

机器学习毕业项目开题报告

项目背景

本项目是来自 Kaggle 过往的 Dogs vs Cats 图像识别比赛。该项目的研究属于深度学习中的计算机视觉类问题。

关于计算机视觉科学的研究，最早追溯到 1963 年美国计算机科学家拉里·罗伯茨 (Larry Roberts) 在 MIT 的博士毕业论文《Machine Perception of Three-Dimensional Solids》，该论文尝试了把输入的二维图像转化成三维模型，然后从另一个角度生成渲染图。所以这算是计算机视觉相关的最早研究。

到了 20 世纪 70 年代，英国神经生理学家和心理学家戴维·马尔 (David Marr) 在 MIT 人工智能实验室任教时期出版的《视觉计算理论》(Vision: A Computational investigation into the human representation and processing of visual information) 一书提出“人类视觉的主要功能是通过大脑进行一系列的变换，来复原真实世界的三维场景，并且这种神经系统里的信息处理过程是可以用计算的方式重现”。由此，开启了计算机视觉作为一门正式学科的研究。

自 2010 年起，由斯坦福大学华人教授李飞飞创办的一个非常庞大的图像数据库 ImageNet 每年都举办一次大规模的视觉识别竞赛 (ImageNet Large Scale Visual Recognition Challenge, ILSVRC)，使计算机视觉的研究与应用受到越来越多的关注。在 ILSVRC 的头两届比赛，主要的算法框架都是使用“手工设计特征+编码+SVM”这种模式进行。直到 2012 年，一名研究生 Alex Krizhevsky 提出了一个 5 个卷积层+2 个全连接层的卷积神经网络 AlexNet，以高于第二名 26.2% 的正确率胜出了比赛，从而促使了计算机视觉的相关研究从之前的 SVM 算法全面转向了深度神经网络算法。

得益于当今互联网技术以及各式影像传感器的发展，影像数据的膨胀已经到了人类无法消化的阶段。因此，只能寄望于计算机视觉技术的发展去消化这些影像数据。计算机视觉技术发展的意义就犹如亿万年前地球原始生物的眼睛进化，将会使得计算机变得更加智能，从而对人类的发展产生不可划缺的深远影响。

问题描述

本项目是通过 CNN 模型构建一个猫狗识别应用，实现手机端对猫狗进行实时识别。最终根据应用的内核模型去对 Kaggle 的 dogs vs cats 测试数据集进行预测，并实现跻身历史排行榜的前 10%。

输入数据

数据来自 Kaggle 的 [Dogs vs Cats](#) 比赛。

分析训练数据集可知，总图片数量为 25000 张，其中 12500 张为猫图，12500 张为狗图。图像大小不等；图像为彩色三通道图像。

解决办法



本项目会利用一个成熟的模型作为基准模型进行迁移学习，然后再在基准模型的基础上进行优化，优化的思路有：数据增强、正则化、模型融合。当内核模型成熟后，把内核整合到手机端，实现手机端实时识别功能。

基准模型

本项目将采用 Inception-v4 基准模型进行迁移学习。

评估指标

本项目属于二分类模型，因此损失函数为：

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

从公开的排行榜可见，最优的 loss 值为 0.03302；而排行榜前 10% 的最差 loss 值为 0.06114。本项目的目标是超过最差 loss 值并挑战最优 loss 值。

设计大纲

1. 数据的转换预处理：主要对图片进行缩放处理统一图片尺寸；
2. 基准模型的构建：使用 Inception-v4 进行迁移学习；
3. 基准模型结果评估与可视化：使用基准模型进行结果预测；
4. 数据预处理优化：把图片进行数据增强；
5. 模型结构及参数优化：尝试其他 CNN 模型；正则化处理；

6. 检验并迭代步骤 4、5；
7. 得到最终模型：生成最终预测并形成内核模型；
8. 做成手机端应用：使用 openCV、Flask 等工具把内核模型整合至手机端。

参考文献

- 《深度学习与计算机视觉》，叶韵，机械工业出版社
- CS231n Winter 2016: Lecture1: Introduction and Historical Context