

# Machine Learning Engineer Nanodegree

## Capstone Proposal

---

毛世杰 2017年3月30日

## Proposal

---

### Domain Background

计算机视觉是机器学习的一个很重要的应用领域。计算机视觉是一门研究如何使机器“看”的科学，更进一步的讲，就是指用摄影机和计算机代替人眼对目标进行识别、跟踪和测量等机器视觉。作为一个科学学科，计算机视觉研究相关的理论和技术，试图建立能够从图像或者多维数据中获取‘信息’的人工智能系统。这里所指的信息指Shannon定义的，可以用来帮助做一个“决定”的信息。因为感知可以看作是从感官信号中提取信息，所以计算机视觉也可以看作是研究如何使人工系统从图像或多维数据中“感知”的科学。

图像识别 (image recognition) 是计算机视觉的一个领域。我们的大脑做图像识别是很容易。人类可以很容易的区别一只狮子和一只美洲虎，识别一个标志，或认出一个人的脸。但对计算机视觉来说这些都是一个难题。近几年机器学习在解决这些困难问题上取得了巨大进步。特别地，我们发现一种称为深度卷积神经网络 (CNN) 的模型在困难的视觉识别任务上有很好的性能。研究人员不断地发明新的CNN模型。从早期的 [LeNet](#), [AlexNet](#) 到最新的 [Inception v4](#), [ResNeXt](#), 这些模型在ImageNet数据集上取得了越来越好的正确率。

Dogs vs. Cats最早来源于Kaggle的一个竞赛。现在已经演变为流行的图像识别初学者练习项目。

### Problem Statement

在Dogs vs. Cats这个项目中。我们需要使用深度学习方法识别一张图片是猫还是狗，并尽可能的提高识别的精确度。整个图片数据集中只包含猫和狗。

### Datasets and Inputs

此数据集可以从kaggle上下载。 [Dogs vs. Cats](#)

训练数据包含25,000张猫和狗的图片，每张图片的文件名标示了这是猫还是狗。测试数据包含12,500张猫和狗的图片。

数据集包含两个文件夹train和test：

- train文件夹中文件名为[dog|cat].n.jpg，文件名标示了样本为dog和cat。

- test文件文件名为n.log，没有样本的类别信息。

dog.2.jpg



cat.5.jpg



数据集中图片的分辨率，清晰度等规格都有所不同，需要考虑数据集的预处理。

另外还有一个数据集[The Oxford-IIIT Pet Dataset](#)，可以作为训练数据的补充。

- 这个数据集中有37种宠物类别的图片，在这个项目中，只考虑使用其中的Dog和Cat类别的数据
- 每个类别大概有200张图片
- 这些图片的分辨率，拍摄角度和光照等都有很大差异

## Solution Statement

项目会使用训练数据训练一个CNN分类器，用这个分类器预测一张图片是猫还是狗：

- 分类器输入：一张彩色图片
- 分类器输出：图片是狗（1）的概率  $p$
- 可选输出：猫狗面部坐标，猫狗身体mask

## Benchmark Model

Benchmark: [Golle, Philippe. "Machine learning attacks against the Asirra CAPTCHA." Proceedings of the 15th ACM conference on Computer and communications security. ACM, 2008.](#)

Golle, Philippe构建了一个SVM分类器。在Asirra CAPTCHA上识别猫和狗，取得了82.7%的精确度。

## Evaluation Metrics

项目使用测试集合上的准确率和logloss作为最终的评价指标：

总体的准确率(accuracy) = 分类正确的样本数 / 总体的样本数

总体的logloss =  $-(y_t \log(y_p) + (1 - y_t) \log(1 - y_p))$ ，其中 $y_p$ 为预测为狗（1）的概率， $y_t$ 属于{0, 1}

## Project Design

项目首先是预处理图片数据，以方便模型训练和结果验证。然后构建多个深度学习模型（AlexNet，Inception等）。使用训练数据训练各个模型。在测试数据上验证模型的精确度；对比各个模型的效果。最后迭代修改模型，尽可能地提高模型精确度。

### 数据预处理

- 处理图片数据，调整图片为统一的大小，并保存为H x W x 3的多维数据，方便后续使用。

### 模型构建

- 参考AlexNet，Inception V4，使用keras构建CNN深度神经网络模型分类器

### 数据集准备

- 训练数据80%用作训练，20%用作交叉验证，
- 在GPU云主机上训练数据
- 调整模型参数，直至收敛

### 评价模型精确度

- 在测试集上评价预测结果
- 计算模型总体精确度和在猫、狗类别上的精确度