

Capstone Project

狗狗分类器

By Yaping

2020/10/8

背景

图像分类是计算机视觉中最基础也是最常见的问题，也是几乎所有的基准模型进行比较的基准。深度学习模型的发展史就是图像分类任务性能提升的发展历史，在 **Imagenet** 这样超过 1000 万图像、2 万类的数据集中，计算机的图像分类水准已经超过了人类。但图像分类并不像你想象的那么简单，也没有被完全解决。

项目说明

在我们这个项目中，我将带领你设计一个狗狗分类器，使其在任何时候输入一副图像，都能分辨出是不是狗狗，并输出狗狗的种类（有 133 个狗狗品种）。有趣的是，如果输入一个人类的照片，这个分类器可以辨识出这个人和哪个品种的狗狗相似呢！在这个项目中，我们主要是使用深度学习模型（**Deep CNN**）来搭建算法，并且会应用迁移学习来提高算法的效率。在算法设计过程中，我们会从最简单的模型开始搭建，逐渐提高算法的复杂度，来提高预测的准确性。



图 1 右图为算法输出示例

项目分析

1. 数据探索

本次使用的数据有两类：一类是 133 个品种的狗狗图片，一共 8351 张；一类是人类的人脸图片，两类图片都是直接从优达学城 Github 上给出的链接直接下载下来的。狗狗数据集直接进行了训练集、验证集和测试集的划分，具体数目如下：

```
There are 133 total dog categories.  
There are 8351 total dog images.  
There are 6680 training dog images.  
There are 835 validation dog images.  
There are 836 test dog images.
```

预测小狗的品种并不简单，是一个非常有挑战性的工作。因为有些狗狗的品种非常相似，即使人类要分辨也非常困难，比如布列塔尼猎（Brittany）和威尔斯激飞猎犬（Welsh Springer Spaniel）。



图 2 相似品种的狗狗

更有挑战的是，有的狗狗品种差异很大，即使是同一个品种，也存在不同的颜色和形态，比如：拉布拉多有黄色、巧克力色和黑色品种。基于视觉的算法需要克服这种同一类别差异很大的问题，并决定如何将所有这些不同肤色的小狗分类为相同的品种。



图 3 不同毛色的同一品种狗狗

在了解了数据集的基本情况之后，观察图片的大小并不相同，这对后续的图像处理显然有不利的影响。因此，我们要先对图片进行预处理，将图片缩放成同一个尺寸。

2. 搭建模型

1) 自己搭建 CNN 模型

Keras 提供了非常方便的模块，可以让我们按照自己的想法随意搭建 CNN 模型，一般的 CNN 网络是一些卷积、非线性、池化层或者 dropout 的堆积。卷积的目的是以小的代价（模型参数）来获取图像特征（可以参数共享），池化层主要是进一步减少图像尺寸，有一

定的预防过拟合的作用，relu 经常被用来做非线性函数，而 dropout 通过随机关闭一些网络节点来达到过拟合的效果，加入 BatchNormalization 也可以起到加速计算的作用。网络最后会通过一个全链接层来实现分类，全链接层的输出数量就是类别数。

图 4 的框架基本包含了 CNN 的一些基本模块，并进行叠加，可以实现权值共享、减少参数、预防过拟合等作用，是一个基本的 CNN 分类网络。

| Layer (type) | Output Shape | Param # | |
|---|----------------------|---------|-------|
| conv2d_1 (Conv2D) | (None, 223, 223, 16) | 208 | INPUT |
| max_pooling2d_1 (MaxPooling2D) | (None, 111, 111, 16) | 0 | CONV |
| conv2d_2 (Conv2D) | (None, 110, 110, 32) | 2080 | POOL |
| max_pooling2d_2 (MaxPooling2D) | (None, 55, 55, 32) | 0 | CONV |
| conv2d_3 (Conv2D) | (None, 54, 54, 64) | 8256 | POOL |
| max_pooling2d_3 (MaxPooling2D) | (None, 27, 27, 64) | 0 | CONV |
| global_average_pooling2d_1 (GlobalAveragePooling2D) | (None, 64) | 0 | POOL |
| dense_1 (Dense) | (None, 133) | 8645 | GAP |
| Total params: 19,189.0 | | | DENSE |
| Trainable params: 19,189.0 | | | |
| Non-trainable params: 0.0 | | | |

图 4 CNN 模型框架

2) 使用迁移学习 (Transfer learning) 模型

简单使用自己搭建的 CNN 框架来训练模型，预测结果并不使人满意。为了在不牺牲准确率的情况下减少训练时间，我们将使用迁移学习的方法训练 CNN。

迁移学习是就是把已经训练好的模型参数迁移到新的模型来帮助新模型训练的方法。由于大部分数据或任务是存在相关性的，所以通过迁移学习我们可以将已经学到的模型参数（也可理解为模型学到的知识）通过某种方式和新模型共享，从而加快并优化模型的学习效率，这是一个被业界证实了的非常好的方法。迁移学习主要有三种形式：

A.冻结预训练模型的全部卷积层，只训练自己定制的全连接层。

在我们这个项目中，有四个迁移模型：VGG-19，ResNet-50，Inception，Xception。这四个模型都是在 IMAGENET 训练集上已经训练好的。我们把卷积层以后的部分去掉，加上全新的未训练的层（相当于把卷积的部分保留并冻结，重新训练分类的部分），然后用我们提供的新的训练集进行二次训练。相比直接在我们的训练集上训练一个全新的模型，迁移学习节省了大量的计算成本，同时因为 IMAGENET 数据集足够强大，可以导致更好的训练效果。

B.抽取特征向量 (Extract Feature Vector)：先计算出预训练模型的卷积层对所有训练和测试数据的特征向量，然后抛开预训练模型，只训练自己定制的简配版全连接网络。

在迁移学习中，迁移的模型（如 ResNet-50）本身就具有大量的参数，即使是进行迁移学习，也需要很强大的计算资源和计算时间（即使是使用 GPU 也需要数小时、数天甚至数周的时间）才能训练的这么深的模型。即使把所有的层都冻结，那么在训练过程中每次更新梯度时，训练集都要和模型中的所有参数进行计算（大量的矩阵相乘），而即使这些参数是定死的、不需要进行更新，这个过程也是非常费时间的。而如果你是使用 CPU 在进行这一步，无疑会需要更多的时间。

抽取特征向量的做法是，把训练集经过预训练模型生成出 bottleneck features，然后直接通过 bottleneck features 进行训练。这种方法相当于是将整个模型拆分成了两个部分：第一步是将所有图片通过 ResNet-50 的卷积结构，将数据“编码”成 bottleneck features；第二步则是用这些 bottleneck features 训练我们后加的新的结构（分类器）。

本项目采用的就是这种方法，因为参数冻结的部分本身就不需要更新，也不需要回传的参数，这种方法和直接训练一个大型的迁移学习网络是差不多的。而这种做法可以省去每次更新过程中和原先模型中参数进行的大量矩阵运算，从而训练的速度就大大加快了。

C. 微调（Fine-tune）：冻结预训练模型的部分卷积层（通常是靠近输入的多数卷积层），训练剩下的卷积层（通常是靠近输出的部分卷积层）和全连接层。

Fine-tune 的形式下分不同程度的解冻原有层参数，甚至可以解冻所有层。实际上，预训练模型的每一层都可以自定义解冻，进行二次训练。相比冻结所有预训练模型卷积层，Fine-tune 可以学到更多的特征知识，可以带来更好的效果，这种方法也是现在迁移学习中最为常用的做法。但是 Fine-tune 的代价就是需要大量的计算成本，包括计算时间和计算性能，因此对于计算资源有限的个人来讲，需要谨慎考虑。

3. 模型结构和训练

1) 模型选择

在本项目中，我首先选择了 ResNet-50。ResNet 对于解决分类问题已经有非常出色的表现，ResNet 的提出是 CNN 图像史上的一件里程碑事件，ResNet 在 ILSVRC 和 COCO 2015 上取得了 5 项第一，并又一次刷新了 CNN 模型在 ImageNet 上的历史，因此非常适合解决目前的问题。

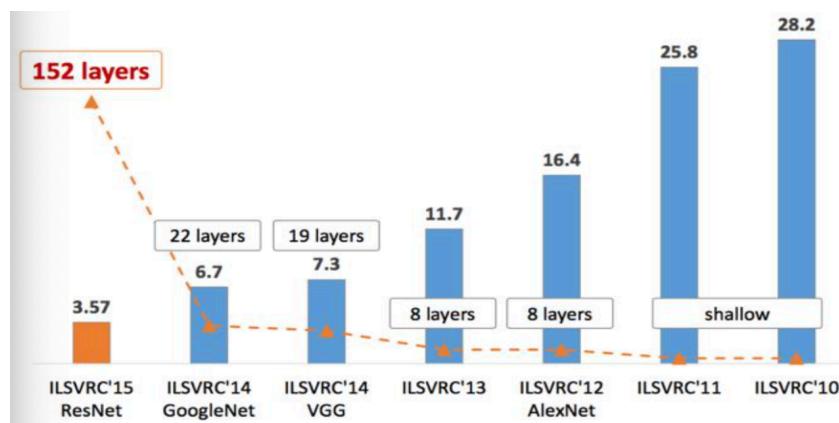


图 5 ILSVRC 不同模型错误率

ResNet 是一个非常有代表性的深度卷积网络 (DNN)，其网络参考了 VGG19，并在其基础上进行了修改，最重要改进是通过短路机制加入了残差单元，残差单元极大地解决了深度 CNN 模型难训练困难的问题。其他改进还体现在 ResNet 直接使用 stride=2 的卷积做下采样，并且用 Global average pool 层替换了全连接层。ResNet 的一个重要设计原则是：当 Feature map 大小降低一半时，其数量就增加一倍，这保持了网络层的复杂度。

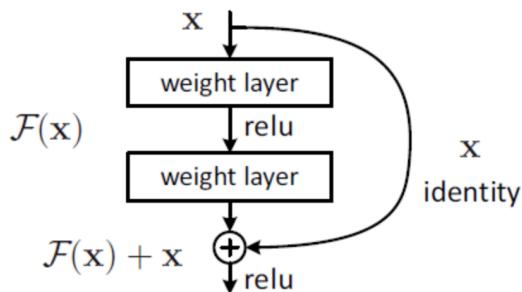


图 6 ResNet-50 残差结构示意图

2) 模型训练

在模型训练中，需要注意的主要问题有：

A. 数据预处理

数据预处理是一个非常重要的事情，数据一定要经过预处理，在送入模型训练时候才能有比较好的效果。

B. 参数的选择

开始训练的时候，我的模型结果非常差，差到完全不能接受，后来仔细观察了 loss 的结果，发现很早就不再下降了。这时候及时的调整了 learning rate，将 learning rate 的值减小，loss 就开始变化了。可见，调参是一个需要耐心的事情。

C. 内存的管理

跑深度学习的算法，最让人崩溃的就是机器跑不动了，有时候还会报错，根本无法运行。这时候一方面要调整自己训练的参数，比如：Batch size，数据格式等，一方面还要寻求更直接的解决方法，比如购买并安装合理配置的 GPU，或者通过远端使用云端的 GPU 等等。

D. 及时地保存结果

我遇到过调了两天的程序，不知道什么原因没有保存，前面的工作全部报废的情况。所以，我除了设置自动保存外，还会经常人工检查程序是否保存成功，并及时做好备份。通常在训练过程中，用模型检查点来保存验证损失最低的模型，以便后面使用，是非常明智的方法。

```
from keras.callbacks import ModelCheckpoint

checkpointer = ModelCheckpoint(filepath='saved_models/weights.best.Resnet50.hdf5',
                               verbose=1, save_best_only=True)

Resnet50_model.fit(train_Resnet50, train_targets,
                    validation_data=(valid_Resnet50, valid_targets),
                    epochs=20, batch_size=16, callbacks=[checkpointer], verbose=1)
```

图 7 检查点保存

3) 评价指标

本项目采用的评价指标是: loss='categorical_crossentropy', metrics=['accuracy']。

结果和结论

经过训练，我们的狗狗分类器可以达到输入任意图片，它可以判断图片中是否含有人、小狗或者两者都没有的效果。如果检测到了小狗，模型将进一步输出小狗的种类，如果检测到人脸，将会把与人脸最匹配的小狗品种输出来，否则会输出错误信息。

我们的分类器达到了 81% 的准确率，当然这不是最优的结果，我们还可以做进一步的优化和提升，如果你也有兴趣的话，就一起来改进吧！本项目的目的只是给你演示了用深度学习的方法，如何一步步根据自己的需要搭建出合理的模型框架，并输出可接受的结果。其中用到了一个很重要的概念就是迁移学习，文章的最后列出了很多很好的参考文献，如果你想更加深入地了解这些概念，可以花时间多读几篇文献，相信一定能给你不少的启发。

反思

我们搭建并测试了基于 ResNet50 的分类模型，但你可以看到，准确率并没有达到预想的效果，要想进一步提高准确率，我们还有很多工作可以做，总结来说包括：

1. **数据增强**: 增加数据集的数量，或者将原有数据集做一些剪裁或旋转，已增加图片的多样性。
2. **模型融合**: 综合各个不同的模型，从而得到不错的效果，这个方法非常常见和有效。
3. **进一步地调参**: 参数调整中深度学习中的地位举足轻重，可以进一步通过仔细观察训练结果的和校验结果的方法，更改学习率，Batch-size 等参数的方法进行微调。

参考文献

你可以参考以下材料来加深对本项目的理解:

- CS231n: Convolutional Neural Networks for Visual Recognition
- Using Convolutional Neural Networks to Classify Dog Breeds
- Building an Image Classifier
- Tips/Tricks in CNN
- Transfer Learning using Keras
- Transfer Learning in TensorFlow on the Kaggle Rainforest competition
- Transfer Learning and Fine-tuning
- Building powerful image classification models using very little data
- 简述迁移学习在深度学习中的应用
- 无需数学背景，读懂 ResNet、Inception 和 Xception 三大变革性架构
- [VGG16] VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION
- [Inception-v3] Rethinking the Inception Architecture for Computer Vision
- [Inception-v4] Inception-ResNet and the Impact of Residual Connections on Learning
- [ResNet] Deep Residual Learning for Image Recognition
- [Xception] Deep Learning with Depthwise Separable Convolutions