Research  
Neural Information Engineering—Review

## 基于双向深度生成模型和功能磁共振成像数据的大脑编码和解码

杜长德<sup>a,b</sup>, 李劲鹏<sup>a,b</sup>, 黄利皆<sup>a,b</sup>, 何晖光<sup>a,b,c,\*</sup><sup>a</sup> Research Center for Brain-Inspired Intelligence and National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences (CAS), Beijing 100190, China<sup>b</sup> School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China<sup>c</sup> Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai 200031, China

## ARTICLE INFO

## Article history:

Received 29 September 2017

Revised 28 March 2019

Accepted 29 March 2019

Available online 1 June 2019

## 关键词

大脑编码和解码

功能磁共振成像

深度神经网络

深度生成模型

对偶学习

## 摘要

通过功能磁共振成像 (fMRI) 进行大脑编码和解码是视觉神经科学的两个重要方面。尽管以前的研究人员在大脑编码和解码模型方面取得了显著进步, 但是现有方法仍需要使用先进的机器学习技术进行改进。例如, 传统方法通常会分别构建编码和解码模型, 并且容易对小型数据集过度拟合。实际上, 有效地统一编码和解码过程可以进行更准确的预测。在本文中, 我们首先回顾了现有的编码和解码方法, 并讨论了“双向”建模策略的潜在优势。接下来, 在体系结构和计算规则方面, 我们证明了深度神经网络和人类视觉通路之间存在的对应关系。此外, 深度生成模型[如变分自编码器 (VAE) 和生成对抗网络 (GAN)]在大脑编码和解码研究中产生了可喜的成果。最后, 我们提出了最初为机器翻译任务设计的对偶学习方法, 该方法通过利用大规模未配对数据提高了编码和解码模型的效果。

© 2019 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. 引言

人类视觉感知与其所诱发的神经活动之间的关系对于计算神经科学领域至关重要[1,2]。通过功能磁共振成像 (fMRI) 进行大脑编码和解码对于理解视觉感知系统很重要[3–5]。编码模型试图根据给定的视觉刺激来预测大脑反应[6,7], 而解码模型则试图通过分析给定的大脑反应来预测相应的视觉刺激[8–22]。大脑编码和解码(图1)已变成两个促进认知神经科学发展的重要途径, 因为它们为理解大脑功能提供了新的方式。

## 1.1. 编码模型

在以前的文献中, 大多数编码模型都是基于特定

的计算规则而建立的。神经科学家认为这些计算规则可能是大脑对特定视觉刺激做出反应的运算基础。例如, Kay等[1]使用金字塔形的Gabor小波滤波器建立编码模型。基于这种编码模型, 作者成功地识别出了特定人类大脑活动对应的自然图像。后来, Kay等[6]进一步提出了基于已建立的面向局部滤波器、除法归一化、压缩空间求和以及类方差非线性的两级级联编码模型。最近, St-Yves和Naselaris [7]基于预训练的深度神经网络 (DNN) 的中间特征图构建了特征加权的感受野模型。该模型可用于预测体素反应并研究每个体素的感受野的形状。此外, Zeidman等[23]建立了Bayesian群体感受野 (pRF) 模型, 用于可解释的大脑编码研究。近年来, DNN在计算机视觉领域已经取得了巨大的成功, 研究

\* Corresponding author.

E-mail address: [huiguang.he@ia.ac.cn](mailto:huiguang.he@ia.ac.cn) (H. He).

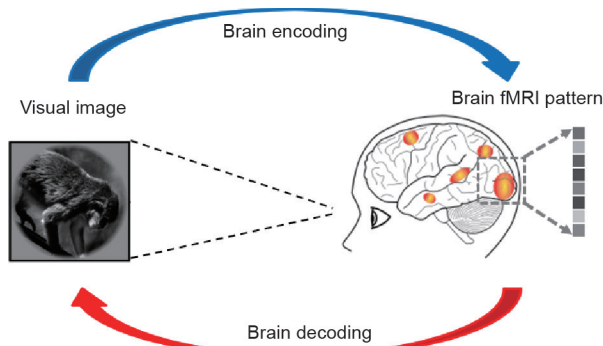


图1. 功能磁共振成像的大脑编码和解码。编码模型用于尝试根据给定的视觉刺激来预测大脑反应，而解码模型则试图通过分析给定的大脑反应来预测相应的视觉刺激。在实践中，编码和解码模型不应该被视为是相互排斥的。有效地统一编码和解码过程可以进行更准确的预测，并有助于我们理解人类大脑中的信息表征。

人员已经开始使用DNN来构建更复杂的大脑编码模型[7,20,24]。除了视觉信息的编码模型，研究人员还对语义信息如何在脑中表达进行了研究。例如，Huth等[25]建立了文本语义和大脑皮层活动之间的映射关系向量，从而提供了详细的大脑皮层语义图。

## 1.2. 解码模型

先前的研究已经证明了从相应的大脑激活模式中解码二进制对比模式[12–14]、手写字符[15,16]、人脸图像[17–19]、自然图片或视频刺激[2,20]和梦[12,21]的可行性。例如，Miyawaki等[12]构建了一种多尺度神经解码模型，重建了大脑反应中感知到的二进制对比模式。Schoenmakers等[15]提出了一种线性解码模型，重建了大脑反应中感知到的手写字符。Güçlütürk等[19]提出将概率推理与对抗训练相结合，以重建大脑反应感知到的面孔。Horikawa和Kamitani[2]指出，可以通过利用多个大脑区域的响应来预测由计算机视觉模型计算出的视觉刺激的分层特征。这些研究表明，视觉皮层与由计算机视觉模型获得的复杂视觉特征之间存在密切的关系。此外，Wen等[20]提出了一种基于深度学习的动态神经解码方法，该方法可以重建人类感知的动态视觉场景并预测其语义标签。Horikawa和Kamitani[21]还表明，大脑活动可以用来预测人类梦中出现的物体。

前面提到的大多数解码研究都是以多体素模式分析(MVPA)方法为基础的[8]。然而，大脑连接模式也是大脑状态的关键特征，可用于大脑解码。先前的解码研究[26–30]表明，大脑的连接性信息可以当作解码过程中的特征。例如，Yargholi和Hossein-Zadeh[29]通过在大脑解码中采用大脑连接性信息，能够成功地从人类大

脑中重建两个手写数字——6和9。Manning等[30]提出了一种概率模型，用于提取大脑活动中的动态功能连接模式，这种模式可用于大脑解码研究。

## 1.3. 混合编码 - 解码双向模型

尽管大脑编码和解码的最新进展[3–21,29,31–33]展示出了良好的前景，但如何从功能磁共振成像数据中重建相应的视觉刺激，构建一个准确的解码模型仍然面临许多挑战。从Bayesian机器学习角度来看，可以通过一个大脑活动的生成模型来获取编码模型。当将此编码模型与有关刺激的先验知识相结合时，在给定大脑活动模式的情况下，可以获得刺激的后验概率分布，即用于解码的预测分布。因此，编码和解码模型不应该被视为是相互独立的。有效地统一编码和解码过程可以进行更加准确的预测，并有助于理解人类大脑中的信息表征[13,34]。例如，Fujiwara等[13]提出了一种“双向”的视觉图像重建方法，该方法假设一组潜在变量与图像像素和fMRI体素相关；这种方法同时支持编码和解码的预测。这些学者采用了Bayesian典型相关分析(BCCA)框架，通过潜在变量计算了图像像素与fMRI体素之间的多重对应。由于可以将每个潜在变量的像素权重定义为图像基，因此使用测量数据训练BCCA模型可自动估计出图像基。尽管对估计的图像基的功能含义进行推测还为时过早，但是这种数据驱动的“双向”方法可以扩展到用于发现大脑的模块化体系结构，以表示在高维空间中定义的复杂的自然刺激、行为和心理感受。

## 2. DNN 与人类视觉系统之间的对应关系

深度学习[35,36]是一大类从输入数据中提取层次表示的机器学习方法。DNN的框架结构首先受到生物神经系统的结构和计算原理的启发[37]。最近，基于DNN的深度学习方法在图像识别、语音识别、自然语言理解等方面取得了巨大的成功。在架构方面，DNN的层次结构与人类大脑腹侧视觉系统的层次结构非常相似[7,35,38]（图2）。在功能方面，深度学习的神经编码和解码研究表明，DNN的浅层表示类似于主视觉区域的功能，而DNN的深度表示则类似于腹侧视觉系统后端的功能[2,24,39,40]。

人类可以通过腹侧视觉通路快速、准确地感知复杂的物体，腹侧视觉通路是一个由大脑区域相互连接组成的系统，可以处理结构逐渐复杂的特征[41,42,43]。然

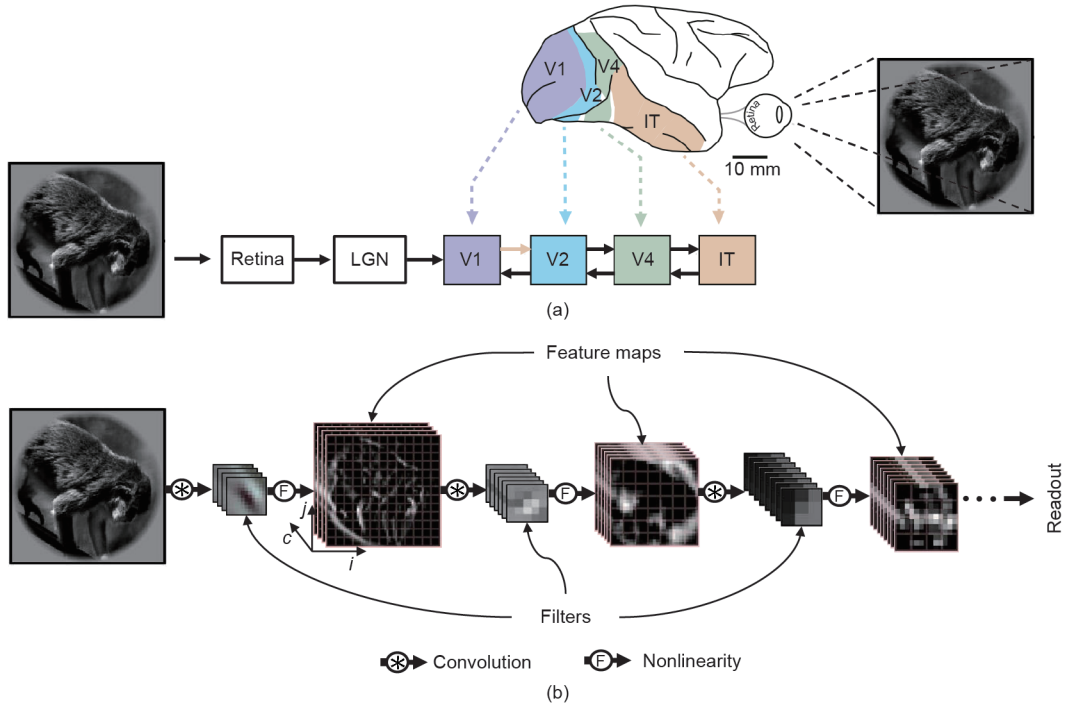


图2. 腹侧视觉系统和深度卷积神经网络 (CNN)。(a) 四个Brodmann区域 (V1、V2、V4和IT) 之间的正向和反向投影；(b) 一个深度前馈CNN的简单示例，它的层次结构被用来模拟腹侧视觉系统的层次表示。LGN：外侧膝状体核。(a) 经Elsevier许可，转自参考文献[38]，© 2014；(b) 经Elsevier许可，转自参考文献[7]，© 2017。

而，在没有监督信息的情况下从视觉图像中自动发现早期视觉概念是机器感知研究中公认的挑战。一方面，从图像中提取的表示形式将有助于在现实世界中更好地执行任务。另一方面，我们希望能够解释这些表示，并使它们能够用于其初始设计任务之外的任务。从传统角度看，使用预训练的DNN模型很难从视觉图像中学习此类表示，因为由该DNN模型从输入图像中自动提取的表示向量的每个维度的语义都是未知的。如果没有分离的表示，我们很难在不同的任务中解释这些表示。所幸的是，Higgins等[44]特别设计的深度生成模型能够学习分离的表示。

### 3. 使用深度生成模型进行大脑解码

一个有前景的研究方向是将深度学习方法应用到大脑解码研究中。诸如变分自动编码器 (VAE) [45,46] 和生成对抗网络 (GAN) [47] 之类的深度生成模型，在图像生成领域取得了巨大的成功。最近，越来越多的注意力集中在了使用深度生成模型进行视觉图像重建的研究上[19,31–33,48,49]。

#### 3.1. 基于 VAE 的方法

最初在参考文献[45,46]中介绍过VAE，它是自动编

码器模型的概率扩展。VAE具有自下而上的编码网络和自上而下的解码网络。研究人员通过共同训练这两个网络，可以使数据似然性的下界最大化，从而将自动编码器模型重新表示为变分推断问题。最近的研究表明，基于VAE的模型能够学习与输入数据中不同的变化因素相对应的解耦表示[43,50,51]。这对于大脑的编码和解码任务非常重要，因为基于VAE的模型学习到的一些视觉概念也可以被人类大脑感知。受这一事实的启发，研究人员探索了基于VAE的模型在大脑活动图像重建中的应用[31,32]。

例如，Du等[31]提出了一种深度生成多视图模型 (deep generate multi-view model, DGMM)，用于重建大脑fMRI活动对应的感知图像 (图3)。DGMM可以看作是线性BCCA的非线性扩展。在DGMM框架下，编码和解码过程同时由两种不同的生成模型表示：

$$p_{\theta}(X|Z) = \prod_{i=1}^N \mathcal{N}\{x_i | \mu_x(z_i), \text{diag}[\sigma_x^2(z_i)]\} \quad (1)$$

$$p(Y|Z) = \prod_{i=1}^N \mathcal{N}(y_i | B^T z_i, \psi) \quad (2)$$

式中， $\mathcal{N}$ 表示正态分布； $X \in \mathbb{R}^{D_x \times N}$ 表示视觉图像； $Y \in \mathbb{R}^{D_y \times N}$ 表示诱发的fMRI活动； $p_{\theta}(X|Z)$ 是具有神经网络参数 $\theta$ 的视觉图像的似然函数； $p(y|z)$ 是所诱发的

fMRI的似然函数； $\psi$ 表示完整的协方差矩阵； $B$ 表示fMRI活动的投影权重； $Z \in \mathbb{R}^{D_z \times N}$ 表示视觉图像和诱发的fMRI活动之间的共享潜变量。 $\mu_x$ 和 $\sigma_x^2$ 分别表示该正态分布的均值和协方差，它们是通过对潜变量进行不同的非线性变换获得的。训练集由 $N$ 个成对样本组成，可以用 $(x_1, y_1), \dots, (x_n, y_n)$ 表示，其中 $x_i \in \mathbb{R}^{D_x}$ ,  $y_i \in \mathbb{R}^{D_y}$ ,  $i=1, \dots, N$ 。具体来说，DGMM使用基于DNN的生成过程对视觉图像的分布进行建模，而使用稀疏线性生成过程对大脑反应数据的分布进行建模。一方面，此处使用的DNN可以有效捕获视觉图像的分层特征，类似于人类大脑腹侧视觉系统的结构[2,24,39,40]。另一方面，这里使用的稀疏线性生成模型不仅符合人类大脑的稀疏表达原理，而且避免了大脑反应数据的过度拟合[52]。请注意，这两个生成过程拥有相同的潜在变量。因此，在测试阶段，使用这些过程可以通过相同的潜在变量从大脑反应中推断出相应的视觉图像。实际上，DGMM框架可以捕获视觉图像和相应的fMRI活动之间的“双向”映射关系，得益于其自动编码的变分Bayesian架构，DGMM可以通过平均场变分推断得到有效的优化，这类类似于经典的VAE求解方法。与非概率扩展的深度多视图学习方法相比，DGMM的Bayesian框架天然地使它更具灵活性和适应性。

### 3.2. 基于GAN的方法

GAN在参考文献[47]中被首次提出。基本的GAN是一种无监督模型，可利用噪声向量生成图像。对抗训练的思想来自博弈论，在博弈论中，两个竞争者为了共同进步而竞争。GAN的典型配置包括生成器和鉴别器。生成器的任务是从噪声中合成图像，以欺骗鉴别器，使它相信合成图像是真实世界的场景。同时，鉴别器试图区分合成数据和真实数据。当二者达到纳什（Nash）平衡时，生成器将学习到真实世界图像的分布，并且鉴别器难以捕捉到真实数据与虚假数据之间的差异。GAN已被广泛应用，包括图像生成[53]、图像到图像翻译[54]和文本到图像合成[55,56]。

不同于VAE，GAN是个无似然模型。也就是说，它没有对数据分布进行任何先验假设，数据分布是完全通过对抗训练学习的。这是对神经编码和解码任务有利的特征。GAN通常要求生成器和鉴别器具有准确的语义信息。然而，在血氧水平依赖性（BOLD）信号中，有用的语义信息和噪声混杂在一起，这对模型训练是一个巨大的挑战。最近的大脑解码研究[19]提出将概率推理与对抗训练相结合，从而从大脑活动中重建感知到的面部结构（图4）。假设 $x \in \mathbb{R}^{h \times w \times c}$ 是视觉图像， $z \in \mathbb{R}^p$ 是其潜在特征， $y \in \mathbb{R}^q$ 是相应的大脑反应，而 $\phi \in \mathbb{R}^{h \times w \times c} \rightarrow$

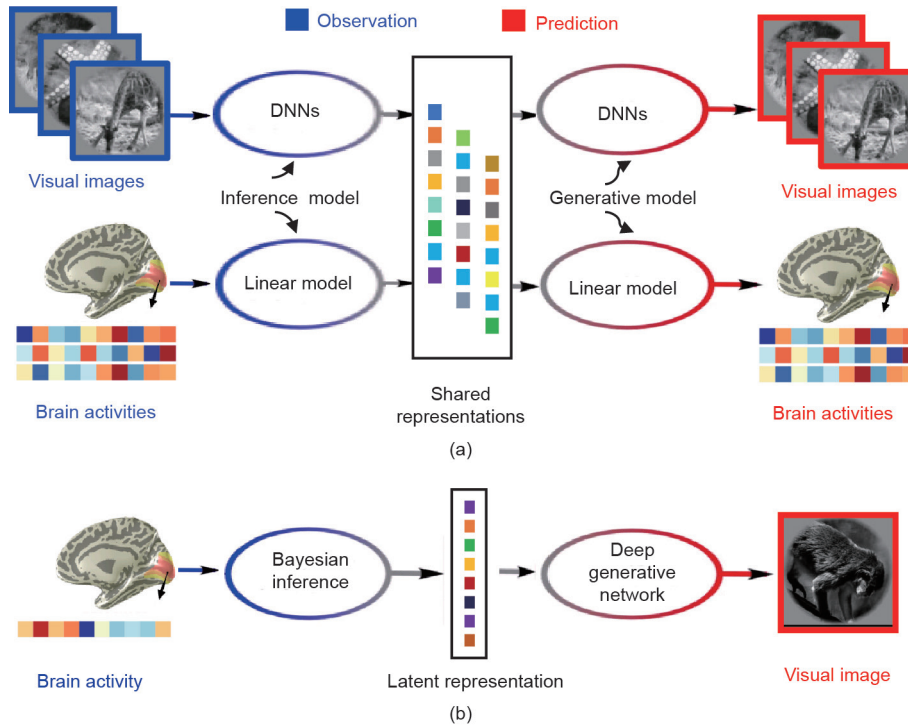


图3. 用于神经解码的深度生成多视图框架。(a) 模型训练：使用特定于视图的生成模型生成数据；其中，DNN模型用于视觉图像建模，线性回归模型用于大脑活动建模。(b) 图像重建：用于与训练无关的大脑活动，该大脑活动被解码成视觉图像。

$\mathbb{R}^p$  是一个潜在特征模型，使得  $z = \varphi(x)$  和  $x = \varphi^{-1}(z)$ 。然后，可以通过以下方程式从大脑反应中重建感知到的视觉图像：

$$\hat{x} = \varphi^{-1} \left[ \underset{z}{\operatorname{argmax}} p(z|y) \right] \quad (3)$$

式中， $p(z|y)$  是潜在变量的后验分布。等式 (3) 可以通过 Bayes 定理来重新表述：

$$\hat{x} = \varphi^{-1} \left\{ \underset{z}{\operatorname{argmax}} [p(y|z)p(z)] \right\} \quad (4)$$

式中， $p(y|z)$  是似然函数； $p(z)$  是潜在变量的先验分布。作者首先以最大后验估计直观地解码了所观察到的大脑对潜在特征的反应。接下来，他们使用对抗学习根据解码后的潜在特征生成了感知图像。这种两步式大脑解码方法可以根据大脑反应准确地生成感知到的面孔。最近，研究人员试图通过利用已在大规模图像数据集上进行过预训练的 GAN，从测得的 fMRI 信号 [33,48,49] 中重建自然图像。

#### 4. 通过对偶学习改善大脑编码和解码

数据驱动的大脑编码和解码方法通常需要获取大量成对的（刺激响应）数据实例，以训练针对单个被试者的模型。但是，在许多编码和解码研究中，最多可以从

单个被试者中收集数千个含有噪声的配对数据实例。为了提高编码和解码模型的泛化能力，有必要充分利用大规模未配对的数据实例（如视觉图像）。

受到最近提出的机器翻译 [57,58] 对偶学习的启发，我们建议可以通过最小化双向映射模型产生的重构损失来同时训练编码和解码模型。编码和解码模型代表一对原始对偶问题，并由其形成一个闭环，从而允许对偶学习的应用 (图 5)。具体而言，在未配对数据 (如视觉图像) 上测得的重建损失将产生信息反馈，以训练双向映射模型。在这种对偶学习框架下，可以利用大规模的未配对视觉图像来提高编码和解码模型的泛化能力。实际上，对偶学习是用于学习从一个数据域  $M_d$  到另一个数据域  $N_d$  [59,60] 的双向映射的通用框架。对于  $M_d \rightarrow N_d$ ，目标是学习编码器映射  $E$ ，以使分布  $E(M_d)$  与分布  $N_d$  在使用对抗损失时无法区分。同样，对于  $N_d \rightarrow M_d$ ，目标是学习解码器映射  $D$ ，以使分布  $D(N_d)$  与分布  $M_d$  在使用另一个对抗损失时无法区分。特别是对于成对的数据，可以将这两个对抗损失和循环一致性损失（对偶损失）结合起来，使得  $D[E(M_d)] \approx M_d$  和  $E[D(N_d)] \approx N_d$ 。

#### 5. 结论

总之，大脑编码和解码是计算神经科学领域的核心，并具有创造更好的脑机接口的潜力。DNN 的体系结构和计算规则与人类视觉通路具有某些相似之处。在

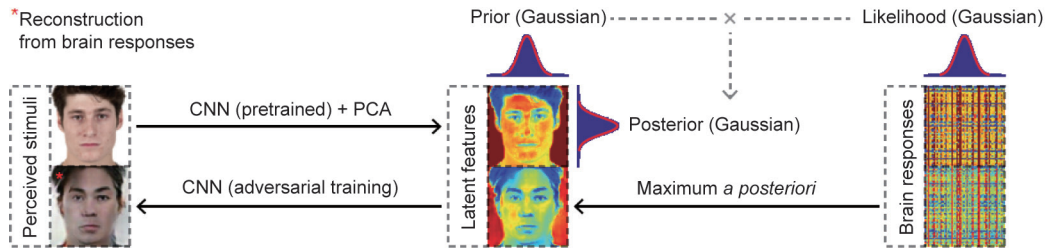


图4. 深度对抗神经解码。该方法将概率推理与对抗学习相结合，能够根据大脑活动清晰地重构出相应的人脸图像。PCA：主成分分析。经 Neural Information Processing Systems Foundation, Inc. 许可，转自参考文献 [19]，©2017。

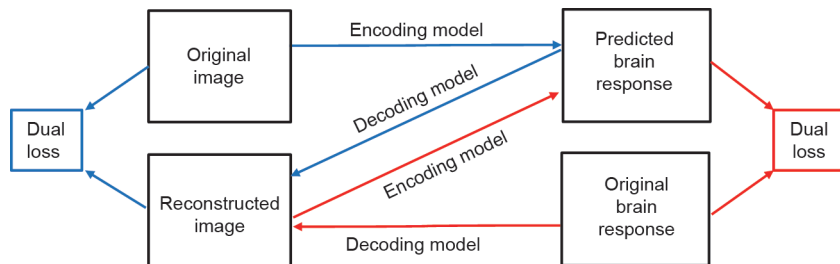


图5. 用对偶学习改进大脑编码和解码。在未配对数据（无论是视觉图像还是大脑反应）上测量的对偶损失产生信息反馈，以训练双向映射模型。在这种对偶学习框架下，可以利用大规模的未配对数据来提高模型的泛化能力。

大脑编码和解码研究中使用深度生成模型（如VAE和GAN），有望为人类了解视觉刺激与诱发的神经活动之间的关系提供更深刻的见解。通过利用大规模的未配对数据，对偶学习有望在开发神经编码和解码模型中发挥重要作用。

## 致谢

这项工作得到了国家重点研发计划（2018YFC2001302）、国家自然科学基金（91520202）、中国科学院科研装备研制项目（YJKYYQ20170050）、北京市科学技术委员会（Z181100008918010）、中国科学院青年创新促进会和中国科学院战略性先导科技专项（XDB32040200）的支持。

## Compliance with ethics guidelines

Changde Du, Jinpeng Li, Lijie Huang, and Huiguang He declare that they have no conflict of interest or financial conflicts to disclose.

## References

- [1] Kay KN, Naselaris T, Prenger RJ, Gallant JL. Identifying natural images from human brain activity. *Nature* 2008;452(7185):352–5.
- [2] Horikawa T, Kamitani Y. Generic decoding of seen and imagined objects using hierarchical visual features. *Nat Commun* 2017;8:15037.
- [3] Naselaris T, Kay KN, Nishimoto S, Gallant JL. Encoding and decoding in fMRI. *NeuroImage* 2011;56(2):400–10.
- [4] Chen M, Han J, Hu X, Jiang X, Guo L, Liu T. Survey of encoding and decoding of visual stimulus via fMRI: an image analysis perspective. *Brain Imaging Behav* 2014;8(1):7–23.
- [5] Van Gerven MA. A primer on encoding models in sensory neuroscience. *J Math Psychol* 2017;76:172–83.
- [6] Kay KN, Winawer J, Rokem A, Mezer A, Wandell BA. A two-stage cascade model of BOLD responses in human visual cortex. *PLoS Comput Biol* 2013;9(5): e1003079.
- [7] St-Yves G, Naselaris T. The feature-weighted receptive field: an interpretable encoding model for complex feature spaces. *NeuroImage* 2018;180(Pt A):188–202.
- [8] Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 2001;293(5539):2425–30.
- [9] Haynes JD, Rees G. Decoding mental states from brain activity in humans. *Nat Rev Neurosci* 2006;7(7):523–34.
- [10] Naselaris T, Prenger RJ, Kay KN, Oliver M, Gallant JL. Bayesian reconstruction of natural images from human brain activity. *Neuron* 2009;63(6):902–15.
- [11] Horikawa T, Tamaki M, Miyawaki Y, Kamitani Y. Neural decoding of visual imagery during sleep. *Science* 2013;340(6132):639–42.
- [12] Miyawaki Y, Uchida H, Yamashita O, Sato MA, Morito Y, Tanabe HC, et al. Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron* 2008;60(5):915–29.
- [13] Fujiwara Y, Miyawaki Y, Kamitani Y. Modular encoding and decoding models derived from bayesian canonical correlation analysis. *Neural Comput* 2013;25(4):979–1005.
- [14] Yu S, Zheng N, Ma Y, Wu H, Chen B. A novel brain decoding method: a correlation network framework for revealing brain connections. 2017. arXiv:1712.01668.
- [15] Schoenmakers S, Barth M, Heskes T, Van Gerven M. Linear reconstruction of perceived images from human brain activity. *NeuroImage* 2013;83:951–61.
- [16] Schoenmakers S, Güçlü U, Van Gerven M, Heskes T. Gaussian mixture models and semantic gating improve reconstructions from human brain activity. *Front Comput Neurosci* 2015;8:173.
- [17] Cowen AS, Chun MM, Kuhl BA. Neural portraits of perception: reconstructing face images from evoked brain activity. *NeuroImage* 2014;94:12–22.
- [18] Lee H, Kuhl BA. Reconstructing perceived and retrieved faces from activity patterns in lateral parietal cortex. *J Neurosci* 2016;36(22):6069–82.
- [19] Güçlütürk Y, Güçlü U, Seeliger K, Bosch S, Van Lier R, Van Gerven MA. Reconstructing perceived faces from brain activations with deep adversarial neural decoding. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al., editors. *Advances in neural information processing systems 30 (NIPS 2017)* La Jolla: Neural Information Processing Systems Foundation. 2017. p. 4249–60.
- [20] Wen H, Shi J, Zhang Y, Lu K, Cao J, Liu Z. Neural encoding and decoding with deep learning for dynamic natural vision. *Cereb Cortex* 2018;28(12):4136–60.
- [21] Horikawa T, Kamitani Y. Hierarchical neural representation of dreamed objects revealed by brain decoding with deep neural network features. *Front Comput Neurosci* 2017;11:4.
- [22] Naselaris T, Olman CA, Stansbury DE, Ugurbil K, Gallant JL. A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. *NeuroImage* 2015;105:215–28.
- [23] Zeidman P, Silson EH, Schwarzkopf DS, Baker CI, Penny W. Bayesian population receptive field modelling. *NeuroImage* 2018;180(Pt A):173–87.
- [24] Güçlü U, Van Gerven MA. Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *J Neurosci* 2015;35(27):10005–14.
- [25] Huth AG, De Heer WA, Griffiths TL, Theunissen FE, Gallant JL. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* 2016;532(7600):453–8.
- [26] Shirer WR, Ryali S, Rykhlevskaia E, Menon V, Greicius MD. Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cereb Cortex* 2012;22(1):158–65.
- [27] Mokhtari F, Hossein-Zadeh GA. Decoding brain states using backward edge elimination and graph kernels in fMRI connectivity networks. *J Neurosci Methods* 2013;212(2):259–68.
- [28] Yargholi E, Hossein-Zadeh GA. Brain decoding-classification of hand written digits from fMRI data employing Bayesian networks. *Front Hum Neurosci* 2016;10:351.
- [29] Yargholi E, Hossein-Zadeh GA. Reconstruction of digit images from human brain fMRI activity through connectivity informed Bayesian networks. *J Neurosci Methods* 2016;257:159–67.
- [30] Manning JR, Zhu X, Willke TL, Ranganath R, Stachenfeld K, Hasson U, et al. A probabilistic approach to discovering dynamic full-brain functional connectivity patterns. *NeuroImage* 2018;180(Pt A):243–52.
- [31] Du C, Du C, He H. Sharing deep generative representation for perceived image reconstruction from human brain activity. In: *Proceedings of the 2017 International Joint Conference on Neural Networks*; 2017 May 14–19; Anchorage, AK, USA. New York: IEEE; 2017. p. 1049–56.
- [32] Han K, Wen H, Shi J, Lu K, Zhang Y, Liu Z. Variational autoencoder: an unsupervised model for modeling and decoding fMRI activity in visual cortex. *NeuroImage* 2019;198:125–36.
- [33] Seeliger K, Güçlü U, Ambrogioni L, Güçlütürk Y, Van Gerven MAJ. Generative adversarial networks for reconstructing natural images from brain activity. *NeuroImage* 2018;181:775–85.
- [34] Kuo PC, Chen YS, Chen LF, Hsieh JC. Decoding and encoding of visual patterns using magnetoencephalographic data represented in manifolds. *NeuroImage* 2014;102(Pt 2):435–50.
- [35] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521(7553):436–44.
- [36] Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw* 2015;61:85–117.
- [37] McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys* 1943;5(4):115–33.
- [38] Cox DD, Dean T. Neural networks and neuroscience-inspired computer vision. *Curr Biol* 2014;24(18):R921–9.
- [39] Cichy RM, Khosla A, Pantazis D, Torralba A, Oliva A. Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Sci Rep* 2016;6(1):27755.
- [40] Eickenberg M, Gramfort A, Varoquaux G, Thirion B. Seeing it all: convolutional network layers map the function of the human visual system. *NeuroImage* 2017;152:184–94.
- [41] DiCarlo JJ, Zoccolan D, Rust NC. How does the brain solve visual object recognition? *Neuron* 2012;73(3):415–34.
- [42] DiCarlo JJ, Cox DD. Untangling invariant object recognition. *Trends Cogn Sci* 2007;11(8):333–41.
- [43] Li J, Zhang Z, He H. Visual information processing mechanism revealed by fMRI data. In: *Proceedings of the 2016 International Conference on Brain and Health Informatics*; 2016 Oct 13–16; Omaha, NE, USA. Chem: Springer; 2016. p. 85–93.
- [44] Higgins I, Matthey L, Glorot X, Pal A, Uria B, Blundell C, et al. Early visual concept learning with unsupervised deep learning. 2016. arXiv:1606.05579.

- [45] Kingma DP, Welling M. Auto-encoding variational Bayes. 2014. arXiv:1312.6114.
- [46] Rezende DJ, Mohamed S, Wierstra D. Stochastic backpropagation and approximate inference in deep generative models. In: Ghahramani Z, Welling M, Cortes C, Lawrence ND, Weinberger KQ, editors. *Advances in neural information processing systems (NIPS 2014)* La Jolla: Neural Information Processing Systems Foundation. 2014. p. 1278–86.
- [47] Goodfellow I, Abadie JP, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: Ghahramani Z, Welling M, Cortes C, Lawrence ND, Weinberger KQ, editors. *Advances in Neural Information Processing Systems (NIPS 2014)* La Jolla: Neural Information Processing Systems Foundation. 2014. p. 2672–80.
- [48] St-Yves G, Naselaris T. Generative adversarial networks conditioned on brain activity reconstruct seen images. In: *Proceedings of the 2018 IEEE International Conference on System, Man, and Cybernetics*; 2018 Oct 7–10; Miyazaki, Japan. New York: IEEE; 2018.
- [49] Shen G, Dwivedi K, Majima K, Horikawa T, Kamitani Y. End-to-end deep image reconstruction from human brain activity. *bioRxiv* 2018:272518.
- [50] Kulkarni TD, Whitney WF, Kohli P, Tenenbaum J. Deep convolutional inverse graphics network. In: Cortes C, Lawrence ND, Lee DD, Sugiyama M, Garnett R, editors. *Advances in neural information processing systems 28 (NIPS 2015)* La Jolla: Neural Information Processing Systems Foundation. 2015. p. 2539–47.
- [51] Eslami SA, Heess N, Weber T, Tassa Y, Szepesvari D, Hinton GE, et al. Attend, infer, repeat: fast scene understanding with generative models. In: Lee DD, Sugiyama M, Luxburg UV, Guyon I, Garnett R, editors. *Advances in neural information processing systems 29 (NIPS 2016)* La Jolla: Neural Information Processing Systems Foundation. 2016. p. 3225–33.
- [52] Norman KA, Polyn SM, Detre GJ, Haxby JV. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci* 2006;10(9):424–30.
- [53] Isola P, Zhu J, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. 2016. arXiv:1611.07004.
- [54] Liu M, Breuel T, Kautz J. Unsupervised image-to-image translation networks. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al., editors. *Advances in neural information processing systems 30 (NIPS 2017)* La Jolla: Neural Information Processing Systems Foundation. 2017. p. 700–8.
- [55] Reed S, Akata Z, Yan X, Logeswaran L, Schiele B, Lee H. Generative adversarial text to image synthesis. 2016. arXiv:1605.05396.
- [56] Hong S, Yang D, Choi J, Lee H. Inferring semantic layout for hierarchical text- to-image synthesis. 2018. arXiv:1801.05091.
- [57] He D, Xia Y, Qin T, Wang L, Yu N, Liu T, et al. Dual learning for machine translation. In: Lee DD, Sugiyama M, Luxburg UV, Guyon I, Garnett R, editors. *Advances in neural information processing systems 29 (NIPS 2016)* La Jolla: Neural Information Processing Systems Foundation. 2016. p. 820–8.
- [58] Xia Y, Qin T, Chen W, Bian J, Yu N, Liu T. Dual supervised learning. In: *Proceedings of the 34th International Conference on Machine Learning*; 2017 Aug 6–11; Sydney, Australia. Brookline: Microtome Publishing 2017. p. 3789–98.
- [59] Xia Y, Tan X, Tian F, Qin T, Yu N, Liu T. Model-level dual learning. In: *Proceedings of the 35th International Conference on Machine Learning*; 2018 Jul 10–15; Stockholm, Sweden. Brookline: Microtome Publishing 2018. p. 5379–88.
- [60] Zhu J, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Proceedings of the 2017 IEEE International Conference on Computer Vision*; 2017 Oct 22–29; Venice, Italy. New York: IEEE; 2017. p. 2242–51.