

基于 ST-CNN 的交通标志实时检测识别算法

曲佳博 秦 勃

(中国海洋大学计算机科学与技术系 山东 青岛 266100)

摘要 深度学习是基于图像的交通标志检测和识别处理的研究热点,已取得了显著的效果。针对基于车载视频的交通标志检测和识别处理问题,文中根据图像序列的帧间时空连续关系构建了时空关系模型(Spatiotemporal Model, STM),并与多尺度卷积神经网络(Convolutional Neural Networks, CNN)相结合,提出了一种基于时空卷积神经网络(Spatiotemporal-CNN, ST-CNN)的交通标志实时检测识别算法。实验结果表明,该算法可对视频图像序列中的同一交通标志实现检测、筛选、追踪和识别处理,在保证高准确率的同时,可有效减少 CNN 的数据输入,降低系统资源占用量,提高计算效率,满足了视频中交通标志检测识别的实时性要求。算法平均每帧耗时 26.82 ms,且识别准确率达到了 96.94%。

关键词 时空关系模型,实时性,多尺度卷积神经网络,交通标志

中图分类号 TP391 **文献标识码** A

Real-time Detection and Recognition Algorithm of Traffic Signs Based on ST-CNN

QU Jia-bo QIN Bo

(Department of Computer Science & Technology, Ocean University of China, Qingdao, Shandong 266100, China)

Abstract At present, deep learning is a research hotspot based on image traffic sign detection and recognition processing, and has achieved remarkable results. Aiming at the problem of traffic sign detection and recognition based on car video, this paper proposed a real-time detection and recognition algorithm for traffic signs based on Spatiotemporal-CNN (ST-CNN). It constructs a Spatiotemporal model (STM) based on the spatiotemporal relationship between frames of image sequences, and combines the STM with Convolutional Neural Network (CNN). The experimental results show that the algorithm can detect, screen, track and identify the same traffic sign in the video image sequence. It can effectively reduce CNN data input and system resource consumption, and improve computational efficiency, while ensuring high accuracy. It satisfies the real-time requirements of traffic sign detection and recognition in video. The algorithm takes an average of 26.82 milliseconds per frame and the recognition accuracy reaches 96.94%.

Keywords Spatiotemporal model, Real-time, Multi-scale convolutional neural network, Traffic sign

1 引言

当今,无人驾驶技术^[1]和智能交通系统^[2]已经成为整个汽车产业的最新发展方向。交通标志的检测和识别(TSR)技术可以提供道路交通标志的当前状态,是驾驶员辅助系统、自动驾驶车辆等实际应用的重要组成部分。为了便于驾驶员阅读和识别,区分标志与背景,交通标志总是以特定的形状和颜色设计。然而,有效地识别交通标志可能面临许多困难,例如,不可预测的复杂场景,改变的照明条件,视角的变化,部分遮挡,相机噪声,低分辨率,运动模糊等。除了上述外部因素之外,交通标志本身还有许多类别,并且一些类别间具有高度相似性。

随着车载行车记录仪的普遍使用,视频图像成为了城市道路交通信息获取的一种快速、简洁方式。然而,城市道路环境下背景复杂、干扰因素多,采用通常的图像处理方法进行交通标志检测与识别的难度较大。交通标志最显著的特征是颜色和几何形态,此外还有空间位置和结构特征。一些研究者通过结合这些特征来实现交通标志的检测与识别。

交通标志检测识别系统的结构分为两大类,一类是直接预测目标类别和包围盒(bounding box)的多目标检测方法^[3-4],另一类方法则分为检测和识别两个阶段。近几年,检测阶段主要出现了基于颜色和形态特征^[5-7]和基于神经网络^[8-10]的检测方法。基于颜色和形态特征检测方法常用的颜色空间有 HSV、标准化 RGB、YCbCr 等,算法简单,但易受环境如光照条件、遮挡和形变影响。基于神经网络的检测方法鲁棒性好,能适应场景变化,但需要对检测阶段的神经网络进行训练,计算成本高。识别阶段主要包含基于卷积神经网络^[11-12]和基于 HOG 特征+SVM 分类器^[13-15]的识别方法。基于卷积神经网络的识别方法已被证明识别准确率高于人类,但此类方法需要消耗大量的计算资源。基于 HOG 特征+SVM 分类器的识别方法的计算效率较高,但其存在正负样本数量不平衡的问题,可能实现局部最优或过度拟合的解决方案,很难实现多类分类。

目前,交通标志识别主要分为 3 个研究方向:实现多类分类,基于文本的交通标志识别和提高交通标志检测识别的计算效率。实现多类分类是交通标志识别的一个重要方向,以

德国 TSR 基准(GTSRB)数据集^[16]为例,其交通标志包含 43 个类别。Lu 等^[17]提出了基于多模态树形结构嵌入式多任务学习的交通标志识别方法。该方法依据不同子集的交通标志在颜色和形状上的差异,按照树形结构为交通标志进行不同深度的类别划分,在识别阶段利用树形结构提高多级交通标志识别的性能。尽管交通标志识别已被研究多年,但大多数现有工作都集中在基于符号的交通标志上,基于文本的交通标志识别是交通标志识别的另一项挑战。Luo 等^[18]提出了基于多任务卷积神经网络的交通标志识别方法。该方法可以识别基于符号和基于文本的交通标志,针对 ROI 的细化和分类任务提出了一个新的卷积神经网络架构。其架构中具有两个决策层,二元分类层用于区分背景和交通标志,交通标志分类层用于对 ROI 进行识别。由于车辆的高速运行,交通标志识别的实时应用不仅要考虑精度,还要考虑计算速度。因此,如何在保持高精度和鲁棒性的同时提高计算效率是交通标志识别的第三个研究方向。

本文提出了一种基于时空卷积神经网络(ST-CNN)的交通标志实时检测识别算法。算法包括 3 部分:

(1)基于颜色的视频图像快速分割压缩算法提取交通标志候选区域;

(2)根据视频图像序列的动态关联性,对同一交通标志的图像序列进行筛选、追踪和补充,保证交通标志图像序列的完整性;

(3)使用多尺度卷积神经网络对交通标志图像进行识别。实验结果显示,在不同光照条件和不同场景下,算法鲁棒性强,检测识别准确率高,平均每帧耗时 26.82 ms,满足了实时性要求。

2 时空卷积神经网络 ST-CNN

交通视频序列中,交通标志具有刚性目标的物理属性,其颜色、几何、空间位置等特征信息在一段时间内变化较小,具有一定的连续变化规律。刚体的出现和消失过程是一个定量变化过程,可以在连续多帧视频图像序列中检测和跟踪。而非刚性目标物体的出现和消失并无规律,如草地、树木和建筑物等,难以构建相应的规则进行连续跟踪识别。本文构建一个时空关系模型并将之与多尺度卷积神经网络相结合,用于交通标志的检测、筛选、追踪和识别处理。

ST-CNN 可对交通视频图像序列中的同一交通标志进行检测、筛选和追踪,将此后检测到的候选区域与时空关系模型保存的候选区域相关联。该模型的主要目的是依据时空序列关系对检测阶段的结果进行简单的学习,丢弃非刚性目标区域,追踪回溯连续出现的目标物体区域,补充目标物体区域的多帧图像序列,并对目标物体区域进行识别。模型在保证使用多尺度卷积神经网络获得高识别准确率的同时,满足了对交通标志检测识别的实时性要求。

本文的网络结构如图 1 所示。

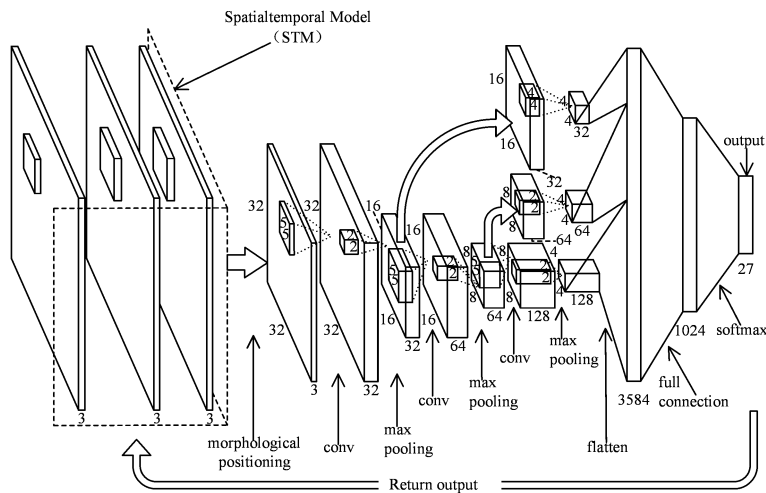


图 1 时空卷积神经网络模型

ST-CNN 依靠 STM 从视频图像序列中提取连续出现的候选区域,以减少多尺度卷积神经网络识别的候选区域数量,并根据形态学检测准确定位候选区域中的交通标志,有效地提高多尺卷积神经网络识别的精度和效率。

3 交通标志检测

颜色和形状是交通标志最显著的特征,也是交通标志区别于其他目标的最重要特征。我国交通标志包含禁令标志、警告标志、指示标志、指路标志、和旅游区标志共 6 个类别。警告标志、禁令标志和指示标志分为红、黄、蓝 3 种颜色,形状主要分为等边三角形、矩形和圆形。指路标志、道路施工安全标志和旅游区标志分为蓝色、绿色和棕色,形状为矩形。

3.1 图像快速分割压缩

为了减少检测计算量,提高处理效率,基于颜色的图像快

速分割压缩算法将符合阈值范围的像素抽取到二值图像中。算法以 $m \times m$ 分块尺寸将 $M \times N$ 大小的输入图像进行分割压缩。图像分块模型如图 2 所示,算法流程如下:

首先,将 RGB 图像转化成 HSV 颜色空间。

对于所有的像素点,其 H, S, V 3 个分量若满足某种颜色各分量的取值范围,则将该像素点在二值图像中对应的像素点置为白色。

如图 2 所示,输入图像被划分为 $M \times N$ 个子图像块 $A_{i,j}$, 每个子图像块为 $m \times m$, m 通常取 8 或者 16,输出图像为 $M \times N$ 。设输出图像的灰度图为 I_r , 每个像素点的初始化值为 0, 且像素点 $I_r(i, j)$ 与图 1 子图像划分模型中的像素块 $A_{i,j}$ 一一对应。

1)对每个子图像块 $A_{i,j}$:统计白色像素点个数 $N_{i,j}$, 设置筛选比例 R , 若图像块 $A_{i,j}$ 中彩色像素点所占比例 $N_{i,j}/(m \times$

$m) \geq R$, 则标记图像块 $A_{i,j}$ 为白色, 即像素点 $I_r(i, j) = 255$ 。

2) 输出分割图像 I_r 。

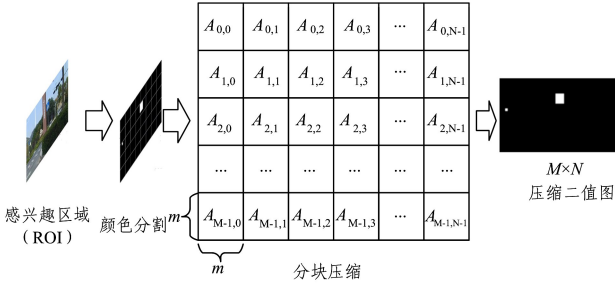


图 2 图像分块模型

3.2 候选区域生成筛选

光照、角度等条件的变化, 导致一个完整的交通标志可能被分割成多个离散的区域块。通过查找标记多个离散的区域块, 进行连通区域合并, 以保持交通标志分割信息的完整性。合并后进行筛选, 最终形成检测候选区域。

3.2.1 连通区域合并

根据块间的欧氏距离, 合并离散块为单一连通区域。设 blk_i 表示标签为 i 的连通区域, W_{blk_i} 为第 i 个连通区域的宽度, H_{blk_i} 为第 i 个连通区域的高度, $center(x_i, y_i)$ 表示第 i 个连通区域的中心点。两连通区域间的距离 $dis_{(i,j)}$ 的计算公式为:

$$dis_{(i,j)} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (1)$$

当两连通区域间的距离满足式(1)时, 将其合并。 T 为连通区域间的距离阈值。

$$dis_{(i,j)} \leq \sqrt{\left(\frac{W_{blk_i}}{2}\right)^2 + \left(\frac{H_{blk_i}}{2}\right)^2} + \sqrt{\left(\frac{W_{blk_j}}{2}\right)^2 + \left(\frac{H_{blk_j}}{2}\right)^2} + T \quad (2)$$

3.2.2 候选区域生成筛选

经过连通区域合并算法处理后, 得到了包含交通标志及背景噪声的候选区域; 计算候选区域的面积和长宽比, 对其进行筛选。计算公式如下:

$$\begin{cases} rate = \frac{W_{blk_i}}{H_{blk_i}} \\ area = W_{blk_i} \times H_{blk_i} \end{cases} \quad (3)$$

检测候选区域的筛选公式:

$$blk_i = \begin{cases} \text{true}, & S_{\min} \leq area \leq S_{\max} \& \& rate < rate_value \\ \text{false}, & \text{else} \end{cases} \quad (4)$$

S_{\min} 为交通标志的最小检测面积, S_{\max} 为最大检测面积。当 blk_i 符合筛选条件时, 将其保留为检测候选区域, 反之将其删除。

3.3 时空关系模型 STM

根据检测目标在多帧交通视频图像序列中的时间和空间位置连续变化关系构建时空关系模型, 并假设摄像头的光轴大致与地面平行, 且摄像头是沿着光轴方向运动的。设 f 是相机焦距, s' 是像平面上交通标志的大小, d 是像平面和交通标志的距离, s 是交通标志的实际尺寸, $camera$ 是摄像头的位置, 空间约束关系的示意图如图 3 所示。

根据三角形相似关系可得公式:

$$\frac{f}{f+d} = \frac{s'}{s} \quad (5)$$

由于 f 远小于 d , 则式(5)可约等于式(6):

$$\frac{f}{d} = \frac{s'}{s} \quad (6)$$

则交通标志在成像平面上的尺寸 s' 为:

$$s' = \frac{f \times s}{d} \quad (7)$$

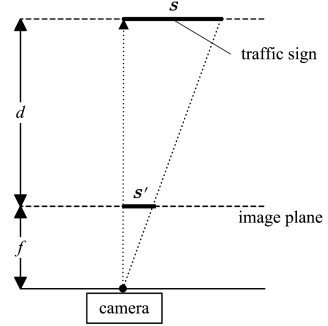


图 3 空间约束关系

安装车载视频的的车辆以一定的速度向距离 d 的交通标志移动, 过程中交通标志的位置和大小会随时间和空间而改变。当车辆向交通标志移动时, d 逐渐减小, s' 逐渐增大, 即随着摄像机与交通标志的空间距离的缩短, 交通标志在图像中的尺寸越来越大。

如图 4 所示, 在 $\Delta T(t_1, t_2, \dots, t_i, \dots, t_j, \dots, t_n)$ 时间段内, $L(l_1, l_2, \dots, l_i, \dots, l_j, \dots, l_n)$ 为 n 帧连续图像对应 l_1, l_2, \dots, l_n 的交通标志检测候选区域集合, $S_{\min} \leq l_1, l_2, \dots, l_n \leq S_{\max}$, $F(f_1, f_2, \dots, f_i, \dots, f_j, \dots, f_n)$, 为检测候选区域存在性标签集合。存在性标签 f 为:

$$f = \begin{cases} 1, & \text{if exists} \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

若在 t_i, \dots, t_j 时段内, 存在性标签序列 f_i, \dots, f_j 满足:

$$C = \frac{\sum_{k=i}^j f_k}{M} > \delta, 0 < M \leq 0 \quad (9)$$

则将其判定为交通标志待识别区域, 反之为背景噪声。其中, $M = |j - i + 1|, 0 < \delta \leq 1$ 。

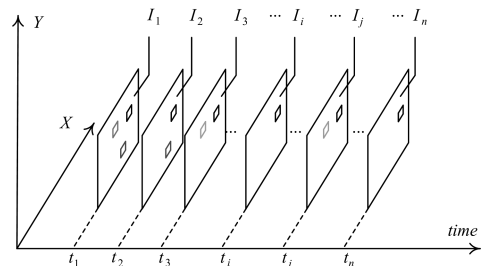


图 4 时空关系模型

假设某候选区域 I_i^j 内第 t_i 帧的存在性标签 $C_i^j > \delta$, 但 t_i 帧并没有检测到该候选区域, 则模型会按照第 t_{i-1} 帧记录的位置对该候选区域框选, 补全其图像序列。回溯补全保证了交通标志图像序列的完整性, 从而确保了识别验证环节的可行性。

检测起始/终止条件分别为:

起始条件: $I_i \geq S_{\min}$; 终止条件: I_i 超出边界

3.4 基于形态学的候选区域定位

交通标志形状多样,包含圆形、矩形、等边三角形和八边形,国外的交通标志数据集还包含菱形。3.2.2节已经根据面积和长宽比对候选区域进行了筛选。本节需要根据形态学来对候选区域进行进一步检测和定位,从而提高识别阶段的准确率和效率。

首先,使用自适应阈值二值化方法对候选区域图像进行二值化操作,得到候选区域中交通标志的轮廓特征。

$$dst(x,y) = f(x) = \begin{cases} max\ value, & \text{if } src(x,y) > T(x,y) \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

为了建立统一的形态学描述,对轮廓特征的面积 S 和周长 L 满足阈值的候选区域进行定位。

在尺度相同的情况下,三角形的面积和周长最小。根据候选区域长宽的最小值计算轮廓的面积阈值 S_{Δ} 和周长阈值 L_{Δ} 。设 a 表示比例系数, w 为候选区域宽度, h 为候选区域高度。

$$S_{\Delta} = \frac{\sqrt{3}}{4} a^2 \min^2(w,h), L_{\Delta} = 3a \min(w,h) \quad (11)$$

图5展示了基于形态学的交通标志定位结果。



图5 基于形态学的候选区域定位结果

4 交通标志识别

本文将交通标志划分为基于符号的交通标志和基于文本的交通标志,其中禁令标志、警告标志和指示标志属于基于符号的交通标志;指路标志、道路施工安全标志和旅游区标志属于基于文本的交通标志。

目前,大多数的交通标志数据集中只包含基于符号的交通标志,其研究侧重于实现基于符号的交通标志的多类分类,缺少基于文本的交通标志类别。而对基于文本的交通标志的研究仅是对其整体进行识别,并没有提取基于文本的交通标志中的符号和文本信息。本文主要针对禁令、警告、指示和指路标志共计4类出现较为频繁的交通标志类别进行检测识别。

4.1 基于符号的交通标志识别

文献[19]使用多尺度卷积神经网络对GTSRB数据集的彩色图像进行训练,达到了98.97%的准确率,使用灰度图像进行训练时达到了99.17%的准确率。

网络有3个卷积层用于特征提取,2个完全连接层用作分类器。网络使用多尺度特征,卷积层的输出不仅被转发到后续层,而且会反馈到完全连接层。每个卷积层的输出都经过额外的最大池化,因此所有卷积在进入分类器之前都按比例进行了二次采样,识别结果如图6所示。针对3种基于符号的交通标志检测识别结果,仅展示了原图中ROI部分。

为了避免网络对训练数据的过度拟合,使用以下3种正

则化技术:Dropout,L2正则化和Early stopping。



(a)



(b)



(c)

图6 基于符号的交通标志检测识别结果(截取ROI部分)

4.2 基于文本的交通标志识别

基于文本的指路标志包含道路、地点、方向和距离4类信息。其中,道路信息指示道路名称和道路编号,地点信息指示目的地和沿途地名,距离信息指示距离,3类均为文本信息。方向信息指示方向,分为标志信息和文本信息,标志信息为箭头形式,文本信息为地理方位,如“南”“北”或“S”“N”等。指路标志包含的语义信息由文本信息、标志信息以及文本与标志的位置关系决定。

4.2.1 符号及文本信息提取

将STM中连续出现的尺度大于 S_{max} 的候选区域认定为指路标志图像,依据HSV颜色空间的阈值对指路标志图像进行二值化,得到边缘图。如果边缘图中的某一连通分量占比超过阈值 T ,则认为是标志信息,并且保留其最大外接矩形的位置。

随后生成标记图。每个边缘图由 $n \times n$ 个像素的窗口按照 n 个像素的步长扫描。这里 n 被设置为2。标记图中的每个块对应于二值图中的窗口。然后,每个窗口分为4个部分:左上角,右上角,左下角和右下角。计算这4个部分中的边缘像素的数量,并将其定义为 n_1, n_2, n_3 和 n_4 。然后,通过 n_1, n_2, n_3 和 n_4 的乘积生成 n :

$$n = \prod_{i=1}^4 n_i \quad (12)$$

标记图中的每个块 p 定义为:

$$p = \begin{cases} 1, & n > 0 \\ 0, & n = 0 \end{cases} \quad (13)$$

标记图表示像素在二值图中的分布情况。由于文本和标志的边缘像素分布密集且具有各向同性,因此倾向于保留标记图中的相应块。生成标记图后,分析连通分量以获得每个分量的最大包围矩形,称为候选块。

这些候选块包含文本信息和标志信息,依据每个候选块占整个候选区域的面积比例,将候选块划分为文本块和标志块,并使用多尺度卷积神经网络对标志块进行识别,识别结果如图7所示。



图 7 指路标志检测识别结果

4.2.2 符号及文本信息分类

本文将指路标志中经常出现的符号划分为 6 类。

图 8 中,“left”“right”和“straight”3 个类别在基于符号的交通标志类别中存在,只需要将样本加入到数据集已知类别中。“down”“T-jun”和“cross”3 个类别在基于符号的交通标志类别中不存在,需要在数据集中增添相应类别。

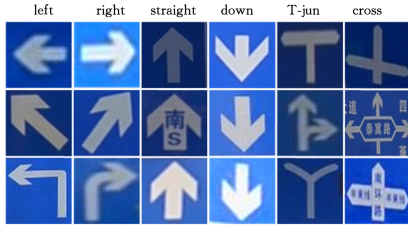


图 8 指路标志中的符号类别

5 交通标志样本

5.1 交通标志样本库的构建

本文标注了 13 346 帧交通视频数据,最终得到含有 20182 个交通标志的训练集,含有 1927 个交通标志牌的测试集。数据集共有 27 个类别,其中基于符号的交通标志 24 类,基于文本的交通标志 3 类,所有图像都被下采样或上采样到 32×32 个像素,图 9 展示了交通标志样本库中的部分样本。



图 9 交通标志样本集

5.2 样本数量的扩充

在实际道路交通视频中,摄像机的参数差异、拍摄光照强度的差异和车辆行驶过程中产生的抖动问题,导致视频中包含的交通标志形态和清晰度可能有很大的差异。此外,采集到的各类交通标志数不均匀,某些交通标志数目过少,容易导致最终的实验结果偏向某些类别。因此,需要对数据集进行扩充,本文采用了以下两种交通标志数据集扩充方式。

1)通过仿射变换以及改变图像的亮度、色度和对比度等方法增强图像。仿射变换可以有效解决车辆行驶过程中由于拍摄角度产生的形变交通标志的识别问题。

2)对于交通标志数据集中数目过少的类别,从清华大学 Tsinghua-Tencent 100 K 数据集的原始图像中随机截取样本。

如图 10 所示,通过以上两种方式扩充交通标志训练集以

后,标注的交通标志数目达到 56714 个,有效地增加了交通标志数目过少类别的数量,使交通标志训练集中各个类别的数目得到了一定程度的平衡。



图 10 交通标志样本拓展

6 实验结果与分析

为测试本文提出算法的性能,使用行车记录仪采集不同时间段、不同光照条件下的城市复杂道路环境视频图像。视频分辨率为 1920×1080,帧率为 30fps。测试环境:CPU Intel (R) Core (TM) i5-4430 @ 3.00 GHz; RAM 64.00 GB; GeForce GTX 1080。

表 1 列出使用 ST-CNN 前后,交通标志检测识别算法的检测准确率、漏检率和平均耗时等测试结果。

表 1 应用 ST-CNN 的实验结果对比

指标	使用	未使用
视频帧数/帧		21 228
交通标志/个		4 358
候选区域/个	47 164	311 025
平均每帧候选区域/个	2.22	14.65
正确检测/个	4 210	4 245
错误检测/个	42 954	306 851
漏检/个	148	113
检测准确率/%	8.93	1.36
漏检率/%	3.40	2.59
平均每帧耗时/ms	26.82	59.27

定义漏检率 LDR,即交通标志检测算法处理后,候选区域中漏检交通标志数占实际交通标志总数的比例,计算公式为:

$$LDR = \frac{LD}{LD + TD} \tag{14}$$

定义准确率 TPR,即交通标志检测算法处理后,候选区域中实际交通标志数占候选区域总数的比例。计算公式为:

$$TPR = \frac{TD}{TD + FD} \tag{15}$$

由表 1 可知,测试视频共有 21 228 帧,其中交通标志总共出现了 4 358 次(最小检测尺度 S_{min} 为 20×20)。未应用 ST-CNN 时,检测阶段共得到 311 025 个候选区域,平均每帧 14.65 个,正确检测交通标志 4 245 个,检测准确率为 1.36%,漏检交通标志 113 个,漏检率为 2.59%,算法平均每帧耗时 59.27 ms;使用 ST-CNN 后,检测阶段共得到 47 164 个候选区域,平均每帧 2.22 个,正确检测交通标志 4 210 个,检测准确率为 8.93%,漏检交通标志 148 个,漏检率为 3.40%,算法平均每帧耗时 26.82 ms,满足了实时性要求。实验结果表明,应用 ST-CNN 后,检测阶段的准确率提高了 6.34%,漏检率提高了 0.81%,平均每帧检测的候选区域减少了 12.43 个,显著减少了输入 CNN 网络的候选区域数量,使得平均每帧耗时降低了 32.45 ms。而漏检率的提高是由于交通标志刚出现时未满足时空关系模型的显示条件。

针对含有不同交通标志个数的视频片段进行测试,以验证在不同复杂程度下应用 ST-CNN 对计算效率的提升。从表 2 可知,在同时出现 4 个交通标志的道路场景中,算法平均每帧耗时 34.05 ms,基本满足 30 fps 的实时性要求。

表2 不同场景下应用 ST-CNN 的耗时对比

场景	标志个数	帧数/帧	未使用 ST-CNN/ms	使用 ST-CNN /ms
场景 1	0	917	43.06	23.77
场景 2	1	293	59.28	25.69
场景 3	2	298	57.01	26.82
场景 4	3	298	64.27	28.92
场景 5	4	447	77.95	34.05

表3为 ST-CNN 检测和识别两个阶段的耗时统计。检测阶段的最大耗时为 28.72ms,最小耗时为 14.18ms,平均耗时为 17.31ms;识别阶段的最大耗时为 13.28ms,最小耗时为 6.43ms,平均耗时为 8.86ms;总计平均耗时为 26.17ms。实验证明,ST-CNN 交通标志检测识别算法的耗时较为稳定。

表3 ST-CNN 各阶段的耗时统计

(单位:ms)			
	最大耗时	最小耗时	平均耗时
检测阶段	28.72	14.18	17.31
识别阶段	13.28	6.43	8.86
总计	—	—	26.17

表4列出不同时间段、不同路段条件下对于交通视频图像序列进行测试的结果。对于总计包含 18230 个交通标志的 46683 帧交通视频图像,ST-CNN 的识别准确率达到 96.94%。在不同光照条件和不同场景下,ST-CNN 的识别准确率较为稳定。

表4 不同光照条件下 ST-CNN 的识别准确率

拍摄时间	帧数/帧	交通标志/个	准确率/%
2016-10-19 07:33	5307	1383	96.39
2016-10-06 08:57	5307	1665	97.54
2016-12-09 11:33	9024	2456	97.04
2017-02-04 13:25	8997	4396	97.49
2016-12-18 15:44	9025	6167	96.85
2017-02-03 16:25	9023	2163	96.83
总计	46683	18230	96.94

表5为 ST-CNN 算法与 YOLOv2 算法和 Faster R-CNN 算法的结果对比。实验使用本文构建的交通标志数据集对 YOLOv2 模型和 Faster R-CNN 模型进行训练,并在相同的测试环境下对同一段交通视频图像序列进行测试。实验统计了 3 种算法的平均耗时,统一使用检测评价函数 IoU 计算平均准确率 *mAP*,并在各算法运行时记录了 GPU 的内存占用率。在交通标志检测识别问题上,与 YOLOv2 算法和 Faster R-CNN 算法相比,本文提出的 ST-CNN 算法能够在保证识别准确率的前提下,满足交通标志检测识别的实时性要求,并且算法的 GPU 资源占用率低,便于向移动端移植。

表5 ST-CNN 与其他算法的效果对比

模型	耗时/ms	<i>mAP</i> /%	GPU 占用率/%
ST-CNN	25.37	62.4	5.7
YOLOv2	23.29	42.3	77.3
Faster R-CNN	88.53	65.2	86.2

结束语 针对复杂城市道路环境下交通标志的检测与识别,提出了一种基于 ST-CNN 的交通标志实时检测识别算法。该算法根据图像序列的帧间时空连续关系构建时空关系模型,并与多尺度卷积神经网络相结合。算法对视频图像序列中的同一交通标志进行检测、筛选、追踪、识别和验证处理,分割提取指路标志中的文本和标志信息。实验结果表明,算法检测识别速度快,准确率高,GPU 资源占用率低,可以向同时检测识别交通信号灯、车辆等其他交通信息的方向进行拓展,从而实现在辅助驾驶系统及移动计算设备中实时检测识别交通信息。

参考文献

- [1] LUETTEL T, HIMMELSBACH M, WUENSCHKE H J. Autonomous Ground Vehicles—Concepts and a Path to the Future [J]. Proceedings of the IEEE, 2012, 100(5): 1831-1839.
- [2] 赵娜, 袁家斌, 徐哈. 智能交通系统综述[J]. 计算机科学, 2014, 41(11): 7-11.
- [3] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector [M] // Computer Vision — ECCV 2016. Chain; Springer, 2016.
- [4] REDMON J, DIVVALA S, GIRSHICK R, et al. You Only Look Once: Unified, Real-Time Object Detection [J]. arXiv: 1506.02640, 2015.
- [5] GAO X W, PODLADCHIKOVA L N, SHAPOSHNIKOV D G, et al. Recognition of traffic signs based on their colour and shape features extracted using human vision models [J]. Journal of Visual Communication & Image Representation, 2006, 17(4): 675-685.
- [6] KUO W J, LIN C C. Two-Stage Road Sign Detection and Recognition [C] // 2007 IEEE International Conference on Multimedia and Expo. IEEE, 2007.
- [7] JIANG Y, ZHOU S, JIANG Y, et al. Traffic sign recognition using Ridge Regression and OTSU method [C] // Intelligent Vehicles Symposium. IEEE, 2011.
- [8] UIJLINGS J R R, VAN DE SANDE K E A. Selective Search for Object Recognition [J]. International Journal of Computer Vision, 2013, 104(2): 154-171.
- [9] 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 & Machine Intelligence, 2015, 39(16): 1137-1149.
- [10] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, PP(99): 1-1.
- [11] STALLKAMP J, SCHLIPSING M, SALMEN J, et al. Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition [J]. Neural networks: the official Journal of the International Neural Network Society, 2012, 32(2): 323-332.
- [12] WU Y, LIU Y, LI J, et al. Traffic sign detection based on convolutional neural networks [C] // International Joint Conference on Neural Networks. IEEE, 2014.
- [13] GREENHALGH J, MIRMEHDI M. Real-Time Detection and Recognition of Road Traffic Signs [J]. IEEE Transactions on Intelligent Transportation Systems, 2012, 13(4): 1498-1506.
- [14] IGEL C. Detection of traffic signs in real-world images; the German traffic sign detection benchmark [C] // International Joint Conference on Neural Networks. IEEE, 2013.
- [15] WANG G, REN G, WU Z, et al. A robust, coarse-to-fine traffic sign detection method [C] // International Joint Conference on Neural Networks. IEEE, 2013.
- [16] STALLKAMP J, SCHLIPSING M, et al. The German Traffic Sign Recognition Benchmark: A multi-class classification competition [C] // International Joint Conference on Neural Networks. IEEE, 2011.
- [17] LU X, WANG Y, ZHOU X, et al. Traffic Sign Recognition via Multi-Modal Tree-Structure Embedded Multi-Task Learning [J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(4): 960-972.
- [18] LUO H, YANG Y, TONG B, et al. Traffic Sign Recognition Using a Multi-Task Convolutional Neural Network [J]. IEEE Transactions on Intelligent Transportation Systems, 2018, PP(99): 1-12.
- [19] SERMANET P, LECUN Y. Traffic Sign Recognition with Multi-Scale Convolutional Networks [C] // The 2011 International Joint Conference on Neural Networks (IJCNN). 2011.