

# Neural Decoding of Visual Information Across Different Neural Recording Modalities and Approaches

Yi-Jun Zhang<sup>1</sup> Zhao-Fei Yu<sup>2,3</sup> Jian. K. Liu<sup>4</sup> Tie-Jun Huang<sup>2,3,5</sup>

<sup>1</sup>Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

<sup>2</sup>School of Computer Science, Peking University, Beijing 100190, China

<sup>3</sup>Institute for Artificial Intelligence, Peking University, Beijing 100190, China

<sup>4</sup>School of Computing, University of Leeds, Leeds LS2 9JT, UK

<sup>5</sup>Beijing Academy of Artificial Intelligence, Beijing 100190, China

**Abstract:** Vision plays a peculiar role in intelligence. Visual information, forming a large part of the sensory information, is fed into the human brain to formulate various types of cognition and behaviours that make humans become intelligent agents. Recent advances have led to the development of brain-inspired algorithms and models for machine vision. One of the key components of these methods is the utilization of the computational principles underlying biological neurons. Additionally, advanced experimental neuroscience techniques have generated different types of neural signals that carry essential visual information. Thus, there is a high demand for mapping out functional models for reading out visual information from neural signals. Here, we briefly review recent progress on this issue with a focus on how machine learning techniques can help in the development of models for contending various types of neural signals, from fine-scale neural spikes and single-cell calcium imaging to coarse-scale electroencephalography (EEG) and functional magnetic resonance imaging recordings of brain signals.

**Keywords:** Neural decoding, machine learning, deep learning, visual decoding, brain-inspired vision.

**Citation:** Y. J. Zhang, Z. F. Yu, J. K. Liu, T. J. Huang. Neural decoding of visual information across different neural recording modalities and approaches. *Machine Intelligence Research*, vol.19, no.5, pp.350–365, 2022. <http://doi.org/10.1007/s11633-022-1335-2>

## 1 Introduction

Every day, various types of sensory information from the external environment are transferred to the brain through different modalities and then processed to generate a series of coping behaviours. Among these perceptual modalities, vision is arguably the dominant contributor to the interactions between the external environment and the brain. Approximately 70 percent of human perception information is derived from vision<sup>[1]</sup>, far more than the auditory system, tactile system, and other sensory systems combined. The visual system is the part of the central nervous system that is required for visual perception, processing, and interpreting visual information to build a representation of the visual environment. It consists of the eye, retina, fibers that conduct visual information to the thalamus, the superior colliculus, and parts of the cerebral cortex. Today, researchers can collect neural

signals using different recording modalities, e.g., spikes, electroencephalography (EEG), and functional magnetic resonance imaging (fMRI), from brain activity in different parts of the visual system, such as the retina, lateral geniculate nucleus (LGN), and primary visual cortex (V1) cortex, etc. Depending on the corresponding collecting devices, different recording modalities differ in their invasiveness, scale, and precision.

Neural coding is an important topic for understanding how the brain processes stimuli from the environment<sup>[2, 3]</sup>. The aim of neural decoding is to read out information embedded in various types of neural signals<sup>[4]</sup>. As for vision, understanding how neurons perceive and respond to rich natural visual information is a major topic of neural encoding<sup>[5, 6]</sup>, whereas, the goal of neural decoding of visual information is to restore the original stimulus from neural responses as much as possible<sup>[7]</sup>, as shown in Fig. 1. It is also critical to the development of artificial vision used by brain-computer interfaces and virtual reality devices.

Much effort has been made to study the various mechanisms underlying neural decoding in the visual pathway in recent decades<sup>[8–12]</sup>. These mechanisms can be roughly divided into three categories depending on the decoding type: 1) Visual stimulus classification, in which a specific

Review

Special Issue on Brain-inspired Machine Learning

Manuscript received February 7, 2022; accepted April 27, 2022; published online July 14, 2022

Recommended by Associate Editor Chun-Hua Shen

Colored figures are available in the online version at <https://link.springer.com/journal/11633>

© The Author(s) 2022

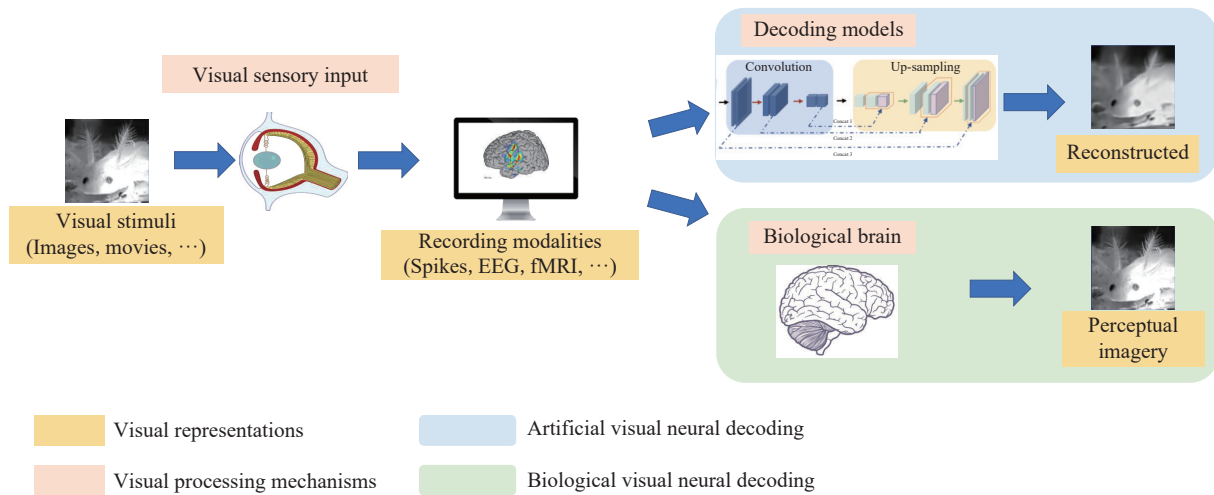


Fig. 1 Illustration of two types of visual neural decoding

stimulus is classified into the best-matched image set; 2) Visual stimulus identification, in which the stimulus is identified with a specific visual object; 3) Visual stimuli reconstruction, in which the corresponding visual stimulus is reconstructed in accordance with the resulting neural responses. Most decoding approaches have depended on linear methods due to their interpretability and computational efficiency<sup>[9, 13, 14]</sup>. Although linear decoding methods are capable of decoding spatially uniform white noise stimuli and the coarse structure of natural scene stimuli from neural responses<sup>[13, 14]</sup>, the recovery of the fine visual details of naturalistic images is difficult for these types of methods. The most recent decoders utilized nonlinear methods for the fine decoding of complex visual stimuli. For instance, optimal Bayesian decoding was leveraged for white noise stimuli, but achieved limited generalizability to a large neural population<sup>[15]</sup>. For natural scene image structures, key prior information was used to perform computationally expensive approximations to Bayesian inference<sup>[16, 17]</sup>. Some researchers have combined linear and nonlinear approaches to generate coarse reconstructions of natural stimuli through calcium imaging data<sup>[10, 18–20]</sup>. Additionally, many researchers have begun to successfully use deep learning techniques for visual neural decoding, leading to the great achievement in artificial vision<sup>[11, 21–24]</sup>.

Visual neural decoding is a significant issue that can help advance engineering applications such as brain-machine interfaces and a more holistic understanding of the brain in neuroscience. Considering the rapid developments of related techniques in visual neural decoding, there is a strong demand for a comprehensive and up-to-date review in this field. In this review, we sorted out the research evolution in visual neural decoding. Various neural recording modalities are introduced in this review, especially for the emerging calcium imaging data. We summarized the advantages and disadvantages of different neural decoding methods. In addition, open resources,

including public neural data and software toolkits, are also provided for the convenience of neural decoding research. Finally, we conclude with our perspective on the open challenges and future directions for the outlook in this study. We aim to provide a review of neural decoding in visual systems that could serve as an inspiration to both neuroscience and multidisciplinary researchers looking to understand the state-of-the-art and current problems in neural decoding, especially regarding the development of artificial intelligence and brain-like vision systems.

## 2 Task evolution in visual neural decoding

Visual neural decoding has been a core topic in computational neuroscience in recent decades. To date, the history of visual neural decoding has involved several stages from the perspective of decoding tasks. Advancements in recording devices of different modalities and increasing efforts in decoding research have allowed more challenging tasks that have been accomplished with researchers' endeavors. Visual neural decoding methods can be coarsely divided into three categories: image classification, image identification, and image reconstruction. In image classification, the stimulus that evoked the neural response recording is classified into a specific stimulus set comprised of similar stimuli. For example, Kamitani and Tong<sup>[25]</sup> successfully predicted which of the eight stimulus orientations the subject was seeing from individual fMRI signals recording trials. Yarholi and Hosseinzadeh<sup>[26]</sup> used an augmented naïve Bayesian classifier to decode the fMRI signals generated in response to handwritten digits into correct classes. In image identification, one stimulus is identified as the specific image (usually the most similar image) according to the corresponding response; this category is seen as more challenging than classification. Kay et al.<sup>[27]</sup> proposed receptive-field mod-

els for different voxels that allowed the identification of the specific image the subject was seeing from a large set of completely novel natural images. Horikawa and Kamitani<sup>[28]</sup> identified the object seen in the most similar category according to a feature vector predicted by measured fMRI activity. Image reconstruction typically involves methods that perform pixel-level reconstructions of stimuli from the recorded responses, which is evidently more difficult than the above two decoding tasks and also the recent focus of neuroscience researchers. The deep learning neural network (DNN) for visual decoding is the most frequently proposed approach to solve the reconstruction problem in recent years<sup>[12]</sup>.

Image classification and identification have been forms of decoding tasks in the early stages of development in the field of neural decoding. In recent years, researchers in the field of neural decoding have mainly regarded reconstruction as the task goal of neural decoding (see Section 4). In order to reflect the latest developments in the field of neural decoding, in this review, we mainly focus on the reconstruction task in neural decoding. Except for special emphasis, the literature and methods mentioned in this paper are generally oriented toward reconstruction tasks of neural decoding. Those emerging decoding tasks beyond reconstruction will be discussed in future directions in Section 6.

### 3 Neural recording modalities

As the input to various types of neural decoders, it is important to understand the characteristics of the signals obtained with neural recording modalities, as shown in Fig 2. The differences in signal types, data structures, and spatial and temporal resolutions, which are summarized in Fig. 3, have a great influence on the design of visual neural decoders.

#### 3.1 Spikes

Action potentials are fast electrical changes that are triggered when the membrane potential of individual neurons depolarizes their thresholds that can be meas-

ured as consistent waveforms (called “spikes”)<sup>[29]</sup>. Spike events are usually detectable with microelectrodes or microelectrode arrays, with a high temporal resolution decoding on a millisecond timescale. Spikes are the most common invasive recording methods, in which electrodes are inserted into or adjacent to neurons to record voltage events<sup>[30–32]</sup>. Because a large number of electrodes need to be implanted in parallel into a single probe in order to record a large number of neurons simultaneously, the neural information collected from subjects depends on the sophistication of the underlying electrophysiology technology<sup>[33, 34]</sup>.

In certain neural decoding systems based on spike signals, spike events are usually converted to firing rates (in a fixed time bin), as shown in Fig. 3, which are then fed into neural decoders<sup>[23]</sup>. The temporal coding of spike events is rarely seen in neural decoding models based on spike signals. Some dimensionality reduction methods are used prior to decoding when a large number of neurons are recorded prior to decoding<sup>[35–37]</sup>. Such dimensionality reduction methods range from the classical machine learning methods, e.g., principal component analysis, to deep learning algorithms.

Recent advances in neural decoding with spike signals collected from visual systems have benefitted from relevant recording technology, such as Neuropixels probes and the increased availability of public datasets<sup>[1, 38–41]</sup>. Zhang et al.<sup>[23]</sup> used spike signals collected from salamander retinas to recover the dynamic movie stimuli. Iqbal et al.<sup>[42]</sup> used a deep neural network to decode natural stimuli from spike responses collected in the mouse cortex. Leveraging the development of large-scale multielectrode recording systems, Kim et al.<sup>[43]</sup> applied the spike signals of 2 000 retinal ganglion cells to develop a multistage decoding approach that exhibited improved accuracy over linear methods. Xu et al.<sup>[44]</sup> proposed a deep spike pattern decoder, which was not only capable of perceiving inputs from noisy environments, but also had a good reconstruction ability of generalizability to images, fMRI brain activities, and sound signals.

Spike signals are still the most widely used neural

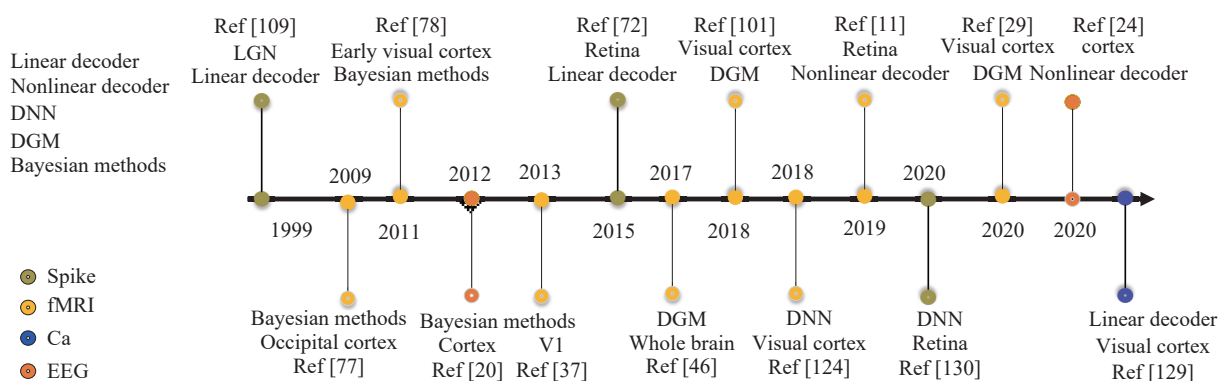


Fig. 2 Key methods in the history of visual neural decoding. On the left, different colored circles represent different neural recording modalities. The decoding methods are indicated with different colored text.

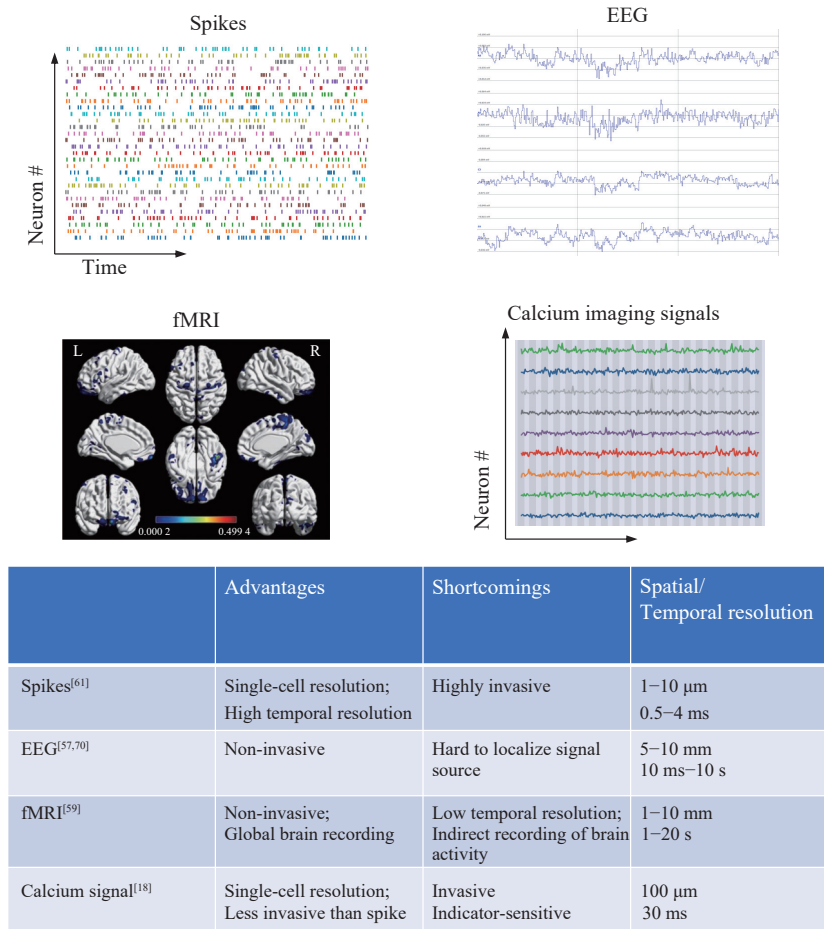


Fig. 3 Multiple recording neural modalities

data in visual neural decoding, and consequently, studies on visual neural decoding based on spike signals are predominant in this field. Signals from other recording modalities can be transferred to spike signals first and then decoded with appropriate neural decoders designed for spikes. For example, in some visual decoding studies, calcium imaging traces are converted to spike events with transcoding algorithms, which are then decoded to extract stimulus information (see Section 3.4).

### 3.2 EEG

EEG is a low-cost, non-invasive neuroimaging technique that provides high temporal resolution recordings of brain activity, and consequently, has been widely applied in various fields, e.g., brain-computer interface systems and brain activity monitoring. However, EEG has limitations resulting from the recording technique. As illustrated in Fig. 3, the complex, high-dimensional EEG signals usually have a low signal-to-noise ratio and nonlinearity and nonstationary properties, resulting in signals that require artificial removal<sup>[45–48]</sup>.

Due to the complexity and information redundancy of EEG, one critical component of EEG neural decoding is the extraction of features from the original EEG signals.

Researchers have proposed a series of feature extraction methods that operate from different aspects of the signal. For example, methods focusing on the time-domain aspects of the signal include independent component analysis and autoregressive models; those investigating frequency-domain information from EEG signals include fast Fourier transform and Welch’s method; and wavelet transform and short-time Fourier transform can be used to conduct time-frequency domain analyses. Recent studies have analyzed the advantages and shortcomings of each feature extraction method, noting that the appropriate feature extraction method must be chosen according to the specific type of task<sup>[49–53]</sup>.

One general decoding goal of EEG acquisition is classification. The EEG classification process pipeline involves data preprocessing and splitting the dataset into a training set for training the classifiers and a test set for predicting the new data classes. In addition to conventional algorithms, e.g., support vector machine and linear discriminant analysis, various machine learning and deep learning algorithms have been used for EEG classification tasks, e.g., motor imagery processing, motion recognition and attention disorder classification<sup>[54–56]</sup>. Beyond these fields, various computational neural decoding models have been proposed<sup>[57–60]</sup>. Garofalo et al.<sup>[61]</sup> found that

the event-related potential signals of prediction errors are modulated by the likelihood of occurrence based on the predicted response-outcome model.

In addition to the classification decoding tasks, EEGs have been proposed for use with inverted encoding models which attempt to reconstruct the contents of memory or attentional focus of EEG brain activity<sup>[62–66]</sup>.

### 3.3 fMRI

fMRI is the most frequently used non-invasive recording modality, in human decoding experiments. As shown in the fMRI recording of Fig. 3, blood oxygenation is measured by fMRI and utilized as a proxy of neural activity. Moreover, fMRI allows the simultaneous recording of whole-brain blood oxygenation-level dependent activity, providing the largest scale coverage among all recording approaches as well as insight into brain function study. fMRI represents the recorded signals in different “voxels” (locations) of the brain during the experiment<sup>[67, 68]</sup>.

Univariate methods were typically used for fMRI brain activity pattern analysis in the early stage, employing a general linear model that was used to estimate each voxel in the brain separately in most univariate methods<sup>[69]</sup>. However, Univariate methods depend on a uniform relation between neural activity and the investigated function, in both individual voxels and across participants, which makes it difficult to detect spatial patterns<sup>[70]</sup>. To overcome this limitation, multivariate pattern analysis was proposed, capable of detecting the activation distribution of the brain and accurately decoding the cognitive state of the patient<sup>[71]</sup>. Thus, multivariate methods are widely used to train classifiers on unsmoothed, voxelwise, patterns of brain activity in different conditions.

In addition, some researchers used their own handcrafted methods to process fMRI data in visual decoding. For example, Zhang et al.<sup>[23]</sup> unravelled the voxel data as a one-dimensional vector to input them into a deep neural network decoder. Du et al.<sup>[72]</sup> assumed that the correlations among fMRI voxels could naturally reflect the characteristics of the corresponding visual stimuli and used a full-covariance matrix to capture these correlations in their decoding experiments.

### 3.4 Calcium imaging signal

Calcium imaging is another invasive technique for recording the activity of activated neurons<sup>[73]</sup>. In calcium imaging data, neuronal activity is measured by the fluorescence intensity of calcium indicators, which is captured by an electron microscope; these calcium indicators are usually chosen according to the type of neuron. As the traces are shown in Fig. 3 calcium signals part, the raw outputs of two-photon electron microscopes in the calci-

um imaging experiment are videos, which record the locations and fluorescence changes of activated neurons during the measuring progress<sup>[74]</sup>. Usually, traces of fluorescence changes in calcium indicators are extracted from the recording video for further analysis<sup>[75]</sup>.

Visual decoding from calcium imaging data is a relatively recent and understudied field in visual information processing. Nevertheless, several methods have been proposed by calcium imaging researchers. Previous work has focused on classification algorithms. For example, Grewe et al.<sup>[78]</sup> attempted to use four machine learning structures to classify the V1 responses of natural stimuli according to the calcium traces acquired during the selected frames. A series of studies converted calcium imaging traces to spike events<sup>[75–81]</sup>, and their findings have been leveraged to conduct several neural decoding studies based on calcium imaging data<sup>[19, 20]</sup>. For example, Garasto et al.<sup>[19]</sup> conducted a pixel-by-pixel reconstruction of a complex natural stimulus from the spike count estimated from the calcium imaging response of the mouse primary visual cortex (V1). In some recent work, the standard ratio of fluorescence change (usually depicted as  $dF/F_0$  form) of each of the regions of interest (ROIs) during a blank screen prior ( $F_0$ ) to stimulus presentation ( $F$ ) was calculated as the neuron’s calcium imaging response. Tang et al.<sup>[82]</sup> evaluated how well the sparse calcium imaging response allowed a decoder to discriminate 2 250 natural stimuli. Yoshida and Ohki<sup>[10]</sup> believed that natural images could be reliably reconstructed by sparse calcium imaging population responses in the mouse visual cortex.

## 4 Decoding approaches

As illustrated in Fig. 1, the visual neural decoding problem can be formulated as the identification of stimulus  $x = (x_1, \dots, x_N)^T$  that produces the observed brain activity  $y = (y_1, \dots, y_M)^T$  best. The stimulus in visual neural decoding tasks is usually presented as an image, one movie frame, etc. Neural decoding specifically refers to predicting  $x$  from  $y$ , in which the representation of the external environment in neural recordings is prospectively revealed.

### 4.1 Linear decoding methods

In recent decades, various methods have been developed. Fig. 2 illustrates a timeline of some key studies in the history of visual neural decoding. Traditionally, a neural decoder can be optimized with linear and nonlinear statistical methods<sup>[8–11]</sup>. In linear decoding methods, the relationship between stimulus set  $X$  and neural response set  $Y$  can be formulated as follows:

$$X = YW. \quad (1)$$

The optimal weight matrix for the linear decoding model  $W$  is calculated using least squares regression:

$$W_{op} = (Y^T Y)^{-1} Y^T X. \quad (2)$$

And the weights  $W_{op}$  are then used to reconstruct the stimulus in a held-out set of brain activity datasets.

For instance, Schoenmakers et al.<sup>[83]</sup> reconstructed BRAINS characters from measurements of human brain activity using the linear reconstruction approach. Linear decoding methods are derived from the reverse of early-stage linear encoding models in neuroscience. Due to the limited representation power of these structures, the decoding tasks were usually simple and the performance in natural image reconstructions was not satisfactory. To overcome this limitation, Brackbill et al.<sup>[14]</sup> combined nonlinear and linear reconstructions to develop a cascade decoding model for retinal ganglion cell (RGC) natural stimulus-response signals.

Most early approaches to visual neural decoding have depended on linear methods due to their interpretability and computational efficiency. Nevertheless, the limited representation power of linear methods makes it difficult to reconstruct the fine visual details of natural images.

## 4.2 Bayesian-based decoding methods

More recent visual decoding studies have incorporated nonlinear methods for better reconstruction performance with complex natural stimuli, as illustrated in Fig. 2. Among these studies, a series of Bayesian methods have been proposed to explore the correlations between neural recording signals and visual stimuli.

In a typical Bayesian-based method, the visual neural decoding problem can be formulated as the following maximum a posteriori probability (MAP) estimation problem:

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} p(\mathbf{x} | \mathbf{y}). \quad (3)$$

Here,  $\hat{\mathbf{x}}$  is the best matching stimulus among all  $\mathbf{x}$ .  $\mathbf{y}$  represents the recorded brain activity signals. Specifically, the stimuli  $\mathbf{x}$  can be then presented as images or movie frames, for example. The interpretation of modality  $\mathbf{y}$  depends on the specific neural recording technique used in the physiological experiment. Typically, for the sake of simplification, both the stimulus and the response are assumed to be zero mean, standard Gaussian distributions. In practice, let the acquired dataset  $\mathcal{D} = \left\{ \left( \mathbf{x}^{(t)}, \mathbf{y}^{(t)} \right) \right\}_{t=1}^T$  consists of  $T$  stimulus-response pairs. Then, by plugging in the predictive density, the MAP solution can be formulated as

$$p(\mathbf{x} | \mathbf{y}, \mathcal{D}) = \int_{\theta} p(\mathbf{x} | \mathbf{y}, \theta) p(\theta | \mathcal{D}) \quad (4)$$

and further simplified to  $p(\mathbf{x} | \mathbf{y}, \mathcal{D}) = p(\mathbf{x} | \mathbf{y}, \hat{\theta})$ , where  $\hat{\theta} = \arg \max_{\theta} p(\theta | \mathcal{D})$ . Notably, in practice, many unique stimuli in  $\mathcal{D}$  are used to sample the stimulus space accurately, and the average value of responses to the same stimulus is used as one of the  $\mathbf{y}$  values to weaken the influence of noise.

Naselaris et al.<sup>[16]</sup> presented a Bayesian framework for accurate reconstructions of the spatial structure of natural images. However, in their framework, a reconstruction was defined as the image with the highest posterior probability that was thought to evoke the corresponding brain activities. Nishimoto et al.<sup>[17]</sup> constructed a Bayesian decoder with a sampled natural movie prior to reconstructing a movie stimulus from fMRI signals. In their framework, a motion-energy encoding model was presented to match the slow fMRI responses to dynamic stimuli.

Gallant first used a Bayesian reconstruction model to decode fMRI data from early visual areas, and the resulting reconstruction is a predefined natural image, selected according to its posterior probability<sup>[16, 17, 84]</sup>. Fujiwara et al.<sup>[84]</sup> proposed a Bayesian canonical correlation analysis in which image bases were automatically learned, and an invertible mapping was obtained between brain activity and image bases. Recently, Du et al.<sup>[72]</sup> leveraged the latent variables inferred by Bayesian reconstruction to capture the correlations among the voxel activities of fMRI signals; this resulted in an increased reconstruction accuracy.

As mentioned above, Bayesian decoding models usually outperformed simple linear decoding models in visual neural decoding tasks. However, there are also some constraints for Bayesian decoding methods. The Bayesian decoding methods usually have to resort to the specific prior information encoded by a specifically designed model. The determination of parameters in the overall decoding process needs to be elaborated. Furthermore, the mapping between the visual stimuli and the corresponding brain activity determined by Bayesian methods does not typically describe the relationship between these two cross-modal data. Consequently, fine natural image details are difficult to be reconstructed with this type of method.

## 4.3 Deep neural network methods

As illustrated in Fig. 2, deep learning techniques, especially deep neural networks, have been implemented in the neural decoding field in recent years. Some studies have revealed that the essential DNN mechanism corresponds to the human visual cortex<sup>[28, 85–87]</sup>. Some designs of essential blocks in deep learning have been inspired by developments in neuroscience. For example, the concept of simple and complex cells in the V1 primate cortex has inspired computation model development from a micro perspective<sup>[88, 89]</sup>, e.g., the popular convolutional neural network (CNN) submodule in deep learning. DNN mod-

els usually comprise simple modules, e.g., simple artificial neurons running matrix multiplication and nonlinearity computations, and complex layered modules, e.g., convolutional neural network CNNs, recurrent neural networks (RNNs) [90–95].

CNN is the most popular module in various deep learning neural network structures. The core idea of the CNN was derived from the visual cortex of cats by Hubel and Wiesel[96]. Specifically, CNN uses convolutional layers with convolutional kernels (analogous to receptive fields in neuroscience) of different sizes to extract visual features from the input image layer by layer[91, 92, 95, 97]. Simplistically, the 2-dimensional image convolution operation can be written as follows:

$$\text{filter } 2d = \sum_{i=0}^k \sum_{j=0}^k g(i, j) \otimes X(k-i, k-j) \quad (5)$$

where  $X$  is the input image matrix to be convolved with the kernel  $h$ . The indices  $i$  and  $j$  depend on the image size, and  $k$  depends on the kernel size. Usually, 3-dimensional convolution is conducted in a CNN, with an additional dimension called channel, i.e., 2D convolutional operations with different kernels. Other layers also contribute to the benefits of CNNs in the computer vision field, such as pooling layers that reduce the size of features fed to the next layer for computational efficiency, dropout layers that ameliorate overfitting by randomly closing neurons during training, and softmax layers that help score the classification problem. Recent advances in CNN algorithms have led to outstanding performance in various tasks, such as image processing and natural language processing[98–101]. Even so, the performance of the CNN is highly dependent on hyperparameter tuning, such as the number of layers and the kernel sizes in the convolutional layers.

The RNN is a time series-based deep learning algorithm that utilizes sequential data with temporal information from network inputs, in contrast to traditional deep neural networks, which regard the input and output as independent[98]. The output of the RNN depends on the prior outputs and the current inputs. Generally, the general formulation of RNN can be formulated as follows:

$$\begin{aligned} \text{net}_t &= Ux_t + Ws_{t-1} \\ s_{t-1} &= f(\text{net}_{t-1}) \end{aligned} \quad (6)$$

where  $x_t$  is the input at time  $t$  and  $s_{t-1}$  is the output of the RNN at time  $t-1$ ,  $U$  and  $W$  are their corresponding weights.  $f(*)$  is a nonlinear operation. Many variants of the RNN have been proposed for specific tasks, e.g., classic long short-term memory (LSTM)[93], gated recurrent units[102], and multiplicative LSTM[103]. RNN architectures have been applied in many fields due to their effective time-series data analysis characteristics, such as speech recognition[104], natural language proces-

sing[105], and signal identification[106]. Given the essential temporal characteristics of neural recording data, there are exciting prospects for applying RNNs in visual neural coding. Even for the retina, it was shown that the role of recurrence is important for encoding of dynamics videos[107].

Generative adversarial networks (GANs) are another popular approach used for visual neural decoding. Goodfellow et al.[92] first proposed the core mechanism of the GAN from game theory, i.e., the generator network attempts to generate data that can fool the discriminator network, while the discriminator network attempts to distinguish the generated (forged) data from real data[108], which can be briefly formulated as follows:

$$\min \max V(D, G) = E_{x \sim p_{\text{data}}(x)}[\log D(x)] + E_{z \sim p_z(z)}[\log(1 - D(G(z)))] \quad (7)$$

where  $D$  and  $G$  are the discriminator and generator, respectively,  $\sim p_{\text{data}}$  is the distribution of the real data,  $p_z$  is the distribution of noise, and  $E$  is the expectation. Here, the generative model captures the distribution of the real data and is trained to maximize the probability that the discriminator will make a mistake. The generator is trained while the discriminator is idle to obtain its predictions when attempting to fool the discriminator. Such steps are repeated as both the discriminator and generator improve in their respective tasks. Various types of GANs have been proposed: Conditional GAN (CGAN)[108], deep convolutional GAN (DCGAN)[109], Laplacian pyramid GAN (LAPGAN)[110], and superresolution GAN (SRGAN)[111].

Transfer learning is an emerging technique in machine learning[112, 113] in which a machine exploits the knowledge gained from a previous task to improve the generalizability of another[114, 115]; usually, the scale of data in the original training task is larger than that in the new problem. In this case, the advantages of transfer learning are obvious, i.e., it reduces training time, results in better neural network performance (in most cases), and does not need much data. The neural coding issue is quite appropriate for transfer-learning applications, given the scarcity of neural recording data. Indeed, a recent study shows that a pre-trained DNN decoder can be used for real-time video decoding via spikes[116]. The use of transfer learning techniques in visual neural decoding is discussed in Section 6.

Although conventional visual neural decoding methods, involving the application of linear decoding and simple nonlinear decoding approaches, have made achievements in the visual neural decoding of white noise and simple artificial pattern stimuli, these methods have encountered a bottleneck in regard to natural scene stimulus decoding. Meanwhile, the great representation power of deep learning techniques has attracted more and more researchers to apply them in various research fields, in-

cluding neuroscience.

In the last several years, DNN-based neural decoding methods have been proposed to address this task. For example, Parthasarathy et al.<sup>[22]</sup> first linearly reconstructed an image from simulated RGC spike signals and then enhanced it through a deep neural network. The reconstruction performance with their decoding framework outperformed that of linear decoding methods. To overcome the limitations of Parthasarathy's methods, Zhang et al.<sup>[23]</sup> proposed an end-to-end decoding framework and reconstructed dynamic movie stimuli from salamander RGC signals. Shen et al.<sup>[117]</sup> designed an end-to-end direct reconstruction model in which fMRI brain signals are decoded to natural stimuli. Furthermore, Shen et al.<sup>[118]</sup> explored the generalizability of their decoding model from natural stimuli to artificial images.

Apart from the supervised learning DNN methods (in which the ground truths for the reconstruction targets are provided) above, as illustrated in Fig. 2, recent studies have also focused on the application of deep generative techniques in visual neural decoding, for example, the variational autoencoder (VAE). Advancing beyond pairwise matching (stimulus-response) for a given real physiological dataset, the VAE uses an encoder to describe the probability distribution of the latent state space. New stimuli that do not exist in real physiological datasets can be created by sampling from the latent state space, resembling the regular patterns of real stimuli. Han et al.<sup>[119]</sup> reconstructed the video inputs of the fMRI activity by converting the latent variables to the video frames through the VAE's decoder. Du et al. proposed a neural decoding model based on the VAE to learn the disentangled image representations. Furthermore, Du et al.<sup>[72]</sup> first introduced the Bayesian deep learning to a neural decoding study and combined it with a multiview deep generative model called DGMM.

Due to the great success of GANs in synthesizing high-fidelity images, some researchers have leveraged GAN-based approaches to generate visual stimuli. St-Yves and Naselaris<sup>[120]</sup> trained their CGANs with reconstructed images conditioned on given fMRI signals. Gerven et al.<sup>[121]</sup> trained a DCGAN separately on image datasets to learn the latent state space and then used it to generate handwritten characters and natural grayscale images from fMRI signals. Some GAN-based methods have been proposed for human face image decoding. For example, Gerven et al.<sup>[122]</sup> first inverted the linear transformation from latent features to neural responses with MAP estimation. Then, adversarial training was used to perform a nonlinear transformation from perceived stimuli to latent features. VanRullen and Reddy<sup>[123]</sup> trained a VAE using a GAN over a large celebrity face dataset and obtained the corresponding latent space. When linear-transformed fMRI signals are input to the VAE, both robust pairwise decoding and accurate gender classification can be achieved. Compared with linear reconstruction

methods, nonlinear methods, especially DNN-based methods, can greatly improve the accuracy of natural image reconstruction, especially for visual details. Although the performances of current DNN-based decoding methods depend greatly on the scale of neural data, deep learning techniques are still one of the most promising methods for the development of visual neural decoding.

## 5 Open resources

The open-source nature of datasets with high-quality neuronal physiological responses is essential for neuroscience research. These open-source data objectively connect different research works through sharing data and play an important role in forming benchmarks in the field of neural decoding. Here, we summarize three large-scale open neural databases widely used worldwide, as well as some neural analysis software toolboxes.

OpenNEURO<sup>[124]</sup> is an open-science neuroinformatics database that is freely available online. All researchers can browse and explore the public datasets, which have been shared by a wide range of global contributors. In addition, the OpenNEURO collaboration is committed to obtaining more public datasets compatible with brain imaging data structure (BIDS), a standardized format of neural imaging data. OpenNEURO is run by a research group led by Russell Poldrack and originated from the OpenfMRI project<sup>[124]</sup> in 2013. Today, the OpenNEURO database comprises MRI (17 940 participants, 509 public datasets), positron emission tomography (PET) (9 participants, 7 public datasets), magnetoencephalography (MEG) (365 participants, 21 public datasets), EEG (2 454 participants, 68 public datasets), and intracranial electroencephalography (iEEG) (202 participants, 12 public datasets).

The Allen Brain Map is a large, open science platform established by the Allen Institute for Brain Science, whose goal is to accelerate neuroscience research with the release of large-scale, publicly available atlases of the brain<sup>[38]</sup>. The Allen Institute publicly shares all the data, products, and findings from their research work. The main projects in the Allen Brain Map consist of the following databases. The Allen Brain Atlases capture patterns of gene expression across the brain in various species. The Allen Cell Types database contains the different types of neurons and other brain cells in human and mouse brains. Connectivity includes neural connections at scales ranging from the whole brain to the subcellular level. Finally, the Allen Brain Observatory contains the result of a series of systematic experiments conducted on mice with a variety of visual stimuli. Multiple recording modalities are the key features of this project, including calcium imaging response from multiple cortical visual areas across hundreds of two-photon imaging sessions and spiking activity from the visual cortex, hippocampus, and thalamus across dozens of electrophysiology sessions. In



addition, a series of software toolkit resources, e.g., the AllenAPI and AllenSDK and modelling tools, are provided for related analysis work.

Collaborative Research in Computational Neuroscience (CRCNS) is a website primarily developed by Jeffrey L. Teeters of both of the Redwood Center for Theoretical Neuroscience at UC Berkeley<sup>[125]</sup> that hosts large-scale and high-quality experimental datasets. The data range from physiological recordings from sensory and memory systems to eye movement data. Specifically, datasets are classified according to their corresponding collection area for navigation, such as the visual cortex, auditory cortex, motor cortex, hippocampus, retina, and LGN.

Some software toolboxes have been developed as auxiliary resources for neuroscience researchers. NeuroRA is an easy-to-use toolbox based on Python, which can conduct representational similarity analysis (RSA) on nearly all kinds of neural data, including behavioral, EEG, MEG, stereoelectroencephalography (sEEG), electrocorticography (ECoG), fMRI and others. In addition, users can perform neural pattern similarity (NPS), classification-based EEG decoding in NeuroRA<sup>[126]</sup>. Brainlife provides an online, community-based platform where users can publish code and data while integrating cloud-computing resources to run their projects<sup>[127]</sup>. MultiVariate pattern analysis in Python (PyMVPA)<sup>[128]</sup> is a free Python package that provides a handy interface to a wide range of algorithms for classification, regression, feature extraction, data input and output, and other data processing demands in the neuroimaging field. Many popular machine learning packages, such as scikit-learn, shogun, the modular toolkit for data processing (MDP), etc., are integrated well into the PyMVPA framework<sup>[129, 130]</sup>. Recently, Huang et al.<sup>[131]</sup> developed an open-source toolbox for human brain mapping and decoding called Easy fMRI. This toolbox includes advanced machine learning techniques and high-performance computing for analyzing task-based fMRI datasets. It provides a friendly GUI-based environment for conducting feature analysis, hyperalignment, multi-voxel pattern analysis (MVPA), RSA, and more.

## 6 Open challenges and future directions

As we mentioned in Section 4 above, the issue of visual neural decoding has been a topic of interest for decades, with rapid advances in the development of both brain-activity recording techniques and neural decoding analysis methods. Here, we highlight several potential directions and open challenges and hope to provide researchers with insight into this issue.

Open challenges:

1) Most of the current visual decoding studies are offline, that is, the whole decoding process is limited by the scale of the recorded neural physiological data. These

methods cannot ensure equivalent performance for new data as that for the test set (derived from the recorded neural data). In practice, medical engineering applications, such as brain-machine interfaces, aim to decode information that is usually detected by specific neuroimaging devices in real-time. Therefore, more challenges remain in online deployment for current state-of-the-art visual neural decoders. Researchers should consider online applicability when designing visual decoders.

2) Despite the availability of some public neuroscience resources, as introduced in Section 5 above, large public physiological datasets will remain an ongoing demand for a long time to come. The main reasons are the scarcity of public large-scale neural recording data and the high costs of physiological experiments in accessing these data.

3) To date, most visual neural decoders have been designed according to the specific neural recording modality available to researchers. This type of modality-specific neural decoder limits the generalization to different neural-decoder applications.

Future directions:

1) Given their strong capability for data fitting, machine learning techniques, especially the deep learning techniques, have greatly improved the performance of visual neural decoding systems, outperforming traditional methods, i.e., linear methods and unsophisticated Bayesian methods. The trajectory of neuroscience research will involve new deep learning techniques, given the rapid advancements in deep learning-based research fields such as computer vision, and the increasing speed of deep-learning and computationally friendly hardware.

2) The “transfer learning” strategy in the computer vision field, i.e., transfer of a model trained on one dataset to another problem (usually involving a smaller scale of data), provides a promising means to ameliorate the decoding performance in the field of computer vision. Some decoding studies have made use of this type of technique. For example, Shen et al.<sup>[117]</sup> used a comparator network trained on the ImageNet dataset<sup>[94]</sup> as part of their deep image reconstruction model from fMRI data. VanRullen and Reddy<sup>[123]</sup> trained a VAE using a GAN over the public, large-scale face dataset CelebA and then mapped the recorded fMRI signal into the latent space in the VAE to achieve face reconstruction.

3) In addition, given the difficulties in large-scale pairwise neural signal communication, semisupervised learning techniques have been used to explore visual decoding without the need for a ground truth<sup>[132–136]</sup>. For example, Belyi et al.<sup>[132]</sup> leveraged a self-supervision technique to train their reconstruction model with unlabeled fMRI signals. Du et al.<sup>[134]</sup> solved the labelled-data-scarcity problem by casting the semisupervised classification problem as a specialized missing data imputation task.

4) Multimodality and cross-modal decoding are also

future directions beyond reconstruction that have become popular in visual neural decoding studies. Multimodality decoding is the decoding of a combination of signals from more than one type of recording modality. For example, Ibayashi et al.<sup>[137]</sup> conducted speech decoding from signals in the ventral sensorimotor cortex, simultaneously leveraging information from spikes and local field potentials. Cross-modal decoding describes joint decoding models compatible with different recording modalities. For instance, Xu et al.<sup>[44]</sup> proposed a transcoding framework in which the stimuli can be reconstructed from spikes transcoded from signals from multi modalities.

5) Although several emerging open large-scale neural datasets have been made available (see Section 5), a great demand persists for standardized benchmarks of datasets and metrics in visual neural decoding, which could pave the way for strong baseline decoding models and widely recognized validation and evaluation standards for incoming studies in this field. Focusing on different fields, the Brain-Score platform<sup>[138]</sup>, where submitted models can be scored on a range of brain benchmarks and new benchmarks can be incorporated to challenge the models, serves as a good example for reference in the development of visual learning decoding benchmarks.

6) Most of the current visual decoding studies are based on neural data collected from single brain areas, e.g., the V1 cortex, the LGN, and the retina. The interaction and association from the whole-brain connectivity are usually disregarded. Here, we suggest that researchers in the visual decoding field focus on the simultaneous study of neuroscience and brain science and regard visual neural decoding from a macro perspective. Recently, Kriegeskorte and Diedrichsen<sup>[139]</sup> proposed that the brain is not a conventional computer, but nothing a dynamic system in his review on brain coding. Haxby et al.<sup>[140]</sup> criticized decodability as a poor guide for revealing the content of neural representations in his decoding opinion. These debates and discussions deserve focus and in-depth consideration by neural decoding researchers.

## 7 Conclusions

In this paper, we first briefly analyzed the evolution of decoding tasks, i.e., classification, identification, and reconstruction, as this research field has developed. And we introduced the main neural recording modalities used in visual neural decoding and analyzed the characteristics of the data they acquire. Then we reviewed the main types of decoding approaches that researchers have proposed in recent decades in this field. Open data resources of data and toolkits, as well as open challenges and potential future directions of visual neural decoding, are suggested as well. The ultimate purpose of visual decoding is to decode the content of our experience in the absence of visual input. However, the scarcity of pairwise neurophysiological stimulus datasets and accurate, large-scale record-

ing neural modalities continue to hinder the development of this discipline. Nevertheless, the importance of visual neural decoding cannot be understated. The development of neural decoding technology will promote the development of neural prostheses and brain-computer interface devices. We hope that our brief review will inspire ideas for future work in the cross-disciplinary field of brain science and neural computing.

## Acknowledgements

This work was supported by National Natural Science Foundation of China (Nos. 62176003 and 62088102), and the Royal Society Newton Advanced Fellowship of the UK (No. NAF-R1-191082).

## Open Access

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made.

The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- [1] M. O. Ernst, M. S. Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, vol.415, no.6870, pp.429–433, 2002. DOI: [10.1038/415429a](https://doi.org/10.1038/415429a).
- [2] A. Borst, F. E. Theunissen. Information theory and neural coding. *Nature Neuroscience*, vol.2, no.11, pp.947–957, 1999. DOI: [10.1038/14731](https://doi.org/10.1038/14731).
- [3] R. Q. Quiroga, S. Panzeri. Extracting information from neuronal populations: Information theory and decoding approaches. *Nature Reviews Neuroscience*, vol.10, no.3, pp.173–185, 2009. DOI: [10.1038/nrn2578](https://doi.org/10.1038/nrn2578).
- [4] S. S. Jia, X. Y. Li, T. J. Huang, J. K. Liu, Z. F. Yu. Representing the dynamics of high-dimensional data with non-redundant wavelets. *Patterns*, vol.3, no.3, Article number 100424, 2022. DOI: [10.1016/j.patter.2021.100424](https://doi.org/10.1016/j.patter.2021.100424).
- [5] J. K. Liu, H. M. Schreyer, A. Onken, F. Rozenblit, M. H. Khani, V. Krishnamoorthy, S. Panzeri, T. Gollisch. Inference of neuronal functional circuitry with spike-triggered non-negative matrix factorization. *Nature Communications*, vol.8, no.1, Article number 149, 2017. DOI: [10.1038/s41467-017-00156-9](https://doi.org/10.1038/s41467-017-00156-9).

- [6] J. K. Liu, D. Karamanlis, T. Gollisch. Simple model for encoding natural images by retinal ganglion cells with nonlinear spatial integration. *PLoS Computational Biology*, vol. 18, no. 3, Article number e1009925, 2022. DOI: [10.1371/journal.pcbi.1009925](https://doi.org/10.1371/journal.pcbi.1009925).
- [7] J. J. DiCarlo, D. Zoccolan, N. C. Rust. How does the brain solve visual object recognition? *Neuron*, vol. 73, no. 3, pp. 415–434, 2012. DOI: [10.1016/j.neuron.2012.01.010](https://doi.org/10.1016/j.neuron.2012.01.010).
- [8] G. B. Stanley, F. F. Li, Y. Dan. Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus. *Journal of Neuroscience*, vol. 19, no. 18, pp. 8036–8042, 1999. DOI: [10.1523/JNEUROSCI.19-18-08036.1999](https://doi.org/10.1523/JNEUROSCI.19-18-08036.1999).
- [9] O. Marre, V. Botella-Soler, K. D. Simmons, T. Mora, G. Tkačik, M. J. Berry II. High accuracy decoding of dynamical motion from a large retinal population. *PLoS Computational Biology*, vol. 11, no. 7, Article number e1004304, 2015. DOI: [10.1371/journal.pcbi.1004304](https://doi.org/10.1371/journal.pcbi.1004304).
- [10] T. Yoshida, K. Ohki. Natural images are reliably represented by sparse and variable populations of neurons in visual cortex. *Nature Communications*, vol. 11, no. 1, Article number 872, 2020. DOI: [10.1038/s41467-020-14645-x](https://doi.org/10.1038/s41467-020-14645-x).
- [11] V. Botella-Soler, S. Deny, G. Martius, O. Marre, G. Tkačik. Nonlinear decoding of a complex movie from the mammalian retina. *PLoS Computational Biology*, vol. 14, no. 5, Article number e1006057, 2018. DOI: [10.1371/journal.pcbi.1006057](https://doi.org/10.1371/journal.pcbi.1006057).
- [12] Z. F. Yu, J. K. Liu, S. S. Jia, Y. C. Zhang, Y. J. Zheng, Y. H. Tian, T. J. Huang. Toward the next generation of retinal neuroprosthesis: Visual computation with spikes. *Engineering*, vol. 6, no. 4, pp. 449–461, 2020. DOI: [10.1016/j.eng.2020.02.004](https://doi.org/10.1016/j.eng.2020.02.004).
- [13] D. K. Warland, P. Reinagel, M. Meister. Decoding visual information from a population of retinal ganglion cells. *Journal of Neurophysiology*, vol. 78, no. 5, pp. 2336–2350, 1997. DOI: [10.1152/jn.1997.78.5.2336](https://doi.org/10.1152/jn.1997.78.5.2336).
- [14] N. Brackbill, C. Rhoades, A. Kling, N. P. Shah, A. Sher, A. M. Litke, E. J. Chichilnisky. Reconstruction of natural images from responses of primate retinal ganglion cells. *eLife*, vol. 9, Article number e58516, 2020. DOI: [10.7554/eLife.58516](https://doi.org/10.7554/eLife.58516).
- [15] J. W. Pillow, J. Shlens, L. Paninski, A. Sher, A. M. Litke, E. J. Chichilnisky, E. P. Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, vol. 454, no. 7207, pp. 995–999, 2008. DOI: [10.1038/nature07140](https://doi.org/10.1038/nature07140).
- [16] T. Naselaris, R. J. Prenger, K. N. Kay, M. Oliver, J. L. Gallant. Bayesian reconstruction of natural images from human brain activity. *Neuron*, vol. 63, no. 6, pp. 902–915, 2009. DOI: [10.1016/j.neuron.2009.09.006](https://doi.org/10.1016/j.neuron.2009.09.006).
- [17] S. Nishimoto, A. T. Vu, T. Naselaris, Y. Benjamini, B. Yu, J. L. Gallant. Reconstructing visual experiences from brain activity evoked by natural movies. *Current Biology*, vol. 21, no. 19, pp. 1641–1646, 2011. DOI: [10.1016/j.cub.2011.08.031](https://doi.org/10.1016/j.cub.2011.08.031).
- [18] R. J. Ellis, M. Michaelides. High-accuracy decoding of complex visual scenes from neuronal calcium responses. [Online], Available: <https://biorxiv.org/content/10.1101/271296v1>, 2018.
- [19] S. Garasto, A. A. Bharath, S. R. Schultz. Visual reconstruction from 2-photon calcium imaging suggests linear readout properties of neurons in mouse primary visual cortex. [Online], Available: <https://biorxiv.org/content/10.1101/300392v1>, 2018.
- [20] S. Garasto, W. Nicola, A. A. Bharath, S. R. Schultz. Neural sampling strategies for visual stimulus reconstruction from two-photon imaging of mouse primary visual cortex. In *Proceedings of the 9th International IEEE/EMBS Conference on Neural Engineering*, IEEE, San Francisco, USA, pp. 566–570, 2019.
- [21] B. C. McCann, M. M. Hayhoe, W. S. Geisler. Decoding natural signals from the peripheral retina. *Journal of Vision*, vol. 11, no. 10, Article number 19, 2011. DOI: [10.1167/11.10.193](https://doi.org/10.1167/11.10.193).
- [22] N. Parthasarathy, E. Batty, W. Falcon, T. Rutten, M. Rajpal, E. J. Chichilnisky, L. Paninski. Neural networks for efficient Bayesian decoding of natural images from retinal neurons. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, USA, pp. 6437–6448, 2017.
- [23] Y. C. Zhang, S. S. Jia, Y. J. Zheng, Z. F. Yu, Y. H. Tian, S. W. Ma, T. J. Huang, J. K. Liu. Reconstruction of natural visual scenes from neural spikes with deep neural networks. *Neural Networks*, vol. 125, pp. 19–30, 2020. DOI: [10.1016/j.neunet.2020.01.033](https://doi.org/10.1016/j.neunet.2020.01.033).
- [24] S. B. Ryu, J. H. Ye, Y. S. Goo, C. H. Kim, K. H. Kim. Decoding of temporal visual information from electrically evoked retinal ganglion cell activities in photoreceptor-degenerated retinas. *Investigative Ophthalmology & Visual Science*, vol. 52, no. 9, pp. 6271–6278, 2011. DOI: [10.1167/iovs.11-7597](https://doi.org/10.1167/iovs.11-7597).
- [25] Y. Kamitani, F. Tong. Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, vol. 8, no. 5, pp. 679–685, 2005. DOI: [10.1038/nm1444](https://doi.org/10.1038/nm1444).
- [26] E. Yargholi, G. A. Hossein-Zadeh. Brain decoding-classification of hand written digits from fMRI data employing Bayesian networks. *Frontiers in Human Neuroscience*, vol. 10, Article number 351, 2016. DOI: [10.3389/fnhum.2016.00351](https://doi.org/10.3389/fnhum.2016.00351).
- [27] K. N. Kay, T. Naselaris, R. J. Prenger, J. L. Gallant. Identifying natural images from human brain activity. *Nature*, vol. 452, no. 7185, pp. 352–355, 2008. DOI: [10.1038/nature06713](https://doi.org/10.1038/nature06713).
- [28] T. Horikawa, Y. Kamitani. Generic decoding of seen and imagined objects using hierarchical visual features. *Nature Communications*, vol. 8, no. 1, Article number 15037, 2017. DOI: [10.1038/ncomms15037](https://doi.org/10.1038/ncomms15037).
- [29] D. Kleinfeld, L. Luan, P. P. Mitra, J. T. Robinson, R. Sarpeshkar, K. Shepard, C. Xie, T. D. Harris. Can one concurrently record electrical spikes from every neuron in a mammalian brain? *Neuron*, vol. 103, no. 6, pp. 1005–1015, 2019. DOI: [10.1016/j.neuron.2019.08.011](https://doi.org/10.1016/j.neuron.2019.08.011).
- [30] A. L. Orsborn, S. Dangi, H. G. Moorman, J. M. Carmena. Closed-loop decoder adaptation on intermediate time-scales facilitates rapid BMI performance improvements independent of decoder initialization conditions. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 20, no. 4, pp. 468–477, 2012. DOI: [10.1109/TNSRE.2012.2185066](https://doi.org/10.1109/TNSRE.2012.2185066).
- [31] A. L. Orsborn, H. G. Moorman, S. A. Overduin, M. M.

- Shanechi, D. F. Dimitrov, J. M. Carmena. Closed-loop decoder adaptation shapes neural plasticity for skillful neuroprosthetic control. *Neuron*, vol. 82, no. 6, pp. 1380–1393, 2014. DOI: [10.1016/j.neuron.2014.04.048](https://doi.org/10.1016/j.neuron.2014.04.048).
- [32] M. M. Shanechi, A. L. Orsborn, H. G. Moorman, S. Gowda, S. Dangi, J. M. Carmena. Rapid control and feedback rates enhance neuroprosthetic control. *Nature Communications*, vol. 8, no. 1, Article number 13825, 2017. DOI: [10.1038/ncomms13825](https://doi.org/10.1038/ncomms13825).
- [33] A. V. Nurmikko. Approaches to large scale neural recording by chronic implants for mobile BCIs. In *Proceedings of the 6th International Conference on Brain-Computer Interface*, IEEE, Gangwon, Korea, pp. 1–2, 2018. DOI: [10.1109/IWW-BCI.2018.8311503](https://doi.org/10.1109/IWW-BCI.2018.8311503).
- [34] N. A. Steinmetz, C. Koch, K. D. Harris, M. Carandini. Challenges and opportunities for large-scale electrophysiology with Neuropixels probes. *Current Opinion in Neurobiology*, vol. 50, pp. 92–100, 2018. DOI: [10.1016/j.conb.2018.01.009](https://doi.org/10.1016/j.conb.2018.01.009).
- [35] C. Pandarinath, D. J. O’Shea, J. Collins, R. Jozefowicz, S. D. Stavisky, J. C. Kao, E. M. Trautmann, M. T. Kaufman, S. I. Ryu, L. R. Hochberg, J. M. Henderson, K. V. Shenoy, L. F. Abbott, D. Sussillo. Inferring single-trial neural population dynamics using sequential auto-encoders. *Nature Methods*, vol. 15, no. 10, pp. 805–815, 2018. DOI: [10.1038/s41592-018-0109-9](https://doi.org/10.1038/s41592-018-0109-9).
- [36] M. Rácz, C. Liber, E. Németh, R. Fiáth, J. Rokai, I. Harmati, I. Ulbert, G. Márton. Spike detection and sorting with deep learning. *Journal of Neural Engineering*, vol. 17, no. 1, Article number 016038, 2020. DOI: [10.1088/1741-2552/ab4896](https://doi.org/10.1088/1741-2552/ab4896).
- [37] M. Emami, M. S. Ardakan, P. Pandit, A. Fletcher, S. Rangan, M. Trumpis, B. Bent, C. H. Chiang, J. Viventi. Low-rank nonlinear decoding of  $\mu$ -ECoG from the primary auditory cortex. In *Proceedings of Conference on Cognitive Computational Neuroscience*, Cognitive Computational Neuroscience, Philadelphia, USA, Article number 1276, 2018.
- [38] Allen Institute for Brain Science. Allen Brain Atlas SDK: Visual Coding Neuropixels, 2019. [Online], Available: [https://allensdk.readthedocs.io/en/latest/visual\\_coding\\_neuropixels.html](https://allensdk.readthedocs.io/en/latest/visual_coding_neuropixels.html). Accessed on: Feb. 2022.
- [39] N. Steinmetz, M. Pachitariu, C. Stringer, M. Carandini, K. Harris. Eight-probe neuropixels recordings during spontaneous behaviors, 2019. [Online], Available: [https://figshare.com/articles/dataset/Eight-probe\\_Neuropixels\\_recordings\\_during\\_spontaneous\\_behaviors/7739750#:~:text=Eight-probe%20Neuropixels%20recordings%20in%20three%20mice%20of%20spontaneous,times%20and%20the%20cluster%20identity%20of%20each%20spike](https://figshare.com/articles/dataset/Eight-probe_Neuropixels_recordings_during_spontaneous_behaviors/7739750#:~:text=Eight-probe%20Neuropixels%20recordings%20in%20three%20mice%20of%20spontaneous,times%20and%20the%20cluster%20identity%20of%20each%20spike.). Accessed on: Feb. 2022.
- [40] S. Musall, M. T. Kaufman, A. L. Juavinett, S. Gluf, A. K. Churchland. Single-trial neural dynamics are dominated by richly varied movements. *Nature Neuroscience*, vol. 22, no. 10, pp. 1677–1686, 2019. DOI: [10.1038/s41593-019-0502-4](https://doi.org/10.1038/s41593-019-0502-4).
- [41] N. A. Steinmetz, P. Zatka-Haas, M. Carandini, K. D. Harris. Distributed coding of choice, action and engagement across the mouse brain. *Nature*, vol. 576, no. 7786, pp. 266–273, 2019. DOI: [10.1038/s41586-019-1787-x](https://doi.org/10.1038/s41586-019-1787-x).
- [42] A. Iqbal, P. Dong, C. M. Kim, H. Jang. Decoding neural responses in mouse visual cortex through a deep neural network. In *Proceedings of International Joint Conference on Neural Networks*, IEEE, Budapest, Hungary, 2019. DOI: [10.1109/IJCNN.2019.8852121](https://doi.org/10.1109/IJCNN.2019.8852121).
- [43] Y. J. Kim, N. Brackbill, E. Batty, J. Lee, C. Mitelut, W. Tong, E. J. Chichilnisky, L. Paninski. Nonlinear decoding of natural images from large-scale primate retinal ganglion recordings. *Neural Computation*, vol. 33, no. 7, pp. 1719–1750, 2020. DOI: [10.1162/neco\\_a\\_01395](https://doi.org/10.1162/neco_a_01395).
- [44] Q. Xu, J. R. Shen, X. M. Ran, H. J. Tang, G. Pan, J. K. Liu. Robust transcoding sensory information with neural spikes. *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 5, pp. 1935–1946, 2022. DOI: [10.1109/TNNLS.2021.3107449](https://doi.org/10.1109/TNNLS.2021.3107449).
- [45] M. A. B. Brazier. *A History of the Electrical Activity of the Brain: The First Half-Century*. Oxford, UK: Macmillan, 1961.
- [46] E. S. Kappenman, S. J. Luck. ERP components: The ups and downs of brainwave recordings. *The Oxford Handbook of Event-Related Potential Components*, E. S. Kappenman, S. J. Luck, Eds., New York, USA: Oxford University Press, pp. 3–30, 2012.
- [47] S. J. Luck. *An Introduction to the Event-related Potential Technique*, 2nd ed., Cambridge, UK: MIT Press, 2014.
- [48] B. J. Roach, D. H. Mathalon. Event-related EEG time-frequency analysis: An overview of measures and an analysis of early gamma band phase locking in schizophrenia. *Schizophrenia Bulletin*, vol. 34, no. 5, pp. 907–926, 2008. DOI: [10.1093/schbul/sbn093](https://doi.org/10.1093/schbul/sbn093).
- [49] A. S. Al-Fahoum, A. A. Al-Fraihat. Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains. *International Scholarly Research Notices*, vol. 2014, Article number 730218, 2014. DOI: [10.1155/2014/730218](https://doi.org/10.1155/2014/730218).
- [50] S. Aggarwal, N. Chugh. Signal processing techniques for motor imagery brain computer interface: A review. *Array*, vol. 1–2, Article number 100003, 2019. DOI: [10.1016/j.array.2019.100003](https://doi.org/10.1016/j.array.2019.100003).
- [51] R. Boostani, F. Karimzadeh, M. Nami. A comparative review on sleep stage classification methods in patients and healthy individuals. *Computer Methods and Programs in Biomedicine*, vol. 140, pp. 77–91, 2017. DOI: [10.1016/j.cmpb.2016.12.004](https://doi.org/10.1016/j.cmpb.2016.12.004).
- [52] E. P. Torres, E. A. Torres, M. Hernández-Álvarez, S. G. Yoo. EEG-based BCI emotion recognition: A survey. *Sensors*, vol. 20, no. 18, Article number 5083, 2020. DOI: [10.3390/s20185083](https://doi.org/10.3390/s20185083).
- [53] Q. Xiong, X. M. Zhang, W. F. Wang, Y. H. Gu. A parallel algorithm framework for feature extraction of EEG signals on MPI. *Computational and Mathematical Methods in Medicine*, vol. 2020, Article number 9812019, 2020. DOI: [10.1155/2020/9812019](https://doi.org/10.1155/2020/9812019).
- [54] P. C. Sen, M. Hajra, M. Ghosh. Supervised classification algorithms in machine learning: A survey and review. In *Proceedings of IEM Graph 2018*, Springer, Singapore, pp. 99–111, 2020.
- [55] S. Kumar, F. Yger, F. Lotte. Towards adaptive classification using Riemannian geometry approaches in brain-computer interfaces. In *Proceedings of the 7th International Winter Conference on Brain-computer Interface*, IEEE, Gangwon, Korea, 2019.

- [56] F. Lotte, L. Bougrain, M. Clerc. Electroencephalography (EEG)-based brain-computer interfaces. *Wiley Encyclopedia of Electrical and Electronics Engineering*. J. G. Webster, Ed., New York, USA: John Wiley & Sons, Ltd, pp. 1–20, 2015.
- [57] S. Y. Chung, H. J. Yoon. Affective classification using Bayesian classifier and supervised learning. In *Proceedings of the 12th International Conference on Control, Automation and Systems*, IEEE, Jeju, Korea, pp. 1768–1771, 2012.
- [58] V. Anand, S. R. Sreeja, D. Samanta. An automated approach for task evaluation using EEG signals. [Online], Available: <https://arxiv.org/abs/1911.02966>, 2019.
- [59] M. Čukić, M. Stokić, S. Simić, D. Pokrajac. The successful discrimination of depression from EEG could be attributed to proper feature extraction and not to a particular classification method. *Cognitive Neurodynamics*, vol. 14, no. 4, pp. 443–455, 2020. DOI: [10.1007/s11571-020-09581-x](https://doi.org/10.1007/s11571-020-09581-x).
- [60] J. V. Riquelme-Ros, G. Rodríguez-Bermúdez, I. Rodríguez-Rodríguez, J. V. Rodríguez, J. M. Molina-García-Pardo. On the better performance of pianists with motor imagery-based brain-computer interface systems. *Sensors*, vol. 20, no. 16, Article number 4452, 2020. DOI: [10.3390/s20164452](https://doi.org/10.3390/s20164452).
- [61] S. Garofalo, C. Timmermann, S. Battaglia, M. E. Maier, G. Mediofrontal negativity signals unexpected timing of salient outcomes. *Journal of Cognitive Neuroscience*, vol. 29, no. 4, pp. 718–727, 2017. DOI: [10.1162/jocn\\_a\\_01074](https://doi.org/10.1162/jocn_a_01074).
- [62] J. J. Foster, D. W. Sutterer, J. T. Serences, E. K. Vogel, E. Awh. Alpha-band oscillations enable spatially and temporally resolved tracking of covert spatial attention. *Psychological Science*, vol. 28, no. 7, pp. 929–941, 2017. DOI: [10.1177/0956797617699167](https://doi.org/10.1177/0956797617699167).
- [63] J. J. Foster, D. W. Sutterer, J. T. Serences, E. K. Vogel, E. Awh. The topography of alpha-band activity tracks the content of spatial working memory. *Journal of Neurophysiology*, vol. 115, no. 1, pp. 168–177, 2016. DOI: [10.1152/jn.00860.2015](https://doi.org/10.1152/jn.00860.2015).
- [64] J. Samaha, J. J. Barrett, A. D. Sheldon, J. J. LaRocque, B. R. Postle. Dissociating perceptual confidence from discrimination accuracy reveals no influence of metacognitive awareness on working memory. *Frontiers in Psychology*, vol. 7, Article number 851, 2016. DOI: [10.3389/fpsyg.2016.00851](https://doi.org/10.3389/fpsyg.2016.00851).
- [65] T. C. Sprague, S. Saproo, J. T. Serences. Visual attention mitigates information loss in small-scale and large-scale neural codes. *Trends in Cognitive Sciences*, vol. 19, no. 4, pp. 215–226, 2015. DOI: [10.1016/j.tics.2015.02.005](https://doi.org/10.1016/j.tics.2015.02.005).
- [66] D. W. Sutterer, J. J. Foster, J. T. Serences, E. K. Vogel, E. Awh. Alpha-band oscillations track the retrieval of precise spatial representations from long-term memory. *Journal of Neurophysiology*, vol. 122, no. 2, pp. 539–551, 2019. DOI: [10.1152/jn.00268.2019](https://doi.org/10.1152/jn.00268.2019).
- [67] H. Keren, G. O'Callaghan, P. Vidal-Ribas, G. A. Buzzell, M. A. Brotman, E. Leibenluft, P. M. Pan, L. Meffert, A. Kaiser, S. Wolke, D. S. Pine, A. Stringaris. Reward processing in depression: A conceptual and meta-analytic review across fMRI and EEG studies. *The American Journal of Psychiatry*, vol. 175, no. 11, pp. 1111–1120, 2018. DOI: [10.1176/appi.ajp.2018.17101124](https://doi.org/10.1176/appi.ajp.2018.17101124).
- [68] J. Radua, A. Schmidt, S. Borgwardt, A. Heinz, F. Schlagenhauf, P. McGuire, P. Fusar-Poli. Ventral striatal activation during reward processing in psychosis: A neurofunctional meta-analysis. *JAMA Psychiatry*, vol. 72, no. 12, pp. 1243–1251, 2015. DOI: [10.1001/jamapsychiatry.2015.2196](https://doi.org/10.1001/jamapsychiatry.2015.2196).
- [69] A. C. Connolly, J. S. Guntupalli, J. Gors, M. Hanke, Y. O. Halchenko, Y. C. Wu, H. Abdi, J. V. Haxby. The representation of biological classes in the human brain. *Journal of Neuroscience*, vol. 32, no. 8, pp. 2608–2618, 2012. DOI: [10.1523/JNEUROSCI.5547-11.2012](https://doi.org/10.1523/JNEUROSCI.5547-11.2012).
- [70] C. Allefeld, J. D. Haynes. Multi-voxel pattern analysis. *Brain Mapping*, A. W. Toga, Ed., Amsterdam, Netherlands: Academic Press, pp. 641–646, 2015.
- [71] C. J. Burke, M. Baddeley, P. N. Tobler, W. Schultz. Partial adaptation of obtained and observed value signals preserves information about gains and losses. *Journal of Neuroscience*, vol. 36, no. 39, pp. 10016–10025, 2016. DOI: [10.1523/JNEUROSCI.0487-16.2016](https://doi.org/10.1523/JNEUROSCI.0487-16.2016).
- [72] C. D. Du, C. Y. Du, L. J. Huang, H. G. He. Reconstructing perceived images from human brain activities with Bayesian deep multiview learning. *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 8, pp. 2310–2323, 2019. DOI: [10.1109/TNNLS.2018.2882456](https://doi.org/10.1109/TNNLS.2018.2882456).
- [73] T. W. Chen, T. J. Wardill, Y. Sun, S. R. Pulver, S. L. Renninger, A. Baohan, E. R. Schreiter, R. A. Kerr, M. B. Orger, V. Jayaraman, L. L. Looger, K. Svoboda, D. S. Kim. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature*, vol. 499, no. 7458, pp. 295–300, 2013. DOI: [10.1038/nature12354](https://doi.org/10.1038/nature12354).
- [74] A. Giovannucci, J. Friedrich, P. Gunn, J. Kalfon, B. L. Brown, S. A. Koay, J. Taxidis, F. Najafi, J. L. Gauthier, P. C. Zhou, B. S. Khakh, D. W. Tank, D. B. Chklovskii, E. A. Pnevmatikakis. CaImAn an open source tool for scalable calcium imaging data analysis. *eLife*, vol. 8, Article number e38173, 2019. DOI: [10.7554/eLife.38173](https://doi.org/10.7554/eLife.38173).
- [75] J. T. Vogelstein, A. M. Packer, T. A. Machado, T. Sippy, B. Babadi, R. Yuste, L. Paninski. Fast nonnegative deconvolution for spike train inference from population calcium imaging. *Journal of Neurophysiology*, vol. 104, no. 6, pp. 3691–3704, 2010. DOI: [10.1152/jn.01073.2009](https://doi.org/10.1152/jn.01073.2009).
- [76] T. Deneux, A. Kaszas, G. Szalay, G. Katona, T. Lakner, A. Grinvald, B. Rózsa, I. Vanzetta. Accurate spike estimation from noisy calcium signals for ultrafast three-dimensional imaging of large neuronal populations *in vivo*. *Nature Communications*, vol. 7, no. 1, Article number 12190, 2016. DOI: [10.1038/ncomms12190](https://doi.org/10.1038/ncomms12190).
- [77] J. Friedrich, P. C. Zhou, L. Paninski. Fast online deconvolution of calcium imaging data. *PLoS Computational Biology*, vol. 13, no. 3, Article number e1005423, 2017. DOI: [10.1371/journal.pcbi.1005423](https://doi.org/10.1371/journal.pcbi.1005423).
- [78] B. F. Grewe, D. Langer, H. Kasper, B. M. Kampa, F. Helmchen. High-speed *in vivo* calcium imaging reveals neuronal network activity with near-millisecond precision. *Nature Methods*, vol. 7, no. 5, pp. 399–405, 2010. DOI: [10.1038/nmeth.1453](https://doi.org/10.1038/nmeth.1453).
- [79] P. Rupperecht, S. Carta, A. Hoffmann, M. Echizen, A. Blot, A. C. Kwan, Y. Dan, S. B. Hofer, K. Kitamura, F. Helmchen, R. W. Friedrich. A database and deep learning toolbox for noise-optimized, generalized spike infer-

- ence from calcium imaging. *Nature Neuroscience*, vol. 24, no. 9, pp. 1324–1337, 2021. DOI: [10.1038/s41593-021-00895-5](https://doi.org/10.1038/s41593-021-00895-5).
- [80] L. Theis, P. Berens, E. Froudarakis, J. Reimer, M. R. Rosón, T. Baden, T. Euler, A. S. Tolias, M. Bethge. Benchmarking spike rate inference in population calcium imaging. *Neuron*, vol. 90, no. 3, pp. 471–482, 2016. DOI: [10.1016/j.neuron.2016.04.014](https://doi.org/10.1016/j.neuron.2016.04.014).
- [81] Z. Q. Wei, B. J. Lin, T. W. Chen, K. Daie, K. Svoboda, S. Druckmann. A comparison of neuronal population dynamics measured with calcium imaging and electrophysiology. *PLoS Computational Biology*, vol. 16, no. 9, Article number e1008198, 2020. DOI: [10.1371/journal.pcbi.1008198](https://doi.org/10.1371/journal.pcbi.1008198).
- [82] S. M. Tang, Y. M. Zhang, Z. H. Li, M. Li, F. Liu, H. F. Jiang, T. S. Lee. Large-scale two-photon imaging revealed super-sparse population codes in the V1 superficial layer of awake monkeys. *eLife*, vol. 7, Article number e33370, 2018. DOI: [10.7554/eLife.33370](https://doi.org/10.7554/eLife.33370).
- [83] S. Schoenmakers, M. Barth, T. Heskes, M. Van Gerven. Linear reconstruction of perceived images from human brain activity. *NeuroImage*, vol. 83, pp. 951–961, 2013. DOI: [10.1016/j.neuroimage.2013.07.043](https://doi.org/10.1016/j.neuroimage.2013.07.043).
- [84] Y. Fujiwara, Y. Miyawaki, Y. Kamitani. Modular encoding and decoding models derived from Bayesian canonical correlation analysis. *Neural Computation*, vol. 25, no. 4, pp. 979–1005, 2013. DOI: [10.1162/NECO\\_a\\_00423](https://doi.org/10.1162/NECO_a_00423).
- [85] U. Güçlü, M. A. J. Van Gerven. Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *Journal of Neuroscience*, vol. 35, no. 27, pp. 10005–10014, 2015. DOI: [10.1523/JNEUROSCI.5023-14.2015](https://doi.org/10.1523/JNEUROSCI.5023-14.2015).
- [86] H. G. Wen, J. X. Shi, Y. Z. Zhang, K. H. Lu, J. Y. Cao, Z. M. Liu. Neural encoding and decoding with deep learning for dynamic natural vision. *Cerebral Cortex*, vol. 28, no. 12, pp. 4136–4160, 2018. DOI: [10.1093/cercor/bhx268](https://doi.org/10.1093/cercor/bhx268).
- [87] R. M. Cichy, A. Khosla, D. Pantazis, A. Torralba, A. Oliva. Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Scientific Reports*, vol. 6, no. 1, Article number 27755, 2016. DOI: [10.1038/srep27755](https://doi.org/10.1038/srep27755).
- [88] K. Fukushima, S. Miyake. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In *Proceedings of Competition and Cooperation in Neural Nets*, Springer, Kyoto, Japan, pp. 267–285, 1982.
- [89] M. Riesenhuber, T. Poggio. Hierarchical models of object recognition in cortex. *Nature Neuroscience*, vol. 2, no. 11, pp. 1019–1025, 1999. DOI: [10.1038/14819](https://doi.org/10.1038/14819).
- [90] K. Cho, B. Van Merriënboer, D. Bahdanau, Y. Bengio. On the properties of neural machine translation: Encoder-decoder approaches. In *Proceedings of SSST@EMNLP 2014, the 8th Workshop on Syntax, Semantics and Structure in Statistical Translation*, Doha, Qatar, pp. 103–111, 2014.
- [91] D. C. Ciresan, U. Meier, J. Masci, L. M. Gambardella, J. Schmidhuber. Flexible, high performance convolutional neural networks for image classification. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, Barcelona, Spain, pp. 1237–1242, 2011.
- [92] I. Goodfellow, Y. Bengio, A. Courville. *Deep Learning*, Cambridge, UK: MIT Press, 2016.
- [93] S. Hochreiter, J. Schmidhuber. Long short-term memory. *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997. DOI: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).
- [94] A. Krizhevsky, I. Sutskever, G. E. Hinton. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017. DOI: [10.1145/3065386](https://doi.org/10.1145/3065386).
- [95] Y. LeCun, Y. Bengio. Convolutional networks for images, speech, and time series. *The Handbook of Brain Theory and Neural Networks*. M. A. Arbib, Ed., Cambridge, USA: MIT Press, pp. 255–258, 1998.
- [96] D. H. Hubel, T. N. Wiesel. Receptive fields of single neurons in the cat's striate cortex. *The Journal of Physiology*, vol. 148, no. 3, pp. 574–591, 1959. DOI: [10.1113/jphysiol.1959.sp006308](https://doi.org/10.1113/jphysiol.1959.sp006308).
- [97] B. A. Richards, T. P. Lillicrap, P. Beaudoin, Y. Bengio, R. Bogacz, A. Christensen, C. Clopath, R. P. Costa, A. De Berker, S. Ganguli, C. J. Gillon, D. Hafner, A. Kepecs, N. Kriegeskorte, P. Latham, G. W. Lindsay, K. D. Miller, R. Naud, C. C. Pack, P. Poirazi, P. Roelfsema, J. Sacramento, A. Saxe, B. Scellier, A. C. Schapiro, W. Senn, G. Wayne, D. Yamins, F. Zenke, J. Zylberberg, D. Therien, K. P. Kording. A deep learning framework for neuroscience. *Nature Neuroscience*, vol. 22, no. 11, pp. 1761–1770, 2019. DOI: [10.1038/s41593-019-0520-2](https://doi.org/10.1038/s41593-019-0520-2).
- [98] A. Al-Saegh, S. A. Dawwd, J. M. Abdul-Jabbar. Deep learning for motor imagery EEG-based classification: A review. *Biomedical Signal Processing and Control*, vol. 63, Article number 102172, 2021. DOI: [10.1016/j.bspc.2020.102172](https://doi.org/10.1016/j.bspc.2020.102172).
- [99] S. Bhattacharya, P. K. R. Maddikunta, Q. V. Pham, T. R. Gadekallu, S. R. Krishnan S, C. L. Chowdhary, M. Alazab, J. Piran. Deep learning and medical image processing for coronavirus (COVID-19) pandemic: A survey. *Sustainable Cities and Society*, vol. 65, Article number 102589, 2021. DOI: [10.1016/j.scs.2020.102589](https://doi.org/10.1016/j.scs.2020.102589).
- [100] K. Fiok, W. Karwowski, E. Gutierrez, M. Saeidi, A. M. Aljuaid, M. R. Davahli, R. Taiar, T. Marek, and B. D. Sawyer. A study of the effects of the COVID-19 pandemic on the experience of back pain reported on twitter® in the United States: A natural language processing approach. *International Journal of Environmental Research and Public Health*, vol. 18, no. 9, Article number 4543, 2021. DOI: [10.3390/ijerph18094543](https://doi.org/10.3390/ijerph18094543).
- [101] K. T. Schütt, H. E. Saucedo, P. J. Kindermans, A. Tkatchenko, K. R. Müller. SchNet – A deep learning architecture for molecules and materials. *The Journal of Chemical Physics*, vol. 148, no. 24, Article number 241722, 2018. DOI: [10.1063/1.5019779](https://doi.org/10.1063/1.5019779).
- [102] S. Roy, I. Kiral-Kornek, S. Harrer. ChronoNet: A deep recurrent neural network for abnormal EEG identification. In *Proceedings of the 17th Conference on Artificial Intelligence in Medicine in Europe*, Springer, Poznan, Poland, pp. 47–56, 2019.
- [103] B. Krause, I. Murray, S. Renals, L. Lu. Multiplicative LSTM for sequence modelling. In *Proceedings of the 5th International Conference on Learning Representations*, Toulon, France, 2016.
- [104] Y. J. Miao, M. Gowayyed, F. Metze. EESN: End-to-end

- speech recognition using deep RNN models and WFST-based decoding. In *Proceedings of IEEE Workshop on Automatic Speech Recognition and Understanding*, IEEE, Scottsdale, USA, pp. 167–174, 2015. DOI: [10.1109/ASRU.2015.7404790](https://doi.org/10.1109/ASRU.2015.7404790).
- [105] W. F. Wiggins, F. Kitamura, I. Santos, L. M. Prevedello. Natural language processing of radiology text reports: Interactive text classification. *Radiology: Artificial Intelligence*, vol. 3, no. 4, Article number e210035, 2021. DOI: [10.1148/ryai.2021210035](https://doi.org/10.1148/ryai.2021210035).
- [106] J. H. Tan, Y. Hagiwara, W. Pang, I. Lim, S. L. Oh, M. Adam, R. S. Tan, M. Chen, U. R. Acharya. Application of stacked convolutional and long short-term memory network for accurate identification of CAD ECG signals. *Computers in Biology and Medicine*, vol. 94, pp. 19–26, 2018. DOI: [10.1016/j.compbiomed.2017.12.023](https://doi.org/10.1016/j.compbiomed.2017.12.023).
- [107] Y. J. Zheng, S. S. Jia, Z. F. Yu, J. K. Liu, T. J. Huang. Unraveling neural coding of dynamic natural visual scenes via convolutional recurrent neural networks. *Patterns*, vol. 2, no. 10, Article number 100350, 2021. DOI: [10.1016/j.patter.2021.100350](https://doi.org/10.1016/j.patter.2021.100350).
- [108] M. Mirza, S. Osindero. Conditional generative adversarial nets. [Online], Available: <https://arxiv.org/abs/1411.1784>, 2014.
- [109] A. Radford, L. Metz, S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. In *Proceedings of the 4th International Conference on Learning Representations*, San Juan, Puerto Rico, 2016.
- [110] E. L. Denton, S. Chintala, A. Szlam, R. Fergus. Deep generative image models using a Laplacian pyramid of adversarial networks. In *Proceedings of the 28th International Conference on Neural Information Processing Systems*, NIPS, Montreal, Canada, pp. 1486–1494, 2015.
- [111] X. T. Wang, K. Yu, S. X. Wu, J. J. Gu, Y. H. Liu, C. Dong, Y. Qiao, C. C. Loy. ESRGAN: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision*, Springer, Munich, Germany, pp. 63–79, 2018.
- [112] S. Bozinovski, A. Fulgosi. The influence of pattern similarity and transfer learning upon training of a base perceptron b2. In *Proceedings of Symposium Informatica*, Gdańsk, Poland, pp. 3–121, 1976.
- [113] S. Bozinovski. Reminder of the first paper on transfer learning in neural networks, 1976. *Informatica*, vol. 44, no. 3, pp. 291–302, 2020.
- [114] L. Y. Pratt, S. Thrun. *Machine Learning-Special Issue on Inductive Transfer*. Berlin, Germany: Springer, 1997.
- [115] L. Y. Pratt. Discriminability-based transfer between neural networks. In *Proceedings of the 5th International Conference on Neural Information Processing Systems*, Morgan Kaufmann Publishers Inc., Denver, Colorado, pp. 204–211, 1992.
- [116] Y. J. Zhang, T. Bu, J. Y. Zhang, S. M. Tang, Z. F. Yu, J. K. Liu, T. J. Huang. Decoding pixel-level image features from two-photon calcium signals of macaque visual cortex. *Neural Computation*, vol. 34, no. 6, pp. 1369–1397, 2022. DOI: [10.1162/neco\\_a\\_01498](https://doi.org/10.1162/neco_a_01498).
- [117] G. H. Shen, K. Dwivedi, K. Majima, T. Horikawa, Y. Kamitani. End-to-end deep image reconstruction from human brain activity. *Frontiers in Computational Neuroscience*, vol. 13, Article number 21, 2019. DOI: [10.3389/fncom.2019.00021](https://doi.org/10.3389/fncom.2019.00021).
- [118] G. H. Shen, T. Horikawa, K. Majima, Y. Kamitani. Deep image reconstruction from human brain activity. *PLoS Computational Biology*, vol. 15, no. 1, Article number e1006633, 2019. DOI: [10.1371/journal.pcbi.1006633](https://doi.org/10.1371/journal.pcbi.1006633).
- [119] K. Han, H. G. Wen, J. X. Shi, K. H. Lu, Y. Z. Zhang, D. Fu, Z. M. Liu. Variational autoencoder: An unsupervised model for encoding and decoding fMRI activity in visual cortex. *NeuroImage*, vol. 198, pp. 125–136, 2019. DOI: [10.1016/j.neuroimage.2019.05.039](https://doi.org/10.1016/j.neuroimage.2019.05.039).
- [120] G. St-Yves, T. Naselaris. Generative adversarial networks conditioned on brain activity reconstruct seen images. In *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, IEEE, Miyazaki, Japan, pp. 1054–1061, 2018. DOI: [10.1109/SMC.2018.00187](https://doi.org/10.1109/SMC.2018.00187).
- [121] K. Gerven, U. Güçlü, L. Ambrogioni, Y. Güçlütürk, M. A. J. Van Gerven. Generative adversarial networks for reconstructing natural images from brain activity. *NeuroImage*, vol. 181, pp. 775–785, 2018. DOI: [10.1016/j.neuroimage.2018.07.043](https://doi.org/10.1016/j.neuroimage.2018.07.043).
- [122] Y. Gerven, U. Güçlü, K. Seeliger, S. Bosch, R. Van Lier, M. Van Gerven. Reconstructing perceived faces from brain activations with deep adversarial neural decoding. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, USA, pp. 4249–4260, 2017.
- [123] R. VanRullen, L. Reddy. Reconstructing faces from fMRI patterns using deep generative neural networks. *Communications Biology*, vol. 2, no. 1, Article number 193, 2019. DOI: [10.1038/s42003-019-0438-y](https://doi.org/10.1038/s42003-019-0438-y).
- [124] C. J. Markiewicz, K. J. Gorgolewski, F. Feingold, R. Blair, Y. O. Halchenko, E. Miller, N. Hardcastle, J. Wexler, O. Esteban, M. Goncavles, A. Jwa, R. Poldrack. The OpenNeuro resource for sharing of neuroscience data. *eLife*, vol. 10, Article number e71774, 2021. DOI: [10.7554/eLife.71774](https://doi.org/10.7554/eLife.71774).
- [125] J. L. Teeters, K. D. Harris, K. J. Millman, B. A. Olshausen, F. T. Sommer. Data sharing for computational neuroscience. *Neuroinformatics*, vol. 6, no. 1, pp. 47–55, 2008. DOI: [10.1007/s12021-008-9009-y](https://doi.org/10.1007/s12021-008-9009-y).
- [126] Z. T. Lu, Y. X. Ku. NeuroRA: A Python toolbox of representational analysis from multi-modal neural data. *Frontiers in Neuroinformatics*, vol. 14, Article number 563669, 2020. DOI: [10.3389/fninf.2020.563669](https://doi.org/10.3389/fninf.2020.563669).
- [127] G. Berto. Benchmark Bundle Segmentation. [Online], Available: <https://doi.org/10.25663/BRAINLIFE.PUB>, 2021.
- [128] M. Hanke, Y. O. Halchenko, P. B. Sederberg, S. J. Hanson, J. V. Haxby, S. Pollmann. PyMVPA: A Python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics*, vol. 7, no. 1, pp. 37–53, 2009. DOI: [10.1007/s12021-008-9041-y](https://doi.org/10.1007/s12021-008-9041-y).
- [129] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011.

- [130] S. Sonnenburg, H. Strathmann, S. Lisitsyn, V. Gal, F. J. I. García, W. Lin, S. De, C. Y. Zhang, frx, tklein23, E. Andreev, JonasBehr, sploving, P. Mazumdar, C. Widmer, P. D. Zora, G. De Toni, S. Mahindre, A. Kislay, K. Hughes, R. Votyakov, khalednasr, S. Sharma, A. Novik, A. Panda, E. Anagnostopoulos, L. Pang, A. Binder, serialhex, B. Esser. Shogun-toolbox/shogun: Shogun 6.1.0. Zenodo, 2017. [Online], Available: [https://zenodo.org/record/1067840#.Yn9UufkyZ\\_p](https://zenodo.org/record/1067840#.Yn9UufkyZ_p).
- [131] S. Huang, L. Sun, M. Yousefnezhad, M. L. Wang, D. Q. Zhang. Temporal information guided generative adversarial networks for stimuli image reconstruction from human brain activities. *IEEE Transactions on Cognitive and Developmental Systems*, to be published, 2021. DOI: [10.1109/TCDS.2021.3098743](https://doi.org/10.1109/TCDS.2021.3098743).
- [132] R. Belyi, G. Gaziv, A. Hoogi, F. Strappini, T. Golan, M. Irani. From voxels to pixels and back: Self-supervision in natural-image reconstruction from fMRI. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, Vancouver, Canada, pp.6514–6524, 2019.
- [133] D. Li, C. D. Du, H. G. He. Semi-supervised cross-modal image generation with generative adversarial networks. *Pattern Recognition*, vol.100, Article number 107085, 2020. DOI: [10.1016/j.patcog.2019.107085](https://doi.org/10.1016/j.patcog.2019.107085).
- [134] C. D. Du, C. Y. Du, H. Wang, J. P. Li, W. L. Zheng, B. L. Lu, H. G. He. Semi-supervised deep generative modelling of incomplete multi-modality emotional data. In *Proceedings of the 26th ACM International Conference on Multimedia*, Lisboa, Portugal, pp.108–116, 2018. DOI: [10.1145/3240508.3240528](https://doi.org/10.1145/3240508.3240528).
- [135] X. Cai, F. P. Nie, W. D. Cai, H. Huang. Heterogeneous image features integration via multi-modal semi-supervised learning model. In *Proceedings of IEEE International Conference on Computer Vision*, IEEE, Sydney, Australia, pp.1737–1744, 2013. DOI: [10.1109/ICCV.2013.218](https://doi.org/10.1109/ICCV.2013.218).
- [136] Z. X. Zhang, F. Ringeval, B. Dong, E. Coutinho, E. Marchi, B. Schüller. Enhanced semi-supervised learning for multimodal emotion recognition. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, Shanghai, China, pp.5185–5189, 2016. DOI: [10.1109/ICASSP.2016.7472666](https://doi.org/10.1109/ICASSP.2016.7472666).
- [137] K. Ibayashi, N. Kunii, T. Matsuo, Y. Ishishita, S. Shimada, K. Kawai, N. Saito. Decoding speech with integrated hybrid signals recorded from the human ventral motor cortex. *Frontiers in Neuroscience*, vol.12, Article number 221, 2018. DOI: [10.3389/fnins.2018.00221](https://doi.org/10.3389/fnins.2018.00221).
- [138] M. Schrimpf, J. Kubilius, H. Hong, N. J. Majaj, R. Rajalingham, E. B. Issa, K. Kar, P. Bashivan, J. Prescott-Roy, F. Geiger, K. Schmidt, D. L. K. Yamins, J. J. DiCarlo. Brain-score: Which artificial neural network for object recognition is most brain-like? *bioRxiv*, Article number 407007, 2020.
- [139] N. Kriegeskorte, J. Diedrichsen. Peeling the onion of brain representations. *Annual Review of Neuroscience*, vol.42, pp.407–432, 2019. DOI: [10.1146/annurev-neuro-080317-061906](https://doi.org/10.1146/annurev-neuro-080317-061906).
- [140] J. V. Haxby, A. C. Connolly, J. S. Guntupalli. Decoding neural representational spaces using multivariate pattern analysis. *Annual Review of Neuroscience*, vol.37, pp.435–456, 2014. DOI: [10.1146/annurev-neuro-062012-170325](https://doi.org/10.1146/annurev-neuro-062012-170325).



**Yi-Jun Zhang** received the B.Eng. degree in communication engineering from Yingcai Honors College, University of Electronic and Technology of China, China in 2017. He is currently a Ph.D. degree candidate in computer science at Department of Computer Science and Engineering, Shanghai Jiao Tong University, China.

His research interests include artificial intelligence and computational neuroscience.

E-mail: yijzhang@sjtu.edu.cn

ORCID iD: 0000-0003-2289-2372

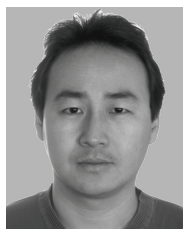


**Zhao-Fei Yu** received the B.Sc. degree in automation from Hong Shen Honors School, College of Optoelectronic Engineering, Chongqing University, China in 2012, and the Ph.D. degree in automation from Automation Department, Tsinghua University, China in 2017. He is currently an assistant professor with Institute for Artificial Intelligence, Peking University, China.

His research interests include artificial intelligence, brain-inspired computing, and computational neuroscience.

E-mail: yuzf12@pku.edu.cn (Corresponding author)

ORCID iD: 0000-0002-6913-7553



**Jian. K. Liu** received the Ph.D. degree in mathematics from University of California at Los Angeles, USA in 2009. He is currently a lecturer with School of Computing, University of Leeds, UK.

His research interests include computational neuroscience and brain-inspired computing.

E-mail: j.liu9@leeds.ac.uk

ORCID iD: 0000-0002-5391-7213



**Tie-Jun Huang** (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in computer science from Wuhan University of Technology, China in 1992 and 1995, respectively, and the Ph.D. degree in pattern recognition and image analysis from Huazhong (Central China) University of Science and Technology, China in 1998. He is currently a professor with Department of

Computer Science, School of Electronics Engineering and Computer Science, Peking University, China, and the director of Beijing Academy of Artificial Intelligence, China. He has published two books and more than 200 peer-reviewed articles in leading journals and conferences. He is a co-editor of four ISO/IEC standards, five national standards of China, and four IEEE standards. He holds more than 50 granted patents. He is also a fellow of Chinese Association for Artificial Intelligence (CAAI) and China Computer Federation (CCF). He received the National Award for Science and Technology of China (Tier-2) three times in 2010, 2012, and 2017. He is also the secretary-general of the Artificial Intelligence Industry Technology Innovation Alliance and the vice-chair of the China National General Group on AI Standardization.

His research interests include visual information processing and neuromorphic computing.

E-mail: tjhuang@pku.edu.cn

ORCID iD: 0000-0002-4234-6099