Engineering 5 (2019) 948–953

Contents lists available at ScienceDirect

Engineering

journal homepage: www.elsevier.com/locate/eng

Research
Neural Information Engineering—Review

Brain Encoding and Decoding in fMRI with Bidirectional Deep Generative Models

Changde Du [a,b], Jinpeng Li [a,b], Lijie Huang [a,b], Huiguang He [a,b,c,*]

[a] Research Center for Brain-Inspired Intelligence and National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences (CAS), Beijing 100190, China
[b] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China
[c] 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

Keywords:
Brain encoding and decoding
Functional magnetic resonance imaging
Deep neural networks
Deep generative models
Dual learning

ABSTRACT

Brain encoding and decoding via functional magnetic resonance imaging (fMRI) are two important aspects of visual perception neuroscience. Although previous researchers have made significant advances in brain encoding and decoding models, existing methods still require improvement using advanced machine learning techniques. For example, traditional methods usually build the encoding and decoding models separately, and are prone to overfitting on a small dataset. In fact, effectively unifying the encoding and decoding procedures may allow for more accurate predictions. In this paper, we first review the existing encoding and decoding methods and discuss the potential advantages of a "bidirectional" modeling strategy. Next, we show that there are correspondences between deep neural networks and human visual streams in terms of the architecture and computational rules. Furthermore, deep generative models (e.g., variational autoencoders (VAEs) and generative adversarial networks (GANs)) have produced promising results in studies on brain encoding and decoding. Finally, we propose that the dual learning method, which was originally designed for machine translation tasks, could help to improve the performance of encoding and decoding models by leveraging large-scale unpaired data.

1. Introduction

The relationship between human visual experience and the evoked neural activity is central to the field of computational neuroscience [1,2]. Brain encoding and decoding via functional magnetic resonance imaging (fMRI) are important in gaining an understanding of the visual perception system [3–5]. An encoding model attempts to predict brain response based on a given visual stimulus [6,7], whereas a decoding model attempts to predict the corresponding visual stimulus by analyzing a given brain response [8–22]. Brain encoding and decoding (Fig. 1) have thus become two significant ways of promoting the development of sensory neuroscience because they provide many insights into brain function.

1.1. Encoding models

In the previous literature, most encoding models have been established based on specific computational rules. Neuroscientists believe that these computational rules may be the mathematical basis for the brain's response to specific visual stimuli. For example, Kay et al. [1] used pyramid-shaped Gabor wavelet filters to build an encoding model. Based on this encoding model, the authors successfully identified the preferred natural images for given human brain activities. Later, Kay et al. [6] further proposed a two-stage cascade encoding model based on the well-established local oriented filters, divisive normalization, compressive spatial summation, and variance-like nonlinearity. Recently, St-Yves and Naselaris [7] constructed a feature-weighted receptive field model based on the intermediate feature maps of a pre-trained deep neural network (DNN); this model can be used to predict the voxel response and study the shape of the receptive field of each voxel. Furthermore, Zeidman et al. [23] built a Bayesian population receptive field (pRF) model for interpretable brain encoding studies. In recent years, DNNs have achieved great success in computer vision, and researchers have begun to use DNNs to construct more complex brain encoding models [7,20,24]. In addition to encoding models for visual information, researchers have studied how semantic information is expressed in the brain. For example, Huth et al. [25] established the mapping relationship between text semantic

* Corresponding author.
E-mail address: huiguang.he@ia.ac.cn (H. He).

https://doi.org/10.1016/j.eng.2019.03.010

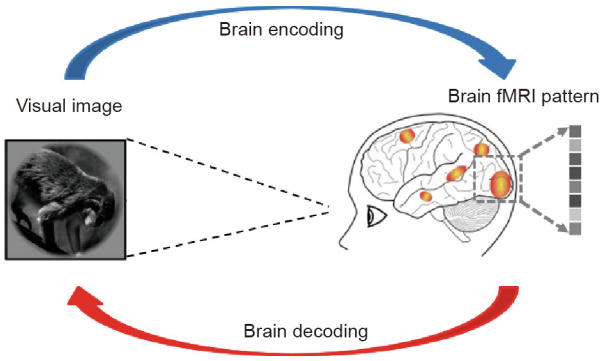

Fig. 1. Brain encoding and decoding in fMRI. The encoding model attempts to predict brain responses based on the presented visual stimuli, while the decoding model attempts to infer the corresponding visual stimuli by analyzing the observed brain responses. In practice, encoding and decoding models should not be seen as mutually exclusive. Effectively unifying encoding and decoding procedures may permit more accurate predictions and facilitate our understanding of information representation in the human brain.

vectors and cerebral cortex activities, thereby providing a detailed semantic map of the cerebral cortex.

1.2. Decoding models

Previous studies have demonstrated the feasibility of decoding the identity of binary contrast patterns [12–14], handwritten characters [15,16], human facial images [17–19], natural picture/video stimuli [2,20], and dreams [12,21] from the corresponding brain activation patterns. For example, Miyawaki et al. [12] constructed a multiscale neural decoding model to reconstruct perceived binary contrast patterns from brain responses. Schoenmakers et al. [15] proposed a linear decoding model to reconstruct handwritten characters from brain responses. Güçlütürk et al. [19] proposed the combination of probabilistic inference with adversarial training for reconstructions of perceived faces from brain responses. Horikawa and Kamitani [2] showed that the hierarchical features of visual stimuli calculated by a computer vision model could be predicted by utilizing the responses of multiple brain regions. These findings indicate that there is a close relationship between the hierarchical visual cortex and the complex visual features obtained by the computer vision model. Furthermore, Wen et al. [20] proposed a dynamic neural decoding method based on deep learning that can reconstruct the dynamic visual scenes perceived by a human and predict their semantic labels. Horikawa and Kamitani [21] even showed that brain activity could be used to predict the objects in humans' dreams.

Most of the aforementioned decoding studies are based on the multi-voxel pattern analysis (MVPA) method [8]. However, brain connectivity patterns are also a key feature of the brain state and can be used for brain decoding. Previous decoding studies [26–30] have shown that brain connectivity information can been utilized as distinguishing features in decoding procedures. For example, by employing brain connectivity information in brain decoding, Yargholi and Hossein-Zadeh [29] were able to successfully reconstruct two handwritten digits—namely, 6 and 9—from human brain activities. Manning et al. [30] proposed a probabilistic model for extracting dynamic functional connectivity patterns in brain activity. The proposed functional connectivity patterns can be used in brain decoding studies.

1.3. Hybrid encoding–decoding with bidirectional models

Although recent developments in brain encoding and decoding [3–21,29,31–33] have shown promising results, many challenges remain in constructing an accurate decoding model in order to reconstruct the corresponding visual stimuli from fMRI data. From the Bayesian machine learning perspective, an encoding model can be acquired with a generative model that accounts for the measured brain activity. When this encoding model is combined with prior knowledge about the stimuli, a posterior probability distribution of the stimuli—that is, a predictive distribution for decoding—could be obtained, given a brain activity pattern. Therefore, encoding and decoding models should not be seen as mutually exclusive. Effectively unifying encoding and decoding procedures may permit accurate predictions and facilitate an understanding of information representation in the human brain [13,34]. For example, Fujiwara et al. [13] proposed a "bidirectional" approach to visual image reconstruction, in which a set of latent variables was assumed to relate image pixels and fMRI voxels; this approach allowed predictions for both encoding and decoding to be generated. These scholars employed the Bayesian canonical correlation analysis (BCCA) framework, which computed multiple correspondences, via latent variables, between image pixels and fMRI voxels. Since the pixel weights for each latent variable can be thought to define an image basis, training the BCCA model using measured data leads to automatic extraction of image bases. Although it is premature to speculate on functional implications of the estimated image bases, this data-driven "bidirectional" approach could be extended to discover the modular architecture of the brain in representing complex natural stimuli, behavior, and mental experience defined in high-dimensional space.

2. Correspondence between DNNs and the human visual system

Deep learning [35,36] is a large class of machine learning methods that extract hierarchical representations from input data. The architectures of DNN were first inspired by the structure and computational principles of the biological nervous system [37]. Recently, DNN-based deep learning methods have achieved great success in image recognition, speech recognition, natural language understanding, and other aspects. In terms of architecture, the hierarchical layers of DNNs are very similar to those of the ventral visual system of the human brain [7,35,38] (Fig. 2). In terms of function, existing research on neural encoding and decoding based on deep learning has shown that the shallow representation of DNN is similar to the function of the primary visual area, while the deep representation of DNN is similar to the back end of the ventral visual system [2,24,39,40].

Humans can perceive complex objects quickly and accurately through the ventral visual stream, a system of interconnected brain regions that processes increasingly complex features in hierarchical structures [41,42,43]. However, the automated discovery of early visual concepts from visual images with no supervised information is a major open challenge in machine perception research. On the one hand, it would be helpful or the representations extracted from the image to perform well in real-world tasks. On the other hand, it would be desirable to be able to interpret these representations, and for them to be useful for tasks beyond those that are explicit in their initial design. From a traditional standpoint, it is difficult to use a pre-trained DNN model to learn such representations from visual images, because the semantic meaning of each dimensionality in the representation vector automatically extracted from the input image by that DNN model is unknown. Without disentangled representations, it is difficult to interpret these representations across different tasks. Fortunately, Higgins et al. [44] have shown that specially designed deep generative models are capable of learning disentangled representations.

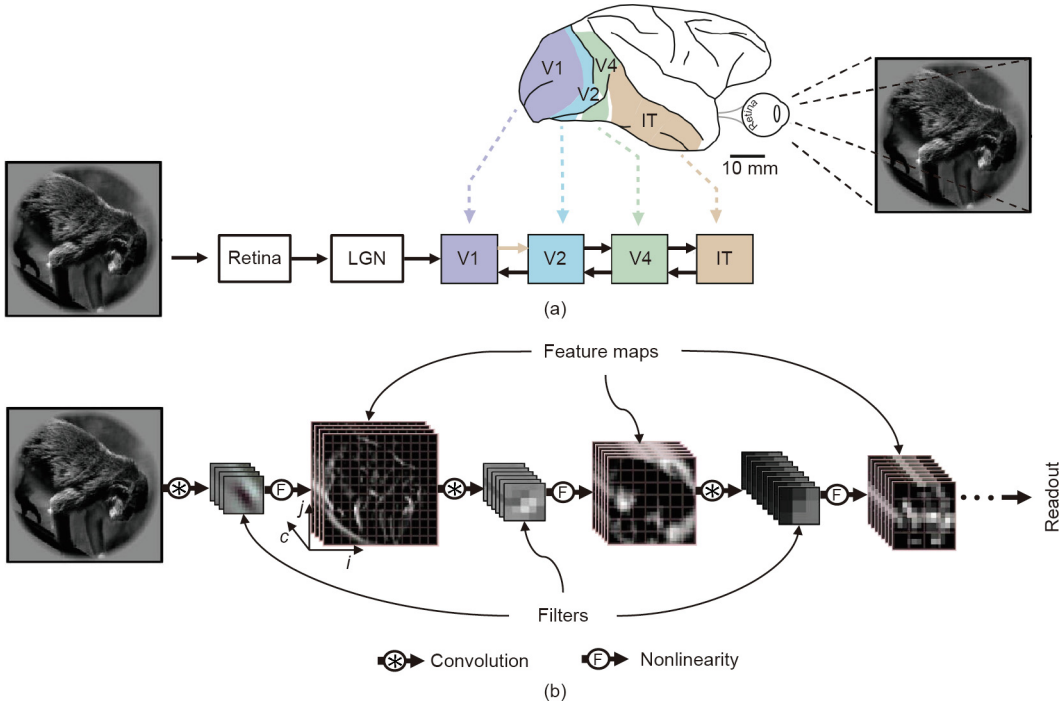

Fig. 2. The ventral visual system and a deep convolutional neural network (CNN). (a) Forward and backward projections between four Brodmann areas (V1, V2, V4, and IT); (b) an illustration of a simple feedforward deep CNN, whose hierarchical structure is used to simulate the hierarchical representation of the ventral visual system. LGN: lateral geniculate nucleus. (a) Reproduced from Ref. [38] with permission of Elsevier, © 2014; (b) reproduced from Ref. [7] with permission of Elsevier, © 2017.

3. Brain decoding with deep generative models

A promising research direction involves the integration of deep learning methods into brain decoding research. Deep generative models such as variational autoencoders (VAEs) [45,46] and generative adversarial networks (GANs) [47] have achieved great success in the field of image generation. An increasing amount of attention has recently been focused on research on visual image reconstruction using deep generative models [19,31–33,48,49].

3.1. VAE-based methods

VAEs—which were originally presented in Refs. [45,46]—are a probabilistic extension of the autoencoder model. A VAE has a bottom-up encoding network and a top-down decoding network. These two networks are jointly trained to maximize the lower bound of the data likelihood, thereby reformulating the autoencoder model as a variational inference problem. Recent works have demonstrated that VAE-based models are capable of learning disentangled representations that correspond to distinct factors of variation in the input data [43,50,51]. This is very important for brain encoding and decoding tasks, since some of the visual concepts learned by VAE-based models are also perceived by the human brain. Inspired by this fact researchers have explored the use of VAE-based models in image reconstruction from brain activities [31,32].

For example, Du et al. [31] proposed a deep generative multiview model (DGMM) for reconstructing the perceived images from brain fMRI activities (Fig. 3). The DGMM can be viewed as a nonlinear extension of the linear BCCA. Under the DGMM framework, the encoding and decoding procedures are simultaneously formulated by two distinct generative models:

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

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

where $\mathcal{N}$ denotes the normal distribution, $X \in \mathbb{R}^{D_x \times N}$ denotes the visual images, $Y \in \mathbb{R}^{D_y \times N}$ denotes the evoked fMRI activities, $p_\theta(X|Z)$ is the likelihood function of the visual images with neural network parameters θ, $p(y|z)$ is the likelihood function of the evoked fMRI activities, ψ denotes the full covariance matrix, B denotes the projection weights of the fMRI activities, and $Z \in \mathbb{R}^{D_z \times N}$ denotes the shared latent variables between the visual images and the evoked fMRI activities. The μ_x and σ_x^2 denote the mean and covariance of that normal distribution, respectively, and they are obtained by different nonlinear transformations with respect to the latent variables. The training set consists of N paired samples, which can be denoted by $(x_1, y_1), \ldots, (x_n, y_n)$, where $x_i \in \mathbb{R}^{D_x}$ and $y_i \in \mathbb{R}^{D_y}$ for $i = 1, \ldots, N$. Specifically, the DGMM uses a DNN-based generative process to model the distribution of visual images, while using a sparse linear generative process to model the distribution of brain response data. On the one hand, the DNN used here can effectively capture the hierarchical features of the visual image, which are similar to the structure of the ventral visual system of the human brain [2,24,39,40]. On the other hand, the sparse linear generative model used here not only conforms to the sparse expression principle of the human brain, but also avoids overfitting of brain response data [52]. Note that these two generative processes share the same latent variables. Therefore, in the test phase, the use of these processes makes it possible to infer the corresponding visual image from the brain response through the shared latent variables. In fact, the DGMM framework can capture "bidirectional" mapping relationships between the visual images and the corresponding fMRI activities. Thanks to its autoencoding variational Bayesian architecture, the DGMM can be optimized efficiently by means of mean-field variational inference, which is similar to the classical VAE solution. Compared with non-probabilistic deep multi-view learning

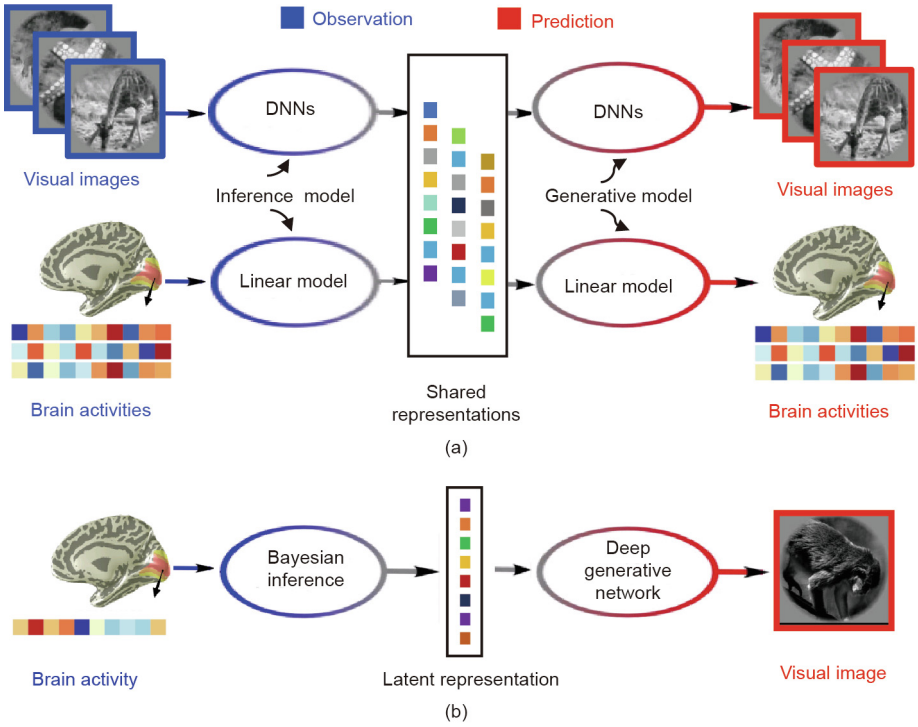

Fig. 3. Illustration of the deep generative multi-view framework for neural decoding. (a) Model training: view-specific generative models are used for data generation; specifically, a DNN is adopted to model visual images, while a linear regression model is used to model brain activities. (b) Image reconstruction: brain activities that are independent of those used for training are decoded into visual images.

methods, the DGMM's Bayesian framework makes it naturally more flexible and adaptive.

3.2. GAN-based methods

GANs were first proposed in Ref. [47]. The basic GAN is an unsupervised model that generates images from a noise vector. The idea of adversarial training comes from game theory, in which two competitors compete in order to make progress together. The typical configuration of a GAN includes a generator and a discriminator. The task of the generator is to synthesize images from noise in order to deceive the discriminator into believing that the synthesized images are real-world scenes. Meanwhile, the discriminator attempts to distinguish between the synthesized data and real data. When the Nash equilibrium is reached, the generator learns the distribution of real-world images, and the discriminator is sensitive to capturing the difference between real and fake data. GANs have been widely used in various applications, including image generation [53], image-to-image translation [54], and text-to-image synthesis [55,56].

Unlike a VAE, a GAN is a likelihood-free model—that is, it does not make any prior assumptions regarding the data distribution, and the data distribution is totally learned through adversarial training. This is a favorable characteristic in neural encoding and decoding tasks. A GAN often requires exact semantic information flow in its generator and discriminator. However, the useful semantic information in the blood-oxygen-level-dependent (BOLD) signal is merged deep in noise, which is a great challenge for model training. Recent brain decoding research [19] has proposed the combination of probabilistic inference with adversarial training for the reconstruction of perceived faces from brain activations (Fig. 4). Assume that $x \in \mathbb{R}^{h\times w\times c}$ is the visual image, $z \in \mathbb{R}^p$ is its latent features, $y \in \mathbb{R}^q$ is the corresponding brain response, and $\varphi \in \mathbb{R}^{h\times w\times c} \rightarrow \mathbb{R}^p$ is a latent feature model such that $z = \varphi(x)$ and $x = \varphi^{-1}(z)$. Then, the perceived visual images can be reconstructed from brain responses by means of the following equation:

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

where $p(z|y)$ is the posterior distribution of the latent variables. Eq. (3) can be reformulated through Bayes' theorem:

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

where $p(y|z)$ is the likelihood function and $p(z)$ is the prior distribution of the latent variables. The authors first intuitively decode the observed brain responses to the latent features with maximum *a posteriori* estimation. Next, they generate the perceived images according to the decoded latent features using adversarial learning. This two-step brain decoding method can accurately generate reconstructions of perceived faces from brain responses. More recently, researchers have attempted to reconstruct natural images from measured fMRI signals [33,48,49] by utilizing GANs that have been pre-trained on large-scale image datasets.

4. Improving brain encoding and decoding with dual learning

Data-driven brain encoding and decoding methods often require the acquisition of a large number of paired (stimulus-response) data instances in order to train a model that is customized to an individual subject. In many encoding and decoding studies, however, it is possible to gather a few thousand noisy paired data instances—at most—from a single subject. To improve the generalization ability of the encoding and decoding models, it is therefore necessary to make good use of large-scale unpaired data instances (e.g., visual images).

Inspired by recently proposed dual learning for machine translation [57,58], we suggest that it is possible to train encoding and

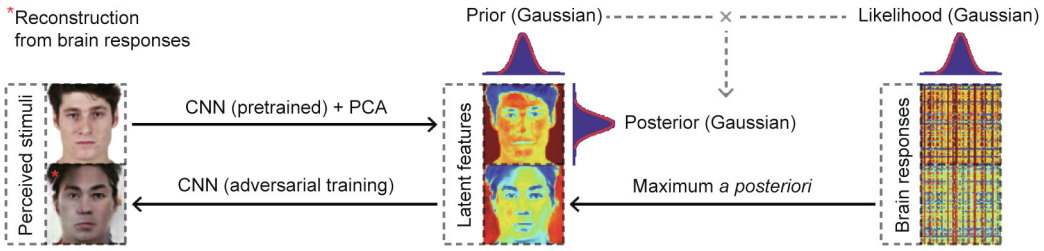

Fig. 4. Illustration of deep adversarial neural decoding. By combining probabilistic inference with adversarial learning, this method can clearly reconstruct the corresponding image of a face from brain activity. PCA: principal component analysis. Reproduced from Ref. [19] with permission of Neural Information Processing Systems Foundation, Inc., © 2017.

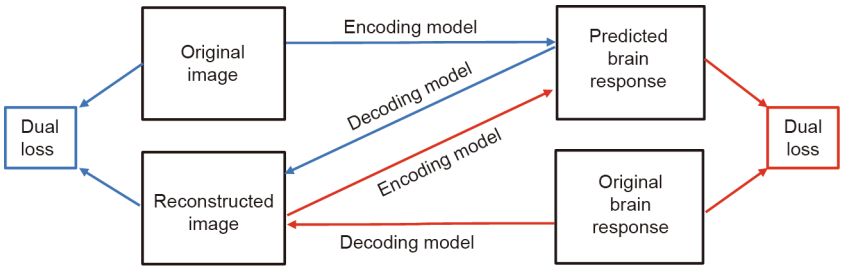

Fig. 5. Improving brain encoding and decoding with dual learning. Dual loss measured over unpaired data (either visual images or brain responses) generates informative feedback to train the bidirectional mapping model. Under this dual learning framework, it is possible to leverage large-scale unpaired data to improve the models' generalization ability.

decoding models simultaneously by minimizing the reconstruction loss resulting from the bidirectional mapping model. The encoding and decoding models represent a primal-dual pair and form a closed loop, allowing the application of dual learning (Fig. 5). Specifically, the reconstruction loss measured over unpaired data (e.g., visual images) would generate informative feedback to train the bidirectional mapping model. Under this dual learning framework, it is possible to leverage large-scale unpaired visual images to improve the generalization ability of the encoding and decoding models. In fact, dual learning is a general framework for learning the bidirectional mappings from one data domain M_d to another data domain N_d [59,60]. For $M_d \rightarrow N_d$, the goal is to learn an encoder mapping E such that the distribution $E(M_d)$ is indistinguishable from the distribution N_d using an adversarial loss. Similarly, for $N_d \rightarrow M_d$, the goal is to learn a decoder mapping D such that the distribution $D(N_d)$ is indistinguishable from the distribution M_d using another adversarial loss. In particular, for the paired data, it is possible to combine these two adversarial losses and the cycle consistency losses (dual losses) to push $D[E(M_d)] \approx M_d$ and $E[D(N_d)] \approx N_d$.

5. Conclusions

In conclusion, brain encoding and decoding are central to the field of computational neuroscience and have the potential to create better brain-machine interfaces. The architecture and computational rules of DNNs share some similarity with human visual streams. The use of deep generative models (e.g., VAEs and GANs) in brain encoding and decoding studies holds promise for providing deeper insight into relationships between human visual experience and the evoked neural activity. By leveraging large-scale unpaired data, dual learning is expected to play an important role in developing neural encoding and decoding models.

Acknowledgements

This work was supported by the National Key Research and Development Program of China (2018YFC2001302), National Natural Science Foundation of China (91520202), Chinese Academy of Sciences Scientific Equipment Development Project (YJKYYQ20170050), Beijing Municipal Science and Technology Commission (Z181100008918010), Youth Innovation Promotion Association of Chinese Academy of Sciences, and Strategic Priority Research Program of Chinese Academy of Sciences (XDB32040200).

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, Lawrenc 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, Lawrenc 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. Front Comput Neurosci 2019;13:21.
[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.