



研究与开发

基于神经网络的相似图像检索

谢仪颀, 杜刚, 朱艳云, 张晨

(中国移动通信集团设计院有限公司, 北京 100080)

摘要: 在大数据的浪潮中, 社会各领域对海量数据分析的热情方兴未艾, 如何准确地在大量图像中检索成为科研工作的新挑战。面对海量数据, 在图像处理过程中可以借助神经网络(neural network, NN)的优势进行特征提取。针对相似图像的检索问题, 提出了卷积神经网络(convolutional neural network, CNN)和图像数据处理结合的方法。首先建立适应场景的卷积神经网络模型, 提取数据特征, 然后利用局部敏感哈希(locality sensitive hashing, LSH)建立索引, 对于系统输入的目标数据, 经过特征提取和哈希映射关系, 在索引中匹配到最适的相似图像。实验结果表明, 该方法的提出, 可以提高运行效率以及检索的精确度。

关键词: 相似图像检索; 神经网络; 局部敏感哈希; 特征向量

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2020233

Similar image retrieval based on neural network

XIE Yidi, DU Gang, ZHU Yanyun, ZHANG Chen

China Mobile Group Design Institute Co., Ltd., Beijing 100080, China

Abstract: In the wave of big data, the enthusiasm of all sectors of society for data analysis runs high. How to retrieve mass images accurately has become a new challenge in the process of scientific research. The neural network (NN) can extract features of data during image processing for big data. Aiming at retrieving similar images, a method that combines Convolutional Neural Network (CNN) and image data processing is proposed. Firstly, a neural network model for specific scenario was established, data features were extracted, and then an index was created based on locality sensitive hashing (LSH). The features of input data of the system were extracted and mapped through hash function. Finally, the most similar image was found by index. The experimental results show that the method proposed in this paper can improve the operating efficiency and accuracy.

Keywords: similar image retrieval, neural network, locality sensitive hashing, feature extraction

1 引言

当下的知识经济社会, 数码设备迅速普及带

动着网络技术的蓬勃发展, 多媒体的形式和功能在不断地推陈出新。其中, 图像在人们生活中发挥着举足轻重的作用(例如在视频帧的处理、医

收稿日期: 2019-09-10; 修回日期: 2020-08-20



学影像分析、网络鉴暴、智能图书馆等众多领域的应用),不断激发着有效检索图像以及图像识别的需求。在这种趋势中人们意识到,海量图像数据中快速检索到用户需要的资源成为了具有社会价值的课题,也成为不小的挑战。

2 综述

2.1 相关背景

2.1.1 图像检索

一般情况下,在已有的图像集合里面搜索指定图像的若干“相似”图像,这个过程定义为“相似性检索”。图像检索技术具有一定的研究历史,从20世纪70年代初期开始,基本分为两类:基于文本的图像检索(text-based image retrieval, TBIR)和基于内容的图像检索(content-based image retrieval, CBIR)。TBIR需要人为地对大量的图像文件进行标注,标注的内容就成为了检索过程中的关键字,描述图像的内容和关键字匹配程度高的文件视为搜索的结果,也可以称之为“以字找图”;CBIR则是以待查找的图像特征为起点,对图像的色彩、纹理等图像视觉内容特征进行分析和检索,在图像集合中找出与之相似的图像,分析的结果主要依据视觉效果上的相似程度,也称作“以图找图”。

在商业应用中,多数的商业系统主要采用TBIR的方式,京东和淘宝购物平台的图片检索是依靠人工标注文本的。TBIR方法的优点是检索的精度高且速度快,但存在的问题是对每一张图片要进行手动地人工标注,在海量的数据压力下,必然投入更多的人力资源,与此同时,不同的人对图像的理解不同,标注信息也就不尽相同,人工标注的误差和主观色彩带来的干扰难以避免。

CBIR系统从原始的图像中提取“特征”数据,然后根据一定的规则定义图像之间相似的程度。面对图像数据的“维度灾难”,特征提取用多维空间的坐标来描述信息,处理结果可能在几维至数

千维之间。用多维空间中的向量关系考察数据之间的相似程度是CBIR的优势,给问题解决提供了新的思路,让形象的图片对比抽象成为数学模型。

2.1.2 卷积神经网络

图像检索的基本前提是图片数据的特征提取,卷积神经网络在这个方面贡献颇多。卷积神经网络是人工神经网络(artificial neural network, ANN)的一种,也属于前馈神经网络的范畴,通常用来处理图像数据。卷积神经网络前向传播、计算输出的过程就是前馈,这个过程仅仅进行前向传播计算输出,后一层的输入就是前一层的输出,按照这个规律推演到最后一层,就能得到最终的输出结果,在前向传播过程中,网络参数不会更改。神经网络还包含反向传播(back propagation)过程,首先利用损失函数计算损失(loss),然后把误差从最后一层往前传递,同时优化每一层的网络参数,也就是调整数据处理的权值。在预先设置好损失函数、输入和网络结构的情况下,卷积神经网络可以逐步处理输入数据并且自动学习网络参数。

文本借鉴了计算机视觉领域中的迁移学习方式,将优秀学者们在海量的图像数据集中训练成熟的网络参数移植到项目中,训练数据或者提取特征,比如大型可视化数据库ImageNet上训练出来的可用网络参数,这样可以降低训练成本。鉴于此,本文采用了卷积神经网络ResNet-50。在本文的场景中,经过前期的数据预处理,加载了ResNet-50的预训练模型,取得了较为理想的效果,结果表明,模型的预测准确率得到了较大地提升。

2.1.3 局部敏感哈希

提取了图像的特征数据之后,才能将其置入图像集合的检索流程中。检索过程是:根据一定的距离度量定义,从大型的高维数据集中找到与目标检索图像接近或者完全一致的若干对象,

然后精确地度量出和目标向量之间的距离。在图像集合中的数据量较小的情况下,传统的顺序遍历检索方案也是可行的,换句话说是将目标图像特征和集合中图像特征逐一地比对,保留相似程度,最后得到形似程度最高的图像集合。但是更多的情况是,图像集合中数据量很大,用传统方法耗费时间过长,无法应用到实际的生产中去,我们参考了索引相关的资料,通过构建合适的索引来缩短检索的时间。常用的高维数据索引技术有:基于树型布局的索引、基于聚类的索引、倒排文件索引、基于哈希的索引。本文目标是寻找相似的图像,基于哈希的索引方式中的局部敏感哈希技术更符合我们的研究要求。在局部敏感哈希系统中,主要的任务就是找到一种 Hash 方程,完成高维特征到低维特征的 Hash 映射,并且原特征空间中接近的点,在映射产生的多维空间中仍然接近。

2.2 国内外研究现状

在图像处理的领域中,神经网络的应用比较广泛。Hinton 等^[1]结合深度神经网络,训练海量数据集然后提取特征,在图像处理的过程中充分发挥了不同结构神经网络的优势,深度发掘图像数据中隐藏的特征。参考文献[2]提到了用数据样

本去训练多个深度神经网络,然后用多个深度卷积网络进行分类。Gayathiri 等^[3]通过 Hadoop MapReduce 框架引入了一种新的局部敏感哈希算法,以提高对大型数据集样本的任意读取的熟练程度。Xia 等^[4]提到卷积神经网络哈希(convolutional neural network hashing, CNNH)方法,先分解相似度矩阵(矩阵中的元素表示样本是否具有相似性),得到样本的二值编码,然后再利用卷积神经网络拟合数据。仿真结果证明检索效果有了明显的提升。参考文献[5-6]均提出将深度学习用作特征提取,并且将特征向量处理成二进制形式。Lai 等^[7]提出了深度神经网络哈希(deep neural network hashing, DNNH)方法,主要完成以下工作:用 3 幅图像数据输入神经网络训练,通过共享的网络结构得到特征;使用 divide-and-module 的模块将图像的特征切分为若干个 branch,每一个 branch 对应 1 bit;在模型编译过程中使用三元组损失函数。

3 算法描述

3.1 算法流程

本文的相似图像检索流程如图 1 所示,首先

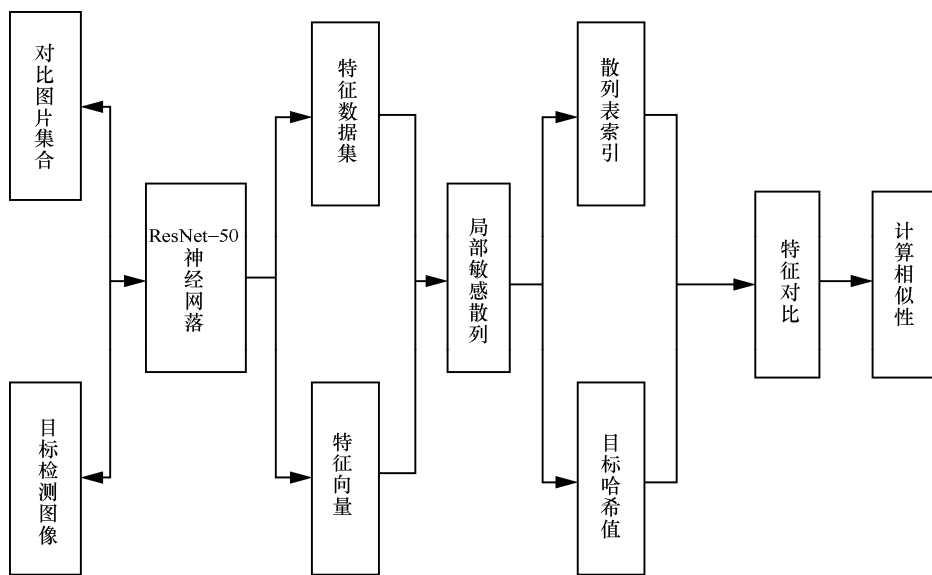


图 1 相似图像检索流程



导入可供检索的图片“对比图像集合”，提取数据库图像，然后进行神经网络的搭建，提取图像的特征并建立数据库图像的特征数据集；然后提取查询图像的全局特征，并生成查询的特征向量；再设置度量方式和返回数目，最终通过相似性计算输出查询结果。

3.2 神经网络模型

卷积神经网络最初阶段是受到了视觉系统的神经机制的启发，是针对二维形状的识别设计，属于生物物理模型，一般由输入层、交替的卷积层和池化层、全连接层和输出层构成。

卷积层也称为“检测层”，通过卷积核来抽取特征，并确定抽取出的特征和其他特征间的关系，卷积核一般是需要训练的，在一些情况下也可以是固定的。对于每一个卷积核，神经元的输入连接上一层的输出。通常一个卷积层会利用一个池化层（下采样层）进行降采样，这种方法也可以增强图像对平移和形变的容忍程度。通常，神经网络中卷积层和池化层的分布状态是交替出现的。

表 1 ResNet-50 网络结构

层	输出维度	卷积核
		7×7, 64 stride2
conv1	112×112	3×3 max_pool, stride2
		$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv2_x	56×56	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$
conv3_x	28×28	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$
conv4_x	14×14	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
conv5_x	7×7	
output	1×1	

残差网络的引入，解决了深度学习中的退化

问题，包含跨层连接，构造了残差模块。其中比较典型的网络包括 ResNet-50，它包含“Bottleneck”单元。“Bottleneck”单元是网络中的基础结构，用于减少计算和参数数量。ResNet-50 模型的结构详解见表 1，“conv”层中的“×”表示 Bottleneck 重复的次数。ResNet-50 尾部的 Dense（全连接层）将数据处理成 1 000 维的形式，不符合笔者的预期，所以对 ResNet-50 模型加以调整，截取网络尾部激活层位置的输出，将其添加到最大池化层，将输出的数据做扁平化处理，就得到了图像的 2 048 维特征。

3.3 数据预处理

本文已经确定了使用 ResNet-50 作为特征提取的模型，其标准输入数据格式是 (224,224,3)，由于需要检测的目标图像的像素分布不确定，为了提高系统的通用性并确保数据集的特征一致，逻辑中把图片统一处理成 (224,224,3) 类型的数据作为神经网络的输入。

3.4 特征提取

特征向量是由若干具有特定意义的数值组成的，特征提取是相似图像检索的关键环节，为了分析相似图像集合，本文加载了 ResNet-50 模型以及预训练权重。实际上，就是将输出层去掉，然后把预训练模型当作一个具有特征提取功能的中间层，输出形如 (1,2 048) 的特征，从而应用到新的数据集中。

3.5 特征度量方式

两个特征向量之间的距离（相似性）定义方式并不唯一，确定它们之间相似度的方法见表 2，以两个特征向量为例，分别表示为 k_p 和 k_q ，其中， $k_p = \{k_{p1}, k_{p2}, k_{p3}, \dots, k_{pn}\}$, $k_q = \{k_{q1}, k_{q2}, k_{q3}, \dots, k_{qn}\}$ 。

本文选择余弦距离作为度量方式，计算两个特征向量之间的相似性。

3.6 LSH 索引建立

局部敏感哈希定义：

一个哈希函数满足以下性质时，被称为

表 2 特征向量相似性计算

欧氏距离	标准化欧氏距离	绝对值距离	余弦距离	相关系数距离
$\sqrt{\sum_{i=1}^N (k_{pi} - k_{qi})^2}$	$\sqrt{\sum_{i=1}^N \left(\frac{k_{pi} - k_{qi}}{S_i} \right)^2}$	$\sum_{i=1}^N k_{pi} - k_{qi} $	$\frac{\sum_{i=1}^N k_{pi} k_{qi}}{\sum_{i=1}^N k_{pi}^2 \sum_{i=1}^N k_{qi}^2}$	$\frac{\sum_{i=1}^N (k_{pi} - \bar{k}_p)(k_{qi} - \bar{k}_q)}{\sqrt{\sum_{i=1}^N (k_{pi} - \bar{k}_p)^2 \sum_{i=1}^N (k_{qi} - \bar{k}_q)^2}}$

(R, cR, P_1, P_2) -sensitive, 对于高维空间中任意的两个点 x, y , 有如下关系存在,

$$\text{当 } \text{dist}(x, y) \leq R \quad P(h(x) = h(y)) \geq p_1 \quad (1)$$

$$\text{当 } \text{dist}(x, y) \geq cR \quad P(h(x) = h(y)) \leq p_2 \quad (2)$$

局部敏感哈希是以 K, L 为参数的随机过程, 其中 K 是哈希映射后数据的大小, L 是哈希表的数量, 经过一个哈希函数的映射, 2 048 维度的特征向量 E 转换成 K 位的哈希数据 H , 每一个哈希函数的映射都是对 2 048 维特征数据进行地加权处理, 强化了局部数据的地位, 也是对特征向量的局部敏感哈希, 在提取局部数据的同时完成了降维工作, 以相同的哈希结果作为标志的图像被放入同一个“桶”中, 一个“桶”内的任意一对图像之间都发生了一次“碰撞”, 这些具有相同映射结果的图像具有一定相似性, 每个图像都会经过 L 次哈希, 分别放入不同的表对应的“桶”中, 一对图像发生“碰撞”的次数越多, 相似度越高。构建局部敏感哈希索引的过程如图 2 所示。

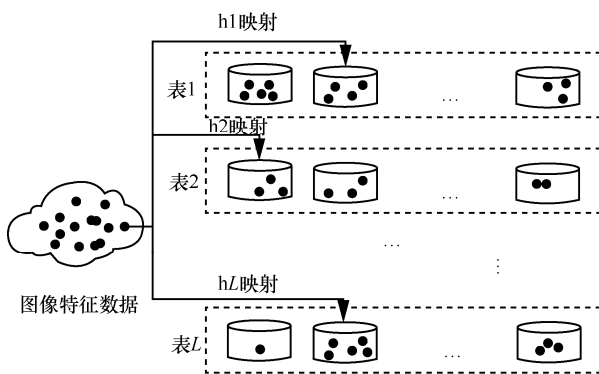


图 2 局部敏感哈希索引构建示意

4 仿真分析

本文选择 Python 作为工具语言进行仿真。利用开源人工神经网络库 Keras, 其作为 Tensorflow、Microsoft-CNTK 和 Theano 的高阶应用程序接口, 在深度学习模型的设计、调试、评估、应用和可视化方面性能优异。Kera 的应用模块 Application 提供了带有预训练权重的 Keras 模型 ResNet-50, 可以用来进行预测、特征提取和微调。对于 ResNet-50 模型的初始化, 使用“image-net”训练的参数, 并导入 400 张像素尺寸不同的图像作为对比对象。

在搜索的时候使用一张目标检测图像, 如图 3 所示。



图 3 图像检索系统输入文件

输入网络进行提取得到如下特征, 然后系统输出图形化的比对结果, 如图 4 所示。

[[1.446 400 3 4.171 128 7 4.595 113 8 ...
1.880 205 9 2.513 447 8 12.270 57 1]]

5 结束语

本文提出了一种基于神经网络的相似图像检索策略, 在准确率、计算时间上具有优势。本文

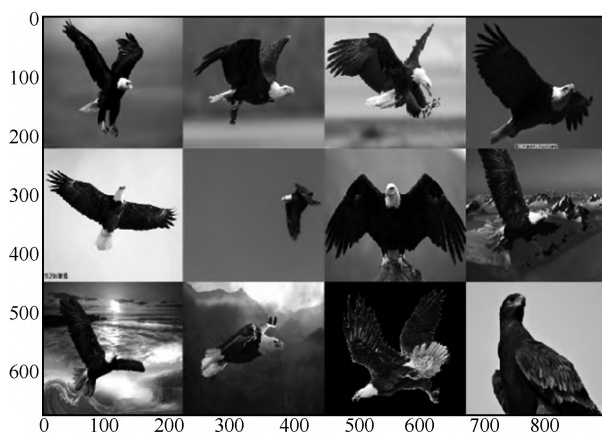


图4 图像检索结果

的系统主要应用了具有特征提取功能的神经网络模型和具有检索功能的哈希索引表。在检索之前,要把对比图像集合转化成可检索的数据集形式,利用 ResNet-50 结构对大量的图片文件进行特征提取,首先对图像数据进行预处理,以适应 ResNet-50 的输入维度,然后利用局部敏感哈希原理映射成二进制形式的字符,整个索引表可以看成由“桶”排列的,其中包含着产生哈希碰撞的相似图片;当系统接收目标检测图像,同样进行特征提取然后在哈希索引表中寻找最相似图像集合,用余弦度量距离得到精确匹配结果,利用可视化界面进行展示。

神经网络的发展促进了图像检索领域技术的提高,让图像的匹配有依据可循,这个方向仍然值得相关领域的科研工作者进一步探索,以融入大数据时代的潮流中。

参考文献:

- [1] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.
- [2] SCHMIDHUBER J, MEIER U, CIRESAN D. Multi-column deep neural networks for image classification[C]//Proceedings of Computer Vision & Pattern Recognition. Piscataway: IEEE Press, 2012: 3642-3649.
- [3] GAYATHIRI N R, NATARAJAN A M. MapReduce-based storage and indexing for big health data[J]. Concurrency and

Computation: Practice and Experience, 2019, 31(14).

- [4] XIA R K, PAN Y, LAI H J, et al. Supervised hashing for image retrieval via image representation learning[C]//Proceedings of AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2014: 2156-2162.
- [5] WANG J, KUMAR S, CHANG S F. Semi-supervised hashing for scalable image retrieval[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2010: 3424-3431.
- [6] JAIN P, KULIS B, GRAUMAN K. Fast image search for learned metrics[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2008: 1-8.
- [7] LAI H J, PAN Y, LIU Y, et al. Simultaneous feature learning and hash coding with deep neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2015: 3270-3278.

[作者简介]



谢仪颀(1994-),女,中国移动通信集团设计院有限公司助理咨询设计师,主要研究方向为信息安全、计算机视觉和语音。



杜刚(1984-),男,博士,中国移动通信集团设计院有限公司高级工程师,主要研究方向为信息安全、人工智能。



张晨(1980-),男,中国移动通信集团设计院有限公司高级工程师,主要研究方向为信息安全、网络安全、内容安全、大数据分析。

朱艳云(1979-),女,中国移动通信集团设计院有限公司高级工程师,主要研究方向为网络与信息安全。