

深度卷积神经网络在图像识别算法中的研究与实现

韩星烁 林 伟

(福州大学 物理与信息工程学院 福建 福州 350002)

摘 要: 深度学习是近些年来人工智能领域取得的重大突破,深度学习与传统模式识别方法的最大不同在于它所提取的特征是从大数据中自动学习得到,而非采用手工设计。现有的深度学习模型属于神经网络,文章引入深度卷积神经网络进行图像识别,该算法在对图像统一转换成固定尺寸后进行处理,具有局部感受野、权值共享和空间下采样等特点,可以有效地提取图像特征。文中使用 Python 爬虫技术采集的 993 张图像数据集,对该方法进行测试,平均识别率达到 92.50%。实验结果表明,基于深度卷积神经网络的图像识别方法是可行的。

关键词: 卷积神经网络; 局部感受野; 权值共享; 图像识别

中图分类号: TP317.4

文献标识码: A

DOI: 10.19358/j.issn.1674-7720.2017.21.017

引用格式: 韩星烁 林伟. 深度卷积神经网络在图像识别算法中的研究与实现[J]. 微型机与应用 2017, 36(21): 54-56.

Design and implementation of image recognition based on convolution neural network

Han Xingshuo, Lin Wei

(College of Physics and Information Engineering, Fuzhou University, Fuzhou 350002, China)

Abstract: Deep learning is a major breakthrough in recent years in the field of artificial intelligence, the biggest difference between deep learning and traditional pattern recognition method is that the extracted features are learned from large data automatically, rather than artificial design. Existing deep learning model belongs to the neural network. This paper introduced the deep convolution neural network for image recognition, the algorithm for image processing after unified into a fixed size, with the characteristics of local receptive field, the weights of sharing and space sampling etc, can effectively extract the image feature. This article uses the crawler technology for the 993 images of the data set, after testing the method, the average recognition rate is 92.50%. The experimental result shows that the proposed method in this paper is feasible.

Key words: CNN; local receptive field; weights of sharing; image recognition

0 引言

图像识别技术是人工智能领域非常核心的一个方向,在国家安全、公安、交通、金融、工业化生产线、食品检测等诸多领域具有广泛的应用前景。

传统图像识别算法首先需要对图像进行预处理,包括图像分割、图像增强、二值化细化等,然后进行人工特征提取,建立相对应特征的识别模型,比如高斯混合模型、支持向量机、隐马尔科夫模型等,然后进行判断匹配。其中手工设计的特征对于多样性的变化并没有很好的鲁棒性,提取的方法对图像中层结构和高层语义信息的表示能力有限^[1],并且传统算法费时费力,很难大规模地应用在实际生产中。

2006 年, Hinton 等人开创性地提出了深度学习的概念^[2]。引入无监督学习的方法训练深度神经网络,使其相比传统机器学习方法得到的特征更有利于数据的分类、回归分析及可视化,揭示出深度学习对于大规模、高维度数据的巨大应用潜力。于是,深度学习迅速成为机器学习领域中最为热门的研究方向,并掀起了一场新的研究热潮。

卷积神经网络是在多层神经网络的基础上发展起来的,针对图像分类和识别而特别设计的一种深度学习方法,具有结构简单、训练参数少和适应性强的特点。所以它的应用领域十分广泛,比如手势识别^[3-4]、人脸识别、虹膜识别^[5]、车型识别^[6]等。

1 卷积神经网络

卷积神经网络是在多层神经网络的基础上发展起来的,针对图像分类和识别而特别设计的一种深度学习方法。它是一种前馈神经网络。卷积神经网络是受人体视觉中感受野的机制影响而提出。比如在人体视觉神经系统中,一个神经元的感受野是指视网膜上的特定区域,只有在这个区域内的刺激才能够激活该神经元。卷积神经网络的最大特点就是局部感受野和权值共享,图 1 为稀疏连接(局部感受野),图 2 为权值共享。稀疏连接和权值共享可以减少训练参数和计算复杂程度。

1.1 卷积神经网络

卷积操作是对图像和滤波矩阵做内积,这也是卷积神经网络的名字来源。它的核心思想是结合稀疏连接、权值

《微型机与应用》2017 年第 36 卷第 21 期

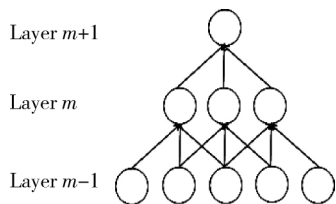


图1 稀疏连接

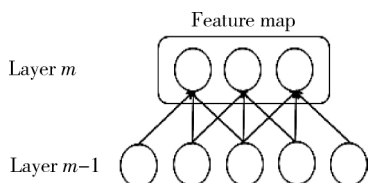


图2 权值共享

共享以及采样这三种结构得到某种程度的位移、尺度、形变不变性。在图像中,输入是 width* height 的区域大小的数据和一组固定权重的卷积核做内积得到新的二维数据。具体操作是一个卷积核(特征矩阵)在输入的图像矩阵上按照一定的步长移动,直至覆盖整个图像,然后与矩阵上对应位置的元素相乘后再相加得到一个值。移动完毕生成新的特征图矩阵。

在卷积神经层中,卷积核的大小影响着图像特征的抽象效果。一般来说,卷积核越大,效果越好,但是训练参数多卷积核则越小,训练参数少处理效果更精细,这需要更多的层数达到相同的效果。在本文的结构中,第一个卷积层使用了 11×11 的卷积核,卷积核较大,虽然能够达到比较好的抽象效果,但是处理得比较粗糙,所以在 Conv 后加入 Norm 归一化层,提升主体与其他部分的区分度。

1.2 ReLU 激励层

通常情况下,卷积层和 ReLU 层是成对出现的。规范的 ReLU 激活函数的表达式为: $y = \{0, \max(x)\}$,当输入 $x > 0$ 时,输出为 x 本身;如果输入小于或等于 0,则输出 0。

在卷积神经网络中,一般情况下,习惯于用 ReLU 激励函数替代之前的 tanh、sigmoid 等激活函数,如图 3 所示,在 $x > 0$ 时,ReLU 函数的导数是恒定的,而 tanh 和 sigmoid 函数不是恒定的,所以 ReLU 函数避免了 tanh 和 sigmoid 函数在两端接近目标时,导数变小,导致在训练神经网络时,BP 反向传播误差而导致收敛减慢。ReLU 的优点是收敛快,求梯度简单,并且训练后具有稀疏性,可减少数据冗余,增强特区特征的表达能力。

1.3 池化 pool 层

池化层也叫空间下采样层,在卷积神经网络中,池化层一般是在卷积层之后,在图像卷积之后使用 pooling 将小邻域内特征点整合之后得到新的特征。通常卷积和池化是以 Conv-Pool 的形式存在,减少卷积之后造成的信息

冗余。pool 层也叫降采样层,一方面是能够达到降维的目的,降低上一个卷积层输出的特征向量的维度。另一方面是能够减少过拟合。一般情况下 pool 层分为三种:一般池化、重叠池化和金字塔池化。一般池化中又分为 max-pooling、mean-pooling 和 Stochastic-pooling,本文采用 max-pooling 降低图像的噪声,减少图像的卷积输出结果对于输入的误差过于敏感的这种过拟合现象。

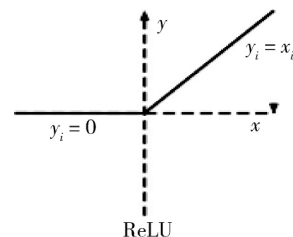


图3 ReLU 函数

本文所采取的 max-pooling 算法,首先对于图像来说,这个操作可以保证特征的位置和旋转不变性,这是一种很好的特性,因为不论这个卷积之后得到的有效特征在哪个位置出现,都不用考虑其出现位置而能把它提取出来。另外,本文中 max-pooling 大大减少了模型的参数数量,而对于 pool 层之后的 norm 层来说,神经元的个数大大减少。

本文所采取的 max-pooling 算法,首先对于图像来说,这个操作可以保证特征的位置和旋转不变性,这是一种很好的特性,因为不论这个卷积之后得到的有效特征在哪个位置出现,都不用考虑其出现位置而能把它提取出来。另外,本文中 max-pooling 大大减少了模型的参数数量,而对于 pool 层之后的 norm 层来说,神经元的个数大大减少。

2 基于深度卷积神经网络的算法与训练

一个典型的卷积神经网络的分类器结构应该包括输入层、卷积层、池化层、全连接层和输出层,其中卷积层和池化层交替出现。本文所采用的网络结构如图 4 所示。

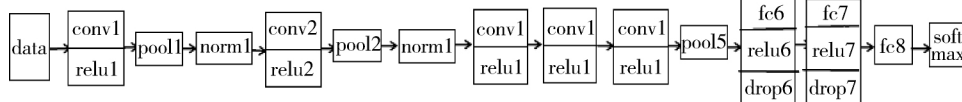


图4 本文网络结构

本文中深度卷积神经网络包含 5 个卷积层(conv), 3 个池化层(pool) 2 个 LRN 层(norm)^[7] 2 个随机采样层(drop) 3 个全连接层(fc) 和 1 个 softmax 分类回归层。

卷积层参数分别是: conv1、conv2、conv3、conv4、conv5 的 blob 类型分别为 $[1, 96, 55, 55]$ 、 $[1, 256, 27, 27]$ 、 $[1, 384, 13, 13]$ 、 $[1, 384, 13, 13]$ 和 $[1, 256, 13, 13]$,步长分别为 4、2、1、1、1。pool 层参数为: pool1: $[1, 96, 27, 27]$, pool2: $[1, 256, 13, 13]$, pool5: $[1, 256, 6, 6]$ 。其中,卷积的计算公式为:

$$X_j^l = \text{ReLU}(\sum_{i \in M_j} X_i^{l-1} * K_{ij}^l + B_j^l) \quad (1)$$

式(1)中 M_j 为输入特征图集合, X_j^l 为当前第 l 层的第 j 个输出, K_{ij}^l 为卷积核,分别对输入特征图 X_i^{l-1} 进行卷积, B_j^l 为偏置,ReLU 表示激活函数,上一节讲到。

卷积层的输出维数的计算公式为:

$$N_2 = (N_1 - F_1 + 2P) / S + 1 \quad (2)$$

其中,输入图片的大小为 $N_1 \times N_1$,卷积核的大小为 $F_1 \times F_1$,步长为 S , P 表示 padding 的像素数,即拓展宽度。输出的图片大小为 $N_2 \times N_2$ 。

pool 池化层的输出维数计算公式如下,其中池化 pool 层的核大小为 F_2 。

$$N_3 = (N_1 - F_2) / S + 1 \quad (3)$$

3 实验结果

本文的图像识别和分类算法是在 ubuntu14.04 环境下, 基于 Python 语言, 采用 CNN 模型框架^[7], Caffe^[8]深度学习框架提供的 CaffeNet 预训练模型, Caffe 是一种流行的清晰高效的深度学习框架, 支持 Python 和 MATLAB 接口和命令行。

3.1 图像的预处理

本文在对模型训练之前, 使用 Python 爬虫技术获取 10 类共 993 张图片, 分为 200 张图片的测试集和 793 张图片的训练集。卷积神经网络对于图像本身来说可以直接输入而不用复杂的预处理操作, 由于硬件条件限制, 本文只是对图像进行了分辨率的统一, 变换成了 256×256 , 然后将 10 类中各个类随机抽取 20 张放入测试集, 剩余图片放入训练集, 之后使用 caffe 提供的均值计算文件, 将图片减去均值再进行训练, 通过计算可以减少图片数据之间的相似性, 从而大大提高训练的精度以及速度。

3.2 实验结果分析

由于硬件条件约束, 本文在纯 CPU 下运行深度神经网络, 运行时间较长。本文使用 anaconda 科学计算环境, 其中可视化工具是 anaconda 自带的 jupyter notebook 工具, 使用其来图形化模型中的一些参数在整个过程中的变化趋势。如图 5 所示, 横坐标为迭代次数, 纵坐标为损失值和准确率。

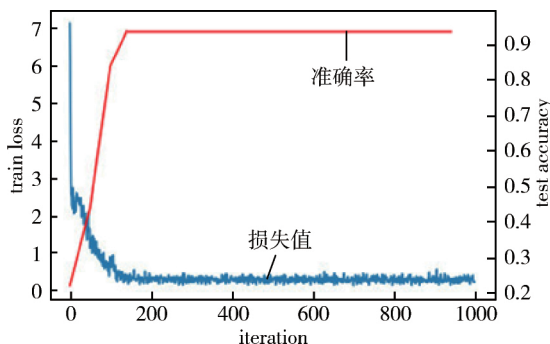


图5 实验结果

本文根据自己的数据集设定了具体的网络参数, 图 5 为迭代 1 000 次, 每 50 次迭代就在测试集上测试一次训练学习的网络, 输出损失值和准确率。对于每 200 次迭代, 取一次当前状态的快照。经过多次修改 Batch-Size 的值后进行多次训练, 在 Batch-Size 设置为 50 时达到最终收敛精度上的最优, 该模型对图像的平均识别率最高, 达到 92.50%。

经过分析得出, Batch-Size 过小会引起识别率震荡过大。调整 Batch-Size 的值能够改善识别准确率的原因在于, 在数据集较小的情况下, 适当增大 Batch-Size, 其确定

的下降方向就越准, 可以降低训练震荡, 提高 CPU 利用率, 提升大矩阵乘法计算效率。由于最终收敛精度会陷入不同的局部极值, 因此 Batch-Size 增大到某个值时, 达到最终收敛精度上的最优。

4 结论

本文使用深度卷积神经网络进行图像分类和识别, 能够有效提高图像识别的准确度, 算法的准确率相对于传统图像识别算法有相当大的提升。本文针对 CaffeNet 网络模型进行了微调, 生成了针对本数据集的最优识别网络模型, 而且总结了深度卷积神经网络的一些优点: 去除了传统图像识别算法的预处理环节, 通过卷积层的稀疏连接 (局部感受野) 和权值共享大大减少了参数, 提升了训练速度, 减少了训练时间。本文存在的不足在于硬件环境较差, 样本较小, 但是算法有较强的扩展性, 下一步可以通过增加样本的类别和图片数量, 以及进一步优化微调本文的网络结构模型, 来提升图像识别的准确度。

参考文献

- [1] 刘炳源. 基于深度学习的特征表示和图像分类方法研究[D]. 北京: 中国科学院大学, 2015.
- [2] HINTON G, SALAKHUTDINOV R. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313 (5786): 504-507.
- [3] 操小文, 薄华. 基于卷积神经网络的手势识别研究[J]. 微型机与应用, 2016, 35(9): 55-57, 61.
- [4] 王龙, 刘辉, 王彬, 等. 结合肤色模型和卷积神经网络的手势识别方法[J]. 计算机工程与应用, 2017, 53(6): 209-214.
- [5] 李志明. 基于卷积神经网络的虹膜活体检测算法研究[J]. 计算机工程, 2016, 42(5): 239-243, 248.
- [6] 邓柳, 汪子杰. 基于深度卷积神经网络的车型识别研究[J]. 计算机应用研究, 2016, 33(3): 930-932.
- [7] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]. Advances in Neural Information Processing Systems, 2012: 1006-1114.
- [8] JIA Y, SHELHAMER E, DONAHUE J, et al. Caffe: Convolutional architecture for fast feature embedding[C]. Proceedings of the 22nd ACM International Conference on Multimedia. ACM, 2014: 675-678.

(收稿日期: 2017-04-14)

作者简介:

韩星烁(1992-), 男, 硕士研究生, 主要研究方向: 机器学习与深度学习、图像处理。

林伟(1968-), 男, 副研究员, 硕士生导师, 主要研究方向: 敏感材料与器件、集成电路设计等。