

# 优达学城机器学习开题报告

## 课题背景

根据 CDC 部门的数据，五分之一的车祸是由一名走神司机引起的。可悲的是，这意味着每年有 425,000 人受伤，3000 人因驾驶走神而死亡。

如果能有效的区分驾驶员是否在专心驾驶，并制定相关的一系列规范，可以从一定程度上降低驾驶事故。

这是 KAGGLE 上两年前的竞赛题目。

对应网页地址：<https://www.kaggle.com/c/state-farm-distracted-driver-detection>

对于图像分类，现在比较有成熟的深度学习算法。对于 IMAGENET 训练集，深度学习的图像分类算法模型准确率已经超过了人类的分辨准确率。

因此使用深度学习分类算法完成项目的实现。

## 问题陈述

首先问题是一个监督学习图像多分类问题，需要根据提供的图片，判断司机当前的状态。

目标是通过程序实现从输入图片到判断出司机状态的功能，可以通过计算交叉熵的方式来描述程序的准确度。

基于图像分类的主要算法模型有：

- [VGGNet](#) 14.09<sup>1</sup>
- [ResNet](#) 15.12<sup>2</sup>
- [Inception v3](#) 15.12<sup>3</sup>
- [InceptionResNetV2](#) 16.02<sup>4</sup>
- [Xception](#) 16.10<sup>5</sup>
- [NASNet](#) 17.07<sup>6</sup>

从中选择一个算法模型作为基准算法模型进行相关的改造和训练。

## 数据集和输入

数据集的下载地址：<https://www.kaggle.com/c/state-farm-distracted-driver-detection/data>

从输入数据集可以看到 3 列数据，分别是司机编号、状态分类、对应的图片文件。

从项目描述可以看到，这是一个 10 分类的问题，共有 10 个状态需要判定。图片和分类之间是有关联性的，司机和分类的联系需要探索。训练数据一共有 22424 张图片。10 个分类：

- 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

共有 26 个司机的数据供训练。

每类图片选择同一个司机可视化展示，人眼观察图片的区别,跟最后的 CAM 可视化做对比。

竞赛提交到 KAGGLE 的测试集共有 79726 张图片。

通过可视化模型展示每个分类有多少张训练图片。

## 解决方案

1. 通过 OPENCV 读取图片，选用基准模型的不同，读取的图片向量也不同。
2. 选择合适的算法模型，如 VGG16,VGG19, Inception V3,Xception 其一作为基础模型。
3. 调整模型，由于图片之间差距很小，考虑放开全部层，全模型训练。
4. 调整模型和训练算法以及各类超参数，需要在训练过程中，看到 VAL\_LOSS 的下降，避免过拟合。
5. 在训练集数据的基础上训练算法，最后通过提交 KAGGLE 判断训练结果。

## 评价指标

KAGGLE 不同的项目评价标准也会有不同。本项目的评判是使用了 Multi class log loss 多分类的对数损失函数。

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

其中 n 是测试集中图像的数目，M 是图像类标签的数目。

为了避免对数函数的极值，用  $\max(\min(p, 1-10^{-15}), 10^{-15})$  代替。

KERAS 有对应的损失参数：

**categorical\_crossentropy**: 亦称作多类的对数损失，注意使用该目标函数时，需要将标签转化为形如 (nb\_samples, nb\_classes) 的二值序列。

目标是将损失函数尽量的降低。

具体可以参看: <https://www.kaggle.com/c/state-farm-distracted-driver-detection#evaluation>

## 项目设计

分以下几步走吧:

1. 数据处理。参见前面数据集和输入的描述。
2. 开始分析原始数据之间可能有什么联系，对数据进行切割重组。可以根据不同的司机来进行训练集和验证集的拆分，查看在不同司机上训练出的结果能否在不认识的司机图片中识别出来。
3. 因为是图像分类问题，选择一个基础模型作为基础模型，并针对性的做预处理。(不同模型的输入图像和预处理方式是不同的，需要对不同的模型采用不同的输入图片大小和输入预处理方式)，先使用 **VGG16** 进行第一步的实践。可以参考 **KAGGLE** **KERNAL** 进行实现。对应地址: <https://www.kaggle.com/c/state-farm-distracted-driver-detection/kernels>
4. 分类图片之间区别不大，放开所有层进行训练。
5. 优化判断结果，使用交叉验证法，对模型预测结果进行增强。
6. 如果选用基础算法模型性能低于预期，需要考虑更换算法模型。同时考虑修改超参数进行调整 (如: 选择不同的优化器如 **ADAM**<sup>7</sup>/**SGD** 等，各种优化器的超参数如学习率等，整个模型的 **BATCH\_SIZE**、**EPOCH** 等)。先从简单的 **VGG16** 开始训练，后续考虑 **XCEPTION** 的模型，开始调参之旅。
7. 最后对预测结果做一下 **CAM**<sup>8</sup> 可视化。

实现目标: 提交到 **KAGGLE** 测试集的 **LOSS** 在 **KAGGLE** 排名 10% 以内，即 0.25634 以内。

## 参考文献:

[1] Karen Simonyan, Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition.(arXiv:1409.1556,2014)

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun.Deep Residual Learning for Image Recognition.(arxiv:1512.03385,2015)

[3] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, Zbigniew Wojna .Rethinking the Inception Architecture for Computer Vision. (arXiv:1512:00567,2015)

[4] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, Alex Alemi. Inception-v4, Inception-ResNet

and the Impact of Residual Connections on Learning. (arXiv:1602.07261,2016)

[5] François Chollet .Xception: Deep Learning with Depthwise Separable Convolutions. (arXiv:1610.02357,2016)

[6] Barret Zoph, Vijay Vasudevan, Jonathon Shlens, Quoc V. Le. Learning Transferable Architectures for Scalable Image Recognition. (arXiv:1707.07012,2017)

[7] Diederik P. Kingma, Jimmy Ba. Adam: A Method for Stochastic Optimization (arXiv:1412.6980,2014).

[8] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba. Learning Deep Features for Discriminative Localization. CVPR'16 (arXiv:1512.04150, 2015).