Ventral and Dorsal Stream EEG Channels: Key Features for EEG-Based Object Recognition and Identification

Daniel Leong[ID], Thomas (Tien-Thong) Do, and Chin-Teng Lin[ID], *Fellow, IEEE*

Abstract—**Object recognition and object identification are multifaceted cognitive operations that require various brain regions to synthesize and process information. Prior research has evidenced the activity of both visual and temporal cortices during these tasks. Notwithstanding their similarities, object recognition and identification are recognized as separate brain functions. Drawing from the two-stream hypothesis, our investigation aims to understand whether the channels within the ventral and dorsal streams contain pertinent information for effective model learning regarding object recognition and identification tasks. By utilizing the data we collected during the object recognition and identification experiment, we scrutinized EEGNet models, trained using channels that replicate the two-stream hypothesis pathways, against a model trained using all available channels. The outcomes reveal that the model trained solely using the temporal region delivered a high accuracy level in classifying four distinct object categories. Specifically, the object recognition and object identification models achieved an accuracy of 89% and 85%, respectively. By incorporating the channels that mimic the ventral stream, the model's accuracy was further improved, with the object recognition model and object identification model achieving an accuracy of 95% and 94%, respectively. Furthermore, the Grad-CAM result of the trained models revealed a significant contribution from the ventral and dorsal stream channels toward the training of the EEGNet model. The aim of our study is to pinpoint the optimal channel configuration that provides a swift and accurate brain-computer interface system for object recognition and identification.**

Index Terms—**Brain-computer interfaces (BCIs), object identification, object recognition, electroencephalogram (EEG), EEG-Net, deep learning, ventral stream, dorsal stream.**

I. INTRODUCTION

IN OUR daily routines, we are faced with the task of decoding a large amount of changing visual information. In order to effectively engage with our surroundings, it is essential for our visual system to rapidly identify and interpret the visual information present in our environment. Impressively, our brains demonstrate an exceptional ability to search for and perceive intricate images from nature with both speed and precision. Despite numerous investigations aiming to decode the functionality of our visual system, there is still a lack of comprehensive understanding of this intricate network. Nowadays, numerous theories and hypotheses have been proposed to explain how our brains recognize objects. The widely recognized two-streams hypothesis [1] is currently regarded as the prevailing model that explains the brain's visual processing mechanisms. This hypothesis suggests that when the occipital lobe, the brain's visual processing region, receives visual data, it splits it into two processing routes: the ventral and dorsal streams. The ventral stream sends its information to the temporal lobe, where the object is identified and recognized. The dorsal stream, on the other hand, is responsible for processing visual-spatial information and determining the object's location relative to the observer. This information is then relayed from the occipital lobe to the parietal lobe. Through fMRI and MEG studies, the activity of the ventral stream has been observed, thus highlighting its role in object recognition in the human brain [2].

Numerous EEG-based studies on object recognition have similar research designs, where participants are expected to respond by pressing a button when a target stimulus is displayed. The main focus of these studies lies in these target stimuli. They are often keywords related to the object, such as its category [3], or they are used to assess the object's meaningfulness [4] or ambiguity [5]. Nevertheless, a handful of studies have modified their experimental design to focus on object identification in the brain instead of object recognition. For example, some studies necessitate participants to view congruent and incongruent scenes where a key object remains constant, and they are asked to identify

that critical object among other objects in the scene [6], [7]. Another study requires participants to spot a target object that is either semantically consistent or inconsistent within a scene and press a button whenever the target object alters its identity, location, or both [8]. These studies highlight that the differentiation between object recognition and identification primarily hinges on the number of objects a person is presented with. When exposed to a single object, individuals will use object recognition. Conversely, when asked to distinguish among multiple objects, they resort to object identification. Despite the similarities between object recognition and identification, object identification is perceived as a unique process, with different brain regions engaged in processing the information [9].

Several researchers have endeavoured to develop Brain-Computer Interface (BCI) systems for object recognition and identification by using salient EEG features. Event-Related Potential (ERP), a prevalent EEG feature, is the brain's response to a specific event or stimulus. In the context of a BCI system for object recognition, ERPs are recorded when an object enters the participant's visual field and subsequently classified according to the object's category ([10], [11], [12]). Conversely, an object identification BCI system relies on a visual stimulus, such as a flash or multiple flashes over the selected object, to provoke the ERP response. For instance, the P3-based BCI identifies objects based on the P3 peak, which occurs approximately 300-500 ms following event onset ([13], [14]). Beyond the P3 peak, the steady-state visually evoked potential (SSVEP) is another feature commonly used in BCI-based object identification systems. This technique involves placing a flicker at a specific frequency over the chosen object ([15], [16], [17]). However, despite the progress made in developing BCI systems that distinguish between object recognition and identification, the majority are not yet ready for practical application. The challenge lies in the systems' inability to discern the user's intention behind targeting an object: whether the user intends for the BCI to recognize the object or whether the user wants the BCI to select an object from their environment.

While Event-Related Potential (ERP) is a prevalent feature utilized in EEG/BCI studies due to the insightful data it provides about cognitive processes and neural activities associated with specific events, its identification and recognition by the naked eye can be challenging, as it can differ significantly across individuals. As a result, machine learning algorithms are employed to facilitate a more precise, efficient, and objective analysis of ERP. Among all machine learning algorithms applied to EEG data analysis, EEGNet has displayed encouraging results in various EEG analysis tasks [18]. EEGNet is a compacted convolutional neural network incorporating depthwise and separable convolutions, enabling the effective capture of both spatial and temporal information in EEG signals. Several studies have confirmed the effectiveness of EEGNet in analyzing object-related ERP ([19], [20], [21], [22], [23]) and other EEG features such as P300 [24]. Nonetheless, comprehending the effectiveness of the EEGNet model requires an explanation of how the model learned from EEG data. Consequently, the utilization of explanation techniques has gained prominence as a means to visualize EEGNet models. Notably, various studies have utilized explanation techniques, including saliency maps [25], [26] and Grad-CAM [27], [28], to highlight the noteworthy EEG channels within the trained EEG models. For better classification results of object-related ERP, researchers often aim to utilize as many channels as possible within the region of interest. However, increasing the number of channels also escalates the complexity and latency of the BCI system, which isn't practical for real-time applications.

The objective of our study is to identify the best channel configuration for a fast and accurate BCI system for object recognition and identification. We examined the model trained using channels that emulate the pathways of the two-stream hypothesis compared to the model trained using all channels. The aim is to determine whether the channels within the ventral and dorsal streams contain information that could facilitate effective learning of the model on tasks related to object recognition and identification.

II. METHODOLOGY

A. Participant and Data Recording

In this study, a total of 25 participants, with an average age of 32.5 ± 10.4 years and either normal or corrected-to-normal vision, were involved. The participants undertook 600 trials each, conducted at the Computation Intelligence and Brain-Computer Interface (CIBCI) Centre situated at the University of Technology Sydney (UTS). Prior to the experiment, the participants were briefed on the instructions and were required to sign a consent form after being informed. The University of Technology Sydney granted ethical approval for this study under the ethics ID ETH20-5519.

We recorded the brain activities of the participants using a 64-channel EEG system produced by Neuroscan Compumedics Australia. This medical-grade device, known for its high-density EEG recordings and high precision, has been extensively utilised in previous neuroscience and neurodiagnostics research. The EEG electrodes were positioned according to the extended 10-20 international system, and the data was referenced to an electrode closest to the standard position FCZ. We maintained the electrode impedance below 5 kΩ and digitally sampled the EEG recordings at a rate of 1000 Hz.

B. Experimental Design

In the course of the experiment, participants were asked to undertake two tasks: object recognition and object identification. The object recognition task consisted of presenting randomly selected images from four categories of the Caltech-256 dataset [29], namely animals, flowers, food, and vehicles. Each category contained five distinct objects, with ten images per object, yielding a total of 200 images used in the experiment. At the beginning of the trial, participants were displayed a target image for a duration of 1 second and then prompted to answer if it was part of the specified category (see Figure 1). The objective of this task was to assess the participant's accuracy in recognizing the target image.

Following the object recognition task, participants were asked to perform an object identification task. For this task, four images were randomly chosen from the dataset and presented in a four-image configuration (up, down, left, and right). However, at least one of the objects displayed was
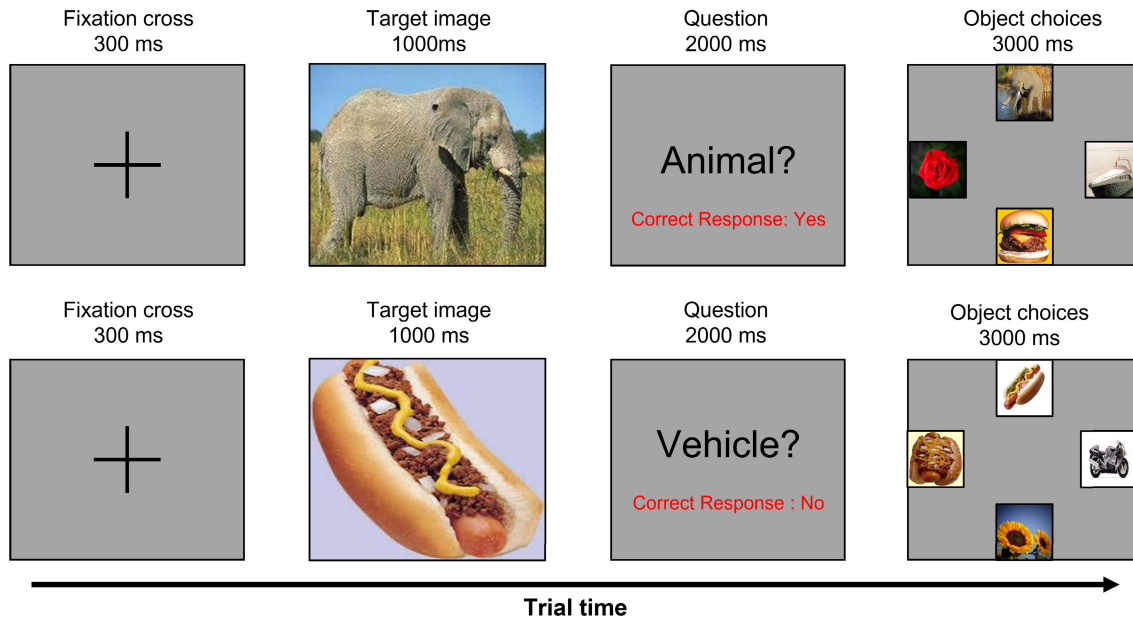
Fig. 1. The structure of our experimental design consisted of the following steps: Every trial began with a 300 ms display of a fixation cross, followed by a target image that remained visible for 1 second. Subsequently, participants were asked to identify the category of the object portrayed in the target image, with a response time window of 2 seconds. Then, an additional set of 4 images was presented, with at least one image being from the same category and subtype as the initial target image. Participants had 3 seconds to select the image they felt most closely resembled the target object by pressing a button that corresponded with the direction of their selected image (up, down, left, or right). The upper section of the figure illustrates a sample trial, which includes the correct response to the question (highlighted in red text) and the arrangement of the four object choices. Conversely, the lower section of the figure represents a trial instance where two out of the four options belong to the same category and subtype as the target object, thus presenting two closely matching alternatives to the target object.

from the same category and subtype as in the preceding object recognition task. Participants were directed to select the image most closely related to the target image by pressing a button corresponding to the up, down, left or right directions within 3 seconds. Note that, due to the random selection of images, more than one image from the same category and subtype could be presented as options. Each trial lasted a total of 6 seconds, with a fixation cross appearing for 300 ms to mark the start of the trial. An example trial is illustrated in Figure 1. Each participant will perform 600 trials of the task, resulting in a total of 15000 trials over 25 participants.

C. EEG Analysis

The processing of EEG signals was carried out using EEGLAB v14.1.2 [30], a MATLAB toolbox. The unprocessed EEG data underwent filtering through a finite impulse response (FIR) filter, consisting of a 1 Hz high-pass and a 50 Hz low-pass filter. Channels identified as noisy were excluded using the EEGALB function 'clean channels' (3 ±2 channels per subject removed), and the data was re-referenced to the average. Following this, the adaptive mixed independent component analysis (AMICA) was implemented on the re-referenced data to decompose it into maximally independent components (ICs). These ICs represent statistically independent sources of EEG variance. Using the IClabel toolbox [31], we removed ICs associated with eye movement and muscle activity (3 ±1 ICs per subject removed). After discarding these undesired components, epochs were extracted. Each epoch spanned the entire trial duration, starting from 300 ms prior to the appearance of the target image (i.e. the event onset) and ending 5 seconds post-event onset. We identified and removed bad epochs by examining their data values

and considering whether they exceeded the specified standard deviation threshold of 150 uV (394 ±57 trials per subject after removal). The epoched data were subsequently divided into two categories based on the object-related tasks. For both object recognition and identification tasks, a one second segment post-stimulation was extracted, resulting in a matrix of dimensions 60 (electrodes) × 1000(sampling points) x number of epochs.

D. EEGNet

1) EEGNet Structure: The EEGNet architecture utilized in this study adheres to a standard block structure comprising a temporal convolution layer, a depthwise convolution layer, and a separable convolution layer. The first layer, the temporal convolution layer, learns temporal filters by applying convolution operations on the input EEG data over time. It possesses filters that cover only a single EEG channel and multiple time points, preventing any mixing of data from different channels. The purpose of this layer is to learn time-dependent features, such as oscillations in the EEG signal, which represent changes in brain activity during object-related tasks. The next layer, the depthwise convolution layer, carries out depthwise convolutions. This layer applies a distinct set of filters to each input channel separately, possessing filters that span multiple channels and time points, allowing the model to learn spatial filters across channels as they evolve over time. This layer's purpose is to learn spatial features, reflecting the distribution of brain activity across various brain areas or channels. These spatial features can aid in identifying patterns associated with specific brain states or tasks. Following the depthwise convolution layer, the model uses the separable convolution layer. This layer applies a depthwise spatial
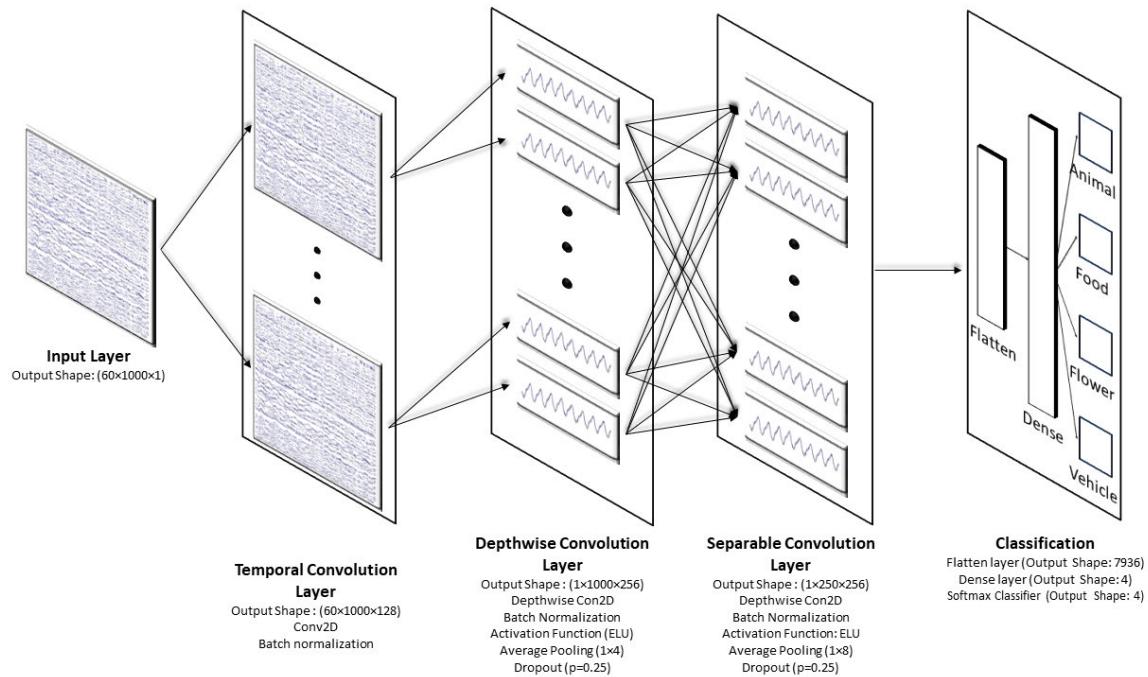
Fig. 2. A comprehensive depiction of the EEGNet architecture is presented. The lines represent the connectivity facilitated by the convolution kernel between the input and output.

convolution followed by a pointwise convolution. Essentially, it applies a separate set of filters to each input channel and uses a 1×1 convolution to mix the output channels. This approach allows the model to learn more complex and abstract features that combine spatial and temporal information. This layer adds an extra layer of complexity to the learned features, potentially enhancing the model's accuracy.

Following the depthwise convolution and separable convolution layers, batch normalization is used to boost the neural networks' speed, efficacy, and stability by normalizing the output from the previous layer. This step aids in learning stability and acceleration. Following batch normalization, the model utilizes the exponential linear unit (ELU) activation function. The ELU activation function's ability to introduce non-linearity into the model is vital for EEG data. Furthermore, the ELU activation function can hasten learning because it generates a balanced output with an average closer to zero and can mitigate the dead neuron issue [32]. The output of the ELU activation function is then subjected to an average pooling operation that reduces its dimensionality and offers a degree of translation invariance. A dropout layer follows, which helps prevent overfitting by providing a form of regularization. The output from the preceding layer is then reshaped via the Flatten layer and passed through a dense layer. This dense layer utilizes the features learned by the preceding layers for the final classification. Ultimately, the softmax function is applied to convert the network output into probability scores for each class. The overall structure is illustrated in Figure 2.

2) Training Procedure: After removing bad epochs, a total of 9835 epochs remained from the collective pool of 25 participants. This epoch dataset was divided into training, testing, and validation sets, with the training set comprising 80% of the entire dataset and the remaining 20% split evenly between testing and validation sets. The model was initially compiled using the Adam optimizer [33] and the categorical cross-entropy loss function [34], which provides the necessary tools and standards to modify the model's parameters during its training phase. Despite the EEGNet structure already incorporating elements designed to help prevent model overfitting, such as batch normalization and dropout, we introduced additional techniques to aid in model training. These included early stopping and a learning rate schedule. Early stopping uses a validation set to assess the model's performance following each epoch and halts the training when the performance on the validation set begins to decline [35]. On the other hand, learning rate schedules offer a mechanism for adjusting the learning rate throughout training by reducing the learning rate based on a predefined schedule, which for this training was an exponential decay. After all, we ensured that the training loss and validation loss were approximately equal and relatively low before progressing to model prediction.

III. RESULTS

A. ERP of Object Recognition and Object Identification

Figure 3 presents the distinctive traits of the ERP signals for both object recognition and object identification tasks. These results were obtained by averaging the ERP outcomes from all 25 participants, with scalp topography visualized using EEGLAB's topoplot function. The upper portion of the figure demonstrates the average ERP throughout the trial for all channels, with ERPs corresponding to channels O1, P7, and T7 represented in blue, green, and red, respectively. The lower half of the figure displays the scalp topography across the trial for all four categories and their average. According to the scalp topography, the occipital area appears to be the most active
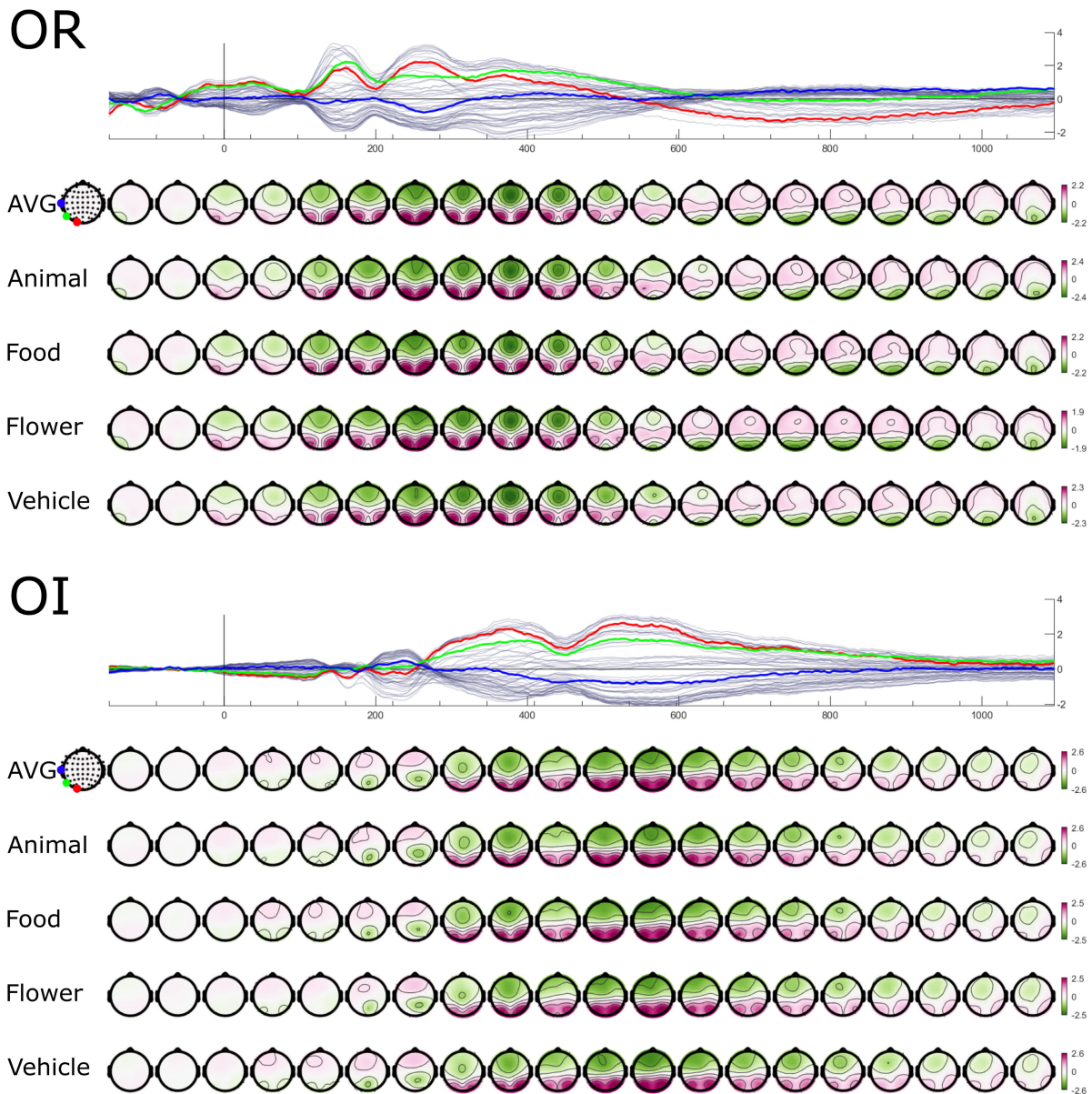
Fig. 3. Characteristics of the object recognition and object identification ERP signals. The top portion of the figure presents the average ERP results derived from all 25 participants, with the ERPs highlighted in blue, green, and red corresponding to channels O1, P7, and T7, respectively. The lower portion of the figure displays the scalp topography across the trial for all four categories, as well as their average. Each topography visualizes a 65ms segment of the trial.

region, from 100ms to 450ms during object recognition and from 300ms to 650ms during object identification.

B. Model Comparisons

Figure 4 presents a comparative analysis of the EEGNet model trained with differing configurations of EEG channels for both the tasks of object recognition and object identification at both group and participant levels. The channel configurations are grouped as follows: the visual region (O1, OZ, O2), the temporal region (T7, TP7, TP8, T8), the ventral stream (T7, T8, TP7, TP8, P7, P8, PO7, PO8, O1, OZ, O2), the dorsal stream (CZ, CPZ, PZ, POZ, OZ), combine both stream and all channels. Figure 4A shows the model accuracy using the grouped data of every participant. For object recognition, the results indicate an accuracy rate of 64% when trained with the visual region, 89% with the temporal region, 95% with the ventral stream, 79% with the dorsal stream, 96% with the combined stream, and 99% when trained with all channels. Regarding object identification, the model reached an accuracy of 65% when trained using the visual region, 85% with the temporal region, 94% with the ventral stream, 82% with the dorsal stream, 96% with the combined stream, and 96% when trained with all channels. Subsequent to the group analysis, an examination of the EEGNet models at the individual participant level was conducted. The outcomes of this examination are shown in Figure 4B. For object recognition, the results demonstrate an average accuracy with standard deviation, as follows: $73.4 \pm 9.4\%$ for models trained using the visual region, $80.7 \pm 7.4\%$ with the temporal region, $93.2 \pm 5.8\%$ with the ventral stream, $84.7 \pm 9.3\%$ with the dorsal stream, $96.9 \pm 3\%$ with the combined stream, and $99.6 \pm 0.2\%$ when utilizing all
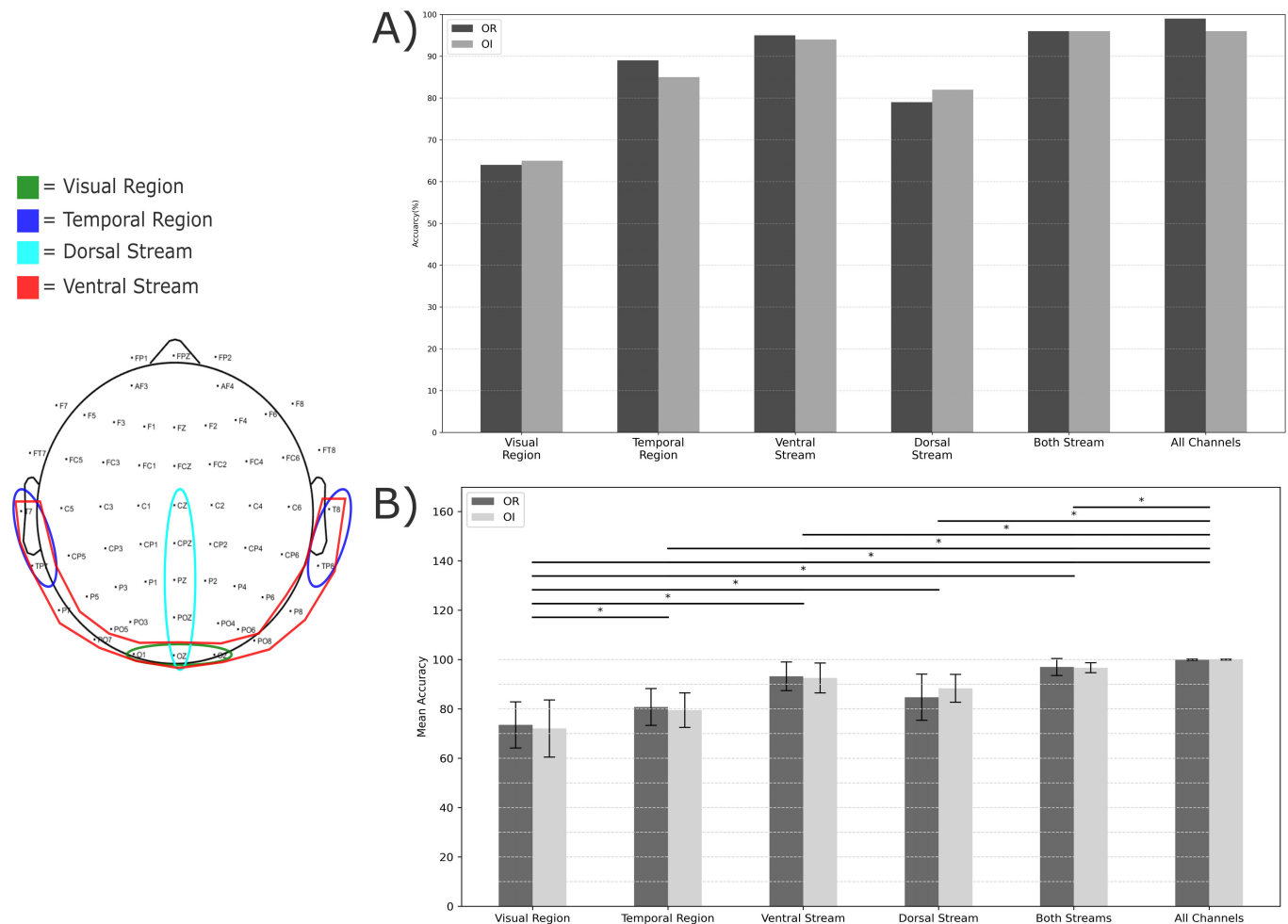
Fig. 4. Comparative result of the accuracy of various models at two distinct levels: group (A) and individual participants (B). Each trained using distinct channel configurations for the task of object recognition (OR) and object identification (OI). Paired t-tests were used to check for significant differences between models (* indicate p<0.05). A Topoplot is provided on the left side of the figure, emphasizing the specific channel configurations utilized in the study.

channels. In terms of object identification, the model achieved an accuracy of $72.04 \pm 11.6\%$ when trained with the visual region, $79.5 \pm 7\%$ with the temporal region, $92.5 \pm 6.1\%$ with the ventral stream, $88.3 \pm 5.6\%$ with the dorsal stream, $96.6 \pm 2\%$ with the combined stream, and $99.6 \pm 0.2\%$ when all channels were utilized. A paired t-test was also implemented to indicate the statistically significant differences between the various EEGNet models.

Furthermore, the trained models were subjected to visualization through the Grad-CAM technique [36]. The resultant Grad-CAM findings were obtained by running the technique on the output generated by the temporal convolution layer. It should be noted that the ERP corresponding to tasks of object recognition and object identification occurs at varying temporal intervals, as illustrated in Figure 3. Consequently, we chose the time frames during which the brain's response exhibited maximum amplitude in both object recognition and object identification tasks. For the model focused on object recognition, data ranging from 0 ms to 500 ms was selected, whereas for the object identification model, the selected data spanned from 300 ms to 800 ms. Figure 5 presents the Grad-CAM visualizations for both models across all channels, targeting four distinct object categories. In the case of the object recognition model,

the Grad-CAM visualizations indicate significant gradient scores primarily localized around the bi-lateral temporal and parietal regions for all categories, barring the flower category, where the significance is comparatively subdued. Conversely, the object identification model reveals pronounced gradient scores around the bilateral temporal and parietal regions for all categories, except for the vehicle category, where the significance is relatively reduced. Additionally, the results demonstrate heightened gradient scores in the frontal brain region across all categories, with the exception again being the vehicle category, where the importance is less notable.

IV. DISCUSSION

Object recognition and object identification play a pivotal role in numerous daily tasks. Both processes necessitate the rapid and precise process of abundant dynamic visual information coupled with the swift retrieval of information from memory. This study delves into these cognitive processes by analyzing EEGNet models trained using diverse EEG channel configurations. The results indicate that models trained solely on visual channels underperformed relative to other configurations in both object recognition and object identification tasks. Although the visual region primarily processes visual inputs from the eyes, its interpretation focuses on basic features of the
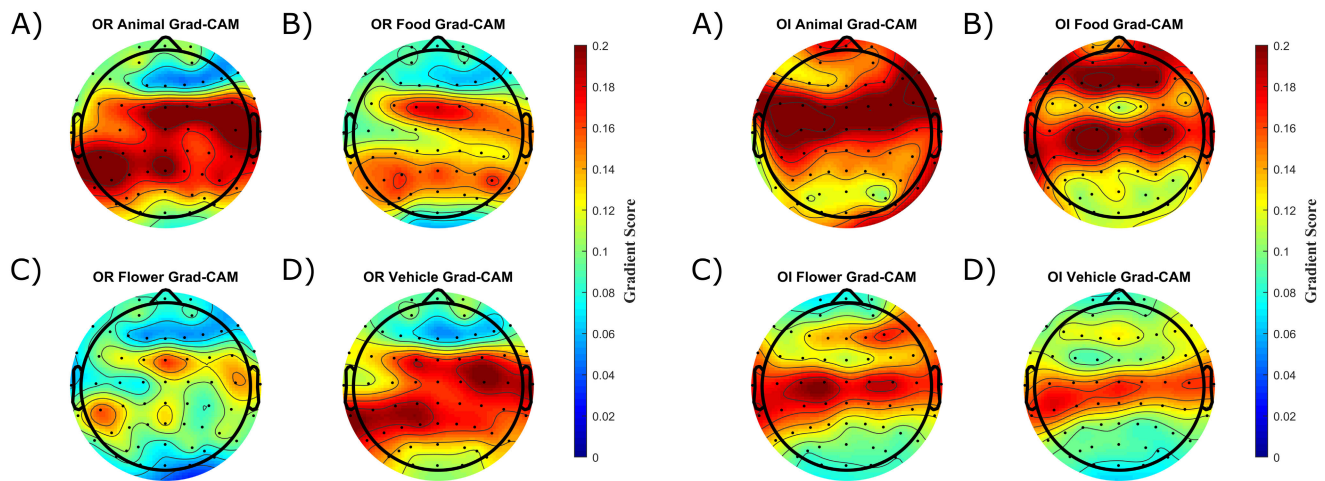
Fig. 5. The Grad-CAM outcomes of the object recognition (Left) and object identification (Right) model, incorporating all channels, were analyzed for the following categories: Animal (A), Food (B), Flower (C), and Vehicle (D).

visual scene, such as edges, lines, and colours [37], [38]. Given that the visual channels lacked comprehensive information distinguishing between categories, the suboptimal performance of this model was anticipated. Conversely, the temporal lobe is instrumental in recognizing and identifying intricate visual stimuli, including objects [39]. Thus, channels located within the temporal region potentially contain data that aids the model in distinguishing between distinct categories. Our findings revealed a marked increase in accuracy when models were trained exclusively on channels from the temporal region for both object recognition and identification tasks compared to models trained on visual channels.

Upon analyzing the ERP signals associated with object recognition and object identification used for model training, we identified several notable traits. As depicted in Figure 3, the average ERP within the visual channel for object recognition exhibited a pronounced double-peak potential approximately between 100 ms and 350 ms post-event onset. This double-peak potential can potentially be linked with the P2a and P2b responses related to object recognition as documented in prior research ([11], [12], [40]). A similar response was identified in the average ERP within the visual channel during object identification. Nonetheless, the temporal scope of this double-peak potential was extended, ranging from approximately 250 ms to 650 ms post-event onset. Such a phenomenon might be attributable to the detection of multiple images, extending the response duration since participants were tasked with recognizing all displayed objects and pinpointing the target image among the object choices.

Besides the visual channel, the temporal channel displayed a negative potential coinciding with the time frame of the double-peak potential. This is indicative of visual processing pertinent to both object recognition and identification ([41], [42], [43]). Similar to the patterns observed in the visual channel, the temporal region also showed a protracted negative potential during the object identification task. In addition to ERP signals, scalp topography illustrated the cerebral responses during object recognition and identification tasks. Upon the onset of object-related ERP throughout the tasks, a synchronized activity was evident in the scalp topography, with the preponderance of the activity manifesting in the

brain's occipital region. As the task progressed, activity was also discerned in the parietal and occipitotemporal regions. This observation accentuates the interrelation between ERP and the information related to objects.

The Grad-CAM findings, as delineated in Figure 5, reveal that multiple brain regions facilitated the training of the EEG-Net model in distinguishing between four object categories. For both the object recognition and identification models, all categories exhibited pronounced gradient scores within the occipitotemporal region. This region encompasses channels from the ventral stream. Such a manifestation suggests that the model effectively assimilated pertinent information regarding the objects primarily through channels within the ventral stream, especially the temporal channels. This observation aligns with prior research indicating the significance of temporal channels in processing intricate visual stimuli [44], [45], [46]. Apart from the temporal channels, channels within the parietal region also displayed considerable contributions to the EEGNet model's training. The parietal brain region is widely recognized for its role in discerning an object's spatial attributes [47], [48], [49]. Given that objects were oriented in four distinct directions in our study, the spatial information is intrinsically vital for object identification. Hence, the pronounced significance of the parietal region in the Grad-CAM results of the object identification model aligns with expectations. However, the emergence of parietal significance in the object recognition model was intriguing, particularly since the target object consistently appeared at the centre of the screen. This suggests that the parietal region's significance might be intricately linked to the spatial characteristics of the target object itself.

It is evidenced that the dorsal visual pathway plays an important role in supporting processes within the ventral pathway [50], [51]. However, the specifics of this interrelation remain relatively obscure. Ayzenberg et al. posited a hypothesis wherein the dorsal stream partakes in object recognition by processing spatial relations of the features of the object, subsequently constructing a global shape precept of the object [52]. This synthesized information is then relayed to the ventral pathway, bolstering object recognition processes. Moreover, research by Jeong and Xu [53] proposes that the

dorsal stream recognizes an abstract representation of object identity, exhibiting a behaviorally pertinent role by closely tracking the perceived facial-identity similarity obtained in behavioural tasks. This involvement of dorsal channels in object recognition might very well be echoed in our Grad-CAM results.

The exploration of the ventral and dorsal streams is by no means a novel undertaking, as many studies have endeavoured to discern the relationship between these streams and object-related tasks. These investigations span both anatomically [54], [55], [56] and computational approaches via deep learning algorithms [57], [58], [59]. Within the area of BCI research, there have been extensive endeavours to decode object-related information from the ventral stream using various EEG features such as ERP [60], power spectral density [45], EEG phase patterns [61], and independent components [62]. Our study, inspired by the two-stream hypothesis, trained models using channels reflecting the ventral and dorsal streams and subsequently combined both streams.

Models emulating the ventral stream exhibited enhanced accuracy in both object recognition and object identification tasks when compared to models oriented around the visual and temporal regions. This improvement underscores the intrinsic value of information within the ventral stream, facilitating more efficient model learning. Conversely, models mimicking the dorsal stream didn't achieve the same efficacy as their ventral counterparts, although they outperformed the visual region models. As mentioned earlier, there's mounting evidence advocating the role of the dorsal stream in object recognition, suggesting that this stream encompasses visual data instrumental for the model's categorization capabilities. Given the growing body of literature on the symbiotic relationship between the ventral and dorsal streams, both anatomically and functionally [63], [64], we developed a model integrating channels from both streams. This composite model demonstrated marginally enhanced accuracy relative to the ventral stream model. Nevertheless, this slight enhancement is arguably attributable to the augmented channel data during model training rather than being a significant functional outcome.

V. Conclusion

In this study, we embarked on a thorough examination of several EEGNet models using the data we collected for both object recognition and object identification tasks, each characterized by different channel configurations, to understand their effectiveness in object recognition and identification tasks. The findings reveal that the model trained utilizing the channels from the ventral stream outperforms those trained using regional channels. Notably, its efficacy is marginally surpassed by the model that was trained using all available channels. Furthermore, a modest enhancement in the model's performance was noted when channels from both the ventral and dorsal streams were combined. To delve into the intricacies of this observation, we used the Grad-CAM visualization technique on the trained model. The Grad-CAM result exposed a pronounced gradient score around the channels that form the ventral stream. Furthermore, a significant contribution from the parietal channels toward the EEGNet model's training was evident. This reinforces the prevailing understanding that the brain's dorsal stream is essential in tasks relating to object recognition and identification. Collectively, the results from our investigation underscore that the ventral and dorsal streams contain crucial information that can be harnessed for the efficient training of models on object recognition and identification tasks. This finding holds potential for the development of a rapid and precise BCI system designed for object recognition and identification.

References

[1] M. A. Goodale and A. D. Milner, "Separate visual pathways for perception and action," *Trends Neurosci.*, vol. 15, no. 1, pp. 20–25, Jan. 1992.

[2] R. M. Cichy, D. Pantazis, and A. Oliva, "Similarity-based fusion of MEG and fMRI reveals spatio-temporal dynamics in human cortex during visual object recognition," *Cerebral Cortex*, vol. 26, no. 8, pp. 3563–3579, Aug. 2016.

[3] T. F. Tafreshi, M. R. Daliri, and M. Ghodousi, "Functional and effective connectivity based features of EEG signals for object recognition," *Cognit. Neurodynamics*, vol. 13, no. 6, pp. 555–566, Dec. 2019.

[4] J. Rizkallah, P. Benquet, A. Kabbara, O. Dufor, F. Wendling, and M. Hassan, "Dynamic reshaping of functional brain networks during visual object recognition," *J. Neural Eng.*, vol. 15, no. 5, Oct. 2018, Art. no. 056022.

[5] M. Dyck and M. B. Brodeur, "ERP evidence for the influence of scene context on the recognition of ambiguous and unambiguous objects," *Neuropsychologia*, vol. 72, pp. 43–51, Jun. 2015.

[6] A. Truman and L. Mudrik, "Are incongruent objects harder to identify? The functional significance of the N300 component," *Neuropsychologia*, vol. 117, pp. 222–232, Aug. 2018.

[7] G. Lu and G. Hou, "Effects of semantic congruence on sign identification: An ERP study," *Hum. Factors, J. Hum. Factors Ergonom. Soc.*, vol. 62, no. 5, pp. 800–811, Aug. 2020.

[8] M. I. Coco, A. Nuthmann, and O. Dimigen, "Fixation-related brain potentials during semantic integration of object–scene information," *J. Cognit. Neurosci.*, vol. 32, no. 4, pp. 571–589, Apr. 2020.

[9] N. K. Logothetis and D. L. Sheinberg, "Visual object recognition," *Annu. Rev. Neurosci.*, vol. 19, no. 1, pp. 577–621, Mar. 1996.

[10] A. M. Chan, E. Halgren, K. Marinkovic, and S. S. Cash, "Decoding word and category-specific spatiotemporal representations from MEG and EEG," *NeuroImage*, vol. 54, no. 4, pp. 3028–3039, Feb. 2011.

[11] C. Wang, S. Xiong, X. Hu, L. Yao, and J. Zhang, "Combining features from ERP components in single-trial EEG for discriminating four-category visual objects," *J. Neural Eng.*, vol. 9, no. 5, Oct. 2012, Art. no. 056013.

[12] Y. Qin et al., "Classifying four-category visual objects using multiple ERP components in single-trial ERP," *Cognit. Neurodyn.*, vol. 10, no. 4, pp. 275–285, Aug. 2016.

[13] B. Choi and S. Jo, "A low-cost EEG system-based hybrid brain-computer interface for humanoid robot navigation and recognition," *PLoS ONE*, vol. 8, no. 9, Sep. 2013, Art. no. e74583.

[14] F. Arrichiello, P. Di Lillo, D. Di Vito, G. Antonelli, and S. Chiaverini, "Assistive robot operated via P300-based brain computer interface," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2017, pp. 6032–6037.

[15] K. Kaspar, U. Hassler, U. Martens, N. Trujillo-Barreto, and T. Gruber, "Steady-state visually evoked potential correlates of object recognition," *Brain Res.*, vol. 1343, pp. 112–121, Jul. 2010.

[16] X. Chen, B. Zhao, Y. Wang, and X. Gao, "Combination of high-frequency SSVEP-based BCI and computer vision for controlling a robotic arm," *J. Neural Eng.*, vol. 16, no. 2, Apr. 2019, Art. no. 026012.

[17] S.-K. Chen, C.-S. Chen, Y.-K. Wang, and C.-T. Lin, "An SSVEP stimuli design using real-time camera view with object recognition," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Dec. 2020, pp. 562–567.

[18] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, Oct. 2018, Art. no. 056013.

[19] H. Ahmed, R. B. Wilbur, H. M. Bharadwaj, and J. M. Siskind, "Object classification from randomized EEG trials," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 3844–3853.

[20] R. Shi, Y. Zhao, Z. Cao, C. Liu, Y. Kang, and J. Zhang, "Categorizing objects from MEG signals using EEGNet," *Cognit. Neurodynamics*, vol. 16, no. 2, pp. 365–377, Apr. 2022.

[21] D. Komolovaitė, R. Maskeliūnas, and R. Damaševičius, "Deep convolutional neural network-based visual stimuli classification using electroencephalography signals of healthy and Alzheimer's disease subjects," *Life*, vol. 12, no. 3, p. 374, Mar. 2022.

[22] D. Huang, X. Wang, J. Liu, J. Li, and W. Tang, "Virtual reality safety training using deep EEG-net and physiology data," *Vis. Comput.*, vol. 38, no. 4, pp. 1195–1207, Apr. 2022.

[23] D. Borra, F. Bossi, D. Rivolta, and E. Magosso, "Deep learning applied to EEG source-data reveals both ventral and dorsal visual stream involvement in holistic processing of social stimuli," *Sci. Rep.*, vol. 13, no. 1, p. 7365, May 2023.

[24] M. Simões et al., "BCIAUT-p300: A multi-session and multi-subject benchmark dataset on autism for p300-based brain-computer-interfaces," *Frontiers Neurosci.*, vol. 14, Sep. 2020, Art. no. 568104.

[25] D. Borra and E. Magosso, "Deep learning-based EEG analysis: Investigating p3 ERP components," *J. Integrative Neurosci.*, vol. 20, no. 4, pp. 791–811, 2021.

[26] D. Borra, E. Magosso, M. Castelo-Branco, and M. Simões, "A Bayesian-optimized design for an interpretable convolutional neural network to decode and analyze the P300 response in autism," *J. Neural Eng.*, vol. 19, no. 4, Aug. 2022, Art. no. 046010.

[27] S. Jonas, A. O. Rossetti, M. Oddo, S. Jenni, P. Favaro, and F. Zubler, "EEG-based outcome prediction after cardiac arrest with convolutional neural networks: Performance and visualization of discriminative features," *Hum. Brain Mapping*, vol. 40, no. 16, pp. 4606–4617, Nov. 2019.

[28] Y. Li, H. Yang, J. Li, D. Chen, and M. Du, "EEG-based intention recognition with deep recurrent-convolution neural network: Performance and channel selection by grad-CAM," *Neurocomputing*, vol. 415, pp. 225–233, Nov. 2020.

[29] G. Griffin, A. Holub, and P. Perona, "Caltech-256 object category dataset," California Inst. Technol., Pasadena, CA, USA, Tech. Rep., 2007.

[30] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004.

[31] L. Pion-Tonachini, K. Kreutz-Delgado, and S. Makeig, "ICLabel: An automated electroencephalographic independent component classifier, dataset, and website," *NeuroImage*, vol. 198, pp. 181–197, Sep. 2019.

[32] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUs)," 2015, *arXiv:1511.07289*.

[33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[34] P.-T. de Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Ann. Oper. Res.*, vol. 134, no. 1, pp. 19–67, Feb. 2005.

[35] R. Caruana, S. Lawrence, and C. Giles, "Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 13, 2000, pp. 381–387.

[36] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.

[37] Z. Kourtzi and N. Kanwisher, "Representation of perceived object shape by the human lateral occipital complex," *Science*, vol. 293, no. 5534, pp. 1506–1509, Aug. 2001.

[38] A. Amedi, "Convergence of visual and tactile shape processing in the human lateral occipital complex," *Cerebral Cortex*, vol. 12, no. 11, pp. 1202–1212, Nov. 2002.

[39] A. Martin, "The representation of object concepts in the brain," *Annu. Rev. Psychol.*, vol. 58, no. 1, pp. 25–45, Jan. 2007.

[40] H. Karimi-Rouzbahani, M. Shahmohammadi, E. Vahab, S. Setayeshi, and T. Carlson, "Temporal variabilities provide additional category-related information in object category decoding: A systematic comparison of informative EEG features," *Neural Comput.*, vol. 33, no. 11, pp. 3027–3072, 2021.

[41] A. M. Proverbio and F. Riva, "RP and N400 ERP components reflect semantic violations in visual processing of human actions," *Neurosci. Lett.*, vol. 459, no. 3, pp. 142–146, Aug. 2009.

[42] A. L. Abreu, L. Fernández-Aguilar, F. Ferreira-Santos, and C. Fernandes, "Increased N250 elicited by facial familiarity: An ERP study including the face inversion effect and facial emotion processing," *Neuropsychologia*, vol. 188, Sep. 2023, Art. no. 108623.

[43] A. Enge, F. Süß, and R. Abdel Rahman, "Instant effects of semantic information on visual perception," *J. Neurosci.*, vol. 43, no. 26, pp. 4896–4906, Jun. 2023.

[44] E. L. Mazerolle, R. C. N. D'Arcy, Y. Marchand, and R. B. Bolster, "ERP assessment of functional status in the temporal lobe: Examining spatiotemporal correlates of object recognition," *Int. J. Psychophysiol.*, vol. 66, no. 1, pp. 81–92, Oct. 2007.

[45] I. D. Rus, P. Marc, M. Dinsoreanu, R. Potolea, and R. C. Muresan, "Classification of EEG signals in an object recognition task," in *Proc. 13th IEEE Int. Conf. Intell. Comput. Commun. Process. (ICCP)*, Sep. 2017, pp. 391–395.

[46] A. Das et al., "Evaluating interhemispheric connectivity during midline object recognition using EEG," *PLoS ONE*, vol. 17, no. 8, Aug. 2022, Art. no. e0270949.

[47] P. R. Brotchie, R. A. Andersen, L. H. Snyder, and S. J. Goodman, "Head position signals used by parietal neurons to encode locations of visual stimuli," *Nature*, vol. 375, no. 6528, pp. 232–235, May 1995.

[48] M. T. de Schotten et al., "Direct evidence for a parietal-frontal pathway subserving spatial awareness in humans," *Science*, vol. 309, no. 5744, pp. 2226–2228, Sep. 2005.

[49] Y. Wamain, F. Gabrielli, and Y. Coello, "EEG μ rhythm in virtual reality reveals that motor coding of visual objects in peripersonal space is task dependent," *Cortex*, vol. 74, pp. 20–30, Jan. 2016.

[50] E. Freud, J. C. Culham, D. C. Plaut, and M. Behrmann, "The large-scale organization of shape processing in the ventral and dorsal pathways," *eLife*, vol. 6, Oct. 2017, Art. no. e27576.

[51] V. Ayzenberg, C. Simmons, and M. Behrmann, "Temporal asymmetries and interactions between dorsal and ventral visual pathways during object recognition," *Cerebral Cortex Commun.*, vol. 4, no. 1, Jan. 2023, Art. no. tgad003.

[52] V. Ayzenberg and M. Behrmann, "The dorsal visual pathway represents object-centered spatial relations for object recognition," *J. Vis.*, vol. 22, no. 14, p. 3233, Dec. 2022.

[53] S. K. Jeong and Y. Xu, "Behaviorally relevant abstract object identity representation in the human parietal cortex," *J. Neurosci.*, vol. 36, no. 5, pp. 1607–1619, Feb. 2016.

[54] J. Wang et al., "Relationship between ventral stream for object vision and dorsal stream for spatial vision: An fMRI+ERP study," *Hum. Brain Mapping*, vol. 8, no. 4, pp. 170–181, 1999.

[55] H.-O. Karnath, J. Rüter, A. Mandler, and M. Himmelbach, "The anatomy of object recognition—Visual form agnosia caused by medial occipitotemporal stroke," *J. Neurosci.*, vol. 29, no. 18, pp. 5854–5862, May 2009.

[56] P. Janssen, B.-E. Verhoef, and E. Premereur, "Functional interactions between the macaque dorsal and ventral visual pathways during three-dimensional object vision," *Cortex*, vol. 98, pp. 218–227, 2018.

[57] U. Guclu and M. A. J. van Gerven, "Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream," *J. Neurosci.*, vol. 35, no. 27, pp. 10005–10014, Jul. 2015.

[58] R. M. Cichy, A. Khosla, D. Pantazis, A. Torralba, and A. Oliva, "Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence," *Sci. Rep.*, vol. 6, no. 1, 2016, Art. no. 27755.

[59] C. Zhuang et al., "Unsupervised neural network models of the ventral visual stream," *Proc. Nat. Acad. Sci. USA*, vol. 118, no. 3, Jan. 2021, Art. no. e2014196118.

[60] G. M. Doniger et al., "Activation timecourse of ventral visual stream object-recognition areas: High density electrical mapping of perceptual closure processes," *J. Cognit. Neurosci.*, vol. 12, no. 4, pp. 615–621, Jul. 2000.

[61] M. Behroozi, M. R. Daliri, and B. Shekarchi, "EEG phase patterns reflect the representation of semantic categories of objects," *Med. Biol. Eng. Comput.*, vol. 54, no. 1, pp. 205–221, Jan. 2016.

[62] A. X. Stewart, A. Nuthmann, and G. Sanguinetti, "Single-trial classification of EEG in a visual object task using ICA and machine learning," *J. Neurosci. Methods*, vol. 228, pp. 1–14, May 2014.

[63] D. J. Kravitz, K. S. Saleem, C. I. Baker, L. G. Ungerleider, and M. Mishkin, "The ventral visual pathway: An expanded neural framework for the processing of object quality," *Trends Cognit. Sci.*, vol. 17, no. 1, pp. 26–49, Jan. 2013.

[64] H. Takemura, A. Rokem, J. Winawer, J. D. Yeatman, B. A. Wandell, and F. Pestilli, "A major human white matter pathway between dorsal and ventral visual cortex," *Cerebral Cortex*, vol. 26, no. 5, pp. 2205–2214, May 2016.