
0.2	16	84	26	42	42	26
0.4	5	95	32	37	53	26
0.6	16	84	16	58	26	16
0.7	21	79	16	63	21	21
0.8	21	79	11	84	11	5
0.9	21	79	11	74	11	16
1.0	0	0	100	0	0	100

当相关性阈值为 0.8 时,取得了最高的正检率,为 84%;同时取得了最低的漏检率,为 5%,效果好于其余对比情况。

## 3 结 语

提出了一种基于监控摄像机的停车检测方法。将传统的运动目标检测和跟踪方法与 CNN 分类模型有效地结合起来,能准确地对高速公路隧道停车事件进行检测。其中,本文设计的 CNN 模型包括了先进的  $3 \times 3$  小卷积核、批处理规范化和 Dropout 技术,大大减少了由车灯、环境光等干扰因素引起的误检情况。本文方法在实际高速公路隧道视频中得到了验证,停车检测准确率达到 84%。

### 参考文献:

- [1] Gazzah S, Mhalla A, Amara N E B. Vehicle detection on a video traffic scene: review and new perspectives[C/OL]//2016 7th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT). New York, USA: IEEE, 2017(2017-06-08)[2020-01-01]. <https://doi.org/10.1109/SETIT.2016.7939912>
- [2] Guler S, Silverstein J A, Pushee I H. Stationary objects in multiple object tracking[C/OL]//2007 IEEE Conference on Advanced Video and Signal Based Surveillance. New York, USA: IEEE, 2008(2008-07-07) [2020-01-01]. <https://doi.org/10.1109/AVSS.2007.4425318>
- [3] Porikli F. Detection of temporarily static regions by processing video at different frame rates[C]//2007 IEEE Conference on Advanced Video and Signal Based Surveillance, September 5-7, 2007, London, UK. IEEE, 2007: 236-241.
- [4] 赵敏, 张波, 唐毅, 等. 露天场景下高速公路违章停车视频检测算法[J]. 计算机测量与控制, 2014, 22(2): 489-492. Zhao M, Zhang B, Tang Y, et al. Detection algorithm for highway illegal parking in open-air scenes[J]. Computer Measurement & Control, 2014, 22(2): 489-492. (in Chinese)
- [5] 隋靓, 党建武. 基于运动目标轨迹的高速公路异常事件检测算法研究[J]. 计算机应用与软件, 2018, 35(1): 246-252. Sui J, Dang J W. Traffic anomaly detection based on moving object trajectory[J]. Computer Applications and Software, 2018, 35(1): 246-252. (in Chinese)
- [6] 胡煦, 黄俊, 袁梅. 基于视频的路侧停驶事件实时检测算法[J]. 信息通信, 2019(3): 4-7. Hu X, Huang J, Yuan M. Real-time detection of roadside parking events based on video[J]. Information & Communications, 2019(3): 4-7. (in Chinese)

- [7] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems. New York, USA: Curran Associates Inc, 2012,1:1097-1105.
- [8] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C/OL]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York, USA: IEEE, 2015(2018-08-15)[2020-01-01]. <https://doi.org/10.1109/CVPR.2015.7298594>.
- [9] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//IEEE Conference on Computer Vision & Pattern Recognition. New York, USA: IEEE Computer Society, 2016,1:770-778.
- [10] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks[C]//Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), June 18-22, 2018, Salt Lake City. IEEE, 2018: 7132-7141.
- [11] Cui B, Créput J C. Moving object detection and tracking based on three-frame difference and background subtraction with laplace filter[M]. Artificial Intelligence and Soft Computing. Berlin, Germany: Springer, 2018:3-13.
- [12] Peng Y, Chen Z, Liu C, et al. Moving human detection and extraction via improved optical flow and adjacent region merger[C/OL]//2017 Chinese Automation Congress (CAC). New York, USA: IEEE, 2017(2017-10-01)[2020-01-01]. <https://doi.org/10.1109/CAC.2017.8244112>.
- [13] Ahmad J, Larijani H, Emmanuel R, et al. An intelligent real-time occupancy monitoring system with enhanced encryption and privacy[C/OL]//2018 IEEE 17th International Conference on Cognitive Informatics & Cognitive Computing (ICCI \* CC). New York, USA: IEEE, 2018(2018-10-08)[2020-01-01] <https://doi.org/10.1109/ICCI-CC.2018.8482047>.
- [14] Fan Y X, Wen G J, Li D, et al. Video anomaly detection and localization via gaussian mixture fully convolutional variational autoencoder[J]. Computer Vision and Image Understanding, 2020, 195:102920.
- [15] 高海壮, 段先华. 基于帧差法和混合高斯的海上运动目标检测[J]. 计算机与数字工程, 2019, 47(5): 1140-1144.  
Gao H Z, Duan X H. A method for detecting maritime moving targets based on three-frame difference method and improved hybrid gaussian background model[J]. Computer and Digital Engineering, 2019, 47(5):1140-1144. (in Chinese)
- [16] Cheng Y, Hongyu W, Xiaohong W. Positioning method research for unmanned aerial vehicles based on Meanshift tracking algorithm[C/OL]//Control & Decision Conference. New York, USA: IEEE, 2017(2017-07-17)[2020-01-01]. <https://doi.org/10.1109/CCDC.2017.7978663>.
- [17] Liu J, Zhong X. An object tracking method based on mean shift algorithm with HSV color space and texture features[J]. Cluster Computing, 2019, 22(3):6079-6090.
- [18] Wang B, Feng K, Yang W, et al. Study on multiple targets tracking algorithm based on multiple sensors[J]. Cluster Computing, 2019, 22(6):13283-13291.
- [19] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]//ICML'15: Proceedings of the 32nd International Conference on International Conference on Machine Learning, New York, USA: JMLR.org. 2015, 37:448-456.
- [20] Hinton G E, Srivastava N, Krizhevsky A, et al. Improving neural networks by preventing co-adaptation of feature detectors[J]. Computer Science, 2012, 3(4):212-223.

(编辑 郑洁)