

# A Novel Brain Decoding Method: a Correlation Network Framework for Revealing Brain Connections

Siyu Yu, Nanning Zheng<sup>\*</sup>, *Fellow, IEEE*, Yongqiang Ma, Hao Wu, and Badong Chen, *Senior Member, IEEE*

**Abstract**—Brain decoding is a hot spot in cognitive science, which focuses on reconstructing perceptual images from brain activities. Analyzing the correlations of collected data from human brain activities and representing activity patterns are two key problems in brain decoding based on functional magnetic resonance imaging (fMRI) signals. However, existing correlation analysis methods mainly focus on the strength information of voxel, which reveals functional connectivity in the cerebral cortex. They tend to neglect the structural information that implies the intracortical or intrinsic connections; that is, structural connectivity. Hence, the effective connectivity inferred by these methods is relatively unilateral. Therefore, we propose in this work a correlation network (CorrNet) framework that could be flexibly combined with diverse pattern representation models. In the CorrNet framework, the topological correlation is introduced to reveal structural information. Rich correlations can be obtained, which contribute to specifying the underlying effective connectivity. We also combine the CorrNet framework with a linear support vector machine (SVM) and a dynamic evolving spike neuron network (SNN) for pattern representation separately, thus provide a novel method for decoding cognitive activity patterns. Experimental results verify the reliability and robustness of our CorrNet framework, and demonstrate that the new method can achieve significant improvement in brain decoding over comparable methods.

**Index Terms**—Brain decoding, functional magnetic resonance imaging (fMRI), connection, correlation network (CorrNet) framework, pattern representation.

## I. INTRODUCTION

ONLY with a good understanding of the brain can we develop more robust artificial intelligence (AI) [1]. Regrettably, the cognitive mechanism of the human brain remains unclear. Brain decoding that focuses on reconstructing perceptual images from brain activities is faced with great challenges. In recent decades, brain decoding with functional magnetic resonance imaging (fMRI) signals has driven plenty of studies [2], [3]. Most relevant methods are based on multi-voxel pattern analysis [4] or voxel-wise modeling [5]. The main idea is to decode the fMRI signals evoked by perceptual stimuli to obtain brain activity patterns, and represent these patterns for perceptual image reconstruction [6], [7]. Hence, there are two key problems in brain decoding with severe noise and high-dimensional fMRI signals. One is how to thoroughly analyze the correlations of collected data, and the other is

S. Yu, N. Zheng(\*correspondence author) Y. Ma, H. Wu and B. Chen are with the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, P.R. China. E-mail: yusiyu, musaqiang, xuan.zhi@stu.xjtu.edu.cn, nnzheng, chenbd@mail.xjtu.edu.cn.

how to effectively represent the brain's activity patterns, which relies on powerful classifiers in machine learning [8]. We believe that the former is more critical to cognitive science.

Complex brain activity states with limited data instances make it necessary to analyze the correlations between fMRI signals and visual stimuli. As is well known, the human brain encodes perceptual images into hierarchical signals, while the reconstruction task aims to decode the brain's responses to the stimuli; hence, the reconstruction process reverts to the process of the human brain's mechanism. In terms of studies on structure-function relationships from the view of brain network theory, it is believed that, as Fig. 1 shows, there are three types of connections between brain nodes: functional connectivity, structural connectivity, and effective connectivity [9]. Actually, the brain nodes are the basic elements of brain network studies, such as neurons in anatomy or voxels in fMRI experiments. The correlations in reconstruction tasks are expected to correspond to the three kinds of connectivity, even for simulating and modeling them, so that the correlation analysis can contribute to revealing the working mechanism of the human brain. An earlier study proposed a modular decoding approach [10]. Later, Bayesian canonical correlation analysis (BCCA) [11] was proposed for brain decoding. Meanwhile, generative models with variational autoencoders (VAEs) [12] were introduced, such as deep canonically correlated autoencoders (DCCAE) [13]. Nevertheless, most existing correlation models [14], [15] may be poor in capturing the correlations because they tend to focus mainly on the strength information of voxel, which reveals the functional connectivity in the cerebral cortex, while neglecting the structural information of voxel, implying the intracortical or intrinsic connections, i.e., the structural connectivity. An early study [16] proved that adjacent neurons in the same cortex tend to present similar activity patterns; that is to say, the functional similarity between neurons is dependent on the structural connectivity. Notably, the effective connectivity expected to be inferred in experiments is exactly constrained by structural connectivity. Therefore, the effective connectivity specified by previous reconstruction methods seems not to be comprehensive and have limited performance. Thus, it is necessary to analyze both the strength values of voxel and the structural information to enrich the correlations and explore the effective connectivity in attention tasks (perceptual tasks). This is the main contribution of this paper.

Moreover, suitable classifiers are required to reveal brain activity patterns as pattern representation models on the basis

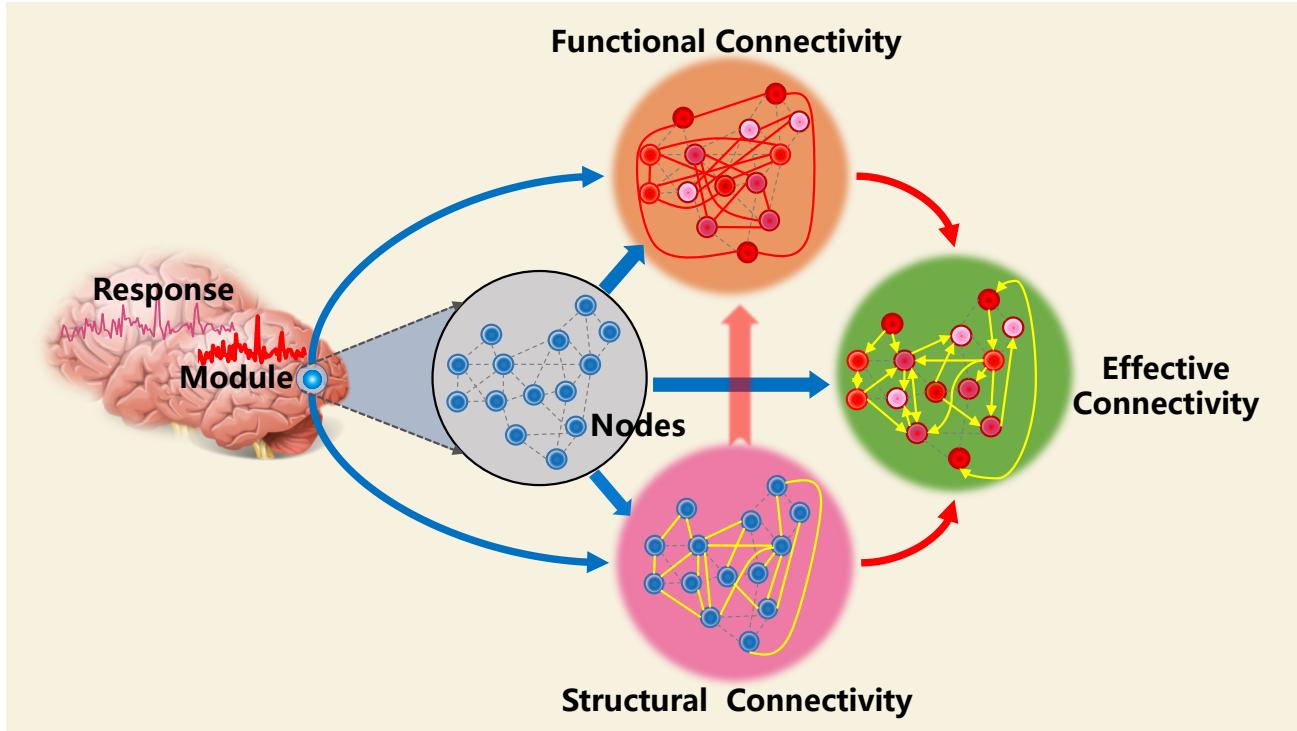


Fig. 1. The three types of connections between brain nodes: functional connectivity, structural connectivity, and effective connectivity. The big blue point in the brain presents a certain module that contains some nodes (small points), and we highlight the module in the gray circle. The dotted gray lines between nodes in the four circles are the imaginary connections used to shape the module. In terms of functional connectivity, nodes in different colors are of different response values. The more red the node, the stronger is its response. Thick red lines are used to connect nodes with the same response strength, revealing functional connectivity. In the pink circle, the thick yellow lines connect nodes that are adjacent according to anatomy. In the green circle, effective connectivity can be inferred by yellow one-sided or two-sided arrows. The one-sided arrows represent the structure-function relationships between two adjacent nodes with different response patterns. The arrow direction means a strong-response node can affect its neighborhood. When two adjacent nodes have the same response pattern, there is a potential interaction effect; this is indicated by the two-sided arrows.

of the correlations, overcoming the perplexing measurement noise. It is hard to make a reliable choice from the numerous machine learning algorithms. Existing studies generally use simple methods (linear models) because fMRI signals have severe noise, are high-dimensional, and often involve small sample sizes. However, most linear reconstruction models may be weak in feature extractions of brain activity patterns. Although some complicated pattern representation models, like DCCAE [13], have been used and have achieved relatively effective performance, they require many intermediate variables and strong hypotheses, thus bringing severe computational problems. Obviously, a single deep learning (DL) model, which has been a topic of active research in recent years, is not suitable for analyzing brain signals with a small sample size in the early visual cortex. A single DL model can result in over-fitting. There are still no powerful, general methods for perceptual image reconstruction for both correlation analysis and brain activity pattern representation. Support vector machines (SVMs) [17] have been used to efficiently represent activity patterns; they are powerful for pattern recognition and classification. Therefore, SVM is one of our optional pattern representation methods. Recently, researchers have tried to use more biological models, such as spike neuron networks (SNNs) [18], to reconstruct perceptual images [19], [20]. Compared with artificial neural networks (ANNs), the learned

connections of SNNs demonstrate the dynamic spatiotemporal relationships in fMRI signals; however, there have not been many studies on this to date. Since we believe that SNNs can be promising for brain decoding, this is the other choice for the pattern representation process in the novel method.

In this paper, we propose a novel brain decoding method composed of a self-adaptive correlation network (CorrNet) framework and a pattern representation model. The CorrNet framework is established by considering three correlations (the correlation among brain activity, the correlation between brain activity and visual stimuli, and the correlation among visual stimuli), and can be combined with diverse pattern representation models. In the CorrNet framework, firstly, the topological correlation and strength correlation are considered. Then, a probabilistic correlation graph in a joint space is generated by iteratively updating the CorrNet framework in a self-learning manner to obtain the pixel-wise correlation pairs (pixelvoxel pairs). Next, the pattern representation model is trained to learn connection weights from the correlation pairs to represent brain activity patterns. Specifically, the linear SVM and the dynamic evolving SNN are both used as pattern representation models separately. This can verify the reliability and robustness of our CorrNet framework. Finally, we test our method and other methods (Miyawaki[10] and BCCA [11]) on a new fMRI dataset from the early visual cortex (mainly area

V1). Experimental results demonstrate that our method can achieve better performance than the other methods in brain decoding, and the CorrNet framework is reliable and robust.

The rest of this paper is organized as follows: Section 2 reviews some related works about brain decoding. Then, the proposed method is introduced in Section 3. Section 4 reports the experimental results. Finally, our conclusion is given in Section 5 as well as some directions for future work.

## II. RELATED WORKS

Much of the literature in cognitive science is devoted to brain decoding. An earlier method combined multi-scale local image bases with multiple scales to decode brain activities [10]. Furthermore, Bayesian inference algorithms and correlation neural networks have been combined [11], [21] to analyze the correlations. The existing correlation methods are almost derived from a voxel receptive-field model, and only rely on the strength information to locate a small perceptual stimulus that elicits activity spatially spread over voxels [22], [23]. As for pattern representation, machine learning algorithms are used, such as linear regression [24], [25], Bayesian classifiers [26], [27], and linear observation models [28], [29]. These methods generally have limited feature extraction power. In addition, some researchers attempt to use deep VAEs [12], like DCCAE [13]. However, brain decoding models with VAEs requires many intermediate variables to model the represent patterns implied in the fMRI signals and corresponding visual stimuli. DCCAE is a brain decoding method proposed in 2015, which contains VAEs as indirect reconstruction models with many intermediate variables. In DCCAE, the assumption distribution within the intermediate variables and fMRI signals is needed as well as the assumption distribution within the intermediate variables and visual stimuli. Then, the joint distributions of the fMRI signals and corresponding visual stimuli are computed in the DCCAE, indirectly. So, the intermediate variables in VAEs and those assumption distributions of DCCAE can easily cause poor generalization and high computational cost. In our work, the proposed method exactly avoided the intermediate variables and assumption distributions by introducing a pixel-wise reconstruction model, and analyzed the correlation between fMRI signals and corresponding visual stimuli, directly. Some researchers found that the linear SVM combined with multivariate voxel selection is effective in classifying fMRI spatial patterns [30]. Recently, a spatiotemporal data machine of evolving SNNs was proposed to represent activity patterns of fMRI signals [20].

Notably, theories about structural and functional brain networks offer insight into connection-to-cognition modeling [9], [31]. Inspired by the cognitive studies, in this paper we propose a CorrNet framework that can be combined with diverse pattern representation models. The CorrNet framework reserves the advantages of past correlation analysis methods in learning the relationships between the strength of fMRI signals and visual stimuli, which reveal the functional connectivity in the human brain. Meanwhile, it considers structure information of voxels by introducing the topological correlation, thus specifying the structural connectivity. Therefore, the effective

connectivity inferred by the CorrNet framework to resolve the dialectic between structure and function is more reliable and comprehensive than previous models. For pattern representation, both a linear SVM model and an SNN model are separately combined with the CorrNet framework, proposing a new method to reconstruct visual images. Our method can avoid many intermediate variables and unnecessary strong hypotheses. To the best of our knowledge, this paper is the first to study topological correlation in perceptual image reconstruction and to discuss the three types of connections (functional connectivity, structural connectivity, and effective connectivity) in attention tasks.

## III. PROPOSED METHOD

Our work rests on how the early visual cortex organizes local interactions to deal with different perceptual stimuli on the basis of functionstructure relationships in the brain. In other words, the proposed method is expected to reveal the functional connectivity by the correlation between brain activities based on blood oxygenation level-dependent (BOLD) fMRI signals acquired during perceptual task performance. Additionally, we attempt to specify the structural connectivity by using the adjacent relationships of voxels in the cortex, instead of analyzing diffusion MRI signals or anatomy. Therefore, the effective connectivity can be inferred by using our model-based method to estimate the model parameters (weight values) that can best reconstruct the perceptual images. As long as we can model the brain activity patterns from the obtained effective connectivity, human-like computer vision in the complex visual scenario may no longer be a mystery.

Actually, for attention tasks, like perceptual tasks, existing studies have not found a reliable way to analyze the structure correlation. Besides, the existing reconstruction experiments only consider the function correlation, and the effective connectivity between voxels is not comprehensive due to the lack of structure correlation; this leads to unsatisfactory experimental results. For example, Bayesian inference algorithms [21], algorithms with VAEs [12] and DCCAE [13] are the most recent methods in brain decoding. However, most of them merely focus on the strength information of fMRI activity, and neglect the structure information by only analyzing the response values of voxels. Reconstruction method with Bayesian network [21] exploits the information of only 10 brain voxels, and may possibly lose some important connection information of voxels. In addition, it does not take the spatial structure of voxels into consideration. To some extent, these methods can be viewed as classifiers for a certain kind of cognitive task, and can be weak in generalization. In this paper, we try to compensate for these past studies by introducing a topological model into our CorrNet framework. Our proposed CorrNet framework can be combined with most brain decoding methods in the correlation analysis stage. Therefore, to some extent, our work explains and simulates the brain's cognitive mechanism.

As Fig. 2 shows, the proposed method contains two parts in terms of the two primary problems in brain decoding. One is the CorrNet framework for obtaining the correlations of



fMRI signals and perceptual images; the other is the pattern representation model for learning brain activity patterns by training a classifier on the basis of the CorrNet framework. The procedure is: First, there are two distinct views ( $\mathbf{X}, \mathbf{Y}$ ),

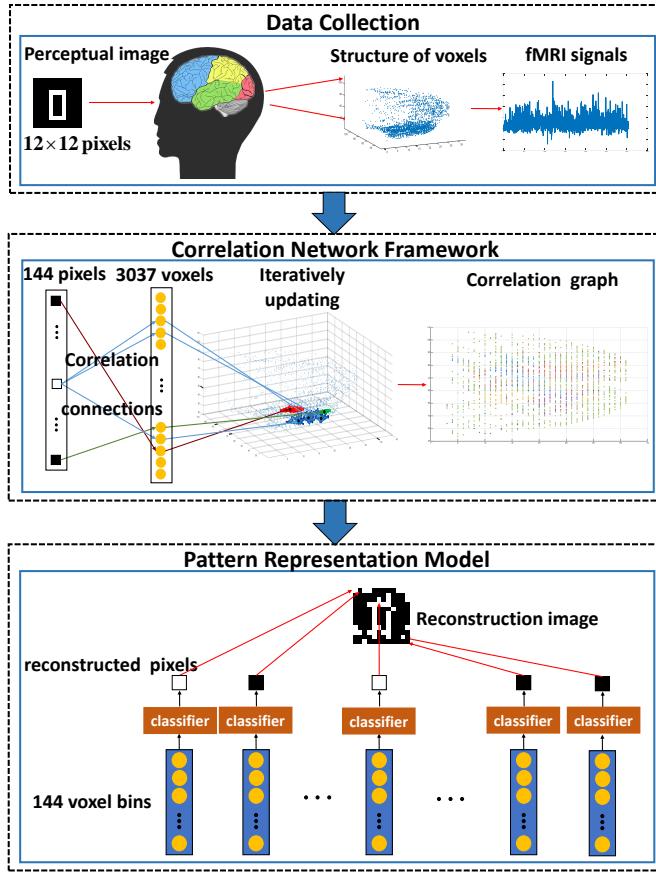


Fig. 2. Flow chart of the proposed method. The complete process of brain decoding contains three steps. In data collection, stimuli, fMRI signals, and the brain structure information of subjects are collected. Then, the CorrNet framework analyzes the data correlations. Finally, with suitable classifier models, such as SVM and SNN, the activity patterns can be presented. To verify the robustness of our CorrNet, we choose different classifiers, like SNN, SVM. Notably, the two Reconstruction image in "the pattern representation model" are reconstructed from the same fMRI signals but with different classifiers. The entire reconstruction images of the testing set will be shown latter.

denoted by  $(\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_N, \mathbf{y}_N)$ , where  $\mathbf{x}_i \in \mathbb{R}^{D_1}$ ,  $\mathbf{y}_i \in \mathbb{R}^{D_2}$  for  $i = 1, \dots, N$ , and  $N$  is the number of training trials.  $\mathbf{X} \in \mathbb{R}^{D_1 \times N}$  and  $\mathbf{Y} \in \mathbb{R}^{D_2 \times N}$  denote fMRI signals with  $D_1$  voxels and visual images with  $D_2$  pixels. In  $N$  training trials,  $N$  visual images with random shape are used. Each image is used just once in the whole experiment, as one trial. That is, each image during the  $N$  training trials is different with others, and is unreplicated. Besides,  $\mathbf{pos}_j \in \mathbb{R}^3$  for  $j = 1, \dots, D_1$  denotes the position of the  $j^{th}$  voxel in the three-dimensional world coordinate frame. In addition,  $\mathbf{v}_j \in \mathbb{R}^N$ ,  $j = 1, \dots, D_1$  denotes the strength value vector of the  $j^{th}$  voxel, and  $\mathbf{p}_k \in \mathbb{R}^N$ ,  $k = 1, \dots, D_2$  presents the pixel value vector of the  $k^{th}$  pixel. The pixel value vector of the  $k^{th}$  pixel is an  $N$ -dimension vector which includes the  $N$  pixel values in the  $N$  trials. For each trial, certain image is used as visual stimulus where the  $k^{th}$

pixel of the images is a binary value. Based on the pixel-wise inverse receptive field, the CorrNet framework was generated to obtain the correlation voxel bins, denoted as  $\mathbf{bin}_k$  for  $k = 1, \dots, D_2$ .

To obtain  $\mathbf{bin}_k$ , a binary matrix  $\mathbf{Corr} \in \mathbb{R}^{D_1 \times D_2}$  is used to define the *corresponding correlation* between voxels and pixels. The notation  $\mathbf{Corr}(\mathbf{v}_j, \mathbf{p}_k)$  presents the element in the  $j^{th}$  row and  $k$  column of the matrix  $\mathbf{Corr}$ , measuring the corresponding correlation between voxel  $\mathbf{v}_j$  and pixel  $\mathbf{p}_k$ . For brain decoding,  $\mathbf{Corr}(\mathbf{v}_j, \mathbf{p}_k) = 1$  indicates that  $\mathbf{v}_j$  and  $\mathbf{p}_k$  are of significant corresponding correlation, that is,  $\mathbf{v}_j$  is in the inverse receptive field of  $\mathbf{p}_k$ ; otherwise,  $\mathbf{v}_j$  is not in the inverse receptive field of  $\mathbf{p}_k$ . In this case,  $\mathbf{bin}_k = \{\mathbf{v}_j | \mathbf{Corr}(\mathbf{v}_j, \mathbf{p}_k) = 1, j \in \{1, \dots, D_1\}\}$ , which presents the set of voxels that are in the inverse receptive field of pixel  $\mathbf{p}_k$ . In our method, the *corresponding correlation probability*  $\text{prob}_{\text{corr}}(\mathbf{p}_k, \mathbf{v}_j | \mathbf{bin}_k)$  is defined to evaluate  $\mathbf{Corr}$ , given by

$$\mathbf{Corr}(\mathbf{v}_j, \mathbf{p}_k) = \begin{cases} 1, & \text{prob}_{\text{corr}}(\mathbf{p}_k, \mathbf{v}_j | \mathbf{bin}_k) \geq \varepsilon_{\text{Corr}} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where  $\text{prob}_{\text{corr}}(\mathbf{p}_k, \mathbf{v}_j | \mathbf{bin}_k)$  is the probability that  $\mathbf{v}_j$  is in the inverse receptive field of  $\mathbf{p}_k$ . Based on the topology of the brain [16] and Bayes theory [11], the computation process of  $\text{prob}_{\text{corr}}(\mathbf{p}_k, \mathbf{v}_j | \mathbf{bin}_k)$  will be detailed latter.  $\varepsilon_{\text{Corr}}$ , a free parameter, is used to determine the degree of corresponding correlation, which is set to 0.5 in the present study. It is worth noting that the corresponding correlation in our method is not of simple linear relationship, and is different from the traditional correlation measure, such as Pearson's Correlation.  $\mathbf{Corr}(\mathbf{v}_j, \mathbf{p}_k)$  is called in this paper the *corresponding correlation coefficient*, which can be viewed as a significance measure of the corresponding correlation between  $\mathbf{v}_j$  and  $\mathbf{p}_k$ . Its value depends on  $\text{prob}_{\text{corr}}(\mathbf{p}_k, \mathbf{v}_j | \mathbf{bin}_k)$  and  $\varepsilon_{\text{Corr}}$ . The voxel  $\mathbf{v}_j$  and pixel  $\mathbf{p}_k$  are of significant corresponding correlation if and only if  $\text{prob}_{\text{corr}}(\mathbf{p}_k, \mathbf{v}_j | \mathbf{bin}_k)$  is equal to or greater than  $\varepsilon_{\text{Corr}}$ .

Then, based on pixel-wise modeling, we used a distributed pattern representation model to build the maps from  $\mathbf{bin}_k$  to  $\mathbf{p}_k$  directly:

$$\mathbf{p}_k = f^{(k)}(\mathbf{bin}_k), k = 1, \dots, D_2. \quad (2)$$

where  $f^{(k)}$  denotes the reconstruction function of  $\mathbf{p}_k$ . Since each pixel is reconstructed by the pixel-wise representation method, the reconstruction image is available.

#### A. Correlation network framework

The CorrNet framework is one of the contributions in this paper. As Fig. 3 shows, the CorrNet framework was proposed to generate the correlation bins  $\mathbf{bin}_k$  and the correlation matrix  $\mathbf{Corr} \in \mathbb{R}^{D_1 \times D_2}$  from the stimuli-response pairs. There are three kinds of correlations considered in this paper: the correlation among brain activities, the correlation between brain activities and visual stimuli, and the correlation among visual stimuli. We believe the three kinds of correlations correspond to the three kinds of connections mentioned above; that is, functional connectivity, structural connectivity,

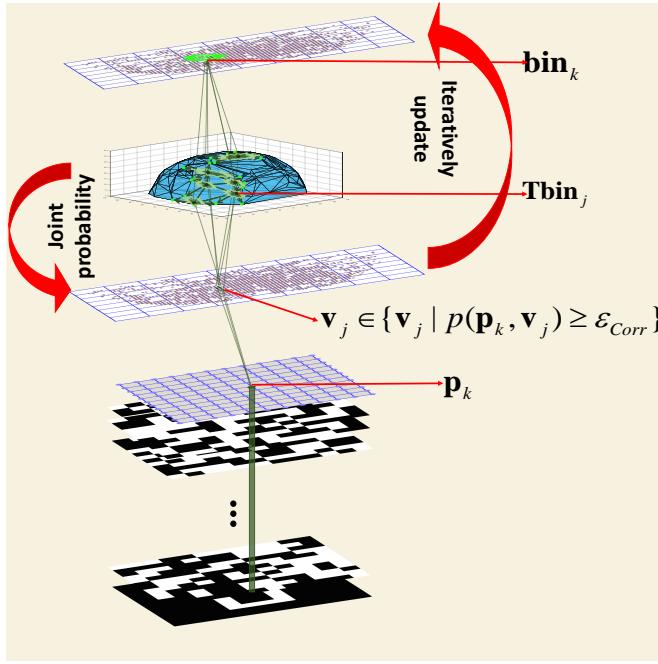


Fig. 3. The updating procedure of the CorrNet framework includes the following: generating the topological structure to obtain the topological neighborhood  $Tbin_j$  of  $v_j$ ; finding the strength correlation set of  $p_k$ ; using the topological correlation and strength correlation to update the probabilistic correlation graph; and getting the  $bin_k$ .

and effective connectivity. In the CorrNet framework, when analyzing the strength information of voxels under diverse stimuli, the correlation between brain activities and visual stimuli can be obtained to reveal the functional connectivity in the early visual cortex. Once we model the topological correlation underlying the structure information of voxels, the structural connectivity can be viewed. Therefore, the effective connectivity can be inferred by the CorrNet framework, which adequately considers the correlations of collected data and contains the reliable model parameters helping to best reconstruct the perceptual images.

First, the correlation among visual stimuli are visualized as reconstruction results, and can be verified through experiments. Then, we generate a probabilistic graph in a joint space to analyze the the correlation among brain activities as well as the correlation between brain activities and visual stimuli. In detail, the correlation among brain activities is denoted as a binary matrix  $Cv \in \mathbb{R}^{D_1 \times D_1}$ .  $Cv$  is determined by the two factors of the voxels (strength information  $S$  and structure information  $POS$ ).  $Cv(j, m)$  is the value of the element in the  $j^{th}$  row and  $m^{th}$  column of the matrix, and defines the correlation between voxel  $v_j$  and voxel  $v_m$ . It is obtained in the three-dimensional world coordinate frame:

$$Cv(j, m) = \begin{cases} 1 & , prob_{sp}(S, POS)_{j,m} \geq \varepsilon_{Cv} \\ 0 & , \text{otherwise} \end{cases}$$

$$prob_{sp}(S, POS) = prob_s(S|POS)prob_{pos}(POS) \quad (3)$$

where  $prob_{sp}(S, POS) \in \mathbb{R}^{D_1 \times D_1}$  is the voxel correlation

probability, measuring the voxel correlation among brain activities. Its element  $prob_{sp}(S, POS)_{j,m}$  is the probability that voxel  $v_j$  and voxel  $v_m$  are of functional and structural correlation.  $prob_s(S|POS) \in \mathbb{R}^{D_1 \times D_1}$  is the strength correlation probability, measuring the strength correlation, whose element  $prob_s(S|POS)_{j,m}$  is the probability that voxel  $v_j$  and voxel  $v_m$  are of strength correlation, given by

$$prob_s(S|POS)_{j,m} = \frac{\left| \sum_{i=1}^N (v_{i,j}v_{i,m}) - \frac{\sum_{i=1}^N v_{i,j} \sum_{i=1}^N v_{i,m}}{N} \right|}{\sqrt{\sum_{i=1}^N v_{i,j}^2 - \frac{\sum_{i=1}^N v_{i,j}^2}{N}} \sqrt{\sum_{i=1}^N v_{i,m}^2 - \frac{\sum_{i=1}^N v_{i,m}^2}{N}}} \quad (4)$$

where  $prob_{pos}(POS) \in \mathbb{R}^{D_1 \times D_1}$  is the topological correlation probability, measuring the topological correlation, whose element  $prob_{pos}(POS)_{j,m}$  is the probability that voxel  $v_j$  and voxel  $v_m$  are of topological correlation. Its computation process will be detailed latter.  $\varepsilon_{Cv}$  is a topological correlation parameter, and is set to 0.1 in our method. To obtain  $prob_{sp}(S, POS)$ , we multiply  $prob_s(S|POS)$  and  $prob_{pos}(POS)$  with Hadamard products. Taking the element  $prob_{sp}(S, POS)_{j,m}$  which is in the  $j^{th}$  row and  $m^{th}$  column in  $prob_{sp}(S, POS)$  as example,  $prob_{sp}(S, POS)_{j,m}$  is a node in the probabilistic map whose value is equal to multiplying the two elements in the  $j^{th}$  row and  $m^{th}$  column of matrix  $prob_s(S|POS)$  and  $prob_{pos}(POS)$ .

1) *Topological correlation*: Because the function of certain brain regions are inevitably dependent on structure, the diverse responses of neurons rely on their various distributions in the cortex. An earlier study proved that adjacent neurons tend to have similar receptive fields and activity patterns [16], [31]. However, almost all existing reconstruction methods neglect the structure information of voxels. Hence, to enrich the correlations in brain decoding, we analyze the structure information of voxels and obtained the prior probability  $prob_{pos}$ . Because there are complex drapes, hierarchies, and numerous neurons in the cortex, it seems impossible to unfold the visual cortex to model the spatial relationship neuron-by-neuron. In this paper, spatial variation is introduced by utilizing the three-dimensional position in the world coordinate frame. In fact, there is exactly a bijective mapping between the voxel position in the cortex and that in the three-dimensional world coordinate frame. Because the visual cortex is a two-dimensional overlapped structure, two voxels that are adjacent in the three-dimensional world coordinate frame may be non-neighboring in the cortex. Inversely, it can be ascertained that voxels adjacent in the cortex are bound to be neighboring in the three-dimensional world coordinate frame. In summary, because the spatial variation from the cortex to the world coordinate frame may shorten the distance between voxels, the topological neighborhood of a certain voxel in the cortex is the subset of that in the world coordinate frame. Based on this conclusion, we compute  $prob_{pos}$  by using the position  $pos_j$  for  $j = 1, \dots, D_1$  directly in the world coordinate frame.

In the world coordinate frame, an undirected graph  $G = \{\mathbb{V}, E\}$  is generated by Delaunay triangulations to present

the topological (structural) connections of all voxels [32].  $\mathbb{V}$  is the set of all voxels  $\mathbf{v}_j$ , and  $E$  presents the set of valid Delaunay edges, which define the topological correlations among voxels. Notably, the reason why we use Delaunay triangulations for the voxels instead of simply the physical distance within the brain is because Delaunay triangulations can reveal hierarchical (superficial) connectivity, consistent with the hierarchical structure of the cerebral cortex. Therefore, the topological method, instead of the anatomy method, is utilized to simulate the structural connectivity in the brain. Only based on the hierarchical connectivity revealed by Delaunay triangulations can the physical distance be meaningful to the simulation. These Delaunay triangulations guarantee that the nearest neighbor graph is a subgraph of the Delaunay triangulation. Thus, the topological neighborhood of the  $j^{th}$  voxel can be found and is denoted as  $\mathbf{Tbin}_j$  for  $j = 1, \dots, D_1$ .  $\mathbf{Tbin}_j$  is defined as  $\mathbf{Tbin}_j = \{\mathbf{v}_m | E(\mathbf{pos}_j, \mathbf{pos}_m) = 1, m \in \{1, \dots, D_1\}\} \cup \{\mathbf{v}_l | d(j, l) \leq \varepsilon_{d_j}, l \in \{1, \dots, D_1\}\}$ , where  $d(j, l)$  presents the Euclidean distance between  $\mathbf{v}_j$  and  $\mathbf{v}_l$  in the world coordinate frame, and  $\varepsilon_{d_j}$  is a self-adaption distance parameter and varies from diverse pattern representation models. The value of  $\varepsilon_{d_j}$  is used to adjust the influence of the topological correlation in the CorrNet framework, and a large value means that the topological correlation has a high influence. Finally,  $prob_{pos}(POS)$  is obtained as below.

$$prob_{pos}(POS)_{j,m} = \begin{cases} 1 & , \mathbf{v}_m \in \mathbf{Tbin}_j \\ 0 & , \text{otherwise} \end{cases} \quad (5)$$

Because topology of the brain and structure connectivity of voxels are partially fixed [16], [9], the topological correlation probability is binary (i.e. there are two possibilities: voxels are of topological neighborhood or not).

2) *Probabilistic correlation graph*: The objective correlation matrix  $\mathbf{Corr}$  is determined by the correlation pairs  $(\mathbf{p}_k, \mathbf{bin}_k)$  for  $k = 1, \dots, D_2$  in a probabilistic correlation graph, denoted  $G_{Corr} = \{\mathbb{V}\mathbb{P}, E_{Corr}\}$ .  $\mathbb{V}\mathbb{P}$  is the set of all voxels and pixels.  $E_{Corr}$  represents the weighted edges that define the correlation connections between  $\mathbf{v}_j$  and  $\mathbf{p}_k$ . The weight values are equal to the probability that  $\mathbf{v}_j$  belongs to  $\mathbf{bin}_k$ , denoted as  $prob_{corr}(\mathbf{p}_k, \mathbf{v}_j | \mathbf{bin}_k)$ . As defined above,  $prob_{corr}(\mathbf{p}_k, \mathbf{v}_j | \mathbf{bin}_k)$  is the corresponding correlation probability between  $\mathbf{v}_j$  and  $\mathbf{p}_k$ . It can be obtained by Bayes rules:

$$prob_{corr}(\mathbf{p}_k, \mathbf{v}_j | \mathbf{bin}_k) = \frac{prob_{bin}(\mathbf{bin}_k | \mathbf{p}_k, \mathbf{v}_j) prob_{corr}(\mathbf{p}_k, \mathbf{v}_j)}{prob_{bin}(\mathbf{bin}_k)} \quad (6)$$

where  $prob_{bin}(\mathbf{bin}_k | \mathbf{p}_k, \mathbf{v}_j)$  is dependent on the topological correlation and will be detailed later;  $prob_{bin}(\mathbf{bin}_k) = 1$  is fixed, and means there must exist  $\mathbf{bin}_k$  for  $\mathbf{p}_k$ .  $prob_{corr}(\mathbf{p}_k, \mathbf{v}_j)$  is the *strength correlation probability*, de-

### Algorithm 1 Generating the Probabilistic Graph $G_{Corr}$ .

#### Input:

Matrix  $\mathbf{Cv}$ , the correlation parameter  $\varepsilon_{Corr}$ , and the strength correlation probability  $prob_{corr}(\mathbf{p}_k, \mathbf{v}_j)$

#### Output:

```

 $prob_{corr}(\mathbf{p}_k, \mathbf{v}_j | \mathbf{bin}_k)$ 
1: Initial  $prob_{bin}(\mathbf{bin}_k | \mathbf{p}_k, \mathbf{v}_j) = 1$ 
    $prob_{corr}(\mathbf{p}_k, \mathbf{v}_j | \mathbf{bin}_k) = prob_{corr}(\mathbf{p}_k, \mathbf{v}_j)$ 
    $L^{(k)} = \emptyset (k = 1 \rightarrow D_2, j = 1 \rightarrow D_1)$ 
2: for  $k = 1$  to  $D_2$  do
3:   for  $j = 1$  to  $D_1$  do
4:     if  $prob_{corr}(\mathbf{p}_k, \mathbf{v}_j) \geq \varepsilon_{Corr}$  then
5:       Update  $L^{(k)} = L^{(k)} \cup \{m | \mathbf{Cv}(j, m) = 1\};$ 
          $k_t = \min_{m \in L^{(k)}, m \neq j} d(\mathbf{v}_j, \mathbf{v}_m);$ 
         for all  $m \in L^{(k)}$ , calculate:
            $prob_{bin}(\mathbf{bin}_k | \mathbf{p}_k, \mathbf{v}_m) = k_t / d(\mathbf{v}_j, \mathbf{v}_m);$ 
            $p_t = prob_{corr}(\mathbf{p}_k, \mathbf{v}_m) \times prob_{bin}(\mathbf{bin}_k | \mathbf{p}_k, \mathbf{v}_m);$ 
           if  $p_t \geq prob_{corr}(\mathbf{p}_k, \mathbf{v}_j | \mathbf{bin}_k)$  then
              $prob_{corr}(\mathbf{p}_k, \mathbf{v}_m | \mathbf{bin}_k) = p_t;$ 
           else
              $prob_{corr}(\mathbf{p}_k, \mathbf{v}_m | \mathbf{bin}_k) = prob_{corr}(\mathbf{p}_k, \mathbf{v}_j | \mathbf{bin}_k);$ 
           end if
         end if
       end for
     if  $L^{(k)} == \emptyset$  then
        $j = \arg \max_{j^* \in \{1, \dots, D_1\}} prob_{corr}(\mathbf{p}_k, \mathbf{v}_{j^*} | \mathbf{bin}_k);$ 
       Return to step 5;
     end if
      $\mathbf{bin}_k = L^{(k)};$ 
17: end for

```

fined by a full covariance matrix:

$$prob_{corr}(\mathbf{p}_k, \mathbf{v}_j) = \frac{\left| \sum_{i=1}^N (v_{i,j} p_{i,k}) - \frac{\sum_{i=1}^N v_{i,j} \sum_{i=1}^N p_{i,k}}{N} \right|}{\sqrt{\sum_{i=1}^N v_{i,j}^2 - \frac{\sum_{i=1}^N v_{i,j}^2}{N}} \sqrt{\sum_{i=1}^N p_{i,k}^2 - \frac{\sum_{i=1}^N p_{i,k}^2}{N}}} \quad (7)$$

Then, in the CorrNet framework,  $prob_{corr}(\mathbf{p}_k, \mathbf{v}_j | \mathbf{bin}_k)$  are iteratively updated. The procedure is shown in Algorithm.1. After the iteration, if  $prob_{corr}(\mathbf{p}_k, \mathbf{v}_j | \mathbf{bin}_k) \geq \varepsilon_{Corr}$ ,  $\mathbf{v}_j$  is in the inverse receptive field of  $\mathbf{p}_k$ ; namely,  $\mathbf{Corr}(\mathbf{v}_j, \mathbf{p}_k) = 1$ , and  $\mathbf{v}_j \in \mathbf{bin}_k$ .

#### B. Pattern representation model

Most reconstruction methods tend to model on the stimulus-response pair  $(\mathbf{X}, \mathbf{Y})$  and decode the most probable image  $\mathbf{Y}$  from the fMRI signal  $\mathbf{X}$ . In this paper, the pixel-wise retinotopy in V1 was utilized to build a map from the voxel bin  $\mathbf{bin}_k$  to its correlation pixel  $\mathbf{p}_k$  directly, finding the corresponding activity patterns. For pattern representation, the linear SVM has been proved to be effective. In addition, compared with ANNs, SNNs have been proved to be more biomimetic in both information transmission and structure [33]. Therefore, both

the linear SVM model and SNN model are trained to learn brain activity patterns from the correlation pairs, verifying the robustness of the CorrNet framework. In the training process, our models serve as distributed multi-voxel receptive-field models. The activity patterns are learned through constructing the maps in pairs  $(\mathbf{p}_k, \mathbf{bin}_k)$ .  $\mathbf{bin}_k$  is the input, and  $\mathbf{p}_k$  is the expected output. As explained before,  $\mathbf{bin}_k$  contains voxels that are in the inverse receptive field of the  $k^{th}$  pixel  $\mathbf{p}_k$ . Given that there are  $M$  voxels in  $\mathbf{bin}_k$  and the number of training trials is  $N$ , the input of the  $k^{th}$  linear SVM model and the  $k^{th}$  spike neuron model is  $N \times M$  matrix, whose element is the strength value of voxels. That is, the number of training samples is  $N$ , and the number of sample attributes is  $M$ . The expected output is the  $N$ -dimension vector whose element is the pixel value. Additionally, the reconstruction procedure monitors  $\mathbf{bin}_k$  corresponding to the retinotopy map of  $\mathbf{p}_k$  by inverting the receptive-field model. Compared with conventional retinotopy, our pixel-wise method solves the problem of missing activity patterns by using the multi-voxel correlation bin  $\mathbf{bin}_k$ . Meanwhile, the proposed CorrNet framework takes the various correlations into full consideration, guaranteeing the accuracy of the reconstruction models. The submodels in Equation 2 work in parallel, and the pair  $(\mathbf{p}_k, \mathbf{bin}_k)$  contains  $N$  trials in the training process.

1) *Linear support vector machine*: We assume that the pattern representation model is given by the linear SVM to find a linear decision boundary in the feature space, such that

$$\mathbf{p}_k = \mathbf{W}_k^T \mathbf{X}^{(k)} + w_0, k = 1, \dots, D_2, \quad (8)$$

where  $\mathbf{W}_k^T$  defines the linear decision boundaries, and each attribution of  $\mathbf{X}^{(k)} = [\mathbf{x}_1^{(k)}, \mathbf{x}_2^{(k)}, \dots, \mathbf{x}_N^{(k)}]^T$ .  $\mathbf{x}_n^{(k)}$  for  $n = 1, \dots, N$  is the  $M$ -dimension vector, which presents the strength value of the  $M$  voxels in  $\mathbf{bin}_k$  for the  $n^{th}$  trial.  $\mathbf{W}_k^T$  is dependent on the support vectors of  $\mathbf{bin}_k$ , which are exactly active voxels corresponding to a certain activity pattern. Therefore, the linear SVM model can highlight the activity patterns that tend to be ambiguous because of measurement noise. For this reason, the linear SVM model is used in this paper. Because  $\mathbf{p}_k = [p_{1,k}, p_{2,k}, \dots, p_{N,k}]^T$  and  $p_{n,k} \in \{0, 1\}$ , the solution of  $\mathbf{W}_k^T$  is a binary optimal problem. The final optimal problem [34] is as below.

$$\begin{aligned} \text{minimize : } & J(\mathbf{W}_k) = \frac{1}{2} \|\mathbf{W}_k\|^2 + C \sum_{n=1}^N \xi_n^{(k)} \\ \text{subject to : } & y_n^{(k)} (\mathbf{W}_k^T \Theta(x_n^{(k)}) + w_0) \geq 1 \\ & \xi_n^{(k)} \geq 0, n = 1, \dots, N \end{aligned} \quad (9)$$

Here,  $\Theta$  is a linear kernel function, and  $C$  is the free parameter that affects the trade-off between complexity and the number of nonseparable samples.  $\xi_n^{(k)}$  denotes the slack variables. Notably, for the linear SVM model, we set the self-adaptive distance parameter in the CorrNet framework as follows:

$$\varepsilon_{d_j} = \min_{E(\mathbf{pos}_j, \mathbf{pos}_m)=1} \{d(j, m)\}. \quad (10)$$

2) *Spike neural network*: Because the learned connections of SNNs [33] represent dynamic spatiotemporal relationships derived from fMRI signals, SNNs are more biomimetic than ANNs in both information transmission and structure. Therefore, in this paper, a dynamic evolving SNN is also used as another pattern representation model. The SNN is built in a brain template by mapping the correlation bin  $\mathbf{bin}_k$  into the input template of the SNN. In the SNN,  $D_2$  spiking neuron submodels are built to train the  $D_2$  correlation pairs. In the  $k^{th}$  spiking neuron submodel, the input is a  $N \times M$  matrix with  $N$  training samples( $N$  training trials) and  $M$  sample attributes( $M$  voxels in  $\mathbf{bin}_k$ ). The input is transformed into spike trains, which reveals the spatiotemporal patterns in the correlation pair  $(\mathbf{p}_k, \mathbf{bin}_k)$ . The strength value vector  $\mathbf{x}_n^{(k)}$  for  $n = 1, \dots, N, k = 1, \dots, D_2$  in  $\mathbf{bin}_k$  constitutes a spike train for the  $n^{th}$  trial. In our study, we design a new way to use the BOLD responses as spike trains. Each pattern relates to  $M$  voxels in  $\mathbf{bin}_k$ , and these voxel values are encoded to  $M$  spike times by the following function:

$$t_i = v_{n,i}^{(k)} + 464, i = 1, \dots, M \quad (11)$$

where  $t_i$  denotes the time of the  $i^{th}$  spike in the spike train,  $M$  is the total number of spikes in the spike train, and  $v_{n,i}^{(k)} \in \mathbf{x}_n^{(k)}$  is the strength value of the  $i^{th}$  voxel of  $\mathbf{bin}_k$  in the  $n^{th}$  trial. Specifically, all spike trains of the  $D_2$  spiking neuron models derived from the encoding module includes  $D_2$  patch states, which are injected into  $D_2$  Tempotron neurons [35] in the leaky integrate-and-fire (LIF) model [36]. Because the expected output is binary pixel value, one Tempotron neuron only performs a binary classification. In the  $k^{th}$  spiking neuron submodel, each spike is transmitted by the synapse between the input neuron and output neuron, and each synapse has its own weight  $w^{(k)} \in \mathbb{R}^M$  equivalent to synaptic efficacy. The  $k^{th}$  spiking neuron submodel for training the correlation pair  $(\mathbf{p}_k, \mathbf{bin}_k)$  can be presented as follows:

$$\tau_m \frac{\partial V^{(k)}}{\partial t} = -V^{(k)}(t) + MI(t) \quad (12)$$

where  $\tau_m$  presents the constant membrane time (we chose  $\tau_m = 4 \times \tau_s = 20ms$ ),  $t$  is the time when the spike is produced,  $I_t$  is the current of a leaky integrator at  $t$  time,  $V^{(k)}(t)$  denotes the membrane potential, and  $M$  is the membrane resistance. Here, the spiking threshold  $V_{thre}$  is introduced. When  $V^{(k)}(t) \geq V_{thre}$ , the neuron will fire the spike, and the membrane potential will immediately reset to the reset potential  $V_{rest} = 0.6$ . Then, it maintains  $V^{(k)}(t) = V_{rest}$  for a short period of time, which refers to the absolute refractory period of the biologic neuron [37]. The subthreshold membrane potential is determined by the weighted sum of postsynaptic potentials (PSPs) from all incoming spikes:

$$\begin{aligned} V^{(k)}(t) &= \sum_i w_i^{(k)} \sum_{t_i} K(t - t_i) + V_{rest}, i = 1, \dots, M \\ K(t - t_i) &= V_0 (\exp(-(t - t_i)/\tau_m) - \exp(-(t - t_i)/\tau_s)) \end{aligned} \quad (13)$$

where  $w_i^{(k)}$  and  $t_i$  are, respectively, the synaptic efficacy and fire time of the  $i^{th}$  afferent neuron. As defined in an earlier study [35],  $K$  is a normalized PSP kernel. The physical significance of  $K$  is, without regard to the synaptic efficacies  $w_i$ , how

postsynaptic membrane impacts the voltage when postsynaptic membrane receives the signals from presynaptic membrane. The factor  $V_0$  normalizes PSP kernels to 1 so that unitary PSP amplitudes are given by  $w_i$ .  $K(t - t_i)$  is a causal filter, vanishing for  $t_i > t$ . Because  $\mathbf{p}_k = (p_{1,k}, p_{2,k}, \dots, p_{N,k})^T$  and  $p_{n,k} \in \{0, 1\}$ , two modes corresponding to the two states (0 and 1) are defined: the bright mode, and the dark mode. The learning procedure of a neuron is conducted by changing its synaptic efficacy  $w_i^{(k)}$  when error happens. There are two error modes: the bright error means that a neuron failed to fire when it should fire; otherwise, the dark error happens. In the training process, based on the analysis of different errors, the learning rule can be demonstrated as below.

$$\Delta w_i^{(k)} = \begin{cases} lr \sum_{t_i < t_{\max}} K(t_{\max} - t_i), & \text{bright error} \\ -lr \sum_{t_i < t_{\max}} K(t_{\max} - t_i), & \text{dark error} \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

Here,  $t_{\max}$  represents the time when the voltage reaches the maximum value, and  $lr = 0.005$  is the learning rate. Considering that the SNN requires plenty of spikes to enable the encoding modes, we extend the  $\text{bin}_k$  by increasing the self-adaptive distance parameter  $\varepsilon_{d_j}$  in the CorrNet framework:

$$\varepsilon_{d_j} = \frac{1}{D_1} \sum_{l=1}^{D_1} d(\mathbf{v}_j, \mathbf{v}_l). \quad (15)$$

#### IV. EXPERIMENTS

##### A. Data preparation

Three male subjects participated in the fMRI study. They had an average age of 23, and normal or corrected-to-normal visual acuity. Informed written consent was obtained from all subjects. Each subject participated in two scan sessions. The first session included 352 patterns with random shapes, and the second session had 80 patterns with regular shapes. The random shapes of patterns were mainly used to train the decoding model, while the other group was used to test the performance of the model. Each stimulus pattern consisted of  $12 \times 12$  pixel patches ( $1.13^\circ \times 1.13^\circ$  each). There were two types of patches: a flickering checkerboard (spatial frequency: 1.78 cycles/ $^\circ$ ; temporal frequency: 6 Hz), and a neutral gray area. Each type of patch was randomly used with equal probability. Each pattern formed by different types of patches was presented on a neutral gray background. There was a fixation spot in the center of each stimulus, and we instructed subjects to fixate on it. There were 20 kinds of stimulus patterns in regular shapes, including digital numbers and geometric shapes.

A rear-projection display device was used to present the stimuli outside the scanner. The resolution of the display device was  $800 \times 600$  at 60Hz. Subjects watched the stimuli presented by the device via a mirror placed above the subject and attached to the receptive field (RF) coil. The visual angle was  $13.5^\circ$ . A laptop controlled stimuli presentation using the E-prime software. Each subject underwent two scan sessions.

Random shape stimuli were displayed in the first session. There were 11 runs in the session. Each run began with a 28s null stimulus, followed by presenting 18 stimulus blocks, and ending with a 12s rest period. In each stimulus block, a 6s stimulus period, followed by a 6s rest period, was presented. A 3060s rest period was inserted between two runs. A total of 352 different stimulus patterns were used in this session. Regular shape stimuli were displayed in the second session. There were ten runs, including five geometric-shape runs and five alphabet runs. The block design was the same as in the first scan session. Each specific shape stimulus was repeated six times during the whole scan session. The color of the fixation spot in the center of each stimulus changed randomly during stimulus presentation. Each subject was asked to press a button when the color changed.

A 3.0-Tesla GE MR Scanner was used to collect functional MRI data at the First Affiliated Hospital of Xi'an Jiaotong University. A T1-Weighted, magnetization-prepared rapid-acquisition gradient-echo (MP-RAGE) sequence (TR: 2250 ms; TE: 2.98 ms; TI: 900 ms; Flip angle:  $9^\circ$ ; FOV:  $256 \times 256$  mm; Voxel size:  $1.0 \times 1.0 \times 1.0$  mm), was firstly used to acquire high-resolution structure images of the same slices in echo-planar imaging (EPI) sequence. Then, a T2\*-weighted gradient-echo EPI sequence (TR: 4000 ms; TE: 30 ms; Flip angle:  $80^\circ$ ; FOV:  $192 \times 192$  mm; Voxel size:  $1.875 \times 1.875 \times 3$  mm; Slice gap: 0 mm; Number of slices: 48) was used to collect functional images covering the whole brain. The first three volumes of each run were discarded in order to avoid the noise caused by the MRI scanner's instability. SPM12 was used to preprocess the MRI data. First, slice-timing correction was implemented to correct the differences in slice acquisition times. Second, slice-timing corrected data underwent head motion correction. Then the structure data of the first scan session were coregistered to all data in this session. Next, the structure data in the second scan session was coregistered to those in the first scan session. Finally, all data in the second scan session were coregistered to the structure data in this session.

##### B. Performance evaluation and analysis

The 352 stimulus-response pairs with random shapes were used as the training set, and the other 80 with regular shapes were used as the test set. In the experiments, to demonstrate the reliability and robustness of the CorrNet framework combined with the linear SVM (Our SVM) and the dynamic evolving SNN (Our SNN), we compared our method with baseline methods, that is, the linear SVM without the CorrNet process (Pure SVM) [30] and the SNN without the CorrNet process (Pure SNN) [20]. Pure SVM and Pure SNN both used all 3037 voxels as input for pixel-wise reconstruction. The fixed-basis model [10] and BCCA [11] were also used for comparison.

In addition, we combine our CorrNet framework with the fixed-basis model to generate another new method (CFFB), which makes use of correlation information of voxels to reconstruct pixel values. The fixed-basis model and CFFB are both pixel-wise reconstruction methods. The difference between the fixed-basis model and CFFB is how to obtain

the reconstructed pixel values. In the fixed-basis model, the weights of voxels are optimized by sparse logistic regression [25]. Then, voxels of high weights are chosen to compute the pixel values in a linear regression model as input. Instead, CFFB chooses the related voxels of certain pixel through the CorrNet framework. The self-adaptive distance parameter in CFFB is defined by Eq. 10. As for the linear combination between the multi-voxel patterns of fMRI signals and multi-scale visual representation, CFFB uses the same process as the fixed-basis model. All experiments were implemented in MATLAB R2016b, and used the LIBSVM Toolbox [38].

Fig. 4 shows the topological distribution of the 144 correlation bins, when the self-adaptive distance parameter was defined by Eq. 10 in the CorrNet framework. The average number of voxels in the bins is 30, and 58% of all 3037 voxels are utilized. Obviously, the CorrNet framework can improve the efficiency of the reconstruction method by condensing data. Besides, a heat map is used to reveal the quantitative distribution of voxels by mapping the number of voxels in  $\text{bin}_k$  into the coordinates of  $p_k$  in the image coordinate frame, as shown in Fig. 5. The middle region contains the most voxels, which conforms to the attention mechanism in human perception. Therefore, our CorrNet framework can reveal the attention mechanism by combining the topological correlation with the strength correlation.

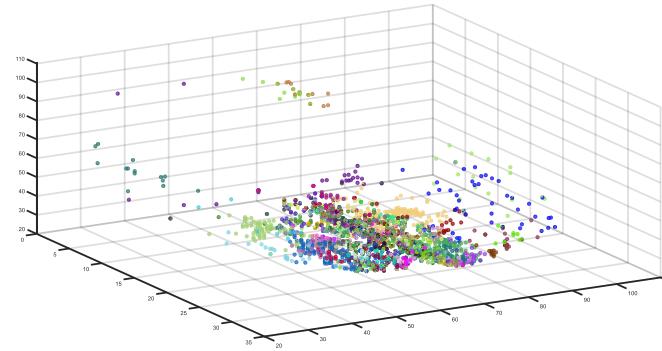


Fig. 4. Topological distribution of all correlation bins. The voxels in the same bin are filled with the same color, and different bins are expressed in different colors.

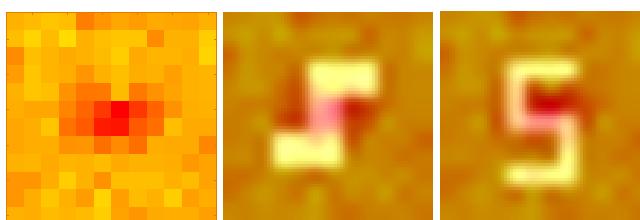


Fig. 5. Quantitative distribution of voxels in the image coordinate frame. The first picture shows the original quantitative distribution, and the more red the pixel patch, the more voxels the corresponding correlation bin contains. The second and third picture shows the perceptual images with a quantitative distribution overlay.

The test set includes geometric shape images and digit images with 144 binary pixels. Different from the baseline methods, the reconstruction results of our method are binary images because the pattern representation is processed as a binary classification problem. Conversely, the results of Miyawaki, BBCA, and CFFB are gray value because the pattern representation is processed as a regression problem. The reconstruction results are shown in Fig. 6, where the first row denotes the original perceptual images, and the rows below are the reconstruction images obtained from our experiments. The reconstruction images obtained by our methods (Our SVM, Our SNN, and CFFB) are much more recognizable than those obtained by other methods (Pure SVM, Pure SNN, Miyawaki and BBCA). Our SVM and Our SNN (in the third and fifth rows) achieve significant improvement over baselines (Pure SVM and Pure SNN in the second and fourth rows); in particular, Our SNN achieves nearly 100% accuracy in the margins of the image coordinate frame. From the perspective of recognizable degree, the CorrNet framework can contribute to the performance promotion for reconstruction tasks. As for the gray images, CFFB (the last row) has successfully reconstructed the relatively legible geometric shapes and digits while Miyawaki and BBCA has nearly failed to separate the meaningful shapes from the background. This result strongly verifies the reliability and robustness of the proposed CorrNet framework. Besides, in the middle area, our methods all perform much better than baseline methods in reconstructing visually recognizable figures, which demonstrates that the attention mechanism can be revealed by our methods (see Figure 5).

Referring to the past reconstruction studies [10], [11], [29], the reconstruction accuracy is computed pixel-by-pixel under Euclidean Distance measurements. The results on the training set, geometry, digit, and the test set are shown in Table I. Both Our SVM and Our SNN can achieve at least a 10% improvement over the baselines (Pure SVM and Pure SNN), indicating the effectiveness of the CorrNet framework. The proposed methods can also avoid the over-fitting problem existing in the Pure SVM. Notably, for pattern representation, the SNN achieves higher average accuracy than the linear SVM, implying the prospect of SNNs in brain decoding. In addition, comparison with Miyawaki and BBCA confirm the superiority of our methods and CorrNet framework. The 85.11% and 84.86% accuracy achieved by Our SNN and CFFB surpasses the state-of-the-art methods (which have around 80% accuracy) in reconstructing perceptual images from fMRI signals.

As mentioned before, we apply SNN model and SVM model to decode cognitive activity patterns from fMRI signals respectively under the CorrNet framework. The main purpose is that, we attempt to illustrate that different classification models can be introduced in the CorrNet framework to realize the reconstruction of visual images. By comparing the two methods, it is proved that it is an effective method to decode fMRI data using SNN model in the CorrNet framework. One of the reasons is that SNN model itself can characterize the neighborhood association of pixels and the dynamic characteristics of the time corresponding neurons. Actually, SNN

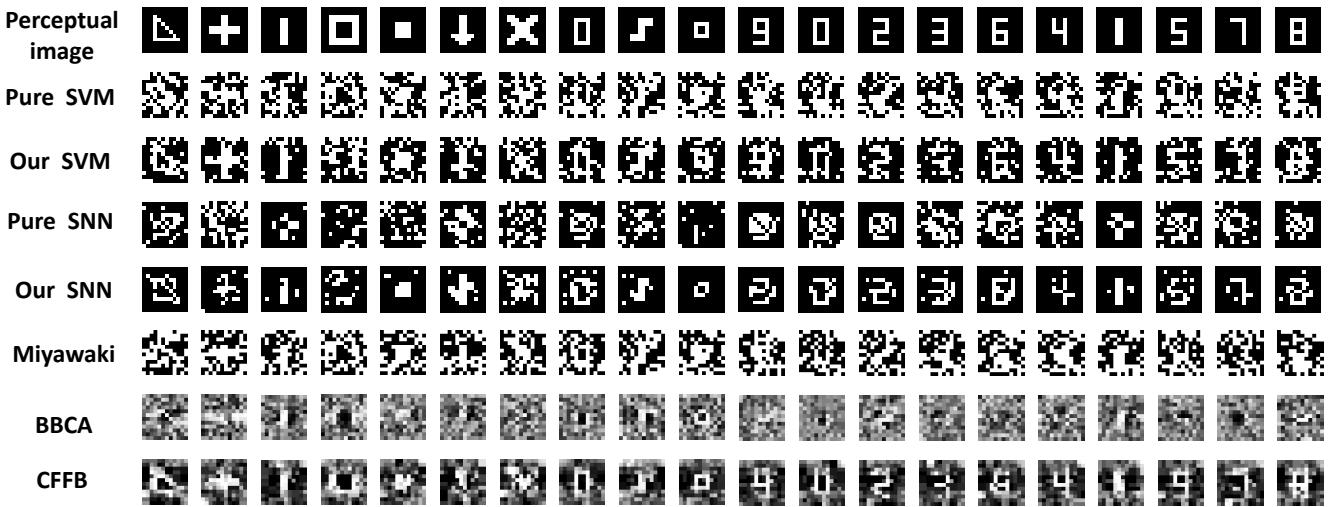


Fig. 6. Image reconstructions of geometric shapes and digits taken from the test set in the experiments. The second to fifth rows are the binary reconstruction images, and the latter rows are the gray reconstruction images

TABLE I  
AVERAGE ACCURACY OF ALL METHODS IN OUR EXPERIMENTS.

Dataset	Pure SVM	Our SVM	Pure SNN	Our SNN	Miyawaki	BBCA	CFFB
Training set	<b>98.87±1.13</b>	78.71±1.04	84.84±2.54	98.02±0.23	95.02±0.67	82.06±1.30	84.94±0.13
Geometry	60.01±3.30	75.60±3.52	73.36±1.27	<b>84.55±0.97</b>	60.36±3.21	73.75±0.27	84.10±0.57
Digit	57.37±3.96	78.09±2.03	78.52±1.67	85.61±1.62	56.17±3.46	73.18±0.96	<b>85.63±0.52</b>
Test set	58.68±3.54	76.85±3.13	75.06±1.41	<b>85.11±1.22</b>	58.92±3.44	73.46±0.57	84.86±0.79

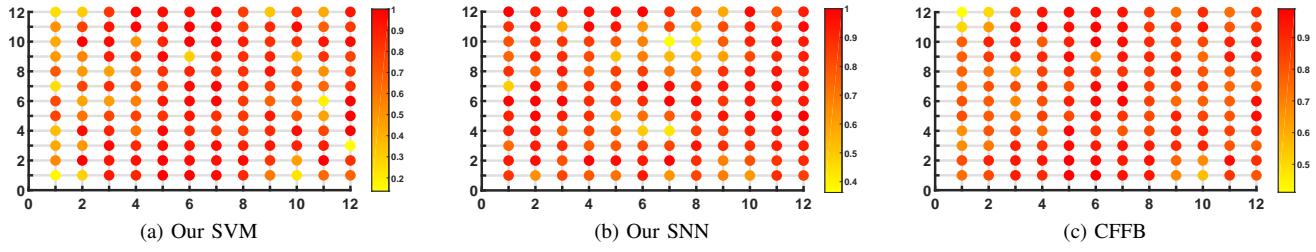


Fig. 7. The accuracy distribution of each pixel. (a)~(c) show the distribution of our SVM, our SNN, and CFFB, respectively. The more red the point, the higher the accuracy of the corresponding pixel.

increases the level of realism in a neural simulation. In terms of neuronal and synaptic state, SNN can incorporate the concept of time into their training process. As it happens with typical multi-layer perceptron networks, neurons in the SNN do not fire at each propagation cycle, but rather fire only when a membrane potential reaches a specific value. When a neuron fires, it generates a signal which travels to others which, in turn, increase or decrease their potentials in accordance with this signal. The operation process of SNN is similar to how voxels in the early visual cortex (mainly V1) response to visual stimulus. The specific process is, when a visual stimulus arrives, voxels in the early visual cortex effect each other based on the structural connectivity. Meanwhile, within the continuous visual tasks, voxels which are sensitive to the visual tasks will present temporal relationships, that is, revealing high correlations in our CorrNet framework. Therefore, in this study, based on the similarity of SNN operation mode and

voxel activity mode, the benefit of the SNN model is that it can simulate biologically-specific neurons in an accurate spike time-encoding operation. However, SVM analyzes data used for classification and regression. In pattern representation process, SVM is used for representing spatial attitude of the brain signals as points in space, and mapping voxels with spatial correlation in one group by a clear gap that is as wide as possible. Therefore, the SVM-based method cannot directly associate the spatial information of neighborhood pixels with the temporal encoding of neuron spikes, and cannot effectively express the spike time-dependent plasticity in synaptic information transmission of neurons.

Also, CFFB, which fuses the CorrNet framework with the fixed-basis model, can achieve relatively satisfactory results. Here are some reasons. First, the multi-scale reconstruction model has been proved more efficient than single-scale models in previous studies [10]. The multi-scale reconstruction

model can extract scale-specific information, analogous to the extraction of orientation information from voxel sampling of cortical columns, potentially. Above all, the proposed CorrNet framework plays the most important role in the great performance promotion (up to 29%). As mentioned before, the fixed-basis model uses pixel-wise sparse logistic regression to optimize the weights of voxels. Then, voxels with high weights are used as input nodes to compute the related pixel values in a linear regression model. Although this method performs well in the earlier studies, it rarely takes the biologic neuron connections into consideration. Differently, our CorrNet framework is based on the perspective of cognition modeling, and chooses the related voxels of certain pixel based on modeling the three connections in human brain. It can be said that, the fMRI data processes of fixed-basis model and our CorrNet framework are derived from totally different thinkings and study directions. The fixed-basis model focuses on classification, but our methods mainly rest on cognition modeling.

Furthermore, in Fig. 7, where the coordinate frame corresponds to the image coordinate frame, the accuracy distribution of a single reconstruction pixel on the test set is shown. The more red points means higher accuracy. Observably, if ignoring some bad points, which can be ascribed to noise, the accuracy of the pixels in the middle of the image coordinate frame is relatively high. Again, the attention mechanism has been revealed. Specifically, more than 80% of pixels reach 0.9 accuracy by Our SNN, and more than 65% of pixels reach 0.75 accuracy by Our SVM and CFFB. Moreover, the proportion of pixels with lower than 0.5 accuracy is reduced to less than 15% by our methods. In particular, Our SNN can achieve comparably comparative accuracy in both the middle and the margin of the image coordinate frame, thus illustrating the topological robustness inside the symmetrical structure of the SNN.

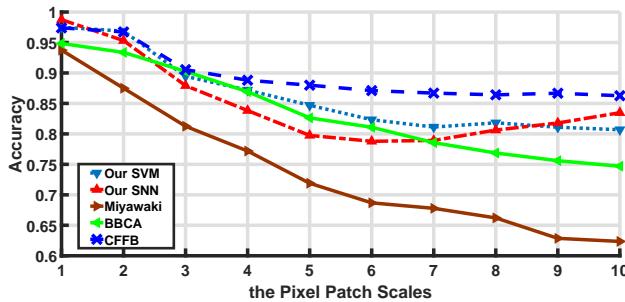


Fig. 8. Reconstruction accuracy of pixel patches with diverse scales in our experiments. The horizontal ordinate denotes the scale values of pixel patches, and the vertical ordinate presents the accuracy. Taking the number 2 in the horizontal ordinate as the instance, the number 2 indicates the middle  $2 \times 2$  pixel region, which is composed of four pixel (6,6), (6,7), (7,6), (7,7) in the  $12 \times 12$  image coordinate frame. Its vertical ordinate presents the average reconstruction accuracy of the four pixels.

As Fig. 8 shows, we compute the pixel-patch accuracy, which denotes the average reconstruction accuracy of the pixel patches with diverse scales ( $1 \times 1, \dots, 10 \times 10$ ). As defined in earlier fMRI study [10], [11], a pixel patch with a certain scale presents a certain region extracted from the

middle to the margin in the image coordinate frame. The accuracy curves of our methods, Miyawaki, and BCCA, are all in decline. Differently, the curves of Our SNN go up after going down. Visibly, considering the obviously distinct gap in accuracy between Miyawaki and CFFB, the proposed CorrNet framework leads to a distinct improvement on every patch scale in the fixed base method. CFFB have obvious superiority over all methods, which implies that multiscale reconstruction has great prospects in brain decoding with compound models. Additionally, our methods tend to provide more smooth curves. This can be owed to the CorrNet framework derived from brain structure and function, which may simulate the robustness in signal process of the human brain.

## V. CONCLUSION

In this paper, we proposed a CorrNet framework for brain decoding in terms of two primary problems now; this framework could be flexibly combined with diverse pattern representation models. As far as we know, topological correlation has been fused with the strength correlation to enrich the correlations for the first time. Then, we represented brain activity patterns by combining the linear SVM and the dynamic evolving SNN separately, with the correlation pairs derived from the CorrNet framework. The experimental results demonstrated that our method could achieve a significant improvement in perceptual image reconstruction over baseline methods. The CorrNet framework was reliable and robust. In addition, the attention mechanism was revealed by the new method. Notably, for the first time, we discussed the three types of connections (functional connectivity, structural connectivity, and effective connectivity) in attention tasks based on fMRI signals.

Our work can contribute to revealing the working mechanism of the human brain, simulating the effective connectivity of the early visual cortex, and establishing mapping from perceptual tasks to voxels. In engineering, this study promotes the development of efficient brain-machine interface. Thus, future work can be focused on hybrid models fusing the CorrNet framework with diverse pattern representation models. Besides, more experiments on other fMRI datasets are required to further improve and generalize the CorrNet framework for practical applications.

## ACKNOWLEDGMENTS

This research was partially funded by the National Natural Science Foundation of China (No. 61773312, L1522023 and No. 91648208), and the 973 Program (No. 2015CB351703).

## REFERENCES

- [1] N.-n. Zheng, Z.-y. Liu, P.-j. Ren, Y.-q. Ma, S.-t. Chen, S.-y. Yu, J.-r. Xue, B.-d. Chen, and F.-y. Wang, "Hybrid-augmented intelligence: collaboration and cognition," *Frontiers of Information Technology & Electronic Engineering*, vol. 18, no. 2, pp. 153–179, 2017.
- [2] J. V. Haxby, M. I. Gobbini, M. L. Furey, A. Ishai, J. L. Schouten, and P. Pietrini, "Distributed and overlapping representations of faces and objects in ventral temporal cortex," *Science*, vol. 293, no. 5539, pp. 2425–2430, 2001.
- [3] Y. Kamitani and F. Tong, "Decoding the visual and subjective contents of the human brain," *Nature neuroscience*, vol. 8, no. 5, pp. 679–685, 2005.

- [4] L. Hausfeld, G. Valente, and E. Formisano, "Multiclass fmri data decoding and visualization using supervised self-organizing maps," *NeuroImage*, vol. 96, pp. 54–66, 2014.
- [5] T. Naselaris, C. A. Olman, D. E. Stansbury, K. Ugurbil, and J. L. Gallant, "A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes," *Neuroimage*, vol. 105, pp. 215–228, 2015.
- [6] K. A. Norman, S. M. Polyn, G. J. Detre, and J. V. Haxby, "Beyond mind-reading: multi-voxel pattern analysis of fmri data," *Trends in cognitive sciences*, vol. 10, no. 9, pp. 424–430, 2006.
- [7] K. N. Kay, T. Naselaris, R. J. Prenger, and J. L. Gallant, "Identifying natural images from human brain activity," *Nature*, vol. 452, no. 7185, p. 352, 2008.
- [8] T. Horikawa and Y. Kamitani, "Generic decoding of seen and imagined objects using hierarchical visual features," *Nature communications*, vol. 8, 2017.
- [9] H.-J. Park and K. Friston, "Structural and functional brain networks: from connections to cognition," *Science*, vol. 342, no. 6158, p. 1238411, 2013.
- [10] Y. Miyawaki, H. Uchida, O. Yamashita, M.-a. Sato, Y. Morito, H. C. Tanabe, N. Sadato, and Y. Kamitani, "Visual image reconstruction from human brain activity: A modular decoding approach," in *Journal of Physics: Conference Series*, vol. 197, no. 1. IOP Publishing, 2009, p. 012021.
- [11] Y. Fujiwara, Y. Miyawaki, and Y. Kamitani, "Modular encoding and decoding models derived from bayesian canonical correlation analysis," *Neural computation*, vol. 25, no. 4, pp. 979–1005, 2013.
- [12] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.
- [13] W. Wang, R. Arora, K. Livescu, and J. Bilmes, "On deep multi-view representation learning," in *Proceedings of the 32nd International Conference on Machine Learning (ICML-15)*, 2015, pp. 1083–1092.
- [14] M. W. Woolrich, M. Jenkinson, J. M. Brady, and S. M. Smith, "Fully bayesian spatio-temporal modeling of fmri data," *IEEE transactions on medical imaging*, vol. 23, no. 2, pp. 213–231, 2004.
- [15] D. Kuang, X. Guo, X. An, Y. Zhao, and L. He, "Discrimination of adhd based on fmri data with deep belief network," in *International Conference on Intelligent Computing*. Springer, 2014, pp. 225–232.
- [16] E. C. Zeeman, "Topology of the brain," 1965.
- [17] S. LaConte, S. Strother, V. Cherkassky, J. Anderson, and X. Hu, "Support vector machines for temporal classification of block design fmri data," *NeuroImage*, vol. 26, no. 2, pp. 317–329, 2005.
- [18] W. Maass, "Lower bounds for the computational power of networks of spiking neurons," *Neural computation*, vol. 8, no. 1, pp. 1–40, 1996.
- [19] N. K. Kasabov, "Neucube: A spiking neural network architecture for mapping, learning and understanding of spatio-temporal brain data," *Neural Networks*, vol. 52, pp. 62–76, 2014.
- [20] N. K. Kasabov, M. G. Doborjeh, and Z. G. Doborjeh, "Mapping, learning, visualization, classification, and understanding of fmri data in the neucube evolving spatiotemporal data machine of spiking neural networks," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 4, pp. 887–899, 2017.
- [21] G.-A. Hosseini-Zadeh *et al.*, "Reconstruction of digit images from human brain fmri activity through connectivity informed bayesian networks," *Journal of neuroscience methods*, vol. 257, pp. 159–167, 2016.
- [22] S. A. Engel, G. H. Glover, and B. A. Wandell, "Retinotopic organization in human visual cortex and the spatial precision of functional mri," *Cerebral cortex (New York, NY: 1991)*, vol. 7, no. 2, pp. 181–192, 1997.
- [23] A. Shmuel, E. Yacoub, D. Chaimow, N. K. Logothetis, and K. Ugurbil, "Spatio-temporal point-spread function of fmri signal in human gray matter at 7 tesla," *Neuroimage*, vol. 35, no. 2, pp. 539–552, 2007.
- [24] O. Yamashita, M.-a. Sato, T. Yoshioka, F. Tong, and Y. Kamitani, "Sparse estimation automatically selects voxels relevant for the decoding of fmri activity patterns," *NeuroImage*, vol. 42, no. 4, pp. 1414–1429, 2008.
- [25] O. Yamashita, "Sparse logistic regression toolbox," 2009.
- [26] G. Flandin and W. D. Penny, "Bayesian fmri data analysis with sparse spatial basis function priors," *NeuroImage*, vol. 34, no. 3, pp. 1108–1125, 2007.
- [27] W. D. Penny, K. J. Friston, J. T. Ashburner, S. J. Kiebel, and T. E. Nichols, *Statistical parametric mapping: the analysis of functional brain images*. Academic press, 2011.
- [28] C. F. Beckmann, M. Jenkinson, and S. M. Smith, "General multilevel linear modeling for group analysis in fmri," *Neuroimage*, vol. 20, no. 2, pp. 1052–1063, 2003.
- [29] S. Schoenmakers, M. Barth, T. Heskes, and M. van Gerven, "Linear reconstruction of perceived images from human brain activity," *NeuroImage*, vol. 83, pp. 951–961, 2013.
- [30] F. De Martino, G. Valente, N. Staeren, J. Ashburner, R. Goebel, and E. Formisano, "Combining multivariate voxel selection and support vector machines for mapping and classification of fmri spatial patterns," *Neuroimage*, vol. 43, no. 1, pp. 44–58, 2008.
- [31] O. Sporns, "Network attributes for segregation and integration in the human brain," *Current opinion in neurobiology*, vol. 23, no. 2, pp. 162–171, 2013.
- [32] L. P. Chew, "Constrained delaunay triangulations," *Algorithmica*, vol. 4, no. 1-4, pp. 97–108, 1989.
- [33] J. Hu, H. Tang, K. C. Tan, and H. Li, "How the brain formulates memory: A spatio-temporal model research frontier," *IEEE Computational Intelligence Magazine*, vol. 11, no. 2, pp. 56–68, 2016.
- [34] C. Cortes and V. Vapnik, "Support vector machine," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [35] R. Güting and H. Sompolinsky, "The tempotron: a neuron that learns spike timing-based decisions," *Nature neuroscience*, vol. 9, no. 3, pp. 420–428, 2006.
- [36] A. M. Andrew, "Spiking neuron models: Single neurons, populations, plasticity," *Kybernetes*, vol. 32, no. 7/8, 2003.
- [37] K. M. Kendrick and R. F. Drewett, "Testosterone reduces refractory period of stria terminalis neurons in the rat brain," *Science*, vol. 204, no. 4395, pp. 877–879, 1979.
- [38] C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines," *ACM transactions on intelligent systems and technology (TIST)*, vol. 2, no. 3, p. 27, 2011.