Coupling Visual Semantics of Artificial Neural Networks and Human Brain Function via Synchronized Activations

Lin Zhao, Haixing Dai, Zihao Wu, Zhenxiang Xiao, Graduate Student Member, IEEE, Lu Zhang, David Weizhong Liu, Xintao Hu, Xi Jiang, Sheng Li, Dajiang Zhu, and Tianming Liu, Senior Member, IEEE

Abstract—Artificial neural networks (ANNs), originally inspired by biological neural networks (BNNs), have achieved remarkable successes in many tasks, such as visual representation learning. However, whether there exist semantic correlations/connections between the visual representations in ANNs and those in BNNs remains largely unexplored due to both the lack of an effective tool to link and couple two different domains, and the lack of a general and effective framework for representing the visual semantics in BNNs such as human functional brain networks (FBNs). To answer this question, we propose a novel computational framework, synchronized activations (Sync-ACTs), to couple the visual representation spaces and semantics between ANNs and BNNs in human brain based on naturalistic functional magnetic resonance imaging (fMRI) data. With this approach, we are able to semantically annotate the neurons in ANNs with biologically meaningful descriptions derived from human brain imaging for the first time. We evaluated the Sync-ACT framework on two publicly available movie-watching fMRI data sets. The experiments demonstrate 1) the significant correlation and similarity of the semantics between the visual representations in FBNs and those in a variety of convolutional neural networks (CNNs) models and 2) the close relationship between CNN’s visual representation similarity to BNNs and its performance in image classification tasks. Overall, our study introduces a general and effective paradigm to couple the ANNs and BNNs and provides novel insights for future studies such as brain-inspired artificial intelligence.

Index Terms—Brain networks, brain-inspired AI, fMRI, human brain function, visual representation, visual semantics.

I. INTRODUCTION

Inspired by the biological neural networks (BNNs), artificial neural networks (ANNs) have achieved great success in a variety of tasks and scenarios due to their powerful representation ability [1]. In the computer vision (CV) field, convolutional neural networks (CNNs) [2] hierarchically learn the visual representations of images/videos as low-level to high-level features in embedding space and have been widely used in many real-world applications [1], [3]. Recent vision transformer (ViT) [4] demonstrates promising performance by representing the image as a sequence of patches and embedding the image patches as latent vectors, based on which the dependencies/correlations among those vectors are modeled. However, the semantics of those embedding spaces for visual representation are not manifest for human perception and challenge us for a comprehensive understanding of representation learning of ANNs. To unveil and describe the semantics of latent space of ANNs for visual representation, increasing efforts have been devoted to interpret the ANNs’ behaviors and annotating their neurons with semantic concepts [5], [6]. For example, Bau et al. [6] proposed to label the hidden units of the convolutional layer with visual concepts from a broad data set. A recent study employed fine-grained natural language description to annotate the semantics of neurons in various ANNs [7]. Despite the remarkable progresses achieved by these methods, whether the visual representation space of ANNs retains biologically meaningful semantics as in the initial inspiration, BNNs, is still an open question.

In the field, researchers now have employed naturalistic functional magnetic resonance imaging (fMRI) to assess the activity and functional mechanism of BNNs [8], [9], [10], [11], e.g., functional brain networks (FBNs), under the naturalistic stimuli, such as real-life images and video streams. This natural stimulus fMRI paradigm provides a powerful tool for investigating the visual perception of human brain and representing the corresponding semantics [9], allowing us to answer the aforementioned question and annotate the neurons in ANNs with biological description even further. However, the current approaches for representing high-dimensional fMRI data, e.g., matrix decomposition based on independent component analysis (ICA) [12] and sparse dictionary learning (SDL) [13], are commonly used for task and...
resting state fMRI. Considering the brain activities encoded by nMRI are dynamic and complex, it is quite challenging to interpret and describe the semantics perceived by the human brain. In addition, the brain responses evoked by naturalistic stimuli exhibit great intersubjects variability [14], [15], while those existing methods do not encode the regularity and variability of different brains, and thus do not offer a general, comparable, and stereotyped embedding space for representing the brain activity and functional semantics. Recently, deep learning approaches demonstrated superior performance in modeling fMRI data [16], [17], [18], [19], [20], [21]. However, as far as we know, these deep learning methods were designed for specific tasks. A more general and effective framework of embedding brain function and representing semantics under naturalistic stimuli is still much needed. In parallel, linking such human brain’s functional embedding and semantics representation with external natural stimulus is very desirable and significant.

In this work, we propose a novel computational framework, synchronized activations (Sync-ACTs), to explore the connections of the visual representation space and semantics between ANNs and BNNs in human brain. Based on Sync-ACT, we describe and annotate the neurons in ANNs with biologically meaningful descriptions for the first time, bridging the gaps between these two drastically different domains. Specifically, as illustrated in Fig. 1, we propose an fMRI embedding framework that represents the brain activity during movie watching as FBNs with temporal activations. The frames at each time point of the corresponding movie are treated as the input into CNN models and the maximum activation values in the features map of each convolutional filter are recorded to form its temporal activation over time. In this way, the activation of convolutional filters and FBNs in the human brain is synchronized, which is supported by the fact that video frames and functional brain responses are intrinsically aligned along the temporal axis during fMRI scan. By correlating the synchronized temporal activation of FBNs and convolutional filters, the embedding spaces of BNNs in the human brain and CNNs are matched and connected. In this way, the semantics in the spaces of two different domains can be used to cross-annotate each other. We evaluate the proposed framework on the publicly available Human Connectome Project (HCP) 7T movie-watching task fMRI data set [22] and StudyForrest movie-watching fMRI data set [23]. The experiments demonstrate significant correlations and the similarity of semantics between the two embedding spaces in a variety of CNN models, suggesting a new paradigm for the interpretability studies of both ANNs and BNNs. It is also found that the similarity of visual representations in CNNs to those in BNNs is closely related to its performances in image classification tasks, which could be potentially utilized for guiding the design of ANNs, i.e., for neural architecture search (NAS) and brain-inspired artificial intelligence.

The contributions of this study are summarized as follows.
1) We proposed a novel Sync-ACT framework to connect the visual representation and semantics between human brain networks and ANNs, bridging the gaps between these two drastically different domains.
2) We introduced a novel fMRI embedding framework that represents brain activity as interpretable FBNs and corresponding temporal activations.
3) We found significant correlations between the two domains’ semantics and the relationship with the CNNs performance in image classification tasks, providing novel insights for future studies, such as brain-inspired artificial intelligence.

II. RELATED WORKS

A. Visual Representation Interpretation

Interpreting the behavior of deep neural networks and the learned visual representation has attracted growing interest in the CV field. As the semantics of visual representation in deep networks are not manifest for human perception, a possible approach is to visualize the activation of neurons and characterize the visual concept they recognize [6], [24], [25], [26]. For example, Bau et al. [6] labeled the neurons (i.e., convolutional filters) of CNNs with visual concepts by aligning the CV field. As the semantics of visual representation in deep networks are not manifest for human perception, a possible approach is to visualize the activation of neurons and characterize the visual concept they recognize [6], [24], [25], [26]. For example, Bau et al. [6] labeled the neurons (i.e., convolutional filters) of CNNs with visual concepts by aligning the CV field. As the semantics of visual representation in deep networks are not manifest for human perception, a possible approach is to visualize the activation of neurons and characterize the visual concept they recognize [6], [24], [25], [26]. For example, Bau et al. [6] labeled the neurons (i.e., convolutional filters) of CNNs with visual concepts by aligning the CV field. As the semantics of visual representation in deep networks are not manifest for human perception, a possible approach is to visualize the activation of neurons and characterize the visual concept they recognize [6], [24], [25], [26]. For example, Bau et al. [6] labeled the neurons (i.e., convolutional filters) of CNNs with visual concepts by aligning the CV field. As the semantics of visual representation in deep networks are not manifest for human perception, a possible approach is to visualize the activation of neurons and characterize the visual concept they recognize [6], [24], [25], [26]. For example, Bau et al. [6] labeled the neurons (i.e., convolutional filters) of CNNs with visual concepts by aligning the CV field. As the semantics of visual representation in deep networks are not manifest for human perception, a possible approach is to visualize the activation of neurons and characterize the visual concept they recognize [6], [24], [25], [26]. For example, Bau et al. [6] labeled the neurons (i.e., convolutional filters) of CNNs with visual concepts by aligning the CV field. As the semantics of visual representation in deep networks are not manifest for human perception, a possible approach is to visualize the activation of neurons and characterize the visual concept they recognize [6], [24], [25], [26]. For example, Bau et al. [6] labeled the neurons (i.e., convolutional filters) of CNNs with visual concepts by aligning the CV field. As the semantics of visual representation in deep networks are not manifest for human perception, a possible approach is to visualize the activation of neurons and characterize the visual concept they recognize [6], [24], [25], [26]. For example, Bau et al. [6] labeled the neurons (i.e., convolutional filters) of CNNs with visual concepts by aligning the CV field. As the semantics of visual representation in deep networks are not manifest for human perception, a possible approach is to visualize the activation of neurons and characterize the visual concept they recognize [6], [24], [25], [26].

B. fMRI Data Representation

A major challenge for fMRI data representation learning is that the number of voxels in 4D spatiotemporal fMRI data is greatly larger than the number of subject brains [27]. To deal with this imbalance, a variety of computational tools have been proposed to select the task-related features and discard the redundant ones as well as the noises [12], [13]. For example, ICA [12] and SDL were employed to decompose the fMRI as two compact matrices (temporal and spatial patterns). However, the temporal and/or spatial patterns obtained from ICA or SDL-based methods are not intrinsically comparable across different individual brains. Recently, deep
learning has been widely employed in fMRI data modeling and achieved superior results over the traditional matrix decomposition methods [16], [17], [18], [19], [20], [21]. However, as far as we know, prior deep learning models of fMRI data were not specifically designed toward a general, comparable, and compact representation of brain function. Instead, prior methods were designed for some specific tasks, such as fMRI time-series classification [18], brain network decomposition [19], [28], and state differentiation [16], among others. Even though some methods derive comparable temporal patterns [21], [28], which might be suitable for our objective, they still rely on matrix decomposition to obtain spatial patterns that are not comparable across different individuals. In this work, we proposed a more general and unified framework to represent the fMRI data from different subjects as FBNs and their temporal activations in a general, comparable, and stereotyped latent space. This design enables us to explore the correlation between the semantics of this latent space and those in CNNs.

C. Connection of ANNs and BNNs

Current ANNs are inspired by the BNNs at the beginning. For example, CNNs are inspired by the hierarchical organization of the vision systems in the cat’s visual cortex [2], [29], [30]. Recently, there is a growing interest in exploring the potential connections between ANNs and BNNs. For instance, the receptive field analysis reveals that the receptive fields of filters in CNNs become progressively larger [31] and more complex, which is similar to the ventral pathway in cerebral cortex [32]. The filters in the last convolutional layer have class-specific receptive fields akin to concept cells in the visual cortex [33]. Yamins et al. [34], [35] synthesized CNNs outputs by linear regression to predict the neural responses in both the V4 and inferior temporal (IT) cortex. This shows a strong correlation between a CNN’s categorization performance and its ability to predict individual IT neural responses, implicitly indicating the potential representation similarity between ANNs and BNNs. Du et al. [36] proposed a hierarchically structured neural decoding framework to predict CNN features and reconstruct the perceived natural images and faces. Zhou et al. [37] evaluated the brain-like properties of CV models from the perspective of explaining brain activities of the human visual cortex triggered by dynamic stimuli. Representational similarity analysis (RSA) [38], [39] is another approach that has been used to examine the similarity between the activation patterns of different neural systems, such as human brain and CNNs [40], [41], [42]. A close correspondence in the representational structure of human visual areas to CNN layers was reported in the RSA-based studies [40], [41]. You et al. [43] proposed a graph-based representation of ANNs called relational graph and found that top-performing ANNs have graph structures similar to those of BNNs. Inspired by these studies, we explore the semantic similarity of visual representations in CNNs and the functional representation of the human brain, providing a novel insight into the connection between ANNs and BNNs.

III. METHODS

A. Formulation of Sync-ACT Framework

Even though the ANNs are originally inspired by the BNNs, the input/output, operating, and reasoning processes of neural networks in the two domains are quite different and not comparable. Our intuition is to avoid being trapped by the remarkable differences but focuses on their responses, such as the activation of neurons, to the external stimuli. In this way, the behavior of the neural networks measured by the responses (i.e., the temporal activation of neurons) in two domains can be directly compared if the stimuli are synchronized, and thus the most similar neurons in two domains can be easily identified and paired. Let \( \mathcal{F} : X_a \rightarrow Y_a \) represents an ANN, and \( f_i(x_a) \) represents the temporal activation of neuron \( f_i \) with respect to stimulus sequence \( x_a \). Similarly, let \( \mathcal{G} : X_b \rightarrow Y_b \) represent a BNN, and \( g_j(x_b) \) denotes the temporal activation of neuron \( g_j \) with the stimulus sequence \( x_b \). If the stimuli \( x_a \) and \( x_b \) are synchronized, the paired neuron of \( f_i \) in the BNN \( \mathcal{G} \) can be defined by

\[
\text{Sync-ACT}(f_i, \mathcal{G}) = \arg\max_{g_j \in \mathcal{G}} \delta(f_i(x_a), g_j(x_b))
\]

where \( \delta(\cdot) \) is the measurement of similarity between two temporal activations. Similarly, we could define the paired neuron of \( g_j \) in the ANN \( \mathcal{F} \) as

\[
\text{Sync-ACT}(g_j, \mathcal{F}) = \arg\max_{f_i \in \mathcal{F}} \delta(g_j(x_b), f_i(x_a)).
\]

In this work, we adopt the Pearson correlation coefficient (PCC) for similarity measurement \( \delta(\cdot) \). Based on (1) and (2), we can obtain the paired neuron \( g_j/f_i \) for any \( f_i/g_j \) by choosing the one with the most significant similarity value. With the neuron pairs, the semantics of one neural network can be used to annotate the other, i.e., the cross-annotation. We define the semantic description of neurons \( f_i/g_j \) as \( d_{f_i}/d_{g_j} \). The cross-annotation of paired neurons is then denoted as \( d_{f_i} \rightarrow d_{g_j} \) and \( d_{g_j} \rightarrow d_{f_i} \).

We adopted nfMRI data to evaluate the Sync-ACT framework by leveraging the fact that, during nfMRI scan, video frames (stimuli for both ANNs and BNNs) and functional brain responses measured by fMRI are temporally aligned in an intrinsic fashion. In Section III-B, we detail a feasible approach about how to derive the BNNs from nfMRI, define neurons in BNNs, and introduce how to obtain the corresponding temporal activations. In Section III-C, we use CNNs as the representative ANNs and convolutional filters as neurons and introduce the way to measure the temporal activations of each filter.

B. fMRI Embedding Framework

In the brain imaging field, it is common to represent the brain function as interactions of FBNs and the corresponding temporal patterns. Thus, the FBNs can be viewed as the neurons of BNN in the human brain, and temporal patterns represent the activations of those FBNs. However, the previous methods including the deep learning ones do not offer a general and stereotyped space in modeling FBNs. The FBNs...
and corresponding temporal activations are not intrinsically comparable across different individual brains.

So, in this section, we propose a general fMRI embedding framework to represent brain function as FBNs and derive the temporal activations in a unified and comparable embedding space. Specifically, the fMRI embedding framework has an encoder–decoder architecture. Fig. 2 illustrates the major components in the encoder. The rearranged 2-D fMRI signal matrix \( S \in \mathbb{R}^{t \times n} \), where \( t \) is the number of time points and \( n \) is the number of voxels, is first embedded as a new feature matrix \( S_f \in \mathbb{R}^{t \times m} \) through a learnable transformation matrix \( W \in \mathbb{R}^{t \times m} \), where \( m \) is the reduced feature dimension \( (m \ll n) \). This transformation can be viewed as compressing the voxels in 3-D volume space into \( m \) components, i.e., \( m \) FBNs, by linear combination. The columns in the transformation matrix \( W \) recorded the contributions of the voxels to each FBN, i.e., the composition of each FBN, which can be mapped back to 3-D volume space for visualizing the spatial pattern of FBN. It is noted that the linear transformation (LT) in the encoder parameterized by \( W \) is optimized in a data-driven manner and consistent for all subjects, which guarantees the comparability of \( S_f \) for all subjects.

The row vectors in matrix \( S_f \) recorded the activation of all resulted FBNs at different time points, and the column vector represents the temporal activation of a specific FBN. We further model the temporal correlations of the column vectors with a neural network \( F(x) \). Here, we explore two popular neural networks for modeling temporal data, long short-term memory (LSTM) [44] and multihead self-attention (MSA) module in the Transformer model [45]. The column vector \( l_i \) in the resulted matrix \( L = F(S_f) \in \mathbb{R}^{t \times m} \) is the temporal activation of the \( i \)th FBNs, which encodes the regularity and variability of different brains in the same latent space. We average the vector \( l_i \) from all subjects in the testing data set as the temporal activation for neuron \( g_i \).

The decoder has a symmetrical architecture as the encoder. All components of the framework are jointly optimized in an unsupervised manner by minimizing the mean square error (MSE) between the original fMRI signals matrix \( S \in \mathbb{R}^{t \times n} \) and their corresponding reconstruction \( S' \in \mathbb{R}^{t \times n} \).

C. Neurons and Activations in CNNs

We adopt CNNs as the representative ANNs in this work because of CNNs’ powerful visual representation ability and wide application in many CV tasks. We recognize the convolutional filters as the neurons in ANN. To derive the temporal activation of CNN’s filters, we adopt a simple but effective strategy by collecting the feature map \( A_{f_i}(x_t) \) of each CNN filter \( f_i \) with image \( x_t \) from the image sequence \( x \) at time point \( t \). Then, the maximum value in feature map \( \text{max}(A_{f_i}(x_t)) \) is extracted to represent the activation degree of filter \( f_i \) at time point \( t \), resulting in the temporal activation \( f_i(x_t) \). When the image sequence \( x \) is the corresponding movie frames of nfMRI, the derived temporal activations are automatically synchronized with the ones in FBNs. This strategy can be easily applied to any pretrained CNN model.

With the obtained temporal activations \( f_i(x_a) \) and \( g_i(x_b) \), given synchronized stimuli \( x_a \) and \( x_b \), we will be able to pair the neurons between ANNs and BNNs and perform cross-annotation following (1) and (2). It is noted that the Sync-ACT is a general framework that is also compatible with temporal activations derived from other representation methods, such as those potentially from other fMRI embedding methods or ViT.

IV. EXPERIMENTS

Data Sets: In this study, we adopt the publicly available HCP 7T movie-watching fMRI data set (http://www.humanconnectomeproject.org/) of S1200 release [58]. The data set contains 184 subjects who were scanned in 4 runs while watching short independent films and Hollywood movie excerpts concatenated into.mp4 files. The important fMRI acquisition parameters are as follows: \( 130 \times 130 \) matrix, 85 slices, \( \text{TR} = 1.0 \) s, \( \text{TE} = 22.2 \) s, 208 mm FOV, flip angle...
TABLE I
AVERAGED (± STANDARD DEVIATION) PCC ON HCP DATA SET FOR THE PAIRS OF FBNs AND THE FILTERS ON CNN MODELS PRETRAINED ON IMAGE-net AND PLACES365 DATA SETS. THE CORRELATIONS MEASURED BY PCC IN THIS TABLE ARE ALL STATistically SIGNIFICANT (p-VALUE = 0.05) FOR DIFFERENT RUNS

| Methods               | ImageNet [46] | Places365 [47] |
|-----------------------|---------------|----------------|
|                       | Run #1        | Run #2        | Run #3        | Run #4        | Run #1        | Run #2        | Run #3        | Run #4        |
| AlexNet [48]          | 0.2323 ± 0.00692 | 0.2223 ± 0.00683 | 0.2558 ± 0.0780 | 0.2607 ± 0.0877 | 0.2651 ± 0.0921 | 0.2374 ± 0.0800 | 0.2788 ± 0.1011 | 0.2774 ± 0.1005 |
| VGG-16 [49]           | 0.2376 ± 0.00686 | 0.2176 ± 0.00559 | 0.2654 ± 0.0880 | 0.2617 ± 0.0715 | -             | -             | -             | -             |
| ResNet-18 [50]        | 0.2415 ± 0.00666 | 0.2267 ± 0.00631 | 0.2516 ± 0.0805 | 0.2530 ± 0.0625 | 0.2703 ± 0.0740 | 0.2536 ± 0.0694 | 0.2896 ± 0.0845 | 0.2931 ± 0.0808 |
| ResNet-50 [50]        | 0.2862 ± 0.00810 | 0.2660 ± 0.07979 | 0.2942 ± 0.0965 | 0.3022 ± 0.0923 | 0.3008 ± 0.0830 | 0.2745 ± 0.0733 | 0.3076 ± 0.0916 | 0.3159 ± 0.0849 |
| DenseNet-161 [51]     | 0.3031 ± 0.00855 | 0.2767 ± 0.0766 | 0.3031 ± 0.0965 | 0.3152 ± 0.0877 | 0.3052 ± 0.0876 | 0.2879 ± 0.0768 | 0.3134 ± 0.1004 | 0.3199 ± 0.0886 |
| Inception V3 [52]     | 0.2720 ± 0.07755 | 0.2615 ± 0.0730 | 0.2747 ± 0.0912 | 0.2895 ± 0.0861 | -             | -             | -             | -             |
| ShuffleNet V2 [53]    | 0.2663 ± 0.0756  | 0.2515 ± 0.0701 | 0.2742 ± 0.0911 | 0.2831 ± 0.0891 | -             | -             | -             | -             |
| MobileNet V2 [54]     | 0.2628 ± 0.0709  | 0.2517 ± 0.0653 | 0.2635 ± 0.0838 | 0.2870 ± 0.0745 | -             | -             | -             | -             |
| ResNeXt-50 [55]       | 0.2774 ± 0.0751  | 0.2607 ± 0.0726 | 0.2904 ± 0.0911 | 0.2940 ± 0.0881 | -             | -             | -             | -             |
| MNASNet [56]          | 0.2612 ± 0.0668  | 0.2422 ± 0.0399 | 0.2669 ± 0.0794 | 0.2669 ± 0.0992 | -             | -             | -             | -             |
| CoRNet-R [57]         | 0.3288 ± 0.0921  | 0.2396 ± 0.0729 | 0.2796 ± 0.0840 | 0.2634 ± 0.0756 | -             | -             | -             | -             |

A. Correlations of Representations in Two Spaces

Correlations With Different CNN Models: We first explored the correlation of temporal activations between the FBNs and convolutional filters in a variety of CNN models pretrained on the ImageNet data set [46] and Places365 data set [47]. Considering that the features in the last convolutional layer are high-level features used for classification and contain the semantics of the image, the filters in the last convolutional layer for all CNNs are selected and paired with FBNs. In Table I, the averaged PCC values over all pairs across different CNN models and different runs of the HCP 7T movie-watching data set are reported. It is observed that for different runs, the averaged PCC values are all larger than 0.2 for different models. However, the PCC values also vary on different pretrained models and runs. To validate such correlations between CNNs and FBNs, we perform a two-tailed significance test (confidence level=0.95) with null hypothesis that there is no correlation between them. The results show that the correlations between the visual representation of CNNs and FBNs are significant for all pairs across different models pretrained on different data sets (ImageNet and Places365).

We perform a similar analysis on the StudyForrest movie-watching fMRI data set for validation, and the results are reported in Table II with CNN models pretrained on the ImageNet data set. However, the averaged PCC values are smaller than those on the HCP 7T movie-watching data set. This might be due to that the number of subjects in the StudyForrest data set (15) is smaller than those in the HCP 7T fMRI data set (184); the spatial/temporal resolution, image quality, and signal-to-noise ratio of nfMRI data are worse than those in the HCP 7T fMRI data set. However, it is still found that the correlations are significant for almost all models and runs of fMRI except the AlexNet, VGG-16, and ResNet-18 on run #1 and/or run #2. Overall, these results consistently suggest that there exists a significant correlation between the convolutional filters in the CNN model and FBNs in the human brain.

Correlations With Different CNN Layers: We further assess the correlations of FBNs with convolutional filters in different layers of four different CNN models. For these models, their architectures can be roughly divided into four blocks. The PCC values averaged over all FBN–filter pairs in each...
block are reported in Table III. We observe that the correlations are significant for pairs in the last two blocks/layers while some of them in the first two blocks/layers are not significant. The PCC values in different layers also show a trend that it increases and reaches a peak at the third layer, which is in line with the literature study [35] reporting that the model's intermediate layers are highly predictive of the brain's neural responses.

B. PCC Variance in Different CNN Models

From Tables I and II, we can observe that the ResNet-50 and DenseNet-161 have higher PCC values than the VGG-16 and ResNet-18. Considering ResNet-50 and DenseNet-161 have a better performance in the image classification task, it may suggest that the similarity of a model’s visual representation to brain's visual representation might be correlated with its performance on image classification task. To verify this hypothesis, we conduct linear regression to model the relationship between the averaged PCC values and the CNN model's top-1 accuracy on the ImageNet classification task. The results are reported in Fig. 3. We found that the relationship between averaged PCC values and top-1 accuracy can be modeled by a linear function, with $R^2$ larger than 0.5 across all four runs of the HCP 7T movie-watching task fMRI data set. These results indicate that if the visual representations of CNN models are more similar to those of human brain function, its performance on image classification tasks will be better. This finding is consistent with a study by You et al. [43], which found that top-performing ANNs have a graph structure similar to that of BNNs.

C. Visualizations of the Cross-Annotation

We conduct the cross-annotation based on (2) to pair each FBN with a filter at the last convolutional layer of ResNet-18 and visualize several sample pairs in Fig. 4. The left panel in Fig. 4 shows the FBNs to be paired and the corresponding semantic description from fMRI meta-analysis. The right panel shows the most activated images obtained by [6] from the movie frame sequence of paired CNN filters. The filter’s corresponding semantic description and representative images are also demonstrated. In Fig. 4, we found some interesting connections between the semantic description of the paired FBN and filters. For example, the description of FBN #25 is related to place and navigation, while the paired filters are labeled as rock and the representative images are related to some natural scenes. Such observation is obvious on some
Fig. 3. Linear regression modeling the relationship between PCC and CNN's top-1 image classification accuracy on ImageNet. Different CNN models are marked as circle with different colors.

Fig. 4. Visualization of FBN–filter pairs obtained from our model. The left panel is the FBNs to be paired and semantic description from fMRI meta-analysis. The colors of different brain regions indicate the activation degree from low (blue) to high (red). The middle panel shows the Sync-ACTs from FBN and paired CNN filter. The right panel shows the most activated frames and the corresponding semantic description and filter’s representative images in [6].

pairs, which is congruent with the results in Section IV-A. We provided more samples in supplementary materials for comparison.

D. Ablation Studies of fMRI Embedding Framework

We conduct the ablation studies for our fMRI embedding framework on three variants: 1) encoder/decoder only has one LT layer; 2) encoder/decoder has one LT layer and two LSTM layers with tanh activation function, and the hidden size of both layers is set as 64; and 3) encoder/decoder has one LT layer followed by MSA module. We measure the similarity of temporal activations of FBN–filter pairs identified by PCC. The averaged values for different metrics over all pairs and runs are reported in Table IV. Overall, the similarity of LT+LSTM and LT+MSA is larger than the LT baseline. The LT+LSTM and LT+MSA have comparable performances in terms of similarity. However, the LT+MSA has better FBNs quality. We provide the details in the supplementary materials. The results and analysis of our work are based on LT+MSA.

E. Delay of Hemodynamic Response

Brain activity measurements using fMRI are consistently delayed compared to the stimuli due to the delay of hemodynamic response (HDR) [62]. In order to determine
TABLE V
AVERAGED (+ STANDARD DEVIATION) PCC FOR THE PAIRS OF FBNs AND FILTERS WITH DIFFERENT HEMODYNAMIC DELAYS AND THE RATIO OF NOT STATISTICALLY SIGNIFICANT PAIRS. THE COLORS RED AND BLUE DENOTE THE HIGHEST AND THE SECOND-HIGHEST PCC VALUE AMONG DIFFERENT DELAYS, RESPECTIVELY.

| Methods | Delays | Run 1 | Run 2 | Run 3 | Run 4 |
|---------|--------|-------|-------|-------|-------|
|         |        | PCC   | Ratio | PCC   | Ratio | PCC   | Ratio | PCC   | Ratio |
| ResNet-18 [50] | 0s     | 0.2415 ± 0.0666 | 0/64 | 0.2267 ± 0.0631 | 0/64 | 0.2516 ± 0.0805 | 0/64 | 0.2530 ± 0.0625 | 0/64 |
|         | 2s     | 0.2686 ± 0.0724 | 0/64 | 0.2705 ± 0.0751 | 0/64 | 0.2744 ± 0.0853 | 0/64 | 0.2773 ± 0.0669 | 0/64 |
|         | 4s     | 0.2808 ± 0.0721 | 0/64 | 0.2715 ± 0.0812 | 0/64 | 0.2920 ± 0.0874 | 0/64 | 0.2821 ± 0.0723 | 0/64 |
|         | 6s     | 0.2730 ± 0.0649 | 0/64 | 0.2586 ± 0.0758 | 0/64 | 0.2725 ± 0.0870 | 0/64 | 0.2639 ± 0.0724 | 0/64 |
| ResNet-50 [50] | 0s     | 0.2862 ± 0.0810 | 0/64 | 0.2660 ± 0.0797 | 0/64 | 0.2942 ± 0.0965 | 0/64 | 0.3022 ± 0.0923 | 0/64 |
|         | 2s     | 0.3139 ± 0.0845 | 0/64 | 0.2950 ± 0.0910 | 0/64 | 0.3202 ± 0.1013 | 0/64 | 0.3234 ± 0.1040 | 0/64 |
|         | 4s     | 0.3387 ± 0.0840 | 0/64 | 0.3223 ± 0.0993 | 0/64 | 0.3440 ± 0.1000 | 0/64 | 0.3386 ± 0.1101 | 0/64 |
|         | 6s     | 0.3254 ± 0.0825 | 0/64 | 0.3095 ± 0.0969 | 0/64 | 0.3247 ± 0.0965 | 0/64 | 0.3214 ± 0.1058 | 0/64 |

TABLE IV
COMPARISON OF TEMPORAL ACTIVATIONS SIMILARITY OF FBN–FILTER PAIRS IDENTIFIED BY PCC. THE COLORS RED AND BLUE DENOTE THE BEST AND THE SECOND-BEST RESULTS, RESPECTIVELY.

| Abbreviations | MAE↓ | MSE↓ | RMSE↓ | DTW↓ | PCC↑ |
|---------------|------|------|-------|------|------|
| a) LT         | 0.9781 | 1.5309 | 1.2360 | 18.0688 | 0.2346 |
| b) LT+LSTM    | 0.9392 | 1.4411 | 1.1986 | 18.4683 | 0.2794 |
| c) LT+MSA     | 0.9658 | 1.5136 | 1.2289 | 18.7328 | 0.2432 |

such delay and investigate its potential impact on our experiment, we synchronize the delayed activation from BNNs with activations from ANNs by assuming that the delay time is 0 s (no delay), 2 s, 4 s, and 6 s, respectively. For each assumption, we utilize ImageNet pretrained ResNet-18 and ResNet-50 models and compare the average PCC values across different delay settings. The comparison results are presented in Table V.

It is observed that the best results are achieved when the delay time is 4 s (in terms of the highest PCC values, highlighted with red color). This experiment result is consistent with previous fMRI study that the fMRI response evoked by visual stimuli peaks 4–6 s following stimulus [63]. Despite the variance in PCC values across different delay settings, the correlation between FBNs and CNN filters across different pairs and delays is consistently significant, demonstrating the connection of semantics between the visual representations in CNNs and those in the human brain. It is noted that 4-s delay might not always be the optimal delay time for natural stimulus studies, and the optimal delay also varies across individuals and task conditions. For unity, we report the results with no delays in our experiments.

V. DISCUSSION
Interpretability: The proposed Sync-ACT framework matches and pairs the neurons in ANNs and BNNs, based on which the cross-annotation is performed to annotate the neurons in one domain with the semantic description in the other. The Sync-ACT framework opens a new paradigm for the interpretability studies of ANN by using the prior knowledge in neuroscience to interpret ANNs. In parallel, we can understand the dynamic function of FBNs with visual and language descriptions from the paired ANNs’ neurons in a direct way, providing a novel way for unveiling the complex brain function.

NAS: One important finding of this study is that the performance of CNNs on image classification tasks is closely related to its visual representation similarity with the human brain. In [64], [65], and [66], the typical evaluation criteria for NAS is the performance of the searched neural network on a specific task. Our Sync-ACT framework provides new inspirations: the ANN’s representation similarity to the human brain could be a reliable and meaningful criterion for NAS and thus guide the NAS approaches to improve interpretability and performance. Our Sync-ACT framework contributes to the emerging field of brain-inspired AI, i.e., using domain knowledge of brain science to inspire and guide the design of AI models.

Comparison With Other Methods: RSA is an important approach that has been used to examine the similarity between the activation patterns of different neural systems, including the human brain and CNNs [38], [39], [40], [41], [42]. Both Sync-ACT framework and RSA method are developed for examine the similarity between two domains, and we do not have ground truth for such similarities. So it is hard to perform a quantitative comparison between Sync-ACT framework and RSA to conclude which method (Sync-ACT or RSA) is better to represent the similarity between the brain and CNNs. Theoretically, one advantage of our work is that we propose a general fMRI embedding framework to represent brain function as FBNs and derive the temporal activations in a unified stereotyped, and comparable embedding space. However, RSA still relies on traditional methods such as standard first-level GLM to estimate individual beta maps as the brain embeddings. Besides, RSA is based on representation dissimilarity matrix (RDM) which focuses on visual representational structures, while Sync-ACT framework predefines the “neurons” and focuses on the temporal activation patterns of those neurons. For the encoding model such as Yamins et al. [35], the analysis of similarity is also based on RSA. Despite the methodological differences, our findings are consistently with those based on RSA. For example, Khaligh-Razavi and Kriegeskorte [40] and Cichy et al. [41] reported a close correspondence in the representational structure of lower and higher...
human visual areas to lower and higher CNN layers, respectively. Meanwhile, some studies also demonstrated that CNNs do not fully capture higher level visual representations of real-world objects [42] and CNN is not optimal for modeling the human visual pathway [37]. These findings are also parallel with our results that PCC values are around 0.2 and 0.3 which is not high. Overall, our results are consistent with previous works based on RSA or other methods, indicating the validity of the proposed framework.

Noise Ceiling Estimation: The noise ceiling of natural stimulus fMRI (ns-fMRI) paradigm such as movie-watching used in this work is relatively hard to estimate. An underlying assumption for noise ceiling estimation is that the stimulus responses are fixed and differ only in noise. However, in the ns-fMRI paradigm, the change in responses is the variable of interest rather than noise, compared with the block-design paradigm. Here, we tentatively borrow the idea from RSA to estimate the noise ceiling. We obtain the embedding matrix for each subject and use the leave-one-out strategy to split the subjects. For each split, we compute the correlation of the left-out embedding matrix to the average of the other embedding matrices as the lower noise ceiling [39]. The average over all splits is our final estimate for the noise ceiling [39]. The average over all splits is our final estimate for the noise ceiling, which is $0.8089 \pm 0.0264$, $0.8172 \pm 0.0254$. However, the response of different subjects to the same stimuli can be different, which should not be considered as noise. So such estimation might be still problematic. And it is still an opening question for estimating the noise ceiling in ns-fMRI, which deserves more attention and efforts in the future.

Limitations: Our approach has several potential limitations.

1) We use the maximum value in the feature map to represent the activation degree of convolutional filters. Currently, how to characterize the activation degree of filters is still an open question.

2) The semantics descriptions from meta-analysis [67] and Bau et al. [6] for neurons in FBNs and CNNs are coarse-grained and ad-hoc (e.g., Fig. 4, unit #440, Bank Vault) due to their intrinsic limitations. More fine-grained descriptions can be explored and adopted in the future.

3) We mainly focus on CNNs for image classification. CNNs and ViTs for other tasks should be investigated in the future.

VI. CONCLUSION

In this article, we proposed a novel computational framework, Sync-ACT, to couple the visual representation spaces and semantics between ANNs and BNNs in the human brain by synchronizing their activations to visual stimuli. We found a significant correlation in the semantics between the visual representations in CNNs and those in the human brain. Also, CNN’s visual representation similarity to the human brain is closely related to its performance on the image classification tasks. In the future, our Sync-ACT model can be easily generalized to other naturalistic stimuli such as natural language and/or audio to explore the connection of the model’s semantic space with the one in the human brain. Overall, our study introduces a general and effective paradigm to couple the ANNs and BNNs and provides novel insight into their connections.

REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, 2015.
[2] Y. LeCun et al., “Convolutional networks for images, speech, and time series,” Handbook Brain Theory Neural Netw., vol. 3361, no. 10, pp. 1995, 1995.
[3] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, “A survey of the recent architectures of deep convolutional neural networks,” Artif. Intell. Rev., vol. 53, no. 8, pp. 5455–5516, 2020.
[4] A. Dosovitskiy et al., “An image is worth 16×16 words: Transformers for image recognition at scale,” 2020, arXiv:2010.11929.
[5] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, “Learning deep features for discriminative localization,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 2921–2929.
[6] D. Bau, B. Zhou, A. Khosla, A. Oliva, and A. Torralba, “Network dissection: Quantifying interpretability of deep visual representations,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 6541–6549.
[7] E. Hernandez, S. Schwettmann, D. Bau, T. Bagashvili, A. Torralba, and J. Andrews, “Natural language descriptions of deep features,” in Proc. Int. Conf. Learn. Rep., 2021, pp. 1–6.
[8] X. Hu et al., “Bridging low-level features and high-level semantics via fMRI brain imaging for video classification,” in Proc. 18th ACM Int. Conf. Multimedia, 2010, pp. 451–460.
[9] T. Liu, X. Hu, X. Li, M. Chen, J. Han, and L. Guo, “Merging neuroimaging and multimedia: Methods, opportunities, and challenges,” IEEE Trans. Human-Mach. Syst., vol. 44, no. 2, pp. 270–280, Apr. 2014.
[10] J. Wang et al., “Test–retest reliability of functional connectivity networks during naturalistic fMRI paradigms,” Human Brain Map., vol. 38, no. 4, pp. 2226–2241, 2017.
[11] Y. Ren, S. Xu, Z. Tao, L. Song, and X. He, “Hierarchical spatio-temporal modeling of naturalistic functional magnetic resonance imaging signals via two-stage deep belief network with neural architecture search,” Front. Neurosci., vol. 15, Dec. 2021, Art. no. 794955.
[12] V. D. Calhoun and T. Adali, “Unmixing fMRI with independent component analysis,” IEEE Eng. Med. Biol. Mag., vol. 25, no. 2, pp. 79–90, Mar./Apr. 2006.
[13] J. Lv et al., “Holistic atlases of functional networks and interactions reveal reciprocal organizational architecture of cortical function,” IEEE Trans. Biomed. Eng., vol. 62, no. 4, pp. 1120–1131, Apr. 2015.
[14] Y. Golland et al., “Extrinsic and intrinsic systems in the posterior cortex of the human brain revealed during natural sensory stimulation,” Cereb. Cortex, vol. 17, no. 4, pp. 766–777, 2007.
[15] Y. Ren, V. T. Nguyen, L. Guo, and C. C. Guo, “Inter-subject functional correlation reveal a hierarchical organization of extrinsic and intrinsic systems in the brain,” Sci. Rep., vol. 7, no. 1, pp. 1–12, 2017.
[16] H. Wang et al., “Recognizing brain states using deep sparse recurrent neural network,” IEEE Trans. Med. Imag., vol. 38, no. 4, pp. 1058–1068, Apr. 2019.
[17] W. Zhang et al., “Identify hierarchical structures from task-based fMRI data via hybrid spatiotemporal neural architecture search net,” in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent., 2019, pp. 745–753.
[18] H. Liu et al., “The cerebral cortex is bisectionally segregated into two fundamentally different functional units of GYRI and SULCI,” Cereb. Cortex, vol. 29, no. 10, pp. 4238–4252, 2019.
[19] Q. Dong et al., “Modeling hierarchical brain networks via volumetric sparse deep belief network,” IEEE Trans. Biomed. Eng., vol. 67, no. 6, pp. 1739–1748, Jun. 2020.
[20] L. Zhao, H. Dai, X. Jiang, T. Zhang, D. Zhu, and T. Liu, “Exploring the functional difference of GYRI/SULCI via hierarchical interpretable autoencoder,” in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent., 2021, pp. 701–709.
[21] Q. Li, W. Zhang, L. Zhao, X. Wu, and T. Liu, “Evolutional neural architecture search for optimization of spatiotemporal brain network decomposition,” IEEE Trans. Biomed. Eng., vol. 69, no. 2, pp. 624–634, Feb. 2022.
[22] D. C. Van Essen et al., “The WU-Minn human connectome project: An overview,” Neuroimage, vol. 80, pp. 62–79, Oct. 2013.
[23] M. Hanke et al., “A studyforest extension, simultaneous fMRI and eye gaze recordings during prolonged natural stimulation,” Sci. Data, vol. 3, no. 1, pp. 1–15, 2016.

[24] F. Dalvi, N. Durran, H. Sajjad, Y. Belinkov, A. Bau, and J. Glass, “What is one grain of sand in the desert? Analyzing individual neurons in deep NLP models,” in Proc. AAAI Conf. Artif. Intell., vol. 33, 2019, pp. 6309–6317.

[25] A. S. Morcos, D. G. Barrett, N. C. Rabinowitz, and M. Botvinick, “On the importance of single directions for generalization,” 2008, arXiv:1803.06959.

[26] J. Mu and J. Andreas, “Compositional explanations of neurons,” in Proc. Adv. Neural Inf. Process. Syst., vol. 33, 2020, pp. 17153–17163.

[27] B. Mwangi, T. S. Tian, and J. C. Soares, “A review of feature reduction techniques in neuroimaging,” Neuroinformatics, vol. 12, no. 2, pp. 229–244, 2014.

[28] Q. Li, Q. Dong, F. Ge, N. Qiang, X. Wu, and T. Liu, “Simultaneous spatial–temporal decomposition for connectome-scale brain networks by deep sparse recurrent auto-encoder,” Brain Imag. Behav., vol. 15, no. 5, pp. 2646–2660, 2021.

[29] D. H. Hubel and T. N. Wiesel, “Receptive fields of single neurones in the cat’s striate cortex,” J. Physiol., vol. 148, no. 3, p. 574, 1959.

[30] D. H. Hubel and T. N. Wiesel, “Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex,” J. Physiol., vol. 160, no. 1, p. 106, 1962.

[31] W. Luo, Y. Li, R. Urtasun, and R. Zemel, “Understanding the effective receptive field in deep convolutional neural networks,” in Proc. Adv. Neural Inf. Process. Syst., vol. 29, 2016, pp. 4905–4913.

[32] D. G. Barrett, A. S. Morcos, and J. H. Macke, “Analyzing biological and artificial neural networks: challenges with opportunities for synergy?” Current Opin. Neurobiol., vol. 55, pp. 55–64, Apr. 2019.

[33] A. Mahendran and A. Vedaldi, “Visualizing deep convolutional neural networks using natural pre-images,” Int. J. Comput. Vis., vol. 120, no. 3, pp. 233–255, 2016.

[34] D. L. Yamins and J. J. DiCarlo, “Using goal-driven deep learning models to understand sensory cortex,” Nat. Neurosci., vol. 19, no. 3, pp. 356–365, 2016.

[35] D. L. Yamins, H. Hong, C. F. Cadieu, E. A. Solomon, D. Seibert, and J. J. DiCarlo, “Performance-optimized hierarchical models predict neural responses in higher visual cortex,” Proc. Nat. Acad. Sci. USA, vol. 111, no. 23, pp. 8619–8624, 2014.

[36] C. Du, C. Du, L. Huang, H. Wang, and H. He, “Structured neural decoding with multitask learning by deep neural network representations,” IEEE Trans. Neural Netw. Learn. Syst., vol. 33, no. 2, pp. 600–614, Feb. 2022.

[37] Q. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, “Places: A large-scale hierarchical image database,” in Proc. Adv. Neural Inf. Process. Syst., vol. 10, no. 4, 2014, Art. no. e1003553.

[38] F. Dalvi, N. Durrani, H. Sajjad, Y. Belinkov, A. Bau, and J. Glass, “Long short-term memory,” Sci. Data, vol. 9, no. 8, pp. 1735–1780, 1997.

[39] A. Vaswani et al., “Attention is all you need,” in Proc. Adv. Neural Inf. Process. Syst., vol. 30, 2017, pp. 6000–6010.

[40] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2009, pp. 248–255.

[41] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, “Places: A 10 million image database for scene recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 6, pp. 1452–1464, Jun. 2018.

[42] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in Proc. Adv. Neural Inf. Process. Syst., vol. 25, 2012, pp. 1106–1114.

[43] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2014, arXiv:1409.1556.

[44] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 770–777.

[45] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 4700–4708.

[46] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 2818–2826.

[47] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, “ShufflenetV2: Practical guidelines for efficient CNN architecture design,” in Proc. Eur. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 116–131.

[48] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “MobileNetV2: Inverted residuals and linear bottlenecks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 4510–4520.

[49] S. Xie, R. B. Girshick, P. Dollár, Z. Tu, and K. He, “Aggregated residual transformations for deep neural networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 1492–1500.

[50] M. Tan et al., “MnasNet: Platform-aware neural architecture search for mobile,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2019, pp. 2820–2828.

[51] J. Kubliius et al., “Brain-like object recognition with high-performing shallow recurrent nets,” in Proc. Adv. Neural Inf. Process. Syst., vol. 32, 2019.

[52] D. M. Barch et al., “Function in the human connectome: Task-fMRI and individual differences in behavior,” Neuroimage, vol. 80, pp. 169–189, Oct. 2013.

[53] M. F. Glasser et al., “The minimal preprocessing pipelines for the human connectome project,” Neuroimage, vol. 80, pp. 105–124, Oct. 2013.

[54] O. Esteban et al., “fMRIPrep: A robust preprocessing pipeline for functional MRI,” Nat. Methods, vol. 16, no. 1, pp. 111–116, 2019.

[55] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2014, arXiv:1412.6980.

[56] G. K. Aguirre, E. Zarahn, and M. D’Esposito, “The variability of human, bold hemodynamic responses,” Neuroimage, vol. 8, no. 4, pp. 360–369, 1998.

[57] B. Bonakdarpour, T. B. Parrish, and C. K. Thompson, “Hemodynamic response function in patients with stroke-induced aphasia: Implications for fMRI data analysis,” Neuroimage, vol. 36, no. 2, pp. 322–331, 2007.

[58] T. Elskens, J. H. Metzen, and F. Hutter, “Neural architecture search: A survey,” J. Mach. Learn. Res., vol. 20, no. 1, pp. 1997–2017, 2019.

[59] H. Liu, K. Simonyan, and Y. Yang, “DARTS: Differentiable architecture search,” 2018, arXiv:1806.09055.

[60] B. Zoph and Q. V. Le, “Neural architecture search with reinforcement learning,” 2016, arXiv:1611.01578.

[61] T. Yarkoni, R. A. Poldrack, T. E. Nichols, D. C. Van Essen, and T. D. Wager, “Large-scale automated synthesis of human functional neuroimaging data,” Nat. Methods, vol. 8, no. 8, pp. 665–670, 2011.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
Haixing Dai received the B.E. degree from the School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu, China, in 2016. He is currently pursuing the Ph.D. degree with the School of Computing, University of Georgia, Athens, GA, USA, under the supervision of Dr. T. Liu.

His current research interests include deep learning and medical image analysis.

Zihao Wu received the B.E. degree from the School of Microelectronics, Tianjin University, Tianjin, China, in 2017, and the master's degree from the Electrical Engineering and Computer Science Department, Vanderbilt University, Nashville, TN, USA, in 2020. He is currently pursuing the Ph.D. degree in computer science with the University of Georgia, Athens, GA, USA, under the supervision of Dr. T. Liu.

His current research interests include brain-inspired AI and deep-learning-based medical image analysis.

Zhenxiang Xiao (Graduate Student Member, IEEE) received the Biomedical Engineering degree from the School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, China, in 2021, where he is currently pursuing the master's degree under the supervision of Dr. X. Jiang.

His research interests include brain functional network analysis and graph convolutional neural networks.

Lu Zhang received the B.S. degree and the M.S. degree in computer science from Northwestern Polytechnical University, Xi'an, China, in 2015 and 2018, respectively. She is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, The University of Texas at Arlington, Arlington, TX, USA, where she is supervised by Dr. D. Zhu.

Her current research interests include computational neuroscience, brain imaging, and deep learning.

David Weizhong Liu is a high school student with Athens Academy, Athens, GA, USA.

He is a Research Intern with the Bio-Sensing and Instrumentation Lab, College of Engineering, University of Georgia. His research interests are robotics and artificial intelligence.

Xintao Hu received the Ph.D. degree from Northwestern Polytechnical University (NWPU), Xi'an, China, in 2011. He is currently a Professor with the School of Automation, NWPU. His research interests include brain image computing and brain-inspired artificial intelligence.

Xi Jiang received the B.E. degree in automation from Northwestern Polytechnical University, Xi'an, China, in 2009, and the Ph.D. degree in computer science from the University of Georgia, Athens, GA, USA, in 2016.

Since 2017, he has been an Associate Professor with the School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, China. His research interests include machine learning/deep-learning-based medical image analysis.

Dr. Jiang was a recipient of the Li Foundation Heritage Prize, USA, for "outstanding research and contributions in the interdisciplinary field of brain science" in 2019.

Sheng Li received the bachelor's and master's degrees from the School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing, China, in 2010 and 2012, respectively, and the Ph.D. degree in computer engineering from Northeastern University, Shenyang, China, in 2017, under the supervision of Prof. Y. R. Fu.

He is an Assistant Professor of Data Science with the University of Virginia (UVA), Charlottesville, VA, USA. Prior to joining UVA, he was an Assistant Professor of Computer Science with the University of Georgia, Athens, GA, USA, from 2018 to 2022, and a Data Scientist with Adobe Research, San Francisco, CA, USA, from 2017 to 2018. His recent research interests include trustworthy representation learning, graph neural networks, visual intelligence, and causal inference.

Dajiang Zhu received the Ph.D. degree in computer science from the University of Georgia, Athens, GA, USA, in 2014.

His current research interests include machine learning, neuroimaging, and computational neuroscience.

Tianming Liu (Senior Member, IEEE) received the Ph.D. degree in computer science from Shanghai Jiao Tong University, Shanghai, China, in 2002. He is a Distinguished Research Professor of Computer Science with the University of Georgia, Athens, GA, USA. His research interests are brain imaging, computational neuroscience, and brain-inspired artificial intelligence. He has published over 400 papers in these areas, his Google Scholar citations are over 10,000, and his H-index is 54.

Dr. Liu is the recipient of the NIH Career Award and the NSF CAREER Award. He serves on the editorial boards of multiple journals, including Medical Image Analysis, IEEE TRANSACTIONS ON MEDICAL IMAGING, IEEE REVIEWS IN BIOMEDICAL ENGINEERING, IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS, and IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS. He is an Elected Fellow of the American Institute for Medical and Biological Engineering.