

基于深度信息和神经网络的 道路图像分割方法

虞钉钉

(华院计算技术(上海)股份有限公司, 上海 201203)

摘要 道路图像分割是自动驾驶场景中的核心任务之一, 其分割效果对于自动驾驶的安全落地起到非常关键的作用; 而单目道路识别又是道路图像分割中十分重要的子任务。目前, 单目图像的道路识别方法主要以采用编码解码结构的网络模型为主, 其对道路的识别效果存在上限。在此背景下, 本文提出了一种基于深度信息的道路识别方法。实验结果表明, 在采用所提方法以后, 不同模型在道路数据集上的识别效果都得到了提升, 充分反映了所提方法的有效性和鲁棒性。

关键词 道路图像分割; 道路识别; 深度信息; 神经网络

中图分类号: TP751

文献标志码: A

DOI: 10.3969/j.issn.2097-3365.2026.04.004

0 引言

随着人工智能技术的飞速发展, 许多计算机领域的难点问题都迎来了新的突破。计算机视觉作为计算机领域的重要分支, 涌现出了很多新的方法。其中, 图像分割作为计算机视觉中的经典问题, 也受到了广泛的关注。图像分割任务本质是一个分类任务, 通过对图像的每个像素点进行分类, 来划分图像中的不同区域。

整体来说, 早期的图像分割技术依赖于规则和数学建模^[1], 通常适用于简单场景, 对于复杂场景的任务, 该技术存在一定瓶颈。随着深度学习的发展, 卷积神经网络(Convolutional Neural Network)被用于图像分割任务。深度神经网络通过使用大量精准标注的数据, 在该任务上取得了重大的突破, 并被应用于很多场景中, 比如自动驾驶中的道路、行人等分割、医学影像中的器官分割等。其中, 道路图像分割作为图像分割的子任务, 其相关技术被广泛应用于自动驾驶中, 是一个非常重要且具有应用价值的问题。

本文从道路图像分割任务的概念出发, 首先介绍道路图像分割任务的基础概念; 然后, 提出一种基于深度信息和神经网络的道路图像分割方法; 再在相关的道路数据集上进行实验, 并进行分析; 最后进行相应的总结和展望。

1 道路图像分割概述

道路图像分割是自动驾驶的核心任务, 旨在分割图像中的道路、车辆、行人等元素。其中, 道路识别

专注于提取图像中的道路区域。根据对道路区域的分类粒度不同, 该任务可进一步细化。例如: 可将道路分为对向多车道、同向多车道及背景三类; 也可将背景与对向车道合并, 仅区分同向车道。其核心目标通常是准确识别出本车所在的同向可行驶区域。此外, 不同任务对同向车道的处理也存在差异, 部分任务仅分割当前车道, 而不区分同向的其他车道。

具体来说, 本文专注于同向多车道的道路识别任务, 即识别出图像中的同向多车道区域。整个图像的像素点被分为两类: 非同向多车道区域和同向多车道区域。

2 基于深度信息的图像分割方法

为提高道路识别精度, 现有研究^[2]常借助多目道路图像或激光雷达数据辅助。多目图像通过多视角匹配可扩大感知范围并估计深度信息; 激光雷达提供的三维点云数据^[3]也能有效提升识别准确率。然而, 额外数据的获取往往成本较高。因此, 本文面向仅具备单目道路图像的同向多车道识别任务, 借鉴上述多源数据的应用思路, 提出从单目图像中估计深度信息^[4], 并以此为基础, 设计一种融合深度信息的道路图像分割方法, 旨在提升单目场景下的道路识别效果。

2.1 深度信息

深度信息是指像素点相对于采集设备的距离信息, 通常采用深度图(Depth Map)来标识。深度图是一张

作者简介: 虞钉钉(1994-), 男, 博士研究生, 工程师, 研究方向: 人工智能相关研究。

与原始图像分辨率相同的单通道图像, 每个像素点的值表示该位置的深度。深度信息目前被广泛应用于如自动驾驶、机器人导航、增强现实等各个场景中。具体来说, 在自动驾驶中, 深度信息通常被用于道路识别、障碍物距离估计等; 在机器人导航中, 深度信息被用于进行避障、获取定位等; 在增强现实中, 深度信息被用于虚实遮挡处理和虚拟物体位置定位等。

那么, 如何来获取深度信息呢? 除了采用设备直接采集以外, 也有很多基于其他数据的深度信息估计方法, 其中有一类就是基于单目图像的深度估计方法。当前主流的单目图像深度估计方法主要采用卷积神经网络或者 transformer 的结构, 然后采用编解码结构来作为整体模型的架构。同时, 主要采用的训练数据包括带真实深度标签的数据集, 以及不带标签的数据集, 其中带标签的数据集可以直接进行训练; 而不带标签的数据集可以通过多帧视频之间的一致性作为监督信号, 也可以采用已经训练好的教师模型预测的结果作为标签进行使用。本文将采用预训练的模型来进行单目深度估计以提供额外的深度信息。

2.2 总体框架

本节将重点介绍本文提出方法的整体框架和相关细节。其整体流程如下: 首先输入单目道路图像, 采用一个预训练的深度估计网络进行深度估计, 其中深度估计网络的整体结构可以分为两部分: 深度信息的编码器和深度信息的解码器; 然后将该深度估计的深度图和单目道路图像进行通道拼接, 再输入到一个分割网络, 最终输出相应的分割图。需特别注意的是, 该整个框架主要包含了深度信息的估计和使用的想法, 并不明确深度估计网络和分割网络。深度估计网络可以采用当前主流的深度信息估计方法, 本文采用一个使用编解码架构的预训练网络^[5], 该网络在大量的带标签和不带标签的数据上进行训练, 可以十分有效地进行单目图像的深度估计, 并且其泛化性能良好, 可以在没有训练过的数据集上进行比较精准的深度估计, 非常适合本文所研究的场景。类似的, 分割网络也可以采用主流的网络方法, 并不限制为某个具体网络。对于分割任务来说, 大部分分割网络都采用经典的编解码架构, 为此, 该框架中用该经典架构来表示分割网络。在训练中, 为了不破坏深度估计网络的性能, 会将整个深度估计网络的参数进行冻结, 只训练分割网络中的参数来得到最终的模型, 具体的相关介绍可以见下一节。

3 实验与结果分析

3.1 实验指标

评价分割模型效果的指标有很多, 本文采用比较经典的几个指标: 查准率(Precision)、召回率(Recall)和准确率(Accuracy)。在介绍这几个指标前, 先介绍一下分割任务中常用的混淆矩阵, 其核心是四个指标: 真阳(True Positive, TP)、真阴(True Negative, TN)、假阳(False Positive, FP)和假阴(False Negative, FN)。具体来说, 真阳表示样本为正类, 预测结果也是正类的数量; 真阴表示样本为负类, 预测结果也是负类的数量; 假阳表示样本为负类, 但是预测结果为正类的数量; 假阴表示样本为正类, 但是预测结果为负类的数量。下面分别给出查准率、召回率和准确率的定义和计算公式。查准率是衡量预测为正类的样本中, 有多少是真的正类, 其计算公式为:

$$\text{查准率} = \frac{TP}{TP+FP} \quad (1)$$

召回率是衡量所有为正类的样本, 被预测为正类的比例, 其计算公式为:

$$\text{召回率} = \frac{TP}{TP+FN} \quad (2)$$

准确率是衡量所有类别预测的正确率, 其计算公式为:

$$\text{准确率} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

以本次道路识别的实验为例, 道路类别就是正类别, 非道路类别就是负类别。

3.2 数据集和实验设置

实验的数据集采用道路分割中非常经典的KITTI数据集, 使用其中的左目图像道路数据集, 一共包含带标签的289张训练集以及290张不带标签的测试集。本次实验使用其中带标签的289张训练集。由于该289张图像尺寸大小存在微小差异, 比如图像宽度存在375和376不同的尺寸, 为此需要先进行预处理。首先筛选出长为1 242、宽为375的202张图像, 然后按约为5:1的比例, 随机抽取得到168张图像作为本次实验的训练数据, 剩下的34张图像作为本次实验的测试数据。

本次实验采用的深度估计网络为论文^[6]中的small模型, 需要注意的是, 该模型并没有在KITTI数据集上进行训练, 不过其在KITTI数据集上仍能取得不错的深度估计效果。分割网络采用了经典的Unet和SegNet两个不同的模型以验证所提框架的有效性。整体的训练采用的学习率为 $1e-4$, 并使用参数为0.5、0.999的Adam算法。在整个训练中, 冻结深度估计网络的参数, 只训练分割网络的参数。另外, 在测试中

模型输出的概率超过阈值 0.5 的被认为是正类,反之,则被认为是负类。实验采用的设备其CPU型号为2.90GHz的 Intel(R) Xeon(R) Gold 6226R CPU,其GPU型号为 Nvidia V100 显卡,显存为 32G。

3.3 实验结果和分析

本次实验的结果可见图1。图1展示了采用原始的 SegNet、Unet 以及采用了本文所述方法的 SegNet (w/depth) 和 Unet (w/depth) 的查准率、召回率和准确率。从实验结果来说,无论是 SegNet 还是 Unet,在增加深度信息以后,在查准率、召回率和准确率三个指标上都有了提升,这充分说明了本文所提方法的有效性和鲁棒性。具体来说,对于 SegNet 网络来说,其查准率提高

了约 0.4%,召回率提高了约 0.3%,准确率提高了约 0.2%;而对于 Unet 网络来说,其查准率提升了约 0.3%,召回率提升了约 0.7%,准确率提高了约 0.2%。另外,由图1可知,所有模型的准确率都比查准率和召回率高,这是因为数据集中负类也就是非道路类别的像素数据量要远远大于道路类别的像素数据量,因此导致了衡量所有类别正确率的准确率比只衡量正类的几个指标数值要高。其次,所有模型的查准率都比召回率要高,也反映了道路图像识别中要找出所有道路是比较困难的,这也与道路边缘区域的复杂性有关。整体来说,该实验说明了两个方面:(1)采用深度信息以后,模型在查准率、召回率和准确率上都有一定的提高,这

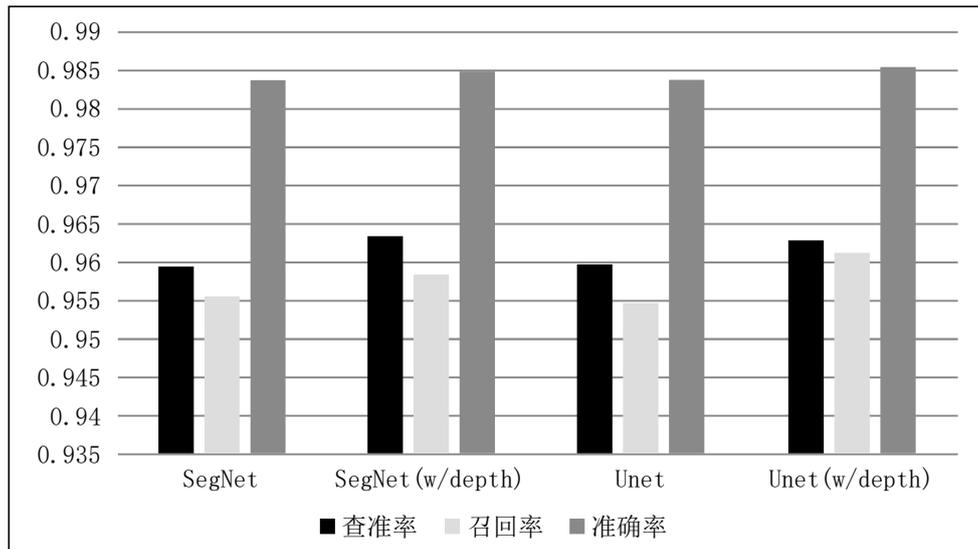


图1 实验对比结果图

表明了本文方法的有效性;(2)不同的模型在采用本文方法后,其结果都有提升,这在一定程度上表明了本文方法的鲁棒性。

4 结束语

本文围绕道路识别问题,提出了一种基于深度信息和神经网络的道路图像分割方法。该方法在只使用单目图像的情况下,通过引入额外网络来估计深度信息,并以此来提高整体的道路识别效果。实验结果表明,所提方法确实提高了不同模型在测试集上的查准率、召回率和准确率,展现了所提方法是有效且具有一定鲁棒性的。

参考文献:

[1] 张富瑜,王艳.图像分割中数学与深度学习融合建模

方法[J].数学建模及其应用,2025,14(03):1-9.

[2] 王龙飞,严春满.道路场景语义分割综述[J].激光与光电子学进展,2021,58(12):36-58.

[3] 王章宇,陈阳,周彬,等.一种面向非结构化道路的点云语义分割方法[J].北京航空航天大学学报,2025,51(02):457-465.

[4] 赵经阳,余昌黔,桑农.RGB-D语义分割:深度信息的选择使用[J].中国图象图形学报,2022,27(08):2473-2486.

[5] 杨丽鹤,康兵义,黄子龙,等.深度估计一切:释放大规模无标签数据的潜力[C]//IEEE计算机视觉与模式识别会议,2024.

[6] 同[5].