

# 机器学习（进阶）纳米学位毕业项目开题报告

## 猫狗大战

---

### 项目背景：

近年来，深度学习的发展突飞猛进，使用深度学习技术解决生活中方方面面的问题也成为可能。计算机视觉作为深度学习最重要的领域之一，应用场景非常丰富，可以用于制造业、检验、文档分析、医疗诊断，和军事等领域中各种智能/自主系统中。

该篇文章讨论的课题是：利用深度学习的方法区分猫和狗的问题，该问题是 kaggle 平台提出的一个比赛问题，旨在使用机器学习，特别是深度学习来实现分类问题。深度学习的方法可以有效提取出图片的特征，并对不同类型的分类进行概率输出，故有很高的可行性。而且解决分类问题，对于各个领域的智能识别都具有重要的意义。

---

### 问题描述：

猫狗大战的项目主要是解决猫/狗两个种类的分类问题，即给定一张图片，模型输出图片上的动物分别是猫和狗的概率。

实现方法：使用卷积神经网络的方法建立模型，顶层分别输出猫和狗的概率。

---

### 数据及输入：

本项目的全部数据均由 kaggle 平台的猫狗大战项目提供（下载地址：<https://www.kaggle.com/c/dogs-vs-cats-redux-kernels-edition/data>）。其中训练数据集中分别包含了 12500 张猫和狗的图片，格式为 jpg，每张图片使用‘类别’+‘序号’的格式进行命名（如‘cat.0’）。测试集有 12500 张图片，格式为 jpg，以序号命名（如‘23’）。

图片的大小不定，需要在训练模型时根据实际情况进行剪裁。输入模型时可以采用 keras 提供的图片预处理的方法——ImageDataGenerator，对图片进行实时增强，创造出水平翻转、旋转、缩放等操作，达到更好地泛化模型的效果。

---

解决方法描述：

确定使用卷积神经网络的方法来解决该项目的问题，可以采用以下两个方法：

1. 通过建立卷积神经网络，对图片的特征进行提取，在模型的顶层添加一层全连接层，两个输出节点，分别输出图片对于猫和狗的预测概率。
2. 通过迁移学习的方法进行模型搭建。利用 keras 提供的预训练模型如 VGGNet/Xception 等结构进行特征提取，去掉预训练模型的输出层，添加符合该项目的全连接层，以输出概率。

其中方法 1 在训练的过程中会更耗费时间，因为所有的权重都是从随机权重开始训练。方法 2 利用在 imageNet 的大数据集上训练的模型作为基础来建立的模型，可以只对顶层的全连接层进行训练，训练难度降低。综合考虑，选择方法 2。

---

评估标准：

根据 kaggle 官方的标准，应该尽量降低 log loss 的值，计算公式如下：

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

where

- $n$  is the number of images in the test set
- $\hat{y}_i$  is the predicted probability of the image being a dog
- $y_i$  is 1 if the image is a dog, 0 if cat
- $\log()$  is the natural (base e) logarithm

此外，还可以判断验证集准确度是否收敛于某一个值，并且这个值尽量接近 100%

---

基准模型：

kaggle Public Leaderboard 前 10%，即最低应该达到 0.06127

---

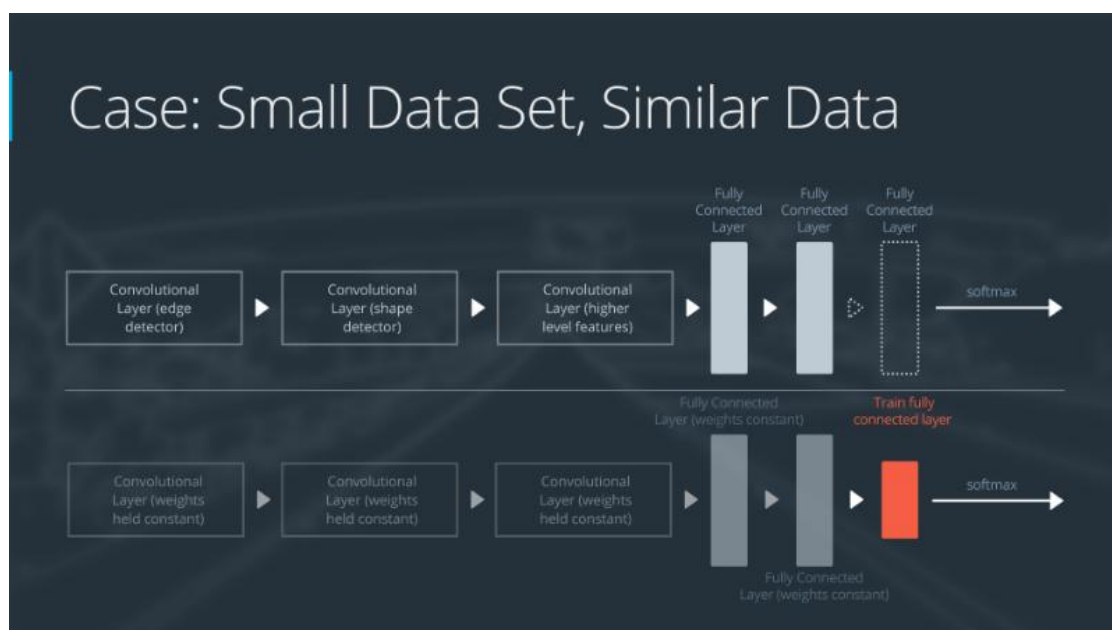
项目设计：

步骤 1：图片预处理。

- 剪裁：训练集中的图片的大小不定，故将图片进行剪裁到固定大小，如采用 Xception 进行迁移学习，则将图片剪裁到（299\*299）。
- 数据增强：通过水平翻转、旋转、缩放对图片进行实时增强。
- 归一化：将图片的每个像素值变成 0-1 之间的浮点数

## 步骤 2：模型建立。

- 采用迁移学习的方法，使用预训练模型，加载预训练权重并去掉顶层的输出层，添加符合该项目的输出层。（由于预训练的模型和该项目的分类数据属于相似数据，且该项目数据属于小数据集，所以只训练顶层权重即可）



（图片来自‘机器学习（进阶）纳米学位课程课件’）

## 步骤 3：模型评估

可视化训练过程中的训练损失、验证损失、验证集准确率的变化趋势。

抽样可视化预测结果和实际图片的对应情况，观察是否一致。

根据‘评估标准’中的公式计算 log loss 的值。