

1. 项目背景

随着道路交通的发展及私家车辆的普及，道路交通事故已经成为了人们关注的社会问题。这些交通事故大部分都或多或少与司机的驾驶行为相关，例如疲劳驾驶、注意力分散等。根据美国国家公路交通安全交通安全署的统计，在美国的公路上，每年由于司机驾驶过程中进入睡眠状态而导致大约 10 万起交通事故，其中约有 1500 起直接导致死亡，7 万起事故导致人员受伤^[1]。根据 CDC 机动车安全部门的统计，每年由于司机开车时精力不集中造成的人员受伤和死亡人数分别为 425,000 人和 3,000 人^[2]。

司机的不良驾驶行为已经对自身和乘客的生命安全造成了及其严重的不良影响，解决该问题具有很强的必要性和紧迫性。

2. 问题描述

基于大量车内司机的 2D 图像，实现一种算法对每张图片中的司机行为进行分类，判定他们是否在专心驾驶，是否系好安全带，是否在驾驶过程中使用了手机，针对每一张图像算法最终输出车内司机可能性最高的行为类别。

3. 输入数据

该项目数据由 State Farm 公司提供，提供的数据包括摄像头采在车内集到的司机 2D 图像和每张图像对应的类别标签。图像的类别标签包括以下 10 种：

- c0: 安全驾驶
- c1: 右手发短信

- c2: 右手打电话
- c3: 左手发短信
- c4: 左手打电话
- c5: 操作收音机
- c6: 喝饮料
- c7: 向后方看
- c8: 整理头发和化妆
- c9: 与乘客交流

以上数据可以在 <https://www.kaggle.com/c/state-farm-distracted-driver-detection/data> 上进行下载，得到的三个文件为：

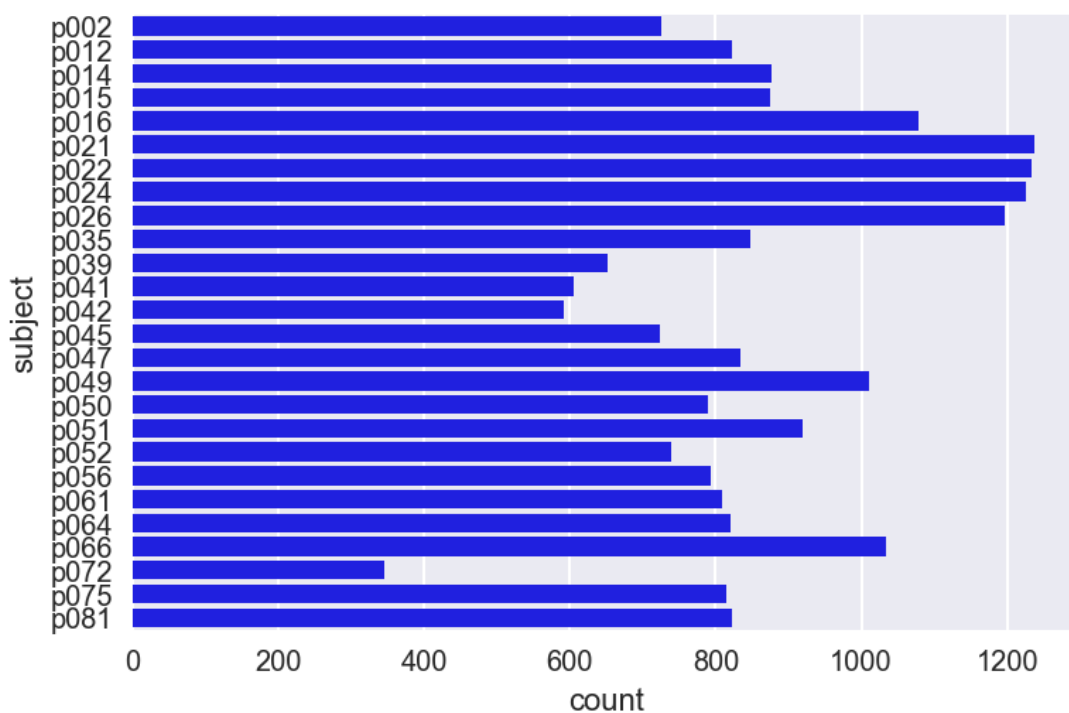
- imgs.zip: 所有图像（训练/测试）文件的压缩包文件
- sample_submission.csv: 正确地提交格式样本
- driver_imgs_list.csv 训练图像列表，司机 id 和类别 id

imgs.zip 文件大小为 4.29GB，解压后得到 train 和 test 两个文件夹，train 文件夹里的数据为训练集数据，共包含 22424 张图像，每张图像的大小为 640*480；test 文件夹里的数据为测试集数据，共包含 79726 张图像，每张图像大小为 640*480。由此可以看出测试集的数据量要多于训练集的数据量。

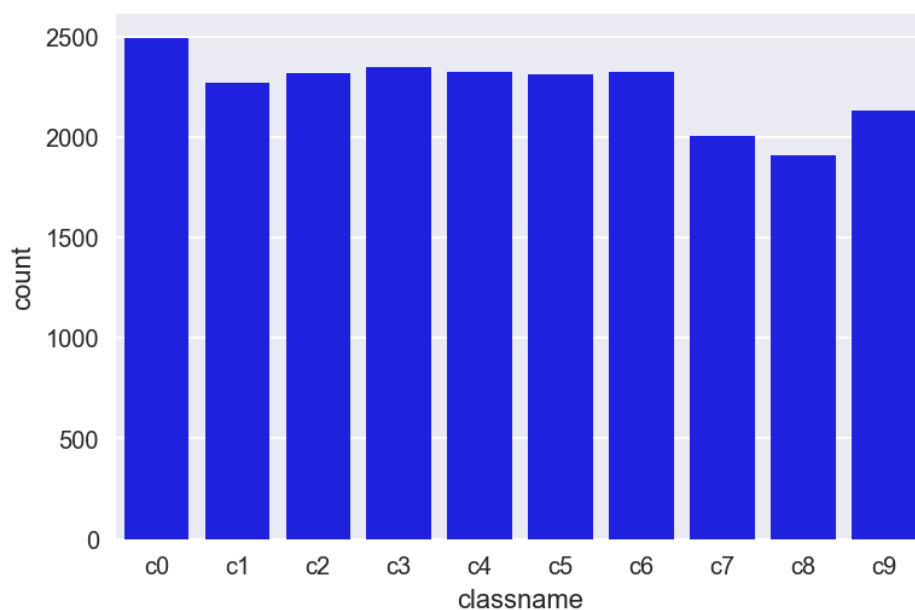
该问题的输入数据为下图所示的图像。



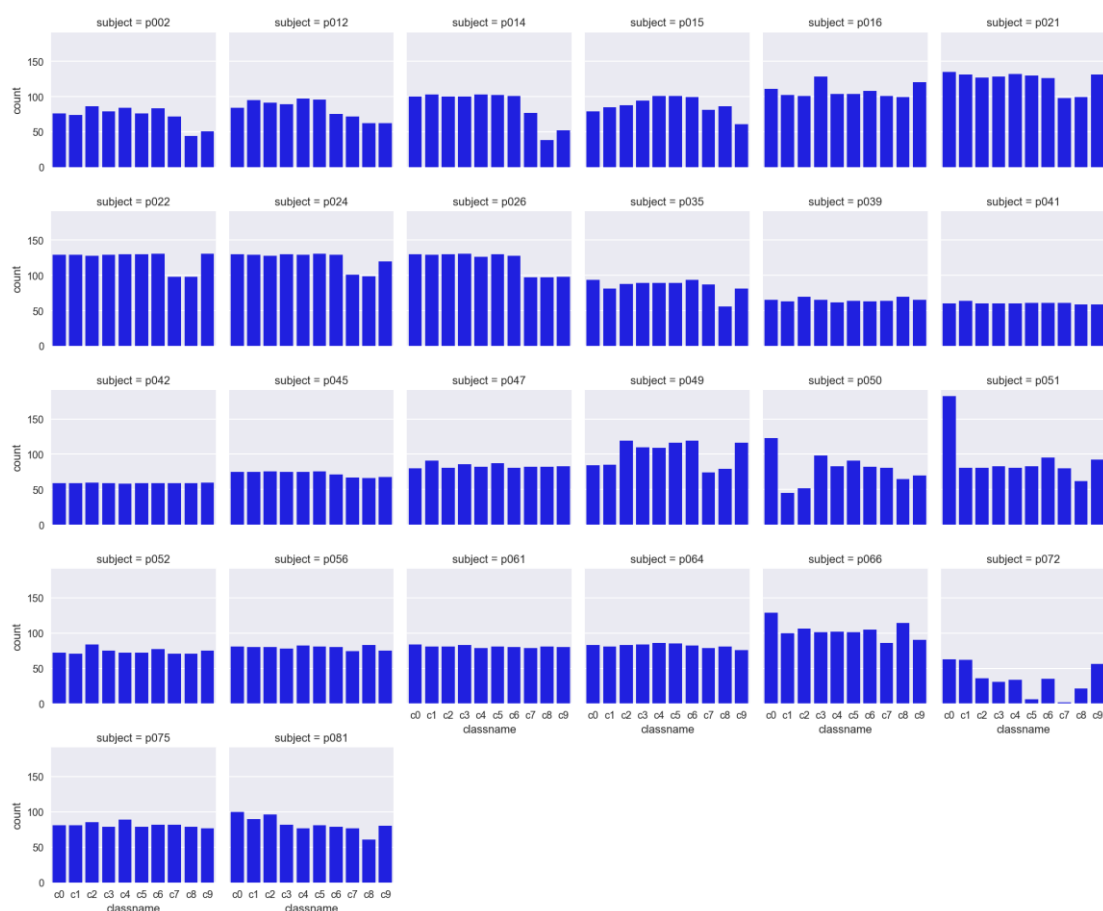
通过 `driver_imgs_list.csv` 文件分析得到训练集中共有 26 位司机，每位司机的图像总数有所不同，P021、P022、P024 的图像数较多超过 1200 张；而 P072 只拥有不到 400 张图像。从下图可以看出，每位司机的图像数量总体来说是平衡的。



通过下面的直方图可以看出每种类别的图像个数相差不大，说明整个数据集在每个类别上是平衡的。



下图展示的是每个司机在每个类别上的图像数量，可以看出每个司机在 c0-c9 中都有图像，除 P072 号司机数据量稍少，总体上每个司机在每个类别上的数据数量是平衡的。



4. 解决方法

利用 keras/tensorflow 设计一个卷积神经网络模型，该模型将图片的像素信息作为模型输入，然后利用卷积操作提取图像的特征，并基于训练数据不断优化模型参数，达到准确分类的目的。

5. 基准模型

我将使用 Keras 中训练好的 VGG16 模型^[3]作为基准模型，模型权重选择在 ImageNet 上训练的模型权重。

由于该项目的图像类别是 10 种，而 keras 中的 VGG16 模型如果使用 ImageNet 权重且包含最顶层的 3 个全连接层，最终的输出类别必须是 1000，所以需要去掉这 3 个全连接层，然后自行训练全连接层。

该模型在测试集上的 logloss 值为 1.49002，我的目标是评估指标低于这个基准模型的 logloss 值。

Name	Submitted	Wait time	Execution time	Score
result.zip	just now	1 seconds	3 seconds	1.49002
Complete				
Jump to your position on the leaderboard				

6. 评估指标

(1) kaggle提高的logloss^[4]

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

N 是测试集图片的数量，M 是图像类别标签，log 是自然对数，在图像 i 属于类别 j 时 $y_{ij}=1$ ，否则 $y_{ij}=0$ ， p_{ij} 是模型预测的图像 i 数据类别 j 的概率

7. 设计大纲

（1）图像数据加载分析

对数据集中的图像数据进行加载，进行可视化分析，说明数据集的特点。

（2）图像数据预处理

利用 `opencv` 对图像数据的大小进行处理，去除无关的数据，保留相关数据；对图像进行去中心化等处理；图像旋转等处理。

对训练数据进行分割，划分出 3 名司机的训练数据作为验证集数据，完成训练集和验证集的划分。

（3）基准模型创建及评估

利用 `keras` 建立基准模型，然后对基准模型进行训练和测试，利用测试集数据对测试结果进行评估，将预测结果提交 `kaggle`，得出评估结果。

（4）模型建立及评估

基于 `keras` 提供的基准模型，例如 `ResNet50`^[5]等建立模型初始模型，对初始模型进行训练评估。

（5）最终模型优化改进

使用迁移学习、模型微调和模型融合等技术对初始模型进行修改，根据第 6 节提到的评估指标对模型进行优化改进，提高模型的准确性，得到最终模型。

（6）模型测试和预测结果提交

根据 `kaggle` 比赛的要求，输出数据集的要求为：

img,c0,c1,c2,c3,c4,c5,c6,c7,c8,c9

img_0.jpg,1,0,0,0,0,...,0

img_1.jpg,0.3,0.1,0.6,0,...,0

...

参考文献

- [1]. 黄创新, 汽车驾驶员的行为状态识别[D], 2007, 北京科技大学
- [2]. <https://www.kaggle.com/c/state-farm-distracted-driver-detection>
- [3]. Simonyan Karen, Zisserman, Andrew. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. eprint arXiv:1409.1556
- [4]. <https://www.kaggle.com/c/state-farm-distracted-driver-detection#evaluation>
- [5]. He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Deep Residual Learning for Image Recognition[J]. eprint arXiv:1512.03385