机器学习工程师纳米学位毕业项目（算式识别）
开题报告

杨学冉

17 November 2018

研究领域背景

深度学习是机器学习领域中的重要分支，近几年深度神经网络在图像识别中取得重大突破。神经网络的发展给计算机视觉领域带来了新的解决方案，在众多神经网络中，CNN是应用最广泛的一种。随着计算机算力的提升，利用大数据训练的CNN网络在计算机视觉领域发挥了越来越重要的作用。特别是自2010年以来，MNIST、CIFAR、ImageNet等规范化的数据集被越来越多的用于图像分类和各种竞赛，催生了像ResNet、VGG16等众多被广泛运用的算法模型。相对于传统的计算机视觉算法，采用深度神经网络特别是卷积神经网络的算法实现了更好的成绩。

硬件方面，GPU的运算能力越来越强。Nvidia公司主推用GPU芯片做通用计算，并提供了CUDA工具包进行深度学习软件的开发，为深度神经网络相关的算法模型提供算力基础。

问题声明

算式识别项目要求使用深度学习算法来识别图像中的算式。算式中的字符种类是有限的，本质上，算式识别是对每张图片中可能出现的字符进行分类。所以该问题可以视为多类别的分类问题。由于每个算式中的字符是储存在一张图片中，我们需要对每个字符进行分类识别，而不是对整张图片中信息的分类识别。

字符序列s = s1,s2, s3,...sn，这里s1,s2,s3,sn均为算式字符集中的一个字符，n是有限的，约为10。我们的目标是通过算法识别每个字符sn，并最终得到完整的算式s。

数据集和输入

数据集来自Udacity提供的连接[1]。数据集是已标注算式内容的10万张计算机生成的RGB彩色图片，每张图片的大小都是300x64像素。每张图片包含一个公式。公式中可能出现的字符有“0～9”10个数字、“+、-、*”3个运算符、一对括号和一个“=”，共计16种字符。“=”出现的次数和数据集中图像总数相同（100000），其他字符出现的次数如图fig.1所示。不同内容的算式共有26341种，可见相同算式内容的样本量很小。

图像中的每种字符呈现不同程度的缩放、旋转，字符的字体、笔画粗细也不相同。字符的背景不是纯净的颜色，存在颗粒状的噪声。

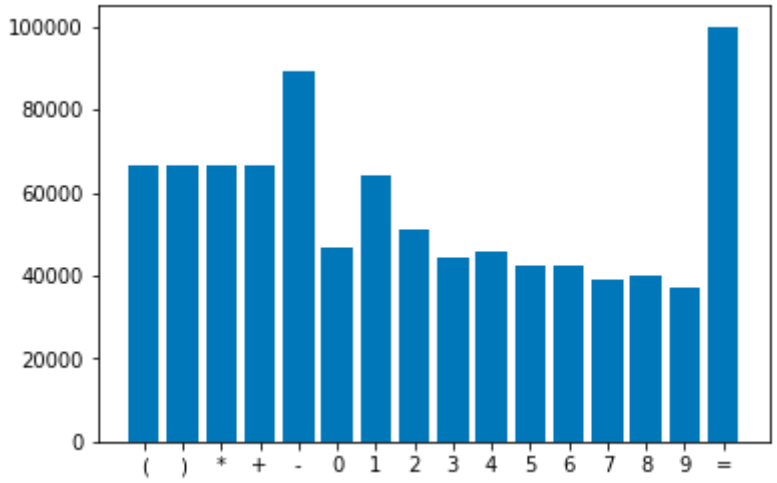

fig.1. 数据集中所有算式的字符出现的次数统计

fig.2. 算式识别数据集中的图像示例

方案声明

这是一个对图像文件中的字符序列进行分类识别的问题。需要提取字符特征，拟采用CNN实现算式中字符特征的提取。由于图像中的算式字符是连续的序列，而我们需要提取其中的每个字符并拼接成算式。有两种潜在的方案：一种是依据字符位置对图像进行分割，分割后的每个片段包含一个字符，并对每个片段用CNN算法实现字符识别。另一种是借助RNN对连续字符序列进行分类识别。Baoguang Shi团队提出了CRNN[2]方法，这一方法是CNN和RNN的结合：使用CNN获取图像的特征图，并将特征序列作为RNN的输入，得到图像中字符的预测分类，并最终转换成字符序列。

基准模型

本项目要求实现至少99%的准确率。

评估矩阵

使用准确率作为分类准确程度的指标，即在测试集中，正确分类的算式数目占测试集中算式总数目的比例。这里正确分类的算式应当是算式内所有字符都需正确识别，否则会使算式内容不成立从而导致识别无意义。

项目设计

总体方案参考CRNN网络架构[2]。利用CNN生成特征序列，传送给RNN，RNN对每个序列进行预测，根据预测结果输出序列的分类结果。
数据集中所有图片大小一致，因此对图片尺寸拟不做处理，RGB文件中的色彩信息对字符识别的作用不大，因此需将整个数据集中的图像转换成灰度图像，并将灰度图像的像素值归一化。在训练前需将数据集按80%、10%、10%的占比随机分成训练集、验证集和测试集。
CNN部分使用几种预训练模型做迁移学习并比较效果。如ResNet50，Xception，VGG16，这些预训练模型是基于ImageNet数据集进行训练的。本项目要训练的算式字符与ImageNet的1000种图像类别相似度较低，需要使用预训练模型的权重初始化模型权重并重新训练。由于CNN生成的特征要传送给RNN，因此CNN部分不再需要全连接层。
CNN和RNN需要在一起训练，并使用同样的损失函数。
借助可视化工具对模型中的图像处理步骤进行可视化处理，如将CNN的卷积核和卷积核的输出可视化。以理解模型如何提取字符特征。

参考

[1] https://s3.cn-north-1.amazonaws.com.cn/static-documents/nd009/MLND+Capstone/Mathematical_Expression_Recognition_train.zip
[2] Baoguang Shi, Xiang Bai and Cong Yao. An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition. 2015