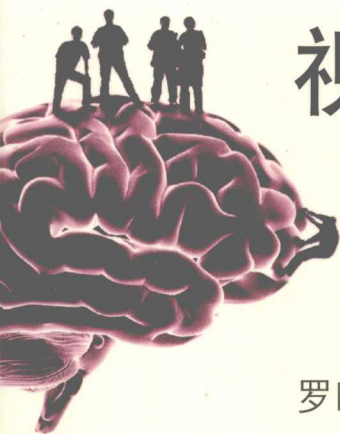


国家高技术研究发展计划 (863计划) 课题
国家自然科学基金资助项目 成果
北京市自然科学基金资助项目



视觉信息 认知计算理论

The Perception Computing of
Visual Information

罗四维 等编著



科学出版社

www.sciencep.com

视觉信息认知计算理论

罗四维等 编著

科学出版社

北京

内 容 简 介

在众多的生物系统中,人脑被认为是最高级的生物智能系统,它具有感知、识别、学习、联想、记忆、推理等功能。而在人脑感知的信息中,大部分来自视觉。视觉是人类获取信息的重要途径,也是人类对自身研究认识最深刻的部分。因此,研究生物体的视知觉功能,解析其内在机理,并用机器来实现,成为科学研究领域的一个重要方面,它可以为提高机器的智能与解决问题的能力提供新的思路。

本书系统地讨论了基于视觉感知和有效编码假说的特征表示、计算模型,从认知心理学出发讨论了半监督学习、聚类、知觉组织,从人类视觉的注意机理角度讨论了模拟视觉注意机制的视觉感知模型等。

本书可作为计算机科学领域人工智能、模式识别等专业的研究生教材,也可供相关专业的研究人员参考。

图书在版编目(CIP)数据

视觉信息认知计算理论/罗四维等编著. —北京:科学出版社,2010.7
ISBN 978-7-03-028134-0

I. ①视… II. ①罗… III. ①视觉-认知-计算技术 IV. ①B842.1
②TP391

中国版本图书馆 CIP 数据核字(2010)第 120282 号

责任编辑:张 濮 潘继敏 / 责任校对:鲁 素
责任印制:张克忠 / 封面设计:耕者设计工作室

科 学 出 版 社 出 版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

骏 杰 印 刷 厂 印 刷

科学出版社发行 各地新华书店经销

*

2010年7月第 一 版 开本:B5(720×1000)

2010年7月第一次印刷 印张:14 1/2 插页:1

印数:1—2 500 字数:290 000

定价:39.00元

(如有印装质量问题,我社负责调换)

前 言

近 20 年来,神经科学、解剖学伴随着科学和技术手段的进步得到了快速发展,使我们对人类大脑有了更深刻的了解。探索神经计算的数理基础,并将其应用于发展新的神经式信息处理模式的神经计算科学已经成为国内外研究的热点课题之一。视觉是人类获取信息的重要途径,也是人类对自身研究认识最深刻的部分,借助和参考生物视觉认知机理来发展和提高机器信息处理与认知计算能力是一种重要的手段。

从理论基础和应用的层面,国外许多学者致力于研究视觉感知和机器学习之间的关系,并取得了很多成果。这种从模拟人类的思维模式出发来指导机器学习的方法符合人类与环境之间不断学习、不断适应的演变过程。因此,结合视觉感知的理论来探讨计算机处理信息的能力具有重要意义。随着神经生理学、认知科学、神经计算等学科不断发展,这种将视觉感知和计算机信息处理结合起来的研究方法,一定会有广阔的应用前景。

目前基于视觉感知有五个热点问题:①基于感知机理的机器学习方法;②基于有效编码假说的初级特征表示;③视觉表象的中级特征表示;④初级视觉表象启发下的知觉组织;⑤注意机制。

本书力求理论与实践统一,从提高计算机对非结构化视听觉感知信息的理解能力和海量异构信息的处理效率,克服图像信息处理所面临的“瓶颈”出发,借助认知心理学、神经生理学、生物学、计算机科学和数理科学的交叉优势,描述在场景理解中基于视觉感知的热点技术问题以及新的计算模型、计算方法。在较为系统地介绍人类视觉感知系统工作机理的基础上,本书着重介绍神经计算领域目前正在发展的一些新理论和新技术,这也正是本书作者在多年来承担该领域国家研究项目,并结合为研究生开设相关课程的基础上编写的。

本书的内容组织如下:

第 1 章是概述,简要介绍基于视觉感知的五个热点问题。

第 2 章对初级视皮层的典型视神经细胞,比如简单细胞、复杂细胞等进行了较为深入的阐述,简要介绍了视神经信息处理机制的生理结构。与生物视觉感知的生理结构相对应,介绍了生物视觉感知的计算模型。最后详细介绍了一种生物视觉启发的特征表示方法,以及它在基于内容的图像检索中的应用。

第 3 章讨论了基于有效编码假说的低层特征表示。有效编码假说作为一种能合理解释复杂外部环境和有限的神经元个数之间关系的理论工具,得到了生物实

验的证明。本章在有效编码框架下,详细介绍了三种常用的衡量编码有效性的编码准则,即稀疏准则、独立准则和慢变准则。

人类处理信息并非如计算机按照大量数据位那样的高维处理方式。流形是一种有效的降维方法。为此,第4章讨论流形的数据处理方式。在这里按数据处理的方式、高维数据几何结构的分析方法以及全局几何与局部几何的表示、算法之间的联系等将流形学习的谱方法进行了分类。在基于谱图的流形学习算法中,分别重点分析、比较了局部保持和全局保持的两类流形学习算法,从算法框架、算法实例、算法特性对比分析表明它们的优点与缺陷。本章还介绍了流形学习的图嵌入(graph embedding)框架,该框架从一个统一的模型对PCA、MDS、ISOMAP、LLE、LE等算法进行合理的诠释。通过线性化、核化和张量化的扩展,将传统的流形学习算法统一到图嵌入框架下,为开拓新的流形学习算法提供了良好的理论基础和框架。

第5章从人类认知行为的心理学实验开始,介绍了半监督学习的有关知识和基本的半监督学习算法,并重点说明了半监督多视图学习。多视图学习针对数据有多个特征集的情况,通过在多个特征集上联合训练分类器来提高性能。而考虑未标记数据后的多视图学习,减少了对标记数据的需求量,在自然语言处理、图像检索、文本分类、生物信息学等许多领域都有广泛的应用。半监督多视图算法主要有协同训练算法、最大化一致算法和多视图特征映射算法。这些算法分别从不同角度寻找问题的解,并在理论分析上也取得了不少进展。参考人类的学习机制,有可能设计出更好的半监督多视图学习算法。如何科学地设计心理学实验来研究人类认知过程的半监督性质,以及如何借鉴人类认知学的实验结果来构建计算模型,都是值得探讨的内容。

第6章讨论了在计算机视觉感知研究中,很多基本的视觉任务以及对图像数据处理都需要用到的聚类。除了聚类分析本身所涉及的基本问题,如目标函数的定义、算法的设计以及数据点之间的距离度量等,计算机视觉感知中所涉及的聚类问题一般需要在基本的聚类方法上加入额外的视觉信息处理技术,使得处理结果更符合人类视觉感知过程所得到的输出。本章首先讨论了聚类分析中的一些主要概念和关键问题,并对常见的聚类算法作了简要概述,主要介绍了谱聚类、异质聚类以及AP(affinity propagation)聚类算法;其次,分别从图像分割、图像聚类这两方面介绍了聚类分析在计算机视觉感知研究中的应用。

第7章属于中级视觉的知觉组织,以轮廓编组为例重点介绍了显著结构的描述,以及局部视觉元素如何组织成完整的目标。在认知心理学的相关工作部分主要介绍了格式塔心理学对视觉完形、知觉分割等现象的解释,以及归纳总结出的可用于指导计算模型的规则和理论。在计算模型部分主要介绍了当前主流的两大类知觉组织方法:以图分割为代表的无监督方法和基于主动轮廓模型的方法。

第 8 章讨论选择性注意机制。人类在感受外界的信息时并不是对所有信息一视同仁的,而是表现出某种特异性。大脑只对那些认为是重要的部分输入信息进行处理,这种特异性处理策略被称为神经系统的注意(attention)机制。注意机制在生物视觉信息处理过程中是一个非常重要,但又很难描述清楚的问题。它是视觉感知模型的一部分,与学习、记忆等模块协同工作,完成将特定目标从背景中分离、在多个目标间注意焦点转移、可能发生形变或有遮挡以及残缺的注意目标与记忆中的模式匹配等任务。本章对注意机制进行了概述,根据生理学和心理学实验的结果,分别讨论了基于空间的注意和基于目标的注意理论及相关研究成果,详细讲解一种结合自顶向下注意和自底向上注意的视觉感知模型——what-where 模型。

参加本书编写的人员来自罗四维教授领导的研究小组。本书正式编写历时两年多,在研究小组充分讨论的基础上,全书由 7 人执笔。第 1 章由罗四维执笔,李清勇副教授执笔第 2 章,黄雅平副教授执笔第 3 章,赵嘉莉副教授执笔第 4 章,刘蕴辉博士执笔第 5 章和第 6 章,邹琪博士执笔第 7 章,田媚博士执笔第 8 章。

本书的数据资料来源于罗四维教授领导的研究小组,以及同一研究领域的他人研究论文。对于他人研究成果在本书中的引用,在此表示感谢。

本书是在国家高技术研究发展计划(863 计划)课题(2007AA01Z168)“基于视觉感知的智能计算模型及其关键技术研究”,国家自然科学基金资助项目(60773016)“视知觉组织的神经计算模型”,北京市自然科学基金资助项目(4092033)“人脑记忆机制启发下的视觉信息的表达与存储”的研究成果的基础上编写的,在此对科技部、国家自然科学基金委员会信息学部、北京市自然科学基金委员会表示衷心的感谢。

作 者

2010 年 3 月

目 录

前言

第 1 章 概述	1
1.1 基于感知机理的机器学习方法	1
1.2 基于有效编码假说的初级特征表示	3
1.2.1 有效编码假说	3
1.2.2 模拟人类方式的有效编码与特征表示	4
1.3 视觉表象的中级特征表示	6
1.4 初级视觉表象启发下的知觉组织	7
1.5 注意机制	8
1.6 智能计算模型在场景识别中的应用	9
参考文献	12
第 2 章 基于视觉感知的特征表示	16
2.1 视觉感知.....	16
2.1.1 外部环境的输入刺激	16
2.1.2 神经信息处理机制	18
2.1.3 视觉感知的输出	18
2.2 生物视觉感知的生理结构.....	18
2.2.1 外周脑	18
2.2.2 初级视皮层	20
2.2.3 纹外皮层.....	21
2.2.4 高级视皮层	22
2.3 生物视觉感知的计算模型.....	22
2.3.1 简单细胞响应模型	23
2.3.2 复杂细胞响应模型	25
2.3.3 高级皮层神经细胞响应模型	27
2.4 生物视觉启发的特征表示及其应用.....	28
2.4.1 独立纹元矩	29
2.4.2 独立纹元矩的图像检索实验	35
2.5 本章小结.....	39
参考文献	40

第 3 章 基于有效编码假说的低层特征表示	43
3.1 有效编码框架	43
3.2 基于稀疏性的有效编码方法——稀疏编码	44
3.2.1 Olshausen 的稀疏编码模型	44
3.2.2 基于稀疏编码的压缩传感	45
3.3 基于独立性的有效编码	57
3.3.1 独立分量分析	57
3.3.2 基于独立分量分析的视觉模型	63
3.3.3 Hyvärinen 研究小组的成果	65
3.4 基于慢变性的有效编码方法——慢变特征分析	68
3.4.1 慢变特征分析简介	68
3.4.2 慢变特征分析的实现	71
3.4.3 慢变特征分析与复杂细胞特性	73
3.4.4 慢变特征分析在手写体识别中的应用	75
3.5 本章小结	77
参考文献	77
第 4 章 流形学习	83
4.1 概述	83
4.2 局部保持流形学习算法分析	85
4.2.1 局部保持的流形学习算法的基本步骤	86
4.2.2 几种典型的局部保持的流形学习算法	86
4.2.3 局部保持的流形学习算法对比	91
4.2.4 全局线性化局部保持的流形学习算法	92
4.2.5 局部保持的流形学习算法实验比较	93
4.3 全局保持的流形学习算法分析	95
4.3.1 几种典型的全局保持流形学习算法	95
4.3.2 全局保持的流形学习算法对比	104
4.3.3 全局保持的流形学习算法的实验比较	105
4.4 图嵌入框架	108
4.4.1 图嵌入框架	108
4.4.2 图嵌入框架下的主成分分析	111
4.4.3 图嵌入框架下的判别分析	112
4.4.4 邻域判别分析	112
4.5 本章小结	114
参考文献	115

第 5 章 半监督学习	117
5.1 半监督学习和人类学习	117
5.2 半监督学习概况	118
5.2.1 半监督学习的概念	118
5.2.2 半监督学习的历史	119
5.2.3 未标记数据起作用的条件	120
5.3 半监督多视图学习算法	121
5.3.1 协同训练算法	121
5.3.2 最大化一致算法	123
5.3.3 多视图特征映射算法	124
5.4 半监督学习的应用	125
5.4.1 文本分类中的半监督学习	125
5.4.2 图像分割中的半监督学习	126
5.5 本章小结	127
参考文献	128
第 6 章 聚类	132
6.1 概述	132
6.2 经典聚类算法	133
6.2.1 划分法	133
6.2.2 层次法	134
6.3 聚类算法中的关键问题	135
6.3.1 距离度量问题	135
6.3.2 聚类数目	136
6.4 聚类分析与算法	137
6.4.1 谱聚类	137
6.4.2 异质聚类	139
6.4.3 基于消息传递的聚类算法	141
6.4.4 关于二元相似关系的假设	142
6.5 聚类分析在计算机视觉感知研究中的应用	143
6.5.1 图像分割	143
6.5.2 图像聚类	145
6.6 本章小结	147
参考文献	147
第 7 章 知觉组织	150
7.1 认知心理学的相关研究成果	151

7.1.1	格式塔知觉组织规则	151
7.1.2	视觉完形	152
7.1.3	非偶然性原则	154
7.2	无监督的知觉组织方法	155
7.2.1	编组线索的描述	155
7.2.2	图分割	157
7.2.3	张量投票	160
7.3	基于主动轮廓的知觉组织方法	165
7.3.1	气球模型	167
7.3.2	距离势能模型	168
7.3.3	GVF 模型和 GGVF 模型	169
7.3.4	T-Snake 模型	172
7.3.5	有形状先验的水平集方法	173
7.4	本章小结	177
	参考文献	178
第 8 章	模拟视觉注意机制的感知模型	181
8.1	注意机制概述	181
8.2	基于空间的注意机制	183
8.2.1	模拟自底向上视觉注意机制的感知模型	184
8.2.2	模拟自顶向下视觉注意机制的感知模型	190
8.3	基于目标的注意机制	196
8.3.1	实验依据	196
8.3.2	基于目标的选择性注意计算模型	198
8.4	基于 what 和 where 信息的视觉感知模型	201
8.4.1	模型框架	202
8.4.2	视觉通路理论	205
8.4.3	what 信息和 where 信息的提取与表示	205
8.4.4	基于一级 where 信息的预注意	209
8.4.5	一级 where 信息驱动的集中注意	212
8.4.6	what 信息与 where 信息的结合	214
8.4.7	实验结果与分析	216
8.5	本章小结	219
	参考文献	219

第 1 章 概 述

在众多的生物系统中人脑是最有效的生物智能系统,它具有感知、识别、学习、联想、记忆、推理等功能。据统计,人类感知的信息有 80% 来自视觉,为此,研究生物体的视知觉功能,解析其内在机理,并用机器来实现,成为科学研究领域的一个重要方面。研究视知觉过程有助于深入理解生物神经的工作机制和人类的认知规律,为模拟这些机制与规律、开发智能化信息处理模式开拓新的途径,为提高机器的智能、增强机器解决问题的能力提供新的思路。

本书将从提高计算机对非结构化视听觉感知信息的理解能力和海量异构信息的处理效率,克服图像信息处理所面临的困难出发,借助心理学、神经生理学、生物学、计算机科学和数理科学的交叉优势,描述在场景理解中基于视觉感知的热点技术问题以及新的计算模型、计算方法。

在第 1 章中,将简要介绍基于视觉感知的五个热点问题。

1.1 基于感知机理的机器学习方法

1. 视觉感知与流形学习

流形学习本身属于基础性的研究方向,但是由于其应用前景的不断扩展,近年来,流形学习日益成为一个热点问题,国内外很多研究人员和机构对其都给予了足够的重视,并积极投身于该研究领域。美国国家自然科学基金分别在 2003 年、2004 年、2005 年连续三年资助有关流形学习的项目。Schlumberger 基金从 1999 年开始连续资助 MIT 的研究项目 Manifold Learning for Nonlinear Inverse Problems。在 CVPR2004(美国华盛顿)和 IJCNN2005(加拿大蒙特利尔)会议上,都有关于流形学习的专题报告,如计算机视觉中的非线性流形、模式识别和图像分析中的非线性流形。很多国际会议也把 Manifold and Geometry in Signal Processing 作为会议的一个主题,如 ICASSP2004。在最近几年的国际会议(如 ICML、NIPS、ICCV 等)上,每年都有大量的相关文章出现。在 Stanford University、Massachusetts Institute of Technology、University of Pennsylvania、University of Toronto、Florida State University 等也都有专门的研究小组。

国内一些研究单位和大学也开展了流形学习方面的研究。例如,清华大学张长水等^[1]在局部线性嵌入算法(LLE)的基础上提出一个从低维嵌入空间向高维空

间的映射方法,并在多姿态人脸图像的重构实验中得到验证;中国科学院自动化所王珏教授、张军平博士^[2]对主曲线、主流形进行了深入研究,并应用于交通流模型中;中国科学院计算技术研究所的山世光博士在流形应用方面取得了很好的成果。

生物激励的神经计算科学最近几年已经得到科学界的广泛关注,如何模拟生物感知能力是一个非常活跃的研究领域。研究发现,除了视觉记忆可能以稳定状态的流形方式存储之外,在听觉神经科学人耳对声音源的定位研究中,来自不同方向的信号频谱在人脑中也可能以流形方式存储。流形学习方法的研究提供了一个新的研究思路,在减少感知输入的维度,保留重要的感知相似和差异的研究中已经取得了重要成果,进一步的探讨有助于研究神经流形和感知流形之间可能存在的某种联系。

流形学习的主要目标是发现嵌入在高维数据空间中观测数据的低维光滑流形,这已经成为机器学习、模式识别、数据挖掘等领域的研究热点之一,形成了大量的基于流形学习的研究方法。流形学习在近年的快速发展与流形可能和人类认知相关联有一定关系。Seung 从神经心理学的角度提出“感知以流形方式存在,视觉记忆也可能是以稳态的流形(或连续吸引子)存储”,为流形提供了与人类认知相关联的理由。近年来,非线性流形学习方法在机器学习领域得到了广泛重视与发展,相继发表的等度规映射算法(ISOMAP)和局部线性嵌入算法(LLE),从算法实现和验证的角度部分支持了这一观点。

维数压缩(dimensionality reduction)或称为降维、维数约简,其目的是从高维数据集构建表征其内在结构的低维表示。Nayar 在对机器人视觉的研究中发现了图像具有内在维数这一有意义的现象。在多数实际应用中,例如图像分析、计算机视觉、生物特征认证、文本挖掘和计算生物学等,获得的数据是高维的。这种高维问题一方面导致了维数灾难问题的出现,另一方面由于这种高维数据不能被人的感知直接理解,所以人们不易发现数据集内在的规律。因此,有必要对高维数据集通过维数压缩得到降维,维数压缩成为当前模式识别和机器学习研究领域中的一个重要研究方向。当然,维数压缩还具备其他一些优点,如压缩数据以节省存储空间、去除噪声、提取特征等。

由于流形学习可以发现嵌入在高维数据空间中观测数据的低维光滑流形,因此,流形学习自然成为生物视觉启发下的维数压缩方法之一,被人们所关注。

2. 人类认知过程与半监督学习

利用大规模的、标注过的训练数据固然可以提高学习算法结果的准确度,但是绝大多数有监督的机器学习方法依赖于标注的训练样本集,忽略了未标注样本的作用,而有效地利用未标注样本无疑将在一定程度上提高学习算法的性能。

Blum 等^[3]提出了 co-training 这一概念,并成功应用到网页文本分类问题上;

南京大学周志华等^[4]在此基础上提出使用三个分类器的 tri-training 算法; Nigam 等^[5]在学习中使用 EM 算法和朴素贝叶斯 (Bayes) 分类器, 并应用于文本分类; Cozman 等^[6]研究了混合模型中半监督学习的偏差方差渐进特性。

近年来, 人们从心理学的角度对机器学习进行了深入研究。一些研究成果表明, 人类认知过程具有半监督性, 机器的半监督学习模型与人类学习过程有相似之处。通过研究半监督学习模型, 或许能对人类认知过程的研究提供启示; 反之, 通过借鉴人类学习行为, 可产生更好的半监督学习技术。

1.2 基于有效编码假说的初级特征表示

人类的神经系统, 无论进化过程还是发展过程, 都能够自适应外界环境, 根据所感受到的信号特性来调节自己的行为。然而, 输入信号是千差万别的, 感知系统对于那些最常出现的信号能够最有效地进行处理。这就是说, 感知系统的信息处理过程与外界信号的统计特性密切相关。这个概念是很多领域的基本原理, 如信源编码、统计和决策理论等, 都需要确定外界信号的先验统计模型。

1.2.1 有效编码假说

尽管很多研究人员认为感知系统的信息处理过程受到外界信号统计模型的影响, 但是很难找到确切的证据。Attneave^[7]提出视觉感知系统的作用是产生输入信号的有效表示。Barlow^[8]提出了有效编码假说, 它从 Shannon 的信息论出发, 认为在神经计算的过程中, 一个重要的约束就是信息(或者编码)的有效性。初级感知系统的作用就是去除输入信号的统计冗余。也就是说, 人脑之所以能够对外界环境自适应, 就是因为复杂的外界刺激存在着冗余, 而人脑的神经元能有效地去除这些冗余, 从而可以利用较少的资源尽可能有效地表达更多的信息。有效编码假说合理地解决了复杂的外部环境和有限的神经元个数之间的矛盾, 并且通过对昆虫和脊椎动物视网膜的一系列实验验证了有效编码假说的正确性^[9-10], 因而它成为一种有效理解人类神经系统的理论工具。

有效编码假说提出以后, 很多研究人员都根据它的思想提出了不同的理论, 使其得到不断发展。主要的思路有两大类。

一类是比较直接的想法, 就是在自然刺激的条件下测试响应的统计特性。例如, Vinje 和 Gallant^[11]发表在 *Science* 上的文章中称, 从神经生理学实验中发现, 对于自然环境中的刺激, 视皮层细胞的响应满足稀疏分布。

另一类是推测感知系统信息处理的模型。首先检测外界信号的统计特性, 然后根据最优化理论推导一个转换工具, 从而能够给出感知系统对外界刺激响应的最佳描述。这方面的工作成果很多, 如 Olshausen 和 Field^[12]、Bell 和 Sejnowski^[13]、

van Hateren 和 van der Schaaf^[14]、Simoncelli 和 Schwartz^[15]等都提出了各自的模型来试图解释感知系统的信息处理过程。这些模型通过线性变换来模拟神经元细胞的相互作用,其缺点在于与人脑神经系统的实际工作机理不一致:复杂细胞并不是以线性方式从刺激中抽取相位和位移的不变特征。最近许多学者提出了用非线性模型模拟复杂细胞的功能,但仍没有给出其自组织映射的算法。

1.2.2 模拟人类方式的有效编码与特征表示

尽管很多研究人员都认为上述研究方法本质上都需要一个特征提取的过程,但实际上,在人类的神经系统中,即使不经过特征提取的预处理阶段,也能够处理不变的特征,如移动的物体、变化的对象等。神经系统能够获取不变特征的性质已经得到了生物物理学和神经生理学的证明^[16]。因此,通过模拟人类神经系统的信息处理方式提取不变特征已经成为另外一种解决有效编码问题的途径。

目前,国内外许多学者在这方面进行了深入的研究工作,研究的方法可以归纳为两大类。

第一大类是归一化方法。代表性的模型是 Olshausen 等^[17]提出的动态开关网络,Postma 等^[18]也提出了类似模型。该模型能够在识别的对象发生平移和尺寸变化时提取出不变的特征。但是由于动态开关网络仅仅使用了归一化方法,对于模式识别问题没有执行特征提取等处理,未能形成层次化的特征,因此无法描述更为复杂的特征,这是该模型的不足之处。

第二大类方法是基于不变特征的神经网络模型,比较著名的模型有以下几种。

Fukushima 等^[19]提出的神经认知机(neocognition)是一个层次化的神经网络,该网络受到生物视觉神经系统模型的启发,每层的神经元具有相同类型,分别是简单神经元层(S层)、复杂神经元层(C层)。神经元层又可以分成若干个子神经元平面,每一个子平面负责识别一个特定的特征,层与层之间有非常稀少并且固定模式的连接。这样随着一层一层的增加,提取的特征也随之增加。神经认知机的优点在于能够识别变形的模式,因此,在手写体识别方面取得了较好的效果,对于字符畸变、位移等具有较高的容错性,并且对字符的大小变化具有一定程度的抗干扰性。但是,神经认知机的计算量很大,随着识别对象数据的增多,模型中所需的神经元个数也随之增加。而且,在有监督的情况下,随着训练样本的增加,参数的选取也越发困难。

与神经认知机类似,高阶神经网络(higher order neural networks, HONN),也称为 Σ - Π 网络^[20],是另外一种能够实现几何不变特征提取的神经网络模型。它的最大优点在于利用输入模式中的已知关系,将所需的不变量直接构造入网络结构,相当于网络被“预训练”。因此,它不需要学习几何变换的不变量,也就不需要基于模式的各种变形来进行训练。二阶形式的 HONN 能够对平移及比例畸变不

敏感。而三阶的 HONN 能够实现平移、比例和旋转不变的对象识别。但是, HONN 也存在随网络的阶数和输入的维数增加, 权值数量迅速增多的问题。

共享权值神经网络(sharing weights neural networks)是另外一种很重要的不变特征提取模型^[21]。由于该模型最初主要应用在语音识别中, 所以也称为时延神经网络(time delay neural networks, TDNN)^[22]。而在图像分析方面, 由于时序关系不明显, 所以称为共享权值神经网络更为直观。它在网络结构方面的主要特点是, 位于同一层的所有单元共享相同的权值, 检测的特征也相同, 不同之处仅在于位置不同。这样, 这些检测到的特征送入高层时就能够实现平移不变的特征检测。同时, 通过子采样策略保证对畸变不敏感。由于该模型的训练采用 BP 算法, 因此易于陷入局部极小的情况, 且收敛速度较慢, 很多研究人员针对这些问题提出了很多改进方法^[23-24]。

另外一种比较重要的不变特征提取模型是 Kohonen 等^[25-27]提出的自适应子空间的自组织特征映射(adaptive-subspace self-organizing map, ASSOM)方法, 它结合了学习子空间方法和自组织特征映射网络的优点, 能够形成平移、缩放和旋转不变的滤波器。在这种方法基础上开发的 WEBSOM^[28], 应用于海量文本的组织 and 聚类, 以方便搜索与查询。WEBSOM 在包含 700 万条专利摘要的大数据集上进行了测试, 取得了很好的效果。也有研究者将 ASSOM 应用于手写字识别, 取得了相当好的识别率^[29]。

在上述几种不变特征提取模型的基础上, 研究者们对其进行了扩展, 添加了注意力控制机制来选择特定位置的局部区域特征^[30]。

由于人类的信息大都来自于视觉, 所以目前大部分的研究都集中在基于人类视觉感知系统的不变特征提取模型方面。因为自然环境下的图像含有大量的冗余, 所以学者们通过分析自然图像的统计特性, 对这个问题做了有益的探索。Field 和 Daugman 分别证明自然图像的高阶统计特性满足非高斯分布^[31-32], 也就是说, 大部分神经元对外界信号的响应很弱, 只有少部分神经元对外界信号有较强的响应, Field 把这种性质称为稀疏性(sparseness), 并且根据稀疏性进行自然图像的有效编码, 称为稀疏编码(sparse coding)。Olshausen 给出了自然图像的稀疏编码模型, 认为每幅图像可以看成多个基函数的线性组合, 这些基函数可以经过多幅图像的训练估计出来^[12, 33]。Bell, van Hateren, Lewicki, Hyvärinen 等采用独立分量分析的方法也取得了相似结果^[13, 14, 34-39]。Hyvärinen 等^[40]将独立特征子空间和多维独立分量分析的方法结合在一起提取自然图像的不变特性, 取得了很好的实验结果。

上述这些研究成果已经成功地应用在图像的特征提取、去噪以及图像的合成和分离等方面。虽然目前有效编码的研究工作主要针对图像信息, 但是许多学者的工作已经证实, 声音信息与图像信息从统计分析的角度看, 有很多相似之处^[41],

因此,相同的方法可以进一步扩展到语音处理领域中。

1.3 视觉表象的中级特征表示

对于视觉通道中神经细胞的特征加工过程,研究者提出了各种理论和假设。例如,神经生理学中的“感受野等级假设理论”,它认为视觉通道中神经细胞的感受野从低级往高级逐渐增大,所加工特征的复杂度也逐渐增高。在计算机科学中,Marr的“视觉计算理论”也体现了类似的观点,他把表达与算法层次的视觉表象(representation)也从低级往高级分为:要素图、二维半(2.5D)简图和三维图。它们都体现了一个共同的特点:分层表示和加工。但是,“感受野等级假设理论”只提供了一个公理性的定性描述,Marr的“视觉计算理论”虽然提出了量化的算法,然而根据图像中的测量值(如灰度)恢复相应表面的三维特性(如深度、方向),从本质上来讲是一个约束不充分(underconstrained)的问题,在实践中遇到了严重困难。按照连接主义的观点,生物视觉的信息处理是高度并行的,中级或者高级特征通过相互连接的多个初级特征(或者说要素)协同作用产生。当前神经科学中的许多研究都是关于探索这些连接和协同作用的。

如果说单个的、初等视神经细胞的编码是视神经通道信息处理的基础,那么中级或者更高级视皮层神经细胞的编码则是其核心和关键。对于中枢神经系统的编码,有两种不同的观点。其一为“祖母细胞”假说,即单细胞编码——以单个神经元的响应表示信息;其二为“动态细胞集群”的理论假设,认为输入刺激的信息是由神经元集群编码的。随着研究的深入,单一神经元编码逐渐受到人们的质疑,越来越多的结果表明:神经系统中信息的编码和处理在很大程度上由大量神经元构成的集群协同活动完成^[42-43]。虽然人们对神经元集群组成回路的宏观机制研究有了一定的进展,但对神经元集群的微观作用机制却缺乏充分的了解。具体说,由哪些神经元协同作用组成神经元集群?怎样找到这种协同响应的模式?神经元集群的输出怎样整合、加工成更加高级的特征?

对于神经元集群编码的协同响应模式,神经科学家的典型方法是通过对生物神经网络响应的观察和实验来得到它,相应的计算模型则是基于这些生理现象的一种硬布线方式以确定哪些神经元协同响应^[44],但是对这种协同响应的形成机制缺乏“定量”的分析。对于神经元集群的输出整合策略,研究者提出了各种模型和策略,包括线性组合模型^[45]、MAX模型^[44]和“分散规范化”模型^[15,46],它们适应于不同的场合,而且都存在相应的生理学依据。但是对于特征抽取来说,缺乏一种与任务相关的自适应策略和模型,指导不同任务下的视觉特征加工。机器学习作为一种有效工具在类似任务中得到成功应用^[47],例如,基于进化论的思想,我们成功地利用遗传规划算法抽取图像语义特征^[48],它有望为构造这种协同响应神经元群

的整合算子和策略提供一些思路,而且,它对中级特征抽取和图像识别都有着比较重要的意义。

1.4 初级视觉表象启发下的知觉组织

人类视觉系统的末梢和大脑皮层高级认知功能的始端,在纹状视皮层(即 V1 皮层)处交汇。V1 皮层是其他皮层视觉信息处理的起点和基础,各通路和所有中、高级感知任务都依靠 V1 皮层的输出。因此,理解 V1 皮层的视觉信息处理是非常重要的问题,对这一课题的研究自然成为计算神经科学领域的热点课题之一。外部世界的物体、场景在视网膜上形成的视像通过特定的变换和编码,以一一对应的拓扑映射方式在 V1 皮层上引起神经元群的发放,其发放模式(也称为活性模式)在计算神经科学中称为“表象”^[49],本质上就是包括编码与解码在内的信息处理过程。

对初级皮层视觉表象的透彻认识,不仅来源于对神经编码机制本身的了解,还依赖于对神经编码在整个视觉信息处理过程中功能延伸的研究,即神经编码是如何对其后继的信息处理作出贡献的。知觉组织正是在时间顺序上与编码相承接的认知过程,它将初级表象编码的原子信息(或时频分量)进行组织和整合,进而把这些信息组织成有明确景象意义的整体。Gray 等近 20 年所进行的一系列实验表明,格式塔心理学家描述的知觉组织现象与神经元以特定的频率同步发放有关^[50]。据此推测,初级视觉表象对知觉组织这一中级表象提供输入是通过同步振荡来实现的。更直接的证据是 V1 皮层的一些细胞对错觉轮廓(illusory contour)有最佳响应^[51],而视错觉是知觉组织的特例。因此,研究时间编码向知觉组织的表象转换,从低层的神经元脉冲发放的时间同步层次上研究知觉组织的内在机理,将会促进理解与静态和运动相关的知觉组织的时间编码的异同,从本质上了解不同知觉组织规则的优先关系,并最终对绑定(binding)和分割(segregation)等认知问题的科学解释提供理论基础。

知觉组织是生物视知觉中最基本的功能之一,用来分辨哪些信息属于一个整体进而把这些信息组织成一个有明确景象意义的目标。基本知觉组织原则是格式塔心理学家对各种知觉组织现象的内在规律的总结,包括邻接律、连续律、相似律、封闭律和协变律。由于知觉组织既涉及下层的特征提取等“初级”视觉信息加工,又涉及上层的检测、识别、语义理解等“高级”任务,在视知觉中起着承上启下的作用,所以对知觉组织的理解是最重要也是最困难的问题之一。困难主要源于两方面:一是人类对生物视知觉的机理理解不够,二是技术手段的限制所致。

在对皮层神经系统的研究中,从对单个细胞的结构和功能的研究占据主流地位,转向通过神经元群的整体行为和集合性质研究皮层分布式信息处理的观点被