

机器学习--猫狗大战App实现

开题报告

何伟华

2019.07.29

项目背景

Cats vs. Dogs来源于Kaggle大数据竞赛的一道赛题(娱乐型竞赛项目):<https://www.kaggle.com/c/dogs-vs-cats-redux-kernels-edition> (<https://www.kaggle.com/c/dogs-vs-cats-redux-kernels-edition>), 利用给定的数据集, 用算法实现猫和狗的识别。这属于计算机视觉领域的一个问题。

kaggle一共举行过两次猫狗大战的比赛, 第一次是在2013年, 那个时候使用的是正确率作为评估标准, 而在2017年第二次举办猫狗大战的比赛时, 使用的是log损失函数。这么做是因为现在深度学习的发展到一定阶段有更好的方法, 而深度学习尤其适合处理图像方面的问题, 如果依旧是使用正确率作为评估标准, 那么大多数选手的模型都是99%的正确率, 不能明显地区分开。如果使用log损失函数, 不仅仅需要分类正确, 还需要对结果有一个较高的可信度, 这样就能明显地区分各个模型的分类效果, 尤其是Top模型的分类效果。

因此参赛者需要训练一个机器学习模型, 输入测试集中的图片, 输出一个概率, 概率越接近1, 表示该图片分类结果是狗的概率越高; 概率越接近0, 表示该图片分类结果是猫的概率越高。

卷积神经网络(Convolutional Neural Network, CNN)是深度学习技术中极具代表的网络结构之一, 在图像处理领域取得了很大的成功, 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。卷积网络是第一个解决重要商业应用的神经网络, 并且仍处于当今深度学习商业应用的前沿。卷积神经网络针对图像识别特别棒。1988年应用于检测医学影像, 人像识别、手势识别等; 随着技术的发展, 卷积神经网络多次成为ImageNet大规模视觉识别竞赛(ImageNet Large Scale Visual Recognition Challenge, ILSVRC)的优胜算法, 到目前出现了很多优秀的算法模型, 比如: VGG、Xception、ResNet50、InceptionV3等等。

问题描述

猫狗大战是Kaggle娱乐型竞赛项目, 我们的目的需要用训练集对模型进行训练, 然后在测试集上“考试”, 提交kaggle取得高分查看考试结果, 现在是2019年, 这个是2年前的比赛, 是不能查看自己排名。本项目使用卷积神经网络识别一张图片是猫还是狗, 这是一个二分类问题。给定一张图片, 算法需要预测出图片属于预先定义类别中的哪一类。在计算机视觉领域, 目前解决这类问题是深度学习(Deep Learning), 特别针对图像类型的数据, 是深度学习中的卷积神经网络CNN架构, 针对图像识别特别棒。

- Kaggle提供的数据存在异常数据吗? 什么算是异常数据?

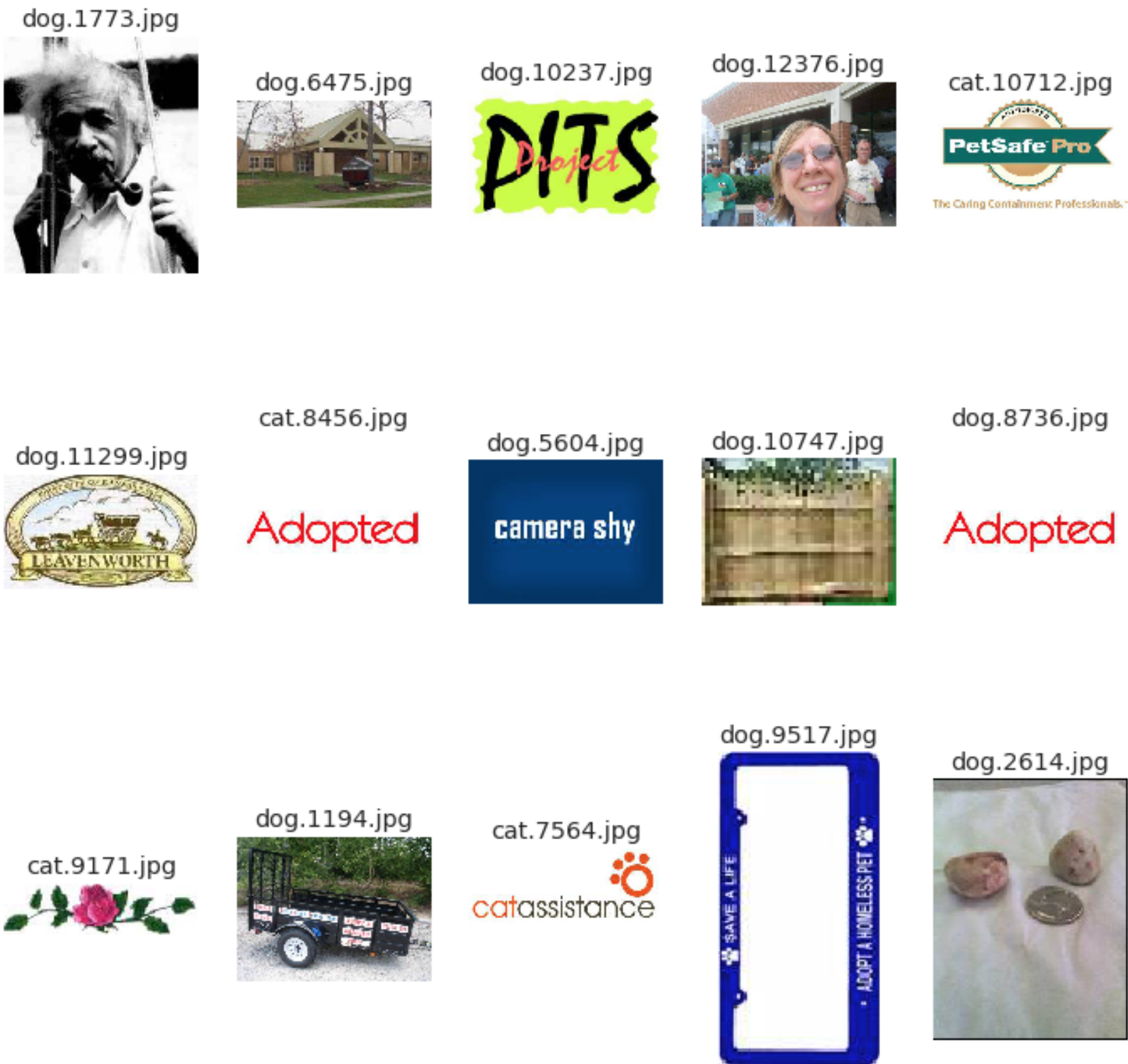
- 哪些图片可以被清理?
- 25000张猫狗的图片，用什么工具或者方法来剔除?
- 没有GPU服务器训练怎么破?

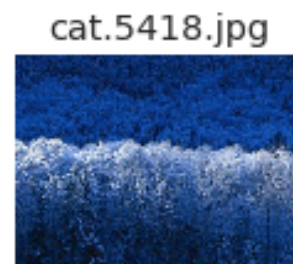
输入数据

数据集来自 kaggle 上的一个竞赛: [Dogs vs. Cats Redux: Kernels Edition](https://www.kaggle.com/c/dogs-vs-cats-redux-kernels-edition/data)
(<https://www.kaggle.com/c/dogs-vs-cats-redux-kernels-edition/data>)。

下载kaggle猫狗数据集解压后分为 3 个文件 train.zip、 test.zip 和 sample_submission.csv。训练集(train)包含了25000张猫狗的图片， 每张图片包含图片信息本身和标签。命名规则为"type.num.jpg"。测试集(test)包含了12500张猫狗的图片， 每张图片命名规则是"num.jpg"， 注意: num从1开始， 训练集的编号从0开始。 sample_submission.csv 需要将最终测试集的测试结果写入xxx.csv文件中， 提交kaggle取得高分查看考试结果。 训练集中大部分图片是正常的， 有少部分异常图片（非猫非狗的图片， 或者影响模型学习的图片-背景复杂错乱等）和低分辨率图片， 对于训练集来说这些异常数据是要剔除掉的。异常图片见下图：

bad pictrues all: 23





训练集中的图像大小是不固定的，但是神经网络输入节点的个数是固定的。所以在将图像的像素作为输入之前，需要将图像的大小进行resize。使用深度学习方法识别一张图片是猫还是狗，这是一个二分类问题。0表示分类结果是猫，1表示分类结果是狗。

- 输入：一张彩色图片
- 输出：猫或者狗的概率

解决办法

项目中使用卷积神经网络（CNN）的方法解决问题。卷积神经网络(Convolutional Neural Network, CNN)是深度学习技术中极具代表性的网络结构之一，在国际标准的ImageNet数据集上，许多成功的模型都是基于CNN的。针对图像识别使用CNN特别棒。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sigmoid计算得到猫狗两个类别各自的概率。

- Kaggle提供的数据存在异常数据吗？什么算是异常数据？
- 答：存在异常数据，在train文件夹中，混杂在25000张图片中。异常数据都是非猫非狗的图片或是背景复杂，低分辨率的图片，干扰项目（背景）非常多的图片。
- 哪些图片可以被清理？
- 答：Train文件夹内的数据可以清理，Test不在清理范围。
- 25000张猫狗的图片，用什么工具或者方法来剔除？
- 答：预处理模型，在ImageNet上有很多大牛调试好的分类器，并且提供了模型和权重文件，可以很好地帮我们完成任务。ImageNet 是要对1000个物品进行分类，当然这个分类不是物种的分类，举例：狗，是物种的分类，但是具体是那种狗，ImageNet细分了118个细分类。评价ImageNet有个指标Top-1和Top-5：

模型	大小	Top-1 准确率	Top-5 准确率	参数数量	深度
Xception	88 MB	0.790	0.945	22,910,480	126
VGG16	528 MB	0.715	0.901	138,357,544	23
VGG19	549 MB	0.727	0.910	143,667,240	26
ResNet50	99 MB	0.759	0.929	25,636,712	168
InceptionV3	92 MB	0.788	0.944	23,851,784	159
InceptionResNetV2	215 MB	0.804	0.953	55,873,736	572
MobileNet	17 MB	0.665	0.871	4,253,864	88
DenseNet121	33 MB	0.745	0.918	8,062,504	121
DenseNet169	57 MB	0.759	0.928	14,307,880	169
DenseNet201	80 MB	0.770	0.933	20,242,984	201

在这里我使用ResNet50来筛选异常图片。

- 没有GPU服务器训练怎么破？
- 答：使用云计算服务，比如：aws、腾讯、阿里云服务，或是使用公司的云计算服务。

基准模型

项目使用ResNet50, Xception, InceptionV3, InceptionResNetV2 四个模型完成。本项目的最低要求是kaggle Public Leaderboard 前10%。在kaggle上，总共有1314只队伍参加了比赛，所以需要最终的结果排在131位之前，131位的得分是0.06127，所以目标是模型预测结果要小于0.06127。

评估指标

kaggle 官方的评估标准是 LogLoss，下面的表达式就是二分类问题的 LogLoss 定义。

$$LogLoss = -\frac{1}{n} \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

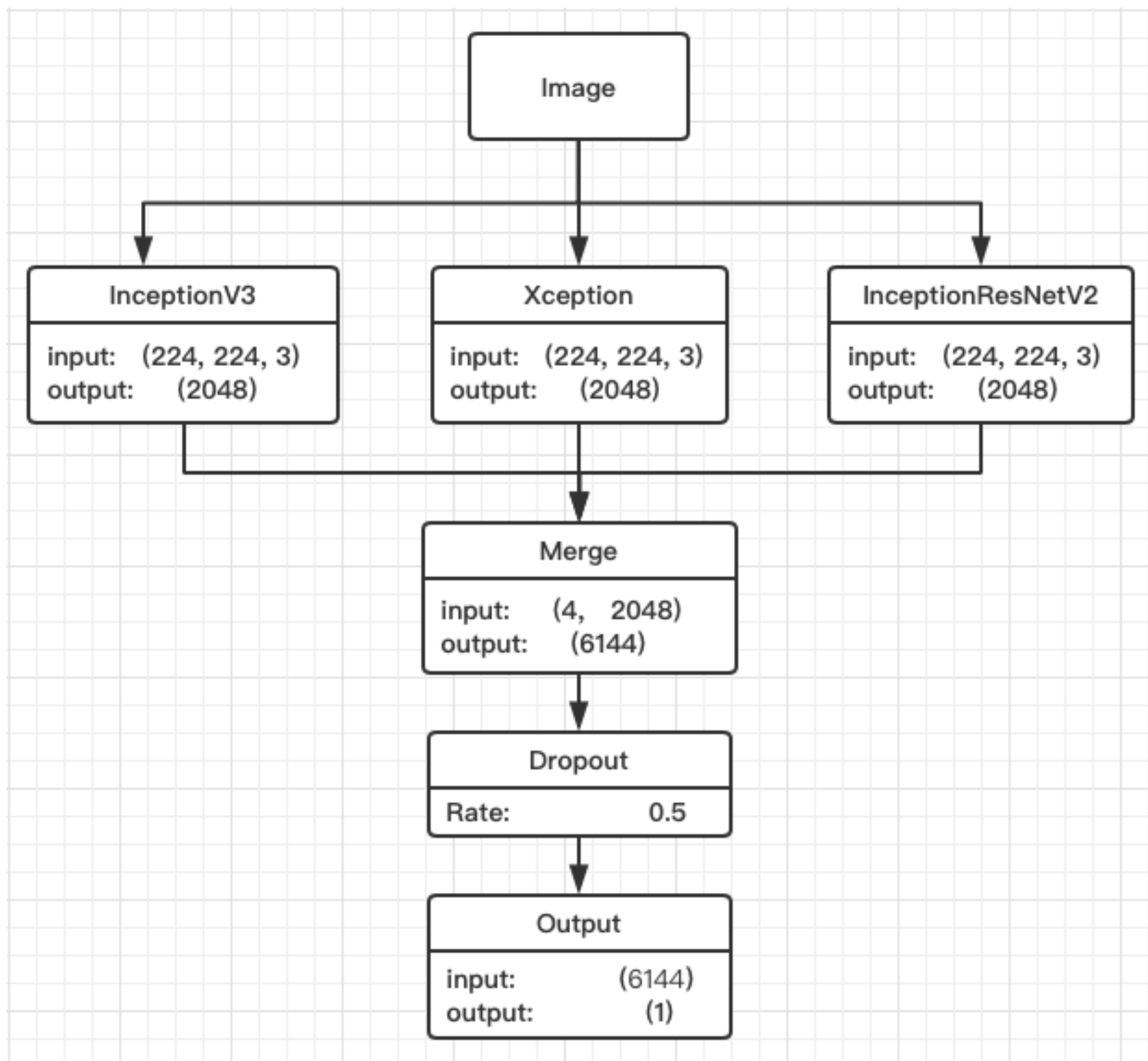
其中：

- n 是测试集中图片数量
- \hat{y}_i 是图片预测为狗的概率
- y_i 如果图像是狗，则为1，如果是猫，则为0
- $\log()$ 是自然（基数 e ）对数

对数损失越小，代表模型的性能越好。上述评估指标可用于评估该项目的解决方案以及基准模型。

设计大纲

- 本项目使用融合模型实现。
- 融合模型方法：首先将特征提取出来，然后拼接在一起，构建一个全连接分类器训练即可。
- 模型融合能提供成绩的理论依据是，有些模型识别狗的准确率高，有一些模型识别猫的准确率高，给这些模型不同的权重，让他们能够取长补短，强强联合，综合各自的优势，为了更高的融合模型，可以提取特征进行融合，这样会有更好的效果，弱特征的权重会越学越小，强特征会越学越大，最后得到效果非常好的模型。



1. 数据预处理

- 从kaggle下载好图片
- 将猫和狗的图片解压分别放在不同的文件夹以示分类，使用创建符号链接的方法
- 对图片进行resize，保持输入图片信息大小一致
- 图像文件分类后的路径如下：
- image
 - |—— test
 - |—— train
 - |—— img_train
 - |—— cat
 - |—— dog

2. 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。

- ResNet50 默认输入图片大小是 $224 \times 224 \times 3$
- Xception 默认输入图片大小是 $299 \times 299 \times 3$
- InceptionV3 默认输入图片大小是 $299 \times 299 \times 3$
- InceptionResNetV2 默认输入图片大小是 $299 \times 299 \times 3$

在Keras中载入模型并进行全局平均池化，只需要在载入模型的时候，设置 `include_top=False`，`pooling='avg'`。每个模型都将图片处理成一个 1×2048 的行向量，将这四个行向量进行拼接，得到一个 1×8192 的行向量，作为数据预处理的结果。

使用正确的预处理函数：

- ResNet50模型，用下面的预处理函数

In [1]:

```
def preprocess_input(x):  
    return x - [103.939, 116.779, 123.68]
```

- Xception、InceptionV3、InceptionResNetV2模型，分别用下面的预处理函数

In [2]:

```
from keras.applications.inception_v3 import preprocess_input  
from keras.applications.xception import preprocess_input  
from keras.applications.inception_resnet_v2 import preprocess_input
```

Using TensorFlow backend.

3. 模型训练&模型调参

载入预处理的数据之后，先进行一次概率为0.5的dropout, 减少参数减少计算量，防止过拟合，然后直接连接输出层，激活函数为Sigmoid，优化器为Adadelta，输出一个零维张量，表示某张图片中有狗的概率。

4. 模型评估

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

5. 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

6. 模型调优

- 训练时，使用交叉验证，打印acc和loss；观察训练结果，使用测试集验证结果；要提高模型效果，可以对训练集进行数据增强，尽量剔除异常图片，排除一些图片对模型的学习掌握，使用更强大的算法来优化模型的表现，例如最近出来的EfficientNets网络等。大胆假设，小心求证。

参考文献

-
- [1] Karen Simonyan and Andrew Zisserman. VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE- SCALE IMAGE RECOGNITION. At ICLR,2015.
 - [2] [译] Deep Residual Learning for Image Recognition (ResNet)
 - [3] 手把手教你如何在Kaggle猫狗大战冲到Top2%: <https://yangpeiwen.com/dogs-vs-cats-2>
 - [4] Keras做图片分类（四）：迁移学习--猫狗大战实战: <https://zhuanlan.zhihu.com/p/51889181>
 - [5] Kaggle猫狗大战准确率Top 2%webapp部署: <https://www.jianshu.com/p/1bc2abe88388>
 - [6] Keras中文文档: <https://keras.io/zh/applications>
 - [7] 毕业设计 Dogs vs Cats For Udacity P7 (异常值检验): <https://zhuanlan.zhihu.com/p/34068451>
 - [8] 面向小数据集构建图像分类模型: https://keras-cn-docs.readthedocs.io/zh_CN/latest/blog/image_classification_using_very_little_data
 - [9] 杨培文 胡博强. 深度学习技术图像处理入门. 北京: 清华大学出版社, 2018 (2019.4 重印) .
 - [10] [美]Ian GoodFellow [加]Yoshua Bengio [加]Aaron Courville 著 赵申剑等人译. 深度学习. 北京: 人民邮电出版社, 2017.8 (2017.12 重印) .