

中图法分类号: TP3 文献标识码: A 文章编号: 1006-8961(2023)02-0335-23

论文引用格式: Zheng Y J, Yu Z F and Huang T J. 2023. A literature review for neural networks-based encoding models of biological visual system. Journal of Image and Graphics, 28(02): 0335-0357 (郑雅菁, 余肇飞, 黄铁军. 2023. 生物视觉系统的神经网络编码模型综述. 中国图象图形学报, 28(02): 0335-0357) [DOI: 10.11834/jig.220461]

生物视觉系统的神经网络编码模型综述

郑雅菁¹, 余肇飞^{1,2}, 黄铁军^{1,2*}

1. 北京大学计算机学院视频与视觉技术国家工程研究中心, 北京 100871; 2. 北京大学人工智能研究院, 北京 100871

摘要: 生物视觉系统的研究一直是计算机视觉算法的重要灵感来源。有许多计算机视觉算法与生物视觉研究具有不同程度的对应关系, 包括从纯粹的功能启发到用于解释生物观察的物理模型的方法。从视觉神经科学向计算机视觉界传达的经典观点是视觉皮层分层处理的结构。而人工神经网络设计的灵感来源正是视觉系统中的分层结构设计。深度神经网络在计算机视觉和机器学习等领域都占据主导地位。许多神经科学领域的学者也开始将深度神经网络应用在生物视觉系统的计算建模中。深度神经网络多层的结构设计加上误差的反向传播训练, 使得它可以拟合绝大多数函数。因此, 深度神经网络在学习视觉刺激与神经元响应的映射关系并取得目前性能最好的模型同时, 网络内部的单元甚至学习出生物视觉系统子单元的表达。本文将从视网膜等初级视觉皮层和高级视觉皮层(如, 视觉皮层第4区(visual area 4, V4)和下颞叶皮层(inferior temporal, IT))分别介绍基于神经网络的视觉系统编码模型。主要内容包括: 1) 有关视觉系统模型的概念与定义; 2) 初级视觉系统的神经网络预测模型; 3) 任务驱动的高级视觉皮层编码模型。最后本文还将介绍最新有关无监督学习的神经编码模型, 并展望基于神经网络的视觉系统编码模型的技术挑战与可能的发展方向。

关键词: 生物视觉系统; 分层结构设计; 深度神经网络; 视觉皮层编码模型; 无监督学习

A literature review for neural networks-based encoding models of biological visual system

Zheng Yajing¹, Yu Zhaofei^{1,2}, Huang Tiejun^{1,2*}

1. School of Computer Science, National Engineering Center of Visual Technology, Peking University, Beijing 100871, China;

2. Institute for Artificial Intelligence, Peking University, Beijing 100871, China

Abstract: The biological visual system, an important part of the brain's nervous system, has evolved over hundreds of millions of years. About 70% of the information that humans obtain from the outside world comes from vision. Its complicated systematic functions are relevant to visual pathways and visual cortex, as well as its mechanism. Human-perceptive and energy-efficient vision ability is better than machine-based vision system like real-time sensor data processing, perception tasks and motion control. To realize a more advanced machine vision paradigm, it is still challenged to learn from the design of biological ingenious vision system effectively. The biological vision systems-contextual researches can be recognized as one of the key aspects for computer vision algorithms. Conventional visual neuroscience to the computer vision domain is focused on the structure of hierarchical processing in the visual cortex. The following artificial neural networks (ANNs) are

收稿日期: 2022-05-23; 修回日期: 2022-10-12; 预印本日期: 2022-10-19

* 通信作者: 黄铁军 tjhuang@pku.edu.cn

基金项目: 国家自然科学基金项目(62176003, 62088102); 中国博士后科学基金(2022M720238)

Supported by: National Natural Science Foundation of China (62176003, 62088102); China Postdoctoral Science Foundation (2022M720238)

targeted on the hierarchical structure design in the visual system. Visual system is mainly composed of the eyes (retina), the lateral geniculate nucleus and the visual cortex (including the primary visual cortex and the striatal cortex). The human visual cortex and its relevance account for about 1/3 area of the cerebral cortex. It has the ability for visual information-related (e.g., extraction, processing and integration) and advanced brain functions-organized (e.g., learning, memory, decision-making, and emotion). For example, for the task of object recognition, the human brain can identify thousands of objects effectively, but this challenging issue is required to be resolved for machine-relevant. In recent years, deep neural networks (DNNs) have been projecting for computer vision and machine learning. To fit more multiple functions of network, the DNN plus multi-layer structure is designed for the back-propagation training. The biological visual system can be used to recognize as the mapping-learned relationship between the external visual information and the internal neuron expression. In addition, the neural network itself is a biological visual system-derived multi-layer structure design. Nowadays, the DNNs are the most accurate model for learning the mapping relationship between visual stimuli and neuron responses. The internal units of the ANN can learn the expressions of the internal subunits of the visual system further. The DNNs-hierarchical can predict the visual representation of visual neural response as well (e.g., V1, V2 and interior temporal of visual cortex). Furthermore, the latest unsupervised learning is employed to visual cortex. To outreach a new generation of general artificial intelligence (AGI), the research and development of ANNs and the exploration of brain function and its structure can be mutual-benefited. Our visual system-based review is focused on neural network-based coding models on the basis of primary visual cortex like retina and advanced visual cortex (e.g., V4, IT area). The main literatures are involved in: 1) concept and definition of the visual system model, 2) the neural network prediction model of the primary visual system, and 3) the goal-driven advanced visual cortex coding model. The latest unsupervised learning researches are reviewed and summarized literately. Technical challenges and future development directions of its neural network-encoding model are predicted further.

Key words: biological visual system; hierarchical processing; deep neural networks; encoding models of visual cortex; unsupervised learning

0 引言

大脑的视觉系统会在复杂的外界环境中提取出有意义的模式 (James 等, 1890)。如果能够准确记住食物、危险的标记或是一些重要伙伴的视觉特征, 将更有利于生存或繁殖。然而, 这些对象在每一时刻的位置、姿势、对比度、背景以及前景都各不相同, 仅利用低级图像属性不容易识别 (Pinto 等, 2008)。为解决这个问题, 在灵长类动物的视觉系统中, 会将外界的图像输入转换一个内部表达, 抽象的高级属性在这个表达中会被更形象的编码, 以便指导行为 (DiCarlo 等, 2012)。

大脑视觉系统是由一系列解剖学上可区分但相互连接的区域组成 (Felleman 和 van Essen, 1991; Malach 等, 2002)。级联的每个单独阶段都执行比较简单的神经变换, 例如输入的线性加权, 或阈值激活和归一化等非线性变换。然而, 复杂的非线性变换可能源于简单变换的串联。由于大脑所接收的输入通常都是各种数据的纠缠, 是高度非线性的表达,

因此, 解码的过程也是高度非线性的。

大脑神经网络所处理的非线性转换空间非常庞大。因此, 理解感知系统的一个主要挑战是系统识别 (system identification)——识别真正的生物回路使用的是哪种转换。虽然生物神经回路的转换函数可能是有用的, 例如, 感受野 (receptive field) 的表征。但解决系统识别问题的最终目标是产生一个编码模型: 一种接受任意刺激输入 (例如, 任何像素图) 并输出对该刺激的神经反应的正确预测的算法。模型不能局限于解释一种狭义的现象, 例如, 面向精心挑选的神经元且高度控制和简化的刺激而定义的。Yamins 和 Dicarlo (2016) 认为能够接受任意输入刺激, 并且能够精确预测某一区域所有神经元的反应是大脑感知模型必须满足的两个核心标准。此外, 一个全面的编码模型不能仅仅预测一个最终区域的神经元的刺激—反应的关系, 例如 (在视觉中) 前颞下皮层 (inferior temporal, IT)。相反, 该模型还必须是可映射的, 即具有与中间皮层区域 (例如 V4) 相对应的可识别组件。模型在每个组成区域的响应应正确预测相应大脑区域内的神经反应模式。

如今,在大量解剖学和生理学证据的支持下,普遍接受这种分层级组织及其在人类和非人类灵长类动物中的双通路结构:腹侧通路(ventral stream)和背侧通路(dorsal stream)(Markov 等,2013;Ungerleider 和 Haxby,1994;van Essen,2003)。如图 1 所示(Gilbert,2013),外界的视觉信息通过两条平行通路(“视网膜—外侧膝状体—皮层”)从视网膜流向初级视觉皮层(V1 区)。其中,大细胞(也称为 M 细胞, Magno 为“大”的拉丁语)通路传递粗略的、基于亮度的空间输入,对区域 V1 的第 4C α 层具有很强的时间敏感性,星状神经元的特征细胞群立即将信

息传输到涉及运动和空间处理的更高皮质区域;而细小细胞(也称为 P 细胞, Parvo 为“小”的拉丁语)通路传递高空间分辨率但低时间敏感性的输入,信息通过 4C β 层进入区域 V1。这种对颜色敏感的输入在 V1 的不同层中缓慢流动,然后流向皮层区域 V2 和参与形式处理的皮层区域网络。在这两条平行的视网膜—LGN—视皮层通路的视觉系统理论中认为(Mishkin 等,1983;Milner 和 Goodale,2008),背侧流专门用于运动感知和视觉场景空间结构的分析,而腹侧流专门用于形成感知,包括对象和人脸识别。

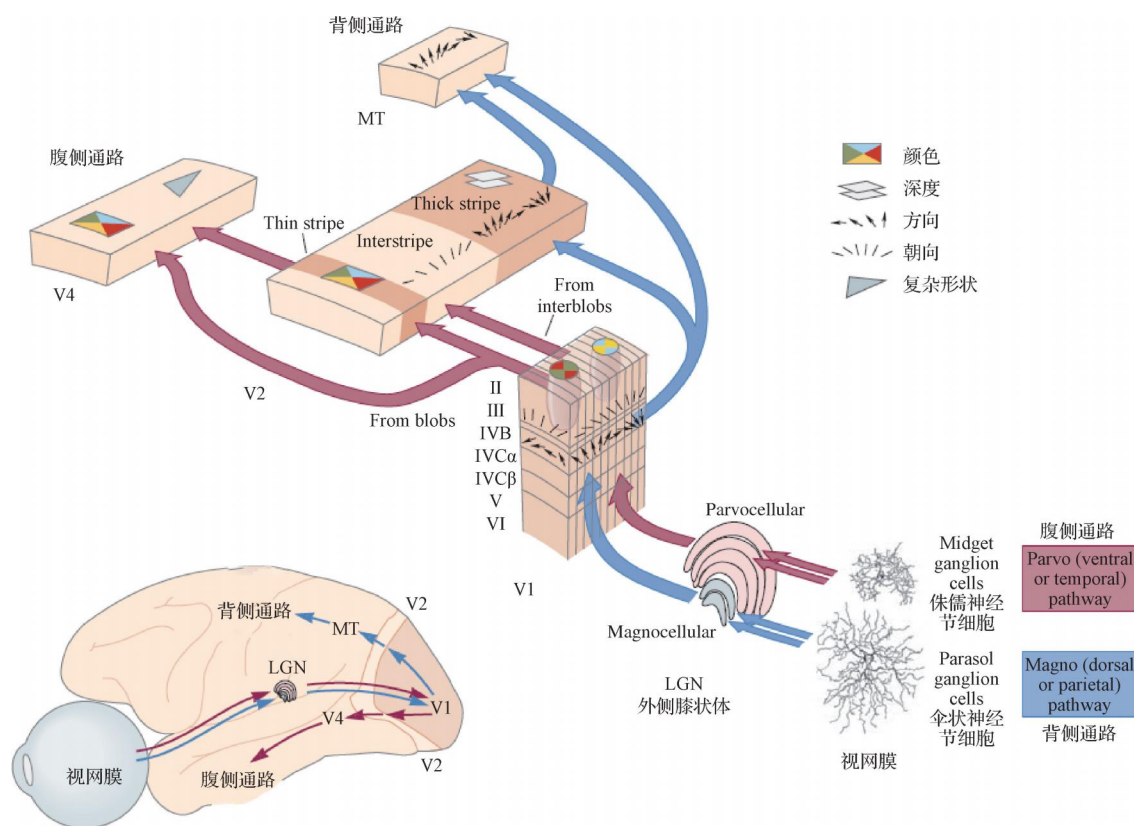


图 1 视觉系统的平行通路:腹侧通路及背侧通路(Gilbert,2013)

Fig. 1 Parallel pathway of visual system: ventral stream and dorsal stream(Gilbert, 2013)

自从 Hubel 和 Wiesel(1962)发现初级视觉皮层 V1 区域的简单细胞和复杂细胞以来,对视觉系统神经科学的相关研究表明,大脑通过一系列分层组织的皮层区域(腹侧视觉流)产生具有不变的对象识别行为(invariant object recognition)。基于 Hubel 和 Wiesel(1962)所提出的开创性工作,许多研究人员已经建立了受生物学启发的神经网络(Fukushima 和 Miyake,1982;Riesenhuber 和 Poggio,1999;Serre 等,2007;Haß 等,2008;Bengio, 2009;Pinto 等,

2009)。随着时间的推移,人们意识到这些模型是一类更通用的计算框架的例子,称为分层神经网络(LeCun 和 Bengio,1998)。分层神经网络是由多层简单的层堆叠而成的,外界的感知输入将会被这些串联的层依次处理。每一层的结构都比较简单,但由这些层组成的深层网络可实现对输入数据的复杂转换,类似于腹侧流中视觉信息的处理。每个神经网络层中的运算也是由传统生物物理模型——线性—非线性(linear-nonlinear, LN)环路的启发(Sah-

ani 和 Linden, 2003; Machens 等, 2004; Carandini 等, 2005)。在线性—非线性模型中主要包含以时空滤波器为主的线性加权操作, 以及以阈值发放、池化和归一化等运算为主的非线性变换。

与神经元的感受野类似, 分层神经网络中的操作基本都在一个固定大小的局部输入区域上进行, 该区域通常小于输入的完整空间范围。例如, 在 256×256 像素的图像上, 网络层的感受野可能是 7×7 像素。由于这些感受野在空间上是重叠的, 所以滤波器和池化操作通常是“跨步的”, 这意味着输出仅保留每个空间维度上的一小部分位置, 例如步长为 2 的卷积核在滑动时将每隔一行/列就跳过。此外, 每个卷积核在所有空间位置进行权值共享, 即每个位置都会有相同的卷积操作。所以即使物体的空间位置发生变化, 网络的物体识别结果也具有不变性。虽然目前腹侧流和其他大脑皮层的生理学结构似乎排除了可以存储共享模板的主导位置存在, 然而, 外界环境中的自然视觉/听觉统计数据在时空域上很大程度上是不变的。因此, 大脑中基于经验的学习过程应该倾向于学得不同空间或时间位置的收敛权重。共享权重可能是大脑视觉系统中央视野的合理近似, 但真实的视觉系统具有很强的中央凹偏差, 因此, 如果对所有位置都采用不同的感受野, 以实现更仿生的视觉信息处理可能会提高模型对神经数据的拟合度。

由于深度学习技术的不断发展, 已经出现了许多基于神经网络的精确预测模型, 例如基于任务优化的深度卷积神经网络 (deep convolutional neural network, DCNN) 是目前编码灵长类动物大脑视觉的最精确预测模型。Yamins 和 Dicarlo (2016) 认为这主要是因为 DCNN 具有以下特性: 1) 视觉皮层启发式的结构, 即神经网络的拓扑结构与视皮层的信息处理流程相似; 2) 网络的训练/优化的目标与动物生存时必须执行的行为一致, 如物体识别等任务。

除了建模高级视觉皮层的物体识别网络外, 还有许多精确预测初级视觉系统 (包括视网膜和 V1 等) 响应的模型, 以及模拟视觉皮层背侧通路的神经网络。本文首先描述有关生物视觉系统预测模型, 及人工神经网络的背景知识, 再对比不同视觉区域的神经网络模型。最后总结基于神经网络视觉计算模型的技术挑战, 并展望未来发展方向。

1 概念与定义

1.1 初级视觉系统的结构及特性

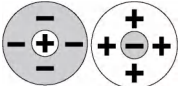
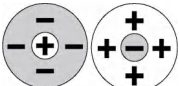
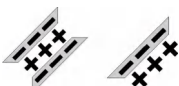
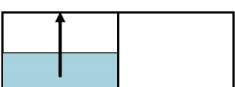
在使用分层神经网络对视觉皮层进行建模之前, 有许多基于手工设计的模型。从 20 世纪 70 年代开始, 有许多神经计算领域的研究者对初级视觉系统进行建模, 例如 V1 区域。这些区域的神经元可能可以通过相对较浅的网络来解释, 例如, 基于一个或多个线性感受野的模型可能实现精准预测神经元对任意输入刺激的响应, 前提是模型包括控制响应性的非线性机制, 基于刺激上下文和历史, 并考虑脉冲生成的非线性。

大多数初级视觉系统的神经元模型都是基于线性感受野的概念。感受野这一概念最初是由 Sherrington 提出用于描述反射活动, 之后由 Hartline 引入到视觉系统中。视觉系统中的神经元感受野定义为光照视网膜时, 能改变神经元活动的区域。表 1 中展示了视网膜中光感受器和神经节细胞, 外侧膝状体的核细胞, V1 区域的简单细胞、复杂细胞和终端抑制复杂细胞的感受野特性。其中视网膜的神经节细胞和外侧膝状体中的核细胞感受野都为中心圆形的感受野, 当给定的光源形状符合感受野特性时, 神经元才会发放脉冲信号, 反之神经元不会发放脉冲信号。V1 区域的简单细胞则是对边缘或者窄条状的刺激会产生响应, 但给定的刺激有明确的“给光”和“撤光”区域。而复杂细胞则无需明确“给光”及“撤光”区, 当符合其朝向的刺激出现时就会产生响应。终端抑制型复杂细胞则是在复杂细胞的基础上, 还会受到边缘端点的影响, 当刺激的光条超过其最佳长度时, 神经元的响应将会减弱。

视网膜和外侧膝状体核 (lateral geniculate nucleus, LGN) 是视觉系统处理外界信息的始发站。但这两种结构在目前很多机器视觉模型中发挥的作用很大程度上被低估了。目前大多数的视觉模型都以图像作为输入, 而不包含生物视觉系统中的视网膜—LGN 变换过程。因为忽略了在这些初级视觉区域上处理的内容, 人们很容易错过一些关键属性, 以了解是什么使生物视觉系统能高效率工作。

在视网膜中, 入射光被转化为电信号。这种转换最初是通过使用线性系统方法来模拟视网膜图像

表 1 不同初级视觉区域神经元的感受野特性
Table 1 Receptive field of different neurons in early visual system

细胞类型	感受野形状	最佳刺激	是否受刺激朝向影响	是否有明确的“给光”和“撤光”区
光感受器		光	否	无
神经节细胞		覆盖中心的小光点或窄条	否	有
外侧膝状体核细胞		覆盖中心的小光点或窄条	否	有
简单细胞		窄条或边缘	是	有
复杂细胞		条形或者边缘	是	无
终端抑制型复杂细胞		有端点的线或者边缘; 尖角或者拐角	是	无

的时空过滤进行描述的 (Enroth-Cugell 和 Robson, 1984)。最近的研究改变了这种观点,并且在不同脊椎动物的视网膜中发现了几种类似皮层的计算 (Kastner 和 Baccus, 2014; Gollisch 和 Meister, 2010)。尽管在不同的空间和时间尺度上工作,但视网膜和皮层水平具有相似的计算原理,这一事实是设计生物视觉模型时需要考虑的重要一点。这种观点的改变将产生重要的后果。例如,与其考虑皮层电路如何实现视觉处理的高时间精度,不如考虑密集互连的皮层网络如何保持视网膜对静态和运动自然图像编码的高时间精度 (Field 和 Chichilnisky, 2007),或者微型眼球运动如何塑造其时空结构 (Rucci 和 Victor, 2015)。

同样地, LGN 和其他视觉丘脑核 (例如,丘脑核) 不应再被视为从视网膜到皮层路径上的纯中继。例如,猫枕神经元表现出一些经典归因于皮层细胞的特性,模式运动选择性 (Merabet 等, 1998)。猴子 LGN 神经元中显示出强烈的中心环绕相互作用,这些相互作用受反馈皮质丘脑连接的控制 (Jones 等, 2012)。这些强大的皮质原发性反馈连接可能解释了为什么平行的视网膜丘脑—皮质通路是高度适应性的动态系统 (Briggs 和 Usrey, 2008;

Cudeiro 和 Sillito, 2006; Nandy 等, 2013)。此外,与视网膜只接受外界视觉刺激不同, LGN 还接受来自视皮层的反馈信息,代表了视觉通路中皮层自上而下的反馈信号可能影响信息处理的第 1 阶段 (O’Connor 等, 2002)。

1.2 初级视觉系统的预测模型

关于初级视觉区域系统的编码研究通常与系统辨识方法密切相关。通过结合神经生物学实验,系统辨识方法主要通过构建模型将视觉场景编码为神经响应,以找到神经系统中的计算单元,例如神经元的感受野和非线性变换等计算组件。随着多电极阵列技术的发展,人们可以使用各种类型的光学图像来操纵光学刺激,包括简单的条形、斑点和光栅,以及复杂的自然图像和视频等,同时可以记录初级视觉区域细胞群体的响应。在这些实验技术的支持下,初级视觉系统的编码模型迅速发展。

脉冲激发平均发放模型 (spike-triggered average model, STA) (Marmarelis 和 Naka, 1972; Chichilnisky, 2001) 是最早应用于视网膜等初级视觉皮层的模型。该模型可以分析视网膜神经节细胞、LGN 和 V1 细胞的感受野。研究人员 (Paninski, 2003; Liu 和 Gollisch, 2015) 提出了脉冲激发协方差模型 (spike-

triggered covariance model, STC), 利用奇异值分解获得基本的时间空间滤波器。除了分析视网膜感受野滤波器的模型外, 还有分析视网膜响应的线性—非线性(LN)模型(Sahani 和 Linden, 2003; Machens 等, 2004)。在 LN 模型中, 假设神经元从接收刺激到决定是否做出响应, 其主要经历两个阶段: 在第 1 阶段, 刺激经过时空线性滤波器, 这些滤波器描述了神经元整合输入的方式, 即描述了神经元感受野的特性; 在第 2 阶段, LN 模型假设原始刺激首先经过时空滤波器过滤之后, 会经过一个非线性的过程, 这个非线性是模仿脉冲信号生成机制, 以获得神经脉冲信号的输出。

到目前为止, 有许多模型通过改进 LN 模型的结构, 从而变得更加复杂, 例如线性—非线性泊松模型(linear-nonlinear Poisson model, LNP)(Schwartz 等, 2006)。LNP 模型在 LN 模型的非线性过程之后增加了一个泊松过程来模拟神经元的发放; Pillow 等人(2008)提出了一种广义线性模型(generalized linear model, GLM)。GLM 模型中增加了更多功能模块, 例如增加历史脉冲滤波来模拟神经元的适应性, 增加耦合滤波来模拟相邻神经元之间的相关性。非线性输入模型(nonlinear input model, NIM)(McFarland 等, 2013)利用非线性滤波器来处理输入信号之间的相关性。除了这些较为简单的结构, 近年来包含子单元组件的模型也越来越多, 如线性—非线性级联网络模型(2-layer linear-nonlinear network model, LNLN)(Maheswaranathan 等, 2018)、脉冲激发非负矩阵分解模型(spike-triggered non-negative matrix factorization, STNMF)(Liu 等, 2017)。

根据 Hubel 和 Wiesel(1962)的经验观察表明, V1 中的神经元类似于 Gabor 小波滤波器, 不同的神经元对应于不同频率和方向的边缘。事实上, 早期使用手工设计的 Gabor 滤波器组作为卷积权重的计算模型在解释 V1 神经反应方面取得了一些成功。后来人们意识到, 使用阈值、归一化和增益控制等非线性可以显著改进模型。此外, Hubel 和 Wiesel(1962)也提出了 V1 中简单细胞和复杂细胞感受野的形成是由其上游细胞感受野组合而成的观点。在这个观点中认为, 视觉系统中复杂的感受野结构是由许多输入以有序的方式会聚而成的。但是, 由于视觉皮层中也存在着许多复杂的反馈或者“跳线”连接, 如复杂细胞也会接收来自 LGN 细胞的直接输

入。并且视觉皮层内部也普遍存在许多水平连接(Gilbert 等, 1990; Souihel 和 Cessac, 2021)。尽管如此, Hubel 和 Wiesel(1962)提出的这种分层组合结构, 启发设计了许多机器视觉模型的产生, 如分层最大池化模型(hierarchical max-pooling models, HMAX)(Riesenhuber 和 Poggio, 1999)和深度神经网络(LeCun 等, 2015)。

1.3 视觉皮层腹侧通路的物体识别模型

Hubel 和 Wiesel(1962)发现了 V1 区域细胞的感受野特性和视觉皮层的层级处理结构, 许多研究者在他们的基础上提出了实现腹侧通路物体识别功能的模型, 其中比较经典且经常被作为基础模型使用的为 VisNet(Rolls 和 Milward, 2000)和 HMAX 模型(Riesenhuber 和 Poggio, 1999)。

1.3.1 VisNet 物体识别模型

大脑视觉系统所解决的主要问题之一是建立视觉信息的表征, 使识别相对独立于大小、对比度、空间频率、视网膜上的位置、视角等。这种识别性能的泛化性无法由简单的某个视觉皮层独立实现。用于物体识别的皮层视觉处理是由腹侧通路中分层连接的皮层区域完成的。如图 2 右侧所示, 腹侧通路至少包括 V1、V2、V4、后下颞叶皮层(posterior inferior temporal cortex, TEO), 以及前颞皮层(anterior inferior temporal cortex, TE)。一个区域的每个小部分都会聚于后续处理区域中的一个局部。因此, 每个后续阶段的神经元的感受野大小都会放大约 2.5 倍。如图 2 右侧所示, V1 中中央凹区域的感受野视角约为 1.3° , V4 为 8° , TEO 为 20° , 下颞叶皮层为 50° (Boussaoud 等, 1991)。这样的会聚区域将不断相互重叠。不同区域神经元感受野的连通性是实现物体平移不变表示的关键。

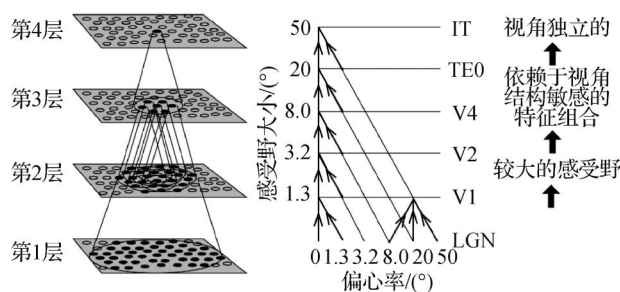


图2 VisNet 的结构及对应的腹侧视觉通路
(Rolls 和 Milward, 2000)

Fig. 2 The structure of VisNet and the corresponding ventral visual pathway (Rolls and Milward, 2000)

基于视觉皮层中面向对象识别的不变性, Rolls 和 Milward (2000) 提出了一个完成物体识别任务的模型——VisNet。如图2左侧所示, VisNet 具有4层前馈层次结构, 并通过使用横向抑制连接提供每层内神经元之间的竞争机制。神经元间权重是根据赫布规则的扩展版本调整, 即当输入轴突可以使得输出神经元强烈发放时, 它们之间的突触权重将增强, 反之, 与不活跃的输入轴突将减弱。

1.3.2 HMAX 模型

HMAX 模型曾经是模拟视觉皮层腹侧流最好的模型, 该模型由 Riesenhuber 和 Poggio 于 1999 年首次提出 (Riesenhuber 和 Poggio, 1999)。该模型的关键元素是一组位置和尺度不变特征检测器。使用了一种非线性最大池化 (max pooling) 机制, 该机制能够为识别杂乱情况提供更稳健的响应。

图3展示了 HMAX 模型的结构图。模型采用多层简单—复杂细胞的交叠结构, 模拟腹侧通路中各个视觉皮层提取从简单到抽象特征的过程。HMAX 模型最后一层的视角调谐细胞 (view-tuned cell) 可表示高度抽象的特征。

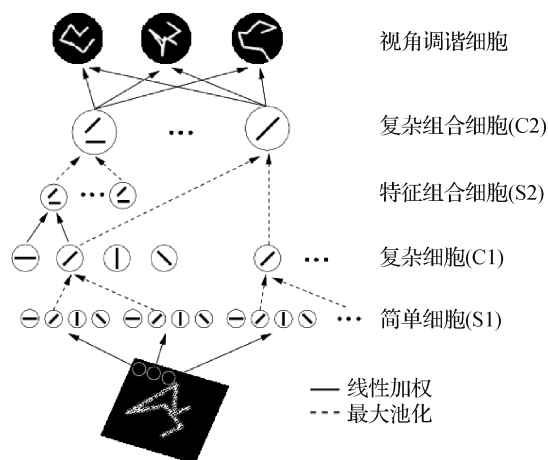


图3 HMAX 模型的结构 (Riesenhuber 和 Poggio, 1999)

Fig. 3 The structure of HMAX model
(Riesenhuber and Poggio, 1999)

1.3.3 卷积神经网络

受神经科学研究的启发 (Hassabis 等, 2017), 典型的卷积神经网络 (convolutional neural network, CNN) 由层次结构组成, 其中每个卷积层最重要的属性之一是可以使用卷积滤波器作为特征检测器提取来自输入图像的有效信息 (Krizhevsky 等, 2012; Simonyan 和 Zisserman, 2015)。因此, 经过训练, 卷

积核中的滤波器具有一定的语义。这些滤波器捕获的特征可以在原始自然图像中表示 (Zeiler 和 Fergus, 2014)。通常, 一个典型特征与训练集里的某些图像块有一些相似之处。而这些相似性是通过使用大量具有标签数据的图像来获得的, 所以卷积核所学得的特征对于一类物体来说具有通用性, 有利于识别。然而, 由于自然图像的复杂统计结构 (Simioncelli 和 Olshausen, 2001), 它也导致了可视化或解释卷积核的意义时存在困难。因此, 卷积神经网络中的神经元的意义也通常不易于解释 (Zeiler 等, 2011)。

2 初级视觉系统的神经网络预测模型

视觉处理始于视网膜和外侧膝状体核 (LGN), 目前很多机器视觉模型中, 这两种结构所发挥的作用很大程度上被低估了。目前的大多数视觉模型都将图像作为输入, 而不包含生物视觉系统中的视网膜—LGN 变换过程。它们并不仅仅是视觉刺激与大脑之间的一个传输中继站, 如视网膜细胞中还具备运动选择性等特征, 而 LGN 中的细胞则有增益控制 (gain control) 的功能, 可以有效减少眼动或者头部移动时伴随的视觉信息的剧烈变化。

传统生物物理模型在研究初级视觉系统对简单人造刺激图像非常有效, 但是对于自然场景, 包括自然图像和自然视频, 却存在许多问题, 例如模型参数量过大时难以拟合。近年来, 深度学习发展迅速, 在对象识别和分类方面取得的效果优于人类的水平。因此, 有许多研究者也开始关注如何使用人工神经网络对初级视觉区域的编码过程进行建模 (Maheswaranathan 等, 2018; Batty 等, 2017; Vance 等, 2018; Yan 等, 2020)。

初级视觉系统中的神经元电路组织相对清晰和简单, 可以使用卷积神经网络进行建模。另一方面, 机器视觉的研究者们也期望这些神经元回路的知识可以为 CNN 提供有用且重要的验证。近年来, 一些神经科学领域的研究者将 CNN 及其变体应用于初级视觉系统, 例如视网膜 (McIntosh 等, 2016; Batty 等, 2017; Vance 等, 2018; Maheswaranathan 等, 2018; Yan 等, 2020)、V1 (Vintch 等, 2015; Antolík 等, 2016; Kindel 等, 2017; Cadena 等, 2019; Klindt 等, 2017; Whiteway 等, 2018; Ukita 等, 2018) 和 V2

(Rowekamp 和 Sharpee, 2017)。这些研究的目标大多数是通过使用前馈和递归神经网络,或结合两者来实现更好的神经反应预测。与传统的线性/非线性模型相比,这些新方法复杂性的增加提高了对视觉系统的辨识能力(McFarland 等, 2013; Chichilnisky, 2001; Liu 和 Gollisch, 2015)。其中一些研究还重构所训练的网络中隐层计算单元的细节,以对比它们是否与生物神经元的功能及结构相对应(Maheswaranathan 等, 2018; Klindt 等, 2017)。

与其他初级视觉皮层系统(如 V1, LGN)等相比,目前对视网膜的解剖结构了解较为清楚(Gollisch 和 Meister, 2010)。如图 4 所示,视网膜可简化为一个 3 层的网络,分别为光感受器、双极细胞和神经节细胞。在这 3 层网络之间穿插着抑制性水平细胞和无长突细胞等结构。视网膜神经节细胞(retinal ganglion cells, RGCs)作为视网膜的最终层输出神经元,通过视束和丘脑将视觉信息发送到皮层区域进行高级认知。每个神经节细胞会接受一些兴奋性双极细胞(bipolar cells, BCs)的输入,并产生脉冲信号。需要强调的是,哺乳类动物的视网膜是一个由至少 60 种不同结构和特性的神经元组成的纵横交错的复杂神经网络(Gollisch 和 Meister, 2010),上

述的 3 层神经网络是一个极为简单的计算模型。

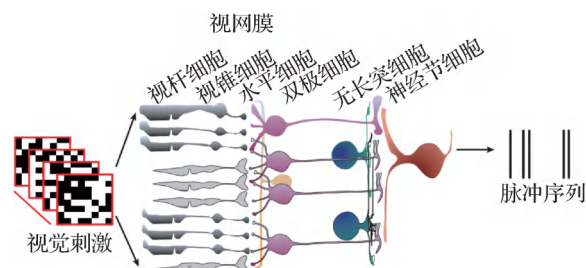
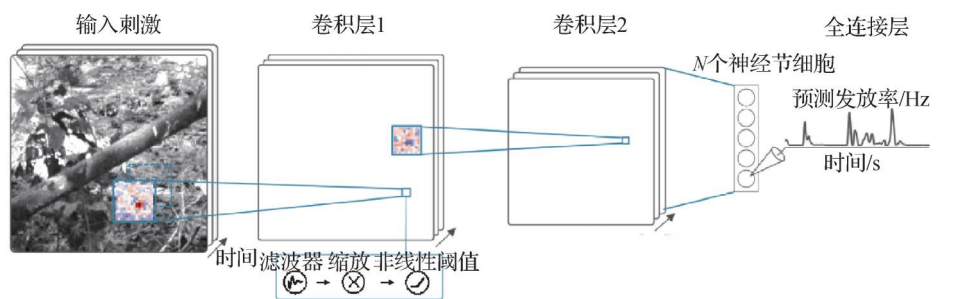


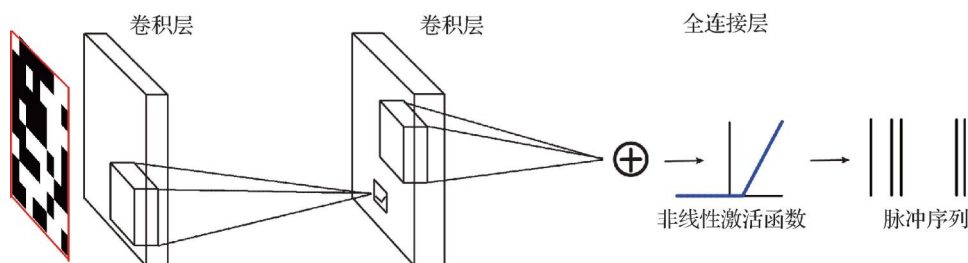
图 4 视网膜的生理结构(Yan 等, 2020)

Fig. 4 Illustration of the retinal structure(Yan et al., 2020)

Yan 等人(2020)也使用 3 层的卷积神经网络建模单个视网膜神经节的编码过程,其模型结构如图 5(b)所示。他们对单个细胞在不同刺激及不同细胞上的迁移学习实验,揭示了一个 3 层的卷积神经网络是在学习一个神经节细胞所需的基本运算。在一个细胞上训练好的卷积神经网络,可以较好地在不同刺激上进行迁移,但在不同细胞上迁移时性能具有较大的差异。最近的一些研究也探索了循环连接在视网膜编码中发挥的作用,Batty 等人(2017)使用循环神经网络(recurrent neural network, RNN)对神经元群体内的共享特征空间进行建模。然而,这种方法的性能主要取决于对神经元初始位置的估计。



(a) 群体神经节细胞的编码模型(McIntosh 等, 2016)



(b) 单个神经节细胞的编码模型(Yan 等, 2020)

图 5 基于 3 层 CNN 的视网膜编码模型

Fig. 5 Retinal encoding model based on 3-layer CNN((a) encoding model of population ganglion cells (McIntosh et al., 2016); (b) encoding model of single ganglion cell(Yan et al., 2020))

虽然卷积神经网络已成功应用于建模简单和静

态场景的视网膜系统,然而,当学习大规模视网膜神

经节细胞对复杂自然场景视频的编码过程时, 基于卷积神经网络的编码模型会随着所编码 RGC 数量的增多而参数量陡然上升, 因而难以得到有效的编码模型。为解决这个问题, Zheng 等人 (2021a) 提出了用于学习群体视网膜神经节细胞编码外界动态视频刺激的卷积循环神经网络 (convolutional recurrent neural network, CRNN)。图 6 中展示了 CRNN 的结

构, 实验结果揭示了网络的循环连接结构是影响视网膜编码的关键因素, 可高精度地预测大规模视网膜神经节细胞对动态自然场景的响应, 并可以同时学习出各个神经节细胞的感受野。Zheng 等人 (2021a) 所提出的卷积循环编码网络除了在结构上更加接近视网膜, 还可以使用更少的参数学习出精度更高的编码模型。

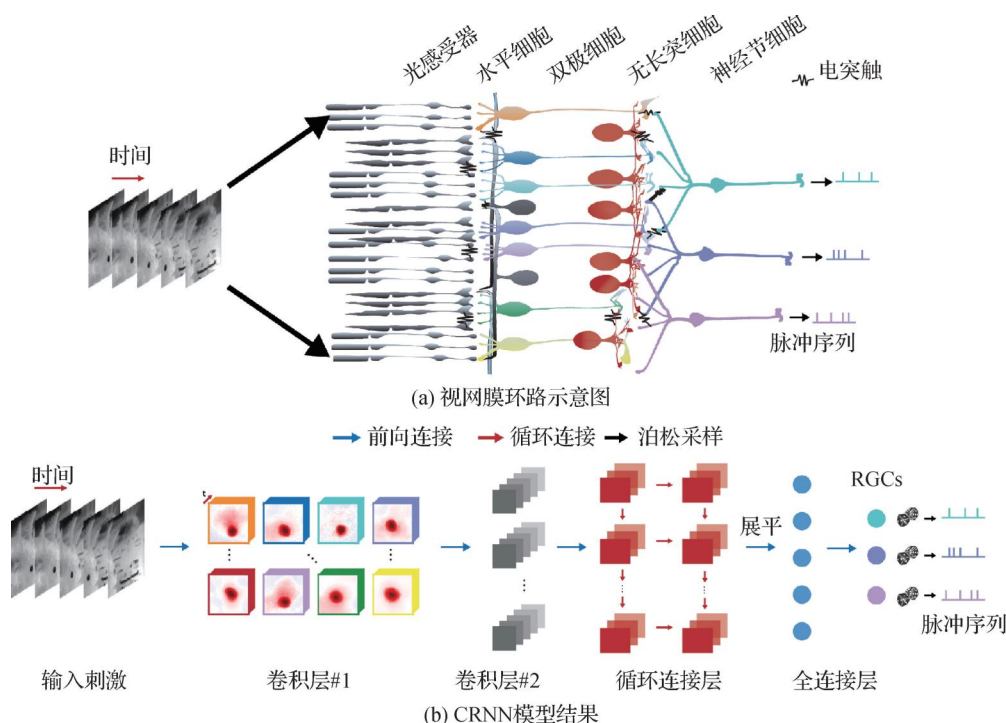


图 6 CRNN 模型架构图例 (Zheng 等, 2021a)

Fig. 6 Illustration of the CRNN model architecture (Zheng et al., 2021a)

((a) schematic diagram of the retinal circuit; (b) structure of the CRNN model)

此外, 即使目前对 V1 神经元的感受野已比较了解, 但传统的预测模型也很难预测它们对自然图像刺激的反应。为了填补这一空白, Kindel 等人 (2017) 训练深度卷积神经网络来预测 V1 神经元对自然图像刺激响应, 并且网络中模拟 V1 神经元的节点经过分析, 可得到具有中心环绕状和 Gabor 小波的感受野, 甚至是具有更高级别纹理特征的复杂感受野。Kindel 等人 (2017) 的网络结构如图 7 所示。

在接受外界视觉刺激时, 有不同类型的神经元在视野的不同位置执行相似的计算。传统的神经系统识别方法没有区分神经元感受野的位置和形状特性。而卷积神经网络中, 卷积核在多个位置的权值

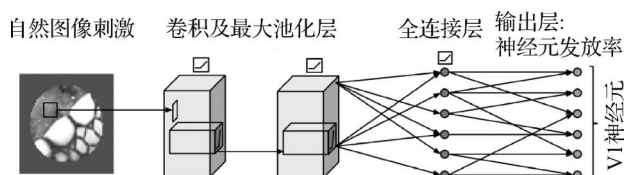


图 7 基于 CNN 的 V1 细胞编码模型 (Kindel 等, 2017)

Fig. 7 CNN encoding models of population V1 neurons (Kindel et al., 2017)

共享, 但架构设计需要考虑数据限制: 虽然新的实验技术能够记录数千个神经元, 但实验时间有限, 因此只能对每个神经元响应空间的一小部分进行采样。Kindt 等人 (2017) 认为使用卷积神经网络拟合神经数据的一个主要瓶颈是对单个感受野位置的估计。因此, 他们提出了一个带有稀疏读出层的卷积神经

网络结构,其可以同时重构出神经元感受野的空间和特征维度。Klindt 等人(2017)提出的网络可以很好地扩展到数千个神经元和简短的记录,并且可以进行端到端的训练。

除了这种直接训练网络预测神经元对输入的数据驱动模型外,还有以物体识别为任务训练一个多层网络后,对网络的浅层网络使用 V1 神经元数据进行迁移学习。Cadena 等人(2019)测试了这两种

方法预测清醒猴子 V1 神经元对自然图像脉冲响应的能力。他们发现以物体识别任务训练的 VGG-19 (Visual Geometry Group) 网络 (Simonyan 和 Zisserman, 2015) 经过迁移学习后,与数据驱动方法的表现相似,并且都优于基于 V1 理论的经典线性—非线性泊松模型(LNP 模型)和基于小波的特征表示(gabor-filter bank, GFB 模型)。Cadena 等人(2019)所使用的网络结构及对应 V1 的选择层如图 8 所示。

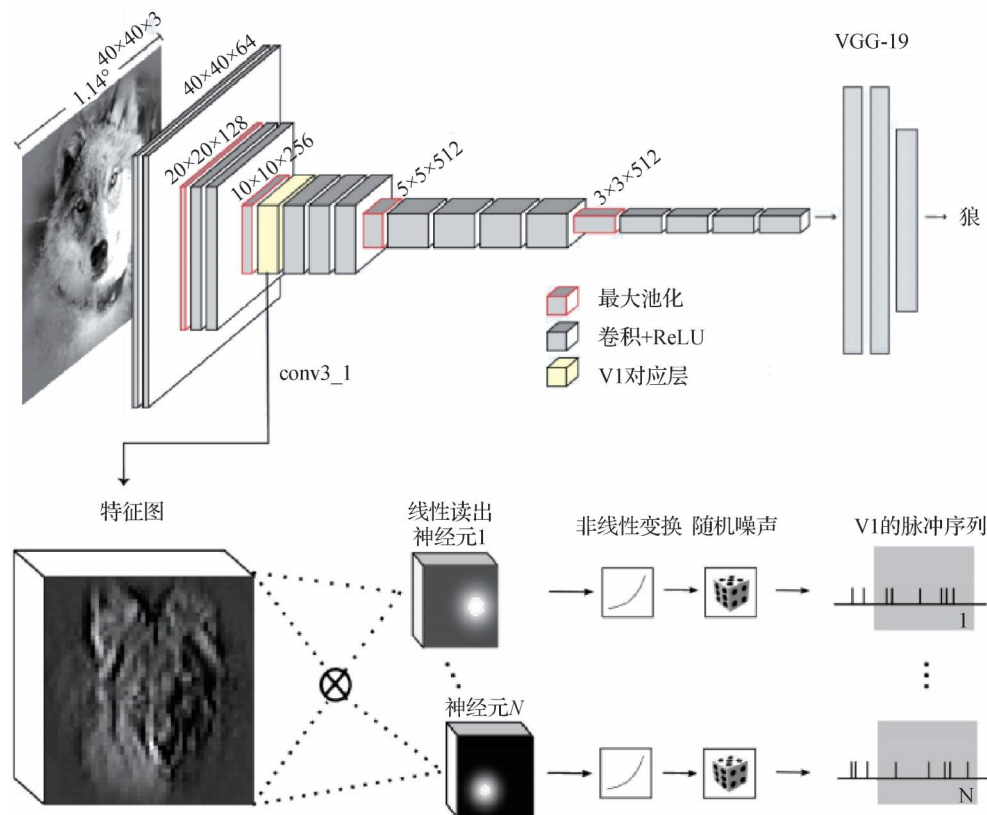


图 8 VGG-19 网络迁移学习的 V1 细胞编码模型 (Cadena 等, 2019)

Fig. 8 V1 cell coding model obtained by transfer learning of VGG-19 network (Cadena et al., 2019)

值得注意的是,使用预训练的特征空间进行迁移学习时,可使用较少的实验时间就实现相同的预测性能。Cadena 等人(2019)的实验结果表明,多层卷积神经网络为预测灵长类动物 V1 中对自然图像的神经反应奠定了新的技术水平,并且比传统滤波器组合成的模型,以对象识别任务为目标的网络学得特征可以更好地解释 V1 细胞的特性。这一发现强化了 V1 模型中具备多重非线性的必要性,并验证了高级功能目标(如物体识别)可影响早期视觉皮层的表征。图 9 中展示了传统 V1 模型 LNP、GFB、V1 数据驱动的浅层卷积网络和以物体识别任务训练的大规模 VGG 网络对 V1 神经元的可解

释性。

除了分别训练视网膜和初级视觉皮层的网络外, Lindsey 等人(2019)还提出了一个同时模拟视网膜和腹侧视觉通路的联合网络。该网络结构如图 10 所示。将在 CIFAR-10 数据集 (Krizhevsky, 2009) 上训练的深度卷积神经网络作为视觉系统的模型,并且认为这种感受野的差异可能是视网膜和皮层网络上不同神经资源限制的直接结果。其中关键的限制是模拟视网膜的网络中输出的神经元数量减少。此外,如图 10 所示,对于简单的下游皮层 VVS (ventral visual system) 网络,视网膜输出的视觉表征表现为非线性和有损特征检测器,而对于更复

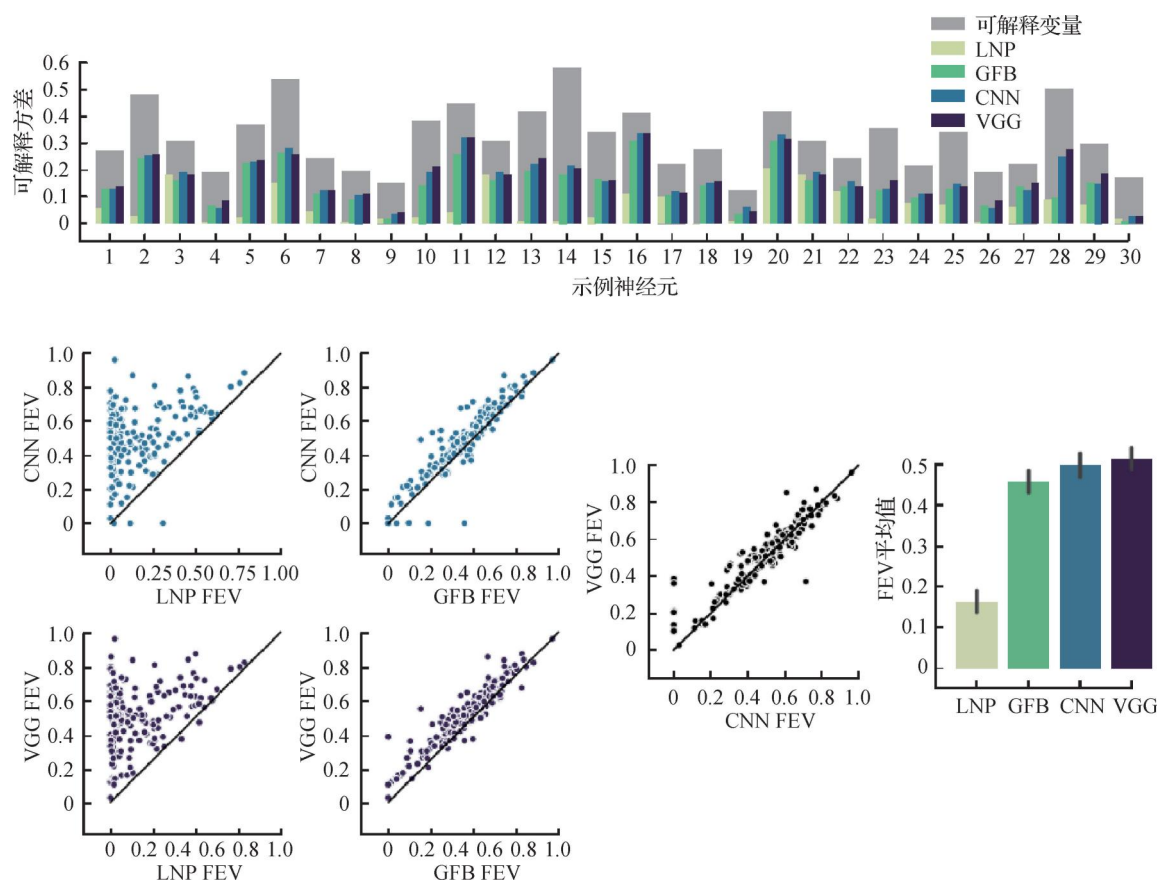


图 9 不同编码模型对真实 V1 神经元的可解释性 (Cadena 等, 2019)

Fig. 9 Interpretability of different encoding models for real V1 neurons (Cadena et al., 2019)

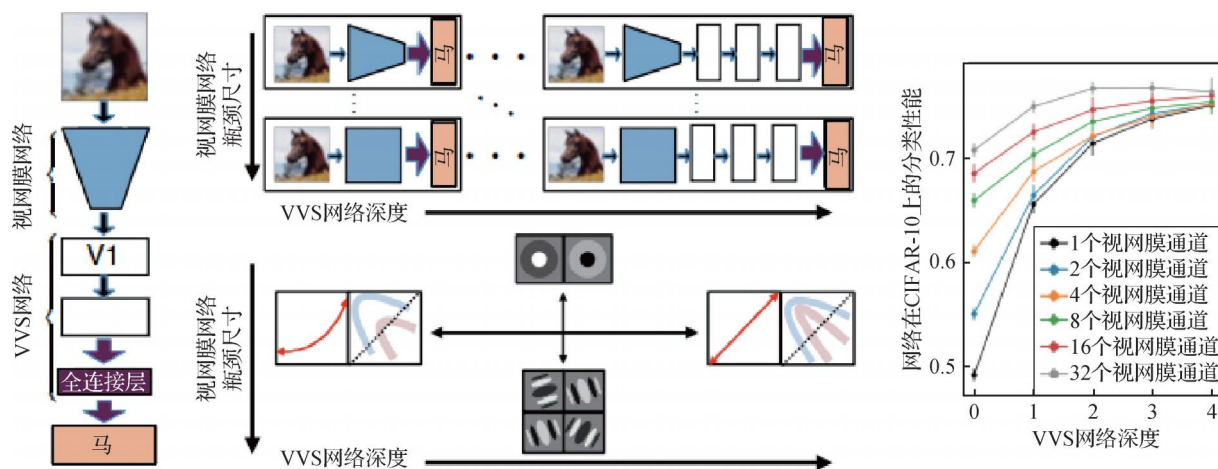


图 10 模拟视网膜与腹侧视觉通路的联合网络结构 (Lindsey 等, 2019)

Fig. 10 Illustration of the framework to model early visual representations (Lindsey et al., 2019)

杂的 VVS 皮层网络, 视网膜表现为视觉场景的简单线性编码器。该结果预测小型脊椎动物 (例如蝾螈、青蛙) 的视网膜应该执行复杂的非线性计算, 提取与行为直接相关的特征。而大型动物 (例如灵长类动物) 的视网膜应该主要对视觉场景进行线性

编码, 并对更广泛的范围做出反应的刺激。这些结果表明, 脊椎动物的视网膜表征取决于分配给其视觉系统的神经资源, 从而决定它们的视网膜表征目标——执行特征提取或对自然场景的高效编码。

3 任务驱动的高级视觉皮层编码模型

除了以电生理驱动的神经网络编码模型,研究人员也采用深度神经网络来研究神经科学的核心问题(Kriegeskorte, 2015; Yamins 和 Dicarlo, 2016)。例如,深度卷积神经网络已用于模拟实现视觉对象识别的腹侧视觉通路(Yamins 等, 2013, 2014; Khaligh-Razavi 和 Kriegeskorte, 2014; Yamins 和 Dicarlo, 2016)。结果证明,物体识别任务驱动训练的深度神经网络中,网络隐层的计算节点可以精确预测神经元的响应。然而,这种网络模型的最终输出代表了在许多层中进行的密集计算,这可能与大脑中信

息处理的生物学基础相关,也可能不相关。鉴于 IT 皮层部分位于人类视觉系统的更高级别,因此很难理解 CNN 的这些网络组件。

通过训练大量自然图片,多层卷积神经网络(hierarchical convolutional neural network, HCNNs)所完成的视觉对象识别可达到人类相当的性能(Zeiler 和 Fergus, 2014; LeCun 等, 2015)。图 11 中展示了 Yamins 和 Dicarlo (2016) 用于建模腹侧视觉通路的 HCNN 模型结构。近期的研究工作表明,基于任务优化的多层卷积神经网络 HCNNs 是灵长类动物大脑视觉编码的精准量化模型(Yamins 等, 2014; Khaligh-Razavi 和 Kriegeskorte, 2014; Güçlü 和 van Gerven, 2015)。

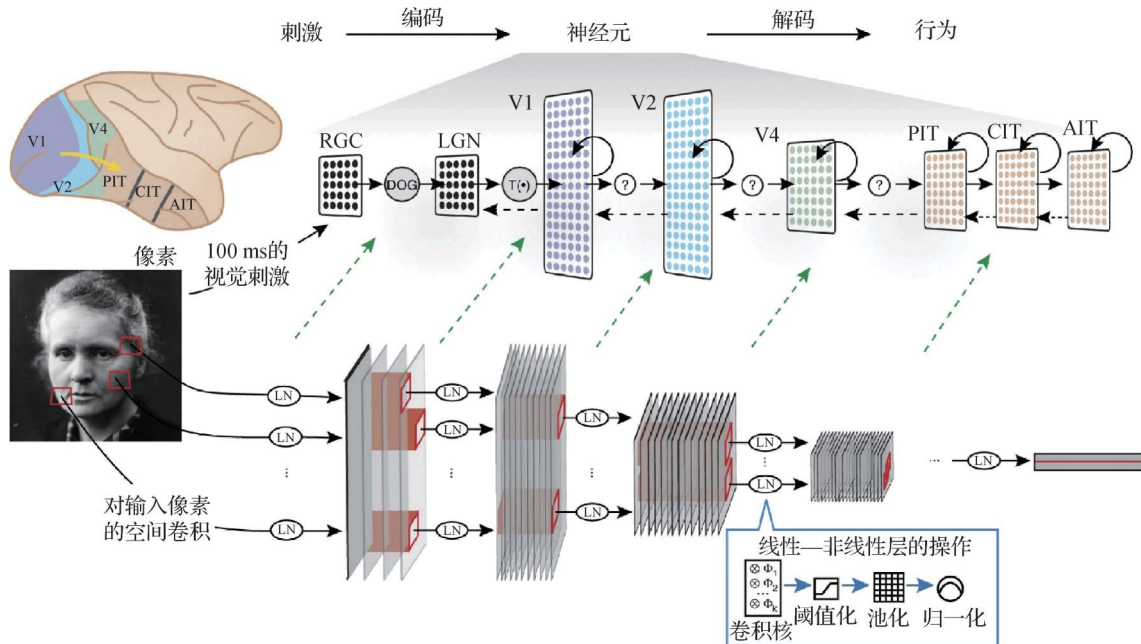


图 11 以物体识别任务学习的 HCNN 编码模型(Yamins 和 Dicarlo, 2016)

Fig. 11 HCNN encoding model learned for object recognition task(Yamins and Dicarlo, 2016)

目前,与其他计算模型类别相比,在 ImageNet 数据集上以物体识别任务训练的 HCNNs 可以更好地预测视觉系统中神经元的平均时间响应(temporally-averaged response)。训练好的网络中底层、中层和高层的卷积层的模型单元分别精准预测初级(Khaligh-Razavi 和 Kriegeskorte, 2014; Cadena 等, 2019)(V1 区)、中级(Yamins 等, 2014)(V4 区)和更高的视觉皮层区域(Khaligh-Razavi 和 Kriegeskorte, 2014; Yamins 等, 2014)(下颞叶皮质,即 IT 区)。这些模型并没有使用真实的生理数据进行拟合,而是直接以高级视觉任务(目标识别)进行训

练,它们能够达到目前视觉信息编码任务的最高性能(state-of-the-art)。如在图 12 中,HCNN 模型可以在获得最佳的物体识别预测性能的同时,精确预测视觉皮层神经元的响应。

在 HCNNs 模型的训练中,选择变化多端、场景比较具有挑战性数据集,以及更加贴近现实生活的视觉任务也很重要,如在种类繁多的 ImageNet 数据集上进行物体识别任务的训练效果会比较好。研究人员(Khaligh-Razavi 和 Kriegeskorte, 2014; Hong 等, 2016)认为使用较简单、种类较少的数据集进行目标识别任务的无监督训练模型(图像自编码器),无

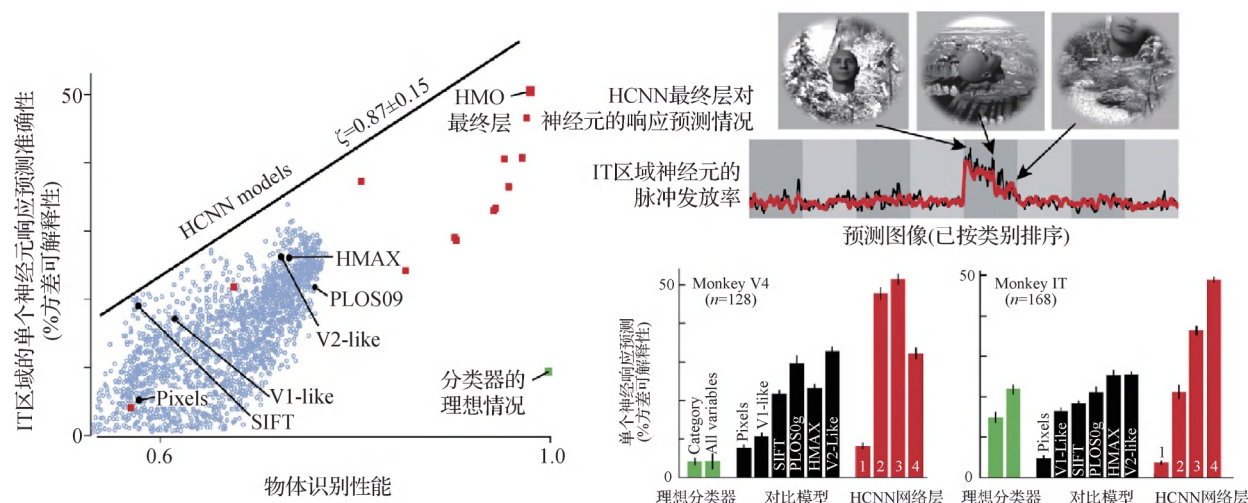


图 12 物体识别任务训练的 HCNN 模型可精确预测高级视觉皮层的神经元响应 (Yamins 和 Dicarlo, 2016)

Fig. 12 Goal-driven optimization yields neutrally predictive models of ventral visual cortex (Yamins and Dicarlo, 2016)

法精确预测神经元对图像的响应,尤其是高级视皮层。这也意味着对灵长类动物来说,捕捉外界环境不计其数的视觉刺激特性对构建一个稳健的视觉系统非常关键。HCNNs 模型的最终输出表示的是先前许多的卷积层中进行的密集的计算,这与大脑中的信息处理的神经科学基础有关与否并不确定,所以需要理解 HCNNs 计算单元与高级视觉神经系统这些组件的对应关系。

虽然 HCNNs 能够在预测视皮层的平均响应上取得傲人的成绩,但灵长类动物的视觉系统的解剖结构中除了现有 HCNNs 中的前馈连接,还有别的连接结构。这些结构包括每个皮层区域内密集的局部循环连接以及不同区域间的远程连接,例如从视觉层次的较高级区域到初级部分的反馈 (Gilbert, 2013)。灵长类动物大脑的视觉系统中循环结构的功能尚未得到充分研究。有研究 (Spoerer 等, 2017; Michaelis 等, 2018; Linsley 等, 2018) 认为循环结构可自动填补缺失数据,如被其他物体遮挡住的物体部分;有研究 (Gilbert, 2013; Lindsay, 2015; McIntosh 等, 2018; Li 等, 2018; Kar 等, 2019) 认为循环结构通过自上而下的注意力特征的细化锐化了表征,以便对特定的刺激因素或特定任务的性能进行解码;有研究 (Rao 和 Ballard, 1999; Lotter 等, 2017; Issa 等, 2018) 认为循环结构允许大脑预测未来的刺激信号 (如电影的帧);而也有研究 (Liao 和 Poggio, 2016; Zamir 等, 2017; Leroux 等, 2018) 认为循环扩展了前馈计算,这意味着展开的循环网络等价于通过多次

重复变换来保存神经元 (和可学习参数) 的更深层前馈网络。

现有的神经生理数据无法排除这些可能性,而计算模型可能有助于对这些假设进行评估。研究人员 (Spoerer 等, 2017; Lotter 等, 2017) 将添加了循环结构的增强 CNN 用于解决相对简单的遮挡变形和预测未来的任务,但这些模型既无法泛化到前馈 CNN 执行的较困难任务 (如识别 ImageNet 数据集中的目标),也无法像在 ImageNet 上优化的 HCNN 一样对神经响应做出解释。在 ImageNet 中进行目标识别是已知仅有可以产生与视觉皮层神经元激活模式相仿的 HCNN 激活模式的任务 (Khaligh-Razavi 和 Kriegeskorte, 2014; Yamins 等, 2014; Cadena 等, 2019)。但事实上,由于多样性和复杂性,ImageNet 中包含许多可以根据上述假设 (例如严重遮挡、出现多个前景目标等) 利用循环结构的图像数据。此外,一些针对 ImageNet 的最有效方法 (如 ResNet 模型 (He 等, 2016)) 是在多个层上重复相同的架构模式,这说明它们可能与较浅的循环网络的展开近似 (Liao 和 Poggio, 2016)。McIntosh 等人 (2018) 将 HCNN 的输出作为 RNN 的输入来解决目标分割等视觉任务。而 Nayebi 等人 (2018) 则认为使用传统人工神经网络中的循环结构 (例如,朴素 RNN、长短期记忆网络 (long short-term memory, LSTM) (Elman, 1990; Hochreiter 和 Schmidhuber, 1997)) 并不能提高计算模型在 ImageNet 上的识别性能,因此他们直接在卷积单元上引入了循环结构,提出了称为 Conv-

RNN(convolutional recurrent neural networks)的模型结构,其结构如图 13 所示。ConvRNN 中主要引入了既带有旁路(bypass),也实现了类似于 LSTM 的长短期记忆功能的门控单元。这种基于目标识别任务训练的模型能够预测高级视皮层区域(V4 和 IT)的神经动力学响应,并且能达到很高的识别准确率。

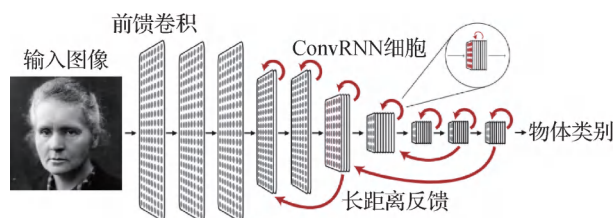


图 13 ConvRNN 模型(Nayebi 等,2018)

Fig. 13 The structure of the ConvRNN model
(Nayebi et al., 2018)

此外, Kar 等人(2019)认为网络层次较多的纯 HCNN 网络与浅层的带循环结构的 CNN 结构功能等价。然而, Spoerer 等人(2017)认为在 HCNNs 网络中加入循环结构后可提高网络训练的收敛速度,即使在与 HCNNs 同等参数量的情况下也能达到较高的目标识别准确率。

4 无监督任务驱动的神编码模型

虽然以物体识别任务有监督地训练网络,可以获得目前最精确的视觉皮层腹侧通路的神经编码模型,但是许多研究者认为,这种有监督式的任务驱动模型不具备生理可解释性。因为大脑需要处理的任务很多,需要在多种任务中都取得较好的性能以生存,而不仅仅是专门处理物体识别等固定任务的网络。例如,目前深度神经网络常用的 ImageNet 数据集中(Krizhevsky 等,2012),含有上百万张带标签数据。若使用同样规模的数据训练一个婴儿学会识别,并且以 5 s 一次的频率指定样本,需要消耗一千多个小时。而人类并不是机器,无法不停地重复训练这一项任务,因此人类显然不是以这种大规模数据样本的形式识别物体。同样,对于其他灵长类动物,这种大规模标签数据的训练模式也显然不可行。因此,寻找更具生物可行性的神经编码模型仍是一个亟须解决的问题。

近年来,许多研究者开始尝试以无监督学习

(Hastie 等,2009)的方式训练深度神经网络编码模型。而目前的无监督网络通常具有以下几种形式:1)无监督学习:旨在表示数据分布或实现数据降维,例如变分自动编码器(variational autoencoder, VAE)(Kingma 和 Welling, 2019);2)自监督学习(self-supervised learning)(LeCun 和 Misra, 2021):旨在通过从输入的任意部分预测其其他部分的任务来找到好的数据表示,例如,对于一幅带遮挡的图像,企图从其不带遮挡的部分预测出缺省的部分;或是从输入序列的过去预测其将来的输入;3)对比学习(contrastive learning):是自监督训练模型的一种特例(Jaiswal 等,2020)。预测样本时只需区分其是正例还是负例,因此称为对比学习。常见的网络有 MoCo(momentum contrast)(Chen 等,2020b), SimCLR(Chen 等,2020a)和 CPC(contrastive predictive coding)(Kharitonov 等,2021)等;4)多模态学习(multimodal learning)(Ramachandram 和 Taylor, 2017):旨在通过相互预测或预测一个公共子空间来寻找不同模态(例如视觉、文本和音频等)的公共子空间。目前常见的多模态学习网络为 OPEN AI 在 2021 年提出的 CLIP(contrastive language-image pre-training)模型(Radford 等,2021)。

Zhuang 等人(2021)最近发现使用无监督或自监督的方式训练网络,可以获得与腹侧视觉通路神经元(例如图 14 中的 V1、V4 和 IT)类似的表达形式。实验结果表明,无监督学习的最新进展在很大程度上弥补了深度网络与腹侧视觉通路的生理可解释差距。并且,发现最近的无监督训练网络,例如 SimCLR 和其他对比学习网络,在腹侧流中的预测精度等于或超过当今最佳的有监督模型。这些结果说明了可用无监督学习来模拟大脑系统,并为生物学上合理的神经编码网络模型提供了强有力的候选者。Konkle 和 Alvarez(2021)也探索了能否用无监督的方式获得与腹侧视觉通路类似的表达。他们的实验结果与 Zhuang 等人(2021)的发现类似,不同的是他们并不是对比单个神经元的表达,而是将模型应用在人脑的功能性磁共振成像(functional magnetic resonance imaging, fMRI)数据上。他们发现模型对 ImageNet 分类准确度与 fMRI 数据匹配度之间没有太大的联系。此外, Konkle 和 Alvarez(2021)还讨论了大脑如何实现自我监督的学习方式。他们认为视网膜扭曲、眼跳、效应复制以及基于海马体的缓

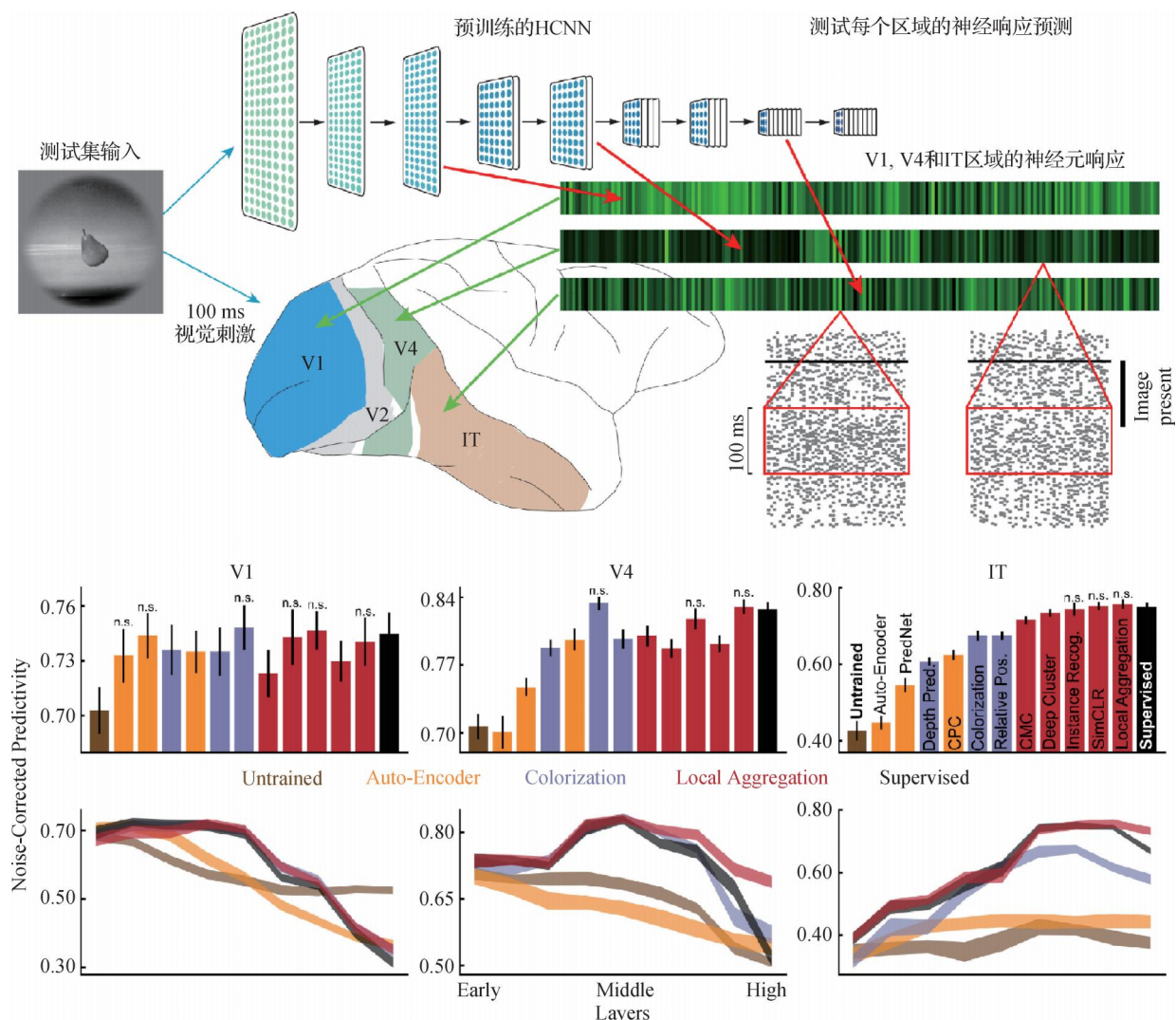


图 14 以物体识别任务无监督学习的 HCNN 编码模型 (Zhuang 等, 2021)

Fig. 14 Unsupervised learning model of the ventral stream (Zhuang et al., 2021)

冲机制是自我监督学习方式的实例化。

除了将无监督方式应用在灵长类动物的腹侧通路建模外,还有一些研究者将无监督方式成功应用在背侧视觉通路、小鼠视觉皮层,及大海马体中概念细胞的建模。例如, Mineault 等人 (2021) 探索了编码运动的神经元特性。他们首先尝试了一些有监督 3D 卷积神经网络来学习背侧通路中不同区域的神经元响应,但是发现这些网络并不能用以解释非人灵长类的单个神经元的响应。然后作者建立了一个智能体,并加入先验——在世界上移动的个体必须根据落在视网膜上的图像模式来确定其自我运动的参数。由此先验限定的网络训练之后的结果从定性和定量上看都很接近背侧视觉通路。而 Bakhtiari 等人 (2021) 使用对比学习网络 CPC (Kharitonov 等, 2021) 同时建模了小鼠的腹侧通路和背侧通路。

Nayebi 等人 (2021) 也探索了小鼠视觉皮层的神经编码过程。使用小鼠视觉皮层对静态图像刺激的数据,并将其与不同结构的监督网络和自我监督网络进行比较。实验结果发现具有平行分支的浅层网络更好地解释了老鼠的数据。这也证实了 Bakhtiari 等人 (2021) 的发现。老鼠的视觉大脑是一种浅层的“通用”视觉机器,可较好地各种任务之间切换。不像灵长类大脑中的深度神经网络高度专一于子任务。这些观点非常符合不同的大脑适应不同的生态环境这一观点。Conwell 等人 (2021) 也将视觉转换器 (transformer)、多层感知器 (multilayer perceptron, MLP) 混合器、任务编码器和自我监督模型应用在小鼠视觉皮层的数据中,实验结果与 Bakhtiari 等人 (2021) 和 Nayebi 等人 (2021) 类似。

Geirhos 等人 (2021) 发现在对图像失真 (旋转、

对比度变化等)的鲁棒性方面,新型的自监督和多模态模型现在已与人类不相上下,例如 CLIP 模型。Choksi 等人(2021)还发现多模态学习模型可以学习出与大脑海马体中概念细胞类似的表达——会对概念或图像的文本表示做出反应。Higgins 等人(2020)发现无监督的深度神经网络可以解耦 IT 皮层中的数据表达,例如,IT 中的一个神经元可能编码了影响面孔外观的多种因素(肤色、亮度和年龄等)。Storrs 等人(2021)发现变分自编码器可以学习出人类感知表面的潜在因素。变分自编码器自然地解开了不同的因素,而且与人类的感知非常吻合。然而,他们发现监督网络在这项任务上的表现不是很理想。

5 结 语

生物视觉研究一直是计算机视觉算法设计的重要灵感来源。有许多计算机视觉算法与生物视觉研究具有不同程度的对应关系,包括从纯粹的功能启发到利用主要用于解释生物观察的模型的方法。从视觉神经科学向计算机视觉界传达的生物视觉处理的经典观点是视觉皮层分层层次处理的结构。

随着新的实验和分析技术的出现,在克服学科交叉研究的技术障碍方面取得了重大进展。脑科学中正在出现大量多尺度功能分析和连接组学信息。令人鼓舞的是,视觉系统的研究在这一快速发展中处于领先地位(None,2013)。例如,现在可以通过结合功能和结构成像来识别选择性神经元群体并在突触水平上剖析它们的电路。应用此类技术的第一系列研究侧重于了解视网膜(Helmstaedter 等,2013)和皮质(Bock 等,2011)水平的视觉回路。在更广泛的范围内,对皮层区域之间连接模式的定量描述现在变得可用,并且视觉皮层网络的研究再次具有开创性(Markov 等,2013)。因此,现在可以使用详细的大规模视觉网络模型来研究多个时间和空间尺度上信息处理的神经生物学基础(Chaudhuri 等,2015;Kim 等,2014;Potjans 和 Diesmann,2014)。随着国际研究计划(例如 BRAIN 和 HBP(human brain project)项目、艾伦研究所)的出现,我们无疑正处于脑科学重大变革的潮流中。与此同时,计算机体系结构的进展也使得模拟大规模模型成为可能。例如,多核架构的出现(Eichner 等,2009)、集群

上的并行计算(Plesser 等,2007)、GPU 计算(Pinto 和 Cox,2012)和神经形态硬件的可用性(Temam 和 Héliot,2011),促进了真正仿生视觉系统的探索(Merolla 等,2014)。然而,计算机和脑科学的这些技术进步需要大力推动理论研究。

各个领域遇到的理论困难需要一种新的跨学科方法来理解如何处理、表示和使用视觉信息。例如,目前尚不清楚皮质区域的密集网络如何充分分析外部世界的结构,部分问题可能来自使用一系列关于中级和高级视觉的框架问题(Cox,2014;Gur,2015;Kubilius 等,2014)。我们不能只过分关注于某项具体的任务(如人脸、物体识别等),而忽视了视觉系统对整个外部世界的认知过程。协调生物视觉和计算机视觉是解决这些挑战的关键。

在技术和工程领域,主要目标是创造能正确预测特定结果的产品,可解释性通常排在第2位(Boon 和 Knuuttila,2009)。虽然有许多研究者认为深度学习仍不具备解释生理的价值,但是他们还是认可其建模的准确性。高预测准确率对建模生物视觉系统也具有科学意义及应用价值(Cichy 和 Kaiser,2019),例如:1)医疗应用:利用 DNN 的预测能力制作神经假体,替换患者受损的视觉皮层或视网膜,模仿视觉皮层的完整神经动力学和物体识别行为(Rajalingham 等,2018;Hong 等,2016;Yu 等,2020);2)大脑的非侵入式实验控制(Yamins 和 Dicarlo,2016):通过使用深度网络合成图像,操纵视觉皮层 V1(Walke 等,2018)和 V4(Bashivan 等,2019)的神经元达到预定的期望状态;3)神经形态视觉芯片的研发:通过模仿生物视觉系统加工处理信息的过程,以获得与神经元类似的脉冲响应,以启发设计具备低时延、高动态范围的神经形态视觉传感器,例如事件相机(Steffen 等,2019)与脉冲相机(Huang 等,2022;Zheng 等,2021b)。

随着学科交叉研究的不断推进,现在已有许多神经计算的研究者用各种深度学习的工具模拟大脑中各个脑区,深度神经网络模型对生物视觉皮层的响应/行为的预测能力可以帮助解释生物视觉系统(Khaligh-Razavi 和 Kriegeskorte,2014;Yamins 等,2014;Cichy 等,2016;Schrumpf 等,2020)。Schrumpf 等人(2020)启动了一个名为 BrainScore 的开源项目,收集和提供了许多视觉皮层神经元对各类视觉刺激的响应。无论是神经计算还是人工神经网络的

研究者,都可以通过上传模型与平台中其他模型对比对神经元的相似性。

BrainScore 提供了一个可量化对比基于神经网络的生物视觉系统编码模型,通过这个排行榜,可以便于了解适用于不同视觉/行为任务的模型,从而找到最精确的模型来用于预测神经元的响应,并应用于医学辅助、神经假体的研制。除了这种实际的经济及工程效益外,模型排行榜的形成也有助于相关领域的研究者,通过对比各种可行的模型与自己提出的模型,揭示何种结构或组件有助于成功预测/建模神经元对视觉刺激的响应。

除了预测能力外,神经网络对认知科学和生物视觉系统的建模同样具备一定的可解释性:1)模型可解释的本质是目的论的(Yamins 等,2014;Marblestone 等,2016):深度神经网络中神经元的表现是为了做出响应后,可以履行它在启用整个系统要完成的特定目标,例如,物体识别。2)网络训练的设置是由先验决定的,且有具体意义的:虽然神经网络模型的训练过程是非透明的,但是其架构和目标函数都涉及具体任务/现象(Cichy 等,2016;Kietzmann 等,2018),与传统的数学理论模型相同。3)精确的神经网络模型参数具有巨大的解释潜力:通过可视化、文本描述或寻找代表实例的方法(Samek 等,2017;Zhou 等,2015;van der Maaten 和 Hinton,2008;Mahendran 和 Vedaldi, 2015;Yosinski 等, 2015;Simonyan 等,2014;Mordvintsev 等,2015;Zhou 等,2019;Girshick 等,2016;Xu 等,2018),可以使得 DNN 模型变得透明且具有生理可解释性(Scholte, 2018)。

自然科学的理想化观点是,从理论中得出假设并在实验中检验。但是,如果缺少成熟且令人信服的理论,就需要为新理论探索创造起点(Steinle, 1997;Burian,1997)。这意味着模型从作为预测工具或类似于解释理论的视角转变为探索新理论的工具(Gelfert,2016)。

观察科学实验表明,探索是一种无所不在的策略。通过构建和操作模型来学习模型,探索它们的行为方式并为观察获得新理论(Kisiel,1973)。神经网络可作为生物视觉系统的探索工具:1)原理验证演示,即通过创建解决问题/完成目标的人工制品来展示特定方案的可行性。例如,第 3 节中在基于对象分类任务训练的前馈神经网络,在特定对象识别

任务上达到了与人类相似的性能水平,并且它们准确地预测了与对象相关的大脑活动(Khaligh-Razavi 和 Kriegeskorte, 2014;Cichy 等,2016;Yamins 等,2014)。网络模型的成功激发了进一步的研究,探索纯自下而上的方法是否可用于解释生物视觉系统。2)在不完全成熟的理论中,实验和理论概念的发展是相互交织进行的(Feest,2012)。理论概念可能根据实验结果进行改进和修改。建模可以具有与实验相同的效果(Sterrett,2014;Waters,2007),它可能改变我们对生物视觉系统的传统概念。Hong 等人(2016)探索了受过对象分类训练的 DNN 如何预测不同的对象属性。正如预期的那样,对象类别在 DNN 的处理层次结构中得到了越来越好的预测。鉴于成功的对象识别需要容忍类别正交属性(例如位置或大小)的变化,假设这种正交对象属性的预测沿着处理层次递减似乎是合理的。令人惊讶的是,作者发现了相反的结果:DNN 预测的类别正交对象属性随着网络的处理层次越来越好。这导致了一个挑战当前物体视觉神经理论的假设:灵长类动物的下颞(IT)皮层(被认为代表跨观察条件的物体类别)也可能代表这种类别的正交属性。模型探索和随后的电生理数据分析证实,研究导致了大脑视觉系统的经典双流假设的重要改进(Hong 等,2016)。

虽然目前有关深度神经网络的理论知识还较为缺乏,但是不可否认,人工神经网络近年来的发展提高了对生物神经元响应的预测能力,并且探索、设计工作方式相仿的人工神经网络可以帮助解释生物视觉系统中结构/组件的工作原理,甚至探索新的生物视觉理论。另一方面,通过结合生物视觉系统的知识,可以帮助了解深度网络的可解释性与意义,帮助促进网络模型的设计。人工神经网络的研发与对大脑功能及结构的探索之间可以相辅相成,为实现新一代的通用人工智能带来更多的启发。

致 谢 本综述的论文整理得到了北京大学视频与视觉技术国家工程研究中心多位成员的帮助,在此表示衷心感谢。

参考文献 (References)

- Antolík J, Hofer S B, Bednar J A and Mrsic-Flogel T D. 2016. Model constrained by visual hierarchy improves prediction of neural responses to natural scenes. *PLoS Computational Biology*, 12 (6): #

- e1004927 [DOI: 10.1371/journal.pcbi.1004927]
- Bakhtiari S, Mineault P, Lillicrap T, Pack C C and Richards B A. 2021. The functional specialization of visual cortex emerges from training parallel pathways with self-supervised predictive learning. *Advances in Neural Information Processing Systems*, 34, 25164-25178.
- Bashivan P, Kar K and DiCarlo J J. 2019. Neural population control via deep image synthesis. *Science*, 364(6439): #eaav9436 [DOI: 10.1126/science.aav9436]
- Batty E, Merel J, Brackbill N, Heitman A, Sher A, Litke A, Chichilnisky E J and Paninski L. 2017. Multilayer recurrent network models of primate retinal ganglion cell responses//*Proceedings of the 5th International Conference on Learning Representations*. Toulon, France: [s. n.].
- Bengio Y. 2009. *Learning Deep Architectures for AI*. Hanover, USA: Now Publishers Inc.
- Bock D D, Lee W C A, Kerlin A M, Andermann M L, Hood G, Wetzel A W, Yurgenson S, Soucy E R, Kim H S and Reid R C. 2011. Network anatomy and in vivo physiology of visual cortical neurons. *Nature*, 471(7337): 177-182 [DOI: 10.1038/nature09802]
- Boon M and Knuuttila T. 2009. Models as epistemic tools in engineering sciences//Meijers A, ed. *Philosophy of Technology and Engineering Sciences*. Amsterdam: Elsevier: 693-726 [DOI: 10.1016/B978-0-444-51667-1.50030-6]
- Boussaoud D, Desimone R and Ungerleider L G. 1991. Visual topography of area TEO in the macaque. *Journal of Comparative Neurology*, 306(4): 554-575 [DOI: 10.1002/cne.903060403]
- Briggs F and Usrey W M. 2008. Emerging views of corticothalamic function. *Current Opinion in Neurobiology*, 18(4): 403-407 [DOI: 10.1016/j.conb.2008.09.002]
- Burian R M. 1997. Exploratory experimentation and the role of histochemical techniques in the work of Jean Brachet, 1938-1952. *History and Philosophy of the Life Sciences*, 19(1): 27-45
- Cadena S A, Denfield G H, Walker E Y, Gatys L A, Tolia A S, Bethge M and Ecker A S. 2019. Deep convolutional models improve predictions of macaque V1 responses to natural images. *PLoS Computational Biology*, 15(4): #e1006897 [DOI: 10.1371/journal.pcbi.1006897]
- Carandini M, Demb J B, Mante V, Tolhurst D J, Dan Y, Olshausen B A, Gallant J L and Rust N C. 2005. Do we know what the early visual system does? *Journal of Neuroscience*, 25(46): 10577-10597 [DOI: 10.1523/JNEUROSCI.3726-05.2005]
- Chaudhuri R, Knoblauch K, Gariel M A, Kennedy H and Wang X J. 2015. A large-scale circuit mechanism for hierarchical dynamical processing in the primate cortex. *Neuron*, 88(2): 419-431 [DOI: 10.1016/j.neuron.2015.09.008]
- Chen T, Kornblith S, Norouzi M and Hinton G. 2020a. A simple framework for contrastive learning of visual representations//*Proceedings of the 37th International Conference on Machine Learning*. Vienna, Austria: JMLR. org: 1597-1607
- Chen X L, Fan H Q, Girshick R and He K M. 2020b. Improved baselines with momentum contrastive learning[EB/OL]. [2020-03-09]. <https://arxiv.org/pdf/2003.04297.pdf>
- Chichilnisky E J. 2001. A simple white noise analysis of neuronal light responses. *Network*, 12(2): 199-213 [DOI: 10.1080/713663221]
- Choksi B, Mozafari M, VanRullen R and Reddy L. 2021. Multimodal neural networks better explain multivoxel patterns in the hippocampus[EB/OL]. [2021-12-11]. <https://arxiv.org/pdf/2201.11517.pdf>
- Cichy R M and Kaiser D. 2019. Deep neural networks as scientific models. *Trends in Cognitive Sciences*, 23(4): 305-317 [DOI: 10.1016/j.tics.2019.01.009]
- Cichy R M, Khosla A, Pantazis D, Torralba A and Oliva A. 2016. Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Scientific Reports*, 6(1): #27755 [DOI: 10.1038/srep27755]
- Conwell C, Mayo D, Buice M A, Katz B, Alvarez G A and Barbu A. 2021. Neural regression, representational similarity, model zoology and neural taskonomy at scale in rodent visual cortex. *Advances in Neural Information Processing Systems*, 34, 5590-5607.
- Cox D D. 2014. Do we understand high-level vision? *Current Opinion in Neurobiology*, 25: 187-193 [DOI: 10.1016/j.conb.2014.01.016]
- Cudeiro J and Sillito A M. 2006. Looking back: corticothalamic feedback and early visual processing. *Trends in Neurosciences*, 29(6): 298-306 [DOI: 10.1016/j.tins.2006.05.002]
- DiCarlo J J, Zoccolan D and Rust N C. 2012. How does the brain solve visual object recognition? *Neuron*, 73(3): 415-434 [DOI: 10.1016/j.neuron.2012.01.010]
- Eichner H, Klug T and Borst A. 2009. Neural simulations on multi-core architectures. *Frontiers in Neuroinformatics*, 3: #21 [DOI: 10.3389/neuro.11.021.2009]
- Elman J L. 1990. Finding structure in time. *Cognitive Science*, 14(2): 179-211 [DOI: 10.1207/s15516709cog1402_1]
- Enroth-Cugell C and Robson J G. 1984. Functional characteristics and diversity of cat retinal ganglion cells. Basic characteristics and quantitative description. *Investigative Ophthalmology and Visual Science*, 25(3): 250-267
- Feest U. 2012. Exploratory experiments, concept formation, and theory construction in psychology//Feest U and Steinle F, eds. *Scientific Concepts and Investigative Practice*. Berlin: Walter de Gruyter, 3: 167-190 [DOI: 10.1515/9783110253610.167]
- Felleman D J and van Essen D C. 1991. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1(1): 1-47 [DOI: 10.1093/CERCOR/1.1.1-A]
- Field G D and Chichilnisky E J. 2007. Information processing in the pri-

- mate retina: circuitry and coding. *Annual Review of Neuroscience*, 30: 1-30 [DOI: 10.1146/annurev.neuro.30.051606.094252]
- Fukushima K and Miyake S. 1982. Neocognitron: a self-organizing neural network model for a mechanism of visual pattern recognition// *Competition and Cooperation in Neural Nets*. Kyoto, Japan: Springer: 267-285 [DOI: 10.1007/978-3-642-46466-9_18]
- Geirhos R, Narayanappa K, Mitzkus B, Thieringer T, Bethge M, Wichmann F A and Brendel W. 2021. Partial success in closing the gap between human and machine vision. *Advances in Neural Information Processing Systems*, 34, 23885-23899
- Gelfert A. 2016. Exploratory uses of scientific models//Gelfert A, ed. *How to Do Science with Models*. Cham: Springer: 71-99 [DOI: 10.1007/978-3-319-27954-1_4]
- Gilbert C D. 2013. The constructive nature of visual processing. *Principles of Neural Science*, 5: 556-576
- Gilbert C D, Hirsch J A and Wiesel T N. 1990. Lateral interactions in visual cortex. *Cold Spring Harbor Symposia on Quantitative Biology*, 55: 663-677 [DOI: 10.1101/sqb.1990.055.01.063]
- Girshick R, Donahue J, Darrell T and Malik J. 2016. Region-based convolutional networks for accurate object detection and segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(1): 142-158 [DOI: 10.1109/TPAMI.2015.2437384]
- Gollisch T and Meister M. 2010. Eye smarter than scientists believed: neural computations in circuits of the retina. *Neuron*, 65(2): 150-164 [DOI: 10.1016/j.neuron.2009.12.009]
- Güçlü U and van Gerven M A J. 2015. Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *Journal of Neuroscience*, 35(27): 10005-10014 [DOI: 10.1523/JNEUROSCI.5023-14.2015]
- Gur M. 2015. Space reconstruction by primary visual cortex activity: a parallel, non-computational mechanism of object representation. *Trends in Neurosciences*, 38(4): 207-216 [DOI: 10.1016/j.tins.2015.02.005]
- Haß J, Blaschke S, Rammsayer T and Hermann J M. 2008. A neuro-computational model for optimal temporal processing. *Journal of Computational Neuroscience*, 25(3): 449-464 [DOI: 10.1007/s10827-008-0088-4]
- Hassabis D, Kumaran D, Summerfield C and Botvinick M. 2017. Neuroscience-inspired artificial intelligence. *Neuron*, 95(2): 245-258 [DOI: 10.1016/j.neuron.2017.06.011]
- Hastie T, Tibshirani R and Friedman J. 2009. *Unsupervised learning//The Elements of Statistical Learning*. 2nd ed. New York, USA: Springer: #485 [DOI: 10.1007/978-0-387-84858-7_14]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Identity mappings in deep residual networks//*Proceedings of the 14th European Conference on Computer Vision*. Amsterdam, the Netherlands: Springer: 630-645 [DOI: 10.1007/978-3-319-46493-0_38]
- Helmstaedter M, Briggman K L, Turaga S C, Jain V, Seung H S and Denk W. 2013. Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature*, 500(7461): 168-174 [DOI: 10.1038/nature12346]
- Higgins I, Chang L, Langston V, Hassabis D, Summerfield C, Tsao D and Botvinick M. 2020. Unsupervised deep learning identifies semantic disentanglement in single inferotemporal neurons [EB/OL]. [2020-06-25]. <https://arxiv.org/pdf/2006.14304.pdf>
- Hochreiter S and Schmidhuber J. 1997. Long short-term memory. *Neural Computation*, 9(8): 1735-1780 [DOI: 10.1162/neco.1997.9.8.1735]
- Hong H, Yamins D L K, Majaj N J and DiCarlo J J. 2016. Explicit information for category-orthogonal object properties increases along the ventral stream. *Nature Neuroscience*, 19(4): 613-622 [DOI: 10.1038/nn.4247]
- Huang T J, Zheng Y J, Yu Z F, Chen R, Li Y, Xiong R Q, Ma L, Zhao J W, Dong S W, Zhu L, Li J N, Jia S S, Fu Y H, Shi B X, Wu S and Tian Y H. 2022. 1 000 × faster camera and machine vision with ordinary devices. *Engineering* [DOI: 10.1016/j.eng.2022.01.012]
- Hubel D H and Wiesel T N. 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160(1): 106-154 [DOI: 10.1113/jphysiol.1962.sp006837]
- Issa E B, Cadieu C F and DiCarlo J J. 2018. Neural dynamics at successive stages of the ventral visual stream are consistent with hierarchical error signals. *Elife*, 7: #e42870 [DOI: 10.7554/eLife.42870]
- Jaiswal A, Babu A R, Zadeh M Z, Banerjee D and Makedon F. 2020. A survey on contrastive self-supervised learning. *Technologies*, 9(1): #2 [DOI: 10.3390/technologies9010002]
- James W, Burkhardt F, Bowers F and Skrupskelis I K. 1890. *The Principles of Psychology*: Vol. 1. London: Macmillan
- Jones H E, Andolina I M, Ahmed B, Shipp S D, Clements J T C, Grieve K L, Cudeiro J, Salt T E and Sillito A M. 2012. Differential feedback modulation of center and surround mechanisms in parvocellular cells in the visual thalamus. *Journal of Neuroscience*, 32(45): 15946-15951 [DOI: 10.1523/JNEUROSCI.0831-12.2012]
- Kar K, Kubilius J, Schmidt K, Issa E B and DiCarlo J J. 2019. Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nature Neuroscience*, 22(6): 974-983 [DOI: 10.1038/s41593-019-0392-5]
- Kastner D B and Baccus S A. 2014. Insights from the retina into the diverse and general computations of adaptation, detection, and prediction. *Current Opinion in Neurobiology*, 25: 63-69 [DOI: 10.1016/j.conb.2013.11.012]
- Khaligh-Razavi S M and Kriegeskorte N. 2014. Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Computational Biology*, 10(11): #e1003915 [DOI: 10.1371/journal.pcbi.1003915]
- Kharitonov E, Rivière M, Synnaeve G, Wolf L, Mazaré P E, Douze M,

- and Dupoux E. 2021. Data augmenting contrastive learning of speech representations in the time domain//2021 IEEE Spoken Language Technology Workshop (SLT). Shenzhen, China; IEEE; 215-222 [DOI: 10.1109/SLT48900.2021.9383605]
- Kietzmann T C, McClure P and Kriegeskorte N. 2018. Deep neural networks in computational neuroscience. *BioRxiv*; #133504 [DOI: 10.1101/133504]
- Kim J S, Greene M J, Zlateski A, Lee K, Richardson M, Turaga S C, Purcaro M, Balkam M, Robinson A, Behabadi B F, Campos M, Denk W and Seung H S. 2014. Space-time wiring specificity supports direction selectivity in the retina. *Nature*, 509(7500): 331-336 [DOI: 10.1038/nature13240]
- Kindel W F, Christensen E D and Zylberberg J. 2017. Using deep learning to reveal the neural code for images in primary visual cortex[EB/OL]. [2022-06-19]. <https://arxiv.org/pdf/1706.06208.pdf>
- Kingma D P and Welling M. 2019. An introduction to variational autoencoders [EB/OL]. [2019-12-11]. <https://arxiv.org/pdf/1906.02691.pdf>
- Kisiel T. 1973. Scientific discovery: logical, psychological, or hermeneutical? //Carr D and Casey E S, eds. *Explorations in Phenomenology*. Dordrecht; Springer; 263-284 [DOI: 10.1007/978-94-010-1999-6_12]
- Klindt D A, Ecker A S, Euler T and Bethge M. 2017. Neural system identification for large populations separating “what” and “where”. *Advances in Neural Information Processing Systems*. Long Beach, USA; Curran Associates Inc.; 3509-3519
- Konkle T and Alvarez G A. 2021. Beyond category-supervision; computational support for domain-general pressures guiding human visual system representation [EB/OL]. [2022-05-08]. <http://biorxiv.org/content/10.1101/2020.06.15.15324703>
- Kriegeskorte N. 2015. Deep neural networks: a new framework for modeling biological vision and brain information processing. *Annual Review of Vision Science*, 1(1): 417-446 [DOI: 10.1146/annurev-vision-082114-035447]
- Krizhevsky A. 2009. Learning Multiple Layers of Features from Tiny Images. University of Toronto
- Krizhevsky A, Sutskever I and Hinton G E. 2012. ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*. Burlington; Morgan Kaufmann Publishers; 1097-1105
- Kubilius J, Wagemans J and Op de Beeck H P. 2014. A conceptual framework of computations in mid-level vision. *Frontiers in Computational Neuroscience*, 8; #158 [DOI: 10.3389/fncom.2014.00158]
- LeCun Y and Bengio Y. 1998. Convolutional networks for images, speech, and time series//Arbib M A, ed. *The Handbook of Brain Theory and Neural Networks*. Cambridge, USA; The MIT Press; 255-258
- LeCun Y and Misra I. 2021. Self-supervised learning; The dark matter of intelligence[EB/OL]. [2022-05-23]. <https://ai.facebook.com/blog/self-supervised-learning-the-dark-matter-of-intelligence>
- LeCun Y, Bengio Y and Hinton G. 2015. Deep learning. *Nature*, 521(7553): 436-444 [DOI: 10.1038/nature14539]
- Leroux S, Molchanov P, Simoens P, Dhoedt B, Breuel T and Kautz J. 2018. IamNN; iterative and adaptive mobile neural network for efficient image classification [EB/OL]. [2018-04-26]. <https://arxiv.org/pdf/1804.10123.pdf>
- Li X, Jie Z Q, Feng J S, Liu C S and Yan S C. 2018. Learning with rethinking: recurrently improving convolutional neural networks through feedback. *Pattern Recognition*, 79: 183-194 [DOI: 10.1016/j.patcog.2018.01.015]
- Liao Q L and Poggio T. 2016. Bridging the gaps between residual learning, recurrent neural networks and visual cortex [EB/OL]. [2016-04-13]. <https://arxiv.org/pdf/1604.03640.pdf>
- Lindsay G W. 2015. Feature-based attention in convolutional neural networks [EB/OL]. [2022-12-09]. <https://arxiv.org/pdf/1511.06408.pdf>
- Lindsey J, Ocko S A, Ganguli S and Deny S. 2019. A unified theory of early visual representations from retina to cortex through anatomically constrained deep CNNs [EB/OL]. [2022-01-03]. <https://arxiv.org/pdf/1901.00945.pdf>
- Linsley D, Kim J, Veerabadran V, Windolf C and Serre T. 2018. Learning long-range spatial dependencies with horizontal gated recurrent units. *Advances in Neural Information Processing Systems*. Montréal, Canada; Curran Associates Inc.; 152-164
- Liu J K and Gollisch T. 2015. Spike-triggered covariance analysis reveals phenomenological diversity of contrast adaptation in the retina. *PLoS Computational Biology*, 11(7): #e1004425 [DOI: 10.1371/journal.pcbi.1004425]
- Liu J K, Schreyer H M, Onken A, Rozenblit F, Khani M H, Krishnamoorthy V, Panzeri S and Gollisch T. 2017. Inference of neuronal functional circuitry with spike-triggered non-negative matrix factorization. *Nature Communications*, 8(1): #149 [DOI: 10.1038/s41467-017-00156-9]
- Lotter W, Kreiman G and Cox D. 2017. Deep predictive coding networks for video prediction and unsupervised learning [EB/OL]. [2022-03-01]. <https://arxiv.org/pdf/1605.08104.pdf>
- Machens C K, Wehr M S and Zador A M. 2004. Linearity of cortical receptive fields measured with natural sounds. *Journal of Neuroscience*, 24(5): 1089-1100 [DOI: 10.1523/JNEUROSCI.4445-03.2004]
- Mahendran A and Vedaldi A. 2015. Understanding deep image representations by inverting them//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA; IEEE; 5188-5196 [DOI: 10.1109/CVPR.2015.7299155]
- Maheswaranathan N, McIntosh L T, Kastner D B, Melander J, Brezovec L, Nayebi A, Wang J L, Ganguli S and Baccus S A. 2018. Deep learning models reveal internal structure and diverse computations in

- the retina under natural scenes. *bioRxiv*; #340943 [DOI: 10.1101/340943]
- Malach R, Levy I and Hasson U. 2002. The topography of high-order human object areas. *Trends in Cognitive Sciences*, 6(4): 176-184 [DOI: 10.1016/S1364-6613(02)01870-3]
- Marblestone A H, Wayne G and Kording K P. 2016. Toward an integration of deep learning and neuroscience. *Frontiers in Computational Neuroscience*, 10; #94 [DOI: 10.3389/fncom.2016.00094]
- Markov N T, Ercsey-Ravasz M, Van Essen D C, Knoblauch K, Toroczkai Z and Kennedy H. 2013. Cortical high-density counterstream architectures. *Science*, 342(6158): #1238406 [DOI: 10.1126/science.1238406]
- Marmarelis P Z and Naka K I. 1972. White-noise analysis of a neuron chain: an application of the Wiener theory. *Science*, 175(4027): 1276-1278 [DOI: 10.1126/science.175.4027.1276]
- McFarland J M, Cui Y W and Butts D A. 2013. Inferring nonlinear neuronal computation based on physiologically plausible inputs. *PLoS Computational Biology*, 9(7): #e1003143 [DOI: 10.1371/journal.pcbi.1003143]
- McIntosh L, Maheswaranathan N, Sussillo D and Shlens J. 2018. Recurrent segmentation for variable computational budgets//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. Salt Lake City, USA: IEEE: 1761-1770 [DOI: 10.1109/CVPRW.2018.00216]
- McIntosh L T, Maheswaranathan N, Nayebi A, Ganguli S and Baccus S A. 2016. Deep learning models of the retinal response to natural scenes. *Advances in Neural Information Processing Systems*. Barcelona, Spain: Curran Associates Inc.: 1369-1377
- Merabet L, Desautels A, Minville K and Casanova C. 1998. Motion integration in a thalamic visual nucleus. *Nature*, 396(6708): 265-268 [DOI: 10.1038/24382]
- Merolla P A, Arthur J V, Alvarez-Icaza R, Cassidy A S, Sawada J, Akopyan F, Jackson B L, Imam N, Guo C, Nakamura Y, Brezzo B, Vo I, Esser S K, Appuswamy R, Taba B, Amir A, Flickner M D, Risk W P, Manohar R and Modha D S. 2014. A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science*, 345(6197): 668-673 [DOI: 10.1126/science.1254642]
- Michaelis C, Bethge M and Ecker A S. 2018. One-shot segmentation in clutter//*Proceedings of the 35th International Conference on Machine Learning*. Stockholm, Sweden: PMLR: 3549-3558
- Milner A D and Goodale M A. 2008. Two visual systems re-viewed. *Neuropsychologia*, 46(3): 774-785 [DOI: 10.1016/j.neuropsychologia.2007.10.005]
- Mineault P J, Bakhtiari S, Richards B A and Pack C C. 2021. Your head is there to move you around: goal-driven models of the primate dorsal pathway. *Advances in Neural Information Processing Systems*, 34, 28757-28771
- Mishkin M, Ungerleider L G and Macko K A. 1983. Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences*, 6: 414-417 [DOI: 10.1016/0166-2236(83)90190-X]
- Mordvintsev A, Olah C and Tyka M. 2015. Inceptionism: going deeper into neural networks. URL <https://research.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html>
- Nandy A S, Sharpee T O, Reynolds J H and Mitchell J F. 2013. The fine structure of shape tuning in area V4. *Neuron*, 78(6): 1102-1115 [DOI: 10.1016/j.neuron.2013.04.016]
- Nayebi A, Bear D, Kubilius J, Kar K, Ganguli S, Sussillo D, DiCarlo J J and Yamins D L K. 2018. Task-driven convolutional recurrent models of the visual system. *Advances in Neural Information Processing Systems*. Montréal, Canada: Curran Associates Inc.: 5295-5306
- Nayebi A, Kong N C L, Zhuang C X, Gardner J L, Norcia A M and Yamins D L K. 2021. Unsupervised models of mouse visual cortex. *bioRxiv* [DOI: 10.1101/2021.06.16.448730]
- None. 2013. Focus on neurotechniques. *Nature Neuroscience*, 16(7): #771 [DOI: 10.1038/nn0713-771]
- O'Connor D H, Fukui M M, Pinsk M A and Kastner S. 2002. Attention modulates responses in the human lateral geniculate nucleus. *Nature Neuroscience*, 5(11): 1203-1209 [DOI: 10.1038/nn957]
- Paninski L. 2003. Convergence properties of some spike-triggered analysis techniques. *Advances in Neural Information Processing Systems*. Cambridge, USA: MIT Press: 189-196
- Pillow J W, Shlens J, Paninski L, Sher A, Litke A M, Chichilnisky E J and Simoncelli E P. 2008. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207): 995-999 [DOI: 10.1038/nature07140]
- Pinto N and Cox D D. 2012. GPU metaprogramming: a case study in biologically inspired machine vision//*GPU Computing Gems Jade Edition*. Amsterdam: Elsevier: 457-471
- Pinto N, Cox D D and DiCarlo J J. 2008. Why is real-world visual object recognition hard? *PLoS Computational Biology*, 4(1): #e27 [DOI: 10.1371/journal.pcbi.0040027]
- Pinto N, Doukhan D, DiCarlo J J and Cox D D. 2009. A high-throughput screening approach to discovering good forms of biologically inspired visual representation. *PLoS Computational Biology*, 5(11): #e1000579 [DOI: 10.1371/journal.pcbi.1000579]
- Plesser H E, Eppler J M, Morrison A, Diesmann M and Gewaltig M O. 2007. Efficient parallel simulation of large-scale neuronal networks on clusters of multiprocessor computers//*Proceedings of the 13th International Euro-Par Conference on Parallel Processing*. Rennes, France: Springer: 672-681 [DOI: 10.1007/978-3-540-74466-5_71]
- Potjans T C and Diesmann M. 2014. The cell-type specific cortical microcircuit: relating structure and activity in a full-scale spiking network model. *Cerebral Cortex*, 24(3): 785-806 [DOI: 10.1093/cercor/bhs358]
- Radford A, Kim J W, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry

- G, Aspell A, Mishkin P, Clark J, Krueger G and Sutskever I. 2021. Learning transferable visual models from natural language supervision [EB/OL]. [2021-02-26]. <https://arxiv.org/pdf/2103.00020.pdf>
- Rajalingham R, Issa E B, Bashivan P, Kar K, Schmidt K and DiCarlo J J. 2018. Large-scale, high-resolution comparison of the core visual object recognition behavior of humans, monkeys, and state-of-the-art deep artificial neural networks. *Journal of Neuroscience*, 38(33): 7255-7269 [DOI: 10.1523/JNEUROSCI.0388-18.2018]
- Ramachandram D and Taylor G W. 2017. Deep multimodal learning: a survey on recent advances and trends. *IEEE Signal Processing Magazine*, 34(6): 96-108 [DOI: 10.1109/MSP.2017.2738401]
- Rao R P N and Ballard D H. 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1): 79-87 [DOI: 10.1038/4580]
- Riesenhuber M and Poggio T. 1999. Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2(11): 1019-1025 [DOI: 10.1038/14819]
- Rolls E T and Milward T. 2000. A model of invariant object recognition in the visual system: learning rules, activation functions, lateral inhibition, and information-based performance measures. *Neural Computation*, 12(11): 2547-2572 [DOI: 10.1162/089976600300014845]
- Rowekamp R J and Sharpee T O. 2017. Cross-orientation suppression in visual area V2. *Nature Communications*, 8: #15739 [DOI: 10.1038/ncomms15739]
- Rucci M and Victor J D. 2015. The unsteady eye: an information-processing stage, not a bug. *Trends in Neurosciences*, 38(4): 195-206 [DOI: 10.1016/j.tins.2015.01.005]
- Sahani M and Linden J F. 2003. How linear are auditory cortical responses? *Advances in Neural Information Processing Systems*. Cambridge, United States: MIT Press: 125-132
- Samek W, Wiegand T and Müller K R. 2017. Explainable artificial intelligence: understanding, visualizing and interpreting deep learning models [EB/OL]. [2022-08-28]. <https://arxiv.org/pdf/1708.08296.pdf>
- Scholtz H S. 2018. Fantastic DNNs and where to find them. *NeuroImage*, 180: 112-113 [DOI: 10.1016/j.neuroimage.2017.12.077]
- Schrimpf M, Kubilius J, Hong H, Majaj N J, Rajalingham R, Issa E B, Kar K, Bashivan P, Prescott-Roy J, Geiger F, Schmidt K, Yamins D L K and DiCarlo J J. 2020. Brain-score: which artificial neural network for object recognition is most brain-like? *BioRxiv*: #407007 [DOI: 10.1101/407007]
- Schwartz O, Pillow J W, Rust N C and Simoncelli E P. 2006. Spike-triggered neural characterization. *Journal of Vision*, 6(4): 484-507 [DOI: 10.1167/6.4.13]
- Serre T, Wolf L, Bileschi S, Riesenhuber M and Poggio T. 2007. Robust object recognition with cortex-like mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(3): 411-426 [DOI: 10.1109/TPAMI.2007.56]
- Simoncelli E P and Olshausen B A. 2001. Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24: 1193-1216 [DOI: 10.1146/annurev.neuro.24.1.1193]
- Simonyan K, Vedaldi A and Zisserman A. 2014. Deep inside convolutional networks: visualising image classification models and saliency maps [EB/OL]. [2022-04-19]. <https://arxiv.org/pdf/1312.6034.pdf>
- Simonyan K and Zisserman A. 2015. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2015-04-10]. <https://arxiv.org/pdf/1406.1556.pdf>
- Souhail S and Cessac B. 2021. On the potential role of lateral connectivity in retinal anticipation. *Journal of Mathematical Neuroscience*, 11: 3 [DOI: 10.1186/s13408-020-00101-z]
- Spoerer C J, McClure P and Kriegeskorte N. 2017. Recurrent convolutional neural networks: a better model of biological object recognition. *Frontiers in Psychology*, 8: #1551 [DOI: 10.3389/fpsyg.2017.01551]
- Steffen L, Reichard D, Weinland J, Kaiser J, Roennau A and Dillmann R. 2019. Neuromorphic stereo vision: a survey of bio-inspired sensors and algorithms. *Frontiers in Neurorobotics*, 13: #28 [DOI: 10.3389/fnbot.2019.00028]
- Steinle F. 1997. Entering new fields: exploratory uses of experimentation. *Philosophy of Science*, 64(S4): S65-S74 [DOI: 10.1086/392587]
- Sterrett S G. 2014. The morals of model-making. *Studies in History and Philosophy of Science Part A*, 46: 31-45 [DOI: 10.1016/j.shpsa.2013.11.006]
- Storrs K R, Anderson B L and Fleming R W. 2021. Unsupervised learning predicts human perception and misperception of gloss. *Nature Human Behaviour*, 5(10): 1402-1417 [DOI: 10.1038/s41562-021-01097-6]
- Temam O and Hélot R. 2011. Implementation of signal processing tasks on neuromorphic hardware//*Proceedings of 2011 International Joint Conference on Neural Networks*. San Jose, USA: IEEE: 1120-1125 [DOI: 10.1109/IJCNN.2011.6033349]
- Ukita J, Yoshida T and Ohki K. 2018. Characterization of nonlinear receptive fields of visual neurons by convolutional neural network. *bioRxiv*: #348060 [DOI: 10.1101/348060]
- Ungerleider L G and Haxby J V. 1994. "What" and "where" in the human brain. *Current Opinion in Neurobiology*, 4(2): 157-165 [DOI: 10.1016/0959-4388(94)90066-3]
- Vance P J, Das G P, Kerr D, Coleman S A, McGinnity T M, Gollisch T and Liu J K. 2018. Bioinspired approach to modeling retinal ganglion cells using system identification techniques. *IEEE Transactions on Neural Networks and Learning Systems*, 29(5): 1796-1808 [DOI: 10.1109/tnnls.2017.2690139]
- van der Maaten L and Hinton G. 2008. Visualizing data using t-SNE.

- Journal of Machine Learning Research, 9: 2579-2605
- Van Essen D C. 2003. Organization of visual areas in macaque and human cerebral cortex//Chalupa L M and Werner J S, eds. The Visual Neurosciences. Cambridge, MA: MIT Press, 1: 507-521
- Vintch B, Movshon J A and Simoncelli E P. 2015. A convolutional sub-unit model for neuronal responses in macaque V1. Journal of Neuroscience, 35 (44): 14829-14841 [DOI: 10.1523/JNEUROSCI.2815-13.2015]
- Walke E Y, Sinz F H, Froudarakis E, Fahey P G, Muhammad T, Ecker A S, Cobos E, Reimer J, Pitkow X and Tolias A S. 2018. Inception in visual cortex; *in vivo-silico* loops reveal most exciting images. bioRxiv: #506956 [DOI: 10.1101/506956]
- Waters C K. 2007. The nature and context of exploratory experimentation; an introduction to three case studies of exploratory research. History and Philosophy of the Life Sciences, 29(3): 275-284
- Whiteway M R, Socha K, Bonin V and Butts D A. 2018. Characterizing the nonlinear structure of shared variability in cortical neuron populations using neural networks. bioRxiv: #407858 [DOI: 10.1101/407858]
- Xu T, Zhan J Y, Garrod O G B, Torr P H S, Zhu S C, Ince R A A and Schyns P G. 2018. Deeper interpretability of deep networks [EB/OL]. [2022-11-20]. <https://arxiv.org/pdf/1811.07807.pdf>
- Yamins D, Hong H, Cadieu C and Dicarlo J J. 2013. Hierarchical modular optimization of convolutional networks achieves representations similar to macaque IT and human ventral stream. Advances in Neural Information Processing Systems. Lake Tahoe, Nevada: Curran Associates Inc. : 3093-3101
- Yamins D L K and Dicarlo J J. 2016. Using goal-driven deep learning models to understand sensory cortex. Nature Neuroscience, 19(3): 356-365 [DOI: 10.1038/nn.4244]
- Yamins D L K, Hong H, Cadieu C F, Solomon E A, Seibert D and DiCarlo J J. 2014. Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proceedings of the National Academy of Sciences of the United States of America, 111 (23): 8619-8624 [DOI: 10.1073/pnas.1403112111]
- Yan Q, Zheng Y J, Jia S S, Zhang Y C, Yu Z F, Chen F, Tian Y H, Huang T J and Liu J K. 2020. Revealing fine structures of the retinal receptive field by deep-learning networks. IEEE Transactions on Cybernetics, 52 (1): 39-50 [DOI: 10.1109/TCYB.2020.2972983]
- Yosinski J, Clune J, Nguyen A, Fuchs T and Lipson H. 2015. Understanding neural networks through deep visualization [EB/OL]. [2022-06-22]. <https://arxiv.org/pdf/1506.06579.pdf>
- Yu Z F, Liu J K, Jia S S, Zhang Y C, Zheng Y J, Tian Y H and Huang T J. 2020. Toward the next generation of retinal neuroprosthesis; visual computation with spikes. Engineering, 6 (4): 449-461 [DOI: 10.1016/j.eng.2020.02.004]
- Zamir A R, Wu T L, Sun L, Shen W B, Shi B E, Malik J and Savarese S. 2017. Feedback networks//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 1808-1817 [DOI: 10.1109/CVPR.2017.196]
- Zeiler M D and Fergus R. 2014. Visualizing and understanding convolutional networks//Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland: Springer: 818-833 [DOI: 10.1007/978-3-319-10590-1_53]
- Zeiler M D, Taylor G W and Fergus R. 2011. Adaptive deconvolutional networks for mid and high level feature learning//Proceedings of 2011 International Conference on Computer Vision. Barcelona, Spain: IEEE: 2018-2025 [DOI: 10.1109/ICCV.2011.6126474]
- Zheng Y J, Jia S S, Yu Z F, Liu J K and Huang T J. 2021a. Unraveling neural coding of dynamic natural visual scenes via convolutional recurrent neural networks. Patterns, 2 (10): #100350 [DOI: 10.1016/j.patter.2021.100350]
- Zheng Y J, Zheng L X, Yu Z F, Shi B X, Tian Y H and Huang T J. 2021b. High-speed image reconstruction through short-term plasticity for spiking cameras//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 6354-6363 [DOI: 10.1109/CVPR46437.2021.00629]
- Zhou B L, Bau D, Oliva A and Torralba A. 2019. Interpreting deep visual representations via network dissection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41 (9): 2131-2145 [DOI: 10.1109/TPAMI.2018.2858759]
- Zhou B L, Khosla A, Lapedriza A, Oliva A and Torralba A. 2015. Object detectors emerge in deep scene CNNs [EB/OL]. [2022-04-15]. <https://arxiv.org/pdf/1412.6856.pdf>
- Zhuang C X, Yan S M, Nayebi A, Schrimpf M, Frank M C, DiCarlo J J and Yamins D L K. 2021. Unsupervised neural network models of the ventral visual stream. Proceedings of the National Academy of Sciences of the United States of America, 118(3): #e2014196118 [DOI: 10.1073/pnas.2014196118]

作者简介

郑雅菁,女,博雅博士后,主要研究方向为类脑视觉计算、计算机视觉。E-mail: yj.zheng@pku.edu.cn

黄铁军,通信作者,男,教授,主要研究方向为视觉信息处理、类脑计算、脉冲视觉。E-mail: tjhuang@pku.edu.cn

余肇飞,男,助理教授,主要研究方向为类脑计算、脉冲神经网络。E-mail: yuzf12@pku.edu.cn