**基础题-猫狗二分类说明文档**

赵为之 人工智能与自动化学院 人工智能本硕博2101班 U202115195

**介绍**

在卷积神经网络(CNN)出现之前，主要采用传统的前馈神经网络进行图像分类任务。而提高性能的方法是收集更多的数据，增加更多的参数和使用技术防止过拟合。在Alex Krizhevsky， Ilya Sutskever， Geoffrey E。 Hinton发表的一篇论文中提出了卷积神经网络。其相比传统网络具有更少的参数，更好训练，而理论最好性能只差一点点。CNN的提出大大增强了人工智能进行图像分类的能力。

何凯明提出的深度残差网络(ResNet)在CNN的基础之上进一步提升了图像分类的性能。ResNet通过将原始输入和卷积层输出求和，再输入下一个卷积层，实现了卷积网络对残差的学习。它很好的解决了当层数增加时神经网络退化的问题，从而使网络的层数大大增加。进一步提升了图像分类的能力。

黄高在2017年又提出了DenseNet，它在ResNet的基础上进行了改进和优化。在DenseNet中，某一层的输入不仅和前一层有关，更和之前每一层都有关。这样的结构使得较早提取出的特征在深层仍可以被使用，而且减少了参数数量，使网络更容易训练还缓解了梯度消失和模型退化问题。其在相同参数情况下分类效果优于ResNet。

在本文中，作为一个新手，作者尝试使用传统CNN和DenseNet两种方法处理猫狗二分类问题。在训练过程中遇到了一系列的问题，一些已解决，一些有待改进。在学习过程中，收获了对计算机视觉和深度学习领域更深入的了解，激发了浓厚的兴趣，锻炼了自学习和动手能力。

**数据集构造**

本次实验共有猫，狗图片各10000张左右，并按照8:2的比例分为训练集和测试集两个集合。在读入数据时对训练集做了数据增强处理，具体处理方法为先进行随机缩放，再剪裁为224\*224大小的图片，并将图片标准化，最后按50%的比例随机水平旋转。

**模型结构**

CNN

输入的图像经过若干层卷积和池化，将图像变为128通道，14\*14大小，再将图像展开，经过两个全连接层输出1个数值，若该数值大于0。5则分类为dog，若小于等于0。5则分类为cat。使用的优化器为Adam，损失函数为交叉商。共训练10个epoch，前5个epoch学习率为0。0001，后五个学习率为0。00001。

DenseNet

整个网络分为卷积层和分类器。

卷积层由若干个block和transition组成。再一个block中有若干个卷积层，每个卷积层的输入为之前所有层的输出之和。由于每个block中图像的大小不能改变，故加入transition，通过池化减小图片大小。

分类器是一个全连接网络，最后输出一个n维向量(n为分类个数)。

使用的优化器为Adam， 初始学习率为0。0001，通过ReduceLROnPlateau动态调整学习率，损失函数为交叉熵。

在本次实验中，调用densenet121，并将其分类器改为两个全连接层，最后输出二维向量，第一维代表cat的概率，第二维代表dog的概率。

**训练过程**

本次实验共进行了3次训练，由于设备和时间的原因，训练不是很充分，尝试的次数也有不足。现得到以下初步结果。

CNN训练速度较快，但准确率较低，经过10个epoch的训练准确率在85%左右，由于水平有限，最终输出结果不是二维向量而是一个数值，不符合图像分类的标准，但对结果并无太大影响。由于时间原因没有再次训练。

DenseNet训练速度较慢，在笔记本上使用CPU训练需要14个小时，准确率在90%左右。第一次训练由于未能理解交叉熵和一些开源代码的误导，错误的同时使用softmax和交叉熵。后在学长指导下发现交叉熵本身已经包含softmax。修改损失函数后进行第二次训练。不过两次训练结果并无明显差别。可能是因为损失函数对训练影响不大，也可能因为训练时常不足没有体现出差别。

由于时间原因，第二次训练并未完成，只有8个有效epochs。所以最终模型请以第一次训练为准。

7月13日更新：经过又一轮训练，DenseNet准确率达到95%左右。

**代码说明**

train.py用于训练网络。

test.py用于测试准确率并生成csv表格。

**未来可改进之处**

由于ResNet相比DenseNet更加难以训练，本次未使用ResNet。但ResNet相较更为基础，可以在以后学习ResNet以稳固知识基础。也可以尝试自己构建神经网络而非调用现成网络，增强自身能力。

7月13日更新：可以考虑把多个网络融合以达到更高准确率，但这也需要更多计算时间。