

机器学习工程师纳米学位开题报告

项目名称

侦测走神司机

项目背景

在驾驶过程中，如果司机的注意力不集中而分心去做其他的事情（比如打电话，看手机，与其他人交谈，伸手去拿后面的东西等），对于驾驶过程中出现的状况就不能做出及时反应，很可能就会引发交通事故。根据美国国家统计与分析中心 2015 年的报告[1]，美国有 15% 的严重交通事故是由于司机分心去做其他事情而造成的，这些事故共造成 3477 人死亡、约 391000 人受伤。所以通过一个辅助驾驶装置去检测司机的驾驶状态，当发现司机有分心做其他事情的行为时能及时提醒司机，这是非常有意义而且有必要的。

本项目是由 State Farm 在 Kaggle 上发起的一个竞赛项目，目的是能通过算法去识别司机的行为，判断司机是否在安全驾驶状态。这样，当司机做出非安全驾驶的行为时，就可以去提醒司机专心驾驶，从而在一定程度上减少交通事故。

问题描述

项目的要求是给定司机驾驶状态的 2 维图像，去识别图像中司机是处于哪种驾驶状态，这实际上是一个分类问题：用给定的带标签的训练数据（图像）训练模型，然后用测试数据集验证模型。需要识别的驾驶状态一共有如下 10 种：

- c0: 安全驾驶
- c1: 右手打字
- c2: 右手打电话
- c3: 左手打字
- c4: 左手打电话
- c5: 调收音机
- c6: 喝饮料
- c7: 拿后面的东西
- c8: 整理头发和化妆
- c9: 和其他乘客说话

输入数据

项目提供分辨率为 640x480、格式为 JPG 的图片，分为训练数据和测试数据，其中训练数据中的图片按照标签 c0~c9 分别存放在对应的文件夹中。图片是从视频中导出的，由摄像机从司机的右侧进行拍摄，这样可以清楚地拍摄到司机动作。这些图片大部分拍的比较清晰，有一小部分稍微有点模糊，光线大部分比较亮，有些则暗一点。每个司机在 c0~c9 中都有由一定数量的图片，这些司机有男的也有女的，有年轻的也有年老的，身材有胖的也有瘦的。

测试数据一共包含 79726 张图片，训练数据一共包含 22424 张图片，其中各类别包含的图片数量如下表所示：

c0	c1	c2	c3	c4	c5	c6	c7	c8	c9
2489	2267	2317	2346	2326	2312	2325	2002	1911	2119

训练数据提供 26 个司机的图片，每个司机的图片都有 c0~c9 共 10 种类别。在项目提供的 driver_imgs_list.csv 文件中，把每个司机都分配了一个 ID，统计了每个司机所有图片的文件名，以及这些图片的类别。这样，在做数据预处理时我们就可以根据 driver_imgs_list.csv 文件提供的信息，对训练数据集进行划分（分为训练集和验证集）。划分训练数据集时应该按照司机来划分，同一个司机的图片不能同时出现在训练集和验证集中，这是因为同一个类别中包含了同一个司机的若干张照片，这些照片是不能同时出现在训练集和验证集中的。

解决方案

本项目的目的是用训练数据集中的图片去训练一个模型，然后用该模型对测试集中的图片进行分类识别。近年来，卷积神经网络（CNN）在图像领域取得了巨大的成功，在大规模视觉识别竞赛（LSVRC）上，出现了一些非常经典的 CNN 模型，比如 AlexNet[2]、VGGNet[3]、GoogLeNet[4]、ResNet[5]等。这些模型在 LSVRC 上取得的成功，说明 CNN 非常适合用来做图像的分类识别，所以本项目也准备采用基于 CNN 的模型去做司机驾驶状态的分类。

如果自己从头搭建 CNN 模型训练的话，需要大量的训练数据才能训练好模型，本项目提供的 10 个类别的训练图片一共只有 22424 张，这是远远不够的。AlexNet、VGGNet、GoogLeNet、ResNet 这些经典的模型都是在 LSVRC 的数据集上得到充分训练的，我们可以利用这些训练好的模型，通过迁移学习的方式来解决本项目的问题。VGG16 模型结构简单，容易理解，可以先基于该模型来做迁移学习。

评估指标

根据项目要求，模型的评估指标为对数损失函数（multi-class logarithmic loss）：

$$\log loss = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{ij} \log(p_{ij})$$

其中 N 是测试集中图片的数量， M 是图片类别的数量， \log 是自然对数，如果输入的图片 i 属于类别 j 则 y_{ij} 为 1 否则为 0， p_{ij} 则是输入图片 i 属于类别 j 的概率。在测试集上模型的对数损失函数的数值越低，说明模型的表现越好。

基准模型

在 Kaggle 上，该项目的 Public Leaderboard 是用大约 31% 的测试数据去测试模型，Private Leaderboard 则是用大约 69% 的测试数据去测试模型，评估指标 $\log loss$ 的值越低，说明模型的表现越好，在 Leaderboard 上的排名也就越靠前。该项目的 Leaderboard 有 1440 组参赛者，做本项目的目标是进入 Private Leaderboard 的 top 10%，也就是 Private Leaderboard 上的 $\log loss$ 要低于 0.25634。

设计大纲

● 相关平台

用 Keras 来搭建和训练 CNN 模型，后端使用 TensorFlow，计算资源使用 AWS EC2 实例。

● 数据预处理

在进行训练之前，需要把训练数据划分为训练集和验证集，具体

的划分方法是：根据司机对数据集进行划分，一共 26 个司机，随机取 21 个司机的数据作为训练集，剩下 5 个司机的数据作为验证集。由于原始图片的分辨率是 640x480，而 VGG16 要求的输入是 224x224，所以还要对图片进行 resize 操作。另外，可能还要对数据进行一定的数据增强。

● 模型训练

首先尝试用 Keras 自带的 VGG16 模型，加载预训练的权重，把 VGG16 的 softmax 1000 分类器换成本项目需要的 softmax 10 分类器。训练的时候，先冻结最后的全连接层之前的层，对该全连接层进行训练，然后再从后往前逐层放开需要的层进行微调。除了用 VGG16,还可以尝试其他的模型，或将多个模型进行融合，这样可以对比用不同的方法所得到的结果。

参考文献

- [1]https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/812_381_distacteddriving2015.pdf
- [2] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]// International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012:1097-1105.
- [3] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.
- [4] C. Szegedy *et al.*, "Going deeper with convolutions," *2015 IEEE*

Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 1-9.

[5] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, United States, 2016, pp. 770-778.