

机器学习工程师纳米学位

猫狗大战项目 开题报告

yijigao Udacity

2018/08/19

项目背景

从图像中判断一张图是猫还是狗，似乎是一个非常容易的问题，恐怕三岁小孩就能做到。让人失望的是，人类拥有最先进的机器和计算设备，依然在图像识别方面表现非常吃力[1]。

所幸的是，随着近些年处理器、图形处理器的性能的飞速提升，带来了机器学习、深度学习领域的飞速发展，计算机已经拥有足够的图形处理性能，可以通过大量的训练标记数据，经过深度学习算法，开始逐步“学会”去识别图片。这也是本项目的目标，让计算机“学会”判断一张图上的动物是猫，还是狗。这是图像识别分类问题。目前主流的图像识别分类方法是采用深度学习卷积神经网络（Convolutional neural net,CNN）[2]。卷积神经网络通过卷积层和池化层来实现对图像特征的提取和筛选。目前流行的深度学习框架有TensorFlow、Caffe、Keras。

问题描述

本项目需要解决的问题是从12500张包含猫或狗的图像中，对图片进行猫和狗分类区分。所以这是一个二分类问题。对于给定的图片，设计算法判断图片中的动物是猫还是狗。

- 输入：一张包含猫/狗的图片
- 输出：图片是猫还是狗

输出值是图片为狗的概率

数据或输入

数据集来源是[Kaggle][2]，训练集是25000张包含猫或狗图像，每张图像都已经在文件名标注好了猫/狗。猫狗图片数量各占一半。而测试集则是12500张未标注猫狗属性的图像。



cat.4490.jpg



cat.4491.jpg



cat.4492.jpg



cat.4498.jpg



cat.4499.jpg



cat.4500.jpg



cat.4506.jpg



cat.4507.jpg



cat.4508.jpg

从训练集中的大部分图来看，照片多为日常拍摄，清晰度不错，人眼能很好的识别出猫和狗。当然也有部分图片是由人抱着动物拍的，而且图中的猫狗太小，很难分辨，这种图片属于“脏数据”，可能需要想办法清洗。另外，也需要注意到图片尺寸各不一样，高度和宽度都从几十到500以上，这对于数据训练来说是不利的，需要预处理使输入尺寸一致。



解决方法描述

1. 先对数据集进行简单的数据探索分析，数据清洗（如果有必要）。目的是对数据集有个初步了解。
2. 数据预处理，标准化，独热编码
3. 构建深度学习神经网络
4. 使用创建的模型训练神经网络，根据结果多次优化模型。
5. 使用最终训练好的模型预测测试集，提交结果

评估标准

模型评估标准是交叉熵LogLoss函数，log loss 越小表示神经网络对于数据集有着较好的分类效果。

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

其中

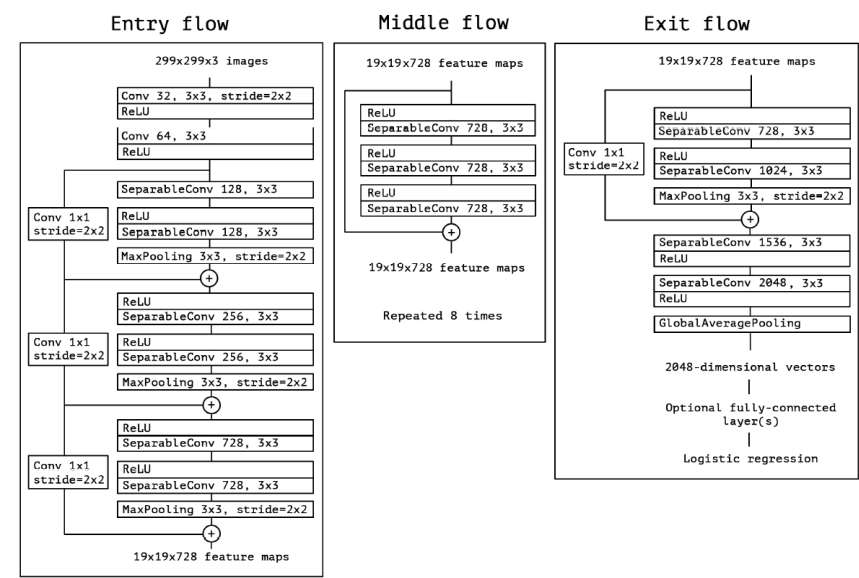
- n 表示测试集的图像数量
- \hat{y}_i 表示图像是狗的概率
- y_i 1 代表图像是狗, 0 表示是猫
- $\log()$ 代表自然对数

基准模型

项目要求是达到kaggle top10%。目前Kaggle该项目Leaderboard一共1314参赛选手，第131名成绩是0.06127。因此基准得分必须小于0.06127。

本项目基础模型采用Xception[3], Xception 是一种轻量化模型，由Google 在2016年10月发表，Xception基于Inception V3[5], 其结构如下图所示，分为Entry flow, Middle flow, Exit flow，其中Entry flow 包含8层卷积，Middle flow 包含24层卷积，而Exit flow 包含4层卷积，共计36层。

Xception的优势是，在给定的硬件资源下，可以尽可能的增加网络效率和训练性能。



项目设计

为了提高准确率，采用迁移-融合学习的方法。迁移学习是一种预训练的方法，使用已有的能够完成特定任务的模型进行修改，将这一模型应用与项目的数据集，完成相似的任务的过程。流程简要如下：

- 数据预处理
 - 按4: 1拆分训练集和验证集，并且都按猫狗用文件夹分开
 - 使用PIL模块处理图片尺寸，先缩放后使图片尺寸统一（统一大小为299 * 299（Xception默认尺寸））
 - 标准化RGB像素值
- 构建模型
 - 构建Xception基础模型，优化器选用 `adadelta`，`batch_size=16`，`fit_generator`参数 `steps_per_epoch=20000//16`，`epochs=5`，`validation_data=validation_generator`，`validation_steps=5000//16`，
 - 训练数据，根据验证集Loss调整参数，保存最佳模型
 - 使用最佳模型预测测试集，获得基础模型的准确率（使用Xception模型跑出得的kaggle得分为0.11315，开放97层以上权重优化后，`val_loss=0.0360`，`val_acc=0.9920`，kaggle得分0.04664）
- 迁移-融合模型
 - 尝试从ResNet50[4]、Xception、InceptionV3[5]、DenseNet201[6]、VGG19[7]等多个候选模型选取若干个构建融合模型
 - 使用融合模型预测，对比基础模型
- 提交结果

参考文献

- [1] : 李飞飞: 如何教计算机理解图片. http://open.163.com/movie/2015/3/Q/R/MAKN9A24M_MAKN9QAQR.html
- [2] : Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.
- [3] : Chollet, François. "Xception: Deep learning with depthwise separable convolutions." arXiv preprint (2017): 1610-02357.
- [4] : He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778. 2016.
- [5] : Szegedy, Christian, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. "Rethinking the inception architecture for computer vision." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818-2826. 2016.
- [6] : Huang, Gao, Zhuang Liu, Kilian Q. Weinberger, and Laurens van der Maaten. "Densely connected convolutional networks." In Proceedings of the IEEE conference on computer vision and pattern recognition, vol. 1, no. 2, p. 3. 2017.
- [7] : Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).