

# 机器学习工程师——猫狗大战

## 开题报告

韩旭 2019年5月22日

### 项目背景

猫狗大战是2013年kaggle上的一个竞赛题目。目的是通过机器学习或者深度学习技术搭建可以识别图片并且能够正确对猫和狗进行分类的模型，这是一个属于计算机视觉领域二分类的模型。

深度学习技术是近几年兴起的技术，也是这些年人工智能领域取得的重大成就之一。现在已经在搜索引擎、机器翻译、图像识别、自然语言处理、无人驾驶这几个领域逐步趋向成熟。

自从Alex和他的导师Hinton在2012年的ImageNet大规模图像识别竞赛（ILSVRC2012）中以超过第二名10个百分点的成绩(83.6%的Top5精度)碾压第二名（74.2%，使用传统的计算机视觉方法）后，深度学习真正开始火热，卷积神经网络（CNN）开始成为家喻户晓的名字，从12年的AlexNet（83.6%），到2013年ImageNet大规模图像识别竞赛冠军的88.8%，再到2014年VGG的92.7%和同年的GoogLeNet的93.3%，终于，到了2015年，在1000类的图像识别中，微软提出的残差网（ResNet）以96.43%的Top5正确率，达到了超过人类的水平。

伴随着图像分类任务，还有另外一个更加有挑战的任务-图像检测，图像检测是指在分类图像的同时把物体用矩形框给圈起来。从14年到16年，先后涌现出R-CNN, Fast R-CNN, Faster R-CNN, YOLO, SSD等知名框架，其检测平均精度，在计算机视觉一个知名数据集上PASCAL VOC上的检测平均精度，也从R-CNN的53.3%，到Fast RCNN的68.4%，再到Faster R-CNN的75.9%，最新实验显示，Faster RCNN结合残差网（Resnet-101），其检测精度可以达到83.8%。深度学习检测速度也越来越快，从最初的RCNN模型，处理一张图片要用2秒多，到Faster RCNN的198毫秒/张，再到YOLO的155帧/秒（其缺陷是精度较低，只有52.7%），最后出来了精度和速度都较高的SSD，精度75.1%，速度23帧/秒。<sup>1</sup>

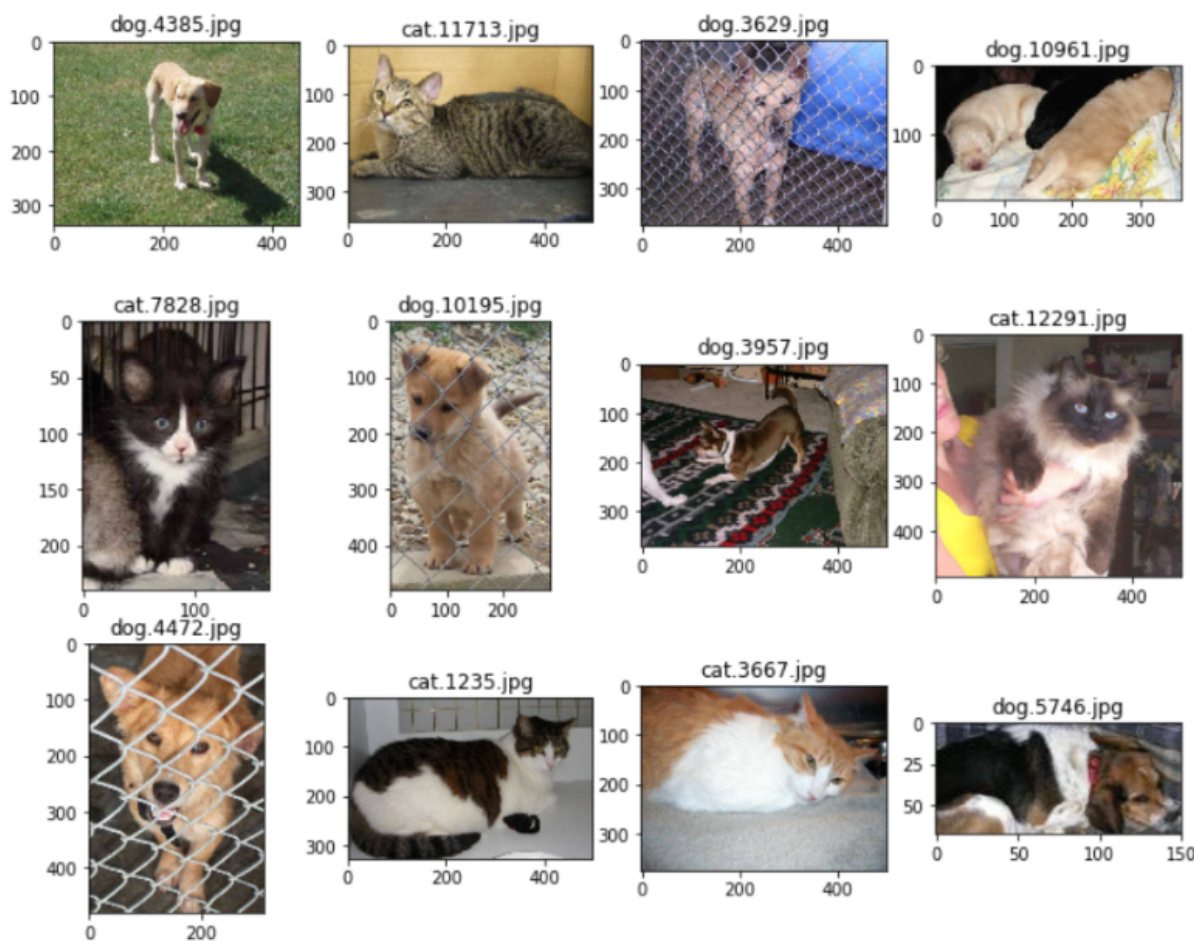
### 问题描述

这个项目需要搭建深度学习模型来识别猫狗，本质上是一个二分类的问题。拟采用深度学习技术在aws上搭建CNN模型配合迁移学习技术。给定一张图片，来通过模型对图片进行预测猫和狗的概率，从而达到分类的目的。

### 输入数据

数据集来由kaggle官方<sup>2</sup>提供，一共有两个压缩文件，分别是train.zip和test.zip。train.zip的文件用来训练数据模型，test.zip的文件是用来评估模型。train.zip中一共有25000张图片，猫和狗各占一半，每张图片都带有标签，因为正负样本的数量相同，所以不用担心样本失衡的问题。另外，需要把图片归一化，把色彩缩放在[0, 1]之间，使用keras的ImageDataGenerator.flow\_from\_directory的参数target\_size设置图片大小进行resize。防止因为图片尺寸问题无法训练神经网络。

随机输出图片



这里可以看出来，图片大小都很不规则，有些时候还有遮挡物，不知道遮挡物会不会对模型的训练产生影响。接下来打算对图片的大小进行归一化，裁剪成（224， 224， 3）的尺寸，为了方便后边神经网络的训练。另外这些图片还要按照4:1的比例划分训练集和验证集。

## 解决方法

这里使用卷积神经网络进行建模解决问题，卷积神经网络是一种前馈神经网络，对于大型图像处理有出色表现。在国际标准的ImageNet数据集上，许多成功的模型都是基于CNN的。CNN相较于传统的图像处理算法的优点之一在于，避免了对图像复杂的前期预处理过程（提取人工特征等），可以直接输入原始图像。CNN网络对图片进行多次卷积层和池化层处理，在输出层给出两个节点并进行softmax计算得到两个类别各自的概率。

## 基准模型

使用基于keras的resnet<sup>3</sup>，Xception<sup>4</sup>，VGG16<sup>5</sup>等网络模型去完成项目。这个比赛总共有1314支队伍参加，而此次毕业项目的要求是 kaggle Public Leaderboard 前10%。所以logloss比第134名的0.006320小就好。<sup>6</sup>

## 评估指标

采用对数损失来衡量：

$$LogLoss = -\frac{1}{n} \sum_i^n [y_i \log(1 - y_i) + (1 - y_i) \log(y_i)]$$

其中：

- $n$ 是图片数量
- $\hat{y}_i$ 是模型预测为狗的概率
- $y_i$ 是类别标签, 1 对应狗, 0 对应猫
- $\log()$  表示自然对数

LogLoss越小的模型也就更加的鲁棒, 所以LogLoss可以用来评估这个项目。

## 设计大纲

### 数据预处理

- 从kaggle下载好图片
- 对图片进行resize, 保持输入图片信息大小一致
- 对训练数据进行随机偏移、转动等变换图像处理, 这样可以尽可能让训练数据多样化

### 模型搭建

Kera的应用模块Application提供了带有预训练权重的Keras模型, 这些模型可以用来进行预测、特征提取和微调整和。

- 使用ResNet50等现有的去掉了全连接层预训练模型
- 添加自己的全连接层到ResNet50网络

### 模型训练&模型调参

- 导入预训练的网络权重
- 搭建卷积层, 池化层, Dropout层, 全链接层
- 优化器使用adam, 二分类的loss使用binary\_crossentropy

### 模型评估

- 使用 $Logloss$ 进行模型评估,上传Kaggle判断是否符合标准

### 可视化

- 进行数据探索并且可视化原始数据
- 可视化模型训练过程的准确率曲线, 损失函数曲线等

## 参考文献

- 
1. 深度学习在计算机视觉领域的前沿进展 [腾讯云](#)
  2. [Dogs vs. Cats Redux: Kernels Edition](#)
  3. [你必须要知道CNN模型: ResNet 知乎](#)
  4. [论文 - Xception, Deep Learning with Depthwise separable Convolution](#)
  5. [一文读懂VGG网络 知乎](#)
  6. 杨培文.深度学习技术图像处理入门. 清华大学出版社.2018年10月 : 134-154