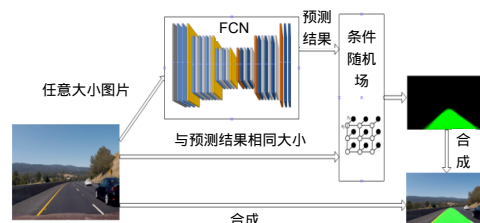




DOI: 10.12086/oe.2019.180274

基于全卷积神经网络与条件随机场的车道识别方法

叶子豪^{1,2}, 孙锐^{1,2*}, 王慧慧^{1,2}¹合肥工业大学计算机与信息学院, 安徽 合肥 230009;²工业安全与应急技术安徽省重点实验室, 安徽 合肥 230009

摘要: 本文针对传统车道识别方法在复杂路面中自适应能力差的特点, 基于图像分割技术提出了一种基于全卷积神经网络与条件随机场的车道识别方法。该方法通过大量数据的训练, 使神经网络模型可以识别出车道, 并且再通过条件随机场使得分割出来的车道覆盖面积及车道边缘的处理更加完善。同时, 本文为了解决高速公路中对检测实时性的高要求, 设计了一个全卷积神经网络, 该网络结构简单, 只有 13 万个参数, 并且做出如下三点改进: 采用 BN 算法提高网络的泛化能力及收敛速度; 采用了 LeakyReLU 激活函数取代了一般使用的 relu 或者 sigmoid 激活函数, 并且采用 Nadam 作为网络的优化器使得该网络具有更好的鲁棒性; 采用条件随机场作为后端处理解决车道边缘处分割不足并且加大了车道覆盖面积。最后本文为了解决城市道路检测中道路环境复杂的问题, 利用 FCN-16s 网络模型加条件随机场的后端处理实现了复杂城市道路的识别。实验证明, 在面对高速公路的高速及车道简单环境下, 本文设计的网络模型更具有实时性且足够胜任车道的识别。在面对城市道路的复杂环境下, FCN-16s 模型加条件随机场更能精确地识别出车道, 并在 KITTI 道路检测基准上取得不错的结果。

关键词: 车道检测; 全卷积神经网络; 条件随机场; 网络优化**中图分类号:** TP301.6**文献标志码:** A**引用格式:** 叶子豪, 孙锐, 王慧慧. 基于全卷积神经网络与条件随机场的车道识别方法[J]. 光电工程, 2019, 46(2): 180274

Lane recognition method based on fully convolution neural network and conditional random fields

Ye Zihao^{1,2}, Sun Rui^{1,2*}, Wang Huihui^{1,2}¹School of Computer and Information, Hefei University of Technology, Hefei, Anhui 230009, China;²Anhui Province Key Laboratory of Industry Safety and Emergency Technology, Hefei, Anhui 230009, China

Abstract: Aiming at the poor adaptability of traditional lane recognition method in complex pavement, this paper proposes a lane recognition method based on full convolutional neural network and conditional random field, according to image segmentation technology. The method can make the neural network model identify the lanes by training a large amount of data, and then make the segmentation of the lanes' coverage and the lane edges more perfect through the conditional random field. At the same time, in order to solve the high requirement of real-time

收稿日期: 2018-05-23; 收到修改稿日期: 2018-07-12

基金项目: 国家自然科学基金项目(61471154); 安徽省科技攻关科技项目(170d0802181)

作者简介: 叶子豪(1993-), 男, 硕士研究生, 主要从事计算机视觉的研究。E-mail: 609833291@qq.com

通信作者: 孙锐(1976-), 男, 博士, 教授, 主要从事计算机视觉与机器学习的研究。E-mail: sunrui@hfut.edu.cn

detection in expressway, a fully convolution neural network is designed in this paper. The network structure is simple with only 130000 parameters and three improvements are made as follows: BN algorithm is used to improve network generalization ability and convergence rate; LeakyReLU activation function is used to replace the commonly used relu or sigmoid activation function, and using Nadam as the network optimizer makes the network have better robustness; Conditional random field is used as the back-end processing solution insufficient lane segmentation and further to increase lane coverage. Finally, in order to solve the problem of complex road environment in urban road testing, this paper uses the back-end processing of FCN-16s network model and conditional random field to realize the recognition of complex urban roads. Experiments show that the network model designed in this paper is more real-time and sufficient for lane identification in the face of high-speed expressways and simple lanes. In the complex environment of urban road, FCN-16s model plus conditional random field can identify lane more accurately and get good results on KITTI road test benchmarks.

Keywords: lane detection; image segmentation; full convolution neural network; condition random field; network optimization

Citation: Ye Z H, Sun R, Wang H H. Lane recognition method based on fully convolution neural network and conditional random fields[J]. *Opto-Electronic Engineering*, 2019, 46(2): 180274

1 引言

近年来随着智能汽车的快速发展,自动驾驶引起了工业与学术界很大的关注。车道识别是实现自动驾驶的一项基础工作,具体而言,准确的车道识别不仅可以使车辆行驶在正确道路还可以提示控制系统的其他信息,如车道标记,行人和其他异常事件。传统的车道检测方法一般是通过预处理、边缘检测、霍夫变换、道线匹配、车道区域分割等环节,这些方法每个环节之间互相影响,很难做到整体最优,同时在实时性以及车道变化适应方面明显不足。深度神经网络是一种强大的视觉分析工具,对比传统的浅层计算结构,它的主要优势在于根据输入数据具有自学习的能力,并且具有端到端的统一结构。因此,本文将车道识别视为一种图像分割任务,设计了一种新颖的全卷积神经网络,实现了稳定高效的车道识别。

现阶段主流的车道检测主要是基于车道线的检测,采用的技术大多都是在霍夫变换的基础上所做的改进^[1],然而传统的霍夫变换并不能达到实时性的要求,同时 Canny 算子中高斯滤波、极大值抑制等过程耗时都比较大^[2]。部分算法使用车道模型来匹配边缘检测的结果,但是都不能在实时性与准确性之间达到较好的平衡^[3-5]。近几年随着深度学习的广泛应用,国内外提出了一些结合深度学习的车道检测方法,例如 ANN(2D/3D artificial neural network)^[6]算法提出了一种基于图像分割的道路检测方法,这种分割是将二维和三维图像处理数据从立体视觉系统中合并而成,最后道路模式识别是由一个人工神经网络进行的,它经

过训练以获得该特征描述符的最终结果,但是这种算法准确率并不高且运行时间达到了 3 s。BM(binary map from color-based road detection)^[9]算法将一个图像转换成一个对数色度空间并提供标识,然后样本像素从假定的“道路”区域随机选择,并且来对图像的像素进行分类,最后基于立体视觉对三维道路轮廓估计和增强检测精度。但 BM^[9]算法的准确性和实时性都只能达到一般的效果。FCN-LC(fully convolutional neural network with large context)^[7]算法采用一个全卷积神经网络,实时性与准确性均达到可观的效果,但是整体效果仍然存在些不足。

本文针对目前车道或车道线检测中对实时性的高要求以及一般深度学习方法在边缘处检测存在缺陷的问题,从图像分割的角度出发,采用了全卷积神经网络。它并不同于传统的卷积神经网络(convolutional neural network, CNN),而是将全连接层换成了反卷积层,从而可以实现端到端输入输出图像的目的。同时本文对高速公路设计的网络做出了些改进:采用了 LeakyReLU 函数取代传统的激活函数,解决了如 relu 函数存在的“神经元死亡”问题;采用 Nadam 优化器使得对网络的学习率约束更强;采用 BN 算法提高网络的泛化能力;并且为了解决单一的神经网络在道路识别中存在的不足,在后端加入了条件随机场(condition random field, CRF)的处理。实验证明,全卷积神经网络在图像分割上具有很好的效果,可以很好地将车道从图片或者视频中分割出来,而 CRF 在道路边缘处以及车道覆盖面积上起到很好的优化效果。且最后使用 FCN-16s^[9]架构和 CRF 后处理来解决城市道

路下环境复杂的问题，并在 KITTI 数据集上取得良好的效果。

2 全卷积神经网络

2.1 基本模型

全卷积神经网络(Fully convolutional neural network, FCN)是由 Long J, Shelhamer E, Darrell T^[8]提出的一种训练端到端的卷积网络结构。他们建立一个“完全卷积”的神经网络，利用任意大小的输入图像，通过有效的推理得到相对应的输出。即传统的卷积神经网络使用一个 CNN 进行图片的分类，而 FCN 则采取使用一个 CNN 来完成物体检测的任务。同时利用位置相关的特征池化来解决图片分类中兼容形变以及物体检测中利用形变的平衡问题。整个学习和推理的过程都是通过密集的前馈计算和反向传播在图像中进行。网络中的上采样层可以进行像素化预测，并在网络中学习。Long J 等人定义了一个新的“跳跃”架构，将深层、粗糙、语义信息和浅层结合起来。如图 1 所示，全卷积网络可以有效地学习对每个像素任务(如语义分割)进行密集预测。

卷积神经网络中的每一层数据都是三维的大小为 $h \times w \times d$ 的数组，其中 h 和 w 是空间维度， d 是特征或通道维度。首先图层是图像，像素大小为 $h \times w$ ， d 个颜色通道。在网络中决定某一层输出结果中一个元素所对应的输出层的区域大小称为感受野。卷积是建立在平移不变性上的。它们的基本组件(卷积、合并和激活函数)在本地输入区域上运行，并且仅依赖于相对空间坐标。在特定层位于 (i, j) 的位置中为数据向量写入 x_{ij} ，通过下式函数计算下一层中的 y_{ij} ：

$$y_{ij} = f_{k_s}(\{x_{si+\Delta i, sj+\Delta j}\}_{0 \leq \Delta i, \Delta j \leq k}) \quad (1)$$

其中： k 称为内核大小； s 是步幅或子采样因子； f_{k_s} 决

定图层类型：卷积运算或平均池化，一个空间最大池化的最大值，或者一个元素的非线性激活功能等其他类型的图层。这种功能形式是在组成下维护，内核大小和步幅遵循如下的转换：

$$f_{k_s} \circ g_{k's'} = (f \circ g)_{k'+(k-1)s', s's'} \quad (2)$$

其中： g 为特征矩阵， k' 为内核大小， s' 为步幅或子采样因子。一个 FCN 的输入可以是任意大小，并产生相应的输出(可能重新采样)空间维度。由 FCN 组成的实值损失函数定义了一个任务。如果损失函数是一个总和 P 在空间上 $\ell(x; \theta) = \sum_{ij} \ell'(x_{ij}; \theta)$ ，其最终层的尺寸梯度将是每个空间分量的梯度的总和。因此，在整个图像上计算的随机梯度下降将与在 0 上的随机梯度下降相同，将所有最终图层接受场作为小批次。当这些接受的领域重叠，两者前馈计算和反向传播在整个层级上逐层计算时效率更高。

在实际场景中，尤其在城市道路环境中，车道上有大量的车辆与行人等障碍物，简单车道线检测并不能将真正的可行路段标识出来。而在很多道路环境下，如山路地区和部分城市道路地区 往往并没有车道线，这就使无人驾驶系统仅仅只能通过行人检测和车辆检测来判断可行路段，这无疑是危险的。所以本文利用图像分割的方法来解决这一实际问题，即无论是什么样的车道类型以及道路环境，都只将可行的道路与其他障碍物和非车道分割出来提供给无人驾驶系统，更加符合实际的需求。而全卷积神经网络在图像分割中有着很大的精度与实时性优势，可以很好地处理实际应用中复杂的道路场景。

2.2 FCN-16s 网络模型

FCN-16s^[9]是一种 FCN 架构模型，它的原理是将 VGG16^[11]模型的全连接层换成卷积层，训练一个端到端的模型结构，VGG16 的模型图如图 2 所示。利用

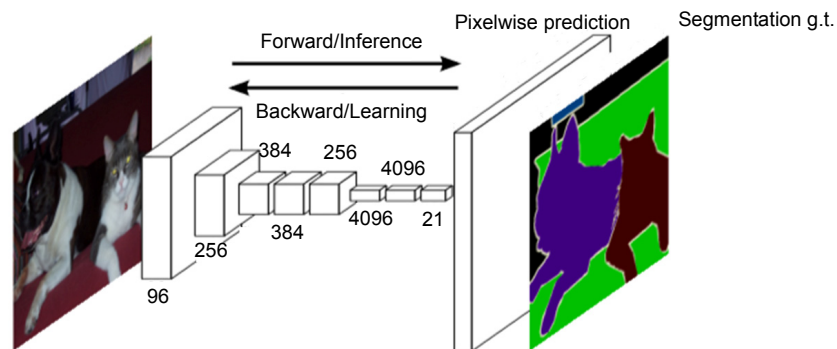


图 1 全卷积神经网络原理图

Fig. 1 Schematic diagram of the full convolutional neural network

VGG16^[11]强大的学习能力来得到准确的结果。而传统的 CNN 的方法需要对输入输出做出一些处理才能够得到需要的结果。而 VGG16^[11]和 AlexNet^[12], GoogLeNet^[13]都是性能比较好的 CNN 模型。对比其他两个模型, VGG16^[11]对图片有更精确的估值以及更省空间, 因此本文选择 VGG16^[11]。而这些 CNN 网络, 如 AlexNet^[12], VGG16^[11], GoogLeNet^[13], 只需在末尾加上上采样层, 参数的学习还是利用 CNN 本身的反向传播原理。其中 VGG16^[11]模型共有 16 层, 其总共参数达到了 138342976 个, 所有的内存加起来大约 24 M, 计算量为 15473144320 次。本文也对比过 FCN-8s^[9]和 FCN-32s^[9]模型结构, 发现 FCN-16s^[9]在现有的一些工作当中是最佳选择。它强大的学习能力使得在复杂场景中的分割效果比一般神经网络具有更高的精度。

3 车道识别算法概述

3.1 算法总体流程

本文将全卷积神经网络作为前端处理, 大致从图片中分割出道路。原始图像作为神经网络的输入, 输出为分割出来的道路图。然后将神经网络的输出以及

原始图像作为后端条件随机场 (conditional random fields, CRF) 的输入, CRF 将会根据神经网络的分割结果与原始图像, 从而输出一个优化后的道路图。最后将优化完成的道路分割图与原始图像叠加 (利用 opencv 中的 addWeight 函数) 达到车道识别的效果, 本文的模型结构亦可以输入为视频 (利用 moviepy), 而输出也将是视频。总体的流程图如图 3 所示。

3.2 网络结构与网络优化

本文采用的神经网络为全卷积神经网络, 它不同于传统的卷积神经网络 (CNN), 它将作为分类的全连接层全部换为卷积层, 这样输出的结果同样是图片, 达到端到端的目的。利用这一点可以实现图像分割的目的, 将车道从图片中分割出来。本文设计的全卷积神经网络由 32 层结构, 包括卷积层、池化层、反卷积层、dropout 层、BN 层、上采样层构成, 网络模型的输入图片为 80 pixels×160 pixels。传统的 FCN-16s 网络的输入图片大小大多设定为 500×500, 该网络的参数达到了 1.4 亿, 对于分割人, 动物, 自行车等事物有着不错的分割效果。但是传统的 FCN-16s 的参数设定适用于大量分割目标的复杂分割任务, 对于高速公路

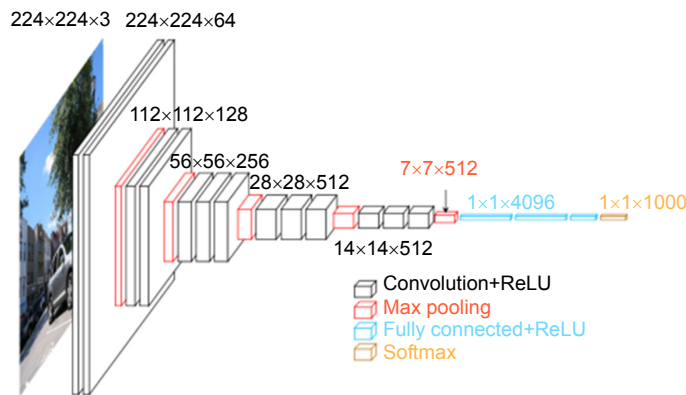


图 2 VGG16 模型图

Fig. 2 VGG16 model diagram

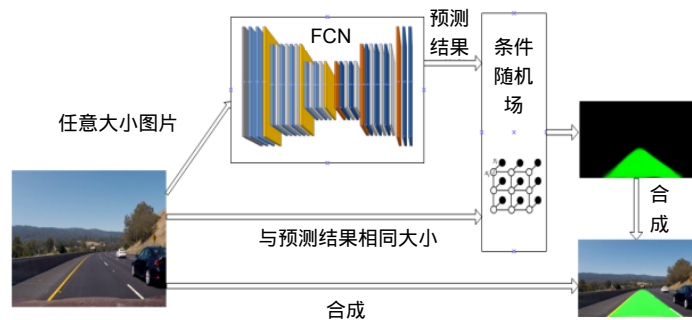


图 3 车道识别流程

Fig. 3 Lane recognition process

中分割车道这一简单场景显然过于复杂,而高速公路上高实时性的要求也更加注重于网络参数尽量少。经过大量实验,我们将网络的参数进行了合理的变化,与原网络相比突出在输入图片的大小以及各卷积层滤波器的数量,这样使该网络的参数只有 13 万,在保证高速公路道路识别的精度同时还具有极好的实时性,这对车辆时速 60 km/h 以上的行驶环境下尤为重要,同时该网络的训练时间也大大的减少。网络结构图如图 4 所示。

本文的网络结构在传统的全卷积神经网络结构上也做出改进。首先本文的网络结构不再使用传统的 sigmoid 激活函数或 relu 激活函数。原因是 sigmoid 函数存在很多不足,它容易饱和,当输入非常大或非常小的时候,神经元的梯度会接近于 0,这会使最终的神经网络中的权值没有多少变化,从而无法递归的学习数据集。而且,当初始值很大的时候,大多数神经元将是 saturation 状态从而影响梯度曲线,最终会使神经网络变得难以学习。另外 sigmoid 函数输出的并不是 0 均值的信号,导致后面层的神经元的输入不是 0 均值的信号,从而产生捆绑的效果,影响梯度,使其收敛缓慢。relu 函数会比 sigmoid 以及 tanh 函数拥有更快的收敛速度,且梯度不会饱和。因为 relu 仅仅通过一个阈值得出激活值,这使得它的计算要简单的多。然而 relu 函数也存在缺陷,它在神经网络训练的时候可能会出现神经元“坏死”的现象,也就是一个神经元之后的梯度一直是 0,不再对任何数据产生响应。

产生这一情况的原因是 relu 激活函数在输入信号小于 0 的时候输出 0。为了解决 relu 激活函数带来的神经元“坏死”问题,我们决定使用 LeakyReLU 函数,其数学表达式如下:

$$f(x) = \alpha x, \quad x < 0, \quad (3)$$

$$f(x) = x, \quad x \geq 0, \quad (4)$$

其中: α 为偏移量。LeakyReLU 函数与 relu 函数不同的是,当输入信号小于 0 时,它的值是一个拥有较小斜率的函数,这样就改变了数据的分布,保留负轴的一些值,从而使得负轴的信息不会全部丢失。

通常提高深度学习系统(不仅仅是 CNN)的方法有三种:第一种方法是改进模型结构,比如在原始的网络结构中再添加一层;第二种方法是改进模型的初始化,保证早期的梯度具有一定的有利特性,或者大规模建设稀疏性;第三种方法是使用一个更加强大的学习算法。本文从第三种方法的角度再次提高模型的性能。

大多数 CNN 的模型都采用梯度下降法(stochastic gradient descent, SGD)优化方法,但 SGD 存在一些不足。比如选取一个好的学习率不是很容易,如果想要使不常出现的特征更新快一些并且让那些常出现的特征更新慢一些,那么 SGD 很难达到这种想要的预期要求。并且 SGD 很容易收敛到局部最优,且有可能困在鞍点。另外 Momentum 和 Nesterov 优化方法可以说使梯度更新灵活一些,但是这些人工设置的学习率和那些自适应学习率的方法比起来显得生硬许多。Adam

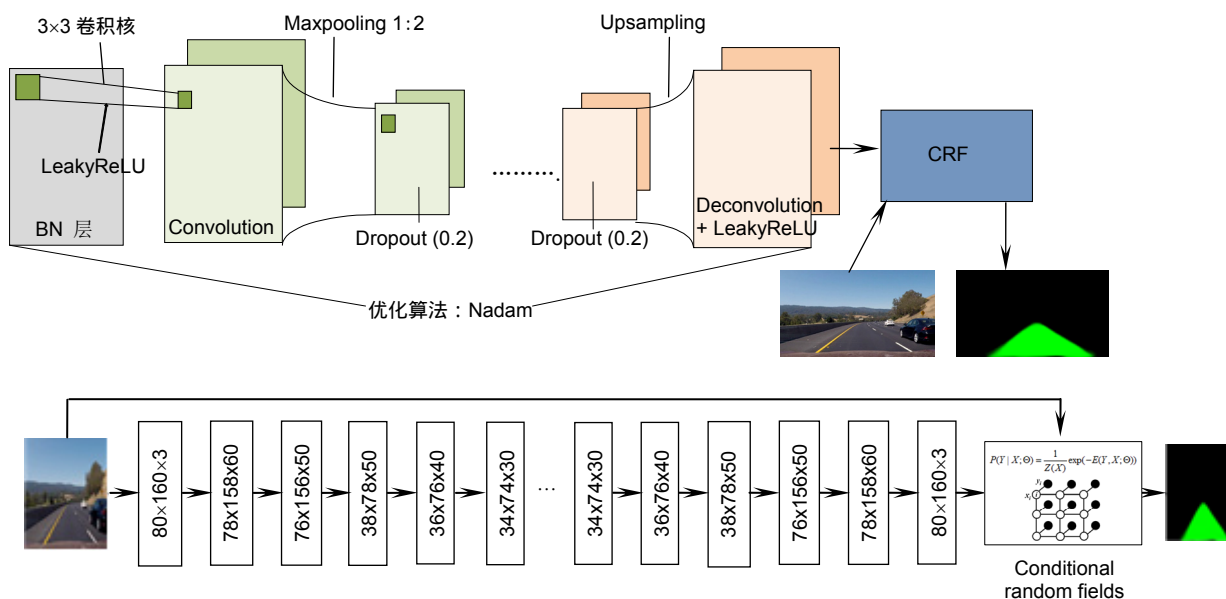


图 4 全卷积神经网络结构
Fig. 4 Full convolutional neural network structure

是一种带动量项的 RMSprop(一种自适应学习率的方法),通过梯度的一阶以及二阶矩阵估计来调试每一个参数的学习率。经过偏置校正后,每次迭代的学习率有一个确定的范围,这让参数较于平稳。Adam 拥有非常明显的优点,它对于处理稀疏密度以及非平稳目标都具有优势,同时它对内存的要求较小,对于每个参数都计算出自己的自适应学习率,在非凸优化、大数据集以及高位空间中也具有很好的效果。但相比于 Adam, Nadam 对于学习率的约束更强,并且它对梯度更新具有更加直接的影响。Nadam 类似于含有 Nesterov 动量项的 Adam。

这些年来,随机梯度下降法是训练神经网络的主流方法。虽然这种算法简单高效,但是依然存在着不足,比如它要求人工选择参数(学习率、参数初始化、权重衰减系数、Dropout 比例等)。这么多的参数都对训练出来的模型起到重要的影响,这使得会花费很多时间在调参上。于是本文使用 BN 算法(batch normalization)。它具有以下几个优势:

- 1) 具有快速收敛的特点,可以选择比较大的初始学习率,这会使训练速度明显加快(如果选取的学习率较小依然会加快);
- 2) 具有提高网络泛化能力的特点,完全可以移除 dropout、L2 正则项参数,或者使用小的 L2 正则约束参数;
- 3) 与 AlexNet^[12]网络不同,不需要局部响应归一化层,而 BN 层就是归一化网络层;
- 4) 可以把数据集完全打乱,这样可以当每批训练时,某些样本经常被挑选到。BN 算法的关键在于定义了两个可以学习的参数 γ 、 β :

$$y^{(k)} = \gamma^{(k)} \hat{x}^{(k)} + \beta^{(k)}. \quad (5)$$

所有的神经元都有这两个参数,当满足 $\gamma^{(k)} = \sqrt{\text{var}[x^{(k)}]}$, $\beta^{(k)} = E[x^{(k)}]$ 时,则可以恢复原始的某一层学习到的特征,其中: $\text{var}()$ 为风险价值函数, E 为数学期望。

随着深度学习的兴起,卷积神经网络(CNN)通过学习更具辨别性和更丰富的特征来提高图像理解能力,而很多新型的优化算法同时也在不断地改善网络的整体性能。FCN 是传统 CNN 的一种变体,在许多应用中取得了巨大的进步,特别是在物体检测和图像语义分割方面,而这些改进的算法与函数依然适用于以 CNN 为基础的 FCN 网络。FCN 网络关注的是对象是什么,但忽略了图像中的基本空间结构和位置信息。

鉴于此,本文将空间结构和位置先验信息引入到传统的 FCN 中。众所周知,图像的轮廓表示物体的重要边缘信息。给定道路的轮廓图像,就能够识别重要的物体及其边界,如道路,车辆等。

3.3 条件随机场

虽然神经网络可以完成道路识别的任务,但是实验发现单纯的神经网络的道路识别在道路的边缘处效果不是很好,并不能很好地覆盖边缘处的车道。于是本文在模型的后端加入了条件随机场。CRF 是一种判别式无向概率图模型,其本质上是一个给定的马尔科夫随机场的观察结果。当给出观察序列时,CRF 对于整个序列的联合概率有统一的指数模型。CRF 模型有很强的推理能力,并能够训练和进行复杂、重叠和非独立的推理。CRF 可以充分利用上下文信息并随意添加其他外部特征以获得丰富的模型信息。

设 G 是 (X, Y) 的因子图,如果对于 X 的任意值 x 都可以根据 G 分配一个 $p(x|y)$ 因式,那么 (X, Y) 是一个条件随机场。设 Y 是可以从 $\{y_1, y_2, \dots, y_n\}$ 取值的标签随机变量, X 是由随机变量 $\{x_1, x_2, \dots, x_n\}$ 形成的矢量,图像中的像素数量是 N 。给定一个无向图 $G=(V, E)$,其中 V 是图的节点集合,而 E 是图节点间边的集合,则可以表示 CRF 模型如下式所示:

$$P(Y|X; \Theta) = \frac{1}{Z(X)} \exp(-E(Y, X; \Theta)), \quad (6)$$

式中: $E(Y, X; \Theta)$ 是能量函数, $Z(X)$ 是配分函数 $Z(X) = \sum \exp(-E(Y, X; \Theta))$ 。能量函数 $E(Y, X; \Theta)$ 的公式如下:

$$E(Y, X; \Theta) = \sum_{p \in N} \Phi^{(1)}(y^p, x; \theta) + \sum \Phi^{(2)}(y^p, y^q, x, \theta), \quad (7)$$

其中: $\Phi^{(1)}$ 是一元能量分量, $\Phi^{(2)}$ 是成对能量分量, p 为相应节点, θ 为 y 与 x 服从的分布参数。在本文的方法中,一元能量分量基于 FCN 的单个像素的特征,是像素与一元能量之间的一对一关系,而成对的能量分量不仅基于本身,相邻像素包括八个像素,四个相邻边和四个相邻节点。但是由于复杂性和计算可行性,并不是所有八个相邻像素都被建模为成对的能量分量。本文中,做四个相邻的边缘像素的联合模型降低计算复杂度并尽可能少地减小影响。最后,本文尽量减少能量函数 E , 通过最大似然估计(maximum likelihood estimate)为给定图像产生最可能的标签分配 X 。一元能量函数如下所示:

$$\Phi^{(1)}(y^p, x) = \lambda_m I(y, y^p) x(l), \quad (8)$$

其中: I 为 0 或 1 的指示函数, $x(l)$ 是特征向量 x 的第

l 输入, λ_m 是状态函数 $s_m = I(y, y^p)x(l)$ 的参数。成对能量函数的定义如下:

$$\Phi^{(2)}(y^p, y^q, x; W) = \mu_m I(y, y^p)I(y', y^q), \quad (9)$$

其中: μ_m 是转换函数 $t_m = I(y, y^p)I(y', y^q)$ 的参数, q 为相应节点, W 为 y 与 x 服从的分布参数。成对能量函数就是两个像素之间的关系。潜在的成对特征映射如下式:

$$\sum_{(p,q) \in S} \Phi^{(2)}(y^p, y^q, x, \theta) = \sum_{i=1}^4 \sum_{(p,q) \in S_i} \Phi_i^{(2)}(y^p, y^q, x, \theta), \quad (10)$$

其中: S_1, S_2, S_3, S_4 是相邻的空间关系, 分别代表“上, 下, 左, 右”。

大量的实验证明, 单纯的全卷积神经网络在解决车道的分割任务时存在着边缘处分割不足的情况, 这不仅在车道线的边缘, 车辆行人等障碍物边缘的处理也依然存在着不足, 这就对实际无人驾驶系统中带来了一定的安全隐患。为了能够使整个网络在分割任务中更加精准, 让驾驶车辆不至于接触到障碍物的边缘, 故本文在后端加入了条件随机场, 利用条件随机场的特性来解决原网络边缘处分割不足的情况。

4 实验结果与分析

4.1 数据集

采用智能手机并以视频的形式采集数据, 视频的像素为 $720 \text{ pixels} \times 1280 \text{ pixels}$, 30 f/s 。在训练模型的时候考虑到时间的问题, 将训练图像缩小到 $80 \text{ pixels} \times 160 \text{ pixels}$ 。为了使我们最终训练出来的模型可以适应更多的道路环境, 采集的视频数据包括各种环境下的车道(雨天、雪天、夜里、晴天), 并且除了天气因素以外, 还包括了一般交通环境以及道路弯曲的环境。本文也从 Udacity 的高级车道项目中提取了简单的项目视频 1252 f 。最后将收集到的道路图像和相应标签的水平翻转使总图像增加了一倍。总共训练的图片数据约 8000 张用于网络训练。

为了可以跟国内外的车道识别技术做对比, 本文采用公开的 KITTI 车道数据集, 由于 KITTI 的车道数据集数量仅为 289 张, 故用到了迁移学习解决训练样本不足的问题。同时使用 FCN-16s^[9]结构代替之前的网络结构而后端处理不变, 并且与其他算法做比较, 包括 ANN^[6], DDN^[10], BM^[9], FCN-LC^[7]。其中本文从图片直观的视角以及 KITTI 中的 7 项指标^[17]来对比。

4.2 模型训练

本文的模型使用 python 开发, 采用的框架是 Keras, 且 Keras 的后台为 Tensorflow GPU 版本。硬件部分, 采用 intel(R)Core(TM)i7-6700K CPU@4.00 GHz×8 的处理器, 内存为 16 G, 显卡使用 GeForce GTX 1080/PCIe/SSE2。算法参数为: 批尺寸为 50, 迭代次数为 10000 次, 池化层大小为(2, 2), 激活函数使用 LeakyReLU, 其中 LeakyReLU 函数中第三象限线段的斜率 α 为 0.3, 优化器使用 Nadam, Nadam 的学习率等参数均为默认参数, 损失函数为 mean_squared_error, 并在网络中间的卷积层和反卷积层中加入 Dropout 层, Dropout 层中需要断开的神经元比列为 0.2。后端 CRF 依然使用 python 开发, 主要使用到的三方库为 pydensecrf。最后车道图与原图结合是使用 openCV, 该部分也是使用 python 开发, 主要使用的三方库为 CV2 即 openCV。

4.3 实验结果分析

在高速公路下, 本文将单一的卷积神经网络与加上条件随机场的模型结果进行对比, 如图 5 所示, 可以看到单一的神经网络的检测结果在边缘处有着明显的缺陷, 且整个道路的覆盖面积要明显小于加上条件随机场的检测结果。而卷积神经网络加上条件随机场的模型, 可以检测到更远的车道, 甚至超过了真实场景下标注的车道。另外该模型在损坏的高速公路以及

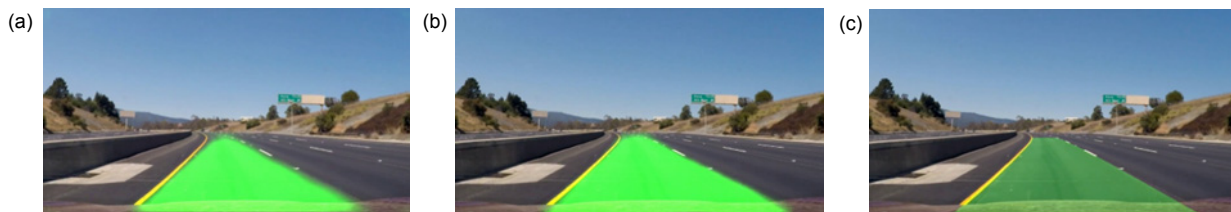


图 5 直线公路对比。(a) FCN; (b) FCN+CRF; (c) 真实场景
Fig. 5 Straight road comparison. (a) FCN; (b) FCN+CRF; (c) Real scene

弯道下依然有着良好的检测效果,如图 6 与图 7 所示。

表 1 为车道线检测方法,如基于线的检测,以及基于神经网络的 ReNet^[20]方法,基于条件随机场的 DenseCRF^[21]方法。这些方法分别利用到了传统的图像处理技术,深度学习技术,机器学习技术。但单一的使用可以发现其在 FP(false positive)指标中并没有很好的精度优势。并且车道线检测并不能分离出车道上的障碍物。

如图 8 为车道线数据集的标注信息。

最后本文使用 KITTI 道路数据集重新训练该模型,由于该数据集中训练图像只有 289 张,且道路环境十分复杂,原网络结构不足以分割出道路,所以这里利用 FCN-16s^[9]来取代本文之前的神经网络,而后端依然采用 CRF 做边缘处等地方的优化。本文将该模型的检测结果与 FCN-LC^[7],ANN^[6],BM^[9]算法进行直观的比较。不难发现这些算法在边缘处的处理存在明显的缺陷,同时分割的结果受环境影响较大,并不一定将车道完全地分割出来,如图 9 与图 10 所示。



图 6 损坏公路对比。(a) FCN; (b) FCN+CRF; (c) 真实场景

Fig. 6 Corrupted road comparison. (a) FCN; (B) FCN+CRF; (c) Real scene

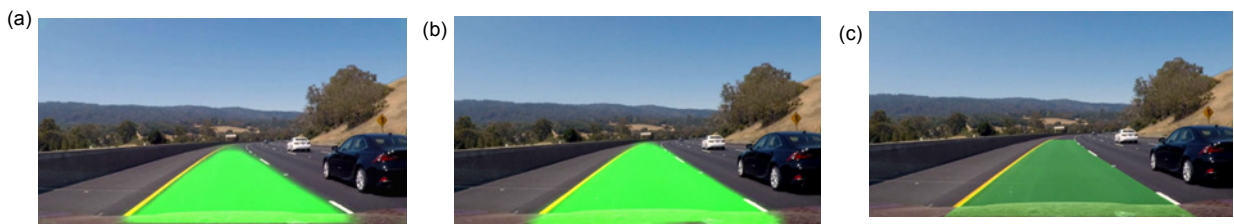


图 7 弯道公路对比。(a) FCN; (b) FCN+CRF; (c) 真实场景

Fig. 7 Curve road comparison. (a) FCN; (B) FCN+CRF; (c) Real scene

表 1 车道线检测方法的性能 FP 指标对比

Table 1 Comparison of performance of FP indicators among lane line detection methods

	Baseline	ReNet ^[20]	DenseCRF ^[21]
Normal	83.1	83.3	81.3
Crowded	61.0	60.5	58.8
Night	56.9	56.3	54.2
No line	34.0	34.5	31.9
Shadow	54.7	55.0	56.3
Arrow	74.0	74.1	71.2
Dazzle light	49.9	48.2	46.2
Curve	61.0	59.9	57.8
Crossroad	2060	2296	2253
Total	63.2	62.9	61.0

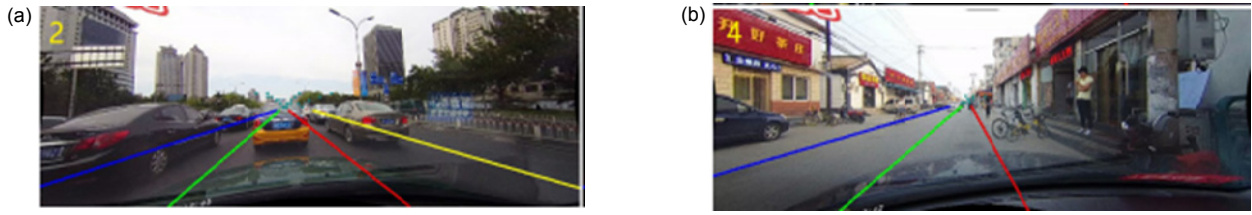


图 8 车道线数据集下的车道线标注
Fig. 8 Lane line labeling under the lane line data set

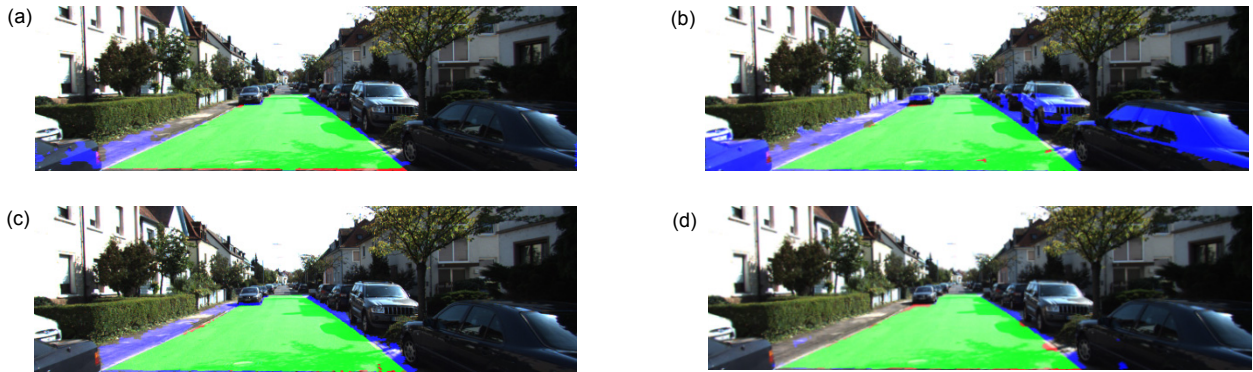


图 9 UU_ROAD_000020 检测对比。(a) FCN-LC; (b) ANN; (c) BM; (d) FCN+CRF
Fig. 9 UU_ROAD_000020 detection comparison. (a) FCN-LC; (b) ANN; (c) BM; (d) FCN+CRF

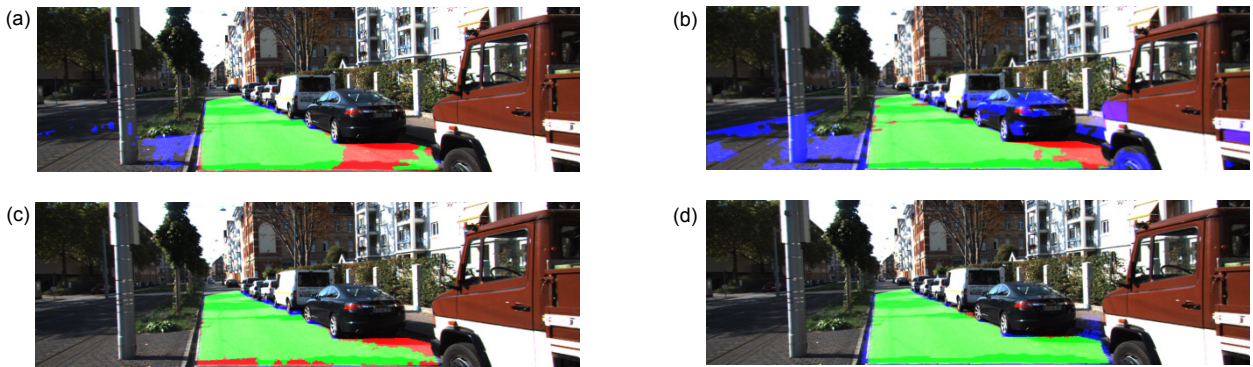


图 10 UU_ROAD_000082 检测对比。(a) FCN-LC; (b) ANN; (c) BM; (d) FCN+CRF
Fig. 10 Detection comparison of 10 UU_ROAD_000082. (a) FCN-LC; (b) ANN; (c) BM; (d) FCN+CRF

关于 KITTI 道路数据集，它有 579 个图像(289 个训练图像，290 个测试图像)组成，图像的分辨率为 375×1242，数据集的构成如表 2 所示，其中 UU 为未标记的城市道路，UM 为城市双向道路，UMM 为城市多车道道路。该数据集的评价基准根据鸟眼视图 (bird eye view)上的 max F-measure 对所有提交的方法进行排序。该基准测试为每个场景提供彩色立体图、GPS 和激光雷达数据。对于这个数据集，本文只利用单眼颜色数据来检测实验中的道路区域。

表 2 KITTI 数据集信息
Table 2 KITTI data set information

场景分类	训练集	测试集
UU	98	100
UM	95	96
UMM	96	94
城市道路总数	289	290

本文将 FCN 模型与 FCN+CRF(本文模型)的模型结果做出量化对比, 将 MaxF(maximum F1-measure), PRE(precision), REC(recall)三项指标进行对比, 并且测试的道路为 UU_ROAD, 可以发现 FCN+CRF 在各项指标上均有所改进, 如表 3 所示。

本文将该模型结构与其他包括深度学习算法以及非深度学习算法的车道检测做量化比较, 如表 4 所示, 将 FCN-LC^[7], ANN^[6], BM^[9], DDN^[10], FCN+CRF 算法通过 MaxF, AP, PRE, REC, FPR, FNR, runtime 这 7 项指标做量化对比。对比的道路是 UU_ROAD 测

试集下的结果, FCN-LC^[7]方法基于 FCN 网络做出了改进, 但是由于单网络性能上的局限使得精度上缺乏优势。DDN^[10]虽然精度上略优于本文算法, 但多达 2s 的运行时间(runtime)使得其在实际应用中存在较大的问题。综合实时性与精度来看可以发现本文的算法相对其他的深度学习算法和非深度学习方法具有一定的优势。

最后本文将对 UMM_ROAD 测试集下的结果, 本文的算法在该测试集下的结果相比另外 3 个测试集效果最好, 对比表与召回率如下所示。

表 3 UU 测试集下 FCN-16s 与 FCN+CRF 算法对比

Table 3 Comparison between FCN-16s and FCN+CRF algorithms under UU test set

算法模型	MaxF/%	PRE/%	REC/%
FCN-16s ^[9]	82.03	80.64	83.41
FCN+CRF	89.87	88.93	90.11

表 4 UU 测试集下的算法对比

Table 4 Comparison among algorithms under the UU test set

算法模型	MaxF/%	AP/%	PRE/%	REC/%	FPR/%	FNR/%	Runtime/s
FCN-LC ^[7]	86.27	75.37	86.65	85.89	4.31	14.11	0.03
ANN ^[6]	54.07	36.61	39.28	86.69	43.67	13.31	3
BM ^[9]	78.43	62.46	70.87	87.80	11.76	12.20	2
DDN ^[10]	91.76	86.84	93.06	90.50	2.20	9.50	2
FCN+CRF	89.87	82.96	88.93	90.11	4.22	9.04	0.08

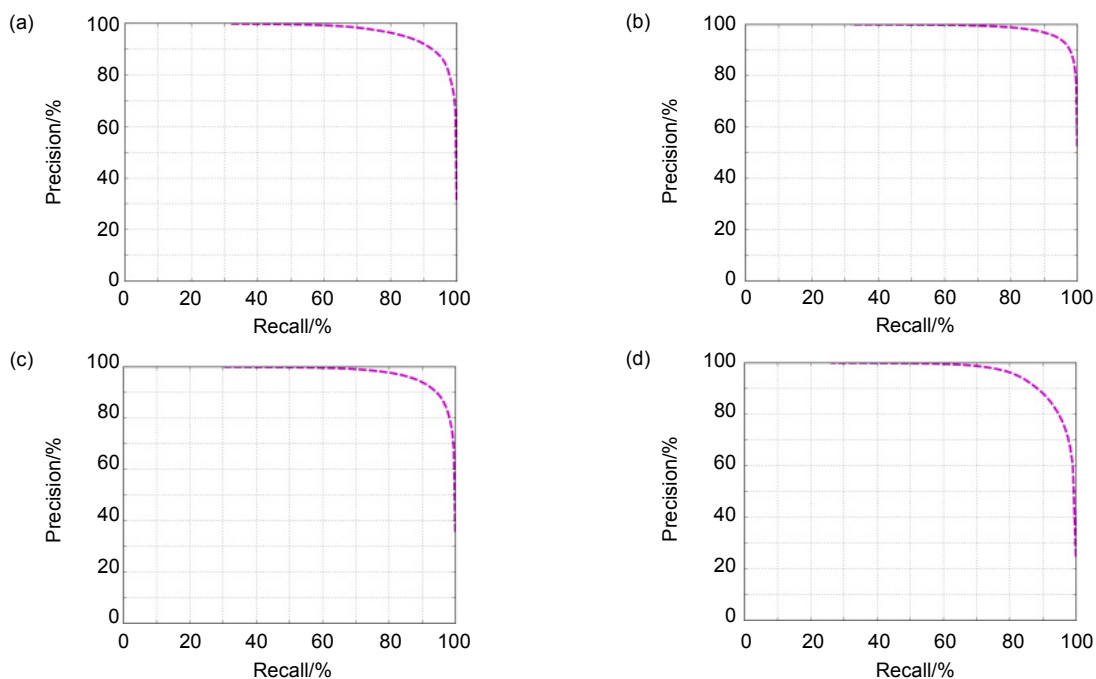


图 11 Kitti 不同测试集下召回率。(a) UM_ROAD; (b) UMM_ROAD; (c) URBAN_ROAD; (d) UU_ROAD

Fig. 11 Kitti recall rate under different test sets. (a) UM_ROAD; (b) UMM_ROAD; (c) URBAN_ROAD; (d) UU_ROAD

表 5 不同测试集结果对比

	MaxF/%	AP/%	PRE/%	REC/%	FPR/%	FNR/%
UM	90.23	86.92	90.99	90.11	3.92	9.27
UMM	94.28	90.87	95.14	93.98	4.25	4.99
UU	89.87	82.9	88.93	90.11	4.22	9.04
URBAN	91.49	86.67	93.10	91.22	4.21	8.04

5 结 论

本文从图像分割的角度出发,提出了一种新型的基于深度学习的车道检测方法,通过大量的数据训练以及全卷积神经网络与条件随机场的结合达到良好的车道检测效果。本文在高速公路的车道检测方面考虑到高速公路的环境简单以及对实时性要求高的两个特点从而设计了一个全卷积神经网络,该网络相对于其他 FCN 模型在满足精度的同时可以极大地改善实时性的问题。最后在城市道路的车道检测方面考虑到城市道路环境复杂以及数据集有限的情况,故使用 FCN-16s^[9]模型来代替之前的网络结构,后端依然采用条件随机场做后处理,同样达到了较好的检测效果。

参考文献

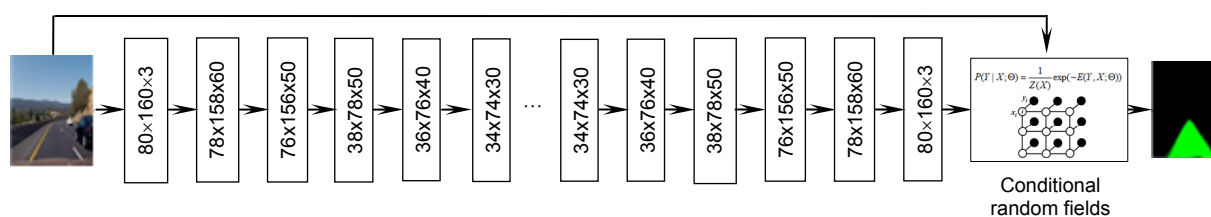
- [1] Li M, Huang H, Xia J G. Research of robust lane detection algorithm based on Hough transform[J]. *Computer Engineering And Design*, 2012, **33**(4):1638–1642.
李明, 黄华, 夏建刚. 基于 Hough 变换的车道检测改进算法研究[J]. *计算机工程与设计*, 2012, **33**(4):1638–1642.
- [2] Sharifi M, Fathy M, Mahmoudi M T. A classified and comparative study of edge detection algorithms[C]//*Proceedings of International Conference on Information Technology: Coding and Computing*, Las Vegas, NV, USA, 2002: 117.
- [3] Jung H G, Lee Y H, Kang H J, et al. Sensor fusion-based lane detection for LKS+ACC system[J]. *International Journal of Automotive Technology*, 2009, **10**(2): 219–228.
- [4] McCall J C, Trivedi M M. An integrated, robust approach to lane marking detection and lane tracking[C]//*IEEE Intelligent Vehicles Symposium*, Parma, Italy, 2004: 533–537.
- [5] 郭磊, 李国强, 王建强, 等. 应用方向可调滤波器的车道线识别方法[J]. *机械工程学报*, 2008, **44**(8): 214–218, 226.
- [6] Vitor G B, Lima D A, Victorino A C, et al. A 2D/3D vision based approach applied to road detection in urban environments[C]//*2013 IEEE Intelligent Vehicles Symposium*, Gold Coast, QLD, Australia, 2013: 952–957.
- [7] Mendes C C T, Frémont V, Wolf D F. Exploiting fully convolutional neural networks for fast road detection[C]//*2016 IEEE International Conference on Robotics and Automation*, Stockholm, Sweden, 2016.
- [8] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, **39**(4): 640–651.
- [9] Wang B, Frémont V, Rodriguez S A. Color-based road detection and its evaluation on the KITTI road benchmark[C]//*2014 IEEE Intelligent Vehicles Symposium Proceedings*, Dearborn, MI, USA, 2014: 31–36.
- [10] Mohan R. Deep deconvolutional networks for scene parsing[EB/OL]. <https://available: arxiv.org/abs/1411.4101>.
- [11] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. <https://available: arxiv.org/abs/1409.1556>.
- [12] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]//*Proceedings of the 25th International Conference on Neural Information Processing Systems*, Lake Tahoe, Nevada, 2012: 1097–1105.
- [13] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions[C]//*2015 IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, 2015: 1–9.
- [14] Brust C A, Sickert S, Simon M, et al. Convolutional patch networks with spatial prior for road detection and urban scene understanding[C]//*VISAPP 2015-10th International Conference on Computer Vision Theory and Applications*, Berlin, Germany, 2015: 510–517.
- [15] Mendes C C T, Frémont V, Wolf D F. Exploiting fully convolutional neural networks for fast road detection[C]//*2016 IEEE International Conference on Robotics and Automation*, Stockholm, Sweden, 2016.
- [16] Dinesh Reddy N, Prateek Singhal, Madhava Krishna K. Semantic Motion Segmentation Using Dense CRF Formulation[C]//*Indian Conference on Computer Vision, Graphics and Image Processing*, ICVGIP, Bangalore, India, 2014:a56.
- [17] Fritsch J, Kühnl T, Geiger A. A new performance measure and evaluation benchmark for road detection algorithms[C]//*16th International IEEE Conference on Intelligent Transportation Systems*, The Hague, Netherlands, 2013: 1693–1700.
- [18] Dan C C, Giusti A, Gambardella L M, et al. Deep neural networks segment neuronal membranes in electron microscopy images[J]. *Advances in Neural Information Processing Systems*, 2012, **25**: 2852–2860.
- [19] Dai J F, He K M, Sun J. Convolutional feature masking for joint object and stuff segmentation[C]//*2015 IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, 2014: 3992–4000.
- [20] Huval B, Wang T, Tandon S, et al. An empirical evaluation of deep learning on highway driving[EB/OL]. <https://arxiv.org/pdf/1504.01716.pdf>.
- [21] Krähenbühl P, Koltun V. Efficient inference in fully connected CRFs with Gaussian edge potentials[C]//*Proceedings of the 24th Neural Information Processing Systems*, Granada, Spain, 2011: 109–117.
- [22] Sun R, Fang W, Gao J. Person re-identification in foggy weather based on dark channel prior and metric learning[J]. *Opto-Electronic Engineering*, 2016, **43**(12): 142–146.
孙锐, 方蔚, 高隽. 暗通道和测度学习的雾天行人再识别[J]. *光电工程*, 2016, **43**(12): 142–146.

Lane recognition method based on fully convolution neural network and conditional random fields

Ye Zihao^{1,2}, Sun Rui^{1,2*}, Wang Huihui^{1,2}

¹School of Computer and Information, Hefei University of Technology, Hefei, Anhui 230009, China;

²Anhui Province Key Laboratory of Industry Safety and Emergency Technology, Hefei, Anhui 230009, China



Full convolutional neural network structure

Overview: In recent years, with the rapid development of smart cars, autonomous driving has attracted great attention from industry and academia. Lane recognition is a fundamental work for achieving automatic driving. Specifically, accurate lane recognition not only allows the vehicle to travel on the correct road but also alerts the control system with other information such as lane markings, pedestrians and other anomalous events. The traditional lane detection method generally adopts the steps of preprocessing, edge detection, Hough transform, lane matching, lane segmentation, etc. These methods interact with each other, and it is difficult to achieve global optimization and real-time optimization simultaneously. Furthermore, the lane change adaptation is obviously insufficient. Deep neural network is a powerful visual analysis tool. Compared with the traditional shallow computing structure, its main advantage lies in its self-learning ability according to input data and its end-to-end unified structure. Aiming at the poor adaptability of traditional lane recognition method in complex pavement, this paper proposes a lane recognition method based on full convolutional neural network and conditional random field, according to image segmentation technology. The method can make the neural network model identify the lanes by training a large amount of data, and then make the segmentation of the lanes' coverage and the lane edges more perfect through the conditional random field. At the same time, in order to solve the high requirement of real-time detection in expressway, a fully convolution neural network is designed in this paper. The network structure is simple with only 130000 parameters and three improvements are made as follows: BN algorithm is used to improve network generalization Ability and convergence rate; LeakyReLU activation function is used to replace the commonly used relu or sigmoid activation function, and using Nadam as the network optimizer makes the network have better robustness. Conditional random field is used as the back-end processing solution insufficient lane segmentation and further to increase lane coverage. Finally, in order to solve the problem of complex road environment in urban road testing, this paper uses the back-end processing of FCN-16s network model and conditional random field to realize the recognition of complex urban roads. Experiments show that the network model designed in this paper is more real-time and sufficient for lane identification in the face of high-speed expressways and simple lanes. In the complex environment of urban road, FCN-16s model plus conditional random field can identify lane more accurately and get good result on KITTI road test benchmarks.

Citation: Ye Z H, Sun R, Wang H H. Lane recognition method based on fully convolution neural network and conditional random fields[J]. *Opto-Electronic Engineering*, 2019, **46**(2): 180274

Supported by National Natural Science Foundation of China (61471154) and Anhui Science and Technology Research Project (170d0802181)

* E-mail: sunrui@hfut.edu.cn