

语义分割与强化学习在自动驾驶中的应用

山东大学 数学与统计学院 数据科学与人工智能实验班

小组成员：刘畅 甘缘 毕嘉栋 唐铭喆

摘要：近年来，自动驾驶技术取得了长足的进步，这主要得益于深度学习等相关算法的快速发展与硬件计算条件的显著提升，随着人们生活方式的改变与道路交通设施的升级，自动驾驶技术成为各大厂商争先研发的前沿领域。本文的主要目的为总结语义分割技术和强化学习算法在自动驾驶领域的常见应用。我们首先介绍了图像语义分割的评价指标、常用数据集，并根据时间的发展，首先介绍了基于阈值分割等多种传统语义分割算法的实现思路、优缺点，再根据深度学习技术的发展，介绍了深度学习与传统方法相结合的图像语义分割算法，最后，在全卷积神经网络的催化下，发展出U-Net、SegNet、PSPNet、Re-fineNet、DeepLab、BiseNet和FPN等基于深度学习的语义分割方法，这些网络成为语义分割技术在自动驾驶领域应用的重要基础，包括对其他车辆、行人、车道、交通标志或交通信号灯等的识别与分类。由于自动驾驶对于实时性的要求，并且考虑到车载嵌入式设备有限的计算能力，我们集中讨论了改进的语义分割模型，包括Enet、ERFNet、ContextNet、Fast-SCNN、FasterSeg、ShelfNet网络，以上高效的架构，能够以适当的准确度实时执行。为了对所考察的架构进行最完整的评估，所有的模型都在一个一致的评估框架中根据其性能进行了比较。我们再视线转向强化学习领域，介绍了三种实现强化学习目标的方法，包括基于价值的方法、基于策略的方法、演员-评论家方法。接下来，分模块的介绍了自动驾驶系统的几种模式，并将强化学习模型应用到自动驾驶系统的几种模式中：控制模式、运动规划模式、行为规划模式、分层规划和控制模式，最后给出了每种模式的优缺点。本文有助于深入了解语义分割和强化学习在自动驾驶方面的基础应用、各个架构模型的优缺点，有助于从宽领域全方位了解自主驾驶的各个组成部分，可以帮助读者在不同的应用场景选择最适宜的模式方法。

关键词：自动驾驶，语义分割，强化学习

1 引言

随着智能交通的迅速发展，自动驾驶技术应运而生。全世界掀起了一股研究开发自动驾驶车辆（AVs）的热潮。经过多年的研究，自动驾驶汽车的技术水平已经成熟，正逐步进入公路试验阶段，并制定相应的法律法规。但自动驾驶汽车的驾驶并非绝对安全，或多或少存在一些潜在的安全隐患，导致各种交通事故的发生，逐渐引起各界的关注和探索。虽然世界各地都有很多交通事故是由人类引起的，但不可否认的是，人们对自动驾驶车辆行驶的安全要求远远高于人类自身。

同时，新兴科技企业对传统汽车行业的强烈冲击让整体的汽车产业格局和生态体系发生了极大的转变，传统汽车企业必须进行产业结构的优化调整和技术领域的不断升级，以应对智能化、现代化的行业发展趋势。因此，汽车的智能化自动驾驶是必然的发展方向，加强人工智能在汽车自动驾驶中的应用质效，是进一步实现汽车行业现代化发展的必然举措。这就需要不断提升自动驾驶技术的准确性与实时性，语义分割与强化学习作为人工智能领域两个炙手可热的方向，在深度学习技术的不断发展下，在自动驾驶技术的发展和进步中起到了举足轻重的作用。本文对相关的研究工作进行了梳理，介绍了语义分割和强化学习在自动驾驶技术领域的研究成果，为自动驾驶技术的进一步研究提供参考。

2 语义分割概述

图像分割处理的结果是将输入图像分割为若干个含有单个物体类别的、具有特定性质的、互不相交的区域，而语义分割是一种更精细的分割，即在像素级别上进行分类，为输入图像中每个像素分配一个语义类别标签，将属于同一类的像素归为一类。

输入图像： $R^{H \times W \times 3}$ ，H和W分别代表输入图像的高与宽。

输出图像： $R^{H \times W \times classNum}$ ，classNum为给定的数据集的类别数，也就是一共classNum个通道的H*W大小的特征图。每一个通道对应一个class,对每一个像素位置，都有classNum数目个通道,每个通道的值对应那个像素属于该class的预测概率。

2.1 语义分割技术评价指标

随着深度学习方法广泛应用于语义分割领域，众多算法都取得了较好效

果，但在对语义分割的结果进行验证和评价时，不仅需要视觉感知上的定性评价，还需要客观的定量评价。下图中展示了两个语义分割的结果A和B，要判断A与B哪一个分割结果更好，即更接近Ground Truth，就需要一些定量的、具有可计算性的、可比较性的、不受人主观意识影响的评价指标。

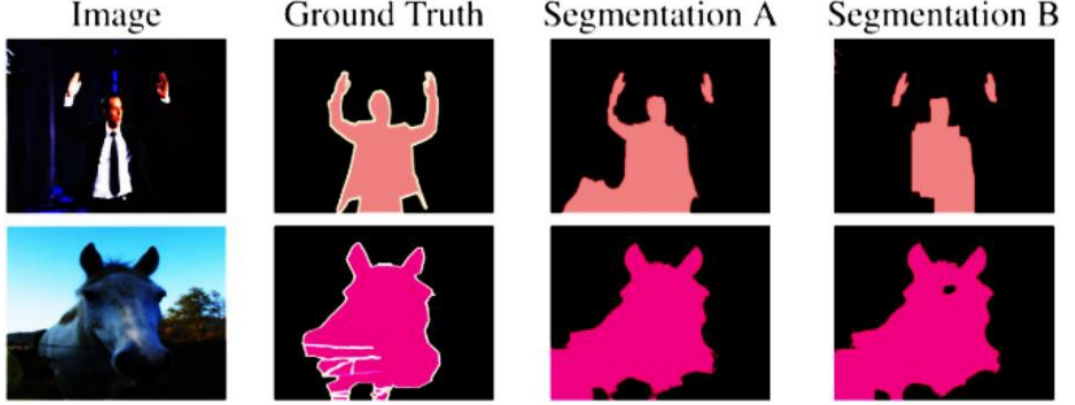


图 1 A\B两种分割结果对比和正确的分割结果

随着语义分割研究的进行，已经提出了许多评价指标。评价指标是评析算法优劣的标准，针对不同的应用场景和算法模型选择恰当、全面的评价指标，对算法优化与改进有着举足轻重的作用[1]。由于语义分割是像素级的分类，评价指标也主要从像素级别的精确度和交并比等组成。

假设分割类别为 k 类， p_{ij} 表示属于类别 i 的像素被分类到类别 j 的像素数量，那么 p_{ii} 即表示 i 类被正确分类的像素数量。

2.1.1 像素精度 (Pixel Accuracy, PA)

像素精度 (Pixel Accuracy, PA)，又称Global Accuracy，表示分割正确的像素数量所占的比例，即正确分类的像素数量占图像空间像素总数的比例：

$$PA = \frac{\sum_{i=1}^k p_{ii}}{\sum_{i=1}^k \sum_{j=1}^k p_{ij}}$$

像素精度描述图像中所有类别的分割准确性，常用于总体分割效果的估计。像素精度越接近1，表示模型性能越好。

由于像素精度包含了所有分类类别的准确度，因此容易忽略某一类别分割效果差的情况，并且不会考虑到单个类别的分割准确性。因此，考虑类别像素精度 (Class Pixel Accuracy, CPA) 分别评估每个类别的分割准确性。对于第 i 个分类：

$$CPA = \frac{p_{ii}}{\sum_{j=0}^k p_{ij}}$$

2.1.2 类别像素精度 (Class Pixel Accuracy, CPA)

平均像素精度，表示所有CPA的平均值。

$$mPA = \frac{1}{k} \sum_{i=1}^k \frac{p_{ii}}{\sum_{j=1}^k p_{ij}}$$

平均像素精度常用于总体的像素分割效果估计，相较于像素精度，结果更加可靠。

2.1.3 交并比 (Intersection over Union, IoU)

交并比主要研究的是相关性，相关性主要指预测分割区域与标签区域之间的相似程度。交并比 (Intersection over Union, IoU) 表示预测掩码与标签像素的交叠率，用于评价预测的前景区域是否精准：

$$IoU = \frac{|A \cap B|}{|A \cup B|} = \frac{\sum_{i=1}^k p_{ii}}{\sum_{i=1}^k (\sum_{j=1}^k p_{ji} + \sum_{j=1}^k p_{ij} - p_{ii})}$$

交并比的取值范围是0~1，最理想情况是预测区域与标签区域完全重叠，此时IoU的值为1。

2.1.4 平均交并比 (mean-IoU, mIoU)

平均交并比是计算每个类别中IoU值的算术平均值，用于总体数据集的像素重叠情况：

$$mIoU = \frac{1}{k} \sum_{i=1}^k \frac{p_{ii}}{\sum_{j=1}^k p_{ij} + \sum_{j=1}^k p_{ji} - p_{ii}}$$

交并比的取值范围是0~1，最理想情况是预测区域与标签区域完全重叠，此时IoU的值为1。

2.2 语义分割常用数据集

本节我们主要介绍了常用的语义分割数据集，同时也是自动驾驶中常用的数据集。选择一个合适的数据集对于训练和评估所建立的模型十分重要。

2.2.1 CamVid(2009)

CamVid[24]是一个欧洲道路场景的图像数据集，一共有700张图像，32个类别。其中一些最重要的类别有：汽车、行人、摩托车、信号灯、交通锥、车道线、交通标志、车道、卡车/巴士以及儿童。

2.2.2 KITTI(2013)

KITTI[25]是自动驾驶领域的一个十分重要数据集，它包含了大量的交通场景，这些数据是由一组不同的传感器收集的，如RGB和灰度相机以及3D激光扫描仪。

2.2.3 COCO(2014)

COCO[26]是一个通用的数据集，常用于物体检测和语义分割等任务。它包含328,000张图像，其中82,873张用于训练，约41,000张用于验证，剩余的用于测试。

2.2.4 Cityscapes(2016)

Cityscapes[27]是语义分割在自动驾驶领域中最常用的数据集之一。数据集是在50个不同城市春夏秋三个季节的晴天拍摄得到的，里面有大量的物体，不同的场景和背景。它提供了8个大类别以及细分后的30个子类别，包含大约5000张精标注的图像和20000张粗标注的图像。

2.2.5 SYNTHIA(2016)

SYNTHIA[28]数据集共包含9400个道路场景图像，这些图像是从一个模拟城市中采集的。它提供了13个类别。图像的分辨率为1280×960。

2.2.6 Mapillary Vistas(2017)

Mapillary Vistas Dataset[29]是一个人工标注且十分详细的道路场景数据集，它提供了25,000张高分辨率的图像，124个类别以及100个特定实例注释的类别，涵盖了6个大洲的场景。

2.2.7 Apollo Scape(2018)

Apollo Scape[30]是一个大型数据集，包括在不同的天气条件下，来自中国不同地点的140,000多个视频帧（73个街道视频）。

2.3 传统语义分割算法

传统的语义分割算法与图像分割算法有相似之处，根据灰度、彩色、空间纹理等特征将图像划分为若干个互不相交的区域，使得这些特征在同一个区域内表现出一致性或者相似性，而在不同的区域间表现出明显不同。划分方法主要有：基于阈值分割[3]、基于区域分割[4]、边缘分割和图论分割[5]。

阈值分割方法是常用的分割技术之一，其实质是根据一定的标准自动确定最佳阈值，并根据灰度级使用这些像素以实现聚类。

基于区域的分割方法是以直接寻找新区域为基础的分割技术，可分为区域生长和区域分裂合并两种基本提取方式。区域生长以单个像素点为基础，将具有相似特征的像素点聚合到一起形成区域，其计算简单，对于均匀分布的图像具有良好效果。区域分裂合并从整体图像出发，通过像素点之间的分裂得到各子区域，四叉树分解法就是其典型代表方法。

基于边缘检测的分割方法通过检测不同区域边缘分割图片，最简单的边缘检测方法是并行微分算子法，它利用相邻区域的像素值不连续的性质，采用导数检测边缘点。

由于传统的方法只是人工提取图像的低级语义，如大小、纹理、颜色等，无法解决复杂场景中的分割任务，且需要研究人员手动调整参数，其训练出的模型效果远差于深度学习方法，传统方法已不适用于实时快速的特征提取。

2.4 深度学习与传统方法相结合的图像语义分割

随着深度学习技术的发展，传统语义分割方法开始尝试与深度学习模型结合，即在传统语义分割方法处理的基础上，更进一步的采用卷积神经网络等方法学习目标特征并训练分类器，对目标区域进行分类并实现语义标注。卷积神经网络模型的提出，为图像语义分割与深度学习的结合奠定了基础，使得图像语义分割技术应用于多个应用领域。

卷积神经网络使用卷积层—激活函数—池化层—全连接层的运行结构，输入图像经卷积层聚拢不同局部区域特征，通过激活函数（Sigmoid、Relu、Tanh等）部分激活，部分抑制从而强化特征。池化层在不改变目标对象的基础上，使输入图片变小，减少训练参数，最后使用全连接层神经元的前向传播与反向传导损失计算函数最优点，使输入图像的分类、分割等更加高效。训练研究人

员以卷积神经网络为基础提出Alex Net、VGGNet、Google Net、Res Net等图像分类网络模型。出现时间以及创新点如下表所示。

年份	事件	结构创新点
2012	Hinton研究组提出 AlexNet	首创深度卷积神经网络模型，成功使用Relu激活函数，训练时使用Dropout避免过拟合
2014	牛津大学的Oxford Visual Geometry Group 提出卷积神经网络 VGG	将网络分为5组，基本使用3x3小卷积核和2x2的小池化核,使得网络模型框架更宽更深
2014	Szegedy等设计第一个 Inception架构的网络 GooleNet	以Inception模块为基础，通过模块串联构成网络，使用大量1x1卷积在有限的感受野中叠加更多卷积层数
2014	He等提出 ResNet 网络	残差模块的使用极大增加了网络深度，跨层连接的提出也解决了梯度消失问题

表 1 经典的深度学习网络模型出现时间以及创新点

2.5 基于深度学习的语义分割方法

随着全卷积神经网络的提出，图像语义分割技术进入新时期，计算机在视觉领域通过深度学习网络进行全卷积后能够极大提高图像分类效率与识别准度，网络框架与语义分割问题进入深度结合快速扩展的时代。研究人员以全卷积神经网络为基础提出了SegNet等图像分割网络结构模型。表2为网络框架的发展历程与创新点。

年份	事件	结构创新点
2015	Jonathan Long等提出 FCN	创造性地使用上采样代替全连接层且接受任意大小的输入图片
2015	Olaf Ronneberger等提出U-Net	网络框架采用左右对称的U形字母结构其池化层被上采样层取代，可运行小批量图片，在医疗图像处理中有更大作用
2015	Badrinarayanan等提出	基于编码器-解码器网络结构，利用

	SegNet	上采样方式恢复图像尺寸，去掉全连接层，大大减少了参数，提升了网络运行速度
2017	Zhao等提出PSPNet	使用空洞卷积改善ResNet结构，并添加了一个金字塔池化模块
2017	Lin等提出使用链式残差连接的RefineNet	对解码器结构进行改进,形成Long-range残差连接能通过上采样方式融合底层和高层语义特征
2018	Chen等提出Deep-Labv3+	V3+版本使用改进版的Xception作为基础网络，加强了图像边缘分割效果
2018	YU等提出双向分割网络BiSeNet	包含SpatialPath和Context Path，分别用来解决空间信息缺失和感受野缩小的问题
2019	何凯明等提出全景FPN	将FCN和Mask R-CNN相结合，使用丰富的多尺度特征，可同时解决语义分割与实例分割任务

表 2 网络框架的发展历程与创新点

2.5.1 全卷积网络（FCN）

全卷积网络（Fully Convolutional Networks, FCN）通过扩展普通卷积网络模型，使其具有更多的参数特征和更好的空间层次。

其结构可以分为全卷积和反卷积两部分，全卷积借鉴卷积神经网络模型，输入图像在参数减少与特征强化后，采用反卷积层对最后卷积层的特征图进行上采样，通过转置卷积恢复输入图像尺寸，从而针对每个像素都产生一个预测，使输入图像达到语义级分割。图2为FCN示意图。

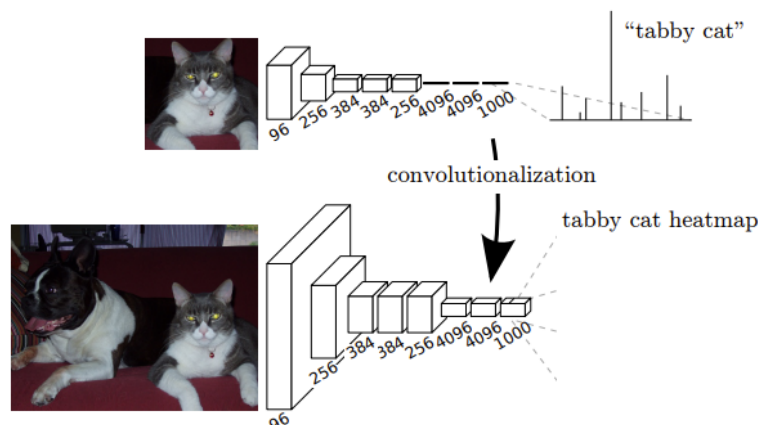


图 2 将全连接层转换成卷积层，使得分类网络可以输出一个类的热图。

通过将经典分类网络（AlexNet、VGG、GoogLeNet）调整为完全卷积网络，实现了数据集PASCAL VOC（2012年将mIU提升至62.2%）、NYUDv2和SIFT Flow的最优分割[6]。

其优点在于将卷积神经网络对于图像的识别精度从图像级识别提升为全卷积神经网络中像素级的识别，但是使用全卷积网络的图像分割仍存在分割结果不够精准、输出图像模糊等问题。

2.5.2 SegNet

SegNet 的新颖之处在于解码器对其较低分辨率的输入特征图进行上采样的方式。具体地说，解码器使用了在相应编码器的最大池化（max pooling）步骤中计算的池化索引来执行非线性上采样。这种方法消除了学习上采样的需要。经上采样后的特征图是稀疏的，因此随后使用可训练的卷积核进行卷积操作，生成密集的特征图[7]。图3为SegNet架构图。

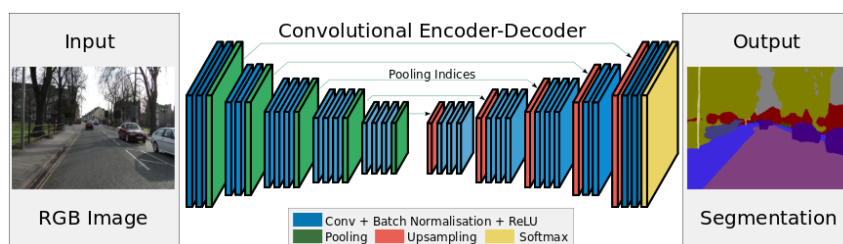


图 3 SegNet 架构

我们将我们所提出的架构与广泛采用的 FCN 以及众所周知的 DeepLab-LargeFOV, DeconvNet 架构进行比较。比较的结果揭示了在实现良好的分割性能时所涉及的内存与精度之间的权衡，即在分割中保持高频细节的完整性，又是一个拥有较少参数的轻量级网络。但是，反池化操作对于低分辨率的特征图，会忽略邻近的信息。

2.5.3 U-Net

U-Net 架构包括一个捕获上下文信息的收缩路径和一个支持精确本地化的对称扩展路径。已经证明了这样一个网络可以使用非常少的图像进行端到端的训练，并且在ISBI神经元结构分割挑战赛中取得了比以前最好的方法（一个滑动窗口的卷积网络）更加优异的性能[8]。图4为U-Net网络架构。

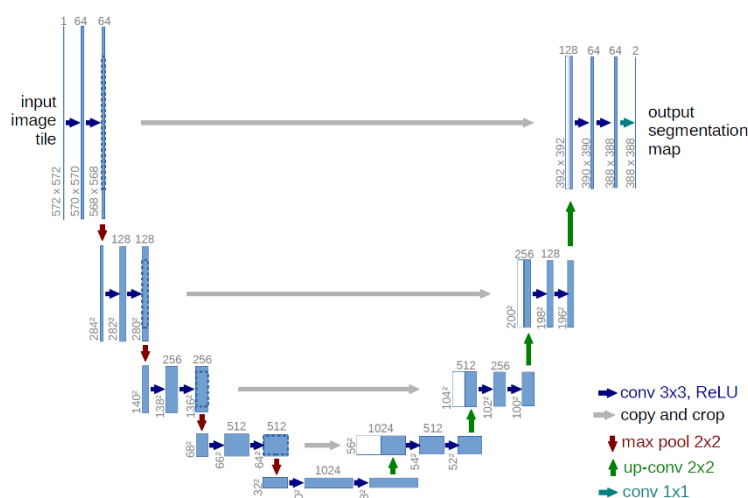


图 4 U-Net 架构

U-Net 在 EM 数据集上取得了最优异的结果，该数据集只有30个密集标注的医学图像和其他医学图像数据集，U-Net 后来扩展到3D版的 3D-U-Net。虽然 U-Net 最初的发表在于其在生物医学领域的分割、网络的实用性以及从非常少的数据中学习的能力，但现在已经成功应用其他几个领域，例如 卫星图像分割，同时也成为许多kaggle竞赛中关于医学图像分割的获胜的解决方案的一部分。

2.5.4 Deeplab 网络模型

DeepLab v1结合 DCNN 和概率图模型来解决像素级分类任务（即语义分割）。提出 空洞卷积（atrous convolution）。在最后两个最大池化操作中不降低特征图的分辨率，并在倒数第二个最大池化之后的卷积中使用空洞卷积。并且使用 CRF（条件随机场）作为后处理，恢复边界细节，达到准确定位效果。最后，附加输入图像和前四个最大池化层的每个输出到一个两层卷积，然后拼接到主网络的最后一层，达到 多尺度预测效果[9]。

Deep Lab v1结合深度卷积神经网络和概率图模型，形成完整端对端网络模型，但是其空间分辨率低，存储空间需求量大。

空洞卷积允许显式地控制在深度卷积神经网络中计算的特征响应的分辨

率。它还允许我们有效地扩大过滤器的视野，在不增加参数数量或计算量的情况下引入更大的上下文。其次，提出了一种空洞空间金字塔池化（ASPP）的多尺度鲁棒分割方法。ASPP 使用多个采样率的过滤器和有效的视野探测传入的卷积特征层，从而在多个尺度上捕获目标和图像上下文。第三，结合 DCNNs 方法和概率图模型，改进了目标边界的定位。DCNNs 中常用的最大池化和下采样的组合实现了不变性，但对定位精度有一定的影响[10]。图5为DeepLab v1与DeepLab v2架构对比图。

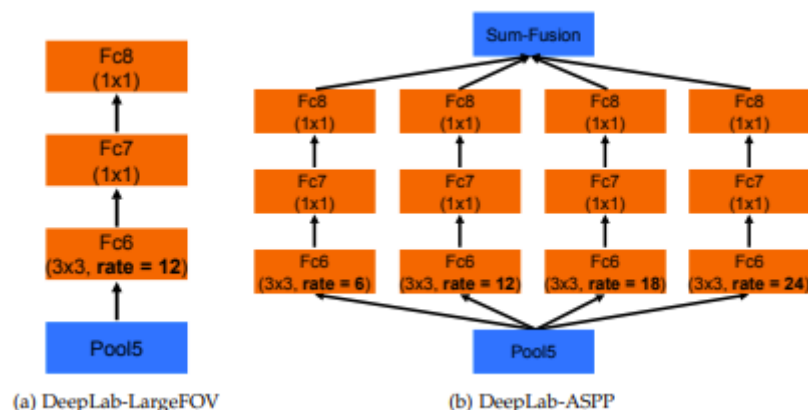


图 5 (a) DeepLab v1, (b) DeepLab v2

Deep Lab v2空洞金字塔的提出，提高了模型优化能力，但图像细节模糊处理能力下降。

为了解决多尺度目标的分割问题，串行/并行设计了能够捕捉多尺度上下文的模块，模块中采用不同的空洞率。此外，增强了先前提出的空洞空间金字塔池化模块，增加了图像级特征来编码全局上下文，使得模块可以在多尺度下探测卷积特征。提出的“DeepLab v3”系统在没有 CRF 作为后处理的情况下显著提升了性能[11]。图6为DeepLab v3网络架构图。

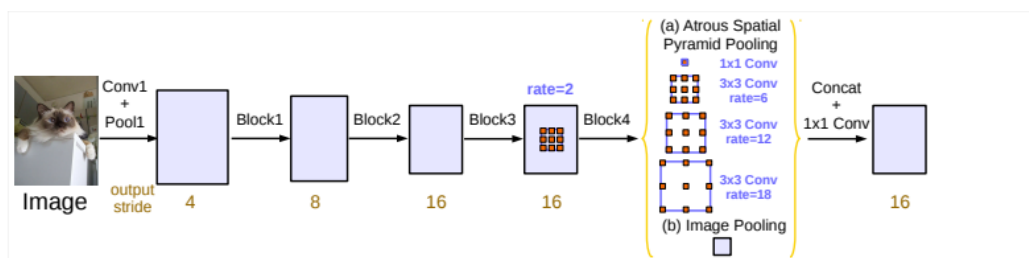


图 6 DeepLab v3网络架构

Deep Lab v3+被提出，其使用编码器解码器结构，使用改进版的Xception作为基础网络，弥补了之前版本网络的缺陷性，但仍需继续提高模型运行速度与性能，如随着Deep Lab算法的发展，深度学习具有更新太快、周期较长、完善缺陷困难等问题。

2.5.5 PSPNet 网络模型

利用基于不同区域的上下文信息集合，通过金字塔池化模块，使用提出的金字塔场景解析网络（PSPNet）来发挥全局上下文信息的能力。全局先验表征在场景解析任务中产生了良好的质量结果，而 PSPNet 为像素级的预测提供了一个更好的框架，该方法在不同的数据集上达到了最优性能[12]。图7为PSPNet网络模型。

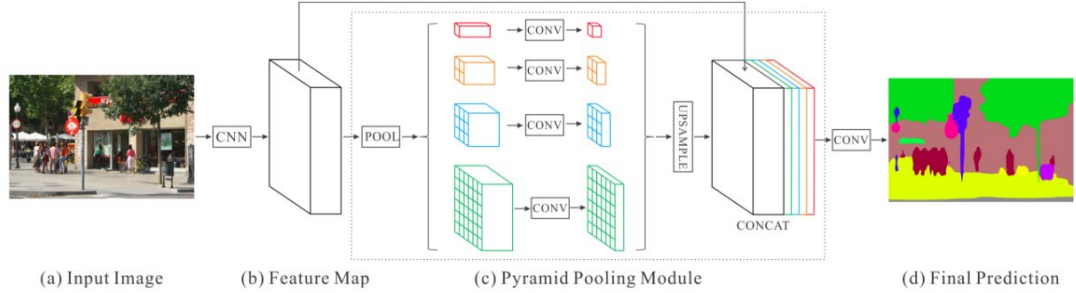


图 7 PSPNet 网络架构

PSPNet通过引入空洞卷积来修改基础的 ResNet 架构，特征经过最初的池化，在整个编码器网络中以相同的分辨率进行处理（原始图像输入的 $\frac{1}{4}$ ），直到它到达空间池化模块。并且在 ResNet 的中间层中引入辅助损失，以优化整体学习。最后，在修改后的 ResNet 编码器顶部的空间金字塔池化聚合全局上下文。

2.5.6 总结

基于深度学习的图像语义分割技术虽然取得了良好分割效果，但是其网络训练需要大量数据集，像素级别的图像质量难以保证，原因是大量使用基于强监督的分割方法，依赖于人工数据标记，且对未知场景适应能力差。

3 强化学习概述

强化学习（Reinforcement learning, RL）是一个原则性的数学框架，智能体通过在探索和利用之间的权衡来与环境进行交互。马尔可夫决策过程

（Markov decision process, MDPs）是强化学习的一种理想形式[13]，其中 (S, A, P, R, γ) ，S和A分别代表状态和动作的集合。 $P(s_{t+1}|s_t, a_t): S \times S \times A \rightarrow [0,1]$ ，将状态-动作对映射到下一步状态的转移函数。数值即时奖励函数 $R(s_t, a_t, s_{t+1}): S \times A \times S \rightarrow R$ ，作为学习信号。折扣系数 $\gamma \in [0,1]$ 决定了未来奖励在当前状态的价值(γ 越低，智能体会出现更多的短视行为)。部分可观察的

马尔可夫决策过程（Partially observable MDPs）将MDP扩展到无法访问完全可观察的马尔可夫属性状态的问题。POMDP有一个观察集 Ω 和一个观察函数 O ，其中 $O(a_t, s_{t+1}, o_{t+1}) = p(o_{t+1}|a_t, s_{t+1})$ 是观察 o_{t+1} 概率，并假设智能体执行动作 a_t ，到达状态 s_{t+1} 。

MDP智能体根据当前的状态 s_t ，在每个时间步 t 处选择一个行动 $a_t \in A$ ，接受奖励值 r_{t+1} ，再访问一个新状态 s_{t+1} 。生成串行的 $s_0, a_0, r_1, s_1, a_1, r_2, \dots$ ，称之为展开或轨迹。未来的预期累计奖励，即时间步长 t 之后的预期收益 G_t ，定义为[13]

$$G_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \dots = \sum_{k=0}^T \gamma^k r_{t+k+1}$$

其中 T 分别表示有界和无界视界问题的有限值或 ∞ 。策略 $\pi(a|s)$ 将状态映射到选择每个动作的概率。值函数 $v_\pi(s)$ 是从状态 s ，执行策略 π 后的预期回报：

$$v_\pi(s) = E_\pi[G_t | s_t = s]$$

相似的，动作值函数 $q_\pi(s, a)$ 定义为：

$$q_\pi(s_t, a_t) = E_{s_{t+1}}[r_{t+1} + \gamma q_\pi(s_{t+1}, \pi(s_{t+1}))]$$

强化学习的目标是确定最大化预期回报的最佳策略：

$$\pi^* = \operatorname{argmax}_\pi E_\pi[G_t | s_t = s]$$

实现强化学习目标的方法主要有三种，包括：基于价值的方法、基于策略的方法、演员-评论家方法。

3.1 基于价值的方法

为了解决强化学习问题，可以识别最优动作-值函数，并从学习到的状态-动作值中恢复最优策略。

$$q_{\pi^*}(s, a) = \max_{\pi} q_{\pi}(s, a)$$

$$\pi^*(s) = \operatorname{argmax}_a q_{\pi^*}(s, a)$$

其中，Q-learning[15]是最流行的方法之一，它通过时间差（TD）来计算Q值：

$$q_{\pi}(s_t, a_t) \leftarrow q_{\pi}(s_t, a_t) + \alpha(Y - q_{\pi}(s_t, a_t))$$

其中 $Y = r_{t+1} + \gamma \max_{a_{t+1} \in A} q_{\pi}(s_{t+1}, a_{t+1})$ 是时间差异目标， α 是学习率。这可以被视为一个标准回归问题，其中要最小化的误差是 $Y - q_{\pi}(s_t, a_t)$ 。Q-learning是off-policy的，因为它根据不一定由派生策略生成的经验更新 q_{π} 。与

TD方法相比，蒙特卡罗方法通过平均多次推出的结果来估计预期收益，这可以应用于非马尔可夫情景环境。TD和蒙特卡罗在 $TD(\lambda)$ 中进一步结合[13]。

最近，普遍使用DNN来近似参数化值函数 $q(s, a; \omega)$ ，其中DQN[4]最具代表性，它通过最小化以下损失函数来学习值：

$$J(\omega) = E_t \left[(Y - q(s_t, a_t; \omega))^2 \right]$$

其中 $Y = r_{t+1} + \gamma \max q(s_{t+1}, a_{t+1}; \gamma^-)$ 是目标， γ^- 表示目标网络的参数， ω 可以根据梯度学习。

$$\omega \leftarrow \omega - \alpha E_t [(Y - q(s_t, a_t; \omega)) \nabla q(s_t, a_t; \omega)]$$

DQN的主要贡献是引入了目标网络和体验回放。为了避免Q值快速波动并稳定训练，目标网络在主Q网络更新期间固定指定次数的迭代，随后更新以匹配主Q网络。

3.2 基于策略的方法

通过直接搜索和优化参数化策略 π_θ 来最大化期望汇报：

$$\max_\theta J(\theta) = \max_\theta v_{\pi_\theta}(s) = \max_\theta E_{\pi_\theta}[G_t | s_t = s]$$

其中 θ 表示策略参数，可以根据策略梯度定理[13]进行优化：

$$\begin{aligned} \nabla J(\theta) &\propto s \mu(s) a q_\pi(s, a) \nabla \pi_\theta(a | s) \\ &= E_\pi \left[\sum_a q_\pi(s_t, a) \nabla \pi_\theta(a | s_t) \right] \\ &= E_\pi [G_t \nabla \ln \pi_\theta(a_t | s_t)] \end{aligned}$$

其中 μ 表示状态的出现频率。强化[17]是一种基于策略的简单蒙特卡罗方法，它选择策略 π_θ 之后的折扣收益 G_t 来估计上式中的策略梯度。参数更新如下[13]：

$$\theta \leftarrow \theta + \alpha G_t \nabla \ln \pi_\theta(a_t | s_t)$$

此更新直观地增加了导致更高回报的行为的出现概率。

基于策略的方法的一个问题是梯度更新不佳可能会导致新更新的策略与以前的策略大相径庭，这可能会降低策略性能。信任区域策略优化（Trust region policy optimization, TRPO）[18]通过优化代理目标函数解决了这个问题，保证了策略性能的单调改进。每个策略梯度更新都通过使用策略之间Kullback-Leibler（KL）散度的二次近似来约束。

3.3 Actor-Critic Methods(演员-评论家方法)

actor-critic方法具有基于价值和基于策略的方法的优点，其中神经网络中的演员 $\pi_{\theta}(a|s)$ 选择动作，神经网络评论家使用 $q(s, a; \omega)$ 或 $v(s; \omega)$ 来估计价值。确定性策略梯度 (DPG)[19]是一种派生确定性策略的离线策略 actor-critic 算法。与随机策略相比，DPG 仅在状态空间上进行集成，在动作空间较大的问题中需要较少的样本。Deep deterministic policy gradient (DDPG)[20]利用 DNN 在高维状态空间上运行，具有经验回放和独立的 actor-critic 目标网络，类似于 DQN。并行计算的开发是经验重放的替代方法。异步优势 actor-critic (A3C)[21] 在 actor critic 框架中使用优势估计而不是折扣回报，并在环境的多个并行线程上异步更新策略和价值网络。

平行的独立环境稳定了学习过程，使更多的探索成为可能。Advantage actor critic (A2C)[22]是 A3C 的同步版本，为简单起见使用单个智能体或等待每个智能体完成其体验以收集多个轨迹。弱演员评论家 (Soft actor critic, SAC)[23]受益于向奖励函数添加熵项以鼓励更好的探索。

4 语义分割在自动驾驶中的应用

自动驾驶是指车辆可以自动感知和理解周围环境并自动导航，已经引起越来越多的汽车行业和科研界的关注。自动驾驶汽车必须了解其周围环境，即其他汽车、行人、车道、交通标志或交通信号灯，语义分割技术是实现这一目标的有力工具。近年来，深度学习的成功对语义分割方法产生了巨大影响，显著提高了其在准确性方面的表现，于是越来越多的研究人员尝试将基于深度学习的语义分割技术落地应用到自动驾驶中去。

但语义分割在自动驾驶中的应用最大的问题不是准确性，而是实时性。不同于在具有强大GPU的服务器或工作站上训练的神经网络，自动驾驶汽车中的嵌入式设备计算能力有限，而且需要对周围新的事件做出即时反应，于是减少语义分割的计算量，提高实时性变得至关重要。相反，如果不能像素级地识别到物体的边界，并不会对交通安全造成特别重大的影响，因此，有必要改进分割模型的设计，以实现高效的架构，仍能够以适当的准确度实时执行。为此，在本文中，我们从准确性和实时性对语义分割的架构进行了回顾。为了对所考察的架构进行最完整的评估，所有的模型都在一个一致的评估框架中根据其性能进行了比较。

4.1 评估指标

在本节中，我们总结了一些常用的评估指标。对于实时语义分割模型的评估主要有两个关键因素：准确性(Effectiveness)和实时性(Efficiency)，其中章节2中的指标用来评估模型的准确性，而实时性的评估有以下几种指标。

帧率(Frames per second): 模型处理一段视频图像帧所需时间，由于便于比较不同的语义分割方法和架构，在实时语义分割领域中帧率是一个常用的评估指标。

推理时间(Inference time): 模型处理一帧图像的执行时间，是用于评估语义分割速度的另一个标准指标。

内存使用量(Memory usage): 在比较深度学习模型的速度和效率时，它是一个重要参数。内存使用量可以用不同的方式测量。一些研究人员使用网络的参数数量。另一种方法是定义内存大小来表示网络，最后，一个经常使用的指标是测量执行所需的浮点运算的数量(FLOPs)的执行。

4.2 实时语义分割模型

在本节中，我们介绍一些经典的实时语义分割模型，并使用"Cityscapes"数据集评估各个模型的准确性和实时性。

4.2.1 ENet

ENet(2016)[31]是最早提出的专门针对实时语义分割的模型，它是一个高效深度神经网络，旨在实现快速推理和高精确度，也是一个里程碑式的研究工作。它使用的是编码器-解码器结构，如表3所示，ENet由一个大的编码器和一个非常简单的解码器组成，它是由几个变化的瓶颈残差块构成，包括降维、卷积和扩展。卷积层可以是标准的 3×3 形式，不对称和扩张的 5×5 形式，或者是解码器中的分层形式。此外，ENet引入了早期下采样以实现低延迟操作了，即在网络的第一层对输入图像进行下采样。这种技术可以使得输入图像的下采样在不损失重要信息的情况下更加有效，从而不牺牲准确性，形成了一个非常紧凑和快速的架构，可用于部署在嵌入式设备上。

Name	Type	Output size
initial		$16 \times 256 \times 256$
bottleneck1.0	downsampling	$64 \times 128 \times 128$
4× bottleneck1.x		$64 \times 128 \times 128$
bottleneck2.0	downsampling	$128 \times 64 \times 64$
bottleneck2.1		$128 \times 64 \times 64$

bottleneck2.2	dilated 2	$128 \times 64 \times 64$
bottleneck2.3	asymmetric 5	$128 \times 64 \times 64$
bottleneck2.4	dilated 4	$128 \times 64 \times 64$
bottleneck2.5		$128 \times 64 \times 64$
bottleneck2.6	dilated 8	$128 \times 64 \times 64$
bottleneck2.7	asymmetric 5	$128 \times 64 \times 64$
bottleneck2.8	dilated 16	$128 \times 64 \times 64$
<i>Repeat section 2, without bottleneck2.0</i>		
bottleneck4.0	upsampling	$64 \times 128 \times 128$
bottleneck4.1		$64 \times 128 \times 128$
bottleneck4.2	upsampling	$64 \times 128 \times 128$
bottleneck5.0		$16 \times 256 \times 256$
bottleneck5.1		$16 \times 256 \times 256$
fullconv		$C \times 512 \times 512$

表 3 ENet结构

4.2.2 ERFNet

ERFNet(2017)[32]是基于编码器-解码器结构的。如图8，它由一个以残差连接和因子化卷积为特征的层构成，旨在维持效率，同时又足够精确。为了加快处理时间，设计者选择了一个小的解码器尺寸和解旋来简化内存和计算成本。

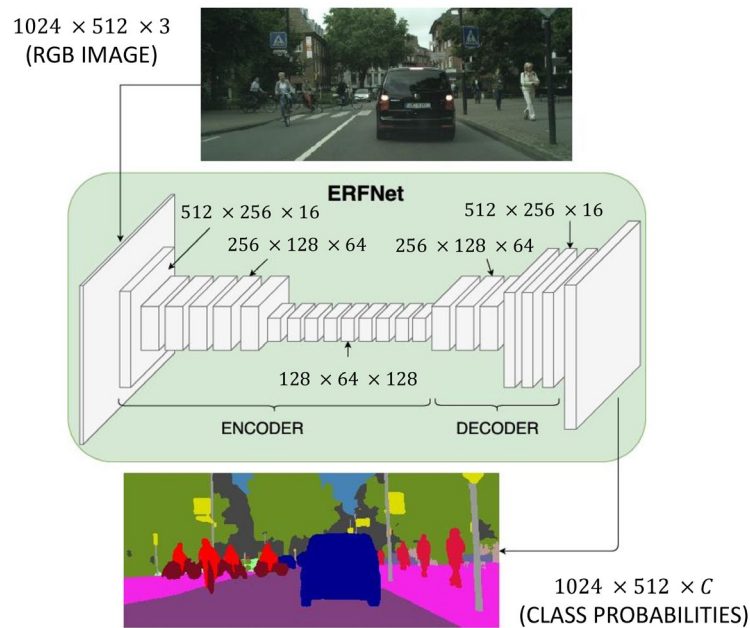


图 8 ERFNet结构

4.2.3 ContextNet

ContextNet(2018)[33]是由两个分支构成的模型。从图9中可以发现，第一个分支在低分辨率下实现了高效和精确的语义分割，而第二个分支在高分辨率下结合了一个子网络，以提供更详细的语义分割结果。ContextNet还使用了深度可分离的卷积，以加快推理时间和瓶颈残留块。另一方面，ContextNet通过使用网络压缩和金字塔表示方法来以较低的内存进行实时分割。此外，ContextNet还通过修剪将其最佳地重新缩减到原来的预定维度，并对上下文分支的输出施加辅助损失以帮助学习。

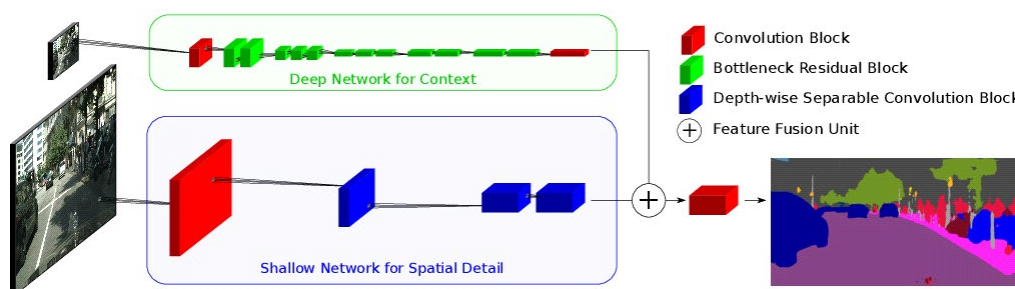


图 9 ContextNet结构

4.2.4 BiSeNet

BiSeNet(2018)[34]是另一种双分支结构，由一个以原始尺寸的1/4输入的上下文路径和一个以全分辨率输入的空间分支组成。通过这种方式，它提高了准确度，同时避免了升采样操作，从而保持了较低的计算成本。BiSeNet V2(2021)[35]是BiSeNet的升级版，在速度和精度之间有了更好的权衡。它的特点是有一个引导聚合层来融合从细节和语义分支中提取的特征，并提出了助推训练策略，在语义分支中使用快速下采样来提高特征表示的水平，并迅速增加感受野。两个网络的结构如图10和图11。

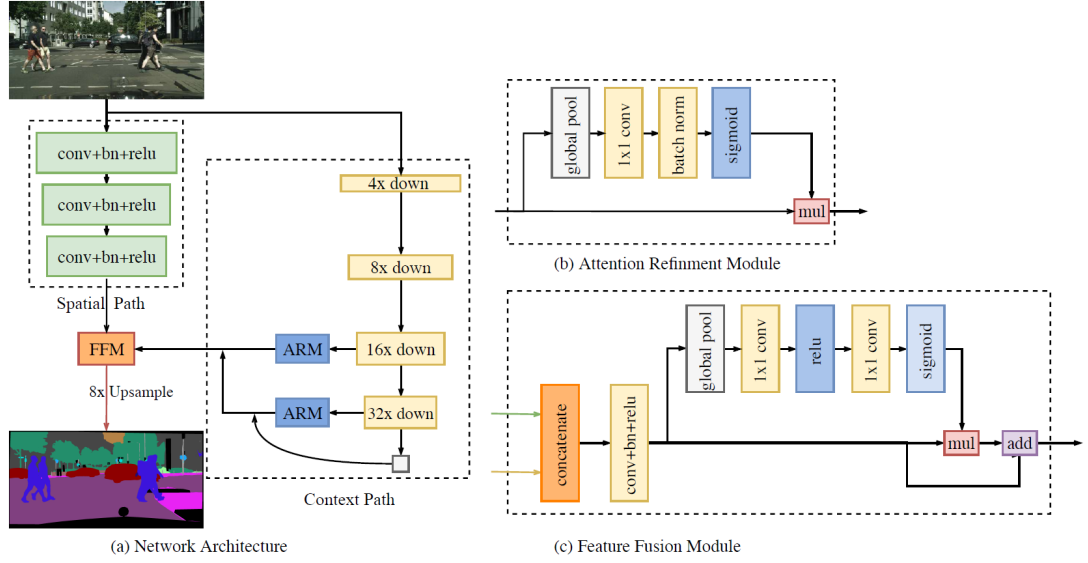


图 10 BiSeNet结构以及特征融合模块和注意力细化模块

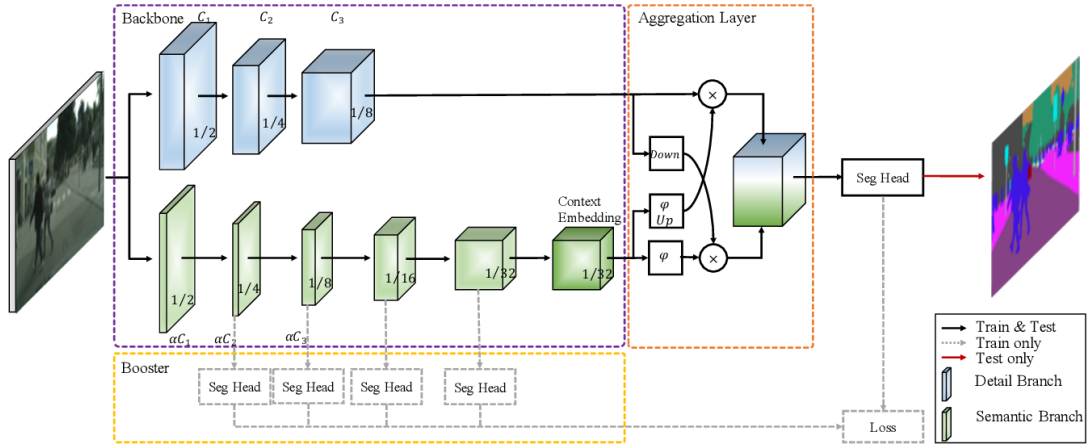


图 11 BiSeNet V2结构

4.2.5 Fast-SCNN

Fast-SCNN(2019)[36]是一个非常轻巧的架构，能够在极高的帧率下提供良好的语义分割结果。如图12所示，作者在文章中提出了一个 "学习降样" 模块，用于并行计算多个分辨率分支的低层次特征。Fast-SCNN还具有一个全局特征提取器，用于捕捉语义分割的全局背景。最后，Fast-SCNN使用深度可分离卷积和剩余瓶颈块来提高速度，其中大部分的卷积都是深度可分的，减少参数数量和计算成本。

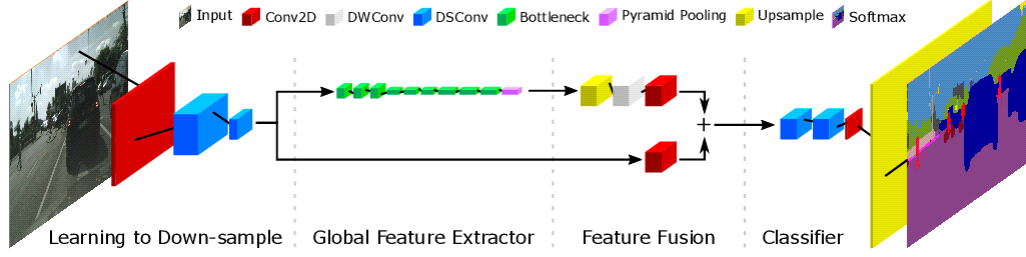


图 12 Fast-SCNN结构

4.2.6 FasterSeg

FasterSeg(2019)[37]采用了神经架构搜索，并提出了一个解耦和细粒度的延迟正则化，以平衡高推理速度和低准确性之间的权衡。神经架构搜索的范围是以自动方式设计深度学习架构，从而减少人为因素的参与。另一个创新点是使用了知识提炼(knowledge distillation)，特别是引入了一个共同搜索的师生框架，从而提高了准确性。知识提炼的目的是将从一个大而复杂的网络（教师网络）中学到的特征转移到一个小而轻的网络（学生网络）。师生共同搜索框架作为一个超级网络工作，教师和学生网络共享相同的权重，不会在内存使用和规模方面产生额外的负担。网络结构如图13所示。

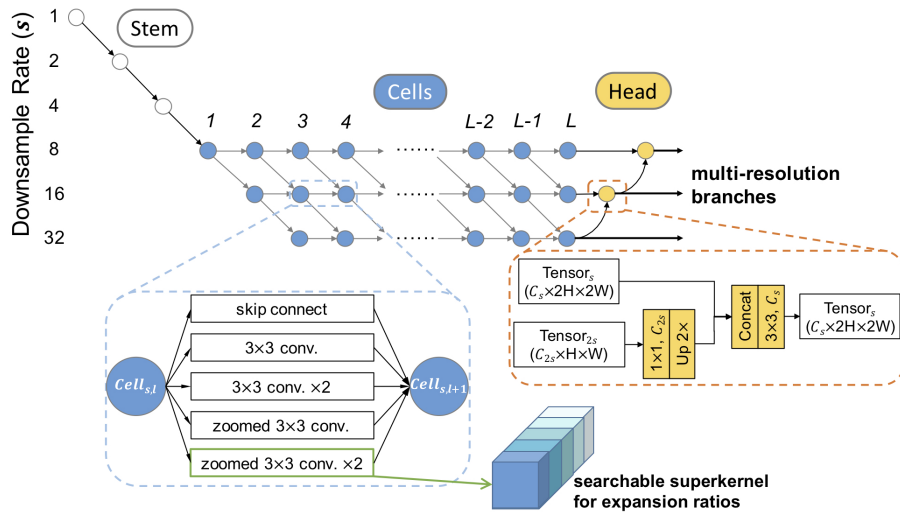


图 13 FasterSeg结构

4.2.7 ShelfNet

ShelfNet(2019)[38]的结构如图14所示，它是由多个编码器-解码器分支组成，但这些分支使用共享权重和一个残差块。为了减少推理时间，ShelfNet提出了通道减少，以有效地减少计算负担。使用不同的编码器-解码器分支改善了计算过程，提高了分割的准确性。同一残差块的卷积层之间共享权重，以减少网

络参数的数量而不降低准确性。

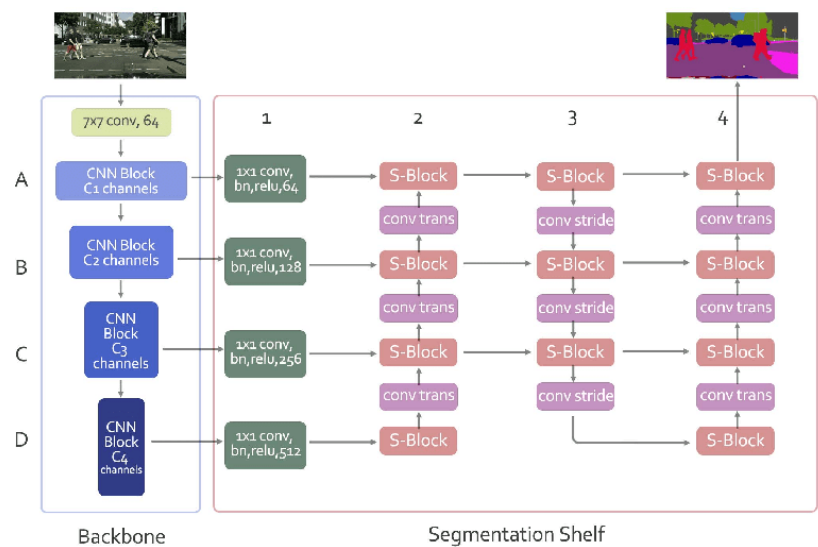


图 14 ShelfNet结构

4.3 模型性能对比

用于训练模型的数据是会很大程度上影响神经网络的性能，因此，在早期模拟实时的自动驾驶过程中，就应选取最合适的数据集进行训练。此外，对不同语义分割模型的比较应使用同一数据集，在实时语义分割领域，大多数模型都是在 "Cityscapes "或 "CamVid "数据集中进行比较的。

Networks	mIoU(%)	FPS
ShelfNet	74.8	59.2
BSeNet V2	72.6	156
FasterSeg	71.5	163.9
ERFNet	69.7	41.6
BiSeNet	68.4	105.8
Fast-SCNN	68	123.5
ContexNet	66.1	23.86
ENet	58.3	46.8

表 4 自动驾驶语义分割模型性能对比

5 强化学习在自动驾驶中的应用

自动驾驶系统的研究已经有了几十年的历史，关于它的研究变得越来越模

块化和细致化，包括但不限于场景信息提取、场景感知定位、场景抽象表示、规划和决策，控制。将强化学习模型应用到自动驾驶系统中大体上可以概括为以下几种模式。表5列出了每种模式的研究分类，以及所利用的强化学习方法、目标自动驾驶任务、模型架构的优缺点等信息。下面，我们将详细介绍各个模式，然后比较每种模式强化学习方法的使用。

模式	优点	缺点	研究	方法	任务
模式1：控制	结构简单，适应各种学习方法	仅限于指定的任务； 绕过计划模块会削弱模型的可解释性和能力	[39][40]	NQL	变道
			[41][42]		汇流
			[43][44]	DQN	交叉路口
			[45][46]	PPO	汇流
			[47][48]	DDPG	循迹
模式2：运动规划	模仿人类的轨迹； 高效预测	不保证安全性或可行性	[49]	DDQN/DQfD	自动泊车
			[50]	SAC	汇流
模式3：行为规划	深度神经网络（DNN）只需要计划高级行为操作	简单而少的动作限制了驾驶方式的控制精度和多样性； 复杂和过多的操作会增加培训成本	[51][52]	DQN	变道
			[53][54]		
			[55][56]		交叉路口
			[57][58]		
			[59][60]	DQN	变道
			[61][62]	Q-learning	变道
			[63][64]	DDQN	车道线
			[65]	DQfD	变道
模式4：分层规划和控制	同时在各种抽象级别进行规划	分层 DNN 会增加训练成本并降低收敛速度	[66]	QR-DQN	高速公路
			[67][68]	DQN	定速巡航
			[69]		变道
			[70]	Hierarchical Policy gradient	信号灯行驶
			[71]	DDPG	变道
			[72]	DDPG	交叉路口
				DDPG	城市道路

表 5 强化学习在自动驾驶中应用的四种模式

5.1 模式 1：控制

如图15，Tang等人提出使用环境栅格化编码，以及相对距离、速度和加速度，作为双分支神经网络的输入，该网络通过近似策略优化训练。

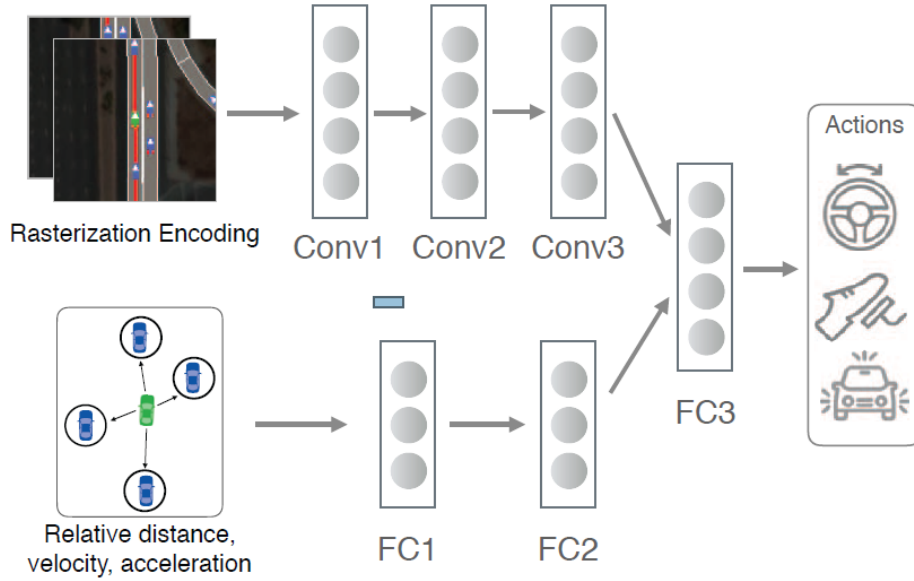


图 15 策略函数的架构图

模式1结构简单，而且能适应大量的学习方法，但它仅限于特定的任务；因此，它难以解决需要以各种情况为条件的多种驾驶技能的情景。此外，绕过和忽略行为规划或运动规划过程可能会削弱模型的可解释性和在复杂任务中的表现（如城市导航）。

5.2 模式 2：运动规划

Sun等人提出通过Sampled-DAgger训练一个模仿长期MPC的神经网络，其中策略网络的输入来自感知（障碍、环境和当前状态）和决策（驾驶决策）。另外，Wulfmeier等人提出将LiDAR点云投影到网格地图上，并将其发送给DNN。DNN负责学习一个指导运动规划的成本函数。

然而模式2的一个主要缺点是，尽管DNN规划的轨迹可以模仿人类的轨迹，但其安全性和可行性不能得到保证。

5.3 模式 3：行为规划

许多研究的重点是将DRL整合到行为规划模块中，并推导出高层次的驾驶策略。DNN决定行为动作，随后的运动规划和控制模块通常利用传统方法。许多研究建立在DQN及其变体的基础上，Yuan[59]等人将行动空间分解为高速公路上的五个典型行为决定。

Wang等人最近的一项研究提出了使用增强的对抗性逆向强化学习（AIRL）来学习高速公路上类似人类的决策，其中的行动空间由横向和纵向决

策的所有可能组合组成。

在模式3中，行为动作的设计是不容易的，必须平衡训练成本和驾驶风格的多样性。简单和少量的行为动作限制了控制精度和驾驶方式的多样性，而许多复杂的动作则增加了训练成本。

5.4 模式 4：分层规划和控制

模式4模糊了规划和控制之间的界限，单一的DNN根据参数化的动作空间输出分层的动作，或者分层DNN输出多层的动作。Rezaee[67]等人提出了一个架构，其中BP（行为规划）被用来对离散状态之间的转换进行高级决策，MoP（运动规划）根据通过BP做出的决策生成目标轨迹。

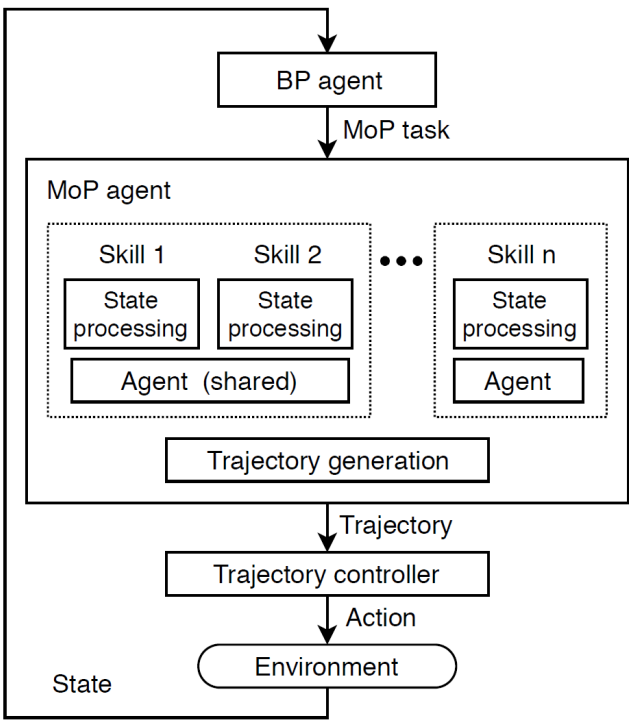


图 16 行为规划和运动规划

Qiao[71]等人在分层马尔可夫决策过程（HMDP）的基础上，通过一个选项框架成功改造了他们的模型。在他们的模型中，分层选项被建模为高层决策（SlowForward 或 Forward）。基于高层决策和当前的观察，得出了低层的控制策略。

模式4同时进行多级规划，低级规划过程考虑高级决策。然而，使用分层的DNN会导致训练成本的增加和收敛速度的降低，因为一个训练不好的高层DNN可能会误导和干扰低层DNN的学习过程。

6 结语

在上文中，我们具体了解了语义分割与强化学习的基础概况及其在自动驾驶领域的应用，给出了各技术的优点和缺点，通过上述的介绍可以知道，汽车的智能化自动驾驶是必然的发展方向，加强人工智能在汽车自动驾驶中的应用质效，是进一步实现汽车行业现代化发展的必然举措。

参考文献

- [1] 于营,王春平,付强,寇人可,吴巍屹,刘天勇.语义分割评价指标和评价方法综述[J/OL].计算机工程与应用:1-14[2023-02-04].<http://kns.cnki.net/kcms/detail/11.2127.TP.20221205.1515.002.html>
- [2] 何家峰,陈宏伟,骆德汉.深度学习实时语义分割算法研究综述[J/OL].计算机工程与应用:1-25[2023-02-04].<http://kns.cnki.net/kcms/detail/11.2127.TP.20221228.1259.001.html>
- [3] MARDIA K V, HAINSWORTH T J.A spatial thresholding method for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1988, 10(6):919-927.
- [4] ADAMS R, BISCHOF L. Seeded region growing[J]. Retranslations on Pattern Analysis and Machine Intelligence,1994,16:641-647.
- [5] ADAMS R, BISCHOF L. Seeded region growing[J]. Retranslations on Pattern Analysis and Machine Intelligence,1994,16:641-647.
- [6] E. Shelhamer, J. Long and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 4, pp. 640-651, 1 April 2017, doi: 10.1109/TPAMI.2016.2572683.
- [7] V. Badrinarayanan, A. Kendall and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 12, pp. 2481-2495, 1 Dec. 2017, doi: 10.1109/TPAMI.2016.2644615.
- [8] Ronneberger, Olaf et al. "U-Net: Convolutional Networks for Biomedical Image Segmentation." *ArXiv* abs/1505.04597 (2015): n. pag.
- [9] Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Trans Pattern Anal Mach Intell. 2018 Apr;40(4):834-848. doi: 10.1109/TPAMI.2017.2699184. Epub 2017 Apr 27. PMID: 28463186.
- [10] Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Trans Pattern Anal Mach Intell. 2018 Apr;40(4):834-848. doi: 10.1109/TPAMI.2017.2699184. Epub 2017 Apr 27. PMID: 28463186.
- [11] Chen, Liang-Chieh et al. "Rethinking Atrous Convolution for Semantic Image Segmentation." *ArXiv* abs/1706.05587 (2017): n. pag.
- [12] H. Zhao, J. Shi, X. Qi, X. Wang and J. Jia, "Pyramid Scene Parsing Network," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 6230-6239, doi: 10.1109/CVPR.2017.660.
- [13] R. S. Sutton and A. G. Barto, *Reinforcement learning: an introduction*. MIT Press, 2018.
- [14] G. Shani, J. Pineau, and R. Kaplow, "A survey of point-based pomdp solvers," Autonomous Agents and Multi-Agent Systems, vol. 27, no. 1,pp. 1-51, 2013.
- [15] C. J. C. H. Watkins and P. Dayan, "Technical note q-learning," Mach.Learn., vol. 8, pp. 279-292, 1992.
- [16] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G.Bellemare, A. Graves,

- M. A. Riedmiller, A. Fidjeland, G. Ostrovski, and et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.
- [17] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," *Mach. Learn.*, vol. 8, pp. 229–256, 1992.
- [18] J. Schulman, S. Levine, P. Abbeel, M. Jordan, and P. Moritz, "Trust region policy optimization," in *International Conference on Machine Learning*, 2015, pp. 1889–1897.
- [19] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. A. Riedmiller, "Deterministic policy gradient algorithms," in *International Conference on Machine Learning*, vol. 32, 2014, pp. 387–395.
- [20] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," in *International Conference on Learning Representations*, 2016.
- [21] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," in *International Conference on Machine Learning*, 2016, pp. 1928–1937.
- [22] J. Wang, Z. Kurth-Nelson, H. Soyer, J. Z. Leibo, D. Tirumala, R. Munos, C. Blundell, D. Kumaran, and M. M. Botvinick, "Learning to reinforcement learn," in the 39th Annual Meeting of the Cognitive Science Society, CogSci 2017, London, UK, 16–29 July 2017. cognitivesciencesociety.org, 2017.
- [23] T. Haarnoja, A. Zhou, K. Hartikainen, G. Tucker, S. Ha, J. Tan, V. Kumar, H. Zhu, A. Gupta, P. Abbeel, and S. Levine, "Soft actorcritic algorithms and applications," *CoRR*, vol. abs/1812.05905, 2018.
- [24] Brostow, Gabriel J., Julien Fauqueur, and Roberto Cipolla. "Semantic object classes in video: A high-definition ground truth database." *Pattern Recognition Letters* 30.2 (2009): 88–97.
- [25] Geiger, Andreas, et al. "Vision meets robotics: The kitti dataset." *The International Journal of Robotics Research* 32.11 (2013): 1231–1237.
- [26] Lin, Tsung-Yi, et al. "Microsoft coco: Common objects in context." *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V* 13. Springer International Publishing, 2014.
- [27] Cordts, Marius, et al. "The cityscapes dataset for semantic urban scene understanding." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
- [28] Ros, German, et al. "The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
- [29] Neuhold, Gerhard, et al. "The mapillary vistas dataset for semantic understanding of street scenes." *Proceedings of the IEEE international conference on computer vision*. 2017.

- [30] Huang, Xinyu, et al. "The apolloscape open dataset for autonomous driving and its application." *IEEE transactions on pattern analysis and machine intelligence* 42.10 (2019): 2702-2719.
- [31] Paszke, Adam, et al. "Enet: A deep neural network architecture for real-time semantic segmentation." *arXiv preprint arXiv:1606.02147* (2016).
- [32] Romera, Eduardo, et al. "Erfnet: Efficient residual factorized convnet for real-time semantic segmentation." *IEEE Transactions on Intelligent Transportation Systems* 19.1 (2017): 263-272.
- [33] Han, Wei, et al. "Contextnet: Improving convolutional neural networks for automatic speech recognition with global context." *arXiv preprint arXiv:2005.03191* (2020).
- [34] Yu, Changqian, et al. "Bisenet: Bilateral segmentation network for real-time semantic segmentation." *Proceedings of the European conference on computer vision (ECCV)*. 2018.
- [35] Yu, Changqian, et al. "Bisenet v2: Bilateral network with guided aggregation for real-time semantic segmentation." *International Journal of Computer Vision* 129 (2021): 3051-3068.
- [36] Poudel, Rudra PK, Stephan Liwicki, and Roberto Cipolla. "Fast-scnn: Fast semantic segmentation network." *arXiv preprint arXiv:1902.04502* (2019).
- [37] Chen, Wuyang, et al. "Fasterseg: Searching for faster real-time semantic segmentation." *arXiv preprint arXiv:1912.10917* (2019).
- [38] Zhuang, Juntang, et al. "Shelfnet for fast semantic segmentation." *Proceedings of the IEEE/CVF international conference on computer vision workshops*. 2019.
- [39] Bouton, Maxime, et al. "Safe reinforcement learning with scene decomposition for navigating complex urban environments." *2019 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2019.
- [40] Wang, Pin, Ching-Yao Chan, and Arnaud de La Fortelle. "A reinforcement learning based approach for automated lane change maneuvers." *2018 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2018.
- [41] Chen, Xin, et al. "A learning model for personalized adaptive cruise control." *2017 IEEE intelligent vehicles symposium (IV)*. IEEE, 2017.
- [42] Wang, Pin, and Ching-Yao Chan. "Formulation of deep reinforcement learning architecture toward autonomous driving for on-ramp merge." *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2017.
- [43] Chae, Hyunmin, et al. "Autonomous braking system via deep reinforcement

- learning." 2017 IEEE 20th International conference on intelligent transportation systems (ITSC). IEEE, 2017.
- [44] Bouton, Maxime, et al. "Cooperation-aware reinforcement learning for merging in dense traffic." 2019 IEEE Intelligent Transportation Systems Conference (ITSC). IEEE, 2019.
 - [45] Tang, Yichuan. "Towards learning multi-agent negotiations via self-play." Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2019.
 - [46] Folkers, Andreas, Matthias Rick, and Christof Büskens. "Controlling an autonomous vehicle with deep reinforcement learning." 2019 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2019.
 - [47] Folkers, Andreas, Matthias Rick, and Christof Büskens. "Controlling an autonomous vehicle with deep reinforcement learning." 2019 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2019.
 - [48] Porav, Horia, and Paul Newman. "Imminent collision mitigation with reinforcement learning and vision." 2018 21st International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2018.
 - [49] Bernhard, Julian, et al. "Experience-based heuristic search: Robust motion planning with deep Q-learning." 2018 21st International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2018.
 - [50] Hart, Patrick, Leonard Rychly, and Alois Knoll. "Lane-merging using policy-based reinforcement learning and post-optimization." 2019 IEEE Intelligent Transportation Systems Conference (ITSC). IEEE, 2019.
 - [51] Alizadeh, Ali, et al. "Automated lane change decision making using deep reinforcement learning in dynamic and uncertain highway environment." 2019 IEEE intelligent transportation systems conference (ITSC). IEEE, 2019.
 - [52] Deshpande, Niranjana, and Anne Spalanzani. "Deep Reinforcement Learning based Vehicle Navigation amongst pedestrians using a Grid-based state representation." 2019 IEEE Intelligent Transportation Systems Conference (ITSC). IEEE, 2019.
 - [53] Tram, Tommy, et al. "Learning when to drive in intersections by combining reinforcement learning and model predictive control." 2019 IEEE Intelligent Transportation Systems Conference (ITSC). IEEE, 2019.
 - [54] Isele, David, et al. "Navigating occluded intersections with autonomous vehicles using deep reinforcement learning." 2018 IEEE international conference on

- robotics and automation (ICRA). IEEE, 2018.
- [55] Li, Changjian, and Krzysztof Czarnecki. "Urban driving with multi-objective deep reinforcement learning." arXiv preprint arXiv:1811.08586 (2018).
- [56] Ronecker, Max Peter, and Yuan Zhu. "Deep Q-network based decision making for autonomous driving." 2019 3rd international conference on robotics and automation sciences (ICRAS). IEEE, 2019.
- [57] Wolf, Peter, et al. "Adaptive behavior generation for autonomous driving using deep reinforcement learning with compact semantic states." 2018 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2018.
- [58] Mirchevska, Branka, et al. "High-level decision making for safe and reasonable autonomous lane changing using reinforcement learning." 2018 21st International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2018.
- [59] Yuan, Wei, et al. "Multi-reward architecture based reinforcement learning for highway driving policies." 2019 IEEE Intelligent Transportation Systems Conference (ITSC). IEEE, 2019.
- [60] Lee, Junho, and Jun Won Choi. "May i cut into your lane?: A policy network to learn interactive lane change behavior for autonomous driving." 2019 IEEE Intelligent Transportation Systems Conference (ITSC). IEEE, 2019.
- [61] You, Changxi, et al. "Highway traffic modeling and decision making for autonomous vehicle using reinforcement learning." 2018 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2018.
- [62] Wang, Long, et al. "A Q-learning foresighted approach to ego-efficient lane changes of connected and automated vehicles on freeways." 2019 IEEE Intelligent Transportation Systems Conference (ITSC). IEEE, 2019.
- [63] Hoel, Carl-Johan, Krister Wolff, and Leo Laine. "Automated speed and lane change decision making using deep reinforcement learning." 2018 21st International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2018.
- [64] Zhang, Yi, et al. "Human-like autonomous vehicle speed control by deep reinforcement learning with double Q-learning." 2018 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2018.
- [65] Liu, Dapeng, et al. "Learning faster to perform autonomous lane changes by constructing maneuvers from shielded semantic actions." 2019 IEEE Intelligent Transportation Systems Conference (ITSC). IEEE, 2019.
- [66] Min, Kyushik, Hayoung Kim, and Kunsoo Huh. "Deep distributional reinforcement learning based high-level driving policy determination." IEEE Transactions on

Intelligent Vehicles 4.3 (2019): 416-424.

- [67] Rezaee, Kasra, et al. "Multi-lane cruising using hierarchical planning and reinforcement learning." 2019 IEEE Intelligent Transportation Systems Conference (ITSC). IEEE, 2019.
- [68] Shi, Tianyu, et al. "Driving decision and control for automated lane change behavior based on deep reinforcement learning." 2019 IEEE intelligent transportation systems conference (ITSC). IEEE, 2019.
- [69] Chen, Jianyu, Zining Wang, and Masayoshi Tomizuka. "Deep hierarchical reinforcement learning for autonomous driving with distinct behaviors." 2018 IEEE intelligent vehicles symposium (IV). IEEE, 2018.
- [70] Chen, Yilun, et al. "Attention-based hierarchical deep reinforcement learning for lane change behaviors in autonomous driving." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2019.
- [71] Qiao, Zhiqian, et al. "Pomdp and hierarchical options mdp with continuous actions for autonomous driving at intersections." 2018 21st International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2018.
- [72] Paxton, Chris, et al. "Combining neural networks and tree search for task and motion planning in challenging environments." 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2017.