A Deep Learning Approach to Self-prioritisation Processes

Zhuoen Lu
School of Psychology
University of Aberdeen
Aberdeen, United Kingdom
z.lu1.22@abdn.ac.uk

Dewei Yi
School of Natural and Computing Sciences
University of Aberdeen
Aberdeen, United Kingdom
dewei.yi@abdn.ac.uk

Xun He
Department of Psychology
Bournemouth University
Poole, United Kingdom
xhe@bournemouth.ac.uk

Jie Sui
School of Psychology
University of Aberdeen
Aberdeen, United Kingdom
jie.sui@abdn.ac.uk

***Abstract*—Self-prioritisation is a ubiquitous psychological phenomenon which occurs when saliency-related processing is linked to self-related information compared to other-related information. Recent research has identified certain brain regions activated during self-saliency processing. However, the dynamic characteristics of the region connections causing such prioritization effects remain unknown. To address this, here we investigated neural couplings at the whole brain level during the self-saliency processing when participants carried out a standardised shape-label matching task widely used to test self-prioritization effects, while electroencephalogram (EEG) data were recorded. Behavioural results indicated the presence of the self-prioritization effect, evident by faster and more accurate responses to shape-label pairs associated with oneself compared to those related to a friend or a stranger. We then applied deep learning models to test and validate the robustness of the two phases of neural couplings. Specifically, in the early top-down regulation phase, the analysis based on SqueezeNet showed higher accuracy performance for the self-related stimuli compared to stimuli associated with others, but this is not the case for the later feedforward. Moreover, the deep learning models performed better in classifying the self-related stimuli in the early top-down phase than in the later feedforward phase, which was not detected in classifying stranger-related stimuli. The data indicated that deep learning networks might offer insights into psychological functioning that would otherwise be unresolved using traditional neuroscience methods; as a result, the latter might aid the development of brain-inspired algorithms for deep learning.**

***Index Terms*—self-prioritisation, EEG, neural coupling, deep learning**

I. Introduction

People will show bias when processing information related to themselves, leading to faster responses and higher accuracies for such information compared to other information, so-called self-prioritisation [1] [2]. One account for such effect is the integrative self model addressing that self-salience will enhance through the activation of self-representation and subsequently facilitate the integration of the levels of information processing [3]. The model has been supported by a wide range of empirical evidence [4] [5] [6] [7].

Recent studies have evaluated the neural characteristics underlying self-saliency processing using neuroimaging techniques [4] [8] [9]. For example, functional magnetic resonance imaging (fMRI) and EEG have been used to evaluate brain activities during self-reference, revealing the fundamental neural features related to self-prioritisation, and quickly transforming self-related information to the improvement of task performance. Specifically, increased neural activities have been observed at frontal regions during a number of self-referencing processing, especially at the medial prefrontal cortex (mPFC) [10] [11] [12] [13] [14] [15] [16] [17] [18] [19] [20] [21] [22] [23]. There was evidence showing that damaged mPFC will abolish the self-reference effect [24]. Moreover, some recent studies have indicated that the activation of self-referential processing was linked to a specific brain circuit including the ventral medial frontal cortex and the posterior superior temporal sulcus [16] [25] [26].

To evaluate the interaction between different brain regions linked to self-referential processing, functional connectivity analysis, which refers to statistical dependence between time series of electro-physiological activities [27], has been extensively applied to neuroscience research, revealing the relationship between self-reference and other psychological effects, such as emotions and neurological disorders [28] [29] [30]. However, conventional functional connectivity only showed the relationship between brain regions in terms of frequency ranges and strength, which causes the limitation in evaluating functional connectivity for highly dynamic neural activities. New methodologies have been proposed and applied in the processing of EEG signals (e.g., time-frequency analysis), as well as measurements aiming to demonstrate dynamic neural activities between brain regions (e.g., the imaginary part of coherency (iCoh) [31]. Time-frequency analysis simultaneously extracts components with the interaction between two signals in temporal, in addition to the detailed connectivity in different frequency components. Accordingly, dynamic connectivity analysis has recently been applied to evaluate the

dynamic neural mechanisms linked to mental issues and neural disorders [32] [33] [34] [35].

The present study aimed to apply the dynamic connectivity analysis to the common self-prioritisation effects. As dynamic progress, it is important to evaluate how functional connectivity changed over time during self-saliency processing, as well as the extraction of dynamic neural features linked to self-saliency processing, which is the first aim of this study. In the first part of this study, the novel multiple wavelet transform algorithm, the so-called "Superlet" [36], was used for time-frequency analysis to previously recorded BioSemi EEG data during self-saliency processing. Since previous studies indicated the strongest background neural activities in the Alpha band [37] [38], as well as the fundamental role of the Alpha band in visual processing [39], the Alpha band (8 – 12 Hz) was chosen as the featured frequency band for further time-frequency analysis and neural coupling evaluation.

Even though the relationship between specialised neural features and self-referencing processing was indicated in previous studies, it is still challenging to validate the robustness of these neural couplings in real data. Therefore, it is necessary to build up a universal model which can validate the robustness of neural couplings potentially linked to self-referencing processing in other data. In the second part of the study, further approaches were applied as the quantification of the relationship between specific neural coupling features and the self-prioritisation effect.

Machine learning (ML) can be applied to the further analysis of EEG data for clinical diagnosis and classification, as well as exploring the neural mechanisms of neural diseases. As the novel algorithms of ML, deep learning models have been extensively applied in bioinformatics and neuroimage processing for the classification of neural diseases and psychological behaviours [40] [41] [42] [43] [44]. High research potential and better performance have been observed in the application of classifying models compared to conventional ML methods [45]. Thus, deep learning methods provide a feasible and straightforward way to quantify the relationship between specialised features of neural couplings and psychological effects.

Neural coupling features between frontal and occipital ROIs in the Alpha band derived from time-frequency analysis were used as training data for constructing automatic classification models based on the deep learning method. In this study, assessment models were built based on convolution neural networks (CNN) using scalogram images of averaging neural couplings between frontal and occipital regions. The performance of fifteen commonly-used CNN models was firstly compared, then chose SqueezeNet [46], which had the fastest training speed, to build up automatic classification models for classifying experimental trials with three matching conditions (Self, Friend and Stranger) and two response conditions (Matched and Mismatched) between two frontal ROIs (Left Frontal and Right Frontal) and Occipital ROI using two types of periods with robust neural coupling features (early top-down and later feedforward).

II. Materials and Methods

A. Participants

Ten pairs of gender-matched close friends (aged 19.9 ± 1.7 years; 8 females, 12 males) were recruited to participate in the experiment. All the participants had a normal or corrected-to-normal vision. The experiment was approved by the university ethics committee.

B. Assessment of Neural Couplings during Self-saliency Processing

1) Measuring Self-saliency Processing: The degree of individual self-saliency processing was evaluated by a psychological procedure [4]. In the experiment, participants were shown three labels (Self, Friend and Stranger) referring to three levels of self-related information that were introduced in the experiment, which were associated with three geometric shapes (circle, square and triangle). Fig. 1(A) illustrates the procedure of the experiment. The stimulus, including a geometric shape and a label of self-related information, was presented in each trial, followed by an interval with a fixation. Correct or incorrect associations were shown with equal probabilities. After the experiment began, participants were required to judge whether the associations were correct by pressing different keys (key mapping counterbalanced). Feedback was provided after the interval period to inform the participants with their responses. The next trial started after another 500-ms interval.

The behavioural results showed more accurate responses to self-related shape-label pairs compared to those related to a friend or a stranger, as well as faster response speeds ($ps < .003$), indicating robust self-prioritisation effects across participants (see Fig. 1 (B)).

2) Data Preprocessing: A set of 128-channel EEG data were recorded following the 10-5 electrode nomenclature [48], with a sampling rate of 512 Hz. EEG was also recorded from the left and right mastoids for the purpose of offline re-referencing. The EEG data were re-referenced to the average voltage at mastoids and segmented the EEG data using a time window of 3,670 ms which covers from 785 ms before to 2,885 ms after the stimulus onsets. Partial results from -200 ms to 1,100 ms relative to stimulus onsets were selected for further investigation.

C. Construction of Assessment Models

1) Feature Extraction - Multiple Wavelet Analysis (Superlet): In the first part of this study, the self-prioritisation effects in visual associations were investigated from the perspective of dynamic neural couplings during self-saliency processing. Wavelet analysis methods allowed us to evaluate how neural couplings between brain regions changed at various periods of self-saliency processing. Time-frequency analysis [49] was applied to EEG data using the method based on continuous wavelet transform (CWT) [50]. Fig. 1(D) demonstrates the definition of the Morlet wavelet and the Gaussian window.

Theoretically, a lower number of cycles would give us more detailed information in temporal, while uncertainty will be

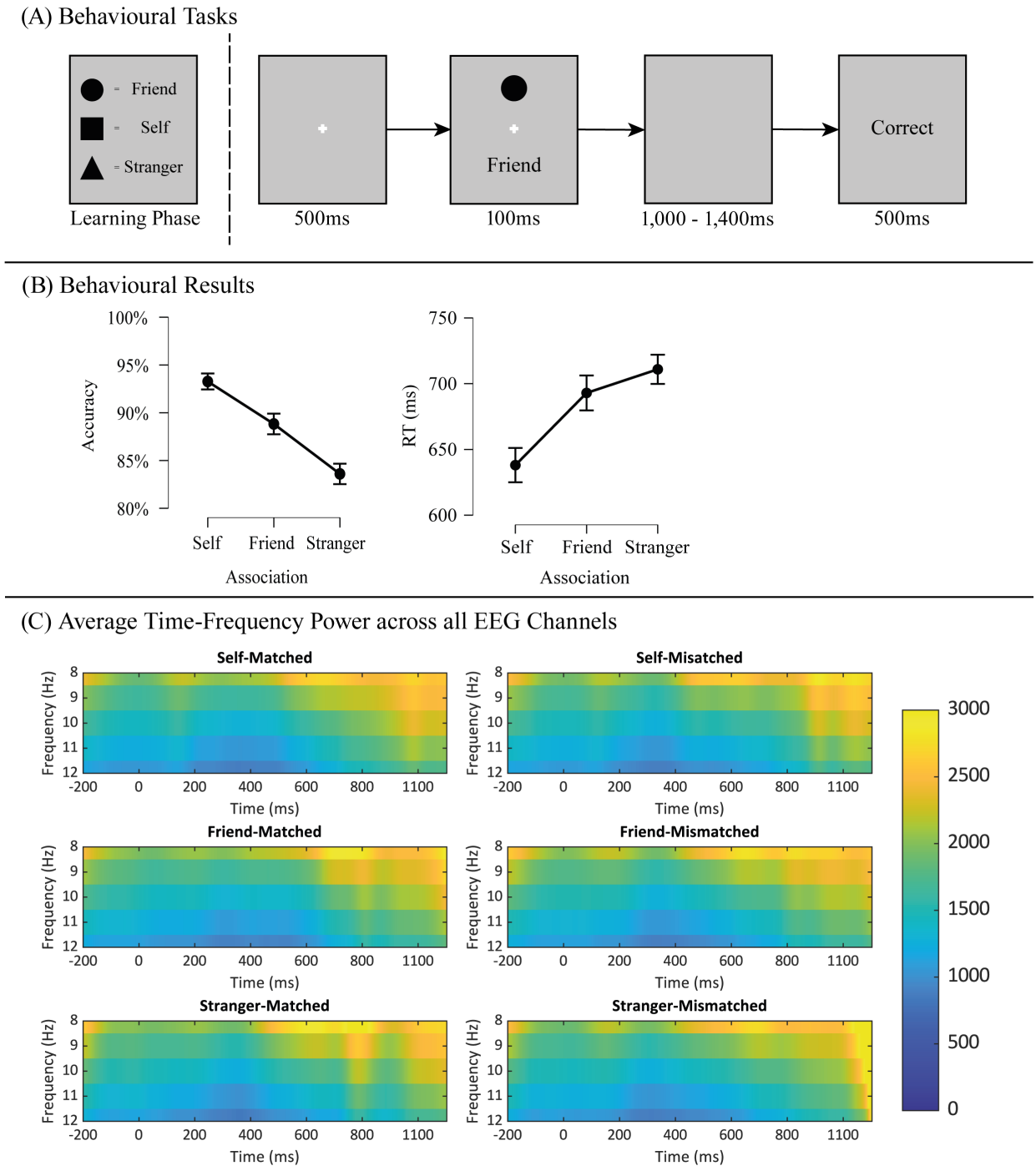

Fig. 1. (A) Trial sequence in the experiment. (B) Behavioural results in matching tasks. Error bars extend to ±SE. Significant main effects were observed in repeat measured ANOVA model across associations in accuracy ($F(2,38) = 23.37$, $p < .001$) and reaction time (RT) ($F(2,38) = 40.32$, $p < .001$) from matched trials. (C) Average time-frequency spectrum in the alpha band (8 – 12Hz) across all EEG electrodes from all matching trials and individuals. (D) The definition of Morlet Wavelet and Gaussian Window. A Morlet wavelet can be defined using a complex sine wave and a Gaussian window [47]. In the Morlet wavelet function, t is the time in seconds and the f refers to the frequency in Hz. In the function of defining a Gaussian window σ, n refers to the coefficient of the number of cycles.

caused in the frequency domain. In this case, a cross-hatched U-shaped unavailable area will cover the bottom of the wavelet spectrograms. To maximise the resolution in both temporal and frequency domains, a novel time-frequency analysis method named "Superlet" [36] was applied in the current study to generate wavelet spectrograms with fewer unavailable areas because it effectively reduces the impact due to the resolution differences in time-frequency domains. This method applies wavelet transform with various numbers of wavelet cycles. Results from multiple wavelets are then combined to generate time-frequency results with the maximum resolution.

In this study, the width of the base wavelets was set at 3, as well as the length of the used wavelets in standard deviations of the implicit Gaussian kernel was set at 3. The range of frequency from 8 to 12 Hz with 1-Hz steps was applied in time-frequency analysis to evaluate time-frequency components and neural couplings in the Alpha band.

Fig. 1(C) illustrates the results of all six conditions, averaging over all EEG electrodes, all experimental trials and all participants. All matching conditions showed increased brain activities after the onset of experimental trials in the Alpha band.

2) Featured Neural Coupling Periods for Self-saliency Processing: Further dynamic connectivity analysis was applied in alpha-band activities (see Fig. 1(C)) aiming to extract featured brain regions with strong neural couplings potentially linked to self-saliency processing. The imaginary part of coherency (iCoh) [31], which reduces the interference of volume conduction from the EEG data, was calculated as the measurement of neural couplings between every two EEG electrodes. Further analysis was based on region-of-interests (ROIs) due to the dense-array channel setting. The ROIs were selected based on the averaging neural couplings over the range of 0 – 1,100 ms after the stimulus onsets across all conditions and all participants. Among a total of 8,128 electrode pairs, 1% of the electrode pairs showing the strongest neural couplings (averaged over all experimental conditions) were used to determine the ROIs.

The electrode pairs having the strongest neural couplings showed consistent clustering mainly between the lateral frontal and occipital regions. Accordingly, three ROIs were identified (see Fig. 2(A)) in which two of them were located symmetrically at lateral frontal regions. After the ROIs were defined, neural couplings between electrodes located in one of the frontal ROIs and the occipital ROI were averaged over for each condition (as there was no neural coupling between the two frontal ROIs). Positive iCoh values represent top-down information flow from the frontal to occipital ROIs, while negative values represent the information feedforward from the occipital to frontal ROIs. Repeated-measures analyses of variance (rm-ANOVA) models were used to evaluate the significant differences of neural couplings between matching conditions, as well as the post-hoc tests between every two matching conditions.

Fig. 2(B) illustrates the neural coupling features between the frontal and occipital ROIs before and during the onset of experiment trials. Before the stimulus onset, there was a weak information flow from the frontal ROIs to the occipital ROI. After the stimulus onset, neural couplings from the frontal to occipital ROIs were observed among all conditions with a high peak of positive neural couplings in all the conditions at approximately 150ms after the onset of the trial, indicating robust top-down communication from the frontal to occipital ROIs. The information flow then reversed with neural couplings from the occipital to frontal ROIs, following a low level of top-down communication again at the end of the experimental trials.

The neural coupling results in the matched conditions and ROI pairs demonstrated robust features of neural couplings after stimulus onsets. Two neural coupling features were identified for training automatic classification models based on a convolutional neural network (CNN) to evaluate the robustness between the features and self-saliency processing (i.e., the early top-down period and the later feedforward period). To identify a universal time window for feature segmentation, the timepoints of reversion of neural couplings from top-down to feedforward were identified, as well as the time points of reversion of neural couplings from feedforward to the resting level (low-level top-down neural couplings) in all experimental trials and then calculated the average turning time points across individuals. The early top-down period was determined to be in the period of 0 to 251ms, and the later feedforward period was determined to be from 251 to 750ms.

3) Assessment of CNN Models: Features of dynamic neural couplings generated in the first part of this study were used as the training data in constructing classification models based on CNN. Commonly used CNN models were assessed in the study by comparing validation accuracies to choose the most robust models for assessing self-saliency processing.

Two groups of deep learning models were built up using different periods of the neural coupling data to evaluate which period was more robustly related to self-saliency processing. For each model group, the neural coupling data from the left and right brain hemispheres was separated to train two types of independent models. Therefore, there were four types of models using data from different brain hemispheres (left frontal and occipital, and right frontal and occipital) and features (early top-down and later feedforward) in this study. In this study, CNN models were constructed and trained using MATLAB (version R2022b) with 15 commonly used CNN models used in performance assessment.

Table 1 shows the comparison of the structure of the CNN models, as well as the performance of using them for training. Depth is defined by the number of layers in a neural network. Size represents the storage space for a neural network. Number of parameters is the number of coefficients of the model. Image input size indicates the standard size of input images required for the neural network. A random selection of 90% of the trials across all conditions and participants was chosen as training data with the rest 10% of the trials being used as validation data. An initial learning rate of 0.0001 was set with the max epochs of 5 to train the models. Each model

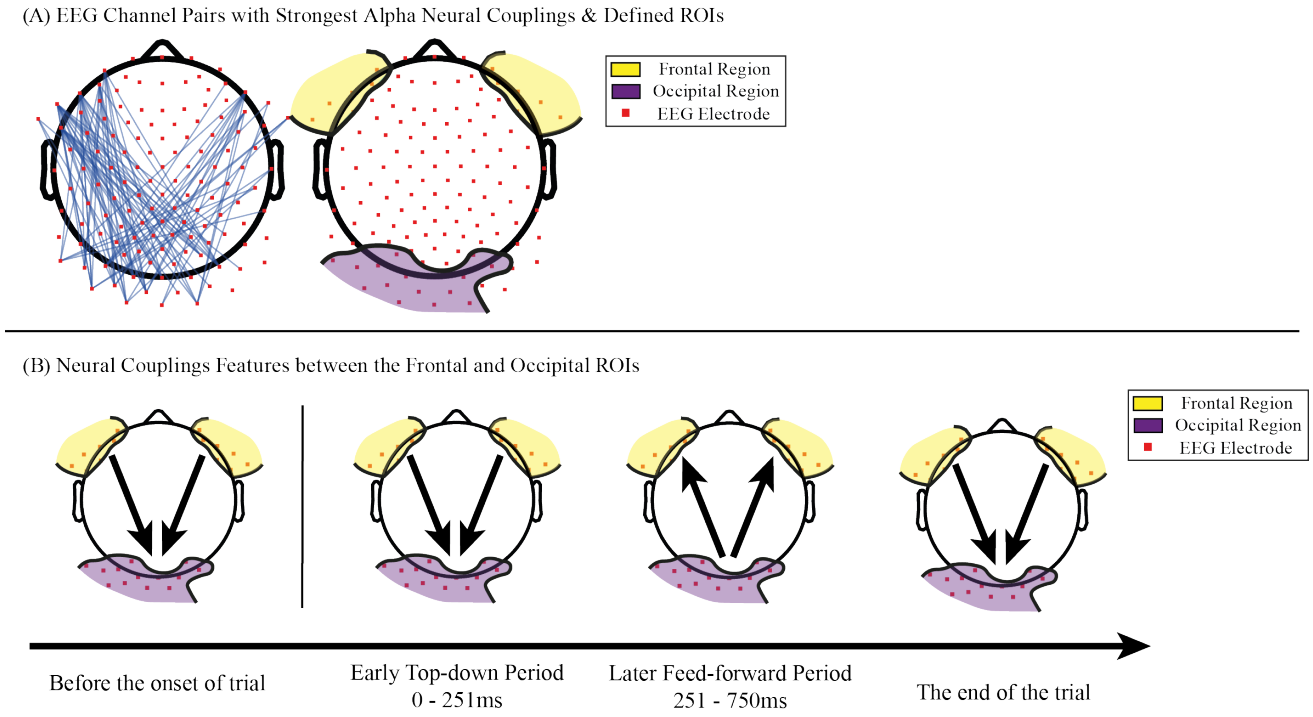

Fig. 2. (A) Head maps of the distribution of EEG electrode pairs with the top 1% average absolute iCoh in the alpha band (8 - 12Hz), and the identified ROIs. (B) Average neural coupling features between Frontal and Occipital ROIs across experimental trials.

was replicated 5 times to calculate the average training time and overall validation accuracies. Some of the models failed to produce results due to running out of memory due to the complex structure.

4) SqueezeNet-based Classification Models: Based on the results of the model assessment, the SqueezeNet network was chosen to build up the CNN models as it got the shortest training time than other networks and relatively high validation accuracy. Proposed in 2016, SqueezeNet is a model that can achieve high accuracy, but with a smaller structure and fewer parameters, which can improve training speed [46] and reduce computational expense to increase the feasibility of on-chip deployment [51].

In SqueezeNet, the architectural simplification strategies are deployed by creating Fire modules (see Fig. 3). A Fire module contains a squeeze convolution layer, feeding into an expanded layer mixing convolution filters. The liberal use of $1x1$ filters can effectively decreases the number of parameters relative to the use of conventional $3x3$ filters. Also, the module structure effectively decreases the number of parameters in a CNN model while accuracy is preserved. In the current study, the last dropout layer '$drop9$' in the network was replaced with a new dropout layer with a probability of 0.6. Because the last learnable layer in SqueezeNet was a 1×1 convolutional layer '$conv10$', the '$conv10$' layer was replaced with a new convolutional layer with the number of filters equal to the number of classes. The learning rate factor of the new layer was increased to 5 and replaced the classification layer with a new classification layer with automatic classification label detection.

Before importing the tempospectrum images into the training models, they were resized to the standard image size required by SqueezeNet (227 x 227 pixels). Also, all the neural coupling images were transformed to the RGB format using the parula colourmap to meet the requirement of SqueezeNet. Moreover, some configurations were applied to the standard network structure.

To train assessment models based on SqueezeNet and evaluate the robustness of neural coupling features in classifying the conditions, 260 iterations were employed with the maximum number of epochs being set at 2 in order to compare validation frequencies of classification across the six conditions, including three correctly matched conditions (Self-matched, Friend-matched and Stranger-matched) and three mismatched conditions (Self-mismatched, Friend-mismatched and Stranger-mismatched). After that, 90% of images from the whole image dataset were randomly chosen as the training data, while images spat from 10% of the dataset were used as the validation data.

The training for each type of model was replicated 30 times to validate the robustness of the models. Validation accuracies were then statistically assessed with rm-ANOVAs, and post-hoc tests between every two matching conditions.

TABLE I

COMPARISON OF COMMONLY USED CNN MODELS IN THE CURRENT RESEARCH CONTEXT. CLASSIFICATION MODELS WERE TRAINED USING IMAGES DEMONSTRATING NEURAL COUPLINGS DURING THE EARLY TOP-DOWN PERIOD FROM EXPERIMENTAL TRIALS IN THE SIX CONDITIONS.

Deep Learning Network	Depth	Size	Parameters (Millions)	Image Input Size (Pixel-by-Pixel)	Training Time (mm:ss)	Overall Validation Accuracy
squeezenet	18	5.2 MB	1.24	227-by-227	02:55	99.83%
googlenet	22	27 MB	7	224-by-224	06:59	99.95%
inceptionv3	48	89 MB	23.9	299-by-299	Failed (Out of GPU Memory)	
densenet201	201	77 MB	20	224-by-224	Failed (Out of GPU Memory)	
mobilenetv2	53	13 MB	3.5	224-by-224	14:52	99.99%
resnet18	18	44 MB	11.7	224-by-224	04:50	99.91%
xception	71	85 MB	22.9	299-by-299	Failed (Out of GPU Memory)	
inceptionresnetv2	164	209 MB	55.9	299-by-299	Failed (Out of GPU Memory)	
shufflenet	50	5.4 MB	1.4	224-by-224	05:35	99.98%
nasnetmobile	*	20 MB	5.3	224-by-224	Failed (Out of GPU Memory)	
nasnetlarge	*	332 MB	88.9	331-by-331	Failed (Out of GPU Memory)	
darknet19	19	78 MB	20.8	256-by-256	36:50	99.99%
efficientnetb0	82	20 MB	5.3	224-by-224	Failed (Out of GPU Memory)	
alexnet	8	227 MB	61	227-by-227	03:50	99.89%
vgg16	16	515 MB	138	224-by-224	Failed (Out of GPU Memory)	

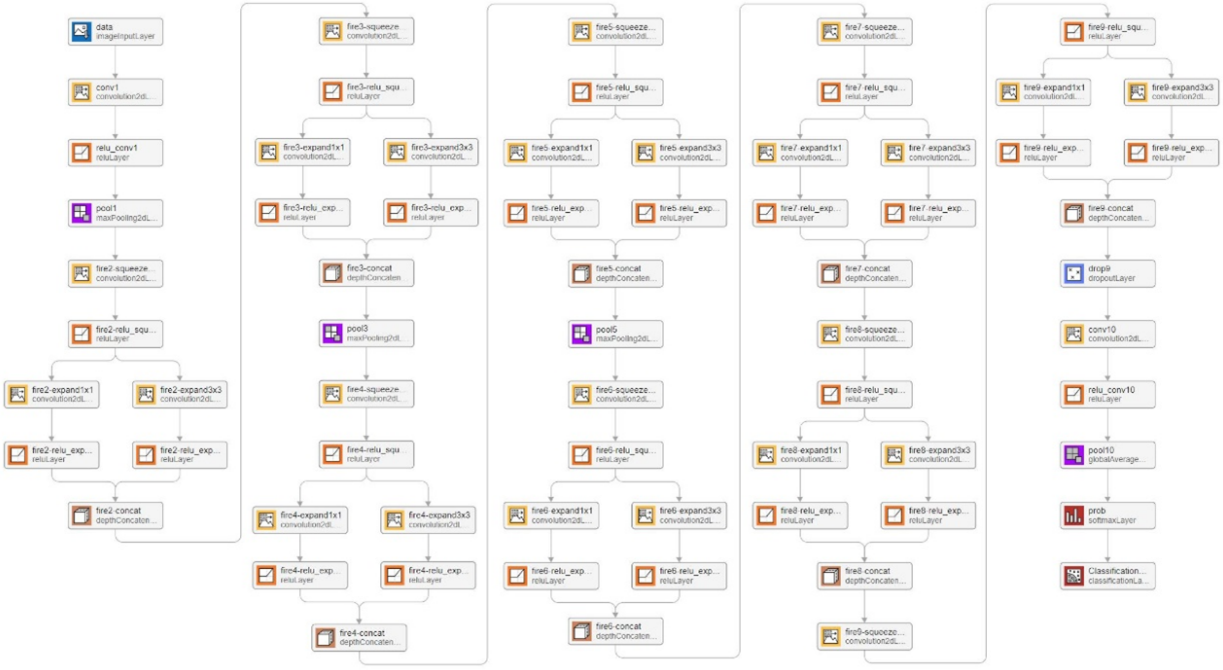

Fig. 3. The architecture of SqueezeNet CNN network. Three hyperparameters $s1x1$, $e1x1$, and $e3x3$ are exposed in a Fire module. The $s1x1$ represents the number of filters in the squeeze layer, the e1x1 represents the number of 1x1 filters in the expanding layer, while the $e3x3$ represents the number of 3x3 filters in the expanding layer. Specifically, the $s1x1$ is less than the sum of $e1x1$ and $e3x3$ while using Fire modules in order to limit the number of input channels to the $3x3$ filters. In this study, the structure of SqueezeNet began with a standalone convolution layer ($conv1$), followed by 8 Fire modules ($fire2$ - $fire9$), ending with a convolutional layer ($conv10$). The number of filters in each Fire module is gradually increased from the beginning to the end of the structure. Also, SqueezeNet performed max-pooling with a stride of 2 after the layers $conv1$, $fire4$, $fire8$ and $conv10$. Relatively late placements of pooling in SqueezeNet can make sure that convolution layers had large activation maps by late down-sampling in the model.

III. Results

A. Higher Validation Accuracies for Self-matched Condition using top-down Data

Four rm-ANOVAs were employed in the four models using different data with post-hoc comparisons across matching conditions.

Table 2 demonstrates the average validation accuracies for the six conditions using models trained by the neural couplings between brain ROIs and during featured time periods. Significant main effects were found in two-way rm-ANOVAs for validation accuracies using models trained by the top-down data (Fs > 7.55, ps $< .001$). Further post hoc tests showed significant differences between the self-matched and other conditions in models using the top-down data (ps $< .003$), as well as significant differences between the two friend-related conditions and stranger-matched conditions in the model using data from the left frontal to occipital regions (ps $< .042$). Overall, the results indicated a consistently higher accuracy of classifying the self-matched condition in models trained by the top-down data (82.19% using data from the left frontal to occipital regions, 89.15% using data from the right frontal to occipital regions) than those for friend-matched (65.45% using data from the left frontal to occipital regions, 75.00% using data from the right frontal to occipital regions) and stranger-related conditions (53.55% using data from the left frontal to occipital regions, 68.32% using data from the right frontal to occipital regions) (see Fig. 4(A) and Fig. 4(B)).

Furthermore, a significant difference was found between the friend-matched and stranger-matched conditions (p = .042), indicating a higher average validation accuracy for the friend-matched condition over the stranger-matched condition in models using data from left frontal to occipital regions. For mismatched conditions, unstable performance was observed. A significant effect was only found between the stranger-matched and stranger-mismatched condition in models using data from the left frontal to occipital regions ($p < .001$), indicating a lower average validation accuracy for the stranger-matched condition (53.55%) than that for the stranger-mismatched condition (75%). For models using the feedforward data, no significant main effect was revealed in the rm-ANOVAs (Fs < 2.04, ps $> .076$).

To validate the reliability of the model, the analysis was also replicated to data collected from another experiment, which measured the self-prioritisation effect while matching tasks in both first- and third-perspective conditions. The procedure of this experiment was identical to the experiment in this study except those stimuli were presented in a social setting where self- or other-related colour was mapped on the T-shirt of an avatar with the first-person (when the avatar was aligned with the participant's body) or third-person perspective. Results showed consistently high accuracies for classifying self-related matching conditions using the model based on SqueezeNet.

B. Performance difference for models using Top-down and Feed-forward Data.

Paired samples t-tests were employed to evaluate the statistical differences in classification performance between models using the top-down and feedforward data for each experimental condition. Most conditions showed similar performance using either the top-down or feedforward models (ps $> .094$). However, significant differences were observed in the self-matched condition using data between the right frontal and occipital regions (p = .004), as well as the self-mismatched and stranger-matched conditions using data between the left frontal and occipital regions (ps $< .011$). Specifically, validation accuracies in classifying the self-matched condition were significantly higher in models using the right-hemisphere top-down data than in models using the right-hemisphere feedforward data (see Fig. 5(A)). Also, better performance for classifying the self-mismatched condition and the stranger-matched condition was observed in models using left-hemisphere feedforward data compared to those from models using the left-hemisphere top-down data (see Fig. 5(B) and Fig. 5(C)).

IV. Discussion

A. Dynamic Neural Coupling Analysis for Self-saliency Processing

In this study, time-frequency analysis was applied with the novel multiple wavelet analysis method to extract features of dynamic neural couplings linked to self-saliency processing, and extended the neural mechanism related to the self-prioritisation effect, extended the neural mechanism related to self-saliency processing to a further level of dynamic interaction between featured brain regions.

The early step of the analysis showed increased power in the Alpha frequency band after the onset of matching trials, indicating increased neural activation in the Alpha band during experimental trials. Furthermore, time-frequency analysis extracted neural couplings between three featured ROIs located at the left and right frontal and the occipital regions. The frontal regions have been revealed as a critical part of neural networks linked to self-referencing processing in recent studies [26] [52], while occipital regions were highly linked to visual processing [53] [54]. Even though it is challenging to locate neural couplings to specific regions such as mPFC or dorsolateral prefrontal cortex (dlPFC) [55], which was highly linked to self-referencing processing, increased neural couplings found around the frontal regions were consistent with the previous finding that functional connectivity was significantly activated at the vlPFC region while processing self-related stimuli, compared to other-related stimuli [56].

All matching conditions in the experiment showed robust features of neural couplings between frontal and occipital ROIs in the Alpha band. Before the onset of experimental trials, low levels of top-down neural couplings were observed across matching conditions. This feature matches the finding in a previous study stating that there was a low level of top-down connectivity from the frontal to occipital brain regions under

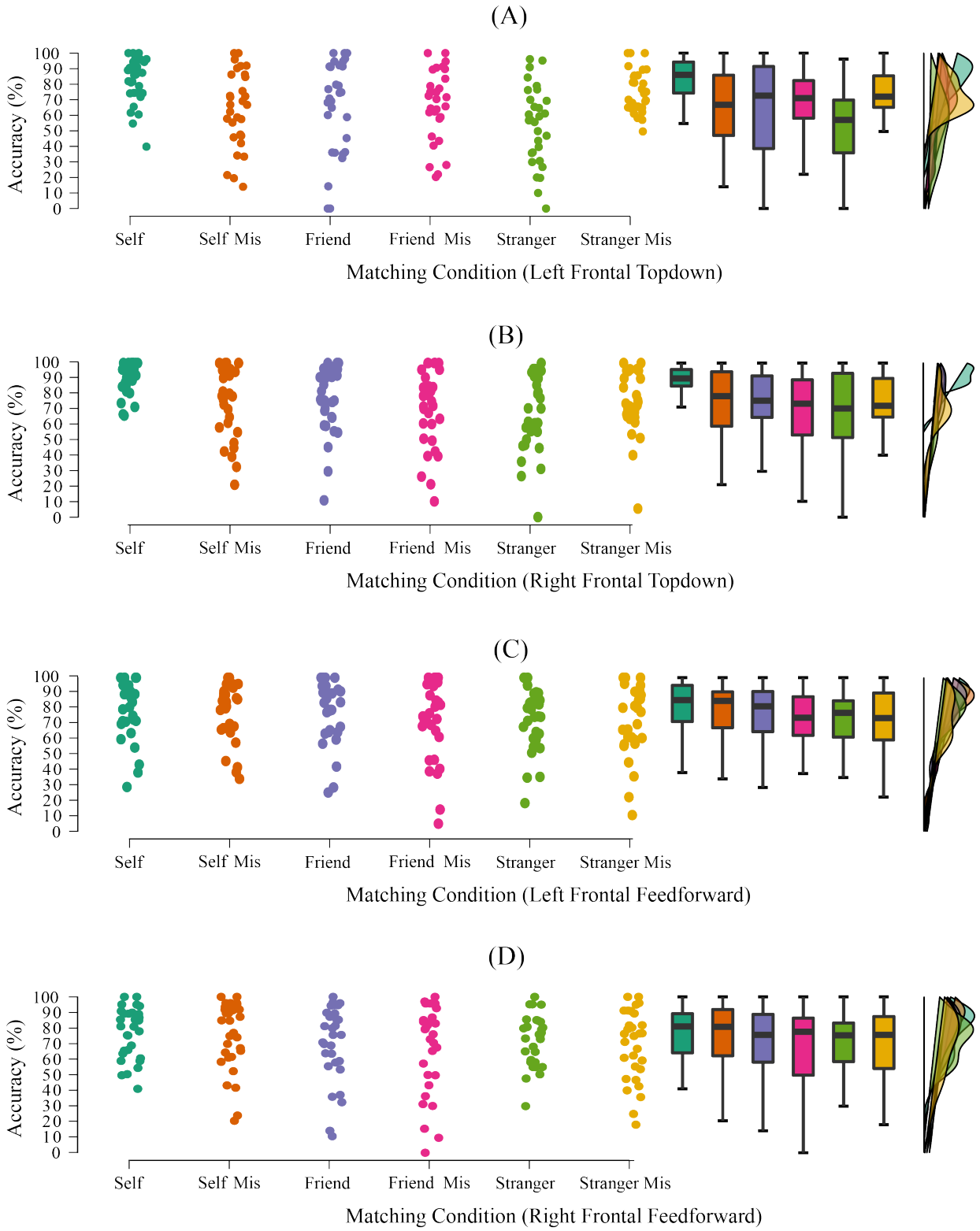

Fig. 4. Validation accuracies over all individual participants in classifying the conditions using SqueezeNet-based models trained by data from top-down and feedforward neural coupling periods. (A) Early top-down neural couplings from the left frontal to occipital regions, (B) Early top-down neural couplings from the right frontal to occipital regions, (C) Later feedforward neural couplings from the left frontal to occipital regions, (D) Later feedforward neural couplings from the right frontal to occipital regions.

TABLE II

THE AVERAGE VALIDATION ACCURACIES OF USING DEEP LEARNING MODELS USING SQUEEZENET TRAINED BY THE NEURAL COUPLING DATA DURING FEATURED PERIODS BETWEEN THE LEFT / RIGHT FRONTAL REGION AND THE OCCIPITAL REGION TO CORRECTLY CLASSIFY DYNAMIC NEURAL COUPLINGS IMAGES FOR DIFFERENT MATCHING CONDITIONS.

	Self-Matched	Self-Mismatched	Friend-Matched	Friend-Mismatched	Stranger-matched	Stranger-Mismatched
Left Frontal to Occipital	82.19	63.92	65.45	66.54	53.55	75
Right Frontal to Occipital	89.15	74.15	75	69.88	68.32	73.4
Occipital to Left Frontal	79.24	77.98	77.32	70.83	72.69	71.03
Occipital to Right Frontal	76.76	74.75	69.95	67.01	72.17	68.84

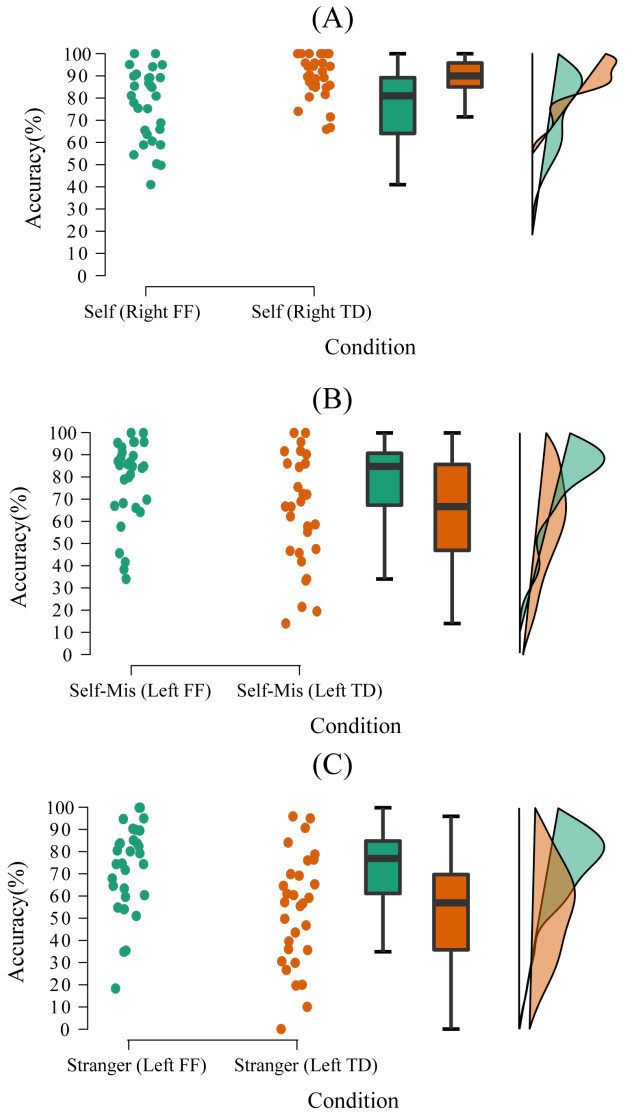

Fig. 5. Performance comparison between models using the top-down (TD) and feedforward (FF) neural coupling data. Validation accuracies for the Self-matched (A), Self-mismatched (B) and Stranger-matched (C) conditions were illustrated.

the resting state [57]. After the onset of experimental trials, two significant periods were observed with robust neural couplings between frontal and occipital ROIs. Top-down neural couplings were firstly increased across matching conditions, with stronger neural couplings for the self-matched condition compared to the friend- and stranger-matched conditions, indicating that the increased top-down couplings were linked to self-saliency processing since it varied according to the self-relevance in matching trials.

Top-down couplings were followed by a period containing neural couplings from occipital to frontal ROIs, revealing the information feedforward between featured brain regions. Partial significant effects were observed in this period across matched conditions from occipital to right frontal ROIs. Interestingly, self-matched conditions did not show consistently stronger couplings than those for other-matched conditions. These couplings could be linked to other advanced neural activities while processing other-related information and giving correct answers for them as self-saliency processing may be influenced by underlying factors such as reward and emotional value [15] [58] [59] [60] [61].

B. Powerful Deep Learning Models for Quantifying Robustness of Neural Coupling Features

As a powerful method, deep learning has been utilised in constructing classification models using features extracted from neurological data [62]. Deep learning methods are extensively used in research of robustness verification since they can give a safety guarantee to the robustness property. In this case, outputs will be invariant even if inputs have perturbations [63]. Therefore, it is suitable to be applied in the research of neuroscience since small perturbations are commonly distributed in different independent trials.

In the second part of the study, deep learning models were constructed using SqueezeNet, a recently proposed CNN algorithm to classify neural coupling data for experimental trials containing stimuli with different degrees of self-reference. A deep learning method was used for evaluating the correlation between specific neural coupling features and self-prioritisation effects. Different models were built up using dynamic neural coupling data with specific neural coupling features (the top-down period vs. feedforward period). The results showed that there were higher accuracies for classifying the self-related condition than those for classifying friend- or stranger-matched conditions in models trained by the early

top-down neural coupling features. The findings provide evidence that top-down neural couplings following the onset of stimuli were initially linked to self-saliency processing, via activation of self-representation in the frontal regions, and then enhanced visual processing in the occipital regions and later decision making.

Interestingly, the models trained by feedforward neural coupling data showed increased accuracies for classifying stranger-matched conditions over those from the models trained by early top-down features, revealing that this period was more likely linked to conditions related to other people.

Previous studies have revealed the critical roles of top-down control in self-saliency processing. The present findings support this by adding new information with specific time windows for dynamic processing. In addition, the accuracy difference between models trained by data from different communication periods provides evidence supporting different neural mechanisms while processing information with self-relevance, revealing the robust link between the specific top-down communication from frontal to occipital regions and self-prioritisation. Furthermore, this research proposes a feasible way of applying deep learning models to evaluate the relationship between specific neural features and psychological effects. Even though the validation accuracies for all matching conditions could hit as high as expected if iteration numbers were high enough, the performance of classifying different conditions would be varied under the condition of low iterations. It was due to the robustness of features in training data to the specialised classification. Classification with more robust features will be easier for models to distinguish, especially in the case of low iterations.

In this study, higher validation accuracies were observed in classifying the self-matched condition from models using the early top-down neural coupling features than those for other-related conditions. Therefore, there is evidence showing that the features used in training models in the early top-down period had higher robustness for the self-matched condition, which contained more self-related information and the corresponding stronger self-saliency processing, compared to other-related conditions. Also, higher validation accuracies for classifying self-related conditions were observed in models using early top-down couplings than those using later feedforward data, revealing more robust features for self-saliency processing compared to those in the later feedforward period. These findings also matched the statistical results of significant effects applied to dynamic neural coupling data, showing enhanced top-down neural couplings for the self-matched condition over those for other-related conditions.

In contrast, higher accuracies for classifying the stranger-matched condition were observed in models trained by the later feedforward data compared to models using the early top-down data, revealing more robust features for the stranger-matched condition in the later feedforward period. The finding matched the results in the first part of this study showing stronger neural couplings for the stranger-matched condition than those for the self-matched condition in this period. In addition, this finding provides evidence to support the previously proposed hypothesis that the strong neural couplings in this period while processing stranger-related stimuli could be related to other advanced neural activities, which were demonstrated via robust features that could lead to models with higher accuracy for classifying stranger-related trials.

V. Conclusion

This study used deep learning approaches to demonstrate a better understanding of psychological functioning – self-saliency processing producing self-prioritisation effects in behaviours. Dynamic connections between the frontal and occipital regions during self-saliency processing provide evidence supporting the down-regulation of the frontal regions to visual processing in the occipital regions and then a follow-up feedforward processing, which led to optimal behaviours. In addition, neural features and specialised classifications derived from the deep learning analyses may inform how to implement knowledge of how the human brain operates, producing optimal behaviour, in developing brain-inspired algorithms.

References

[1] C. S. Symons and B. T. Johnson, "The self-reference effect in memory: a meta-analysis." *Psychological bulletin*, vol. 121, no. 3, p. 371, 1997.

[2] T. B. Rogers, N. A. Kuiper, and W. S. Kirker, "Self-reference and the encoding of personal information." *Journal of personality and social psychology*, vol. 35, no. 9, p. 677, 1977.

[3] J. Sui and G. W. Humphreys, "The integrative self: How self-reference integrates perception and memory," *Trends in cognitive sciences*, vol. 19, no. 12, pp. 719–728, 2015.

[4] J. Sui, X. He, and G. W. Humphreys, "Perceptual effects of social salience: evidence from self-prioritization effects on perceptual matching." *Journal of Experimental Psychology: Human perception and performance*, vol. 38, no. 5, p. 1105, 2012.

[5] S. J. Cunningham, J. L. Brebner, F. Quinn, and D. J. Turk, "The self-reference effect on memory in early childhood," *Child Development*, vol. 85, no. 2, pp. 808–823, 2014.

[6] A. Hamami, S. J. Serbun, and A. H. Gutchess, "Self-referencing enhances memory specificity with age." *Psychology and aging*, vol. 26, no. 3, p. 636, 2011.

[7] W. Zhang, I.-T. Hung, J. D. Jackson, T.-L. Tai, J. O. S. Goh, and A. Gutchess, "Influence of culture and age on the self-reference effect," *Aging, Neuropsychology, and Cognition*, vol. 27, no. 3, pp. 370–384, 2020.

[8] C. Desebrock, J. Sui, and C. Spence, "Self-reference in action: Arm-movement responses are enhanced in perceptual matching," *Acta Psychologica*, vol. 190, pp. 258–266, 2018.

[9] Y. Sun, L. J. Fuentes, G. W. Humphreys, and J. Sui, "Try to see it my way: Embodied perspective enhances self and friend-biases in perceptual matching," *Cognition*, vol. 153, pp. 108–117, 2016.

[10] C. Burden, R. C. Leach, A. M. Sklenar, P. Urban Levy, A. N. Frankenstein, and E. D. Leshikar, "Examining the influence of brain stimulation to the medial prefrontal cortex on the self-reference effect in memory," *Brain and Behavior*, vol. 11, no. 12, p. e2368, 2021.

[11] E. D. Leshikar and A. Duarte, "Medial prefrontal cortex supports source memory for self-referenced materials in young and older adults," *Cognitive, Affective, & Behavioral Neuroscience*, vol. 14, no. 1, pp. 236–252, 2014.

[12] F. Lian and G. Northoff, "The lost neural hierarchy of the autistic self—locked-out of the mental self and its default-mode network," *Brain Sciences*, vol. 11, no. 5, p. 574, 2021.

[13] G. Northoff and F. Bermpohl, "Cortical midline structures and the self," *Trends in cognitive sciences*, vol. 8, no. 3, pp. 102–107, 2004.

[14] S. Schäfer and C. Frings, "Searching for the inner self: Evidence against a direct dependence of the self-prioritization effect on the ventro-medial prefrontal cortex," *Experimental brain research*, vol. 237, no. 1, pp. 247–256, 2019.

[15] A. Yankouskaya and J. Sui, "Self-positivity or self-negativity as a function of the medial prefrontal cortex," *Brain sciences*, vol. 11, no. 2, p. 264, 2021.

[16] J. Sui and G. W. Humphreys, "The ubiquitous self: What the properties of self-bias tell us about the self," *Annals of the New York Academy of Sciences*, vol. 1396, no. 1, pp. 222–235, 2017.

[17] J. Sui, C. H. Liu, and S. Han, "Cultural difference in neural mechanisms of self-recognition," *Social Neuroscience*, vol. 4, no. 5, pp. 402–411, 2009.

[18] J. Sui, Y. Zhu, and S. Han, "Self-face recognition in attended and unattended conditions: an event-related brain potential study," *Neuroreport*, vol. 17, no. 4, pp. 423–427, 2006.

[19] A. Yankouskaya and J. Sui, "Self-prioritization is supported by interactions between large-scale brain networks," *European Journal of Neuroscience*, vol. 55, no. 5, pp. 1244–1261, 2022.

[20] C. N. Macrae, J. M. Moran, T. F. Heatherton, J. F. Banfield, and W. M. Kelley, "Medial prefrontal activity predicts memory for self," *Cerebral cortex*, vol. 14, no. 6, pp. 647–654, 2004.

[21] P. Martinelli, M. Sperduti, and P. Piolino, "Neural substrates of the self-memory system: New insights from a meta-analysis," *Human brain mapping*, vol. 34, no. 7, pp. 1515–1529, 2013.

[22] G. Northoff, A. Heinzel, M. De Greck, F. Bermpohl, H. Dobrowolny, and J. Panksepp, "Self-referential processing in our brain—a meta-analysis of imaging studies on the self," *Neuroimage*, vol. 31, no. 1, pp. 440–457, 2006.

[23] D. Stendardi, F. Biscotto, E. Bertossi, and E. Ciaramelli, "Present and future self in memory: the role of vmpfc in the self-reference effect," *Social Cognitive and Affective Neuroscience*, vol. 16, no. 12, pp. 1205–1213, 2021.

[24] C. L. Philippi, M. C. Duff, N. L. Denburg, D. Tranel, and D. Rudrauf, "Medial pfc damage abolishes the self-reference effect," *Journal of cognitive neuroscience*, vol. 24, no. 2, pp. 475–481, 2012.

[25] J. Sui, P. Rotshtein, and G. W. Humphreys, "Coupling social attention to the self forms a network for personal significance," *Proceedings of the National Academy of Sciences*, vol. 110, no. 19, pp. 7607–7612, 2013.

[26] J. Sui and P. Rotshtein, "Self-prioritization and the attentional systems," *Current opinion in psychology*, vol. 29, pp. 148–152, 2019.

[27] P. Babaeeghazvini, L. M. Rueda-Delgado, J. Gooijers, S. P. Swinnen, and A. Daffertshofer, "Brain structural and functional connectivity: A review of combined works of diffusion magnetic resonance imaging and electro-encephalography," *Frontiers in human neuroscience*, p. 585, 2021.

[28] K. S. Damme, A. Pelletier-Baldelli, H. R. Cowan, J. M. Orr, and V. A. Mittal, "Distinct and opposite profiles of connectivity during self-reference task and rest in youth at clinical high risk for psychosis," *Human brain mapping*, vol. 40, no. 11, pp. 3254–3264, 2019.

[29] R. J. Murray, M. Debbane, P. T. Fox, D. Bzdok, and S. B. Eickhoff, "Functional connectivity mapping of regions associated with self-and other-processing," *Human brain mapping*, vol. 36, no. 4, pp. 1304–1324, 2015.

[30] M. L. Ries, D. G. McLaren, B. B. Bendlin, H. A. Rowley, R. Birn, E. K. Kastman, M. A. Sager, S. Asthana, S. C. Johnson *et al.*, "Medial prefrontal functional connectivity—relation to memory self-appraisal accuracy in older adults with and without memory disorders," *Neuropsychologia*, vol. 50, no. 5, pp. 603–611, 2012.

[31] G. Nolte, O. Bai, L. Wheaton, Z. Mari, S. Vorbach, and M. Hallett, "Identifying true brain interaction from eeg data using the imaginary part of coherency," *Clinical neurophysiology*, vol. 115, no. 10, pp. 2292–2307, 2004.

[32] A. K. Bonkhoff, F. A. Espinoza, H. Gazula, V. M. Vergara, L. Hensel, J. Michely, T. Paul, A. K. Rehme, L. J. Volz, G. R. Fink *et al.*, "Acute ischaemic stroke alters the brain's preference for distinct dynamic connectivity states," *Brain*, vol. 143, no. 5, pp. 1525–1540, 2020.

[33] I. Buller-Peralta, J. Maicas-Royo, Z. Lu, S. M. Till, E. R. Wood, P. C. Kind, J. Escudero, and A. Gonzalez-Sulser, "Abnormal brain state distribution and network connectivity in a syngap1 rat model," *Brain Communications*, vol. 4, no. 6, p. fcac263, 2022.

[34] N. de Lacy and V. D. Calhoun, "Dynamic connectivity and the effects of maturation in youth with attention deficit hyperactivity disorder," *Network Neuroscience*, vol. 3, no. 1, pp. 195–216, 2018.

[35] Z. Fu, A. Caprihan, J. Chen, Y. Du, J. C. Adair, J. Sui, G. A. Rosenberg, and V. D. Calhoun, "Altered static and dynamic functional network connectivity in alzheimer's disease and subcortical ischemic vascular disease: shared and specific brain connectivity abnormalities," *Human brain mapping*, vol. 40, no. 11, pp. 3203–3221, 2019.

[36] V. V. Moca, H. Bârzan, A. Nagy-Dăbâcan, and R. C. Mureșan, "Time-frequency super-resolution with superlets," *Nature communications*, vol. 12, no. 1, pp. 1–18, 2021.

[37] F. L. Da Silva, T. Van Lierop, C. Schrijer, and W. S. Van Leeuwen, "Organization of thalamic and cortical alpha rhythms: spectra and coherences," *Electroencephalography and clinical neurophysiology*, vol. 35, no. 6, pp. 627–639, 1973.

[38] J. Mo, Y. Liu, H. Huang, and M. Ding, "Coupling between visual alpha oscillations and default mode activity," *Neuroimage*, vol. 68, pp. 112–118, 2013.

[39] O. Jensen, B. Gips, T. O. Bergmann, and M. Bonnefond, "Temporal coding organized by coupled alpha and gamma oscillations prioritize visual processing," *Trends in neurosciences*, vol. 37, no. 7, pp. 357–369, 2014.

[40] Z. Aslan and M. Akin, "A deep learning approach in automated detection of schizophrenia using scalogram images of eeg signals," *Physical and Engineering Sciences in Medicine*, vol. 45, no. 1, pp. 83–96, 2022.

[41] F. Demir, N. Sobahi, S. Siuly, and A. Sengur, "Exploring deep learning features for automatic classification of human emotion using eeg rhythms," *IEEE Sensors Journal*, vol. 21, no. 13, pp. 14 923–14 930, 2021.

[42] J. Sun, R. Cao, M. Zhou, W. Hussain, B. Wang, J. Xue, and J. Xiang, "A hybrid deep neural network for classification of schizophrenia using eeg data," *Scientific Reports*, vol. 11, no. 1, pp. 1–16, 2021.

[43] K. Tziridis, T. Kalampokas, and G. A. Papakostas, "Eeg signal analysis for seizure detection using recurrence plots and tchebichef moments," in *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 2021, pp. 0184–0190.

[44] H. Zeng, C. Yang, G. Dai, F. Qin, J. Zhang, and W. Kong, "Eeg classification of driver mental states by deep learning," *Cognitive neurodynamics*, vol. 12, no. 6, pp. 597–606, 2018.

[45] S. Min, B. Lee, and S. Yoon, "Deep learning in bioinformatics," *Briefings in bioinformatics*, vol. 18, no. 5, pp. 851–869, 2017.

[46] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size," *arXiv preprint arXiv:1602.07360*, 2016.

[47] M. X. Cohen, "A better way to define and describe morlet wavelets for time-frequency analysis," *NeuroImage*, vol. 199, pp. 81–86, 2019.

[48] R. Oostenveld and P. Praamstra, "The five percent electrode system for high-resolution eeg and erp measurements," *Clinical neurophysiology*, vol. 112, no. 4, pp. 713–719, 2001.

[49] R. M. Hutchison, T. Womelsdorf, E. A. Allen, P. A. Bandettini, V. D. Calhoun, M. Corbetta, S. Della Penna, J. H. Duyn, G. H. Glover, J. Gonzalez-Castillo *et al.*, "Dynamic functional connectivity: promise, issues, and interpretations," *Neuroimage*, vol. 80, pp. 360–378, 2013.

[50] S. Mallat, *A wavelet tour of signal processing*. Elsevier, 1999.

[51] J. Qiu, J. Wang, S. Yao, K. Guo, B. Li, E. Zhou, J. Yu, T. Tang, N. Xu, S. Song *et al.*, "Going deeper with embedded fpga platform for convolutional neural network," in *Proceedings of the 2016 ACM/SIGDA international symposium on field-programmable gate arrays*, 2016, pp. 26–35.

[52] G. W. Humphreys and J. Sui, "Attentional control and the self: the self-attention network (san)," *Cognitive neuroscience*, vol. 7, no. 1-4, pp. 5–17, 2016.

[53] M. Glickstein, "The discovery of the visual cortex," *Scientific American*, vol. 259, no. 3, pp. 118–127, 1988.

[54] S. Kalenzaga, M. Sperduti, A. Anssens, P. Martinelli, A.-D. Devauchelle, T. Gallarda, M. Delhommeau, S. Lion, I. Amado, M.-O. Krebs *et al.*, "Episodic memory and self-reference via semantic autobiographical memory: insights from an fmri study in younger and older adults," *Frontiers in Behavioral Neuroscience*, vol. 8, p. 449, 2015.

[55] Q. Liang, B. Zhang, S. Fu, J. Sui, and F. Wang, "The roles of the lpsts and dlpfc in self-prioritization: A transcranial magnetic stimulation study," *Human brain mapping*, vol. 43, no. 4, pp. 1381–1393, 2022.

[56] B. T. Denny, H. Kober, T. D. Wager, and K. N. Ochsner, "A meta-analysis of functional neuroimaging studies of self-and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex," *Journal of cognitive Neuroscience*, vol. 24, no. 8, pp. 1742–1752, 2012.

[57] C. Wang, X. Wang, M. Zhu, Y. Pi, X. Wang, F. Wan, S. Chen, and G. Li, "Spectrum power and brain functional connectivity of different eeg frequency bands in attention network tests," in *2021 43rd Annual*

International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2021, pp. 224–227.

[58] K. Kim, A. M. Banquer, S. N. Resnik, J. D. Johnson, and L. Fernandez, "Self-reference and cognitive effort: source memory for affectively neutral information is impaired following negative compared to positive self-referential processing," *Journal of Cognitive Psychology*, pp. 1–13, 2022.

[59] Y. Ma and S. Han, "Why we respond faster to the self than to others? an implicit positive association theory of self-advantage during implicit face recognition." *Journal of Experimental Psychology: Human Perception and Performance*, vol. 36, no. 3, p. 619, 2010.

[60] G. Northoff and D. J. Hayes, "Is our self nothing but reward?" *Biological psychiatry*, vol. 69, no. 11, pp. 1019–1025, 2011.

[61] J. Sui and G. W. Humphreys, "The interaction between self-bias and reward: Evidence for common and distinct processes," *The Quarterly Journal of Experimental Psychology*, vol. 68, no. 10, pp. 1952–1964, 2015.

[62] G. Sinha, "Study of assessment of cognitive ability of human brain using deep learning," *International Journal of Information Technology*, vol. 9, no. 3, pp. 321–326, 2017.

[63] R. Li, J. Li, C.-C. Huang, P. Yang, X. Huang, L. Zhang, B. Xue, and H. Hermanns, "Prodeep: a platform for robustness verification of deep neural networks," in *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, 2020, pp. 1630–1634.