

机器学习（进阶）纳米学位

Capstone Proposal

张宏海

2018年4月8日

Proposal

项目背景

在本次毕业项目中，选择的项目是“猫狗大战”。该项目是通过输入一张图片，图片内容为猫或者狗，随后输出“猫”或者“狗”。该项目之所以能够实现，或者说更广的层面，计算机视觉分类之所能实现，主要是因为计算机经过多年的发展，在有足够多的数据去进行模型训练的同时，还有足够强大的硬件、算力在底部做支持。

同时，在计算机视觉方面，也有很多相关的学术论文、理论模型。这些论文及模型，都能大大降低本项目的实现难度。在本项目中，计划使用VGGNet模型进行构建。

问题描述

本项目主要属于二分类问题。通过图像识别，对图片进行分类，最终分为“猫”、“狗”。

数据或输入

在本项目的数据中，猫狗的图片涉及场景非常丰富，草地、笼子、马路，还有一部分是跟人合照（如图1），并且部分图片是复数的动物同时出现（如图2）。同时，还发现部分数据存在异常值（如图3）。



图1.源自训练集cat.26



图2.源自训练集cat.712



图3.源自训练集dog.2877

由于项目提供的‘test’集是没有标签的，需要上传预测才能知道结果。为了能够更方便的验证训练模型的结果，还应当将训练集拿出20%的数据作为验证集使用。并且由于图片的尺寸不一，在预处理环节可以通过先将图片转成一致的尺寸，从而保证输入尺寸相同。

解决方法描述

通过卷积神经网络，在卷积层，通过权值矩阵，从图像中提取特征。同时为了减少训练参数的数量，加快训练的进程，通过池化层来减少图像的空间大小。最后再通过输出层，输出结果。

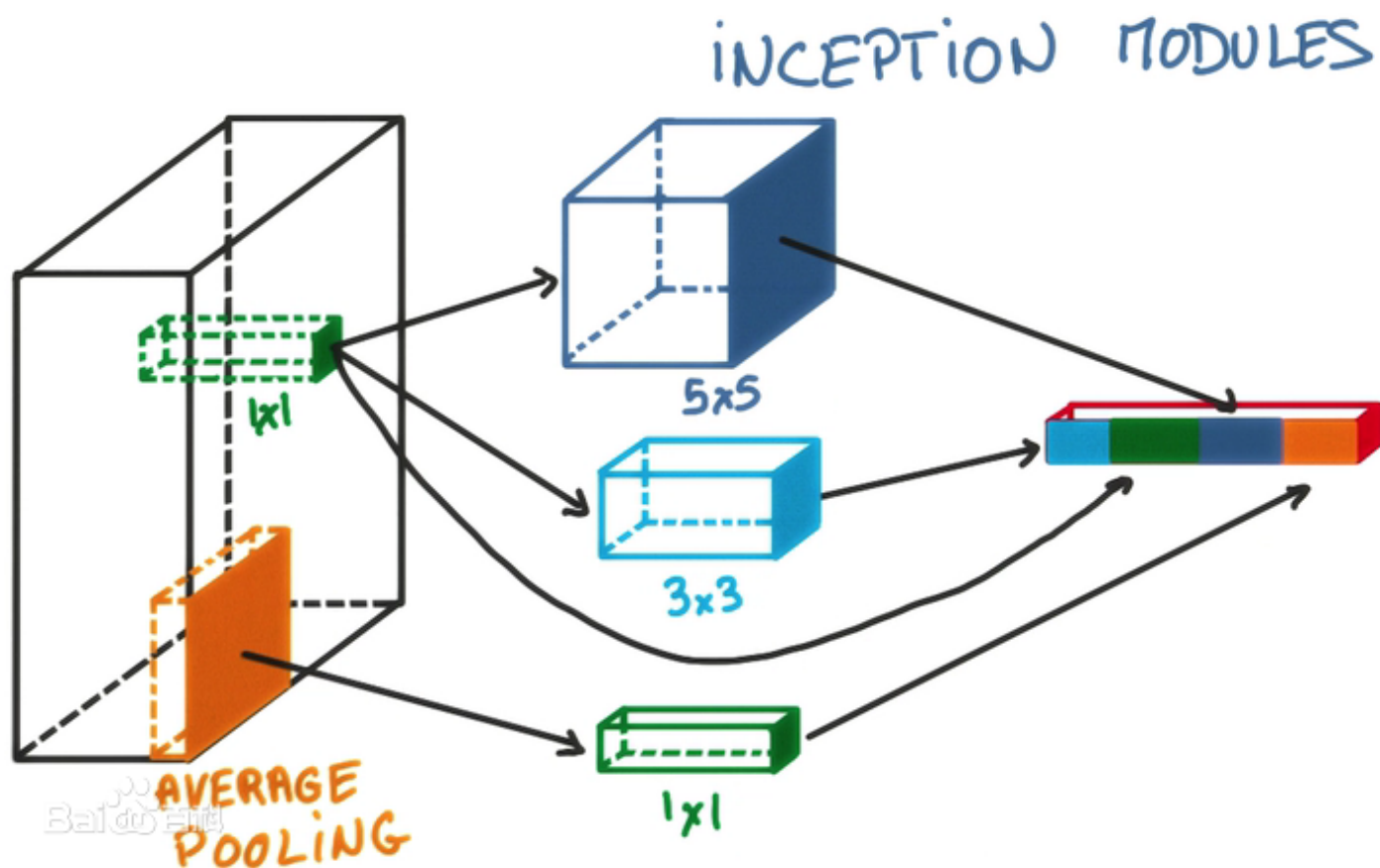


图4.卷积神经网络示意图

评估标准

通过 *LogLoss* 对模型进行评估，如下：

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

其中：

- n 为在测试集的图片数量
- \hat{y}_i 是该图片为“狗”的概率
- y_i 的值为”1“时，则图片为”狗“，”0“时为”猫“
- $\log()$ 的底为自然对数 e

LogLoss 的值越小越好。

基准模型

参考Kaggle排行榜前10%，也就是在“Public Leaderboard”上的 *LogLoss* 要低于0.06127。

项目设计

- 数据预处理
 - i. 调整图片大小（resize），将所有图片调整成尺寸一样的图片；
 - ii. 在读取图片是，同时获取图片文件名，生成label，狗为1，猫为0；
 - iii. 打乱数据，由于所有样本是按照顺序排列的，故应先将其打乱；
 - iv. 分割'train'集，将'train'中的数据按照2:8，分为验证集以及训练集。
- 模型搭建

使用Keras搭建VGG16卷积神经网络模型，结构如下的模型D：

Table 1: **ConvNet configurations** (shown in columns). The depth of the configurations increases from the left (A) to the right (E), as more layers are added (the added layers are shown in bold). The convolutional layer parameters are denoted as “conv<receptive field size>-<number of channels>”. The ReLU activation function is not shown for brevity.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Table 2: **Number of parameters** (in millions).

Network	A,A-LRN	B	C	D	E
Number of parameters	133	133	134	138	144

图5.VGG系列结构图解

其中：

- conv3-64 指的是卷积核的尺寸为 3*3，同时卷积核深度为 64的卷积层，其他以此类推
 - maxpool 表示最大池化层
 - FC-4096 表示神经元个数为 4096的全连接层，其他以此类推
- 模型训练及调参
 - i. 训练模型，根据验证集的 loss，来对模型进行调参优化；
 - ii. 在调参过程中，获取到满意的结果后，尝试训练完整的训练模型。
- 模型评估

- i. 完整训练的模型后，输入测试集，测试集的分评估模型。
- 可视化
 - i. 将准确率、loss进行可视化。

参考资料

[1] Karen Simonyan, Andrew Zisserman. Very deep convolutional networks for large-scale image recognition.