

优达学城机器学习（进阶）毕业项目开题报告

Kaggle: Dog Breed Identification

1. 项目背景

本项目是 Kaggle 近期上线的一个 PlayGround 项目。训练数据来自于 ImageNet 关于狗品种的子集。通过这个项目可以让参与者了解图像识别算法，找到自己正在开发或者学习的算法中的问题^[1]。

本项目属于图像识别领域新出现的一个研究方向，细粒度视觉分类(Fine-Grained Visual Categorization, FGVC)，其目的是将某一大类事物的子类进行分类。本项目中将狗这一大类事物，识别图片中的狗属于 80 个品种中的哪一种。

2. 问题描述

本项目的具体问题是识别 10357 张狗的图片，识别出每张图片中的狗是属于 120 个品种的哪一个品种。

3. 输入数据

本项目的输入数据包括三个部分：

- 1) Keras 预训练模型的权重，计划使用 InceptionV3 这 1 个模型^[2]。
- 2) 训练数据包括，10222 张已标记的彩色图片。
- 3) 测试数据包括，10357 张未标记的彩色图片，这些图片对应的分类的结果将作为输出，提交 Kaggle，验证最终算法的正确率。

4. 解决办法

本项目的主要方法由文章《面部识别技术能用来识别鲸鱼》^[3]启发，先定位要识别的目标，然后再进行识别。因为 YOLO V2^[4]提供了对于狗的识别，所以，可以不用人工的标记数据，实践证明这是非常方便的。本项目中，使用 YOLO V2 裁剪图片中的狗所在的矩形区域。然后，使用预训练 InceptionV3 提取特征。最后，将这些特征输入一个多层神经网络训练，得到模型。^{[5][6][7][8][9]}

5. 基准模型

将使用 InceptionV3 作为基准模型，代表最基本的图片特征提取方法，并以此模型获得的 Score 作为基准模型的 benchmark，约为 0.39893。也可以选取其他的 Keras 提供的预训练模型。但是，因为本文进行的是对照组实验，所以使用相同的模型是必须的。

6. 评估指标

评价指标，选用 Kaggle 官方的 Multi Class Log Loss^[10]。

7. 设计大纲

本项目的解决方案将主要会先使用 YOLO V2 提取图片中的狗区域。然后，使用 InceptionV3 提取特征。最后，将这些特征输入一个多层神经网络训练，得到模型。与只用 INceptionV3 提取特征对比。本项目表明，先提取图片中有效区域，然后提取特征，识别效果更好。

引用：

- [1] <https://www.kaggle.com/c/dog-breed-identification>
- [2] <https://github.com/fchollet/deep-learning-models/releases>
- [3] 面部识别技术能用来识别鲸鱼，
https://mp.weixin.qq.com/s?srcid=0717jR0shJpsHITF5EjsY84z&scene=1&mid=2651651403&sn=f0360e5cdcabf938fd8139a18c9d557f&idx=4&__biz=MjM5MTQzNzU2NA%3D%3D&chksm=bd4ddad88a3a53ce9e84c03dbce03674df9e7e6b36cf6221dce4ed0d0e07222205a98c42282a&mpshare=1#rd
- [4] <https://arxiv.org/abs/1612.08242>
- [5] <https://www.kaggle.com/gaborfodor/use-pretrained-keras-models-lb-0-3>
- [6] https://github.com/ypwhs/dogs_vs_cats
- [7] Rethinking the Inception Architecture for Computer Vision, <https://arxiv.org/abs/1512.00567>
- [8] <https://keras.io/applications/>
- [9] <https://keras.io/>
- [10] <https://www.kaggle.com/wiki/MultiClassLogLoss>