E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035
中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号:TP305　　**文献标识码**:A　　**文章编号**:1006-8961(2023)02-0372-13

论文引用格式:Du C D, Zhou Q Y, Liu C and He H G. 2023. Review of visual neural encoding and decoding methods in fMRI. Journal of Image and Graphics,28(02):0372-0384(杜长德,周琼怡,刘澈,何晖光. 2023. fMRI 的视觉神经信息编解码方法综述. 中国图象图形学报,28(02):0372-0384)[DOI:10.11834/jig.220525]

fMRI 的视觉神经信息编解码方法综述

杜长德[1,2],周琼怡[1,3],刘澈[1,3],何晖光[1,2,3*]

1. 中国科学院自动化研究所脑图谱与类脑智能研究中心,北京　100190;2. 中国科学院自动化研究所多模态人工智能系统全国重点实验室,北京　100190;3. 中国科学院大学人工智能学院,北京　100049

摘　要:视觉神经信息编解码旨在利用功能磁共振成像(functional magnetic resonance imaging, fMRI)等神经影像数据研究视觉刺激与大脑神经活动之间的关系。编码研究可以对神经活动模式进行建模和预测,有助于脑科学与类脑智能的发展;解码研究可以对人的视知觉状态进行解译,能够促进脑机接口领域的发展。因此,基于 fMRI 的视觉神经信息编解码方法研究具有重要的科学意义和工程价值。本文在总结基于 fMRI 的视觉神经信息编解码关键技术与研究进展的基础上,分析现有视觉神经信息编解码方法的局限。在视觉神经信息编码方面,详细介绍了基于群体感受野估计方法的发展过程;在视觉神经信息解码方面,首先,按照任务类型将其划分为语义分类、图像辨识和图像重建 3 个部分,并深入阐述了每个部分的代表性研究工作和所用的方法。特别地,在图像重建部分着重介绍了基于深度生成模型(主要包括变分自编码器和生成对抗网络)的简单图像、人脸图像和复杂自然图像的重建技术。其次,统计整理了该领域常用的 10 个开源数据集,并对数据集的样本规模、被试个数、刺激类型、研究用途及下载地址进行了详细归纳。最后,详细介绍了视觉神经信息编解码模型常用的度量指标,分析了当前视觉神经信息编码和解码方法的不足,提出可行的改进意见,并对未来发展方向进行展望。

关键词:神经编码;神经解码;图像重建;视觉认知计算;深度学习;脑机接口(BCI)

Review of visual neural encoding and decoding methods in fMRI

Du Changde[1,2], Zhou Qiongyi[1,3], Liu Che[1,3], He Huiguang[1,2,3*]

1. Research Center for Brain Mapping and Brain-Inspired Intelligence, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China; 2. State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China; 3. School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: The relationship between human visual experience and evoked neural activity is central to the field of computational neuroscience. The purpose of visual neural encoding and decoding is to study the relationship between visual stimuli and the evoked neural activity by using neuroimaging data such as functional magnetic resonance imaging (fMRI). Neural encoding researches attempt to predict the brain activity according to the presented external stimuli, which contributes to the development of brain science and brain-like artificial intelligence. Neural decoding researches attempt to predict the infor-

收稿日期:2022-05-26;**修回日期**:2022-07-17;**预印本日期**:2022-07-24

*通信作者:何晖光　huiguang.he@ia.ac.cn

基金项目:国家自然科学基金项目(62206284,61976209,62020106015);中国人工智能学会-华为 MindSpore 学术奖励基金;北京市自然科学基金项目(J210010,7222311)

Supported by: National Natural Science Foundation of China (62206284,61976209,62020106015); CAAI-Huawei MindSpore Open Fund; Beijing Municipal Natural Science Foundation (J210010,7222311)

mation about external stimuli by analyzing the observed brain activities, which can interpret the state of human visual perception and promote the development of brain computer interface (BCI). Therefore, fMRI based visual neural encoding and decoding researches have important scientific significance and engineering value. Typically, the encoding models are based on the specific computations that are thought to underlie the observed brain responses for specific visual stimuli. Early studies of visual neural encoding relied heavily on Gabor wavelet features because these features are very good at modeling brain responses in the primary visual cortex. Recently, given the success of deep neural networks (DNNs) in classifying objects in natural images, the representations within these networks have been used to build encoding models of cortical responses to complex visual stimuli. Most of the existing decoding studies are based on multi-voxel pattern analysis (MVPA) method, but brain connectivity pattern is also a key feature of the brain state and can be used for brain decoding. Although recent studies have demonstrated the feasibility of decoding the identity of binary contrast patterns, handwritten characters, human facial images, natural picture/video stimuli and dreams from the corresponding brain activation patterns, the accurate reconstruction of the visual stimuli from fMRI still lacks adequate examination and requires plenty of efforts to improve. On the basis of summarizing the key technologies and research progress of fMRI based visual neural encoding and decoding, this paper further analyzes the limitations of existing visual neural encoding and decoding methods. In terms of visual neural encoding, the development process of population receptive field (pRF) estimation method is introduced in detail. In terms of visual neural decoding, it is divided into semantic classification, image identification and image reconstruction according to task types, and the representative research work of each part and the methods used are described in detail. From the perspective of machine learning, semantic classification is a single label or multi-label classification problem. Simple visual stimuli only contain a single object, while natural visual stimuli often contain multiple semantic labels. For example, an image may contain flowers, water, trees, cars, etc. Predicting one or more semantic labels of the visual stimulus from the brain signal is called semantic decoding. Image retrieval based on brain signal is also a common visual decoding task where the model is created to "decode" neural activity by retrieving a picture of what a person has just seen or imagined. In particular, the reconstruction techniques of simple image, face image and complex natural image based on deep generative models (including variational auto-encoders (VAEs) and generative adversarial networks (GANs)) are introduced in the part of image reconstruction. Secondly, 10 open source datasets commonly used in this field were statistically sorted out, and the sample size, number of subjects, types of stimuli, research purposes and download links of the datasets were summarized in detail. These datasets have made important contributions to the development of this field. Finally, we introduce the commonly used measurement metrics of visual neural encoding and decoding model in detail, analyze the shortcomings of current visual neural encoding and decoding methods, propose feasible suggestions for improvement, and show the future development directions. Specifically, for neural encoding, the existing methods still have the following shortcomings: 1) the computational models are mostly based on the existing neural network architecture, which cannot reflect the real biological visual information flow; 2) due to the selective attention of each person in the visual perception and the inevitable noise in the fMRI data collection, individual differences are significant; 3) the sample size of the existing fMRI data set is insufficient; 4) most researchers construct feature spaces of neural encoding models based on fixed types of pre-trained neural networks (such as AlexNet), causing problems such as insufficient diversity of visual features. On the other hand, although the existing visual neural decoding methods perform well in the semantic classification and image identification tasks, it is still very difficult to establish an accurate mapping between visual stimuli and visual neural signals, and the results of image reconstruction are often blurry and lack of clear semantics. Moreover, most of the existing visual neural decoding methods are based on linear transformation or deep network transformation of visual images, lacking exploration of new visual features. Factors that hinder researchers from effectively decoding visual information and reconstructing images or videos mainly include high dimension of fMRI data, small sample size and serious noise. In the future, more advanced artificial intelligence technology should be used to develop more effective methods of neural encoding and decoding, and try to translate brain signals into images, video, voice, text and other multimedia content, so as to achieve more BCI applications. The significant research directions include 1) multi-modal neural encoding and decoding based on the union of image and text; 2) brain-guided computer vision model training and enhancement; 3) visual neural encoding and decoding based on the high efficient features of large-scale pre-trained models. In addition, since brain signals are characterized by com-

plexity, high dimension, large individual diversity, high dynamic nature and small sample size, future research needs to combine computational neuroscience and artificial intelligence theories to develop visual neural encoding and decoding methods with higher robustness, adaptability and interpretability.

Key words: neural encoding; neural decoding; image reconstruction; visual cognitive computing; deep learning; brain computer interface (BCI)

0 引 言

视觉是人类感知和理解外部世界的最重要途径之一。视觉系统作为人类和外部世界进行交互的桥梁,能够稳定、高效、鲁棒地处理复杂的视觉刺激信息,具有当前计算机视觉所无法比拟的优越性。基于功能磁共振成像(functional magnetic resonance imaging, fMRI)等神经影像的视觉神经信息编解码是理解、破译和模拟大脑视觉系统运作机制的重要研究途径,对于类脑智能技术的发展具有重要意义。

视觉神经信息编码以大脑视觉感知机制为基础,通过建立大脑视觉信息处理的计算模型来描述大脑对外界刺激的响应过程,以实现对大脑活动的预测。其中计算模型的输入是图像刺激,输出为大脑对图像刺激的响应。研究视觉神经信息编码,对于探索大脑视觉信息加工机制,提高人工视觉模型的感知和认知能力具有重要意义。与此相反,视觉神经信息解码则主要通过分析脑信号数据,从中找到其与外界视觉刺激的对应关系,实现利用脑信号对外界视觉刺激进行分类、辨识或重建(参见图 1)。

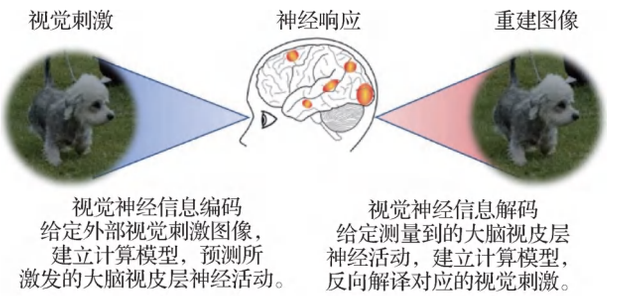

图 1 视觉神经信息编码与解码

Fig. 1 Visual neural encoding and decoding

视觉神经信息编解码模型是建立在统计机器学习框架上的(参见图 2)。基于 fMRI 数据的视觉神经信息编解码通过测量被试大脑的血氧水平依赖性(blood-oxygen-level-dependent, BOLD)信号,得到一系列 3 维脑图像,每个 3 维脑图像都包含上万个体素(对应于 2 维图像中的像素),每个体素的信号对应于该区域内神经元活动所引起的 BOLD 信号。在视觉神经信息解码研究中,多体素模式分析(multi-voxel pattern analysis, MVPA)方法将大脑中的多体素激活模式看做高维空间(不同体素的响应代表不同的维度)中的一个样本点,利用统计机器学习算法解码多体素激活模式中所蕴含的信息。在统计机器学习中,估计条件概率 $P(\boldsymbol{y}\mid\boldsymbol{x})$ 的模型是判别式模型,而估计联合分布 $P(\boldsymbol{x},\boldsymbol{y})$ 的模型是生成式模型。被训练用来从一个变量 $\boldsymbol{x}$ 预测另一个变量 $\boldsymbol{y}$ 的模型是有监督模型,而估计单个变量分布的模型是无监督模型。现有的视觉神经信息编解码研究工作大都可以归纳到上述范畴中的一种。

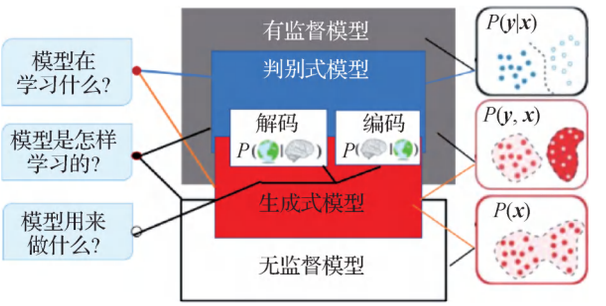

图 2 视觉神经信息编解码与统计机器学习方法间的关系

Fig. 2 Relationships between encoding and decoding of visual neural information and statistical machine learning methods

无论是视觉神经信息编码还是解码研究都依赖于“视觉刺激—大脑响应”成对数据。现有方法大都是基于数据驱动的原理,通过训练数学模型来拟合成对数据的输入和输出之间的关系来确定的。下面从视觉神经信息编码、视觉神经信息解码、公开数据集、度量指标、挑战以及未来展望等角度对该领域进行综述。

1 视觉神经信息编码

随着对神经元信息编码机制研究的不断深入,以视觉感知为目的、信息处理为过程的计算模型,推

动了人们对大脑视觉系统的探索与认知。一方面，研究视觉系统结构特性以及视觉皮层编码特性，探索神经机制在不同层次上的运作规律，将有助于揭示大脑的运作机理，特别是大脑感知、学习和记忆等高级功能；另一方面，视觉系统高效、鲁棒地处理外界信息的能力，启发了信息科学的研究。模拟视觉系统结构和视觉皮层编码特性，开发新一代类脑计算模型，将大幅度提高机器智能处理视觉信息的能力。因此，视觉信息编码机制的研究不仅是脑科学、神经科学的研究热点，也逐渐成为信息科学领域的一个重要研究方向。

近年来，图像物体识别方面的突破进展，进一步激发了人们研究深度学习的热情。神经网络因其强大的特征提取能力和非线性函数刻画能力，已经在计算机视觉、自然语言处理等领域取得了飞跃性的进展。尤其在视觉信息处理领域，神经网络层次化的网络结构、神经元感受野的逐层增加机制，都与视觉系统编码特性有很强的相似性。深度神经网络应用于视觉信息编码模型中，有望更加精细地刻画大脑视觉皮层的信息处理过程，进一步促进现阶段视觉信息编码模型的研究和揭示大脑视觉系统的运作规律。同时，随着脑成像技术的不断进步，fMRI 凭借着极高的空间分辨率和良好的时间分辨率，成为观测大脑活动的主要工具，推动了视皮层信息处理研究的发展。

基于 fMRI 的视觉信息编码模型是描述大脑对于外界视觉刺激如何响应的计算模型（如图 3 所示）。视觉神经编码研究主要由 4 个部分构成（Naselaris 等，2011）。第 1 部分是视觉刺激。这些刺激或是离散的不同类别的刺激图像，或是连续的自然图像，亦可为其他形式的刺激。第 2 部分是刺激的特征提取。不同的特征反映不同的图像所包含的视觉信息，可以通过卷积神经网络、Gabor 滤波器等方法提取。第 3 部分是大脑中的感兴趣区域。这些区域中包含参与建立模型的体素，能够产生对刺激图像的响应。第 4 部分是估计模型参数的算法，通常是采用线性回归。完整的视觉信息编码模型的计算过程主要由两个映射构成：第 1 个映射是从刺激空间到特征空间的映射，是非线性映射；第 2 个映射是特征空间到体素空间的映射，通常是线性映射。

视觉信息编码模型研究方式主要有两种：体素感受野模型（receptive field，RF）（Kay 等，2008）和

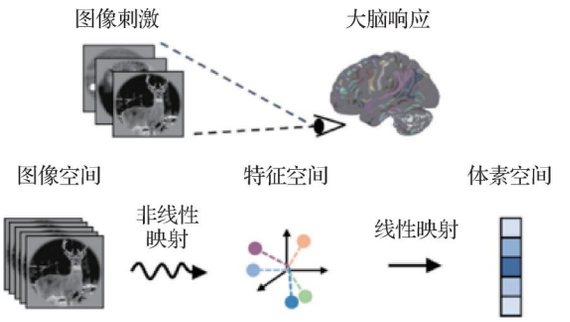

图 3　基于 fMRI 的视觉神经信息编码模型

Fig. 3　Visual neural encoding based on fMRI

表征相似性分析方法（representational similarity analysis，RSA）（Kriegeskorte 等，2008）。体素感受野模型基于体素构建视觉信息编码模型，代表性工作是 Kay 等人（2008）的工作，利用金字塔结构的 Gabor 基函数来模拟视觉区域简单细胞的感受野，再经过投影计算得到 Gabor 感受野编码模型。该模型成功模拟了从图像到脑激活模式的构建过程，实现了对视觉刺激图像的预测。表征相似性分析方法是一种高阶数据分析方法，能够实现不同模态数据之间的比较分析（Kriegeskorte 等，2008）。其核心为表征差异性矩阵（representational dissimilarity matrix，RDM）。利用表征差异性矩阵作为某个脑区对视觉刺激响应的标签，通过计算不同矩阵之间的关系可以分析大脑的不同脑区对于相同视觉刺激的不同激活模型，亦可分析某个脑区的不同计算模型之间的差异性。

由于 fMRI 的体素信号测量的是许多神经元的汇集响应，因此这些模型通常又称为群体感受野（population receptive field，pRF）模型。如图 4 所示，我们重点关注这个特别活跃的研究领域，即使用 pRF 建立视觉神经信息编码模型，来描述人类视觉皮层在各种任务中对一系列刺激的响应过程。下面主要介绍几个经典建模方法。

1）OG 模型。2008 年，Dumoulin 和 Wandell（2008）提出了单高斯（one Gaussian，OG）pRF 估计方法，这种方法第一次利用 fMRI 来构建视皮层群体感受野模型，定量地测量了人类视觉皮层群体感受野的属性。实验者测量了一系列视野位置处由显式对比度定义的环和楔的响应，用来估计每个体素产生最大 fMRI 响应的视野位置。该模型将感受野形状在视野中建模为圆对称（各向同性）高斯模型，组合刺激位置进而预测 fMRI 响应。这种 pRF 估计

方法将毫米级的 fMRI 测量与微米尺度的神经元特性联系起来,减少了功能磁共振成像和电生理学之间的差距。

2) DOG 模型。2012 年, Zuiderbaan 等人(2012)在 pRF 结构上做出了进一步的拓展,将原先的二元单高斯模型拓展成二元双高斯模型。由于经典感受野区域的刺激响应能够被在其他感受野区域的刺激所抑制,会降低功能磁共振幅度(通常称降低到基线以下的幅度为“负面”响应或抑制)。因此,Zuiderbaan 等人(2012)设计出了双高斯模型,使用圆对称高斯差(difference-of-Gaussians, DOG)函数,即通过在 OG 模型上增加一个负高斯函数,可以允许 pRF 分析捕获低于基线和环绕抑制的 fMRI 信号。考虑了环绕抑制的 DOG 模型显示出了 fMRI 数据拟合上的进步,在生物学上更合理地表征了 pRF 结构,进一步提高了视觉信息编码的性能。

3) Topography 模型。2013 年,Lee 等人(2013)提出了一种新的数据驱动方法来估计 pRF 的结构。pRF 结构建模为结构向量,对于每个体素,可以通过求解一系列线性模型从 fMRI 时间序列中来估计结构向量。估计完整的 pRF 拓扑结构后,通过设计阈值,模型选择了中心区域内的结构向量,然后再用二元高斯函数去拟合,得出最佳中心位置。这种方法没有对具体的 pRF 形状做出先验假设,因此是揭示不同空间位置的潜在 pRF 结构的有用工具。由于该模型先估计 pRF 的拓扑结构,所以它可以更好地优化模型中 pRF 中心落在刺激空间之外的体素。

4) 贝叶斯 pRF 模型。2018 年,Zeidman 等人(2018)提出了贝叶斯群体感受野估计模型,提供了一个处理任意维度刺激的通用框架。这个框架将感受野模型建为一个多元正态分布,对于分布中的参数进行约束是通过给每个参数引入给定先验分布的潜在变量。仅对参数空间的采样,就可以估计出先验条件下的 pRF。随着观测数据的进入,参数先验概率密度的估计转化为后验概率,样本信息逐渐修正了参数的初始估计值。贝叶斯方法的关键优点是估计每个参数的不确定性(方差)以及参数之间的协方差。贝叶斯群体感受野模型不会对 pRF 参数施加强烈的先验,而是将它们保持在合理的范围内。这种合理的约束使其能够调整参数和演变来验证两个重要问题:1)体素的反应是用单一多元正态分布(OG)还是具有兴奋中心和抑制环绕的高斯差模型(DOG);2)感受野的形状是圆形、椭圆形还是旋转的椭圆形。

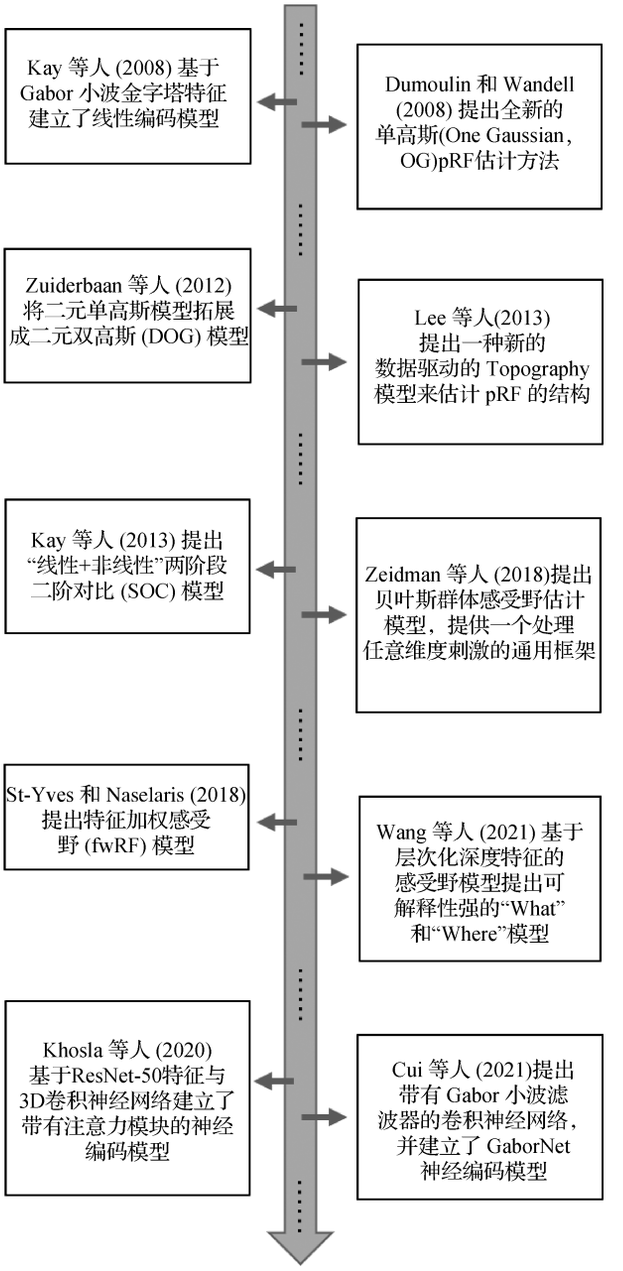

图 4 基于 fMRI 的视觉神经信息编码发展历程

Fig. 4 Development of fMRI-based visual neural encoding

5) fwRF 模型。2018 年, St-Yves 和 Naselaris (2018)提出了特征加权感受野模型(feature-weighted receptive field, fwRF),这个模型有 3 个成分:特征图、权重向量和特征采样区域。关键假设是每个体素的活跃都跨越多个特征图来在空间局部区域编码差异,对所有特征层而言这个区域是固定的。这个模型构建各向同性 2 维高斯的特征采样区域,在自然图像形成的每个特征图中采样整合,通过采用

最小误差平方和函数,对每个体素最优化特征图层权重和高斯采样区域的参数,形成最合适的编码模型。该模型利用 pRF 的大小和特征采样区域的大小之间的关系,刻画了初级视觉皮层中群体感受野分布。当这种方法应用于具有数千个特征图层的深度神经网络时,所得到的编码模型在视觉系统中大多数体素的预测精度高于相比的编码模型。

6)“What”&“Where”模型。2020 年,基于层次化深度特征的感受野模型研究深度学习和大脑视觉通路之间的关系,Wang 等人(2021)开发了可解释性强的“What”和“Where”编码模型,从两个角度对大脑视觉通路进行解释。“What”指的是研究大脑视觉处理通道会产生什么类型的特征,“Where”指的是研究大脑神经元的群体感受野位置在哪。为了自动学习到每个体素的感受野位置及形状,作者使用了带有拉普拉斯正则化约束的稀疏线性回归模型。模型的输入是多层次深度特征的加权组合,拟合目标是各个体素的响应信号。层次化特征的组合系数即可反映体素编码过程中的“What”信息。最终,编码训练使得各个脑区的每个体素会对层次化的深度特征产生选择效应,根据此体素对深度特征所在层的选择倾向和体素所属的感兴趣区域(regions of interest,ROI),可以定量分析深度神经网络和大脑视觉通路之间的对应关系。该模型不仅有效利用了深度神经网络的层次化表征而且符合神经科学对大脑视觉通路研究的基本结论,具有较好的编码效果和可解释性。

2　视觉神经信息解码

近年来,随着人工智能技术的不断进步,基于 fMRI 的神经信息解码研究也得到了快速发展。如图 5 所示,目前国内外已经有很多视觉神经信息解码方面的研究,涵盖了对初级视觉特征(方向、对比度)、中级视觉特征(轮廓)以及高级视觉特征(语义)的分类、辨识和重建。

2.1　基于脑信号解码的语义分类

基于功能磁共振成像的多标签语义解码是一项有挑战性的任务,具有重要的科学意义和应用价值。从机器学习方法角度来讲,语义解码是一个单标签或者多标签(multi-label learning, MLL)的分类问题。简单的图像刺激只包含单个物体,复杂的图像刺激中往往含有多个语义标签,如一幅图像中可能同时含有花、水、树木和汽车等。根据大脑信号,预测出图像刺激的一个或多个标签即为语义解码。早在 2001 年,Haxby 等人(2001)利用 MVPA 方法成功实现了根据 fMRI 信号对呈现给被试者的 8 个不同类别的图像进行分类。Kamitani 和 Tong(2005)将不同方向的条纹作为视觉刺激,根据 fMRI 信号实现了对不同条纹刺激的分类,也证明了在初级视觉区域含有外界图像刺激的信息。Norman 等人(2006)通过给被试者观看不同种类物品,采集相应的任务态 fMRI 数据,并使用其训练支持向量机(support vector machine, SVM)模型,用于最终的分类任务。SVM 在解决小样本、非线性和高维模式识别问题方面有明显的优势。Schmah 等人(2008)基于 fMRI 数据首次使用受限玻尔兹曼机(restricted Boltzmann machine, RBM)实现了大脑状态的解码。Huth 等人(2012)研究了大脑对于动态视觉刺激中的 1 000 多种物体和行为类别的语义表征空间,探索了语义在大脑皮层上的分布地图。Stansbury 等人(2013)使用基于隐含狄利克雷分配(latent Dirichlet allocation, LDA)的神经编解码方法研究人脑如何聚合有关对象的信息来表示场景类别。Huth 等人(2016)利用层次逻辑回归模型以及 WordNet 数据库成功地从大脑信号中解码出了动态视觉刺激中包含的多种语义及语义之间的关联信息。

上述语义解码工作大多是基于单标签的,语义信息单一,而现实世界的视觉刺激往往包含多个物体的信息。研究人脑对于复杂视觉刺激的感知和解码,尤其是考察大脑如何对多个物体进行同时表征对于研究人脑视觉加工机制具有重要意义。Li 等人(2018)提出了多标签语义解码方法,用于从大脑信号中解读多种共生的语义。此外,针对单被试小样本的问题,Li 等人(2021)还提出了基于多模态对抗学习的多被试数据增广方法。该方法基于子空间和多个生成对抗网络将目标被试少量数据与其他被试数据相结合,有效地克服了单被试样本数目少和多被试差异大的问题,提高了单被试的解码精度。在多标签语义解码研究中,现有方法主要集中在标签学习上,忽略了样本本身所包含的信息量,尤其是脑数据,从而限制了方法的性能。另外,大脑信号的多标签标注也是一项费时费力的工作。针对这些问题,Li 等人(2022)提出了一种基于模态辅助和协同

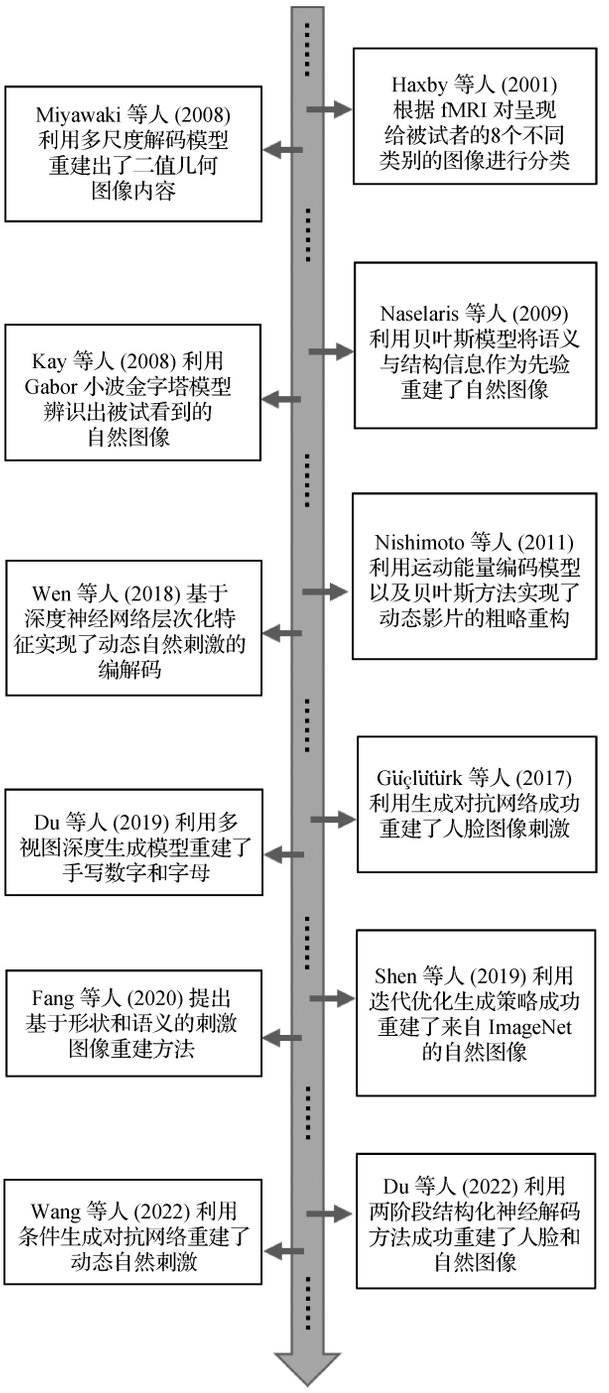

图 5　基于 fMRI 的视觉神经信息解码发展历程

Fig. 5　Development of fMRI-based visual neural decoding

训练的半监督多标签神经解码方法。该方法利用成对有标签的图像模态和脑信号模态(非图像模态)以及大量的非成对无标签的互联网图像数据进行多标签识别,从而利用图像模态来辅助脑信号模态进行多标签学习。

2.2　基于脑信号解码的图像检索

基于脑信号解码的图像检索也是一类常见的视觉解码任务。简单地说,就是利用创建的模型和fMRI,通过解码模型,检索出一个人刚刚所看到的图片。主要有两种思路解决这一问题。

2.2.1　基于编码模型的图像检索

Kay 等人(2008)首先建立神经编码模型,然后将大量候选图像依次输入到神经编码模型,根据编码模型的预测结果和待解码脑信号之间的相关性来确定哪幅图像更有可能是激发该脑信号的刺激。首先给被试者看一千余幅图像,记录他们每一次的磁共振功能成像,然后从这一千余次图像和脑信号的成对数据中估计出一套比较普适的规律,这一步叫做模型估计。接下来就要将这套规律运用于全新的一套图像上,预测出被试者看到这其中每幅新图像的大脑反应是什么样子。当被试者看到一幅新图像,测试者并不知道是哪一幅,但是可以把受试者脑信号的记录与之前的预测相比较,选取预测值与本次实测值最相近的一幅图像,也就是“推测”被试者所看到的究竟是哪一幅图像。

2.2.2　基于解码模型的图像检索

Horikawa 和 Kamitani(2017)首先根据图像特征—大脑信号成对数据集训练一个特征解码器,这可以将脑信号转换为图像特征,如采用卷积神经网络提取的图像特征,然后根据解码得到的特征与候选图像的特征做一一匹配,根据相关性大小返回最匹配的图像。

2.3　基于脑信号解码的图像重建

基于脑信号解码的图像重建,也就是像素级神经解码问题,是解码研究中最难的一种。目前的研究在简单的字母或数字图像上的重建效果较好,在复杂的自然场景图像的精确重建方面仍然非常困难,重建效果还有很大提升空间。解决视觉重建问题的思路有两个关键步骤:首先利用 BOLD 信号变换得到一个特征表示,然后利用该特征表示通过图像生成网络进行图像像素的预测,可以分两阶段完成,也可以端到端的方式完成。视觉刺激重建算法在技术上涉及机器学习中的自编码器(auto-encoder)架构(基于 MLP (multilayer perceptron)、CNN(convolutional neural network)或 VAE(variational auto-encoder)等)以及利用对抗学习进行图像生成方面的研究(如 GAN 等)。由于视觉重建问题本质上是建立两个空间的映射关系,所以也可以利用 CCA (canonical correlation analysis)、多视图学习 (multi-

view learning, MVL) 等方法。总之,这里要建立大脑体素和图像像素之间的映射关系,如何建模这个映射才更好更有效,是研究人员要探索的核心问题。

对于图像重建任务,Haynes 和 Rees(2006)的研究表明可以利用大脑视觉区域的信号重建出被试实际观察到的刺激图像。Miyawaki 等人(2008)采用了多尺度的思想与多变量重构方法,实现了对二值几何图像视觉刺激的重构。Naselaris 等人(2009)利用贝叶斯模型,将语义信息与结构信息作为模型的先验信息,较好地重构了自然视觉刺激图像。Nishimoto 等人(2011)利用运动能量编码模型以及贝叶斯模型成功实现了对动态影片的粗略重构。这项研究的难点在于相对于缓慢变化的血氧依赖水平而言,动态影片的变化速度很快,大量细节信息很难通过血氧依赖水平信号来恢复。Fujiwara 等人(2013)利用贝叶斯典型相关分析(Bayesian canonical correlation analysis, BCCA) 建立了从视觉图像预测大脑响应及从大脑响应预测视觉图像的双向生成式编解码模型。Schoenmakers 等人(2013)使用线性模型对手写字母进行了重建。Cowen 等人(2014)利用主成分分析 (principal component analysis, PCA) 和偏最小二乘回归方法从 fMRI 脑活动中重建出了被试看到的人脸图像。Wen 等人(2018)基于深度神经网络层次化中间特征实现了动态自然刺激的编解码。Güçlütürk 等人(2017)利用回归方法和生成对抗网络从 fMRI 信号重建出了人看到的人脸图像。Du 等人(2019)首次提出了基于多视图变分自编码生成式模型的视觉信息编解码的研究框架,即假定大脑信号和外部刺激是由同一隐含变量生成。通过学习一个多视图变分自编码器可以建立外部刺激到脑信号的双向映射关系。将视觉信息编解码问题看做多视图学习中缺失视图的推断问题。该方法较好地重建出了人看到的字母图像信息。Shen 等人(2019)首先将大脑信号解码到深度神经网络中层次化的视觉特征,然后利用梯度下降和自然图像先验成功重建出了选自于 ImageNet 数据集的自然图像。但是这些重建结果还是较为模糊,且不含语义信息。VanRullen 和 Reddy(2019)基于 VAE 和 GAN 的混合模型也实现了人脸图像重建。Fang 等人(2020)提出了一个基于形状和语义的刺激图像重建方法,该方法分别从低级视觉皮层和高级视觉皮层解码出形状和语义表征,然后将形状和语义信息输入到图像生成网络中进行图像生成。Du 等人(2022)提出了一种新的结构化神经信息解码方法。新研究通过多任务特征解码的方式揭示了多个典型计算机视觉模型(如 VGG(Visual Geometry Group)、ResNet(residual neural network))与人脑腹侧视觉通路在层次化特征表达方面的联系。通过高效结构化地利用这种层次化特征与人脑视觉皮层信号表达之间的关系,该方法能够根据采集到的少量人脑 fMRI 数据清晰地重建出被试所感知到的复杂自然图像和人脸刺激内容。该方法由两个阶段组成,即 Voxel2Unit 和 Unit2Pixel。在 Voxel2Unit 阶段,Du 等人(2022)首先使用矩阵变量高斯先验来建立结构化多输出回归模型,将高维 fMRI 数据解码到卷积神经网络的层次化中间单元特征。在 Unit2Pixel 阶段,作者进一步建立了自省条件生成模型,将预测到的 CNN 中间特征作为条件反演回对应的视觉图像。最近,Wang 等人(2022)进一步利用条件视频生成对抗网络实现了对动态自然刺激的重构,但重建结果仍面临轮廓模糊、语义不清晰等问题。为了改善基于 VAE 模型的图像重建模糊的问题,Zhou 等人(2022a)提出基于可逆归一化流(normalizing flow, NF)的神经编解码框架,在 fMRI 信号上实现了手写数字刺激的重建。

上述 fMRI 解码研究表明,深度神经网络在视觉信息处理方面与人类大脑的视觉处理过程在一定程度上具有类似的表现。因此,深度神经网络是研究神经信息解码的强有力工具。尽管上述神经信息解码工作已经在语义分类、图像辨识与重建方面取得了一定的效果,但是该领域仍然处于发展过程中,尚有很大的提升空间。此外,现有的视觉神经信息解码研究依赖于大量的“视觉刺激—大脑响应”成对数据,神经解码模型能否取得成功很大程度上取决于数据集的大小和质量。

3 公开数据集

视觉神经信息编解码研究中常用的公开数据集如表 1 所示,这些数据集为该领域的发展做出了重要的贡献。

1) 69 (van Gerven 等,2010)。该数据集包含一名被试在 100 幅灰度手写数字图像刺激下的 BOLD 信号。每幅图像以 6 Hz 的频率闪烁 12.5 s。

表 1　视觉神经信息编解码研究中常用的公开数据集

Table 1　Public datasets commonly used in visual neural encoding and decoding

名称	视觉刺激	样本量/对	下载地址
69	手写数字	100	https://repository.ubn.ru.nl/handle/2066/203799
BRAINS	手写字母	360	http://sciencesanne.com/research/
Binary Contrast Patterns	二值图像	1 400	http://brainliner.jp/data/brainliner/Visual_Image_Reconstruction
Vim-1	自然图像	1 750	https://crcns.org/datasets/vc/vim-1
Vim-2	自然视频	7 200	https://crcns.org/data-sets/vc/vim-2
BOLD5000	自然图像	5 254	https://bold5000.github.io/download.html
Generic Object Decoding	自然图像	1 200	https://figshare.com/articles/dataset/Generic_Object_Decoding/7387130
Deep Image Reconstruction	自然图像	6 000	https://figshare.com/articles/dataset/Deep_Image_Reconstruction/7033577
Faces	人脸图像	8 000	https://openneuro.org/datasets/ds001761/versions/2.0.0
NSD	自然图像	10 000	https://natural-scenes-dataset.s3.amazonaws.com/index.html

刺激图像从 MNIST 手写数字数据集中挑选,包含 50 幅手写数字 6 和 50 幅手写数字 9,分辨率为 28×28 像素。采集设备为 3T MRI 采集系统,扫描间隔 TR=2.5 s,体素尺寸为 2 mm×2 mm×2 mm。包含了初级视觉皮层 V1、V2、V3 脑区的体素。

2) BRAINS (Schoenmakers 等,2013)。该数据集以 360 幅灰度手写字母图像做刺激,包含 B、R、A、I、N 和 S 共 6 种字母,每类字母有 60 个样本,分辨率为 56×56 像素。用 3T MRI 采集系统记录了共 3 名被试的 BOLD 信号,扫描间隔 TR=1.74 s,体素尺寸为 2 mm×2 mm×2 mm。为了更准确地估计被试的 BOLD 响应,每个刺激重复呈现两次。包含了初级视觉皮层 V1 和 V2 脑区的体素。

3) Binary Contrast Patterns (Miyawaki 等,2008)。该数据集有两种刺激,一种是 440 幅随机图像,用于模型训练,每个刺激重复呈现 20 次;一种是包含 5 种几何形状和 5 种字母的人工图像,每个刺激重复呈现两次,用于模型测试。图像分辨率为 10×10 像素。实验用 3T MRI 采集设备记录了共 2 名被试的 BOLD 信号,扫描间隔 TR=2 s,体素尺寸为 3 mm×3 mm×3 mm。体素来自初级视觉皮层 V1、V2、V3 和 V4。

4) Vim-1 (Kay 等,2008)。该数据集的视觉刺激为 1 870 幅灰度自然图像,其中训练集的 1 750 幅图像重复呈现 2 次,测试集的 120 幅图像重复呈现 13 次。刺激图像的分辨率为 500×500 像素。采集设备为 4T MRI 扫描仪,扫描间隔 TR=1 s,体素尺寸为 2 mm×2 mm×2 mm。共记录了两名被试的 V1、V2、V3、V4、V3a、V3b 和 LO 视觉区的体素。

5) Vim-2 (Nishimoto 等,2011)。该数据集所用的刺激是动态彩色电影刺激,分辨率为 512×512 像素。共采集了 3 名被试的 BOLD 信号,采集设备为 4T MRI 扫描仪,扫描间隔 TR=1 s,体素尺寸为 2 mm×2 mm×2.5 mm。视频被裁剪成 10～20 s 长度的片段,训练集中的电影片段总时长 120 min,播放一次;测试集中的电影总时长 9 min,重复播放 10 次。训练集和测试集中的电影不同。体素来自 V4、LO、STS、FFA、PPA、OFA 和 RSC。

6) BOLD5000 (Chang 等,2019)。该数据集是大规模的人类 fMRI 数据集,包含从主流计算机视觉数据集,如 SUN (Xiao 等,2010)、微软 COCO(common objects context) (Lin 等, 2014) 和 ImageNet (Deng 等,2009)中挑选出的 4 916 幅自然图像刺激,囊括了丰富的语义类别。刺激大小为 375×375 像素,采集设备为 3T 核磁共振成像系统,TR=2 s,体素尺寸为 2 mm×2 mm×2 mm,记录了 4 名被试 LO、OPA、PPA 和 RSC 脑区的体素活动。

7) Generic Object Decoding (Horikawa 和 Kamitani,2017)。该数据集包括图像刺激实验和想象实验。从 ImageNet 数据集中挑选图像刺激实验中的刺激,在 12×12°的视角范围内呈现刺激。训练集包含 150 类的 1 200 幅图像,每幅仅呈现一次,测试集包含 50 类的 50 幅图像,每类一幅图像,每幅重复呈现 35 次。训练集和测试集的图像类别不同。想象实验中被试被要求根据提示想象测试集中的某一

类图像。共采集了 5 名被试的 BOLD 信号,采集设备为 3T 核磁共振成像系统,TR = 3 s,体素尺寸为 3 mm × 3 mm × 3 mm。体素来自初级视觉区 V1—V4 和高级视觉区 LOC、FFA 和 PPA。

8）Deep Image Reconstruction（Shen 等,2019）。该数据集包括图像刺激实验和想象实验。图像刺激包括 Generic Object Decoding 中的刺激（Horikawa 和 Kamitani,2017）,Miyawaki 等人（2008）所用的 5 种人工形状与 8 种颜色构成的 40 种组合,以及 10 种字母。训练集的 1 200 幅自然图像刺激被重复呈现 5 次,测试集的 50 幅自然图像刺激、40 幅人工形状刺激和 10 幅字母刺激分别重复呈现了 24、20 和 12 次。在想象环节,被试需要根据提示想象测试集中的自然刺激或人工形状刺激。共采集了 3 名被试的 BOLD 信号,采集设备为 3T 核磁共振成像系统,TR = 2 s,体素尺寸为 2 mm × 2 mm × 2 mm。记录了初级视觉区 V1—V4 和高级视觉区 LOC、FFA、PPA 的体素响应。

9）Faces（VanRullen 和 Reddy,2019）。该数据集以人脸作为图像刺激,人脸图像来自 CelebA 数据集（Liu 等,2015）。为了使被试关注面孔的信息或者模型易于重建的背景信息,呈现的刺激图像是原始图像经过 VAE-GAN 自编码器后的重建图像。在每次运行中,图像呈现环节,每名被试在 8 × 8°的视角内被呈现 88 张训练集人脸和 20 张测试集人脸,不同被试的训练集和测试集都不相交;在想象环节,被试需要从 20 张测试刺激中选出一张人脸进行想象。在整个数据采集过程中,平均每名被试观看 8 000 多张人脸刺激。采集设备为 3T 核磁共振成像系统,TR = 2 s,体素尺寸为 3 mm × 3 mm × 3 mm,共有 4 名被试的 BOLD 信号。

10）NSD（Allen 等,2022）。该数据集是大规模的自然场景刺激 fMRI 数据集,包含 73 000 幅刺激图像,这些图像来自 COCO 数据集（Lin 等,2014）。刺激在 8.4 × 8.4°的视角范围内被呈现。在 7T 磁共振成像系统下,TR = 1.6 s,体素尺寸为 1.8 mm × 1.8 mm × 1.8 mm,记录了 8 名被试的高空间分辨率、高信噪比的 BOLD 响应。每名被试在 40 次扫描的过程中总共被呈现 10 000 幅图像,每幅重复呈现 3 次,其中 1 000 幅是所有被试共有的刺激,剩余 9 000 幅刺激在被试间没有交集。

4　度量指标

4.1　编码模型度量指标

对模型编码表现的度量通常通过逐体素地计算预测误差或精度,并在全部体素上求平均。常用的评估单个体素拟合精度的指标有均方误差（mean square error, MSE）,皮尔逊相关系数（Pearson correlation coefficient, PCC）,可决系数（coefficient of determination, R^2）等。

4.2　解码模型度量指标

在评估模型的解码性能时,主要通过评估原始刺激图像和重建图像的相似度,并在所有样本上取均值。常用的评估图像重建质量的指标有均方误差 MSE,峰值信噪比（peak signal to noise ratio, PSNR）,皮尔逊相关系数 PCC 以及结构相似性指数（structural similarity index, SSIM）等。相比 MSE 和 PCC,SSIM 在更高的层面上对图像的相似性进行了衡量（Wang 等,2004）。该指标综合对比一个图像对的亮度、对比度和结构。

5　问题与展望

5.1　视觉神经信息编解码存在的问题

在神经编码方面,现有研究往往是基于预训练卷积神经网络的中间层特征建立体素回归模型。尽管相较传统的基于 Gabor 小波特征的编码方法具有更多的解释性和更好的效果,但是和真实的脑响应数据之间还有很大的拟合误差,且如何进一步减少这种拟合误差目前仍没有解决。现有的视觉神经编码方法还存在以下缺陷:1）计算模型多基于现有的神经网络架构,不能反映真实的生物视觉加工过程;2）由于每个人在视觉加工过程的选择性注意以及 fMRI 数据采集过程中不可避免的噪声导致个体差异大;3）fMRI 数据采集昂贵,对被试的要求较高,不适合长时间采集,因此现有的数据集样本量不足;4）研究者大多基于固定类型的预训练神经网络（例如 AlexNet）来构建神经编码模型的特征空间,造成了视觉特征多样性不足等问题。

在神经解码方面,传统的基于多体素模式分析的方法直接在高维的 fMRI 体素空间和视觉图像像素空间建立映射关系,这种解码方法很容易造成对

冗余或噪声体素的过拟合。尽管现有的视觉信息编解码模型在对大脑信号的分类、辨识任务上表现良好,但是试图建立视觉刺激和大脑视觉皮层信号之间的精确映射关系仍然非常困难,图像重建结果往往不清晰且缺乏明确语义。此外,现有的视觉信息解码方法大多数基于对视觉图像的线性变换或者深度网络变换特征,缺乏对新型特征的探索。阻碍人们有效地进行视觉信息解码、重建图像或视频的因素主要包括 fMRI 数据维度高、样本量小以及噪声严重等。

5.2 展望

随着功能磁共振成像和人工智能技术的进步(叶慧慧 等,2022),现有研究已经可以较为有效地建模被试在观看数字、人脸和自然场景等视觉刺激时的大脑神经活动,并实现了相应的视觉刺激重建和语义解码。未来要通过采用更加先进的人工智能技术研发更为有效的大脑视觉神经信息编解码方法,并尝试将大脑信号翻译成图像、视频、语音和文字等多媒体内容,实现更多的脑—机接口功能。有意义的研究方向包括:1)基于图像和文本联合的多模态神经信息编解码;2)大脑神经信号指导的计算机视觉模型训练与提升;3)基于大规模预训练模型高效特征的视觉神经编解码(张浩宇 等,2022);4)基于神经编解码方法评估现有人工神经网络模型的类脑特性(Zhou 等,2022b)等。此外,由于大脑信号具有复杂性、高维度、个体差异性大、高度动态性和样本量小等特点,因此未来的研究还需要更多结合计算神经科学和人工智能理论研发高鲁棒性、适应性、准确性和解释性的视觉神经信息编解码方法。

6 结 语

基于 fMRI 的视觉神经信息编解码方法的研究,有利于理解视觉系统感知机制和探索人脑视觉皮层高效的信息处理过程。事实上,机器智能和人脑的研究可以相互促进、相辅相成,先进的机器智能有助于探索大脑信息处理的内在神经机制,而大脑的运行机理也将启发新一代类脑计算模型,提高机器智能感知和处理外部信息的能力。因此,基于深度学习和 fMRI 的视觉神经信息编解码方法研究,不仅对脑科学、神经科学有重要意义,更对人工智能领域有深远影响。

本文总结了基于 fMRI 的视觉神经信息编解码方法的研究进展,从视觉神经信息编解码的定义、与统计机器学习之间的关系、研究方法和进展、公开数据集和度量指标等方面进行深入阐述。首先,详细介绍了基于群体感受野估计的视觉神经信息编码方法的发展过程。其次,将视觉神经信息解码分为语义分类、图像辨识/检索和图像重建 3 个部分,并详细介绍了每种解码任务类型的差异。之后,列举了该领域常用的公开数据集以及视觉神经信息编解码算法的度量指标。最后,提出视觉神经信息编解码研究方法的不足,并对未来的研究方向进行了展望。结合更多计算神经科学和人工智能理论研究高鲁棒性、适应性、准确性和解释性的视觉神经信息编解码方法是未来的发展方向。

参考文献(References)

Allen E J, St-Yves G, Wu Y H, Breedlove J L, Prince J S, Dowdle L T, Nau M, Caron B, Pestilli F, Charest I, Hutchinson J B, Naselaris T and Kay K. 2022. A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. Nature Neuroscience, 25(1): 116-126 [DOI: 10.1038/s41593-021-00962-x]

Chang N, Pyles J A, Marcus A, Gupta A, Tarr M J and Aminoff E M. 2019. BOLD5000, a public fMRI dataset while viewing 5000 visual images. Scientific Data, 6(1): #49 [DOI: 10.1038/s41597-019-0052-3]

Cowen A S, Chun M M and Kuhl B A. 2014. Neural portraits of perception: reconstructing face images from evoked brain activity. NeuroImage, 94: 12-22 [DOI: 10.1016/j.neuroimage.2014.03.018]

Cui Y B, Qiao K, Zhang C, Wang L Y, Yan B and Tong L. 2021. GaborNet visual encoding: a lightweight region-based visual encoding model with good expressiveness and biological interpretability. Frontiers in Neuroscience, 15: #614182 [DOI: 10.3389/fnins.2021.614182]

Deng J, Dong W, Socher R, Li L J, Li K and Li F F. 2009. ImageNet: a large-scale hierarchical image database//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE: 248-255 [DOI: 10.1109/cvpr.2009.5206848]

Du C D, Du C Y, Huang L J and He H G. 2019. Reconstructing perceived images from human brain activities with Bayesian deep multiview learning. IEEE Transactions on Neural Networks and Learning Systems, 30(8): 2310-2323 [DOI: 10.1109/tnnls.2018.2882456]

Du C D, Du C Y, Huang L J, Wang H B and He H G. 2022. Structured

neural decoding with multi-task transfer learning of deep neural network representations. IEEE Transactions on Neural Networks and Learning Systems, 33(2): 600-614 [DOI: 10.1109/tnnls.2020.3028167]

Dumoulin S O and Wandell B A. 2008. Population receptive field estimates in human visual cortex. NeuroImage, 39(2): 647-660 [DOI: 10.1016/j.neuroimage.2007.09.034]

Fang T, Qi Y and Pan G. 2020. Reconstructing perceptive images from brain activity by shape-semantic GAN//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: 13038-13048

Fujiwara Y, Miyawaki Y and Kamitani Y. 2013. Modular encoding and decoding models derived from Bayesian canonical correlation analysis. Neural Computation, 25(4): 979-1005 [DOI: 10.1162/neco_a_00423]

Güçlütürk Y, Güçlü U, Seeliger K, Bosch S, van Lier R and van Gerven M A J. 2017. Reconstructing perceived faces from brain activations with deep adversarial neural decoding//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc.: 4249-4260

Haxby J V, Gobbini M I, Furey M L, Ishai A, Schouten J L and Pietrini P. 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science, 293(5539): 2425-2430 [DOI: 10.1126/science.1063736]

Haynes J D and Rees G. 2006. Decoding mental states from brain activity in humans. Nature Reviews Neuroscience, 7(7): 523-534 [DOI: 10.1038/nrn1931]

Horikawa T and Kamitani Y. 2017. Generic decoding of seen and imagined objects using hierarchical visual features. Nature Communications, 8(1): #15037 [DOI: 10.1038/ncomms15037]

Huth A G, Lee T, Nishimoto S, Bilenko N Y, Vu A T and Gallant J L. 2016. Decoding the semantic content of natural movies from human brain activity. Frontiers in Systems Neuroscience, 10: #81 [DOI: 10.3389/fnsys.2016.00081]

Huth A G, Nishimoto S, Vu A T and Gallant J L. 2012. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron, 76(6): 1210-1224 [DOI: 10.1016/j.neuron.2012.10.014]

Kamitani Y and Tong F. 2005. Decoding the visual and subjective contents of the human brain. Nature Neuroscience, 8(5): 679-685 [DOI: 10.1038/nn1444]

Kay K N, Naselaris T, Prenger R J and Gallant J L. 2008. Identifying natural images from human brain activity. Nature, 452(7185): 352-355 [DOI: 10.1038/nature06713]

Kay K N, Winawer J, Rokem A, Mezer A and Wandell B A. 2013. A two-stage cascade model of BOLD responses in human visual cortex. PLoS Computational Biology, 9(5): #e1003079 [DOI: 10.1371/journal.pcbi.1003079]

Khosla M, Ngo G H, Jamison K, Kuceyeski A and Sabuncu M R. 2020. Neural encoding with visual attention. Advances in Neural Information Processing Systems, 33: 15942-15953

Kriegeskorte N, Mur M and Bandettini P A. 2008. Representational similarity analysis-connecting the branches of systems neuroscience. Frontiers in Systems Neuroscience, 2: #4 [DOI: 10.3389/neuro.06.004.2008]

Lee S, Papanikolaou A, Logothetis N K, Smirnakis S M and Keliris G A. 2013. A new method for estimating population receptive field topography in visual cortex. NeuroImage, 81: 144-157 [DOI: 10.1016/j.neuroimage.2013.05.026]

Li D, Du C D, Huang L J, Chen Z Q and He H G. 2018. Multi-label semantic decoding from human brain activity//Proceedings of the 24th International Conference on Pattern Recognition (ICPR). Beijing, China: IEEE: 3796-3801 [DOI: 10.1109/icpr.2018.8545855]

Li D, Du C D, Wang H B, Zhou Q Y and He H G. 2022. Deep modality assistance co-training network for semi-supervised multi-label semantic decoding. IEEE Transactions on Multimedia, 24: 3287-3299 [DOI: 10.1109/tmm.2021.3104980]

Li D, Du C D, Wang S P, Wang H B and He H G. 2021. Multi-subject data augmentation for target subject semantic decoding with deep multi-view adversarial learning. Information Sciences, 547: 1025-1044 [DOI: 10.1016/j.ins.2020.09.012]

Lin T Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P and Zitnick C L. 2014. Microsoft COCO: common objects in context//Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland: Springer: 740-755 [DOI: 10.1007/978-3-319-10602-1_48]

Liu Z W, Luo P, Wang X G and Tang X O. 2015. Deep learning face attributes in the wild//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE: 3730-3738 [DOI: 10.1109/iccv.2015.425]

Miyawaki Y, Uchida H, Yamashita O, Sato M A, Morito Y, Tanabe H C, Sadato N and Kamitani Y. 2008. Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. Neuron, 60(5): 915-929 [DOI: 10.1016/j.neuron.2008.11.004]

Naselaris T, Kay K N, Nishimoto S and Gallant J L. 2011. Encoding and decoding in fMRI. NeuroImage, 56(2): 400-410 [DOI: 10.1016/j.neuroimage.2010.07.073]

Naselaris T, Prenger R J, Kay K N, Oliver M and Gallant J L. 2009. Bayesian reconstruction of natural images from human brain activity. Neuron, 63(6): 902-915 [DOI: 10.1016/j.neuron.2009.09.006]

Nishimoto S, Vu A T, Naselaris T, Benjamini Y, Yu B and Gallant J L. 2011. Reconstructing visual experiences from brain activity evoked by natural movies. Current Biology, 21(19): 1641-1646 [DOI: 10.1016/j.cub.2011.08.031]

Norman K A, Polyn S M, Detre G J and Haxby J V. 2006. Beyond

mind-reading: multi-voxel pattern analysis of fMRI data. Trends in Cognitive Sciences, 10(9): 424-430 [DOI: 10.1016/j.tics.2006.07.005]

Schmah T, Hinton G E, Zemel R S, Small S L and Strother S. 2008. Generative versus discriminative training of RBMs for classification of fMRI images//Proceedings of the 21st International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: 1409-1416

Schoenmakers S, Barth M, Heskes T and van Gerven M. 2013. Linear reconstruction of perceived images from human brain activity. NeuroImage, 83: 951-961 [DOI: 10.1016/j.neuroimage.2013.07.043]

Shen G H, Horikawa T, Majima K and Kamitani Y. 2019. Deep image reconstruction from human brain activity. PLoS Computational Biology, 15(1): #e1006633 [DOI: 10.1371/journal.pcbi.1006633]

Stansbury D E, Naselaris T and Gallant J L. 2013. Natural scene statistics account for the representation of scene categories in human visual cortex. Neuron, 79(5): 1025-1034 [DOI: 10.1016/j.neuron.2013.06.034]

St-Yves G and Naselaris T. 2018. The feature-weighted receptive field: an interpretable encoding model for complex feature spaces. NeuroImage, 180: 188-202 [DOI: 10.1016/j.neuroimage.2017.06.035]

Van Gerven M A J, De Lange F P and Heskes T. 2010. Neural decoding with hierarchical generative models. Neural Computation, 22(12): 3127-3142 [DOI: 10.1162/neco_a_00047]

VanRullen R and Reddy L. 2019. Reconstructing faces from fMRI patterns using deep generative neural networks. Communications Biology, 2(1): #193 [DOI: 10.1038/s42003-019-0438-y]

Wang C, Yan H M, Huang W, Li J Y, Wang Y T, Fan Y S, Sheng W, Liu T, Li R and Chen H F. 2022. Reconstructing rapid natural vision with fMRI-conditional video generative adversarial network. Cerebral Cortex, 32(20): 4502-4511 [DOI: 10.1093/cercor/bhab498]

Wang H B, Huang L J, Du C D, Li D, Wang B and He H G. 2021. Neural encoding for human visual cortex with deep neural networks learning "What" and "Where". IEEE Transactions on Cognitive and Developmental Systems, 13(4): 827-840 [DOI: 10.1109/tcds.2020.3007761]

Wang Z, Bovik A C, Sheikh H R and Simoncelli E P. 2004. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4): 600-612 [DOI: 10.1109/tip.2003.819861]

Wen H G, Shi J X, Zhang Y Z, Lu K H, Cao J Y and Liu Z M. 2018. Neural encoding and decoding with deep learning for dynamic natural vision. Cerebral Cortex, 28(12): 4136-4160 [DOI: 10.1093/cercor/bhx268]

Xiao J X, Hays J, Ehinger K A, Oliva A and Torralba A. 2010. SUN database: large-scale scene recognition from abbey to zoo//Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE: 3485-3492 [DOI: 10.1109/cvpr.2010.5539970]

Ye H H, He H J, Fang J W, Tong Q Q, Zhou Z H and Liu H F. 2022. Research progress of quantitative multimodal brain imaging technology. Journal of Image and Graphics, 27(6): 1944-1955 (叶慧慧, 何宏建, 方静宛, 童琪琦, 周子涵, 刘华锋. 2022. 大脑多模态成像技术定量研究进展. 中国图象图形学报, 27(6): 1944-1955) [DOI: 10.11834/jig.220153]

Zeidman P, Silson E H, Schwarzkopf D S, Baker C I and Penny W. 2018. Bayesian population receptive field modelling. NeuroImage, 180: 173-187 [DOI: 10.1016/j.neuroimage.2017.09.008]

Zhang H Y, Wang T B, Li M Z, Zhao Z, Pu S L and Wu F. 2022. Comprehensive review of visual-language-oriented multimodal pre-training methods. Journal of Image and Graphics, 27(9): 2652-2682 (张浩宇, 王天保, 李孟择, 赵洲, 浦世亮, 吴飞. 2022. 视觉语言多模态预训练综述. 中国图象图形学报, 27(9): 2652-2682 [DOI: 10.11834/jig.220173]

Zhou Q Y, Du C D, Li D, Wang H B, Liu K J and He H G. 2022a. Neural encoding and decoding with a flow-based invertible generative model. IEEE Transactions on Cognitive and Developmental Systems. Early Access, https://ieeexplore.ieee.org/document/9780264 [DOI: 10.1109/TCDS.2022.3176977]

Zhou Q Y, Du C D, He H G. 2022b. Exploring the brain-like properties of deep neural networks: a neural encoding perspective. Machine Intelligence Research, 19(5): 439-455 [DOI: 10.1007/s11633-022-1348-x]

Zuiderbaan W, Harvey B M and Dumoulin S O. 2012. Modeling center-surround configurations in population receptive fields using fMRI. Journal of Vision, 12(3): #10 [DOI: 10.1167/12.3.10]

作者简介

杜长德,男,助理研究员,主要研究方向为模式识别、深度学习、神经信息编解码、计算机视觉、脑科学与类脑智能。
E-mail: changde.du@ia.ac.cn
何晖光,通信作者,男,研究员,主要研究方向为模式识别、脑机接口、脑科学与类脑智能。E-mail:huiguang.he@ia.ac.cn
周琼怡,女,博士研究生,主要研究方向为模式识别、视觉神经信息编解码和类脑智能。
E-mail:zhouqiongyi2018@ia.ac.cn
刘澈,男,硕士研究生,主要研究方向为模式识别、视听觉神经信息编解码和类脑智能。E-mail:liuche2022@ia.ac.cn