谨以此论文献给我的恩师、父母及朋友

—— 杨攀

深度哈希特征及其应用研究

学位论文答辩日期：\_\_\_\_\_\_\_\_\_\_\_\_\_\_

指导教师签字：\_\_\_\_\_\_\_\_\_\_\_\_\_\_

答辩委员会成员签字：\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**独 创 声 明**

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。据我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含未获得 （注：如没有其他需要特别声明的，本栏可空）或其他教育机构的学位或证书使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

学位论文作者签名： 签字日期： 年 月 日

---------------------------------------------------------------------

**学位论文版权使用授权书**

本学位论文作者完全了解学校有关保留、使用学位论文的规定，有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许论文被查阅和借阅。本人授权学校可以将学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。同时授权中国科学技术信息研究所将本学位论文收录到《中国学位论文全文数据库》，并通过网络向社会公众提供信息服务。（保密的学位论文在解密后适用本授权书）

学位论文作者签名： 导师签字：

签字日期： 年 月 日 签字日期： 年 月 日

深度哈希特征及其应用研究

# 摘 要

关键词：

# Abstract

Keywords：

目 录

[摘 要 I](#_Toc444520180)

[Abstract II](#_Toc444520181)

[1 绪论 1](#_Toc444520182)

[1.1 研究的背景及意义 1](#_Toc444520183)

[1.2 国内外研究和应用现状 2](#_Toc444520184)

[1.2.1 深度学习研究 2](#_Toc444520185)

[1.2.2 基于内容的图像检索研究 2](#_Toc444520186)

[1.2.3 图像深度估计研究 2](#_Toc444520187)

[1.3 论文的创新性 2](#_Toc444520188)

[1.4 论文的主要内容及结构安排 3](#_Toc444520189)

[2 深度学习网络模型 4](#_Toc444520190)

[2.1 深度卷积神经网络(CNN) 4](#_Toc444520191)

[2.1.1 卷积运算 4](#_Toc444520192)

[2.1.2 池化操作 6](#_Toc444520193)

[2.1.3 卷积神经网络设计机理 9](#_Toc444520194)

[2.2 深度生成模型 13](#_Toc444520195)

[2.2.1 玻尔兹曼机 13](#_Toc444520196)

[2.2.2 受限玻尔兹曼机（RBMs） 14](#_Toc444520197)

[2.2.3 实值受限玻尔兹曼机 16](#_Toc444520198)

[2.3 小结 17](#_Toc444520199)

[3 深度哈希特征学习 19](#_Toc444520200)

[3.1 深度特征学习 19](#_Toc444520201)

[3.2 深度哈希特征 19](#_Toc444520202)

[3.3 小结 19](#_Toc444520203)

[4 深度哈希特征在图像检索中的应用 19](#_Toc444520204)

[4.1 图像检索系统 19](#_Toc444520205)

[4.1.1 基于内容的图像检索系统 19](#_Toc444520206)

[4.2 图像检索中常用特征提取算法 21](#_Toc444520207)

[4.2.1 感知哈希算法 21](#_Toc444520208)

[4.2.2 颜色直方图算法 21](#_Toc444520209)

[4.2.3 内容特征算法 21](#_Toc444520210)

[4.3 深度哈希特征图像检索 21](#_Toc444520211)

[4.4 实验结果 21](#_Toc444520212)

[4.5 小结 21](#_Toc444520213)

[5 深度哈希特征在水下图像深度估计中的应用 22](#_Toc444520214)

[5.1 水下光学成像模型 22](#_Toc444520215)

[5.1.1 水下图像深度估计模型 22](#_Toc444520216)

[5.1.2 22](#_Toc444520217)

[5.2 水下图像深度估计哈希模型 22](#_Toc444520218)

[5.3 实验结果 22](#_Toc444520219)

[5.4 小结 22](#_Toc444520220)

[6 总结与展望 23](#_Toc444520221)

[参考文献 24](#_Toc444520222)

[致谢 26](#_Toc444520223)

[个人简历 27](#_Toc444520224)

# 1绪论

## 1.1 研究的背景及意义

近年来，随着分布式计算技术，高性能计算，大数据技术的飞速发展，结合机器学习技术领域取得的突破性进展，社会已经从之前的大数据认识、积累阶段进入了以“大数据+智能数据分析”为主题的智能大数据分析、应用阶段。谷歌，微软，亚马逊，阿里巴巴，百度等互联网巨头相继成立人工智能实验室，结合自身研究方向，开展自然语言处理，计算机视觉，机器翻译等方面的研究；MetaMind, Clarifai, Deepomatic等一大批初创公司也加入机器学习研究，开发人工智能产品，这些研究背后，都有深度学习的支撑。深度学习作为机器学习领域的研究热点，自从被提出以来，就受到了广泛的关注和研究，《Nature》等顶级期刊相继推出专栏来推动其发展。特征表示学习是机器学习中的一个重要研究方向，表示学习算法是一种给计算机输入原始数据，然后能自动发现数据的表达，以用来检测或分类。深度学习正在取得重大进展，在图像识别，语音识别和机器翻译等领域打破了记录，并且在其他领域挑战传统的机器学习技术，包括医药学分子活性分析，粒子加速器数据分析，大脑回路重建，情感分析等。

互联网的迅速发展，计算机技术的普及，各行各业都积累了海量的业务数据。例如Google每天处理24PB的数据，Facebook注册用户超过10亿，每天生成300TB以上的日志文件，每天有大约3万小时的视频上传到Youtube,淘宝上每天产生数百万订单，而每笔订单的交易信息都有截图记录。根据国际数据公司（IDC）监测，人类每天产生的数据正在呈指数级增长，大约每两年翻一番，并且在2020年前保持这个速度，这意味着人类最近两年产生的数据量相当于之前数据总和。图像作为大数据的重要组成部分，包含了内容丰富的多媒体信息（如颜色，纹理等），在很多领域都有广阔的应用前景。图像数据具有高维度，大体量，内容复杂等特点，虽然借助人的抽象能力可以给图像设置标注、索引等，但网络数据量巨大，大部分数据是不包含标签等信息的，如何选择合理的方式来表示、存储、检索这类数据是急需解决的问题，其中图像检索在互联网中有广泛应用，相比于文字，图像包含更多信息，人的视觉系统经过长期进化，已经发展成为人大脑中最重要的系统之一。基于内容的图像检索技术是图像检索问题的中一个新的研究热点，与传统的基于概念的图像索引不同，基于内容的检索技术尝试分析图像的内容，而不是图像的元数据，如关键字，标签，描述等。“内容”指图像的颜色、形状、纹理，或者其他可以由图像自身得到的信息，借助于深度学习模型的特征表示能力，基于内容的图像检索技术是解决图像自动索引问题的研究重点和趋势。

三维重建是指利用二维图像重建图像中对象的三维模型的过程，是计算机视觉领域中一个重点研究方向。三维模型反映了对象在三维空间中的形状信息，能够全面的展示对象的空间特性，在虚拟现实，医学辅助，计算机动画等多个领域有广泛的应用场景。深度估计是理解场景中几何关系的重要部分，这种几何关系能提供对象和对象所处环境得更丰富的表达，从而帮助提升对象识别等任务的准确度。

## 1.2国内外研究和应用现状

### 1.2.1深度学习研究

深度学习方法是一种特征学习方法，一般由多个简单的非线性模块组成，从输入开始，每个模块将前一层表达变换为更高层次的另一种表达，从而形成了多层的特征表达，高层的表达更抽象。通过组合许多的变换，这种模型能够近似十分复杂的函数。与传统的机器学习技术相比，深度学习的核心是不利用人工设计特征，而是使用一种通用的学习策略，从数据中学习到数据的特征表达。

在文献[1]中，Hinton用一个多层的信念网络来实现高维数据降维，通过在网络中设置一个小的中间层来实现对高维数据的重构，然后用梯度下降方法来对这种类似自编码器的模型进行调优，这种训练方法只有在模型具有较好的初始参数时才能取得好的效果，而文中提出用玻尔兹曼机来学习初始参数，实验证明这种方法比主成分分析方法更有效。在文献[2, 3]（CNNs）中，卷积网络被设计用来处理二维数据，如图像，视频等。CNNs受文献[4]中权值共享思想的启发，通过减少连接来降低模型训练的计算量，是第一种真正意义上取得成功的深度学习模型。2012年Krizhevsky和Alex在文献[5]中利用一种深度卷积网络来对120万张高分辨率的图像进行分类，这些图像包含1000个类别，并且取得了最好的成绩。这个模型包含6千万个参数和65万个神经元，由5个卷积层和3个全连接层组成，为了更快的训练模型，作者利用了GPU来进行加速。同时为了避免过拟合，采用了Dropout技术和数据扩充等技术。文中提出的深度卷积模型作为一种新的特征学习模型，为特征表示学习提供了一种新的思路。Sainath在文献[6]中，利用CNN来实现大词汇量连续语音识别，利用卷积神经网络来降低音谱成分的多变性，对信号中存在的不同谱段之间的关系进行建模。通过选取合适数量的卷积层，恰当的隐含节点数目和最佳的池化策略，然后利用大词汇量连续语音识别任务来评估CNN特征的效果。在文献[7]中，Goodfellow等利用深度卷积模型设计了一个完整的多目标识别任务框架，而不是按照传统的方法将任务分为：定位、分割、识别等部分，在谷歌街景门牌号数据集中达到了97.84%的准确率。Collobert等在文献[8]中利用卷积神经网络和语言模型来预测给定句子中的词性标记，主语，近义词及句子语法上与语义上的合理性可能性。在文献[9]中，作者将卷积神经网络和递归神经网络结合，实现了从图像到语句的映射，让计算机真正学习理解图像，图1-1展示了文中的机器翻译结果。Hinton等在文献中[10]就深度神经网络在语音识别任务中的应用做了详细的介绍，深度神经网络的效果在多个语音识别任务中已经超过高斯混合模型，基于深度神经网络的语音识别系统已经广泛应用于谷歌，百度，科大讯飞等公司开发的系统中。

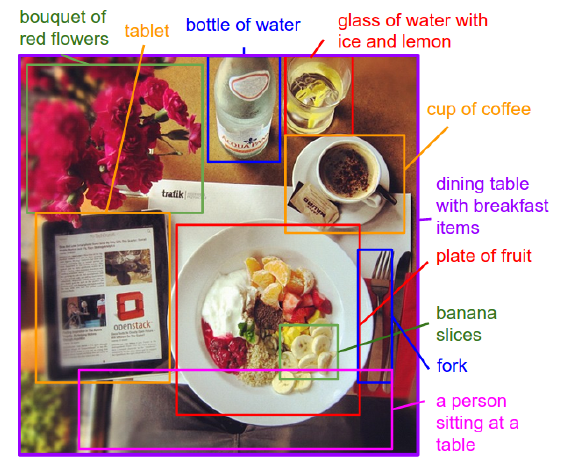


图1-1 机器自动翻译系统

深度学习理论表明深度网络相比于传统的不利用分布式特征的学习算法有两大优势，首先，学习分布式特征表示可以利用新学到的特征泛化出不同的组合（如 元特征可以得到出 种组合）；其次，深度网络中的各层表示的组合带来另一种指数级的优势。

### 1.2.2基于内容的图像检索研究

近年来，随着数字技术的发展，数字图像的收集和应用快速增长。每天都有海量的图像被存储，为了解决海量图像的查找和检索，图像检索成为了一个热点问题。主流的图像检索方法分为基于文本的检索方法和基于内容的检索方法。

早起的基于文本的方法可以追溯到20世纪70年代，一个典型的基于文本的图像检索系统包括以下步骤：（1）人工给图像数据库每个图像添加标注，建立图像文本数据索引，（2）人工给查询图像添加标注，（3）在图像文本数据库中查找，（返回结果），代表文献有[11-13]。然而，当数据量过大时基于文本的检索系统有两个主要的缺陷：首先，数据库中的所有图像都需要人工标注，当数据量太大，图像标注问题成为系统的发展瓶颈；其次，由于人的感知差异，不同人对同一图像的标注不同，导致系统结果的不一致性。

在20世纪90年代早期，由于大量图像数据的积累和基于文本的检索系统的缺陷，基于内容的图像检索系统应运而生。基于内容的图像检索系统提取图像中的颜色、纹理、形状等信息，根据特定的特征描述子获取图像的特征表示，然后检索得到相似图像。检索步骤包括：输入图像呈递给图像描述子，将特征提取算子应用到图像上，然后返回一个（或一组）特征向量，用来量化输入图像的内容。选取一个距离度量或者相似性函数之后，即可以通过比较特征向量来获取相似性。距离度量或者相似性函数以两个特征向量作为输入，然后返回一个反映两个特征向量“相似”程度的实数，文献[14-16]等介绍了相关工作。图1-2 展示两个图像比较的过程。

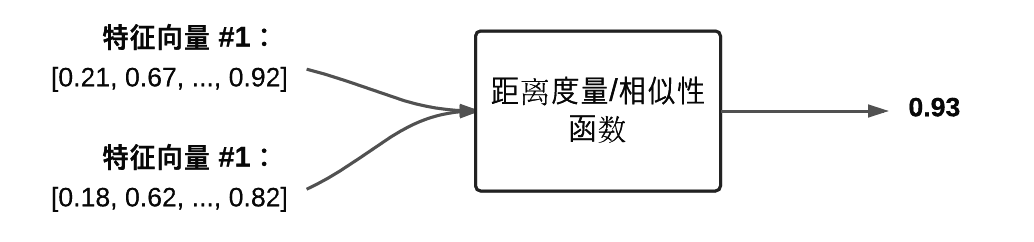


图1-2 两幅图片的相似性计算

任何一种基于内容的图像检索系统都包含以下四部分：

（1）定义图像描述子。在这个部分需要确定我们所要描述的图像内容，是对图像颜色感兴趣，还是对图像中对象的形状感兴趣，亦或者对图像中的纹理感兴趣？

（2）对数据库中数据建立索引。第一步已经确定了图像描述子，接下来我们需要做的是将选定的图像描述子应用到数据库中的每幅图像，提取图像特征，将特征向量写入数据仓库（如csv文件，关系型数据库Redis等），以便之后用来计算相似性。

（3）定义相似性度量。经过第一，二步计算得出一些列特征向量，但在没有定义一个统一的相似性度量之前，无法衡量两个特征向量的距离。常见的选择有欧式距离，余弦距离和卡方距离等，但实际的选择高度依赖于数据类型和所提取得到的特征。

（4）检索。系统的最后一步是运行实际的查找算法。用户可能会提交一幅（或多幅）图像到检索系统（例如通过个人电脑或移动软件上载），系统的任务是：（a）提取查询图像的特征，（b）应用相似性函数来计算查询图像与系统已经存储的特征，（c）根据相似性函数返回最相似的结果。

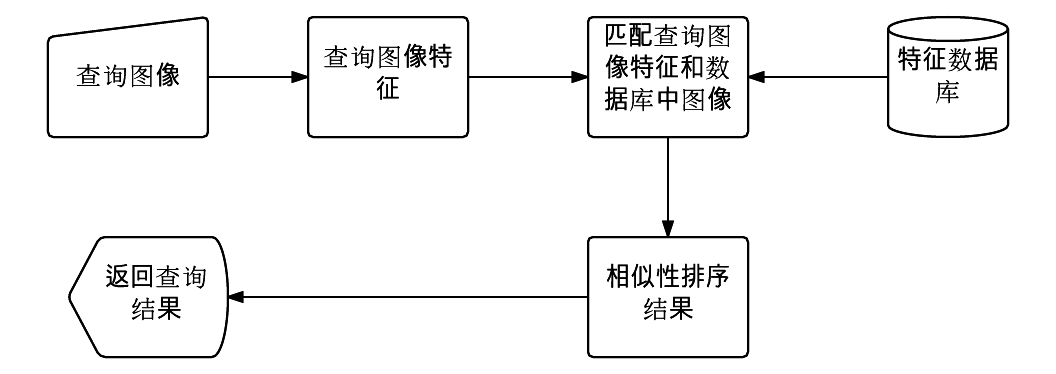


图1-3 在CBIR系统中执行图像查找

以上是任意基于内容的图像检索系统最基本的四部分，随着更多的、更复杂的特征表达的应用，系统的分块会有所增加。在基于内容的图像检索系统中查找相似图像，用户需要提交查询图像，从查询图像中获取“查询特征”后，将查询特征与库中存储的特征做匹配。由于特征数据库中已经建立了索引，查询结果将会按照相关性大小排序并呈现给用户，图1-3 描述了整个检索的过程。

随着基于内容的图像检索研究的发展，一些图像检索系统相继出现，表1-1列举了几个主要的基于内容的图像检索引擎。

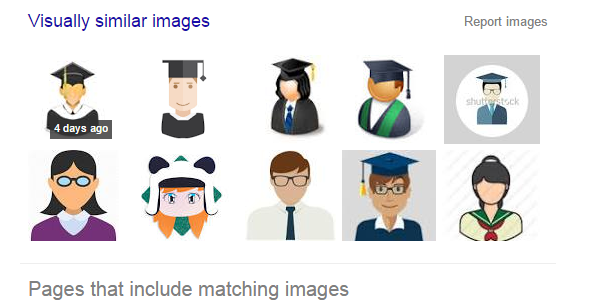
表1-1 商业CBIR引擎

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 名称 | 描述 | 是否支持库外图像检索 | 是否支持元数据检索 | 数据库尺寸 |
| Superfish | 图像对图像检索 | 是 | 是 | Billions |
| TinEye | 基于互联网图像的CBIR系统 | 是 | 否 | 12100M |
| Yandex Image Search | Yandex CBIR系统 | 是 | 是 | 10000M |
| Baidu Image Search | Baidu CBIR系统 | 是 | 是 | 1000M |
| PicScout | 基于互联网图像的CBIR系统 | 是 | 否 | 200M |
| MiPai | 在线图像检索系统 | 是 | 是 | 100M |
| Incogna | Incogna CBIR系统 | 否 | 是 | 100M |

图1-4 展示了谷歌的图像检索系统，用户可以上传图像，在互联网上检索与图像近似的图像和相关的网页链接，同时结合最新的图像识别技术，可以对查询图像进行分类，分类的结果可以帮助搜索引擎提高检索的准确度。



（a）



（b）

图1-4 谷歌的图像检索系统，（a）输入查询图像，

（b）返回相似图像和相关页面

哈希学习使用机器学习方法，利用哈希函数和量化函数将数据映射为一个二进制串，从而降低图像数据的存储开销和计算开销，进而提高图像检索系统的效率。哈希学习的目的是在保持数据在原数据空间的邻近关系的前提下，将数据映射到低维的二进制空间上，即保相似性映射。如图1-5，原始两张相似的图像经过特征学习阶段提取出良好的特征后，再经过量化得到的哈希码也是比较相似的。图中的特征 并不是特征提取算法得到的特征，而是经过压缩函数得到图像的低维空间的表达，最后被量化为一个8位的二进制串，容易得出，原空间相似的图像被映射到相似的2个二进制编码（海明距离为1），原空间中不相似的图像被映射到不相似的二进制编码（海明距离为6）。哈希学习方法不仅能够保持数据间的相似性，还能大幅降低数据特征所需的存储空间，以上文的CNN特征为例，AlexNet第7层网络包含4096个节点，假设最终用第7层输出作为输入图像的特征表示，对ImageNet数据做分类实验，1400万张图片，每张图片特征需要使用32768B的空间，整个数据集需要427GB的存储空间，这个数量级的数据已经超出了一般工作站的处理能力。但如果用哈希特征（64位）表示，整个数据集只需要106M的空间，即使使用个人电脑，也可以运行常用的支持向量机等算法。因此，哈希学习方法在智能大数据分析应用中有重要位置。

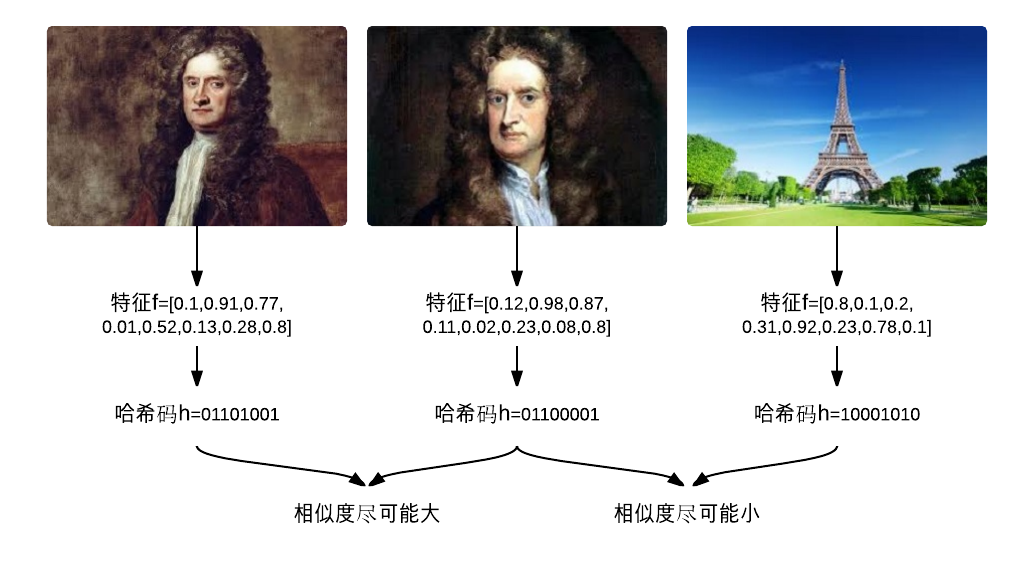


图1-5 哈希学习示意图

根据机器学习的分类，哈希学习也分为监督学习核非监督学习。早期的方法多属于非监督学习方法，非监督方法也被称为数据独立算法，文献[17]中提出局部敏感哈希算法（LSH），是最早用来应用到近似近邻搜索任务中的哈希算法，LSH利用随机投影，保相似性的将数据映射到哈希空间，随机投影不利用额外的计算，使得LSH算法适合大规模的近邻查找任务。在文献[18]和[19]中，Kulis和Brain又相继提出了基于核方法的局部敏感哈希和基于变化相似性尺度的哈希算法。与非监督算法相比，监督算法依赖于数据，所以也成为数据关联算法。语义哈希[20]用一个深层图模型来学习哈希映射，通过设置网络中中间层的节点个数，来达到降维的目的。基于锚点的图哈希[21]和谱哈希[22]使用图分割的方法来获得相似性度量，文献[23]通过最小化原始数据的重建误差和最小化对应二元映射的海明距离来学习哈希函数。迭代量化的主成分分析方法[24]运用主成分分析来初始化投影矩阵，来使得量化误差达到最小。以上所有哈希方法都使用手工设计特征或者原始数据作为哈希函数的输入特征，然而，手工特征只能提取到数据的底层表达，同时哈希函数和特征提取阶段相互独立，因此，需要新的特征提取算法和哈希学习框架来提升哈希学习的效果。

哈希学习适应智能大数据分析的要求，能够处理海量数据。在图像检索领域，哈希学习方法能够提高图像检索的精度和速度，是值得研究的一个方向。

### 1.2.3 图像深度估计研究

从单幅单目图像中估计深度是计算机视觉领域中一个重要问题，许多视觉问题已证明依赖于图像的深度信息，例如，语义标注，姿态估计，尽管深度感应器设备如Kinect等已经可以获取深度信息，但视觉领域使用的数据集仍是彩色图像。之前已经有很多基于立体视觉的深度估计研究，从单幅图像估计深度只有较少进展，与立体视觉方法不同，从单幅图像估计深度需要用到的线索包括图像中线的角度和视点，对象大小，图像方位，空气影响等。

早期的工作主要使用几何假设，如盒式模型，来推断房间的空间分布[25, 26]或室外场景[27]。这些模型具有固有限制：模型只能应用于特定场景，而不能应用到一半场景的深度估计中，文献[28]提出非参数模型，使用候选图像来进行检索，首先用模板图像来匹配预测图像，然后用平滑性限制来优化图像深度估计，这种方法基于相似图像具有相似深度值分布假设，然而这类方法有在不同阶段传播误差的缺陷，并且依赖于足够全面的数据集来进行候选图像检索。在文献[29]Ladicky等证明将深度估计和语义标注结合在一起，可以相互提高。文献[30, 31]等尝试将三维重建技术应用到水下。

## 论文的创新性

随着深度学习在计算机视觉、自然语言处理等各个领域的广泛应用，深度特征表达作为一种新的特征提供了一种分布式的特征表示方式，其特征值的组合和不同特征层的组合带了相比于传统特征的指数级优势。通用的深度卷积特征维度大约为4000维（如4096），对于不同的任务来说，特征维度是有不同需求的。例如针对不同尺度图像的分类任务，可以用不同大小维度的特征来达到较好的分类任务；对于图像检索任务，图像的特征表示维度越低，对提高检索效率和降低存储消耗越有利。为了探究低维的哈希特征是否能满足图像分类，检索等任务的需求，本文设计了深度哈希特征，并将哈希特征应用于图像检索和图像深度估计中，以验证深度哈希特征的有效性。本文的创新性包括以下几点：

（1）基于深度卷积神经网络，提出深度哈希特征的学习模型，这种特征具有维度低（如48维），表达能力强（在常用分类任务中能达到与高维特征近似的分类结果）的特征。

（2）将深度哈希特征应用于图像检索任务，在多个数据集上的实验表明，

深度哈希特征可以有效降低特征的存储空间，同时能够大幅提高图像检索系统的准确度和覆盖率。

（3）将深度哈希学习模型和玻尔兹曼机结合，提出深度图像估计模型，

单幅图像中恢复出物体深度信息，通过使用迁移学习的方法，在小数据集上也取得较好的结果。同时将这种深度哈希模型应用到水下图像的深度估计中，为水下三维重建提供了一种新的思路。

## 1.4 论文的主要内容及结构安排

本文提出深度哈希模型并将哈希特征应用于图像检索和图像深度估计任务中。首先根据VGG网络设计了深度哈希网络，在ImageNet模型的基础上，重新训练对应的图像数据库，通过设置哈希层的节点数和训练参数，获取不同维度大小的哈希特征，训练完成后，将哈希特征量化，用来进行检索，并利用准确率，召回率等来评估检索系统的有效性。深度估计设计方案与检索实验类似。

本文在结构上共分为六章：

第一章 绪论。首先说明了本文研究内容的背景及意义，并对相关领域的国内外研究和应用现状进行了大致介绍，同时阐述了本设计的创新点，最后列出了文章的主要工作内容及结构安排。

第二章 深度学习网络模型。简单介绍了本文使用的两种深度网络模型，说明网络模型的特点和思想，为本文的设计方案提供理论依据。

第三章 深度哈希特征。首先介绍了深度哈希特征的原理及应用场景，然后针对通用哈希特征的缺点提出了深度哈希特征，并对哈希特征的工作原理进行详细分析。

第四章 深度哈希特征应用于图像检索。将本文提出的深度哈希特征应用于图像检索任务，在多个数据集上和传统方法进行对比评估，以验证深度哈希特征的有效性。

第五章 深度哈希特征应用于图像深度估计。将本文提出的深度哈希模型应用于图像深度估计任务中，并结合受限玻尔兹曼机，构建了图像深度预测模型，最后将此模型推广到水下三维重建中，同时展示了图像深度估计的结果。

第六章：总结与展望。对本文中的工作进行了全面的总结，结合本文中未解决的问题和深度学习的发展趋势，提出了相关工作今后的研究方向。

# 2 深度学习网络模型

## 2.1 深度卷积神经网络(CNN)

卷积网络（又名卷积神经网络或CNNs）是一种特殊的人工神经网络，主要被设计用来处理具有网格拓扑的数据。例如时间序列数据，可以认为是在固定时间区间上采样获得的1D网格；图像数据，可以认为是像素点组成的2D网格。卷积网络在实际应用中已经取得了惊人的成功，“卷积神经网络”的名称表明这种网络应用了一种数学运算——卷积，卷积是一种特定的线性运算。卷积网络可以看做是在简单网络的某些层中用卷积替换一般的矩阵乘法的神经网络，图2-1展示了一个用于手写数字分类的卷积神经网络。

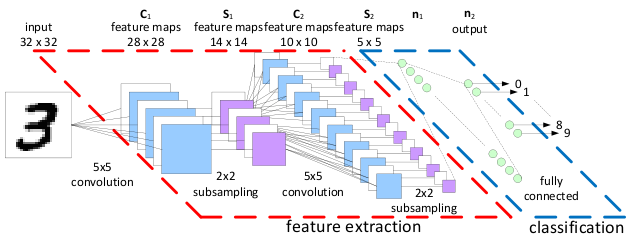


图2-1 卷积神经网络

### 2.1.1 卷积运算

在大部分形式中，卷积是两个函数上的一种运算。假设用激光感应器来定位一艘太空飞船，激光感应器提供单一输出 ，代表 时刻太空飞船的位置。 和 都是实数，即任意时刻我们可以从激光感应器读取到不同的读数。噪声不可避免的出现在各种感应器中，为了尽量避免噪声对估计结果的影响，取多次测量的平均值是常用的方法。显然，越近的测量应该具有更高的相关性，所以加权平均是更合适的方法。设权值函数为 ， 是测量的时间阶段，通过采用这种加权平均方法，得到一个新的函数 ，提供飞船位置的更平滑的估计（由公式（2-1））



这种运算称为卷积，通常卷积运算用星号表示：



在以上例子中， 需要满足条件密度函数的要求，否则输出不是加权平均的结果；同时， 在参数为负数时值为0。这些限制都是针对上述示例，一般来说，卷积运算不限定函数类型，加权平均只是一个简单的应用。

卷积网络术语中，第一个参数（如 ）被称为输入，第二个参数（如 ）被称为核，输出则称为特征图。实际应用中，数据在计算机中表示需要被离散化，上例中，更符合的假设是激光感应器在特定时间区间提供一次测量（如每秒一次测量），而不是在任意瞬时时刻给出测量值。时间索引 取整数值，相应的，在新的 和 上定义离散卷积：



在机器学习应用中，输入往往是多维数组形式的数据，核是一个对应的多维数组，由可以通过学习获得的参数组成，我们用张量来代表多维数组。因为输入张量和核张量中的元素必须独立存储，假设 和 只在有限点集上取值，在其余位置值为0，因此，在实际应用中，无穷求计算和转换为有限点集上的求和计算。最后，以二维张量为例，用二维图像 作为输入，用二维张量 作为核：



由于卷积运算满足交换律，公式（2-4）与下式（由公式（2-5））等价。



在机器学习算法中公式（2-5）具有更直观的理解和应用，因为 和 的有效取值范围相对较小。因为核相对于输入数据做了翻转，不失一般性，假设 增大，输入数据中的索引随着增大，但是核中的索引减小，所以，翻转核的唯一原因是为了利用卷积的可交换性质方便理论证明。实际上，很多机器学习库用相关运算来实现变换，相关和卷积相似除了不翻转核张量：



图2-2是一个在二维张量上应用卷积的示例，在示例中，只有当完整的核映射到输入图像上，才进行卷积运算，这种限制保证卷积运算的有效性。图中方框和箭头展示了输出张量中左上角元素是如何通过核张量和输入张量中对应的左上角区域卷积而来。在机器学习领域，学习算法会在适当的空间中学习核张量中的元素值，因此，一个基于翻转核的算法学习到的核张量相对于一个基于不翻转核算法学习到的核张量，是一种翻转对应关系。

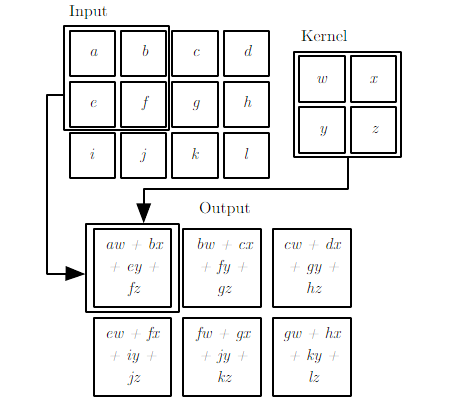


图2-2 在2D张量上应用卷积示例（不翻转核）

离散卷积可以看作是矩阵乘法，卷积运算在任何包含矩阵乘法并且不依赖于特定的矩阵结构神经网络算法中都适用，而不需要对神经网络算法做其他改变。为了更有效的应对复杂的输入，典型的卷积神经网络确实经过特定的处理，从理论角度严格来说，这些特殊化处理并不是必须的。

### 2.1.2 池化操作

卷积神经网络中典型的层包含三种类型（由图2-3）。在第一阶段中，网络层并行执行多个卷积运算，得出一系列线性响应；在第二阶段中，每个线性响应都作为输入，通过一个非线性激活函数，例如修正线性函数，这个阶段也被称为检测阶段。在最后一个阶段，池化函数被用来进一步修正结果。

在网络的特定层之后，通过统计汇总方法，池化函数[32]用来将特定节点的输出替换为其附近节点的统计值。例如，最大值池化返回方形领域内节点输出最大值作为输出，其他常用池化包括均值池化， 范数池化和基于距离中心节点距离的加权平均池化等。在所有情况中，池化使得特征表达保持对输入的微小变换（如平移，旋转等）的不变性，假设在输入上进行微小的变换，大部分经过池化操作的节点值不会改变，图2-3展示池化是如何工作的。在实际应用中，对局部变换的不变形是一种十分有用的特性，因为往往判定某种特征是否存在比了解这种特征是什么更重要。例如，当需要判定一幅图像中是否包含人脸，算法需要做到的不是知道眼睛的具体像素位置，而是需要判断是否是否在人脸左边和右边都有一只眼睛。在其他上下文环境中，获取特征的位置或许更重要，比如寻找由两条边交叉所定义的角，边的位置需要被精确的保持以用来测试两条边的交点。

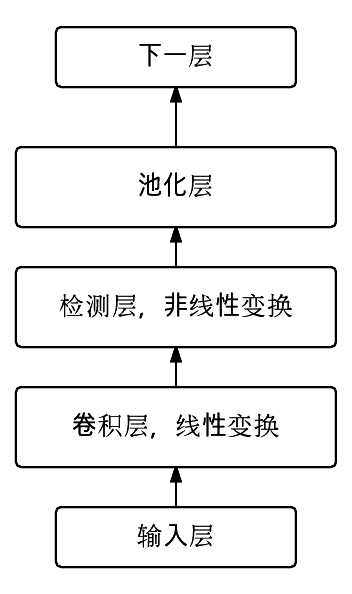
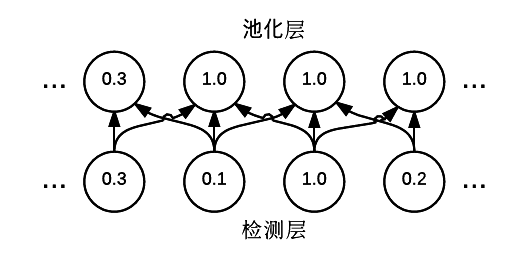
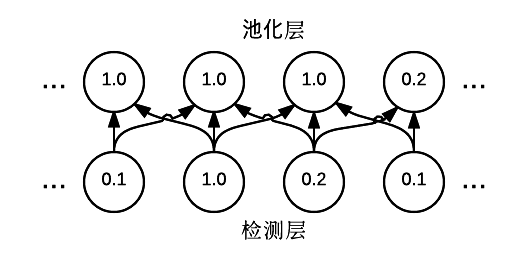


图2-3 典型卷积神经网络组件

池化可以被看作一种人为加入的先验知识，即网络层所学习到的映射必须保证对微小变换的不变性。如果这种假设成立，那么网络的统计效率将会被大大提高。



1. (b)

图2-4 最大值池化引入不变性

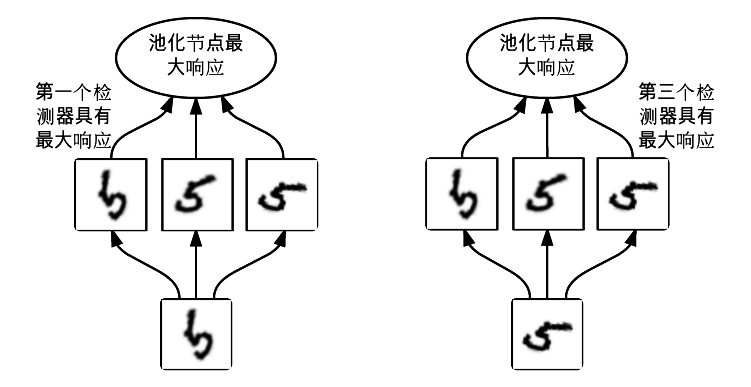
图2-4 描述最大值池化如何引入局部特征不变性。（a）中下面的行是检测层的非线性映射输出，上面的行展示最大池化的结果，池化步长为单个像素，池化区域为3邻域。（b）图描述和（a）图相同的网络，但输入层向右做了一个像素的平移变换，池化的输入数据全部变化，但池化结果只有一半发生变化，因为最大值池化节点只对输入邻域中最大值敏感，而不是确定位置的输入。

图2-5 学习不变性示例

一个池化单元处理多个相互独立学习到的特征，通过池化操作，可以使算法学习到对输入的不变性。图2-5 中展示了三个学习到的滤波器和一个最大值池化节点如何学习到不变性。所有滤波器都用来检测手写数字“5”，每个滤波器尝试去匹配有轻微不同方向的数字5，当数字5出现在输入中，对应的滤波器会匹配它并且引起检测节点高响应输出，然后最大值池化节点会得到一个大的激活响应，不管哪个检测节点具有最大值。图中展示了同一个卷积神经网络处理两个不同的输入，导致两个不同的检测节点具有最大响应，然而池化节点输出大致相似。这个原则在Maxout networks[33]和其他卷积神经网络中广泛使用，最大值池化在空间位置上对变换有固有的不变性。

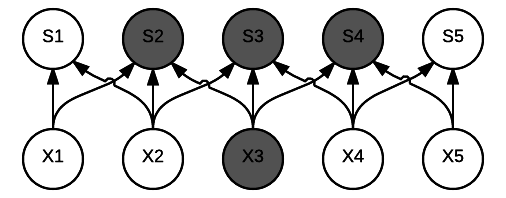
因为池化在整个邻域上汇总检测节点的响应，因此相对于检测节点，可以使用较少的池化节点。这样可以使得卷积神经网络更有效率，因为下一层将减少 倍的处理节点，如果下一层的参数个数是输入个数的函数（如全连接层），池化操作这种运算可以提升统计效率，并且减少参数的存储消耗。在处理不同大小输入的任务中，池化必不可少，例如需要对不同大小的图像进行分类，分类器的输入必须是固定大小，通过池化运算，不同大小的输入总可以通过设置不同步长而变到统一尺寸。

许多理论研究给出了针对不同环境应该采取的池化策略，如文献[34]，文献[35]通过在感兴趣的位置对特征使用聚类算法，动态的将特征池化在一起，这种方法在每幅图像上得到了一组池化区域。文献[36]试图在所有图像上学习一种通用的池化结构。

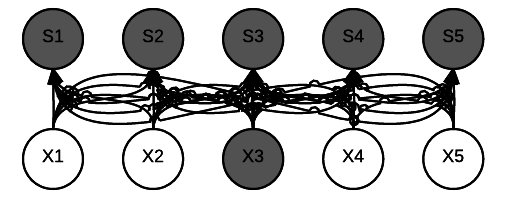
### 2.1.3 卷积神经网络设计机理

卷积运算可以从三个角度提升机器学习系统：稀疏的交互，参数共享和等变化的表达。此外，卷积运算为处理不同大小输入问题提供了一种新的方式。

传统神经网络结构中层与层之间使用矩阵乘法来描述输入单元和输出单元的交互作用，这意味着每个输出节点都与每个输入节点有连接，这种结构也称作全连接网络结构。然而，卷积网络具有稀疏的交互（也指稀疏连接或者稀疏权重），网络通过限制卷积核的大小，使其小于输入张量大小，来实现连接稀疏。例如，处理图片时，输入图片可能有成千上万个像素，但卷积网络可以利用卷积核检测很小的，有意义的特征（如边缘），这些特征只有数十个或数百个像素点。稀疏连接意味着只有少量的参数需要存储，这不仅降低了模型存储要求，并且提高了模型的统计效率；计算输出结果也不再需要更多的运算。相对于全连接网络，这些效率和性能方面的提升是巨大的，假设有 个输入节点和 个输出节点，矩阵相乘包含 个参数，处理每个样本需要的计算量是 。假如限制每个输出节点与 个输入节点相连，这种稀疏的连接方式只有 个参数，计算量降低到 。在众多实际应用中，使用比 小多个数量级的参数 也能获得足够好的结果，从图2-6和图2-7中见稀疏连接的图解。在一个深度卷积神经网络中，深层的神经元可能间接的与大部分输入节点交互（由图2-8），通过在简单的结构模块间构建这种连接，使得网络可以更有效的描述不同变量之前的复杂关系。



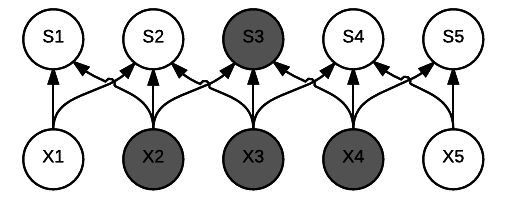
（a）



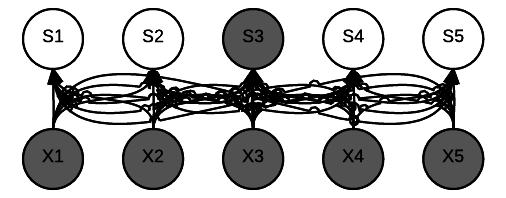
（b）

图2-6 稀疏连接（输入视角）

在图2-6中，输入节点 高亮显示，输出节点中与有连接的也高亮显示。子图（a）中， 集合由输入节点经过宽度为3的卷积核卷积得到，只有三个节点受 影响，子图（b）中， 由矩阵乘积得到，连接是稠密的，因此所有输出节点都受 影响。同理，图2-7展示了针对特定输出节点 的连接情况，稀疏连接时，只有局部的接受域影响 ，而全连接时，所有输入节点都会影响 。



（a）



（b）

图2-7 稀疏连接（输出视角）

图2-8 说明了卷积网络中深层次节点的接受域大于浅层节点的接受域。如果网络中包含大步长的卷积或者池化操作，或者网络深度较大，接受域的大小差距越大。因此，即使卷积网络中直接连接是稀疏的，深层节点仍然会与输入图像中的大部分或全部节点相连。

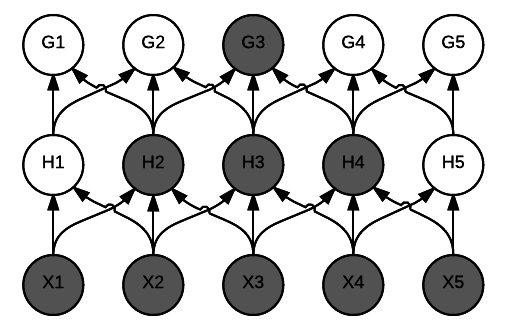


图2-8 网络深层节点的接受域

参数共享（权值共享）指在同一个模型的不同的函数中使用同样的参数。在传统的神经网络的某一层中，权值矩阵中的每个元素只使用一次，它仅仅用来与输入图像中的一个元素相乘，然后不再被访问。网络中的参数共享也可以理解为网络中有固定的权值，因为应用到某个输入上的权值和应用到其他输入上的权值是相同的，在卷积神经网络中，卷积核中的每个元素都被应用到输入图像上的每个位置（根据设计规则，边界上的像素可能除外）。在卷积运算中使用参数共享指：不同于在图像的每个位置学习不同的核参数，参数共享方式只学习一组核参数。运行时，参数共享在前向传播时不会影响计算量（ ），但将模型需要存储的参数降低到 个，考虑到 比 小数个数量级，并且 和 往往具有同样大小， 相对于 几乎可以忽略不计，因此卷积运算相对于矩阵乘法极大的提升了模型学习的效率。图2-9 描述了参数共享的工作原理。

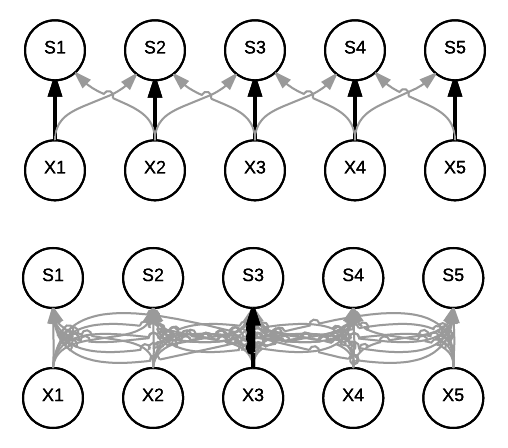


图2-9 参数共享

图2-9中，黑色箭头表示这个连接使用特定参数，分别在卷积模型和全连接模型中。图上部分的黑色箭头表示在卷积模型中应用三元素核中的中间元素，由于参数共享，这个参数在输入图像的所以位置都会使用；图下部分的黑色箭头表示在全连接模型中使用权值矩阵中的中间一个元素，由于没有参数共享，这个参数有且只有一次使用机会。

在卷积网络中，这种特定形式的参数共享使得网络层具有映射的同变性。一个函数具有同变性指函数的输出和输入具有相同的变化方式，特别的，若 ，则函数 对函数 等变化。在卷积网络中，假设 是对输入的任意函数映射，则卷积函数对 等变化，设 是给出图像中整数坐标处的亮度值函数， 是将一个图像函数变换到图像函数的映射，使得 满足 变换 将 中每个像素向右移动一个单位。将 平移变换后，再进行卷积，结果和我们先对 进行卷积，再用 做平移变换得到的结果一样。对于时间序列数据，卷积得出一种类似时间轴的结构，说明不同特征在输入出现的时间，如果将输入中某一事件后移，事件的表达在输出中也同样向后移动。在图像数据中，当特定的特征出现在输入中，卷积得出一个2D特征图，如果将输入中对应对象移动到其他位置，对象的表达则会在特征图中移动相应距离，这表明同一个局部函数在输入中的任何位置都有用。例如图像处理中，同样的边缘或多或少会出现在图像的任何位置，因此在整幅图像中共享边缘检测器参数是有可行的。但卷积对尺度变换，旋转等变换并不具有等变化性，这些变换需要其他机理来处理。

## 2.2 深度生成模型

### 2.2.1 玻尔兹曼机

玻尔兹曼机最早出现在文献[37]中，作为一种通用的方法来学习任意基于二元向量的概率分布。玻尔兹曼机形成了大量的流行的变种算法，实际上，这些变种算法的普及已经远远超过原始算法和其典型模型。

在 维二元随机向量 上定义玻尔兹曼机，玻尔兹曼机是一个基于能力的模型，即使用能量函数（由公式（2-7））来定义模型参数符合的联合概率分布。



 表示能量函数， 是配分函数，使得 。玻尔兹曼机的能量函数由公式（2-8）给出：



其中 是模型参数中的权值矩阵， 是每个 的偏置参数。

在玻尔兹曼机的通常设置中，有一个观察集，每个观察值都是 维向量。公式（2-7）描述了观察值的联合概率分布，当不是所有变量都是可观察到时，玻尔兹曼机变得更有用，在这种情况下，不可观察到的变量（潜变量）可以被看作是多层感知机中的隐含节点，从而被用来对可见节点进行高阶交互关系的建模。类似通过增加隐含节点来将逻辑回归转化为多层感知机，使得多层感知机成为通用的函数近似器，拥有隐含节点的玻尔兹曼机不再局限于对变量之间线性关系进行建模，而是成为概率密度函数的通用近似器[38]

形式上，所有节点被被分为两个子集：可见节点集 和隐含（或潜在）节点集 ，不失一般性，通过分解为子集 和 将公式（2-8）转化为公式（2-9）：



作为概率模型，玻尔兹曼机一般使用最大似然估计来学习范式。根据最大似然估计原理，算法尝试选取能够使得可见节点概率达到最大时的参数。给定一个包含 个样本的数据集 ，模型学习的目标在满足玻尔兹曼机定义时最大化似然估计，假设样本满足独立同分布条件，即最大化下式：



玻尔兹曼机并不能精确的将可见节点的分布参数化为 ，实际上，玻尔兹曼机是通过一个定义在可见节点 和隐含节点 共同体上的能量函数来参数化数据分布的。为了求得，需要将 的影响边缘化掉：



公式（2-10）和公式（2-11）给出了算法需要最大化的目标函数，然而，由于 是模型参数的函数，所以无法得到最大似然估计问题的解析解，即使使用大多数深度学习使用的迭代法也不能直接使用，因为目标函数对特定参数的偏导数没有解析公式。

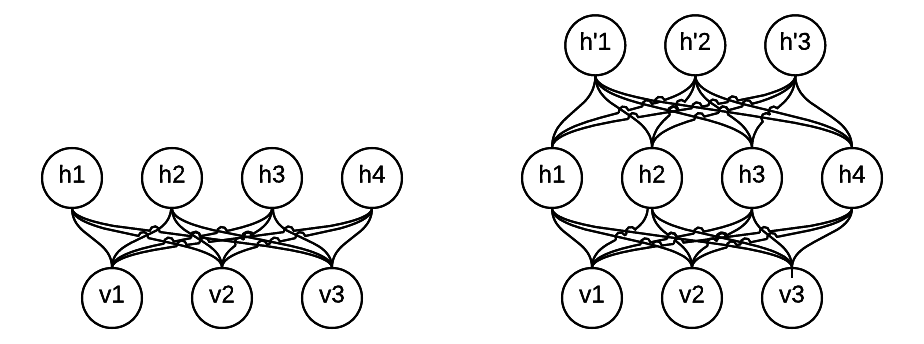
### 2.2.2 受限玻尔兹曼机（RBMs）

在文献[39]首次提出，受限玻尔兹曼机是深度概率模型中最常用的构建模块。RBMs是无向概率图模型，包含一层可观察变量和一层隐含变量，在同层之间没有连接（由图2-10），通过堆叠RBM可以搭建深度概率模型。一般的RBM限制输入 是二元的，下一节会介绍连续类型的RBM。

假设可观察层由 个二值随机变量组成，用向量 表示，隐含层的二值随机变量组成的向量用 表示。和一般玻尔兹曼机一样，受限玻尔兹曼机也是基于能量的模型：



能量函数为： ，配分函数为： 。



1. （b）

图2-10 RBMs和由RBMs构建的模型

从定义可以发现，用直接法（对所有状态进行加和）计算配分函数 的计算复杂度是无法接受的，除非有一种新的算法可以发掘概率分布中的规律来使得 的计算更快。Long and Philip [40]2010年正式提出配分函数 是难解的，这也表示包含规则化项 的联合概率分布 也是难解的。

尽管 是难解的，但双边图结构的RBM具有特别的性质：条件概率分布 和 是可分解的，相对容易计算和采样的。从联合概率分布中得到条件概率分布：



公式（2-13）中条件概率以可见节点 作为条件，而 可以被看作是常数， 可因子分解的性质，结合节点二值化假设及独立性假设，可以得出单个节点的条件概率：



其中 为Logistic函数，结合公式（2-13）和公式（2-14）：



RBM可因子分解的的性质使得采用吉布斯方法从联合分布中高效采样变得可能，同时吉布斯采样过程中可以使用批处理的方式，同时对多个变量进行采样。在RBM中，吉布斯采样分为两步：1）对 采样，得益于RBM分解性质，给定 时，可以同时对 中所有元素进行独立采样；2）对 采样，同理，对 的采样也可以是同时独立进行的。

### 2.2.3 实值受限玻尔兹曼机

最初的玻尔兹曼机被设计用来处理二值数据，但是应用于图像和语音等数据的模型需要具有表达实数上数据概率分布的能量。在文献[41]中，RBM被用来对手写数字进行建模，将训练数据集中的灰度图像规划为 之间的概率值，，但考虑到二值图像独立采样过程会出现噪声，这种设计并不是理论上最适合的方案。

高斯-伯努利RBMs（gbRBM）[42]由二值的隐含节点和实值的可视节点组成，用高斯分布来拟合可视节点的分布，可以处理如图像等实值数据。但由于自然图像中很多内容信息更多的存在于像素之间关系，而不是单个像素中，即是：像素间的关系，而不是像素大小的绝对值才是图像中有用信息的载体。为了解决这个问题，其他替代模型被提出，如均值-方差RBM（mcRBM）[43]，均值乘积 分布模型（mPoT）[44]等。以mcRBM为例，mcRBM模型由两个能量函数定义，均值能量函数和方差能量函数：



其中 表示 个均值隐含节点组成的向量， 表示 个方差隐含节点组成的向量。 和 分别对均值信息和方差信息进行建模。进一步推到可得出：



由于其非对数结构的条件概率结构，通过对比分歧的方法来训练mcRBM模型是十分复杂的，从 中采样需要每次迭代都消耗大量的计算，当有大量可见节点时，可能会成为算法的瓶颈。Ranzato和Hinton[44]通过直接从边缘分布 中使用混合的蒙特卡罗方法[45]采样，从而避免从 中采样需要的大计算量。

玻尔兹曼机可以根据训练准则的不同演化出很多变种，常见的玻尔兹曼机都集中于尝试用最大化生成概率 来训练模型，文献[46]尝试用最大化 来学习玻尔兹曼机。在众多玻尔兹曼机模型中，节点间最多只有二阶交互会体现在能量函数中，高阶玻尔兹曼机会中的隐含节点除了对可视节点进行建模，还可能同时对可视节点和对象标签进行建模。一般来说，玻尔兹曼机框架包含丰富的模型，目前设计出的模型只是一小部分，设计一种新的玻尔兹曼机模型相比设计一种新的网络层更加需要创造力，因为找到一种具有良好计算性质的条件概率模型十分困难。

## 2.3 小结

在本章中，我们介绍了深度学习模型中两类常用的模型框架。深度卷积神经网络是一种监督的前馈式神经网络，卷积网络所具有的权值共享，稀疏连接和等变换的表达等性质使得其相比于传统的全连接网络更易于训练，并且具有更好的泛化性能，当人工神经网络发展限于瓶颈时期，卷积神经网络在实践中取得了很多成就，现在已经被计算机视觉团队广泛应用。在本章的前半部分，通过对卷积神经网络中主要的卷积层和池化层进行介绍，阐述了卷积神经网络的工作方式，和其不同于传统全连接网络的性质。卷积神经网络在计算机视觉方面的应用，展示了其强大的特征表示能力和泛化能力，本文设计的深度哈希特征正是建立在卷积网络的这两条性质上。

本章后半部分介绍了深度生成模型，玻尔兹曼机作为深度深生成模型的基础模块，在文章中有细致的介绍。生成模型是一种非监督的学习模型，这种方法可以通过隐含网络层来提取特征而不使用带标签的数据，隐含层因此被用来重构数据或者对特征检测器的活动进行建模，通过预训练，深度网络的权值可以被初始化为接近最优解的一个局部值，再辅以反向传播算法进行微调，可以在很多应用中取得较好的结果。

接下来的章节将详细介绍深度哈希特征的学习过程及这种哈希特征在图像检索和图像深度估计问题中的应用。

# 3 深度哈希特征学习

## 3.1 深度特征学习

视觉特征是目前最好的分类系统和检测系统的特征选择，事实上，过去十年中图像理解领域的大多数进步都得益于特征表示的发展，从视觉词袋模型（BoVW）[47, 48]到Fisher向量[49]。这些模型最大的特点是人工设计，并且相对简单，通过采样获得局部图像块，用视觉描述子（如SIFT）来将其编码为高维的特征表示。近年来，这类手工设计特征被深度卷积网络特征[5]所替代，CNNs具有更复杂的网络结构，由多层的非线性特征提取器组合而成（由图2-1）。尽管网络结构由需要手工设计，但是网络包含大量的需要学习的参数。在大规模的图像分类和检测任务中，CNNs取得了完美的表现，显著优于标准的图像编码[50]。

在文献[5]中提出的网络（AlexNet）基础上，文献[51]设计了多种卷积神经网络，与BoVW和Fisher向量等传统方法进行了全面的对比。从文献[52, 53]可以发现，在大规模数据集（如ImageNet）上CNN的倒数第二层的激活值向量 ，可以当做一种有效的图像描述子，并且应用到其他数据集。更进一步，用一个训练好的CNN模型在目标数据集上调优可以显著的提高效果，因为经过调优以后，图像特征在一个通用的模式下变为数据特定的模式。本文利用在AlexNet基础改进的卷积网络模型VGG\_S作为基本模型，设计深度哈希网络，表3-1描述了VGG\_S的网络配置，相比于AlexNet, VGG\_S具有同样的网络深度，但具有更宽的网络结构。使用10个样本裁剪测试，VGG\_S在ILSVRC-2012（大规模视觉识别挑战赛）上的Top 5分类错误率为13.1%， 使用1个中心样本裁剪，Top 5分类错误率为15.4%。

表3-1 本文使用的改进CNN配置

|  |  |
| --- | --- |
| Layer | Configuration |
| Conv1 | 96\*7\*7, st.2, pad 0, LRN, \*3 pool |
| Conv2 | 256\*5\*5, st.1, pad 1, \*2 pool |
| Conv3 | 512\*3\*3, st.1, pad 1, - |
| Conv4 | 512\*3\*3, st.1, pad 1, - |
| Conv5 | 512\*3\*3, st.1, pad 1, \*3 pool |
| Full6 | 4096, dropout |
| Full7 | 4096, dropout |
| Full8 | 1000, softmax |

## 3.2 哈希学习

由于从原数据特征空间中，经过哈希映射得到二元哈希编码是NP难问题，因此常用哈希学习方法都采样两步学习策略：

（1）使用常用的特征提取算法提取样本的特征表示，并且采用降维算法将特征向量压缩到一个低维的实数特征表示向量。

（2）采用量化算法，对第一步得到的低维实数向量进行离散化，得到一个低维的二进制编码

哈希学习第一步中使用的特征提取算法主要由常用的SIFT算法，BoVF和Fisher向量等方法，降维可以采用度量学习算法实现。一般来说，一个好的哈希编码满足以下三个条件：（1）将相似的原始空间中的数据对映射到哈希空间中后具有相似的二元码字，（2）用较小数量的比特位来编码每个数据样本，（3）对输入数据映射的计算仅需要较少的计算量。

对一个给定数据集 ，设 为 的二进制编码。假设二进制编码中不同位的编码是独立的，即是

## 3.3 深度哈希特征学习

## 3.4 小结

# 4 深度哈希特征在图像检索中的应用

## 4.1 图像检索中常用特征提取算法

### 4.1.1 感知哈希算法

### 4.1.2 颜色直方图算法

### 4.1.3 内容特征算法

## 4.2 深度哈希特征图像检索

## 4.3 实验结果

## 4.4 小结

# 5 深度哈希特征在水下图像深度估计中的应用

## 5.1 水下光学成像模型

深度估计在

### 5.1.1 水下图像深度估计模型

与之前的方法不同，我们将深度估计问题转化为典型的回归问题，不依赖于任何几何先验知识或额外信息。

深度估计的目标是推断图像中每个像素点的深度信息，与先前方法类似，我们假设一副图像是由一系列同质的小区域（超像素）组成，并且以超像素作为节点构建图模型，每个超像素由区域中心点近似。 表示彩色图像， 表示 个超像素点的深度值。与传统CRF类似，我们用条件概率来对像素深度值的概率分布进行建模：



 为能量函数， 为

### 5.1.2

## 5.2 水下图像深度估计哈希模型

## 5.3 实验结果

## 5.4 小结

本章中，

# 6 总结与展望

随着计算机计算能力。

本文首先介绍了当前一些常见的

水下手持操作提供了可能，因此具有较大的应用前景。

本文还存在一些不足和可以进行进一步的研究的地方，主要可以总结为以下几点：

1）本文

2）本文

3）

4）

# 参考文献

1. Hinton, G.E. and R.R. Salakhutdinov, *Reducing the dimensionality of data with neural networks.* Science, 2006. **313**(5786): p. 504-507.

2. LeCun, Y., et al., *Gradient-based learning applied to document recognition.* Proceedings of the IEEE, 1998. **86**(11): p. 2278-2324.

3. Huang, F. and Y. LeCun. *Large-scale learning with svm and convolutional netw for generic object recognition*. in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2006.

4. Waibel, A., et al., *Phoneme recognition using time-delay neural networks.* Acoustics, Speech and Signal Processing, IEEE Transactions on, 1989. **37**(3): p. 328-339.

5. Krizhevsky, A., I. Sutskever, and G.E. Hinton. *Imagenet classification with deep convolutional neural networks*. in *Advances in neural information processing systems*. 2012.

6. Sainath, T.N., et al. *Deep convolutional neural networks for LVCSR*. in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. 2013. IEEE.

7. Goodfellow, I.J., et al., *Multi-digit number recognition from street view imagery using deep convolutional neural networks.* arXiv preprint arXiv:1312.6082, 2013.

8. Collobert, R. and J. Weston. *A unified architecture for natural language processing: Deep neural networks with multitask learning*. in *Proceedings of the 25th international conference on Machine learning*. 2008. ACM.

9. Karpathy, A. and L. Fei-Fei. *Deep visual-semantic alignments for generating image descriptions*. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015.

10. Jaitly, N. and G.E. Hinton. *Vocal tract length perturbation (VTLP) improves speech recognition*. in *Proc. ICML Workshop on Deep Learning for Audio, Speech and Language*. 2013.

11. Chang, N.-S. and K.-S. Fu, *Query-by-pictorial example.* IEEE Transactions on Software Engineering, 1980. **6**(6): p. 519.

12. Chang, N.-S. and K.S. Fu, *A relational database system for images*. 1980: Springer.

13. Chang, S.-K., et al., *An intelligent image database system.* Software Engineering, IEEE Transactions on, 1988. **14**(5): p. 681-688.

14. Gudivada, V.N. and V.V. Raghavan, *Content based image retrieval systems.* Computer, 1995. **28**(9): p. 18-22.

15. Rui, Y., T.S. Huang, and S. Mehrotra. *Content-based image retrieval with relevance feedback in MARS*. in *Image Processing, 1997. Proceedings., International Conference on*. 1997. IEEE.

16. Müller, H., et al., *A review of content-based image retrieval systems in medical applications—clinical benefits and future directions.* International journal of medical informatics, 2004. **73**(1): p. 1-23.

17. Datar, M., et al. *Locality-sensitive hashing scheme based on p-stable distributions*. in *Proceedings of the twentieth annual symposium on Computational geometry*. 2004. ACM.

18. Kulis, B. and K. Grauman. *Kernelized locality-sensitive hashing for scalable image search*. in *Computer Vision, 2009 IEEE 12th International Conference on*. 2009. IEEE.

19. Kulis, B., P. Jain, and K. Grauman, *Fast similarity search for learned metrics.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2009. **31**(12): p. 2143-2157.

20. Salakhutdinov, R. and G. Hinton, *Semantic hashing.* International Journal of Approximate Reasoning, 2009. **50**(7): p. 969-978.

21. Liu, W., et al. *Hashing with graphs*. in *Proceedings of the 28th international conference on machine learning (ICML-11)*. 2011.

22. Weiss, Y., A. Torralba, and R. Fergus. *Spectral hashing*. in *Advances in neural information processing systems*. 2009.

23. Kulis, B. and T. Darrell. *Learning to hash with binary reconstructive embeddings*. in *Advances in neural information processing systems*. 2010.

24. Gong, Y. and S. Lazebnik. *Iterative quantization: A procrustean approach to learning binary codes*. in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. 2011. IEEE.

25. Gupta, A., et al. *Estimating spatial layout of rooms using volumetric reasoning about objects and surfaces*. in *Advances in neural information processing systems*. 2010.

26. Hedau, V., D. Hoiem, and D. Forsyth, *Thinking inside the box: Using appearance models and context based on room geometry*, in *Computer Vision–ECCV 2010*. 2010, Springer. p. 224-237.

27. Gupta, A., A.A. Efros, and M. Hebert, *Blocks world revisited: Image understanding using qualitative geometry and mechanics*, in *Computer Vision–ECCV 2010*. 2010, Springer. p. 482-496.

28. Karsch, K., C. Liu, and S.B. Kang, *Depth transfer: Depth extraction from video using non-parametric sampling.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2014. **36**(11): p. 2144-2158.

29. Ladicky, L., J. Shi, and M. Pollefeys. *Pulling things out of perspective*. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014.

30. Bruno, F., et al., *Experimentation of structured light and stereo vision for underwater 3D reconstruction.* ISPRS Journal of Photogrammetry and Remote Sensing, 2011. **66**(4): p. 508-518.

31. Narasimhan, S.G., et al. *Structured light in scattering media*. in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*. 2005. IEEE.

32. Zhou, Y.-T., et al., *Image restoration using a neural network.* Acoustics, Speech and Signal Processing, IEEE Transactions on, 1988. **36**(7): p. 1141-1151.

33. Goodfellow, I.J., et al., *Maxout networks.* arXiv preprint arXiv:1302.4389, 2013.

34. Boureau, Y.-L., J. Ponce, and Y. LeCun. *A theoretical analysis of feature pooling in visual recognition*. in *Proceedings of the 27th international conference on machine learning (ICML-10)*. 2010.

35. Boureau, Y.-L., et al. *Ask the locals: multi-way local pooling for image recognition*. in *Computer Vision (ICCV), 2011 IEEE International Conference on*. 2011. IEEE.

36. Jia, Y., C. Huang, and T. Darrell. *Beyond spatial pyramids: Receptive field learning for pooled image features*. in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. 2012. IEEE.

37. Ackley, D.H., G.E. Hinton, and T.J. Sejnowski, *A learning algorithm for Boltzmann machines.* Cognitive science, 1985. **9**(1): p. 147-169.

38. Le Roux, N. and Y. Bengio, *Representational power of restricted Boltzmann machines and deep belief networks.* Neural computation, 2008. **20**(6): p. 1631-1649.

39. Smolensky, P., *Information processing in dynamical systems: Foundations of harmony theory*. 1986, DTIC Document.

40. Long, P.M. and R. Servedio. *Restricted Boltzmann machines are hard to approximately evaluate or simulate*. in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*. 2010.

41. Hinton, G.E., *Training products of experts by minimizing contrastive divergence.* Neural computation, 2002. **14**(8): p. 1771-1800.

42. Welling, M., M. Rosen-Zvi, and G.E. Hinton. *Exponential family harmoniums with an application to information retrieval*. in *Advances in neural information processing systems*. 2004.

43. Dahl, G., A.-r. Mohamed, and G.E. Hinton. *Phone recognition with the mean-covariance restricted Boltzmann machine*. in *Advances in neural information processing systems*. 2010.

44. Ranzato, M.A. and G.E. Hinton. *Modeling pixel means and covariances using factorized third-order Boltzmann machines*. in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. 2010. IEEE.

45. Neal, R.M., *Probabilistic inference using Markov chain Monte Carlo methods.* 1993.

46. Larochelle, H. and Y. Bengio. *Classification using discriminative restricted Boltzmann machines*. in *Proceedings of the 25th international conference on Machine learning*. 2008. ACM.

47. Csurka, G., et al. *Visual categorization with bags of keypoints*. in *Workshop on statistical learning in computer vision, ECCV*. 2004. Prague.

48. Sivic, J. and A. Zisserman. *Video Google: A text retrieval approach to object matching in videos*. in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*. 2003. IEEE.

49. Perronnin, F., J. Sánchez, and T. Mensink, *Improving the fisher kernel for large-scale image classification*, in *Computer Vision–ECCV 2010*. 2010, Springer. p. 143-156.

50. Chatfield, K., et al. *The devil is in the details: an evaluation of recent feature encoding methods*. in *BMVC*. 2011.

51. Chatfield, K., et al., *Return of the devil in the details: Delving deep into convolutional nets.* arXiv preprint arXiv:1405.3531, 2014.

52. Donahue, J., et al., *Decaf: A deep convolutional activation feature for generic visual recognition.* arXiv preprint arXiv:1310.1531, 2013.

53. Zeiler, M.D. and R. Fergus, *Visualizing and understanding convolutional networks*, in *Computer vision–ECCV 2014*. 2014, Springer. p. 818-833.

# 致谢

时光飞逝，转眼间两年的硕士研究生生活即将结束，在此谨向在攻读研究生期间对我的学习、生活给予帮助和关心的所有老师和同学表示深深的谢意。

在此，我要衷心的感谢我的导师董军宇教授，董老师在我的研究生求学期间，在生活、学习以及工作等方面都给予了很大的帮助、支持和鼓励。还要特别感谢亓琳老师，从毕业论文选题开始，我在背景知识学习、代码编写及论文修改各个阶段都遇到了很多问题，是他的悉心指导和严格要求，让我顺利完成了毕业设计。两位老师都是我两年求学过程中的良师益友。

同时也要感谢实验室已毕业的同学前期所做的工作对本文研究的启发，是他们留下的资料让我以最快的速度了解了研究相关的背景知识；感谢吴则举博士制作了实验器材，让我的理论研究能够在实验中验证结果。

感谢实验室各位伙伴，是他们的热情帮助，让我能够顺利的进行科研，他们在生活上对我的关心和爱护，让我的研究生求学期间充满了欢乐。

最后要感谢的是我的父母，感谢他们多年来对我的养育和培养。

# 个人简历

1990年03月30日出生于四川省遂宁市。

2009年9月考入中国海洋大学数学科学学院信息与计算科学专业，2013年6月本科毕业并获得理学学士学位。

2013年9月保送入中国海洋大学信息科学与工程学院计算机应用技术专业攻读硕士学位至今。