

机器学习工程师毕业项目开题报告

udacity

问题描述



机动车驾驶员在形式的过程中时常会做出一些不符合安全驾驶的行为，例如：接打电话，发信息，和后排乘客谈话等等，这些将影响驾驶员的驾驶行为。本项目通过车内摄像头对车内驾驶员进行拍摄并通过**深度神经网络**处理每一帧的图片来确定驾驶员的状态，确定驾驶员当前是否正在安全驾驶，帮助驾驶员提高驾驶安全性。

项目背景

本项目是一个计算机视觉的多分类任务，属于有监督学习。具体是为了解决在机动车驾驶过程中，通过摄像头拍摄的图片确定驾驶员的行为状态是否符合安全驾驶。由于摄像头居中拍摄驾驶员捕获的图像位于捕获图片的正中央，所以本项目无需一些**全卷积网络**算法进行处理。通过深度神经网络，例如：*VGG*，*SSD*，*GoogLeNet*以及*ResNet*等成熟的网络对已经截获的样本做训练并确定当前的状态。

数据或输入

本项目的数据均来自*udacity*学城，[百度云链接](#)。

下载压缩包之后，解压分别有一个*csv*文件，两个存储图片的目录。

1. *driver_imgs_list.csv*
2. *train*
3. *test*

driver_imgs_list.csv

此文件类似以下内容，用于表示训练集与测试集的种类。

subject	classname	img
p002	c0	img_44733.jpg
...

- **subject**: 驾驶员的ID
- **classname**: 行为标号
- **img**: 对应的图片文件名称

下表列出了当前的所有的行为标号以及对应的意义。

状态列表：

行为标号	说明
c0	安全驾驶

行为标号	说明
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

以下列出了每个状态对应的训练图片样例：

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



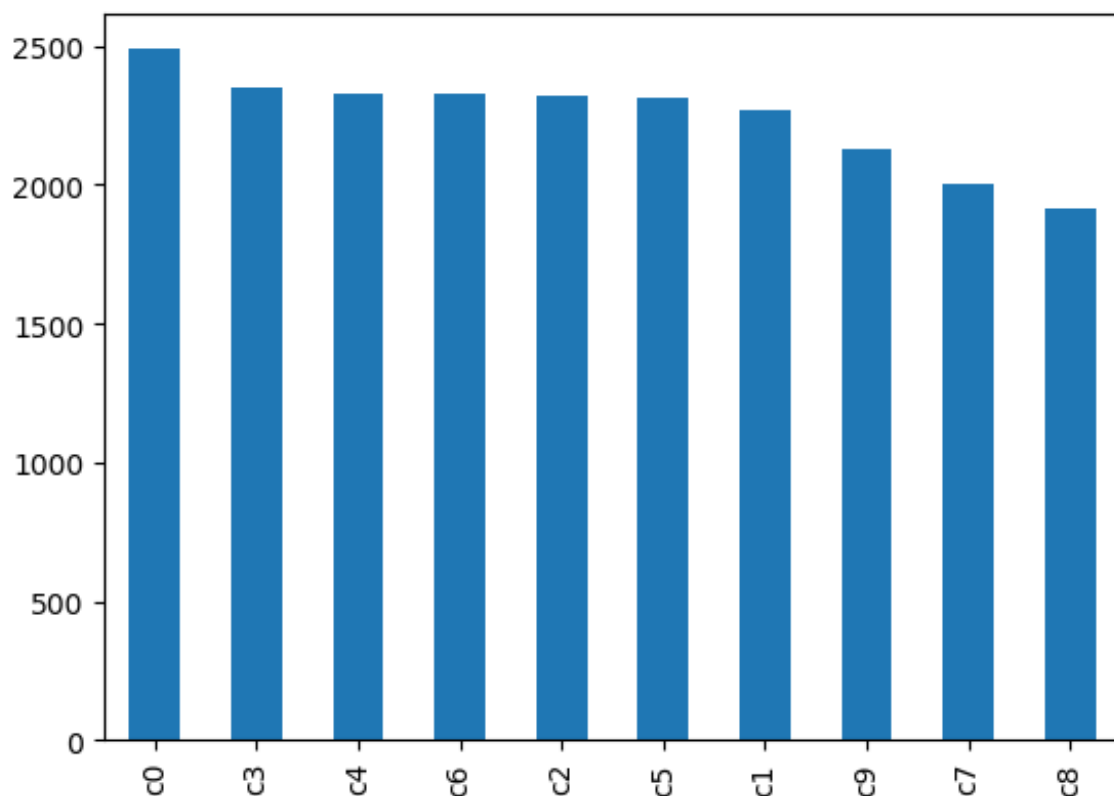
C9 和其他乘客说话



train目录

- 图片规格： $640 \times 480 \times 3$
- 样本总数量：**22424**

其中 **c0** – **c9** 十种状态的训练集差不多是按照 $\frac{1}{10}$ 的比例进行分配。其中 **c8** 最少只有 **1911** 张图片，其余的图片数量在 **2100** – **2300** 之间。



通过读取_driver_imgs_list.csv_文件并以行为状态来划分，也可以看出每种状态的训练图片的数量都基本相同，可以确定本项目属于一个个**类别平衡的多分类问题**。

test目录

测试样本具有**79,726**张图片。

在本项目中，首先会对拿到的图片做**灰度化**，**正则化**等数据预处理，来去掉数值计算对结果影响以及去掉一些颜色带来的噪声等影响。

由于图片都是居中设置，所以我会使用当前成熟的深度神经网络网络并做迁移学习，将提取特征的卷积层集中在提取图片中央的位置，已达到更符合本项目的目的。

valid目录

样本集中没有提供**验证集**的图片，我会挑选两位司机的所有图片作为验证集。

评估标准

本项目将最终使用对识别的精确度进行最终的评估。我要达成的目标 $accuracy > 0.95$, $logloss < 1$ 并且在 *Kaggle* 的排名在前1/4。

!!! 我看 *kaggle* 此项目已经关闭，不知道是否还可以继续进行提交。

这里我采用 $logloss$ 来对项目做最终的评估。 $logloss = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{ij} \log(p_{ij})$

基准模型

现在已经存在大量成熟的图片识别的深度学习模型。这里我准备采用 **VGG16** 卷积网络来作为基准模型来确定图片的状态。

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224×224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION

项目设计

本项目将遵循以下流程来进行工作。

训练

1. 对每次训练的图片进行**裁剪，灰化，正则化**处理。
2. 尝试对 **VGG16**的一些参数进行调整。
3. 使用 **VGG16**进行训练。
4. 训练完毕后通过验证集进行评估。
5. 如果流程4评估后的结果不满意则返回流程2对参数进行调整。
6. 如果流程4评估后的结果满意则保存模型。

检测

1. 对要检测的图片进行**裁剪，灰化，正则化**处理。
2. 使用训练好的模型进行鉴别。

参考资料

1. Karen Simonyan* & Andrew Zisserman+V.:VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION
2. Kaiming He Xiangyu Zhang Shaoqing Ren Jian Sun.: DeepResidualLearningforImageRecognition
3. Wei Liu¹, Dragomir Anguelov², Dumitru Erhan³, Christian Szegedy³, Scott Reed⁴, Cheng-Yang Fu¹, Alexander C. Berg¹.: SSD:SingleShotMultiBoxDetector
4. Uijlings, J.R., van de Sande, K.E., Gevers, T., Smeulders, A.W.: Selective search for object recognition. IJCV (2013)
5. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks. In: NIPS. (2015)
6. He,K.,Zhang,X.,Ren,S.,Sun,J.: Deepresiduallearningforimagerecognition. In:CVPR. (2016)
7. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: CVPR. (2016)