

基于 CNN 的图片分类

于汶平¹⁾

¹⁾(西安邮电大学 计算机学院, 西安市 710061)

摘 要 本文针对图像的低层视觉特征和高层语义特征之间的鸿沟, 利用一个双输出的 CNN 卷积神经网络, 提取图像主要特征。CNN 是一种多层神经网络, 基于人工神经网络, 在人工神经网络前, 用滤波器进行特征抽取, 使用卷积核作为特征抽取器, 自动训练特征抽取器, 就是说卷积核以及阈值参数这些都需要由网络去学习。图像可以直接作为网络的输入, 避免了传统识别算法中复杂的特征提取和数据重建过程。

关键词 CNN 神经网络; 图像分类

Image Classification Based on CNN

Yu Wenping¹⁾

¹⁾(Department of Computer Science, Xi'an University of Post and Telecommunications, Xi'an 710061, China)

Abstract Based on image of the gap between the low-level visual features and high-level semantic features, using a double output convolution neural network (CNN), extracting image features. CNN is a kind of multilayer neural networks, based on artificial neural network, in front of the artificial neural network, feature extraction, using filter using convolution kernels as feature extraction, automatic feature extraction, training is convolution kernels and threshold parameter which is needed by the network to learn. Images can be directly as the network input, and avoids the complex in traditional recognition algorithm of feature extraction and data reconstruction process.

Key words CNN neural network, image classification

1 引言

传统的基于内容的图像检索是通过提取图像的颜色、纹理、形状等视觉特征来实现检索的。然而, 这些图像的低层信息反应的只是图像的一些客观统计特性, 并不能真正被人们理解。过去几年, 深度学习在解决诸如视觉识别、语音识别和自然语言处理等很多问题方面都表现出色。在不同类型的神经网络当中, 卷积神经网络等深度学习方法得到了大量的研究。早期由于缺乏训练数据和计算能力, 要在不产生过拟合的情况下训练高性能卷积神经网络是很困难的。标记数据和近来GPU的发展, 使得卷积神经网络研究涌现并取得一流结果。

最近, 随着深度学习^[1]的发展, 卷积神经网络^[2]在图像识别领域取得了重要进展, 如图像分类和物体检测。卷积神经网络(CNN)是当前图像识别领域

的研究热点, 利用预训练的CNN网络提取的图像特征展示出了较强的图像识别能力。

1.1 CNN模型简介

卷积神经网络(Convolutional Neural Networks, CNN)是多层感知机(MLP)的变种。CNN 由纽约大学的 Yann LeCun 于 1998 年提出。CNN 本质上是一个多层感知机, 其成功的原因关键在于它所采用的局部连接和共享权值的方式, 一方面减少了的权值的数量使得网络易于优化, 另一方面降低了过拟合的风险。CNN 是神经网络中的一种, 它的权值共享网络结构使之更类似于生物神经网络, 降低了网络

模型的复杂度,减少了权值的数量。该优点在网络的输入是多维图像时表现的更为明显,使图像可以直接作为网络的输入,避免了传统识别算法中复杂的特征提取和数据重建过程。在二维图像处理上有众多优势,如网络能自行抽取图像特征包括颜色、纹理、形状及图像的拓扑结构;在处理二维图像问题上,特别是识别位移、缩放及其它形式扭曲不变性的应用上具有良好的鲁棒性和运算效率等。

CNN 具有一些传统技术所没有的优点:良好的容错能力、并行处理能力和自学习能力,可处理环境信息复杂,背景知识不清楚,推理规则不明确情况下的问题,允许样品有较大的缺损、畸变,运行速度快,自适应性能好,具有较高的分辨率。它是通过结构重组和减少权值将特征抽取功能融合进多层感知器,省略识别前复杂的图像特征抽取过程。

CNN 的泛化能力要显著优于其它方法,卷积神经网络已被应用于模式分类,物体检测和物体识别等方面。利用卷积神经网络建立模式分类器,将卷积神经网络作为通用的模式分类器,直接用于灰度图像。CNN 是一个前馈式神经网络,能从一个二维图像中提取其拓扑结构,采用反向传播算法来优化网络结构,求解网络中的未知参数。

1.2 Cifar-10数据集介绍

Cifar-10 是由 Hinton 的两个大弟子 Alex Krizhevsky、Ilya Sutskever 收集的一个用于普适物体识别的数据集。Cifar 是加拿大政府牵头投资的一个先进科学项目研究所。Hinton、Bengio 和他的学生在 2004 年拿到了 Cifar 投资的少量资金,建立了神经计算和自适应感知项目。这个项目结集了不少计算机科学家、生物学家、电气工程师、神经科学家、物理学家、心理学家,加速推动了 DL 的进程。从这个阵容来看,DL 已经和 ML 系的数据挖掘分的很远了。DL 强调的是自适应感知和人工智能,是计算机与神经科学交叉。DM 强调的是高速、大数据、统计数学分析,是计算机和数学的交叉。

Cifar-10 由 60000 张 32*32 的 RGB 彩色图片构成,共 10 个分类,如图 1-1。50000 张训练,10000 张测试(交叉验证)。这个数据集最大的特点在于将识别迁移到了普适物体,而且应用于多分类。本文在搭建 CNN 模型时只使用了其中的 airplane 和 automobile 两个分类的数据来进行实验。

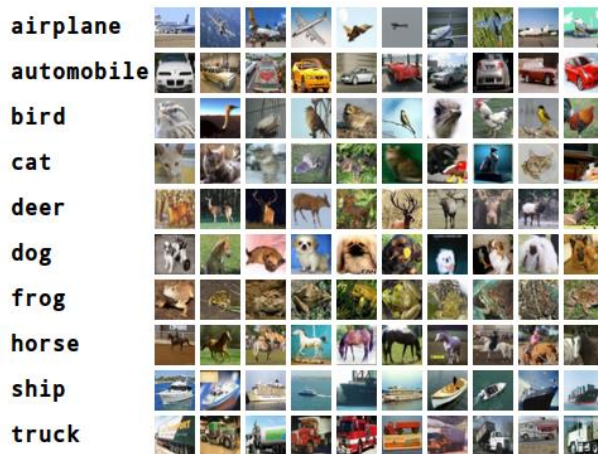


图 1-1 Cifar-10 数据集分类

1.3 Keras环境介绍

Keras 是基于 Theano 的一个深度学习框架,它的设计参考了 Torch,用 Python 语言编写,是一个高度模块化的神经网络库,支持 GPU 和 CPU。

Theano 是一个 Python 库,专门用于定义、优化、求值数学表达式,效率高,适用于多维数组。因为是一个代数符号系统,所以数学表达式里面的一个符号(symbol, Theano 也管这个叫 variable)是一个 Variable 对象,对象之间用加减乘除等操作符连接起来,就变成了一个图(Theano 管这个叫 graph)。在建模一个神经网络的时候将网络表示成为一个巨大的公式(graph)。所以不仅是神经网络,市面上大部分参数化模型都可以用 Theano 实现。

2 基于 CNN 的图像特征提取

一般地,CNN 的基本结构包括两层,其一为特征提取层,每个神经元的输入与前一层的局部接受域相连,并提取该局部的特征。一旦该局部特征被提取后,它与其它特征间的位置关系也随之确定下来;其二是特征映射层,网络的每个计算层由多个特征映射组成,每个特征映射是一个平面,平面上所有神经元的权值相等。特征映射结构采用影响函数核小的 sigmoid 函数作为卷积网络的激活函数,使得特征映射具有位移不变性。此外,由于一个映射面上的神经元共享权值,因而减少了网络自由参数的个数。卷积神经网络中的每一个卷积层都紧跟着一个用来求局部平均与二次提取的计算层,这种特有的两次特征提取结构减小了特征分辨率。

卷积神经网络的卷积层工作流程如图 2-1。

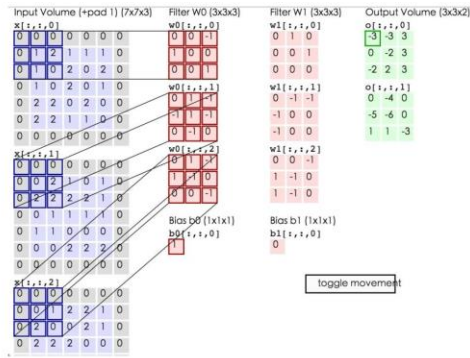


图 2-1 卷积层工作流程

卷积层的 map 个数是在网络初始化指定的,而卷积层的 map 的大小是由卷积核和上一层输入 map 的大小决定的,假设上一层的 map 大小是 $n \times n$ 、卷积核的大小是 $k \times k$, 则该层的 map 大小是 $(n-k+1) \times (n-k+1)$ 。本文构建的 CNN 模型中共存在三个卷积层。第一个卷积层, 32 个卷积核, 每个卷积核大小 5×5 。1 表示输入的图像的通道, 灰度图为 1 通道; 第二个卷积层, 32 个卷积核, 每个卷积核大小 3×3 ; 第三个卷积层, 64 个卷积核, 每个卷积核大小 3×3 。

CNN 是一种特殊的深层的神经网络模型, 它的特殊性体现在两个方面, 一方面它的神经元的连接是全连接的, 另一方面同一层中某些神经元之间的连接的权重是共享的 (即相同的)。它的非全连接和权值共享的网络结构使之更类似于生物神经网络, 降低了网络模型的复杂度 (对于很难学习的深层结构来说, 这是非常重要的), 减少了权值的数量。

3 CNN 的构建与训练

3.1 神经计算及神经网络简介

神经计算^[4]就是通过对人脑的基本单元 - 神经元的建模和联结, 来探索模拟人脑神经系统功能的模型, 并研制一种具有学习、联想、记忆和模式识别等智能信息处理功能的人工系统。

神经元的六个基本特征:

- (1) 神经元及其联接。
- (2) 神经元之间的联接强度决定信号传递的强弱。
- (3) 神经元之间的联接强度是可以随训练改变的。
- (4) 信号可以是起刺激作用的, 也可以是起抑制作用的。
- (5) 一个神经元接受的信号的累积效果决定该神经元的状态。

(6) 每个神经元可以有一个“阈值”。

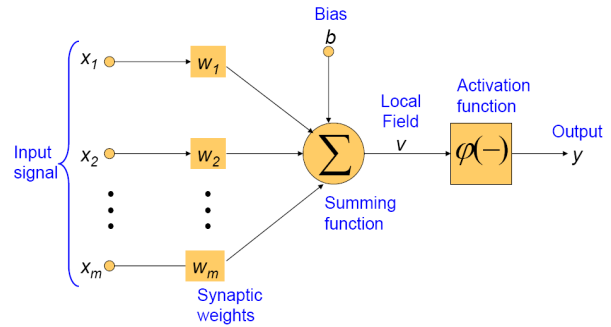


图 3-1 神经网络建模图

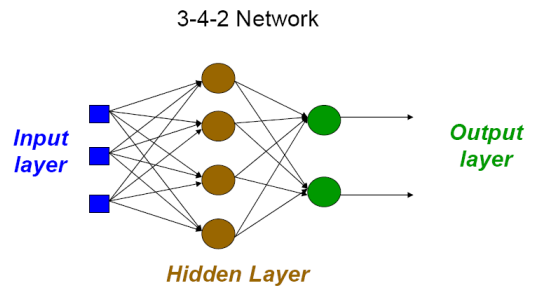


图 3-2 3 层神经网络图

(1) 人工神经网络最具有吸引力的特点是它的学习能力。

(2) 人工神经网络的学习过程就是对它的训练过程。

(3) 学习就是学习神经网络中的联结权重。

3.2 基于 CNN 的激活函数对比与选择

基于 CNN 模型的激活函数目前主要有 Relu、Logistic-Sigmoid、Tanh-Sigmoid 这三种主流的方法, 接下来将对于这三种方法进行对比, 从而选出本文所采用的激活函数。

传统神经网络中最常用的两个激活函数,

Sigmoid 系 (Logistic-Sigmoid、Tanh-Sigmoid) 被视为神经网络的核心所在。从数学上来看, 非线性的 Sigmoid 函数对中央区的信号增益较大, 对两侧区的信号增益小, 在信号的特征空间映射上, 有很好的效果。从神经科学上来看, 中央区酷似神经元的兴奋态, 两侧区酷似神经元的抑制态, 因而在神经网络学习方面, 可以将重点特征推向中央区, 将非重点特征推向两侧区。无论是哪种解释, 看起来都比早期的线性激活函数 ($y=x$), 阶跃激活函数 ($-1/1, 0/1$) 高明了不少。

2001 年, 神经科学家 Dayan、Abott 从生物学角度, 模拟出了脑神经元接受信号更精确的激活模型, 这个模型对比 Sigmoid 系主要变化有三点: ①

单侧抑制 ②相对宽阔的兴奋边界 ③稀疏激活性
同年, Charles Dugas 等人在做正数回归预测论文^[5]中偶然使用了 Softplus 函数, Softplus 函数是 Logistic-Sigmoid 函数原函数。Softplus(x)=log(1+ex) 按照论文的说法, 一开始想要使用一个指数函数 (天然正数) 作为激活函数来回归, 但是到后期梯度实在太太, 难以训练, 于是加了一个 log 来减缓上升趋势。加了 1 是为了保证非负性。

所以根据对比分析, 我们可以得出结论, 相较于原有的 Sigmoid 系激活函数, Relu 激活函数^[6]具有更好的信息解离性, 可以使特征具有更好的鲁棒性; 使稀疏特征有更大可能线性可分, 或者对于非线性映射机制有更小的依赖; 假如神经元激活具有稀疏性, 那么不同激活路径上: 不同数量 (选择性不激活)、不同功能 (分布式激活), 两种可优化的结构生成的激活路径, 可以更好地从有效的数据的维度上, 学习到相对稀疏的特征, 起到自动化解离效果。因此本文采用的是 Relu 和 Softmax 激活函数。

3.3 基于CNN训练Cifar-10数据集

已有的测试推出了一个 Cifar-10 的 CNN 深度、广度基本结构, 理论上这个网络容量能够支持把验证集错误率降到 25% 左右。结构如下图 3-3:

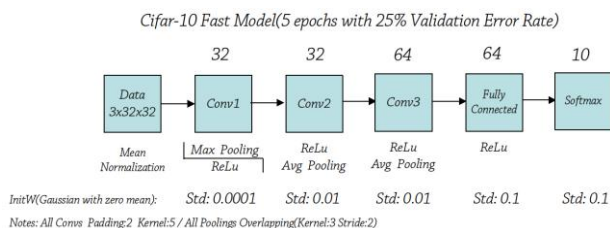


图 3-3 CNN 训练模型结构图

训练算法主要步骤包括:

- (1) 导入各类需要的包;
- (2) 加载数据;
- (3) 测试数据归一化;
- (4) 打乱数据;
- (5) label 为 0~1 共 2 个类别, keras 要求格式为 binary class matrices(即原始标签为 0 则转化为(1,0), 原始标签为 1 则转化为(0,1)), 转化一下, 可以直接调用 keras 提供的这个函数;
- (6) 构建 CNN 模型;
- (7) 生成一个 model;
- (8) 第一个卷积层, 32 个卷积核, 每个卷积核大小 5*5。1 表示输入的图片的通

道, 灰度图为 1 通道, 3 表示 RGB3 通道图像, 这里采用 3 通道图像进行卷积。激活函数用 relu, 采用 Maxpooling2D, poolsize 为(2,2), 为防止训练结果过拟合, 还可以在

model.add(Activation('relu'))后加上 dropout 的函数:

model.add(Dropout(0.5));

- (9) 第二个卷积层, 32 个卷积核, 每个卷积核大小 3*3。激活函数用 relu, 采用 AveragePooling2D, poolsize 为(2,2);
- (10) 第三个卷积层, 64 个卷积核, 每个卷积核大小 3*3。激活函数用 relu, 采用 AveragePooling2D, poolsize 为(2,2);
- (11) 全连接层, 先将前一层输出的二维特征图 flatten 为一维的, 然后输入全连接层;
- (12) softmax 分类, 输出是 2 类别;
- (13) 开始训练模型, 使用随机梯度下降法;
- (14) 使用 SGD + momentum, model.compile 里的参数 loss 就是损失函数(目标函数);
- (15) 调用 fit 方法, 就是一个训练过程。训练的 epoch 数设为 10, batch_size 为 100。batch_size 为随机梯度下降法求解最优权重的计算大小即每次迭代数量大小。

由此整个训练结束, 上文提到的随机梯度下降算法为: 一个典型的机器学习的过程, 首先给出一组输入数据 X, 我们的算法会通过一系列的过程得到一个估计的函数, 这个函数有能力对没有见过的新数据给出一个新的估计 Y, 也被称为构建一个模型。

我们用 X1、X2...Xn 去描述 feature 里面的分量, 用 Y 来描述我们的估计, 得到一下模型:

$$h(x) = h_{\theta}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 \quad (13)$$

我们需要一种机制去评价这个模型对数据的描述到底够不够准确, 而采集的数据 x、y 通常来说是存在误差的 (多数情况下误差服从高斯分布), 于是, 自然的, 引入误差函数:

$$J(\theta) = \frac{1}{2} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 \quad (14)$$

$$\min_{\theta} J_{\theta} \quad (15)$$

关键的一点是如何调整 θ 值, 使误差函数 J 最小化。 J 函数构成一个曲面或者曲线, 我们的目的是找到该曲面的最低点。

假设随机站在该曲面的一点, 要以最快的速度到达最低点, 我们当然会沿着坡度最大的方向往下走(梯度的反方向), 用数学描述就是一个求偏导数的过程:

$$\frac{\partial}{\partial \theta} J(\theta) = \frac{\partial}{\partial \theta} \frac{1}{2} \sum_{i=1}^m (h_{\theta}(x) - y)x^{(i)} \quad (16)$$

这样, 参数 θ 的更新过程描述为以下:

$$\theta_i = \theta_i - \alpha \frac{\partial}{\partial \theta} J(\theta) = \theta_i - \alpha (h_{\theta}(x) - y)x^{(i)} \quad (17)$$

3.4 基于CNN的模型测试

将预测标签与原始标签相减, 若结果为零, 则认为分类正确。其中训练数据占数据集 90%, 测试数据占数据集的 10%。根据训练模型导入测试数据, 通过交叉测试, 得到基于 CNN 的训练模型准确率为 94.1%, 总共经过了 10 轮训练。

4 实验结果

训练轮次	训练正确率	测试正确率
第一轮	0.6860	0.7940
第二轮	0.8511	0.8720
第三轮	0.8856	0.8900
第四轮	0.9144	0.9200
第五轮	0.9284	0.9280
第六轮	0.9399	0.9270
第七轮	0.9509	0.9360
第八轮	0.9547	0.9420
第九轮	0.9633	0.9480
第十轮	0.9641	0.9420

参 考 文 献

- [1] Yaz E E. Linear Matrix Inequalities In System And Control Theory[J]. Proceedings of the IEEE, 2002, 86(12):2473-2474.
- [2] Young Soo Moon, Poogyeon Park, Wook Hyun Kwon, & Young Sam Lee. (2001). Delay-dependent robust stabilization of uncertain state-delayed systems. International Journal of Control, 74(14), 1447-1455.
- [3] XIE Wen-lan, SHI Yue-xiang, XIAO Ping. Classification of natural image based on BP neural network. Computer Engineering and Applications 2010 46(2):163-166.
- [4] Fabrikant A, Koutsoupias E, Papadimitriou C. Heuristically optimized trade-offs: A new paradigm for power laws in the Internet [J]. Automata, Languages and Programming. 2002: 781-781.
- [5] Anonymous. Reviewers for The Auk 2010[J]. The Auk, 2010, 127(4):.
- [6] Glorot X, Bordes A, Bengio Y. Deep Sparse Rectifier Neural Networks[J]. Jmlr W & Cp, 2011, 15.
- [7] Fabrikant A, Koutsoupias E, Papadimitriou C. Heuristically optimized trade-offs: A new paradigm for power laws in the Internet [J]. Automata, Languages and Programming. 2002: 781-781.
- [8] 李钊, 卢苇, 邢薇薇, 等. CNN视觉特征的图像检索[J]. 北京邮电大学学报, 2015, 38(b06):103-106.
- [9] 刘万军, 姜庆玲, 张闯. 基于CNN彩色图像边缘检测的车牌定位方法[J]. 自动化学报, 2009, 35(12):1503-1512.