# 定义

## 项目背景

随着社会的发展，汽车已成为人们生活的必备交通工具，马路上的汽车也越来越多，而汽车驾驶的安全性，很大程度上与驾驶员的状态有关，酒驾、疲劳驾驶、驾驶过程中接打电话等虽然有关法律命令禁止，但是实际当中仍然有驾驶员违反从而造成安全事故的行为。而无人驾驶汽车，可以彻底把人从驾驶工作中解放出来，可以彻底杜绝酒驾、疲劳驾驶等问题。

本项目是MIT的无人驾驶课程学习项目，实现无人驾驶汽车中，根据前置摄像头所拍摄的实时录像，预测车辆转向角度的功能。本项目使用MIT 6.S094 这门公开课中的Tesla数据集，训练深度卷积神经网络学习模型，再使用模型对车辆的前置相机所拍摄的路况图像，对车辆转向角度进行预测。

## 问题描述

要得到训练模型，并实现对车辆转向角度预测，需要解决以下问题：

1. 从Tesla数据集的视频中，得到训练图片。从数据集的车辆转向记录中读取转向数据，将转向数据和训练图片做一一对应，得到训练数据集。
2. 图片数据和转向角度数据的预处理。
3. 使用交叉验证的方法，训练深度卷积神经网络，得到测试结果最佳时的CNN网络模型。

## 解决办法

对第一个问题，使用opencv库来读取视频中的图片，读取的图片还要进行标准化处理，将图片的像素值转化为0到1之间的浮点值。使用pandas库来读取车辆转向记录，由于车辆转向记录表中，已经有帧数和对应的转向角度数据，所以不用再做数据转换。

对第二个问题，由于车辆前置摄像头所拍摄的图像，有训练时用不到的冗余部分，如靠近底部显示车身的部分，和顶部有大约1/3的天空部分，需要在训练之前裁减掉。使用opencv库来做图片的裁剪。由于车辆转向角度是-20到20之间的离散值，间隔为0.5，需要对车辆转向角度数据做one-hot编码。为了后续训练方便，还要将预处理后的图片数据存到本地磁盘。使用pickle做图片数据的存储。

对第三个问题，使用10%的训练数据作为验证集。由于训练图片来自前置摄像机录制的视频，有一定的顺序性，所以有必要对数据随机化处理，打乱原本的图片顺序，增强训练模型的泛化能力。使用numpy库的random.permutation 方法来对训练图片和转向角度数据做随机化处理，确保图片和转向角度对应关系不打乱。

建立深度神经网络是训练模型的关键，在本项目中，将会分别使用以下CNN模型的基础上，做迁移学习，从中选择测试结果最好的一个模型最终使用。

1. 自己在训练CIFAR-10数据集中使用过的简单CNN模型。
2. 在2014年ILSVRC大赛上，获取分类项目冠军的InceptionNet网络，获得了top-5错误率6.67%。
3. 在2015年ILSVRC大赛上，获取分类项目冠军的ResNet网络，获得了top-5错误率3.57%。

# 分析

## 输入数据

本项目采用的数据，是MIT 6.S094 这门公开课中的Tesla数据集。其中视频图像是前置摄像机拍摄的10段高速公路上驾驶员行驶过程的录像，车辆转向信号数据集，来自驾驶过程中汽车内置存储记录 [1]。这些数据都是驾驶过程中的实际数据，驾驶过程中可能遇到的干扰图像，如路过阴影面积较大的大桥等都包括在训练数据集中，所以训练集训练的网络可以应用在实际驾驶当中。

由于实际训练时，使用的视频的图片是连续的，前后图片之间有很大的相关性，训练结果不利于泛化，所以在训练前要对训练图片做随机化处理。

# 方法

## 基准模型

ILSVRC大赛是图像识别领域世界级的权威赛事，获奖的参赛模型代表了图像识别领域的最先进的水平，2016 年 ILSVRC 的图像识别错误率已经达到约 2.9%，远远超越人类（5.1%），但是本项目涉及无人驾驶，识别错误意味着可能引发安全事故。不过好在与ILSVRC大赛相比，大赛需要在有大量随机干扰的图片中，识别确定的可能实际就很相似的物品。而驾驶过程中前置摄像机拍摄的内容，对于判断转向角度来说，重点只在视野前方的交通指示线、有无行人等道路状况内容，对于从中学习车辆转向角度来说，图片干扰比识别物体来说要少很多。所以，结果的错误率应该小于 ILSVRC中能达到的最小值2.9%，本项目预设的基本目标为测试错误率1%。

## 设计大纲

数据预处理

1. 从视频读取图片
2. 图片和转向几率做随机化处理
3. 图片规格化处理
4. 转向数据one-hot编码
5. 图片裁剪

简单神经网络

InceptionNet网络

ResNet网络

得到最佳模型导入测试视频

### 系统整体结构：

### 其中迁移学习使用的简单神经网络，结构为：

输入图片数据

卷积核：

* 大小：3\*3，深度：128，间距：1\*1
* 使用非线性激活
* 添加池化层，大小：2\*2，间距：2\*2

卷积核：

* 大小：1\*1，深度：256，间距：1\*1
* 使用非线性激活
* 添加池化层，大小：2\*2，间距：2\*2

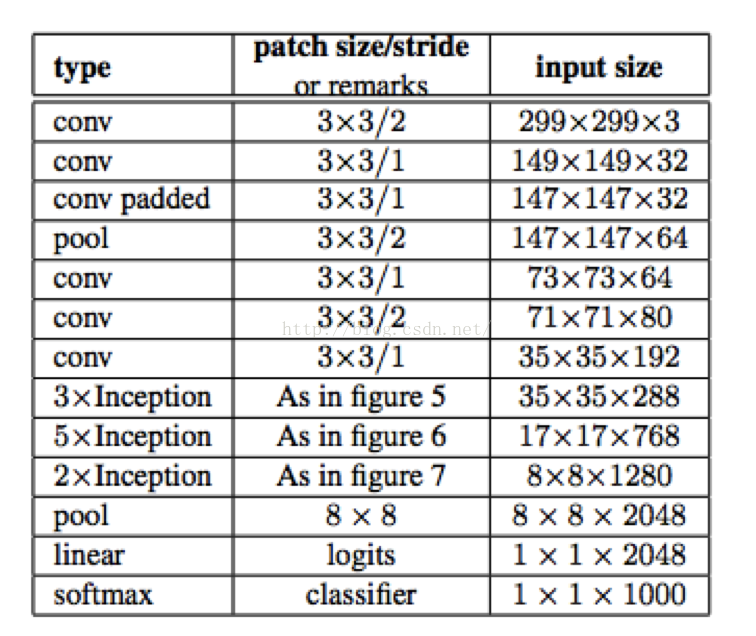
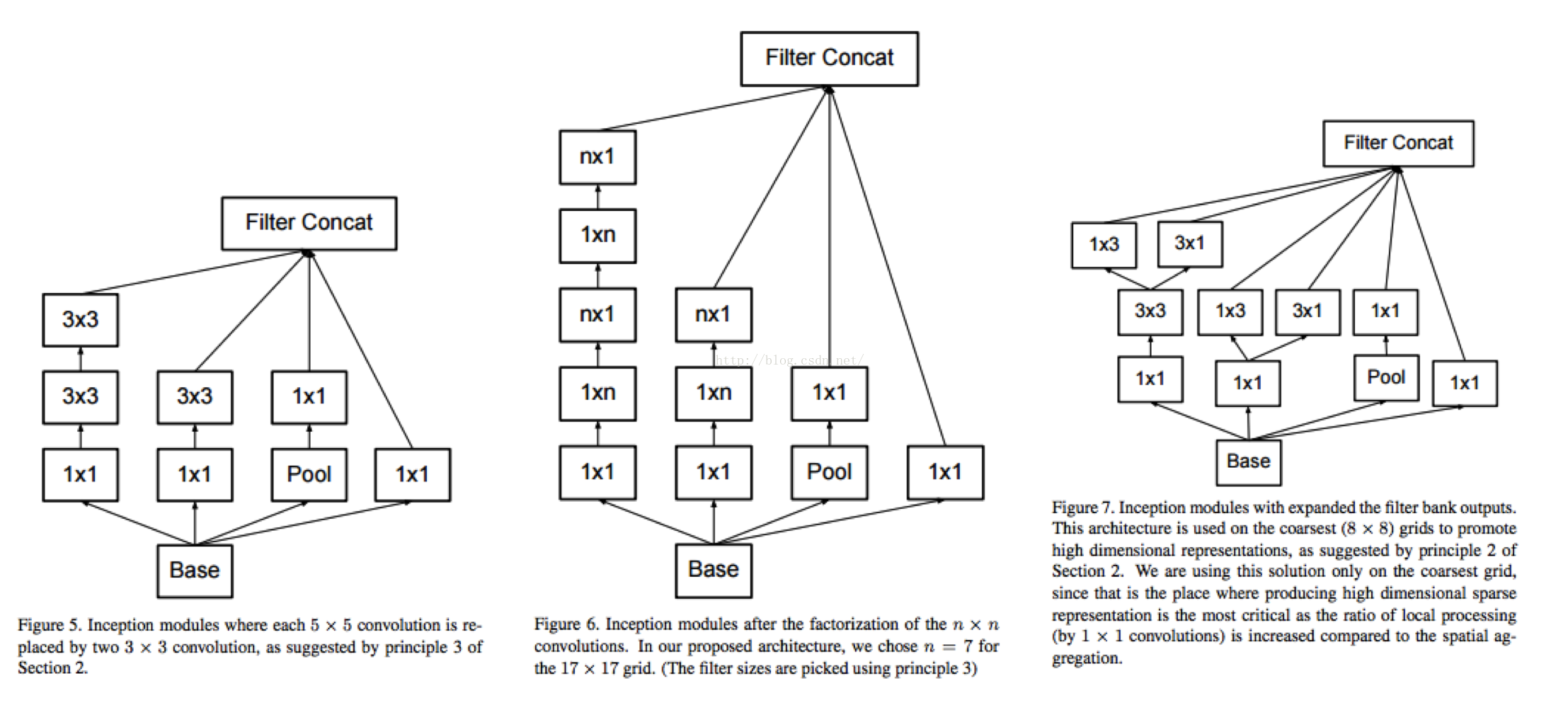
卷积核：

* 大小：9\*9，深度：384，间距：1\*1
* 使用非线性激活
* 添加池化层，大小：2\*2，间距：2\*2

卷积核：

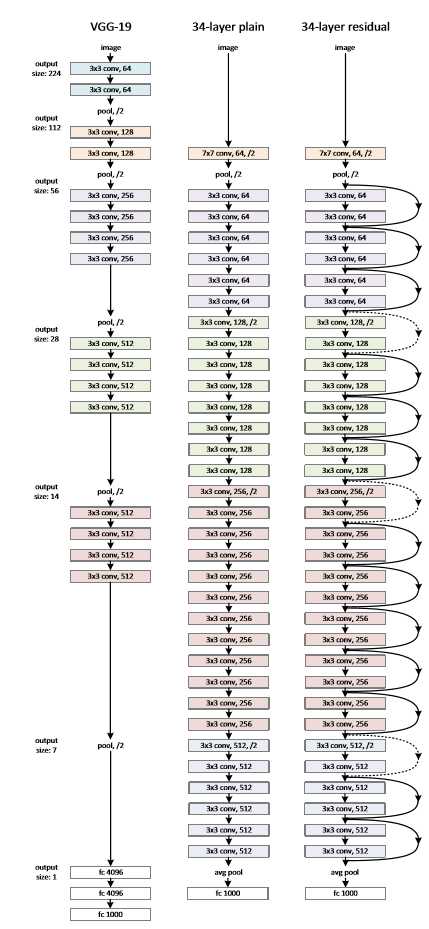
* 大小：1\*1，深度：512，间距：1\*1
* 使用非线性激活
* 添加池化层，大小：2\*2，间距：2\*2
* 扁平化层
* 深度1024的全连接层
* dropout保留率0.6
* 输出层

### 其中迁移学习使用的InceptionNet v2网络结构为[2]：



inceptionNet V3的结构为inceptionNetV2结构的基础上，做了进一步优化，inceptionNet提出了Inception层，将较大size的卷积核拆分成小的卷积连接，大大减少了模型参数，却保留了精度。本项目使用keras.applications.inception\_v3，来做inceptionNet V3网络的迁移训练。

### 其中迁移学习使用的RestNet的网络[3]



plain网络结构是在VGG19的基础上做了加深和优化，而RestNet则是在plain的基础上，增加了残差结构。有证据显示，卷积神经网络层数对于训练网络的表现有至关重要的作用，然而卷积神经网络越深，越容易导致梯度消失，从而阻止模型收敛，而残差结构可以解决深度神经网络容易梯度消失的问题，从而大幅提高模型的表现。本项目使用keras自带的restNet做迁移学习。

# 结果

## 评估指标

本项目的评估指标主要是最终测试结果的正确率。

另外本项目搭建的的深度神经网络，在做梯度下降时，将使用交叉墒代价函数计算训练中正向传播的结果和标注样本之间的差值loss（使用Tenserflow的softmax\_cross\_entropy\_with\_logits方法计算），并对这个差值做随机梯度下降计算，从而反向传播更新模型权重，来不断减少这个差值loss，从而缩小正向传播的结果和标注样本间的差距，达到减小图像识别错误率的目的。为了达到基准模型小结中设定的1%的错误率，训练中loss值应该需要减少到至少0.05。

参考文献

[1] MIT公开课6.S094: Deep Learning for Self-Driving Cars 讲义

[2] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, Zbigniew Wojna , Rethinking the Inception Architecture for Computer Vision

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Microsoft Research , Deep Residual Learning for Image Recognition