

开题报告

Weilin Tang

项目背景

猫狗识别源自于kaggle上的一个竞赛：[Dogs vs. Cats](#)。训练集有25000张，猫狗各占一半。从机器学习的角度，这属于一个计算机视觉中的图像分类任务。随着深度学习的不断发展，卷积神经网络已经在图像分类上取得了非常不错的效果，比如[微软研究员在ImageNet计算机视觉识别挑战中实现里程碑式突破](#)。因此我们选择利用CNN的相关技术来解决这个猫狗分类的问题是非常合适的。

问题描述

在这个项目中，我将实现一个算法，能够检测到图像中是否包含狗或者猫，以及分辨出是狗还是猫。算法将在测试集上运行并计算出分类的准确率，并通过模型优化提升准确率至一个能够接受的结果。

输入数据

在这个项目中用到的数据集有如下几个来源：

- kaggle官方提供的数据集：包含25000张的训练集（猫狗各占一半），12500张测试集（未标注）
- 在kaggle官方的训练集上利用 `data augmentation` 的方式扩展用于模型训练的数据
- 通过搜索引擎爬取一定量的数据
- 从公开数据集 `stanford dogs datasets` 中获取一定的数据

收集到的数据集按照一定的比例分为训练集和验证集，分别存放在两个目录下。

解决办法

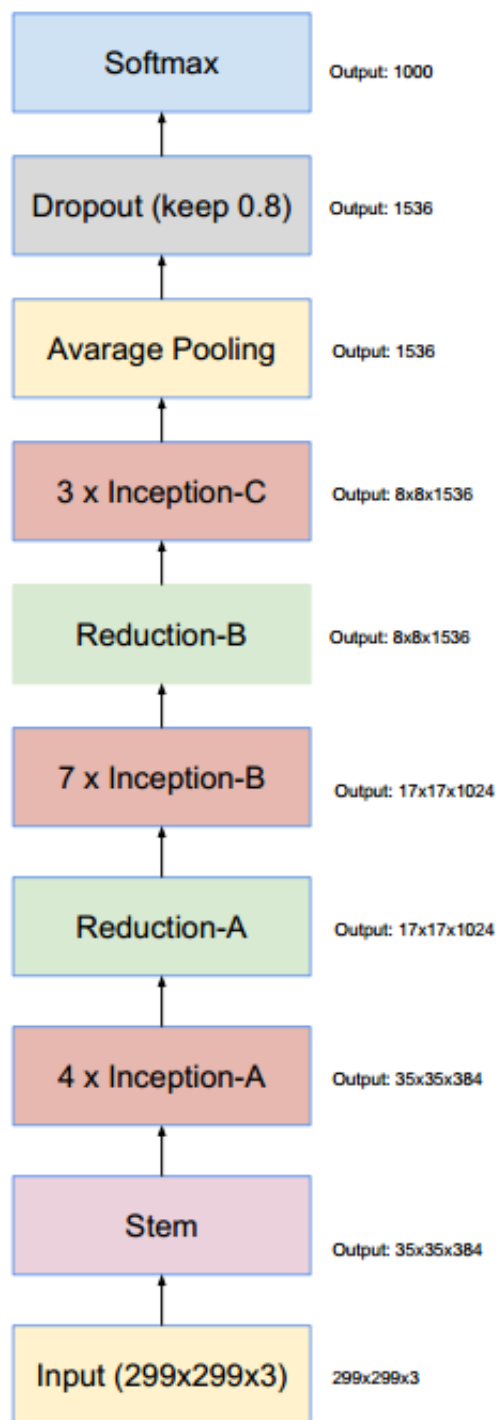
因为该问题可以当作是一个图像分类问题，因此有多种成熟的cnn网络结构（如vgg19、resnet、googlenet等等）可供使用。但是考虑到我手头上的猫和狗的数据是有限的，不足以从头训练一个能够收敛得很好的模型，所以可以考虑在已有的训练好的模型上进行 `fine tuning`。

整个流程可以分为：

- 加载预训练模型
- 准备好自己的训练数据
- 修改新的类别数以及输出

基准模型

基准模型我们选取GoogLeNet的Inception v4。该模型是Inception模块结合了ResNet的结构从而得到的新的网络架构，极大地提升了在之前Inception v3基础上的训练速度，性能也得到了提升。其结构如下图所示：



评估指标

利用kaggle上提供的测试集对本解决方案进行评估，统计猫狗的总体分类准确率、猫和狗的分别的分类准确率。

设计大纲

基于上面的分析，本项目的方案总结为：

1. 选择卷积神经网络作为猫狗分类的方法
2. 采用fine tuning的方式，在已有的预训练好的模型上进行迁移学习
3. 选取两到三个主流的state-of-art的网络架构进行实验，选择表现最好的一个框架
4. 模型训练：调节超参数，优化分类结果。