交通场景下的图像识别

09020104 林志涛，09021102 郭天琦，JS121232 蒋志强，JS121233 谭彦廷
（排名不分先后）

摘要—随着社会经济的进步，道路拥堵、交通安全等问题愈发凸显. 为此，交通场景下的图像识别应用愈发广泛，而这又与无人驾驶车辆的设计紧密联系. 本文即以无人驾驶车辆为主线去理解交通场景下的图像识别. 首先，本文阐述了图像拍摄设备的选取方案；然后，介绍了图像预处理的方法，该方法可以去除图中的光照、阴影等干扰信息；接着，以两篇论文为例，分析了车辆、交通标志等物体的识别策略和特征提取策略；最后，分析了该领域的现实意义，并对未来的研究方向做出展望.

I. 引言

视觉在人与人、人与物的交互中都起到了重要作用，让计算机终端最终获得与人相似的“视觉”能力和相应的交互能力是人工智能领域的核心挑战之一. 通过阅读一篇对计算机视觉目标检测方向的综述性论文 [1]，我们锁定了其中一个重要应用——交通场景分析，并决定对这个方向进行整体性理解.

然而对任何一个领域进行整体性理解必须有一条足够清晰的主线. 我们了解到，交通场景下的图像识别直接作用于无人驾驶车辆的设计 [2]. 据此，我们搜索找到了其余相关论文，并在三遍阅读法的概览阶段列出了对应的论文关系图（图 1）. 遗憾的是，为紧紧围绕无人驾驶车辆这一主线，我们寻找论文时一定程度上牺牲了论文的权威性，查找到的部分论文较为冷门.

下文中，我们将从无人驾驶车辆的视角，逐步解释图像获取、图像预处理、特征提取这整个过程，分析如何让它“看到”世界.

II. 图像获取

车辆“看到”世界需要感觉器官，[2] 介绍了几种不同的感知模式. 目前该领域所拥有的感知模式非常丰富，包括激光雷达检测、单相机成像、双相机立体成像、GPS 等.

单相机是最常用的感知模式. 这种基于视觉的感知模式之所以在车道和道路检测任务中占据主导地位，主要有两个原因:

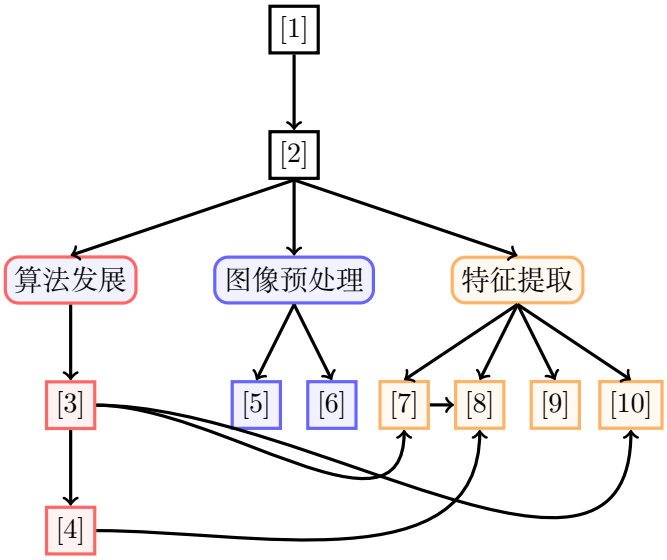


图 1. 论文关系图

第一，无人驾驶车辆的感官功能归根结底是用于驾驶车辆. 而交通设施都适合于人类. 在驾驶时，人类司机所获取最多的是视觉数据，车道标志和道路边界等的设计都必须保障人类驾驶员在所有驾驶条件下都能看到它们. 因此使用摄像机来让无人驾驶车辆获得与人相同的视觉信息就很有意义.

第二，很明显，相机目前是汽车应用中最便宜、最稳定的模式. 消费级相机大规模生产链的成熟，和其在机器视觉方面投入的大量适配，使得良好的成本效益解决方案成为可能.

之后的论文也紧紧围绕单相机感知模式进行。

III. 图像预处理

A. 预处理的目的

单相机感知模式虽然划算，但相比其他方式有明显的缺点. 如激光雷达检测能直接检测出 3D 结构，但是

单相机所给出的只是几组连续的图片，这些图片中还有许多干扰因素影响识别.

图像预处理的目标就是去除杂乱而具有误导性的成像伪影和不相关的图像部分，剩下的处理过的图像部分将作为输入数据，稍后将从中提取特征.

B. 光照一致化

无论天气情况和时间如何，即无论是阳光明媚的正午还是夜晚的人工照明，一个稳健（robust）的道路检测系统应该能够应对不同的照明条件. 这便带来了处理图像中不同光照条件的第一种思路：光照一致化，这里我们选取了 [5] 这篇经典论文来进行讲解.

1) 论文发布背景: 已有的基于尺度不变特征变换（Scale-invariant feature transform，SIFT）算法的视觉导航系统作用广泛. 它的本质是在不同的尺度空间上查找关键点（特征点），并计算出关键点的方向. 通过它能够查找到一些十分突出的关键点（如边缘点）. 这些点不会因光照、仿射变换和噪音等因素而变化. 然而这种特征不足以在户外环境中提供真正的光照不变性，并且图像传感器中的大部分信息得不到利用.

2) 研究过程: 在这篇论文中，作者试图利用自动驾驶车辆平台上相机的光谱属性的全部信息，并提出从中提取出一个光照不变的颜色空间，以减少原始 RGB 图像中存在的阳光和阴影的影响.

通过使用光照不变的颜色空间来增加不同时间图像的相似度，大大减少了由于阳光和阴影而产生的变化和影响。光照不变处理的结果是一个灰度图像，其中灰度值主要取决于场景中物体的材质属性.

这个转换过程基于相机光谱响应的原理. 通过图像传感器和照明光源的响应关系以及普朗克源的维恩近似等物理规律可推导出，转换所需的参数 α 满足以下约束:

$$\frac{hc}{k_B T\lambda_2} - \frac{\alpha hc}{k_B T\lambda_1} - \frac{(1-\alpha)hc}{k_B T\lambda_3} = 0$$
$$\text{即，} \frac{1}{\lambda_2} = \frac{\alpha}{\lambda_1} + \frac{(1-\alpha)}{\lambda_3} \tag{1}$$

其中 h 是普朗克常数，c 是光速，k_B 为玻尔兹曼常数，T 为黑体源的相关色温.

由此，我们只需要知道每个传感器通道的峰值光谱响应，就可以为给定的三通道相机确定唯一的 α 参数，这样大大降低了传感器套件的成本.

表 I
三个点灰工业相机的数据表估计的峰值光谱响应，以及相应的 α 参数

相机	Grasshopper2	Bumblebee2	Flea2
图像传感器	ICX285AQ	ICX204AK	ICX267AK
λ_1	470nm	460nm	470nm
λ_1	540nm	540nm	535nm
λ_3	620nm	610nm	610nm
α	0.4642	0.3975	0.4706

使用 α 参数将三通道浮点 RGB 图像转换为相应的光照不变图像，效果如图 2 所示.

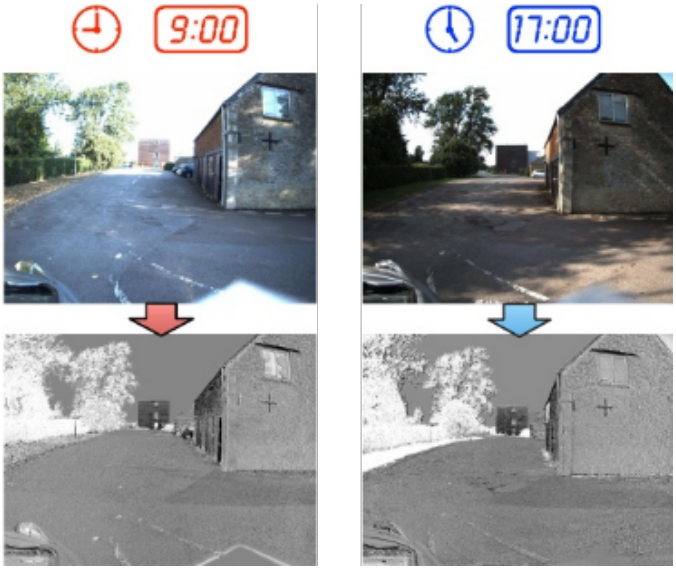


图 2. 将 RGB 图像转换为光照不变的颜色图像

3) 评价: 为了评估光照不变颜色空间的性能，作者选择多个位置，于不同时间获取 RGB 图像和对应的光照不变图像，收集了一个 24 小时的视觉数据集，并分别计算其平均零均值归一化互相关系数（zero-mean normalised cross correlation, ZNCC）. 结果表明在白天时光照不变的颜色图像明显更加一致；然而当违反了黑体照明的必要假设，即夜间时，原始的 RGB 图像更加一致. 因此，通过组合，在白天使用光照不变图像并在夜间使用原始的 RGB 图像，可以获得最佳的光照条件一致结果.

该论文以精简的方式，通过日照、照相机成像等原理构建起物理模型，就解决了照明条件这一必须面对的问题. 它的主要目标是生成一个图像，无论场景光照强度、方向或光谱如何，其中像素值单纯对应于在场景中看到的物体的材质属性. 这使得它很自然地成为交通场

景下图像识别的一个预处理步骤.

C. 阴影分割

上一篇论文的思路是让图像的光照变得均匀，以此抹除不同光照条件对图像识别的影响. 这篇 [6] 给出另一种处理光照条件的思路：阴影分割——不需要均匀化光照，通过光照把阴影分割出来. 由于在交通场景中，目标多处于运动状态，产生一组图像序列，静态图像分割阴影的方法不再适用. 这篇论文于是利用了空间（阴影与本体的空间相邻性）、时间（前几帧的信息）、色彩（阴影与光照下存在某种比例转换）来搭建图片像素的概率模型. 通过该概率模型可以判断出像素是否属于背景，接着通过色彩比例判断是否属于阴影. 这种方法属于传统机器学习领域，具有很强的可解释性.

1) 论文发布背景: 要想给帧中的像素进行阴影、背景、前景这样的分类，需要估计像素关于这三个类别的概率密度函数（probability density function, PDF）.

该过程常采用的方法是背景减法：从数据中构建出背景模型，如果对象与背景出现显著差异，则进行分割. 不幸的是，用这种方法处理运动对象时，阴影通常会随着对象一起被提取出. 这会导致物体定位的巨大误差，并可能对使用分割结果作为基准的算法造成严重影响.

2) 研究过程: 作者认为可以识别出三种有助于检测物体和阴影的信息源.

第一种信息来源于像素显示的变化：像素点被阴影覆盖时看起来会更暗；第二种信息来源于空间：物体和它的阴影位于图像中邻近的区域；第三种信息来源于时间：物体和阴影的位置可以从之前的帧中预测.

利用这三种信息源，作者提出了一种算法来将像素分类为阴影、前景和背景.

作者发现用对角矩阵来近似这种效果是令人满意的. 令 $v = [R\ G\ B]^{\mathrm{T}}$ 是曲面上某一像素点在光照下的相机响应, 那么 Dv 是该点在阴影下的相机响应，其中 $D = diag(d_R, d_G, d_B)$ 是对角矩阵. 在交通场景中，这样的抽象会使阴影显示更蓝. 而且对于不同的像素，只要在平坦的表面上，其 D 近似恒定；即使背景在整个图像上不是平坦的，我们也可以将图像划分为多个子区域，在这些子区域中分别进行建模，这样 D 依然能保持相对恒定. 有了这个阴影下像素显示的变化模型，我们很容易推导出估计阴影下三种颜色分量的均值和方差的规则：

$$\mu_{SH}^{i} = \mu_{IL}^{i} d_i \tag{2}$$

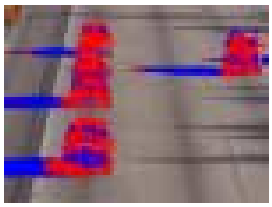
$$\sigma_{SH}^{i} = \sigma_{IL}^{i} d_i, i \in \{R, G, B\} \tag{3}$$

其中，μ 表示均值，σ 表示方差，SH 表示阴影下的颜色分量，IL 表示光照下颜色分量.

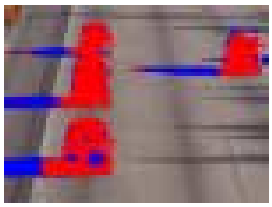
我们将每个像素的颜色分量与模型中该位置的颜色分量均值进行比较来开始分割. 如果没有显著差异(差异小于均值的 10%)，则该像素被分类为背景类. 否则，我们根据差异所计算出的概率进行分类.

通过这种基于像素显示的算法，大多数像素被正确分类（大约 73%，如图 3(a)）. 然而，由于部分像素分类错误，物体和阴影区域有不少噪点. 通过施加空间平滑度，可以显著提高结果.

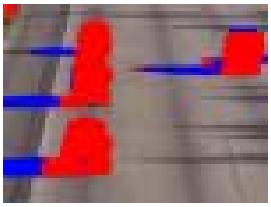
论文中研究了两种空间平滑方法.



(a) 处理前



(b) 用迭代的概率松弛法来传播邻域信息



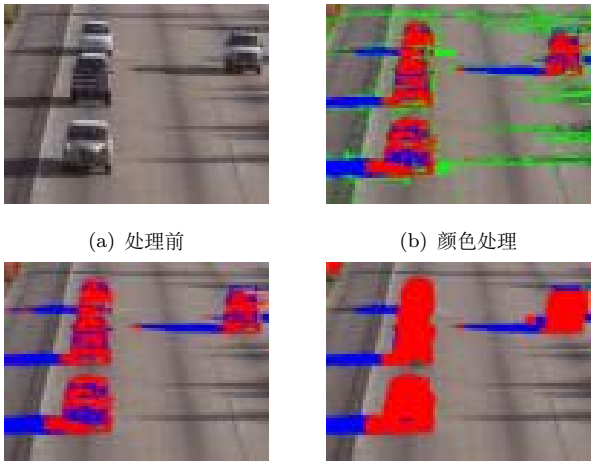
(c) 在 3(b) 的基础上进行空间滤波后处理

图 3. 两种空间平滑方法对比

第一种方法是用迭代的概率松弛法来传播邻域信息. 为了改变给定像素的最终分类，必须大幅改变像素对应的概率. 基于已有像素分类，可以得到某像素的相邻像素被分类最多的类别. 通过为该像素分配这一类别相关联的概率将会改善这一结果（78% 的像素被正确分类，如图 3(b)）.

第二种方法是对分割后的图像进行空间滤波后处理. 然而，这需要在第一种方法后进行. 如此处理后的

结果已经非常接近设想结果，(约 90% 的像素被正确分类，如图 3(c))．因此，通过一个简单的后处理进行空间平滑是最有效的．



(a) 处理前　(b) 颜色处理

(c) 像素分类　(d) 空间滤波处理

图 4. 阴影分割流程

3) 评价: 如图 4(c)．红色、蓝色的像素是那些与背景 PDF 的平均值相差超过 10% 的像素．这种方法能正确分类阴影和闪烁的背景像素，使得计算出的物体位置的准确性——特别是在有长阴影的场景中——大大提高．

本论文就这样提出了一种在图像序列中分割运动目标和其阴影的实时算法．这种算法行之有效、可长时间运行，为交通场景下图像识别提供了一种稳健的测量方案．

IV. 特征提取

经过预处理，可以排除图像中的多数干扰信息．接下来，应当对图像中的车辆、道路等物体进行检测，并提取这些物体的特征．目前的物体检测算法一般是从 R-CNN [10] 或 YOLO [4] 模型发展而来．R-CNN 是一种二阶段的检测方法，即先提出检测区域的建议，再进行物体识别和分类；YOLO 是一种一阶段的检测方法，将检测和分类整合在同一个神经网络中．

A. DAVE: 车辆检测与标注系统

在复杂的视频中进行车辆检测与标注是一项具有挑战性的任务．在 [7] 中，作者提出了一种车辆检测和标注系统 (Detection and Annotation for Vehicles, DAVE)．DAVE 由两个卷积神经网络组成，一个是快速车辆建议网络 (Fast Vehicle Proposal Network, FVPN)，用于寻找可能存在车辆的区域，提出检测建议；另一个是属性学习网络 (Attributes Learning Network, ALN)，用于验证每个建议区域是否为车辆，并标注车辆的颜色、类型和朝向．ALN 在训练过程中学习到的知识可以用于指导 FVPN 的训练．当整个系统训练完成后，就可以实现高效的车辆检测和标注．

1) 论文发布背景: 传统的车辆检测方法可以分为基于帧的方法和基于运动的方法．基于运动的方法包括帧间差分法、背景减法、光流法等等．然而，因为视觉信息的利用率不高，这种方法容易将其他运动的物体误检为车辆．为了提高检测性能，近年来的一些研究采用了基于部件的可变形模型 (Deformable Part Model, DPM)．这种算法将目标对象建模成几个部件的组合，并对这些部件进行检测，即使目标物体被部分遮挡，DPM 也可以有效地完成检测任务．然而，由于 DPM 在检测中使用了滑动窗口法，这是一种类似于遍历图像的方法，导致它的计算成本较高．

2) 研究过程: DAVE 的结构如图 5所示．它由两个卷积神经网络组成，分别是 FVPN 和 ALN．FVPN 的功能是预测图中的类车物体的位置，并给出这些位置的建议．之后，这些建议被传递给 ALN，ALN 负责验证该区域是否为车辆，并判断该车辆的朝向、颜色和类型．在训练阶段，FVPN 和 ALN 采用了联合优化．即，将深层网络 ALN 学到的知识进行提炼，再用来指导浅层网络 FVPN 的训练．实验表明，该方法能够在一定程度上提升 FVPN 的性能．

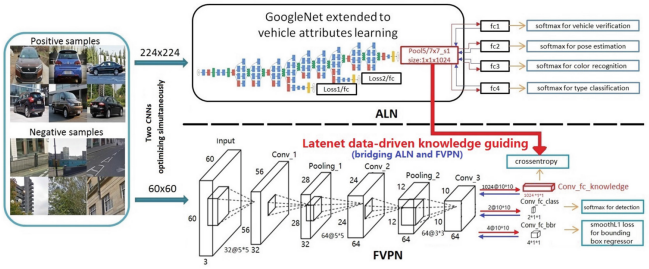


图 5. DAVE 训练架构图

在训练中，采用了 CompCars 数据集中的 10 万余条数据作为样本．该数据集标注了边界框和车辆的各种属性，包括车辆类型、朝向等．经过筛选，在数据集中选取了 6 种类型、5 种朝向的车辆样本，并增加了少量的不含车辆的空街景样本，来进行神经网络的训练．

训练完成后，DAVE 将按照两阶段方案进行推理. 推理过程如图 6 所示. 首先，将 10 层图像金字塔输入到 FVPN，将所有的热图整合到一张建议得分图上. 然后，根据阈值来过滤这一得分图，滤去得分较低的亮斑. 再用圆形扫描器来检测得分图上的局部峰值，这些局部峰值点可以作为建议区域的中心坐标. 这些中心所对应的边界框则作为建议区域的大致尺寸. 最后，将这些建议区域放大 1.5 倍，输入 ALN，来得到精细的边界框. 得到所有边界框后，执行非极大值抑制，消除重复检测的车辆对象.

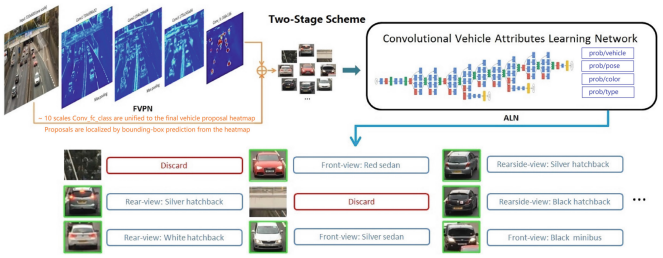


图 6. DAVE 推理流程图

3) 评价： 专用于车辆检测的 DAVE 系统性能十分优秀，对车辆属性的标注也较为准确. 然而，该模型还难以识别极为多样的颜色和车型. 没有经过训练的颜色和车型在检测中总是被误归类到相似的类型，或者直接被归为“其他”类. 为了解决这一问题，在未来的研究中，应当扩展训练数据，使用更丰富的车辆类别进行训练.

B. 交通标志检测

特征提取的另一方面是交通标志识别（Traffic Sign Recognition, TSR），这一功能可以用于提醒驾驶员注意交通标志，以便遵守交通规则. 论文 [9] 对 TSR 技术进行了综述，并对这一领域的未来研究方向提出了建议.

1) 论文发布背景： 交通标志通常被设计成醒目的样式，与周围环境有较大的区别，另外，标志的设计一般会遵循一定的标准，这使得检测任务十分明确. 但是，世界各地交通标志的设计标准各不相同，带来了一定的困难. TSR 还面临着如下挑战：

1) 不同种类的标志外形相似（如图 7）.
2) 标志可能褪色或变脏，使得颜色与设计标准产生偏差.
3) 标志没有正对道路，在图像中发生形变.
4) 照明条件可能使颜色产生偏差.
5) 低对比度的环境可能使形状检测困难.
6) 在杂乱的城市环境中，其他物体可能看起来与标志相似.
7) 不同天气条件的影响.

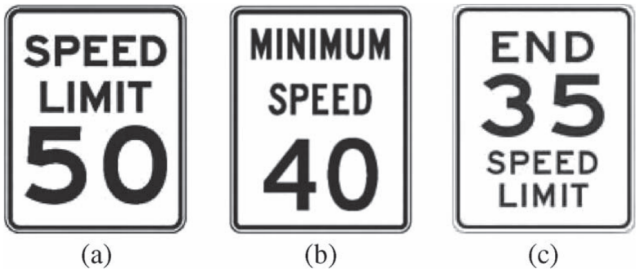


图 7. 相似的路标

一般而言，TSR 任务分为两个阶段：检测和分类（如图 8）. 检测任务的重点是在图像中确定标志的位置，而分类的重点是确定标志的类型. 这两者通常可以视为互相独立的任务，但在某些情况下，分类器会依赖检测器提供的信息做出推理，例如标志的大小、形状信息等. 在一个完整的系统中，检测与分类两个阶段是相互依赖的.

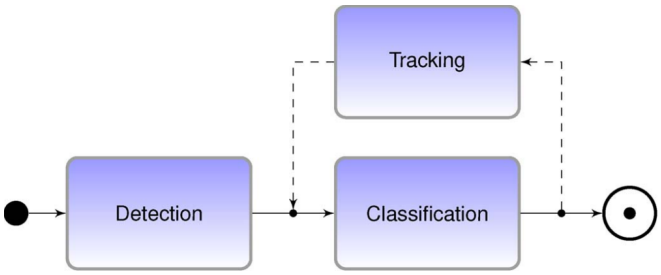


图 8. TSR 基本流程

2) 研究过程： 传统的交通标志检测方法有两种：基于颜色的方法、基于形状的方法.

基于颜色的方法利用了这样一个事实：交通标志在视觉上很容易与周围的环境区分开，它们通常采用对比度很高的颜色. 这些颜色可以作为检测的基础，因为在图像中搜索特定的颜色是简单的. 标志上的颜色块一般还有着明确的形状，这也可以作为检测的补充. 几乎所有基于颜色的方法在检查完颜色之后都会把形状考虑进去.

相比之下，基于形状的检测方法则较少地依赖颜色，该方法更倾向于直接检测标志的特征形状. 标志的形状不受光照和标志年龄的影响，但标志的部分区域可能会被遮挡，对边缘检测造成障碍，使检测变得困难.

在检测标志是，不同的研究选用了各种各样的检测特征. 最流行的检测特征是边缘，有些是直接从原始图片中获得的边缘，有些是从预分割图像中获得的边缘. 除此之外，也有研究使用方向梯度直方图（Histogram of Oriented Gradient, HOG），或是各种简单特征的组合.

3) 评价: 这样的做法虽然让计算机能够识别交通标志，但没能排除其他道路上标志的影响，这些标志与当前的行驶是无关的. 如图 9，这些无关标志应当被忽略. 而当车辆变道时，系统需要能检测到这种情况，并改变对有效路标的判断标准.

图 9. 虽然两个标志都可能被检测到，但只有右边的标志与司机相关, 左边的标志属于另一条路.

V. 领域意义

无人驾驶车辆终于“看到”了世界！但正如引言所述，无人驾驶车辆只是我们理解这个领域——交通场景下的图像识别技术——所抓取的一条主线. 它是辅助驾驶和自动驾驶技术的一项核心功能，具有重大的现实意义.

第一，该技术能提高驾驶的安全性. 交通事故每年在全球大约会造成相当于 6000 亿美金的经济损失. 我国作为作为全球人口密度最大和最为拥堵的国家之一，交通安全问题尤为突出. 自动驾驶技术可以作为人类驾驶的辅助系统，提供道路信息、预警功能，亦可在必要的情况下完全接管驾驶，从而有效地避免人为因素导致的交通事故，如醉酒、瞌睡、走神等. 让车辆学会理解道路的状况，能够让行驶更加安全可靠.

第二，提高交通系统的效率，减少拥堵. 大部分的交通拥堵是由于司机个人原因引发的交通事故. （如不专心驾驶、红绿灯前缓慢启动等.）引入自动驾驶有望可以帮助司机做出决策，减少交通拥堵，尤其是减少司机个人原因导致的拥堵. 此外，也有研究提出将自动驾驶与物联网结合，使道路上的车辆形成互联. AI 可以根据各个车辆的信息，形成最佳的整体通行方案.

第三，减轻人类驾驶的负担. 驾驶需要司机长时间集中精力，让人工智能帮助或代替人类司机，可以将人类的双手从方向盘上解放出来. 从经济效益角度考虑，物流行业如果采用自动驾驶方案，将节省大量的人力成本.

第四，为弱势群体创造驾驶条件. 例如，图像识别技术与面向盲人的操控设备结合，能够让盲人群体进行驾驶. 智能道路感知的实现，为困难人群提供了驾驶车辆的可能.

VI. 未来展望

目前，计算机的道路感知能力取得了长足的进步. 道路感知主要有两种发展趋势. 一是基于视觉的车道偏离警告（Lane Departure Warning）系统，该系统有望逐渐成为商业产品. 它使得车辆可以依靠视觉信息，进行一些高级的推理，在不同的环境下都具有一定的可靠性；二是以美国国防部高级研究计划局（Defense Advanced Research Projects Agency）为代表的全自动驾驶方案，该方案对视觉的依赖性较小. 方案放弃了车载全道路感知，而依赖 GPS 等探测设备，获得车辆定位信息和高分辨率地图.

然而在道路感知上，目前的研究仍有需要改进之处.

第一，需要扩大道路理解的范围. 下一代汽车的研究重点是需要理解多车道，学会识别远前方和车后方，以及识别道路的分、并. 为了完成这个任务，需要开发新的道路表示方法，该表示必须足够丰富，可以充分描述多车道的结构，并且能方便地从视频中提取和追踪.

第二，需要提高道路理解的可靠性. 当前识别系统的可靠性足以用于警报系统，但可能还不足以用于自动驾驶. 自动驾驶要求的错误率比目前还要低几个数量级. 为了进一步降低识别的错误率，提高可靠性，未来可能

可以做出如下改进：让车辆装载多个识别系统，串行或并行使用；采用视觉以外的辅助感知设备，如 GPS、卫星图和街景；更广泛地采用机器学习技术，以较少的工作量解决更复杂的问题.

参考文献

[1] Z. Zou, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *arXiv preprint arXiv:1905.05055*, 2019.

[2] A. Bar Hillel, R. Lerner, D. Levi, and G. Raz, "Recent progress in road and lane detection: a survey," *Machine vision and applications*, vol. 25, no. 3, pp. 727–745, 2014.

[3] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," pp. 580–587, 2014.

[4] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[5] W. Maddern, A. Stewart, C. McManus, B. Upcroft, W. Churchill, and P. Newman, "Illumination invariant imaging: Applications in robust vision-based localisation, mapping and classification for autonomous vehicles," in *Proceedings of the Visual Place Recognition in Changing Environments Workshop, IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China*, vol. 2, no. 3, 2014, p. 5.

[6] I. Mikic, P. C. Cosman, G. T. Kogut, and M. M. Trivedi, "Moving shadow and object detection in traffic scenes," in *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, vol. 1. IEEE, 2000, pp. 321–324.

[7] Y. Zhou, L. Liu, L. Shao, and M. Mellor, "Dave: A unified framework for fast vehicle detection and annotation," in *European conference on computer vision*. Springer, 2016, pp. 278–293.

[8] T. Huang, D. Koller, J. Malik, G. Ogasawara, B. S. Rao, S. Russell, and J. Weber, "Automatic symbolic traffic scene analysis using belief networks," in *AAAI*, vol. 94, 1994, pp. 966–972.

[9] A. Mogelmose, M. M. Trivedi, and T. B. Moeslund, "Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1484–1497, 2012.

[10] M. Braun, S. Krebs, F. Flohr, and D. M. Gavrila, "Eurocity persons: A novel benchmark for person detection in traffic scenes," vol. 41, no. 8. IEEE, 2019, pp. 1844–1861.