# 猫狗大战开题报告

## 项目背景

图像分类是根据图像的语义信息将不同类别图像区分开来，是计算机视觉中重要的基本问题，也是图像检测、图像分割、物体跟踪、行为分析等其他高层视觉任务的基础。ImageNet上的图像分类竞赛是计算机视觉领域中的重要比赛，该项比赛需要对1000个类别进行识别。在2012年，Alex Krizhevsky、Ilya Sutskever 和 Geoffrey E.Hinton利用CNN网络进行图像分类，使Top-5的错误率降到15.3%，而第二名的错误率则为26.2%，领先了10.9%[1]。在后续的比赛中，不断的涌现出各种结构的CNN模型，大幅提升了图像分类的准确率，降低了错误率。2014年的GoogLeNet将Top-5的错误率降至6.77%[2]，2015年的ResNet将Top-5错误率降至3.57%[3]，这一错误率已经超越了人类水平。

## 问题描述

本项目来源于2016年9月Kaggle举办的竞赛（[Dogs vs. Cats Redux: Kernels Edition](https://www.kaggle.com/c/dogs-vs-cats-redux-kernels-edition)），是一个典型的监督学习中的图像分类问题。

通过训练机器学习模型，来分辨图像为猫或狗的的概率（ 1代表狗，0代表猫）。使用训练后的模型对测试集进行预测，最终通过将预测结果提交至Kaggle平台进行评分。

## 输入数据

该项目所使用的数据集是Kaggle所提供的猫狗图片，包含训练集和测试集。

训练集包含25,000张已经标注猫和狗类别的图片，其中有12500张标注为猫的图片和12500张标注为狗的图片。图片的命名格式为**type.num.jpg**，其中type为图片类别（cat或dog），num代表图片序号（从0至12499）。图像包含了大小从42x62到1052x702内的各种尺寸，以及猫狗的各种动作和方向，其中大部分图像以猫狗为主体，部分图片不以猫狗为主体，甚至还有些不是含猫狗的异常值。



以猫狗为主体的图片



不以猫狗为主题的图片

E:\python\udacity\capstone\dog_vs_cat\data\train\dog.4367.jpgE:\python\udacity\capstone\dog_vs_cat\data\train\cat.9171.jpg

不是猫狗的异常值

测试集包含12,500张未标注图片，尺寸范围从37x50到500x499，其分布情况与训练集类似，存在以猫狗为主体的图像和不以猫狗为主体的图像，以及异常值。



以猫狗为主体的图像



不以猫狗为主体的图像以及异常值

## 解决办法

由于该项目为图像分类问题，目前在图像处理领域内，相比SVM等传统机器学习方法，使用深度学习中的深层卷积神经网络能够取得良好的效果。同时，由于目前在图像分类领域有很多非常好的CNN模型，因此，针对该项目，决定使用卷积神经网络结合迁移学习进行解决，具体使用Keras进行实现。

## 基准模型

进入Kaggle比赛公共排行榜10%（131名以内），即得分（Score）小于0.06127。

## 评估指标

该项目Kaggle比赛统一规定评估指标为log loss，具体公式如下：

这里：

： 为测试集中图片数量；

： 预测图像为狗的概率；

： 如果图像为狗则为1，为猫则为0；

： 底数为e的对数函数；

## 设计大纲

项目的整体方案大致可以分为以下几个步骤：

* 数据预处理
* 搭建并训练CNN模型
* 预测测试集
* 提交Kaggle进行评分

### 数据预处理

数据集从Kaggle下载下来后，经过观察可以发现以下几个问题：

1. 默认的数据集没有按照cat和dog进行文件夹分类。若使用Keras的图像预处理函数ImageDataGenerator，则需要将图像按照标签归类到不同的文件夹中。
2. 图像的尺寸大小不一致，对于该问题，可通过使用Keras自带的预处理函数对图片进行尺寸更改，同时使用Keras提供的preprocess\_input函数进行像素级缩放,将RGB的颜色从0~255缩放至模型所需的范围（根据模型不同，该范围也不同，可能是0~1，可也能是-1~1）。
3. 异常数据的存在，在训练集中存在部分既不是猫也不是狗的异常值，对于这部分数据需要剔除。

**具体剔除方法：**由于猫和狗是存在于ImageNet上的分类的，所以使用在ImageNet上训练好的模型和权重对所有训练集进行预测，取Top-N中不存在猫或狗的数据为异常值，进行剔除。为避免单个模型的局限性，计划采用Xception、InceptionResNetV2、InceptionV3三个模型分别进行预测，取三个模型异常值的并集作为最终异常值。[4]

上述三个问题解决后，还可以根据需要对图像进行增强，例如旋转图片、调整图像的亮度等。

### 搭建并训练CNN模型

#### 1）模型搭建：

由于是采用迁移学习的方法，计划采用Xception、InceptionResNetV2、InceptionV3这三个已有模型和权重实现迁移学习，之所以采用这三个模型，是因为这三个模型在ImageNet数据集上的Top-5准确率比较高，分别达到0.945、0.953和0.944。但是，由于这个三个模型在ImageNet数据集上训练的，而ImageNet数据集有1000种分类，因此若要使用这三种模型，还需要对其最后的全连接层进行更改，将其改为单神经元输出。具体的实现方法计划采用以下两种方案：

**方案1：**在三个模型现有的ImageNet权重的基础上，将每个原模型的全连接层替换为两层全连接层，神经元数量分别为256、1，在两个全连接层中增加Dropout,取参数0.5，并放开所有层进行训练，最后使用sigmoid函数进行输出概率。对于每一个图像，最后会的到3个结果，将这3个求平均值作为最后预测结果。



**方案2：**在三个模型现有ImageNet权重基础上，去除原全连接层，将最后的卷积层的输出进行全局平均池化（GlobalAveragePooling），作为特征向量（2048,1）。将三个模型的特征向量进行连接，形成一个1维向量（6144,1），并在该向量后面接两层全连接层，神经元数量分别为256、1，在两个全连接层中增加Dropout,取参数0.5，最后使用sigmoid函数输出概率。



#### 2）数据集划分

对预处理并打乱后的数据进行训练集和验证集的划分，取验证集占所有数据集的0.2，即：用于训练模型的图像有20000张，用于验证的图像有5000张。

#### 3）训练模型

模型的训练采用Adam优化方法，学习率初步给定0.001。

### 预测测试集

对测试集分别使用两种方案的模型进行预测，输出两份csv文件。

### 提交Kaggle进行评分

将预测结果提交Kaggle平台进行评分，并检验是否可以进入公共排行榜前10%以内。

## 参考文献

[1] Krizhevsky A, Sutskever I, Hinton G E, et al. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 67: 2361-2367.

[2] Szegedy C, Liu W, Jia Y, et al. Going Deeper with Convolutions[J]. arXiv preprint arXiv:1409.4842, 2014.

[3] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[J]. arXiv preprint arXiv:1512.03385, 2015.

[4] 凌蓝风, 毕业设计 Dogs vs Cats For Udacity P7 (异常值检验). <https://zhuanlan.zhihu.com/p/34068451>. Published: 2018-05-21.