

文章编号: 1000-5862(2017)05-0476-08

一种改进的深度卷积神经网络的精细图像分类

杨国亮, 王志元, 张 雨

(江西理工大学电气工程与自动化学院, 江西 赣州 341000)

摘要: 精细图像分类不同于传统的图像分类, 由于精细图像自身的类间相似性和类内差异性, 传统的基于手工特征和局部特征组合方法已经很难完整地表达精细图像的特征, 因此提出了一种基于改进的深度卷积神经网络模型。由于深度卷积神经网络结构参数和神经元数量巨大, 训练模型困难, 所以采用高斯分布对前6层参数初始化, 其中激活函数采用校正之后的 Relu-Softplus 函数, 在花卉图像数据库 OXford-102 flowers 中 TOP1 准确率达到 85.75%, TOP3 准确率达到 94.50%。实验结果表明: 该模型在中等规模数据集上比传统方法优势明显, 且比未改进的 CNN 模型识别率高。

关键词: 精细图像分类; 深度卷积神经网络; 激活函数; 特征提取

中图分类号: TP 391.41 文献标志码: A DOI: 10.16357/j.cnki.issn1000-5862.2017.05.06

0 引言

在视觉领域中, 精细图像分类一直是研究的重点^[1-2], 不同于传统的图像分类, 比如 caltech101^[3]、caltech256^[4] 图像数据集分别包含 101 类、256 类不同种类的图片, 类与类之间语义结构差别较大, 而精细图像语义及结构特征非常相似。在传统的方法中, 如有通过使用监督模板矩阵迭代来分析 (Grab-Cut 算法) 图片中感兴趣的区域, 还有对图像特征进行编码操作 (词袋 (BOW)) 等, 对于这些传统的方法, 一般需要大量的人工区域标注信息, 其特征信息表达程度不够, 因此分类能力有限。

2006 年, Hinton 等提出了一种基于深度信念网络的快速学习算法 (DBN), 引起了深度学习的一度浪潮。深度学习无需人工提取特征, 不仅在人工智能方面取得了质的飞越, 而且在计算机视觉的领域上取得了较大突破, 如语音识别^[5]、人脸识别^[6] 等。对于传统的神经网络, 有 2 个缺点: (i) 经典的 BP 神经网络, 在进行误差传播时若层数过多则会出现梯度弥散现象; (ii) 整个神经网络需要通过大量的非线性变换, 参数过多, 容易出现过度拟合现象。所以传统的神经网络结构层数不多, 提取图片特征信息不够全面。

本文尝试使用改进的深度卷积神经网络进行精细图像分类, 认为在构建精细图像分类问题的 CNN 网络结构时, 有以下 2 个问题值得考虑。

问题 1 当数据库中数据量偏小时, 是否能体现出深度卷积神经网络的优势?

文献 [7] 指出在大量的数据训练情况下, 使用深度 CNN 网络效果良好, 如在 2012 年 Alex 在 ImageNet 比赛中使用的数据集 ILSVRC, 其中图片有 120 万幅左右, 类别在 1 000 个。

然而在精细图像分类问题上, 图像的数据集难以达到以上要求, 原因有: (i) 精细图像一般是属于一个级别, 比如, 在花类中要找到 1 000 个不同类的花卉类别, 其工作量太大, 数据集也很难获取; (ii) 在精细图像数据集类别中, 不管是不同类别还是同一类别的图片, 语义相似、背景复杂、类之间的差别较小等都会引入较大的噪声, 这不利于精细图像分类。

那么如何在数据集和类别数较少的情况下, 充分发挥深度 CNN 网络对精细图像分类任务中的作用, 是值得去思考的问题。

问题 2 在精细图像分类中, 深度 CNN 网络模型是否能够学习到不同图像的不同特征?

深度卷积神经网络是根据生物视觉系统进行设计的, 对图像的特征提取更加抽象, 更有利于分类。

收稿日期: 2017-03-27

基金项目: 国家自然科学基金 (51365017, 61305019) 资助项目。

作者简介: 杨国亮 (1973-) 江西丰城人, 男, 教授, 博士, 主要从事模式识别与图像处理、智能控制的研究。E-mail: ygliang30@

126.com

从已有图像分类研究表明^[8],深度 CNN 网络结构学习到的图像特征比传统方法学习到的图像特征更易于分类。

精细图像分类与图像分类都是对图像特征进行学习,从中获取图像特征的共性与个性,最大化的分出不同的类别。

对于以上 2 个问题的考虑,本文提出一种改进的深度卷积神经网络模型,以便对精细图像进行分类。在进行深度 CNN 网络学习图像特征时,本文进行了以下尝试:(i) 问题 1 采用 2012 年 Alex 在 ImageNet 数据库上训练得到的模型,用来初始化模型参数,解决数据集较少、类别数目少以及难以发挥深度 CNN 网络优势的问题;(ii) 问题 2 采用深度卷积神经网络进行模型训练。

1 深度卷积神经网络

目前,深度卷积神经网络已经成为众多领域的研究热点之一。卷积神经网络主要有 3 大优点:局部感受野、权值共享、下采样。其中局部感受野指的是每一层的神经元只与上一层的一部分区域的神经元相连;权值共享是指采用相同的卷积核对同一张图片进行卷积操作;下采样是指把输入的特征图像划分为不同的子模块,然后对每个子模块进行池化操作。

1.1 卷积层

卷积层主要是通过输入的 2 维图像矩阵与每个卷积核进行卷积操作,采用的方式是局部感受野,每一组神经元都有一组权值,这些权值与前一层神经网络矩阵块神经元相关联,即权值共享。一般地,通过不同的卷积核可以得到不同的特征图,其卷积层的计算公式为 $X_j^l = f(\sum_{i \in M_j} X_i^{l-1} K_{ij}^l + b_j^l)$, l 表示层数, X_i^{l-1} 表示第 $l-1$ 层的第 i 张特征图, K_{ij}^l 表示第 l 层卷积核, $f(\cdot)$ 表示激活函数, b_j^l 表示偏置参数。

1.2 下采样层

卷积层紧接着下一层是下采样层,下采样的基本原理是根据图像的相对不变性的属性,对图像相邻区域的特征信息进行聚合统计。经过下采样层之后其中特征图的个数不会减少,但是输出的特征图的大小会改变(一般变为原来的 1/2),不仅可以减少运算量,而且可以降低图像的分辨率。另一方面,图像经过下采样之后具有平移不变性,从而增强网络对图像变化的适应性。其中下采样一般形式为 $X_j^l = f(\beta_j^l d_{\text{own}}(X_j^{l-1}) + b_j^l)$, 其中 $d_{\text{own}}(\cdot)$ 表示下采样

函数,一般有 3 种下采样的方式:最大下采样(Max-Pooling)、均值下采样(Mean-Pooling)、随机下采样(Stochastic-pooling)。 β_j^l 表示第 l 层第 j 个特征图, b_j^l 表示第 l 层第 j 个偏置项。

1.3 激活层

激活函数的选取对网络结构训练的收敛性有较大影响,因此选取合适的激活函数至关重要。除了输入层神经元和下采样层一般没有激活函数,其他各层神经元都有激活函数作为传递函数。常用的激活函数有饱和非线性函数(如 Sigmoid 函数、Tanh 函数)和不饱和非线性函数(如 Relu 函数、Softplus 函数)。

1.4 全连接层

全连接层相当于一种前馈人工神经网络模型,特点是本层的神经元与上层的神经元两两相连,但是本层的神经元之间是不相连的结构模型。其中每个神经元输出表示为 $h_{wb}(x) = f(W^T x + b)$ 。经过全连接层之后输出的局部特征信息作为下一层(Hingeloss 等分类器)的输入,图像特征由原来的 2 维矩阵变为 1 维矩阵。由于卷积运算是 2 维的,所以后面不需要卷积层操作,为了防止过度拟合现象,提高网络结构的泛化能力,采用了 Dropout 技术^[9]。

1.5 输出层

本文设计的模型有 2 个输出层:(i) Softmax 分类器输出层,(ii) Accuracy 准确率输出层。Softmax 分类器是逻辑回归模型在多个类别分类问题上的推广,但是不适合多个标签样本。在 Logistic regression 中,所学习的系统方程为 $h_{\theta}(x) = (1 + \exp(-\theta^T x))^{-1}$, 对应的损失函数为 $J(\theta) = -\frac{1}{m}(\sum_{i=1}^m y^{(i)} \log h_{\theta}(x^{(i)}) + (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)})))$, 可以看出,给定样本输出一个概率值,该概率值表示的含义是这个样本属于类别 1 的概率。因为总共才有 2 个类别,若有 K 个类别,则在 Softmax regression 中系统方程为

$$h_{\theta}(x^i) = \begin{bmatrix} p(y^i = 1 | x^i); & \theta_1 \\ p(y^i = 2 | x^i); & \theta_2 \\ \vdots & \vdots \\ p(y^i = k | x^i); & \theta_k \end{bmatrix} = \frac{1}{\sum_{j=1}^k e^{\theta_j^T x^i}} \begin{bmatrix} e^{\theta_1^T x^i} \\ e^{\theta_2^T x^i} \\ \vdots \\ e^{\theta_k^T x^i} \end{bmatrix},$$

其中 $(\sum_{j=1}^k e^{\theta_j^T x^i})^{-1}$ 表示归一化项, $\theta_1, \theta_2, \dots, \theta_k$ 是模型的可学习参数。

此时损失函数为

$$J(\theta) = -\frac{1}{m} \left(\sum_{i=1}^m \sum_{j=1}^k I(y^i = j) \log \frac{e^{\theta_j^T x^i}}{\sum_{l=1}^k e^{\theta_l^T x^i}} \right),$$

其中 $I\{\cdot\}$ 表示指示性指标函数, 当 $y^i = J$ 值为真时输出为 1, 反之输出为 0.

当求最优解时, 可以选用随机梯度下降法、牛顿法、L-BFGS 法等. Logistic regression 中的损失函数的偏导函数为 $\nabla_{\theta_j} J(\theta) = \frac{1}{m} \sum_{i=1}^m (p(y^i = j | x^i; \theta) - x^i(I(y^i = j)))$, 其中 $\nabla_{\theta_j} J(\theta)$ 表示损失函数对第 j 个类别的第 i 个参数的偏导, 本文采用随机梯度下降法求出最优解.

2 改进型的深度卷积神经网络模型

2.1 模型原理

整个实验过程可分为 2 个阶段: 训练阶段和测试阶段. 在训练阶段可以分为前向传播和反向传播, 比如卷积、下采样等一系列过程都是前向传播提取特征的过程, 其中前向传播输出函数表示为 $h = f_L(X_L; W_L)$, $f(\cdot)$ 表示激活函数, X_L 表示前一层的输出值, W_L 表示权值矩阵.

反向传播的过程采用传统的 BP 算法, 使用链式求导法则更新权值和偏置, 在最后一层采用 Softmax 回归模型, 本文使用的是 Log-loss 误差函数. 比如在第 L 层, 损失函数可表示为 $l(x_i, y_i) = -\frac{1}{n} \sum_{i=1}^n (y_i - \ln(g_L(x_i; w_i))) + \lambda \sum_{k=1}^L \sum_{j=1}^K \|W_j\|$, 其中 n 表示的是样本的总数, x_i, y_i 分别对应的是训练样本数和训练样本标签, 为了防止过度拟合, 也采用了 L_2 正则化项.

此时整个网络的输出函数表示为 $O = \arg_{w_L} \min \frac{1}{n} \cdot$

$\sum_{i=1}^n l(g_L(x_i; w_L), y_i)$ 其中 w_L 权值矩阵通过下列函数进行更新 $\frac{\partial l}{\partial w_L} = -\frac{1}{n} (Y_L - g(X_L; w_L)) g'(X_L; w_L) + \lambda w_L$.

本文采用随机梯度下降法进行迭代, 不断更新权重以及偏置使得误差值最快达到收敛状态. 在整个

深度 CNN 网络的结构中, 有 2 层在误差反向传播的方式是与传统的 BP 是不相同的, 分别为卷积层和下采样层.

若卷积层的前一层为下采样层时, 则需要从后一层卷积层的误差来求该下采样层的误差. 其中输出表达式为 $y_{L-1}^j = (\sum_{i=1}^M y_L^i K_{ij}) g'(x_L^j; w_L)$, 其中 y_L 表示卷积误差, K_{ij} 表示卷积层的卷积核, $g'(x_L^j; w_L)$ 表示该激活函数的导数.

若卷积层的后一层为下采样层, 则该卷积层的误差可表示为 $y_{L-1}^j = u_{psample}(y_L^j) g'(x_L^j; w_L)$, 其中 $u_{psample}(y_L^j)$ 表示进行下采样所使用的方法与前向传播时的下采样方式有关, 若下采样的方式是随机最大概率, 则需要查找出在下采样时随机最大概率的位置, 并且在进行下采样时该位置被置为相对应的数值, 其他的位置被置为 0.

2.2 模型架构

本文提出的改进的深度卷积神经网络结构模型如图 1 所示(与 AlexNet 模型相同, 但是采用的激活函数不同), 包括 5 个卷积层, 卷积核大小为 11×11 , 滑动步长为 4, 3 个 Pooling 层和 3 个 LRN 层, 激活函数采用的是通过 Relus 函数与 Softplus 函数校正之后的 Relus-softplus 函数, 3 个全连接层, 最后一层为全连接层, 即需要提取的特征向量. 本文最大的特点在于将不饱和非线性 Relus 函数和 Softplus 函数各自的稀疏表达能力和光滑特性结合起来, Relus 函数和 Softplus 函数的公式表示如下及各自函数曲线如图 2 所示. Relus 函数为 $f(x) = \text{MAX}(0, x)$, Softplus 函数为 $h(x) = \log(1 + e^x)$.

从图 2 可看出, Relus 函数在小于零的情况下全部置为零, 大于零的数就保持不变, 这说明该函数具有稀疏表达能力但并没有非线性纠正能力, Softplus 函数是对全部的数据进行非线性映射, 但是没有稀疏表达能力. 先将 Softplus 函数曲线向下平移 $\log 2$ 个单位长度, 并且把小于零的数据归为 0.

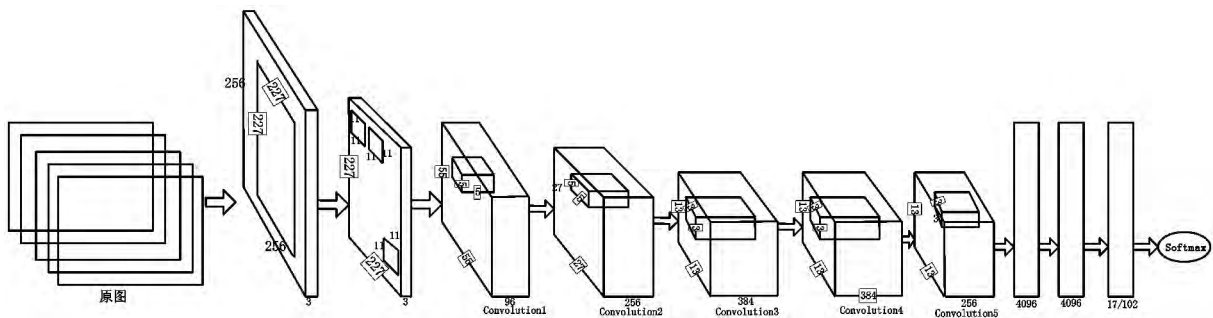


图 1 深度卷积神经网络结构图

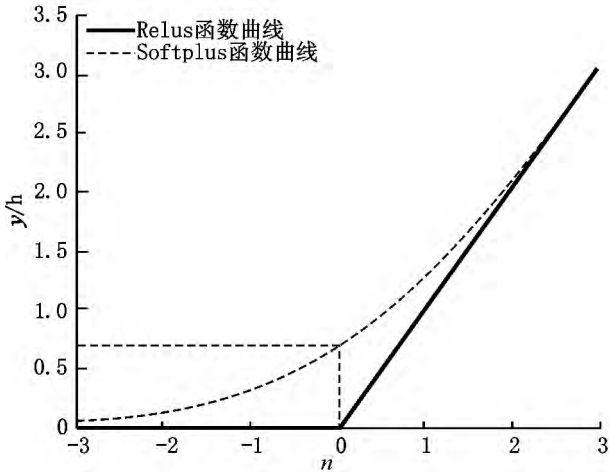


图 2 Relus 与 Softplus 函数曲线

该函数可记为 Relus-Softplus(rectified nonlinear units-softplus) ,其函数曲线如图 3 所示 ,函数表达式可表示为 $f(x) = \text{MAX}(0, \log(1 + e^x) - \log 2)$,从该函数的图形来看 ,该函数不仅具有稀疏表达能力 ,而且可以去除数据中的冗余 ,最大可能的保留数据的特征.

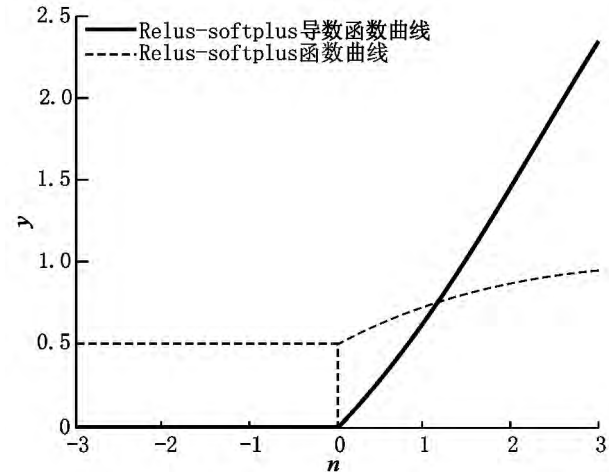


图 3 Relus-Softplus 函数及导数曲线

在反向传导时 ,采用的是 BP 算法推导 ,当误差从输出层反向传播算梯度时 ,在各层都要乘上当前层的输入神经元值 ,激活函数的一阶导数 ,即梯度表达式可为 $\text{Grad} = \text{Error} \cdot (\text{Relus-softplus})'(x) \cdot x$.这样 ,在经过每一层时 Error 都是成倍的衰减 ,一旦进行递推式的多层反向传播 ,梯度就会不停地衰减 ,消失 ,使得网络学习变慢 .如图 3 所示 ,Relus-softplusd 函数在大于零的数据都进行了非线性映射 .且该激活函数的导数在大于 0 的情况下 ,呈现的是递增趋势 ,与 Relus 一样克服了梯度消失问题.

为了验证改进后的激活函数是否真的能提高分类准确率 ,本文直接采用 Relus、softplus、Relus-soft-

plus 3 种激活函数对 102 类花卉进行分类 ,如图 4 所示 ,改进后的激活函数 Relus-softplus 分类准确率有所提高;如表 1 所示 ,分类错误率明显下降;如图 5 所示 ,在迭代 500 次之后 ,采用 Relus-softplus 激活函数的 loss 从 5.457 降到 1.699. 通过以上实验可以得出改进后的激活函数不仅可以提高网络的收敛速度 ,而且对精细图像分类识别率有所提高.

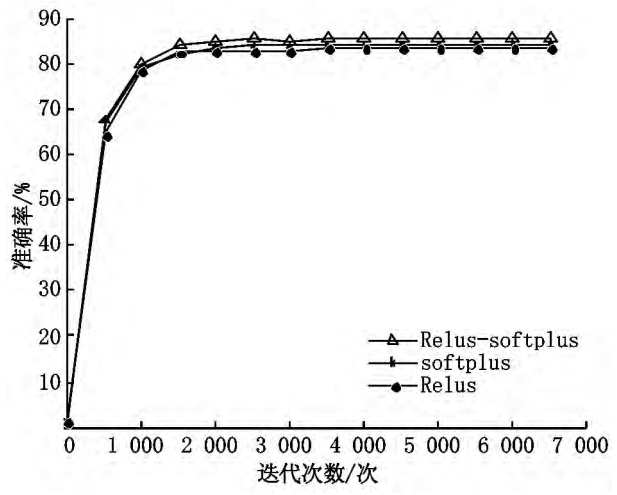


图 4 不同激活函数网络模型对 102 类花卉分类正确率随着训练次数变化曲线

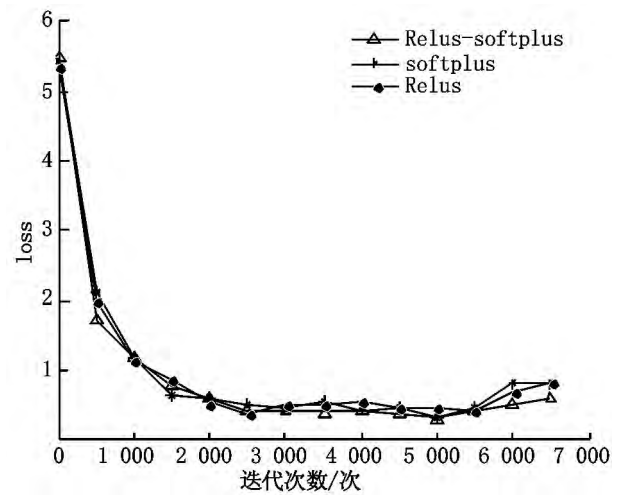


图 5 不同激活函数网络模型对 102 类花卉分类生成 loss 随着训练次数变化曲线

表 1 不同激活函数网络模型对 102 类花卉分类错误率

| 激活函数 | 不同激活函数的 TOP1 分类错误率(%) |
|----------------|-------------------------|
| Relus | 16.57 |
| Softplus | 15.69 |
| Relus-softplus | 14.25 |

2.3 模型分析

卷积层以及全连接层各参数设定不当均会对深度 CNN 网络模型有较大的影响 ,比如计算加速优化以及参数进行优化、权值裁剪等.

在进行单样本前向传播时,假设卷积层有 L 个输出通道和 K 个输入通道,卷积核大小为 $I \cdot J$,每个输出通道的特征图大小均为 $M \cdot N$,则计算量记为 $C_{calculations}(M_{AC}) = I \cdot J \cdot M \cdot N \cdot K \cdot L$. 注意,实际中通常一次送入一批样本(batch),此时计算量应该乘上批量尺寸.

通过上述参数,可以计算出卷积层的学习参数数量,也就是卷积核数目乘上卷积核尺寸 $P_{params} = I \cdot J \cdot K \cdot L$,这里定义计算量-参数数量之比为 C_{PR} (Calculations to Parameters Ratio) $= M_{AC}/P_{params} = M \cdot N$,可以看出若卷积层的输出特征图尺寸越大,则 C_{PR} 值越大,参数的重复率越高,若输入一批数据(B 个样本)则 C_{PR} 值可再提高 B 倍.在全连接层中单样本前向传播计算量统计为 $Calculations(MAC) = V \cdot D$,其中 D 表示输入节点组成向量的维度, V 表示输出节点组成向量的维度.

在参数量的统计为 $Params = V \cdot D$,则 C_{PR} 值为 1,可见,全连接层的 C_{PR} 值始终为 1,与输入输出维度无关,权值重复率很低,若将一批(B 个)样本 x_i 逐列拼接成矩阵 X ,则全连接层前向传播计算量提高了 B 倍,而参数量是不变的,因此 C_{PR} 提高了 B 倍,权值矩阵在多样本之间实现了共享,可提高计算速度,也就是说卷积层在输出特征图维度实现了权值共享,这是降低参数量的重要举措,与此同时,卷积层局部连接也大幅度减少参数量,但是计算量大,而后几层的全连接层参数量占比大,计算量小,大多数 CNN 模型都是符合这个特点,因此在进行计算加速优化时,重点放在卷积层;而当参数进行优化、权值剪裁时,最好放在全连接层.

3 实验与分析

3.1 精细图像分类数据库

为了检测该模型的有效性,选取了 Oxford-17 flowers、Oxford-102 flowers 数据库,该数据来自于英国牛津大学的视觉几何实验室(Visual Geometry Group)所创建的花卉图像数据库,Oxford-17 flowers 数据库包含 17 类花卉图像,一共有 1 360 张图片,每个类别是 80 张,Oxford-102 flowers 数据库包含了 102 类,每一类花卉图片在 40 ~ 258 张之间,总共是 8 189 张图片.该数据库也兼顾了在图像识别领域中所有的难点,比如光照的变化、视觉的变化、背景复杂、花卉类别形态以及颜色变化,加上部分不同的花卉高度相似,因此研究精细图像分类具有重要的意

义,本节将给出试验结果与分析.

3.2 改进型的卷积神经网络结构实验

3.2.1 实验设置 本实验采用的硬件设备 Inter(R) Core(TM) i7-4790 CPU 3.60 GHz 内存 16.0 GB 和 Nvidia GeForce GT750 M GPU,软件环境是在 Ubuntu14.04 系统下采用深度学习 Caffe 框架进行训练和测试.选用的是 Oxford-17 flowers、Oxford-102 flowers 数据库,在对图像进行预处理时,因为 17 类花卉图片分辨率不均等,首先是将所有的图片进行缩放到 256×256 像素,在程序文件 prototxt 中设置 scale 为 227,对图片中间进行截取 227×227 像素图像块,然后对图片进行镜像翻转.值得注意的是,本文数据在进行训练和测试时都是用训练数据处理为均值文件,可以提高文件存取效率.在 Oxford-102flowers 数据库训练时,其 Dropout ratio 设置为 0.5,对 Oxford-17 flowers 数据集,其 Dropout ratio 设置为 0.6.原因是训练样本数据过少需要抑制更多的神经元,防止过拟合现象.在 pooling 层, pool 参数选择的是 MAX,即为最大下采样.本文选择的激活函数为校正之后的 Rule-softplus 函数,不仅可以加快收敛的速度,而且对识别率有所提高.由于样本数据不多,初始学习率设置为 0.005,2 个数据库都是采用 step 方式迭代,当达到 544 次或者 1 500 次时,都是以 1/10 的方式下降.使用训练好的模型文件,作为前 6 层网络的初始化,并且将前 6 层学习率调为 0.001,后全连接层学习率调为 0.005,加快全连接层训练速度.具体的参数需要根据所选择哪种梯度下降法进行设置.在 17 类花卉分类时,对比采用不同 Batchsize 进行测试.在 102 类花卉中训练集的 Batchsize 为 64,测试集的 Batchsize 为 50.

3.2.2 结果分析 表 2 列出了不同激活函数所对应的 17 类花卉的正确率,其中训练集的 Batchsize 为 64,测试集的 Batchsize 为 34,迭代次数是 1 088 次,其中在 Sigmoid 激活函数下 TOP1 和 TOP3 分类准确率都是最低的.在通过 Relus 和 Softplus 函数校正之后的激活函数 Relus-Softplus 上分类比其他激活函数分类效果好.

表 3 与表 2 不同的地方在于训练集的 Batchsize 为 32,测试集的 Batchsize 为 17,如表 3 所示,在降低 Batchsize 情况下,对图像分类还是有一定的影响,在 TOP1 分类中达到了 93.95% 的正确率,比文献[12]中所提到的 2 种方法(直接融合特征方法和多特征融合方法)分类精度提高了 0.81% 左右.

在进行 102 类花卉分类实验时,直接对比 Relus

函数、Softplus 函数及校正之后的 Relus-Softplus 函数分类正确率,通过实验对比,其中训练集的 Batchsize 为 64,测试集的 Batchsize 为 50,测试的迭代次数为 32 次,总共迭代次数为 6 500 次。从表 4 可以看出,校正之后的 Relus-Softplus 函数比 Relus 与 Softplus 激活函数准确率分别高出 2.32%、1.44% ,

在对 102 类花卉进行分类时得出准确率为 85.75% ,针对实验一些部分错分比较严重以及准确率比较高的部分花卉进行分析(见图 6 和图 7)。

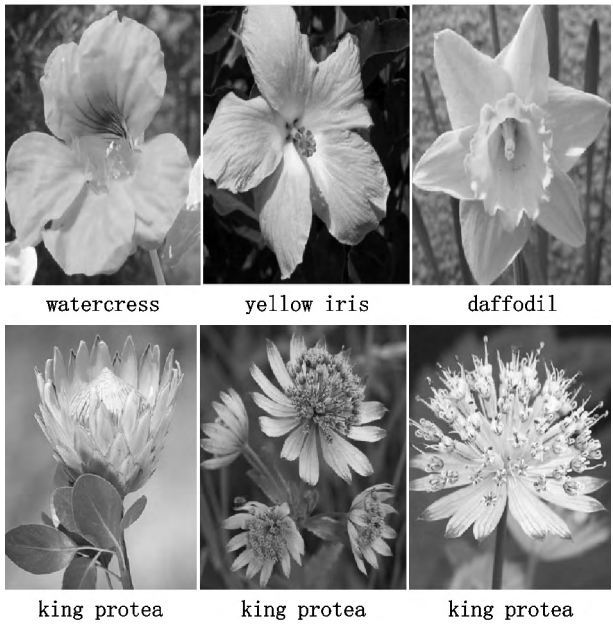


图 6 试验中部分错误分类的花卉图像示例



图 7 分类准确率比较高的部分花卉示例

由图 6 对比发现 watercress、yellow iris 以及 daffodil 不仅在颜色而且形状都是比较相似的,king protea 在不同时期中颜色形状都不一样,从而影响分类效果。由图 7 对比发现分类准确率比较高的图片的颜色和形状均是有明显差异的。

针对花卉图像分类,目前已经取到不少成果,如文献 [15、17-18] 采用的是先对图像进行分割再进行分类方法,文献 [13-14、16] 研究特征在图像表示方面。具体的准确率如表 5 和表 6 所示。

据表 5 知,通过对比传统方法和未改进的 CNN 模型方法,改进后的 CNN 模型对花卉图像分类准确率明显提高。但是对于 17 类花卉图像的分类效果来看改进后的 CNN 模型分类效果不是很明显,由于采用 CNN 模型需要大量的数据集进行训练,得到更抽象的特征,从表 6 可以发现改进后的 CNN 模型分类准确率明显高于其他方法,进而可以看出样本数量的大小对深度 CNN 模型分类性能有一定的影响。因此对于模型的泛化能力的强弱以及分类效果的好坏,训练样本的多少很重要。

表 2 不同激活函数的分类正确率
(train_batchsize = 64 ,test_batchsize = 32) %

| 激活函数 | 不同激活函数的分类正确率 | |
|----------------|--------------|-------|
| | TOP1 | TOP3 |
| Tanh | 48.16 | 79.04 |
| Sigmoid | 5.83 | 17.64 |
| Relus | 91.17 | 99.26 |
| Relus-Softplus | 91.91 | 98.89 |

表 3 不同激活函数的分类正确率
(train_batchsize = 32 ,test_batchsize = 17) %

| 激活函数 | 不同激活函数的分类正确率 | |
|----------------|--------------|-------|
| | TOP1 | TOP3 |
| Tanh | 47.79 | 72.79 |
| Sigmoid | 5.88 | 17.64 |
| Relus | 92.64 | 98.16 |
| Relus-Softplus | 93.95 | 99.63 |

表 4 102 类花卉分类准确率 %

| 激活函数 | 不同激活函数的分类准确率 | |
|----------------|--------------|-------|
| | TOP1 | TOP3 |
| Relus | 83.43 | 92.52 |
| Softplus | 84.31 | 92.95 |
| Relus-Softplus | 85.75 | 94.50 |

表 5 OXford-17 flowers 数据库实验结果对比 %

| 方法 | 不同方法的分类准确率 | |
|-----------------------------------|------------|--|
| | TOP1 | |
| Baseline ^[12] (特征直接融合) | 87.50 | |
| Gehler ^[13] | 85.50 | |
| Khan ^[14] | 89.00 | |
| Chai ,BicosMT ^[15] | 90.40 | |
| Chai ,Bicos ^[15] | 91.10 | |
| Femando ^[16] | 93.00 | |
| 多特征融合 ^[12] | 93.14 | |
| 本文方法 | 93.95 | |

* 此表数据引用自文献 [12] 的结果。

表6 Oxford-102 flowers 数据库实验结果对比 %

| 方法 | 不同方法的分类准确率 |
|-------------------------------|------------|
| | TOP1 |
| Nilsback ^[17] | 72.80 |
| Nilsback ^[15] | 76.30 |
| Chai ,Bicos ^[15] | 76.30 |
| Chai ,BicosMT ^[15] | 80.00 |
| Angelova ^[18] | 80.66 |
| 多特征融合 ^[12] | 79.10 |
| 本文方法 | 85.75 |

* 此行数据引用自文献[12]的结果.

4 结论

深度学习不仅在思想上已经成为机器学习领域的重要分支,而且在技术上是应用大数据分析的重要工具.随着大数据概念的提出以及计算机硬件要求的不断提高,卷积神经网络应用更加广泛.本文设计的模型就是在 Alexnet 的基础上进行模型优化,利用校正之后的非线性修正函数对神经网络中的神经元进行映射,提高模型的分类正确率,其性能比未改进的网络模型分类效果更好,且比传统方法的分类效果更好.

然而深度学习模型只有在大量数据集的前提下进行训练模型,拟合能力才会更强,若数据量少则很容易造成过拟合的现象.因此本文设计的模型需要在更大的数据集上进行验证.该模型用于精细图像分类也存在一些需要思考的地方,比如如何提取更有效的特征等.更值得注意的是,近年来所有有关卷积神经网络应用的文献,其中大部分是针对某一类的数据库而设计的模型,以其泛化能力不强,因此对于不同的识别任务,设计出通用性强的模型,提高效率是目前深度学习需要解决的问题之一.

5 参考文献

- [1] Wang Xiaoyu ,Yang Tianbao ,Lin Yuanqing. Object-centric fine-grained image classification [EB/OL]. [2016-10-11]. <http://www.google.com/patents/US20160140424>.
- [2] Akata Z ,Reed S ,Walter D ,et al. Evaluation of output embeddings for fine-grained image classification [EB/OL]. [2016-10-11]. <http://www-personal.umich.edu/~reed-scot/CVPR15.pdf>.
- [3] Zheng L ,Zhao Y ,Wang S ,et al. Good Practice in CNN Feature Transfer [EB/OL]. [2016-10-11]. <http://128.84.21.199/pdf/1604.00133.pdf>.
- [4] Grohs P ,Wiatowski T ,Bölcskei H. Deep convolutional neural networks on cartoon functions [EB/OL]. [2016-10-11]. 10.1109/ISIT.2016.7541482.
- [5] Kim C ,Stern R M. Power-normalized cepstral coefficients (PNCC) for robust speech recognition [J]. ACM Transactions on Audio , Speech and Language Processing (TASLP) 2016 24(7):1315-1329.
- [6] Pentland A ,Moghaddam B ,Starner T. View-based and modular eigenspaces for face recognition [EB/OL]. [2016-10-11]. <http://www.ijsr.net/archive/v5i9/9091603.pdf>.
- [7] Sun Y ,Wang X ,Tang X. Deep learning face representation from predicting 10000 classes [EB/OL]. [2016-10-11]. <http://www.ee.cuhk.edu.hk/~xgwang/papers/sun-WTcvpr14.pdf>.
- [8] Rastegari M ,Ordonez V ,Redmon J ,et al. XNOR-Net: imagenet classification using binary convolutional neural network [EB/OL]. [2016-10-11]. https://pjreddie.com/media/files/papers/xnor_arxiv.pdf.
- [9] Schmitz A ,Bansho Y ,Noda K ,et al. Tactile object recognition using deep learning and dropout [J]. IEEE-RAS International Conference on Humanoid Robots ,2015 ,11(3):1044-1050.
- [10] RenShaoqing ,He Kaiming ,Girshick R ,et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence 2016 39(6):1137.
- [11] Nilsback M E ,Zisserman A. Delving into the whorl of flower segmentation [EB/OL]. [2016-10-11]. <http://www.robots.ox.ac.uk/~vgg/publications/papers/nilsback07.pdf>.
- [12] 谢晓东. 面向花卉图像的精细图像分类研究 [D]. 厦门: 厦门大学 2014.
- [13] Angelova A ,Zhu S. Efficient object detection and segmentation for fine-grained recognition [J]. Computer Vision & Pattern Recognition 2013 9(4):811-818.
- [14] Zou J ,Nagy G. Evaluation of model-based interactive flower recognition [J]. International Conference on Pattern Recognition 2004 2(2):311-314.
- [15] Nilsback ,Maria Elena. An automatic visual flora-segmentation and classification of flower images [D]. Oxford: Oxford University 2009.
- [16] Chai Y ,Lempitsky V ,Zisserman A. BiCoS: a Bi-level co-segmentation method for image classification [J]. IEEE International Conference on Computer Vision ,2011 ,58(11):2579-2586.
- [17] Yang Shulin ,Bo Liefeng ,Wang Jue ,et al. Unsupervised template learning for fine-grained object recognition [C]. Advances in Neural Information Processing Systems , 2012: 3122-3130.

[18] Hong Anxiang ,Chen Gang ,Li Junli ,et al. A flower image retrieval method based on ROI feature [J]. Journal of

Zhejiang University: Science A 2004 5(7) : 764-772.

An Improved Depth Convolutional Neural Network for Fine Image Classification

YANG Guoliang ,WANG Zhiyuan ,ZHANG Yu

(School of Electrical Engineering and Automation ,Jiangxi University of Science and Technology ,Ganzhou Jiangxi 34100 ,China)

Abstract: Fine image classification is different from traditional image classification. Due to the similarity between intra-class and inter-class differences of fine-grained images themselves ,it is difficult to express the characteristics of fine image based on manual feature and local feature combination method. Based on the improved depth convolution neural network model ,due to the large number of deep convolution neural network structure parameters and the large number of neurons ,the training model is difficult ,and the Gaussian distribution is used to initialize the first six parameters. The activation function is used after the correction of the Relus-Softplus function ,the TOP1 accuracy rate of the flower image database Oxford-102 flowers is 85.75% ,and the TOP3 accuracy rate is 94.50% . The experimental results show that the model has obvious advantages over the traditional method ,and the recognition rate is higher than that of the unmodified CNN model.

Key words: fine-grained image classification; deep convolutional neural network; activation function; feature extraction

(责任编辑: 冉小晓)

(上接第 469 页)

[38] 涂冬波 ,蔡艳 ,丁树良. 认知诊断理论、方法与应用 [M]. 北京: 北京师范大学出版社 2012.

[39] 高椿雷 ,罗照盛 ,郑蝉金 等. 具有认知诊断功能的多阶段自适应测验及其影响因素研究 [J]. 心理科学 2015 39(6) : 1492-1499.

[40] Timminga E. Solving infeasibility problems in computerized test assembly [J]. Applied Psychological Measurement ,1998 22 (3) : 280-291.

The Test Assemble Methods of Multistage Adaptive Test

LI Guiyu ,TU Dongbo* ,DAI Buyun ,ZONG Yitao ,GAO Xuliang ,MIAO Ying

(School of Psychology ,Jiangxi Normal University ,Nanchang Jiangxi 330022 ,China)

Abstract: The key to implement multistage adaptive test(MST) is to build multiple parallel tests(or panels) that meet statistical and nonstatistical constraints required. Automated test assemble(ATA) provides a way to achieve parallel tests. Existing assemble methods are mainly based on linear programming ,heuristic algorithm ,Monte Carlo and on-the-fly method. Future studies should focus on the comparison and improvement of these methods and the development of ATA based on cognitive diagnosis tests.

Key words: multistage adaptive Test; automated test assemble; test assemble methods; test specification

(责任编辑: 冉小晓)