Accepted Manuscript

Accepted Manuscript (Uncorrected Proof)

Title: Encoding neural activity During Visual Processing: A CNN-Based Model for ECoG Signal Reconstruction Using Convolutional and Regression Networks

Authors: Mohammadamin Lotfi[1], Farimah Soltani[1], Fatemeh Zareayan Jahromy[1,*]

1. *Biomedical Engineering Department, School of Electrical Engineering Iran University of Science and Technology (IUST), Tehran, Iran.*

***Corresponding Author**: Fatemeh Zareayan Jahromy. Biomedical Engineering Department, School of Electrical Engineering Iran University of Science and Technology (IUST), Tehran, Iran. Email: fzareayan@iust.ac.ir

To appear in: **Basic and Clinical Neuroscience**

Received date: 2025/09/28

Revised date: 2026/05/19

Accepted date: 2026/05/31

Please cite this article as:

Lotfi, M., Soltani, F., Zareayan Jahromy, F. (In Press). Encoding neural activity During Visual Processing: A CNN-Based Model for ECoG Signal Reconstruction Using Convolutional and Regression Networks. Basic and Clinical Neuroscience. Just Accepted publication Jul. 10, 2026. Doi: http://dx.doi.org/10.32598/bcn.2026.5461.1

DOI: http://dx.doi.org/10.32598/bcn.2026.5461.1

Abstract

Background. Advances in processing technology have enabled researchers to explore brain mechanisms in greater depth, utilizing methodologies such as deep learning to simulate the electrical activity of the brain with high resolution.

New method. In this study we designed an end-to-end convolutional neural networks (CNNs) model which directly reconstructs electrocorticography signals from image inputs. Image features were extracted specifically for data reconstruction. Additionally, a classification-regressor model was introduced, where a CNN was first trained to classify images into five conceptual categories, and then its extracted features were used with a regressor to reconstruct the brain signals.

Result. Both models were found to be capable of reconstructing ECoG data in the occipital region, which has an important role in the processing of visual information. Furthermore, as the distance from this area increased, the reconstruction accuracy decreased. Another noteworthy finding was that features relevant to the classification of conceptual categories of visual stimuli had more information about the signal, resulting in a greater performance enhancement for regressors (Correlation=0.56, $p<0.01$) relative to an end-to-end (Correlation=0.46, $p<0.05$) learning paradigm.

Comparison with existing methods. This paper demonstrated the importance of the feature extraction objective, showing that correctly choosing the model's goal is crucial for enhancing signal reconstruction accuracy.

Conclusions. CNN-based models can effectively simulate the brain network's behavior in generating outputs from input stimulus. When we use image classification to get features, it is better for reconstructing neural signals than when we learn things end to end.

Keyword: convolutional neural networks, electrocorticography, regressor, visual processing, brain simulation

1. Introduction

One can state that visual processing is one of the main parts of human cognition, a relevant example of an intricate cross-interaction of neural mechanisms that has fascinated neuroscientists, cognitive psychologists, and engineers for many decades. It enables the brain to parse complex visual scenes, recognize objects, and make sense of our visual world, with its exceptional computational power [1]. Whereas this is far from being a process confined to one particular area of the brain, it rather implies a network of regions implicated in efforts to decode and form meaning from visual stimuli by making responses[2]. First in line for processing in the case of vision is the occipital lobe, housing the primary visual cortex, V1. Recent investigation, however, has illuminated the much wider network involved: regions in the temporal, parietal, and even frontal lobes. This distributed processing underscores the brain's efficiency in dealing with the complexity of visual information, thereby leading to specialized areas that process it. As demonstrated in Figure 1, the visual data pipeline within the brain is illustrated, providing a comprehensive overview of visual data processing. The figure also illustrates the stages of visual data projection from the retina to the cortex. The course of visual information through the brain is well-defined, each contributing uniquely to our visual experience[3]. The ventral stream, often called the "what" pathway, snakes through the temporal lobe and provides object recognition and form representation. Running in parallel, the dorsal stream or the "where" pathway courses through the parietal lobe and is involved in spatial processing and motion detection. Although separate, these pathways do not exist in isolation; rather, they interact and integrate information to build our rich visual percepts. Probably one of the most important quests in neuroscience has been to understand how different brain regions along these pathways encode various features of visual stimuli[4]. This quest has been immensely helped by the development of neuroimaging techniques. Therein, ECoG has emerged as one of the strongest tools, offering high spatial and temporal resolution by placing electrode arrays directly on the cortical surface [5]. It enables the investigator to record electrical activity from a specific brain region with less signal attenuation, thereby providing unparalleled insight into the neural correlates of visual processing. Such neuroimaging studies have provided insights that go far beyond our knowledge of brain function; they have inspired the development of computational models that attempt to emulate the brain's strange and marvelous processing of vision. Perhaps best known is the so-called HMAX model, an abbreviation for Hierarchical Model, with X standing for open-ended, which simulates hierarchical processing in the visual cortex[6]. Its original form and various derivatives have performed exceptionally well in object recognition tasks, a good example of the potential benefit of neurobiological principles to an artificial vision system. But the HMAX model is hardly an exception in neuroscience; it contributes to AI. Computer vision stands to gain enormously from a set of brain-inspired architectures. Convolutional Neural Networks (CNNs) were inspired by the hierarchical organization of the visual cortex and have revolutionized image recognition and related tasks [7]. These involve a hierarchy of layers, each successively extracting more complex features of the visual input, much like the brain builds up representations from simple edges and contours to complex object shapes and scene layouts. While successful deep learning models, including CNNs, have provided an important foundation in this field, they are certainly not the only approach to understanding and replicating how visual processing works. Other useful techniques include representational similarity analysis, multivariate pattern analysis, and Bayesian decoding, which have contributed significantly to our understanding of how the brain represents visual information [8]. Such diversity underlines a set of complementary insights that help create a broad picture of visual cognition. Indeed, another paper introduced a model that approximated the brain's response to images across different semantic categories. First, images were filtered by their model using Gabor wavelets; these were then used as inputs to a neural

network to predict the intracranial field potential area. Then, the model's accuracy was evaluated using both Pearson's correlation and normalized root-mean-square error. This model surely is good at predicting brain activity. Another K-nearest neighbor classifier decoded brain signals, revealing semantic categories with significantly better-than-chance accuracy[9]. Maybe the most exciting frontier involves the reconstruction of visual experiences from brain activity. Besides promising even more profound insights into neural representations, this line of research offers a host of potential applications, including brain-computer interfaces and medical diagnostics. Pioneering work by [10] and [11] proved that it is possible to reconstruct static and dynamic visual experiences using fMRI data. These early successes opened the door to even more sophisticated approaches driven by deep learning. In recent years, there has been tremendous improvement in this area. One such breakthrough was achieved by [12] CNNs. Reconstructed visual stimuli recorded from the EEG using a deep learning approach by[13]. These and many other works have shown great potential in combining neuroimaging data with advanced machine learning techniques to understand neural representations of visual experiences and to decode and reconstruct them. The field has been growing rapidly, with many research investigations examining different neural mechanisms underlying imaging and various computational models to yield increasingly accurate, high-resolution reconstructions. For example, [14] introduced a new two-stage framework, "Brain-Diffuser," which builds diffusion models to reconstruct natural scenes from fMRI signals. Here is a classic case of using state-of-the-art generative models in neuroscience research. Among other breakthrough studies enabled by work on brain activity, [15] reconstructed both viewed and imagined images. This research is also important for showing not only the feasibility of "mind-reading" technology but also the similarities and differences between the brain's perception and imagination. These kinds of advances really push the boundary of what is possible in decoding and reconstructing neural representations of visual experiences. Furthermore, combining large-scale neuroimaging datasets with high-order machine learning methods, including generative models such as Stable Diffusion, promises to expand our understanding of the inner workings of the human brain. The implications of such progress in neuroscience, artificial intelligence, and even brain-computer interfaces themselves are monumental [14]. This could be translated into more intuitive, responsive brain-computer interfaces, enabling improved decoding of visual experience, for instance, enabling revolutionary assistive technologies for people with impaired vision or motor function. The knowledge obtained on how the brain processes visual information is guiding the development of computer vision systems that are increasingly efficient and robust. By understanding how the brain achieves such remarkable efficiency and generalization in visual tasks, it will be possible to design artificial systems that are far more adaptable, energy-efficient, and capable of learning from limited data, a number of key open challenges in current AI systems [15]. The study of visual processing also informs key issues in cognitive development and neuroplasticity. If one investigates how the visual system adapts to different experiences and learns to process complex scenes, one can uncover general principles of learning and adaptation in the brain [16]. Looking ahead, visual neuroscience and computational modeling stand at an exciting threshold. Converging advances in neuroimaging techniques, sophisticated computational models, and strong machine learning algorithms are opening new avenues toward understanding the language of the brain. We are rapidly approaching a comprehensive understanding of precisely how the brain processes, encodes, and reconstructs visual information a quest that promises to unlock new frontiers in science, technology, and medicine. The more one learns of the enigma that is visual cognition, the greater one's understanding not only of the basic operation of the brain but also of the potentially transformative applications it may reach, extending human capabilities, enhancing medical diagnostics, and improving artificial intelligence. And it's not as if this journey of discovery into the visual neurosciences is at an end; every new discovery brings us

closer to decoding that subtle language the brain ventures, promising a future in which the boundaries between mind and machine will blur.

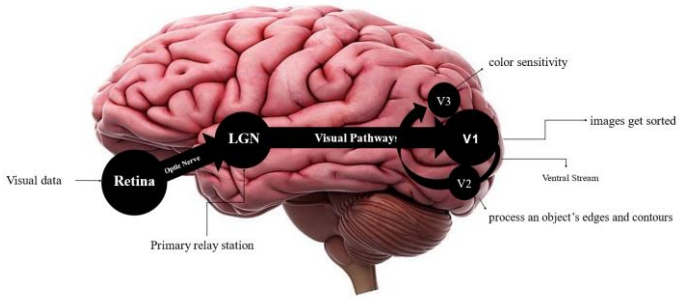

Figure 1 – Schematic illustration of the hierarchical processing of visual information in the human brain, demonstrating the pathway from retinal input through early visual cortices to higher-order cortical regions involved in visual perception and cognitive processing.

2. Dataset

The dataset consists of cortical recordings from human participants performing a one-back visual object recognition task, in which images of objects were presented at varying sizes, viewpoints, and locations[17]. Recordings were obtained from a total of 15 subjects; however, four subjects were excluded from analysis. Subjects 3 and 9 were excluded because the occipital region was not covered, which is central to this study's aims. Subjects 1 and 12 were excluded because they had the lowest signal-to-noise ratios (SNR) in the cohort (Figure 2). The final analysis, therefore, included 11 subjects. Each subject had between 48 and 126 intracranial recording sites (mean ± SD: 80.4 ± 18.4). Raw signals were amplified 32,500× and bandpass filtered between 0.1 and 100 Hz. Sampling rate varied by recording facility: 256 Hz (CHB, XLTEK) and 500 Hz (BWH, Bio-Logic). For consistency across subjects, all signals were resampled to a common rate prior to model input. ECoG signals were epoched relative to stimulus onset. Each epoch was baseline-corrected by subtracting the mean pre-stimulus activity (−200 to 0 ms). Channels with excessive noise or artifact (defined by variance exceeding 3 standard deviations from the channel mean across trials) were excluded. Remaining channels were normalized (z-scored) per channel across trials. Visual stimuli (images) were resized to a uniform resolution of 224×224 pixels and normalized using ImageNet mean and standard deviation values prior to input into the CNN. The dataset was partitioned as follows: 80% of trials were used for training, 10% for validation (used for hyperparameter tuning and early stopping), and 10% for held-out testing. Partitioning was performed at the trial level to prevent data leakage.

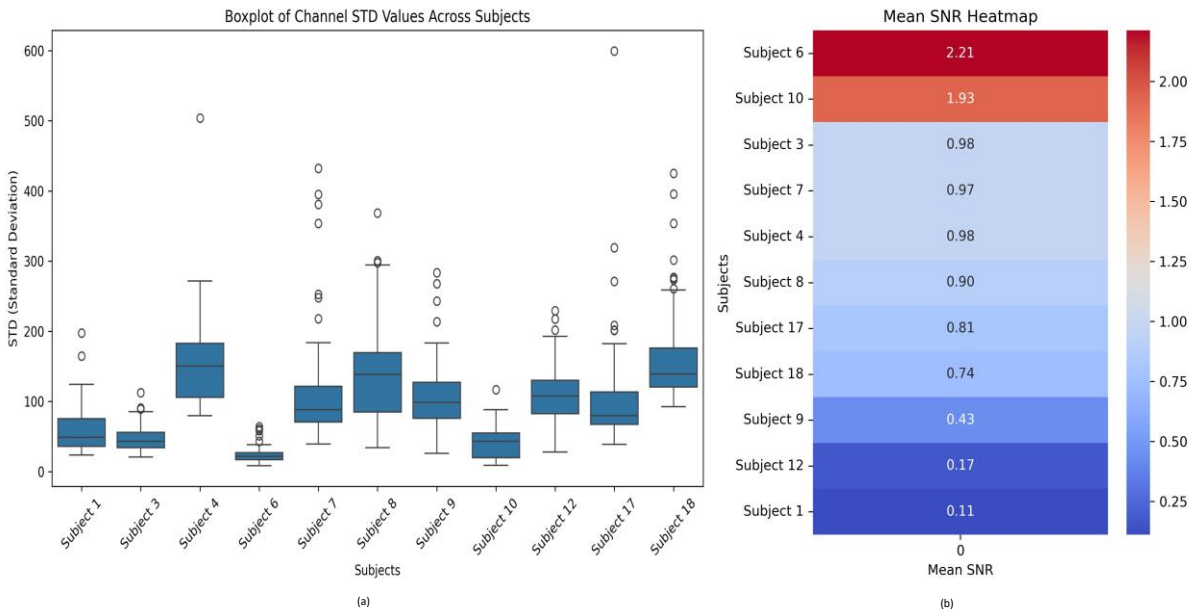

Figure 2 – (a) Boxplot illustrating the variability and distribution of neural response amplitudes across subjects based on the recorded ECoG signals. (b) Heatmap representation of the ranked signal-to-noise ratio (SNR) across subjects and channels, highlighting differences in recording quality and signal reliability used for subsequent analyses.

3. Deep Learning Model

The objective of this study was to investigate whether visually evoked cortical activity could be predicted from image stimuli using deep learning–based computational models. To achieve this goal, a framework was developed to reconstruct electrocorticography (ECoG) signals from visual inputs by combining convolutional neural networks (CNNs) with nonlinear regression networks. The proposed framework was designed to learn mappings between visual representations extracted from images and the corresponding neural responses recorded from distributed cortical regions. The overall architecture consisted of two major components: (1) a CNN-based feature extraction module responsible for learning hierarchical representations of visual stimuli, and (2) a regression network designed to estimate multi-channel ECoG activity from the extracted image features. Two different learning strategies were investigated in this study: an end-to-end reconstruction framework and a classification–regression framework. The primary difference between these approaches was the way visual features were learned and transferred to the regression stage. CNN architectures were selected for their strong performance in image representation learning and for their biological inspiration from hierarchical visual processing in the cortex. Through successive convolutional operations, CNNs progressively transform low-level image information such as edges, textures, and contours into increasingly abstract semantic representations. These representations are particularly well-suited for modeling visually evoked neural activity because they preserve structural and category-related information relevant to object recognition. The regression module consisted of a multilayer fully connected neural network that received visual feature vectors as input and predicted the corresponding ECoG activity recorded from intracranial electrodes. Since electrode coverage varied across subjects, the regression

network's output dimensionality was adjusted for each participant based on the number of valid recording channels. Channels exhibiting excessive noise, unstable activity, or recording artifacts were excluded during preprocessing, and the associated output nodes were removed from the model accordingly. All models were implemented using the PyTorch deep learning framework and trained using the Adam optimizer with an initial learning rate of 10^{-4}. Training was performed using mini-batches of 32 samples. Early stopping based on validation loss was applied to reduce overfitting, with a patience window of 10 epochs. The maximum number of training epochs was set to 100. Hyperparameters, including the learning rate, dropout probability, hidden layer dimensions, and loss weighting parameters, were selected empirically based on validation performance.

3.1 End To End

In the end-to-end approach, the CNN feature extractor and the regression network were jointly optimized within a unified architecture. In this framework, the model directly learned a mapping from raw image stimuli to corresponding ECoG activity, without an intermediate classification objective. The purpose of this strategy was to enable the CNN to learn visual representations specifically optimized for neural signal reconstruction rather than for conventional image recognition tasks. The architecture of the proposed end-to-end framework is illustrated in Figure 3. The model establishes a direct computational relationship between visual inputs and cortical neural responses by integrating hierarchical feature extraction with nonlinear signal regression.

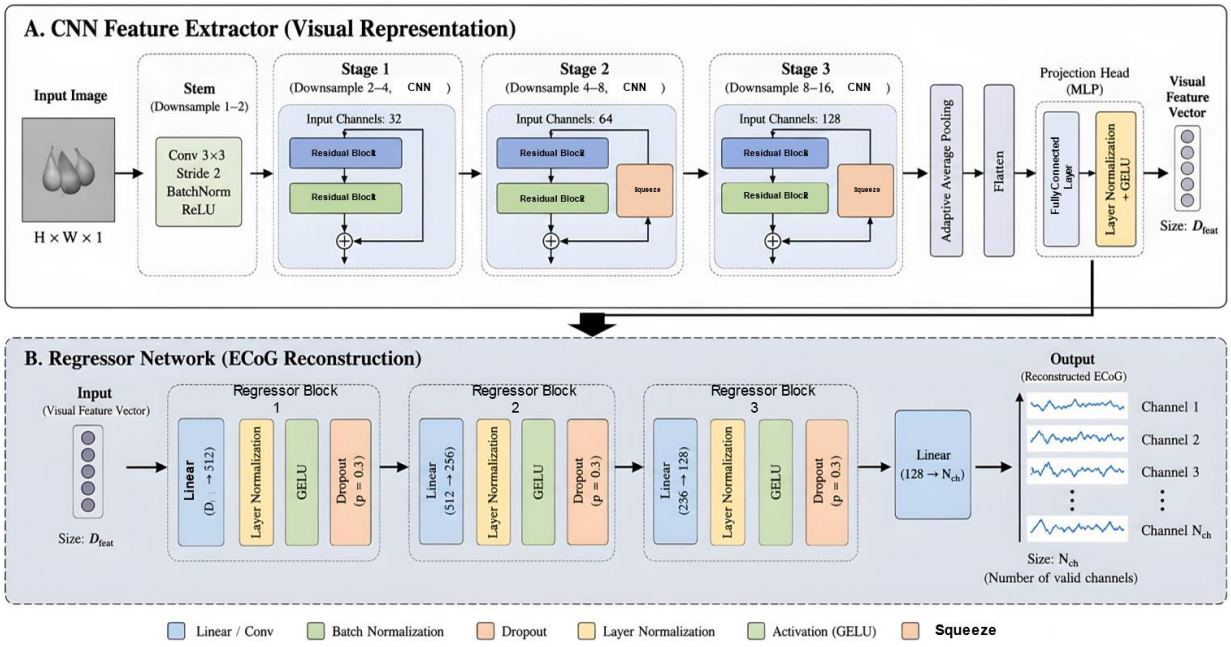

Figure 3 – Architecture of the proposed end-to-end reconstruction framework. Input images are processed through a CNN composed of convolution, normalization, activation, and pooling layers to extract hierarchical visual representations. The resulting feature vectors are then provided to a fully connected regression network that predicts multi-channel ECoG activity. The entire model is jointly optimized using a composite loss function that combines mean-squared reconstruction error with a correlation-based similarity loss.

Image Feature Extraction (CNN Backbone): The visual stimuli are processed by a CNN architecture composed of three convolutional layers with batch normalization and pooling layers. This CNN is trained to extract relevant high-dimensional feature representations from images, which are then used for ECoG reconstruction. Each convolutional layer applies filters to the input images to capture hierarchical patterns and spatial features, followed by pooling operations to reduce the dimensionality while preserving critical information. The final output of the CNN is a feature vector:

$$\boldsymbol{f}_{img} = CNN(I)$$

Where I is the input image and $fimg$ is the feature vector, which captures essential aspects of the visual input. This feature vector is then passed to the regressor for ECoG signal reconstruction.

Regressor for ECoG Signal Reconstruction: The second part of the model consists of a regressor that maps the visual features extracted by the CNN to ECoG signals. The regression module consisted of multiple fully connected layers with hidden dimensions of 512, 256, and 128 neurons, respectively. GELU nonlinear activation functions were used between layers to improve gradient flow and nonlinear representation learning. To improve generalization and reduce overfitting, layer normalization and dropout (p=0.3) were applied after each hidden layer. The regressor predicts the corresponding ECoG signal $\hat{\boldsymbol{E}}$ based on the input visual features $fimg$:

$$\hat{\boldsymbol{E}} = \textbf{Regressor}(\boldsymbol{f}_{img})$$

Here, $\hat{\boldsymbol{E}}$ represents the reconstructed ECoG signal, which is expected to resemble the real ECoG signals recorded from the brain regions of interest. To optimize the model's performance, we designed a composite loss function that integrates two components: mean squared error (MSE) and correlation loss. The rationale behind this design is to balance the reconstruction accuracy (MSE) and the alignment of the predicted and true signals in terms of their correlation.

Mean Squared Error (MSE) Loss: This loss function penalizes the difference in magnitude between the reconstructed signals $\hat{\boldsymbol{E}}$ and the true signals $\boldsymbol{E}$

$$L_{\text{MSE}} = \frac{1}{N}\sum_{i=1}^{N}(\hat{\mathbf{E}}_i - \mathbf{E}_i)^2$$

where N is the number of samples, $\hat{\boldsymbol{E}}_{\boldsymbol{i}}$ is the reconstructed signal, and $\boldsymbol{E}_{\boldsymbol{i}}$ is the ground truth signal.

Correlation Loss: Because neural signals contain important temporal and structural characteristics that may not be fully captured by MSE alone, a correlation-based loss term was additionally incorporated. This component encourages the reconstructed signals to preserve waveform similarity and temporal dynamics:

$$L_{\text{corr}} = 1 - \frac{\text{Cov}(\hat{\mathbf{E}}, \mathbf{E})}{\sigma_{\hat{\mathbf{E}}}\sigma_{\mathbf{E}} + \epsilon}$$

where Cov represents the covariance between $\widehat{\boldsymbol{E}}_{\boldsymbol{i}}$ and $\boldsymbol{E}_{\boldsymbol{i}}$, σ denotes the standard deviations, and ϵ is a small value added to avoid division by zero. The final loss function is a weighted sum of the two components:

The final training objective combined both loss terms:

$$\mathrm{L} = \lambda \mathrm{L}_{\mathrm{MSE}} + \mathrm{L}_{\mathrm{corr}}$$

where λdetermines the relative contribution of reconstruction accuracy and correlation preservation. The weighting parameter was empirically set to $\lambda = 0.7$based on validation experiments. This configuration provided the most stable reconstruction performance and achieved the best balance between minimizing signal error and preserving the temporal structure of neural activity. The combined optimization strategy improved the model's ability to reconstruct ECoG signals that closely resembled recorded cortical activity in both amplitude and temporal dynamics.

3.2 Classification-Regressor

In the classification regression framework, the learning process was divided into two sequential stages in order to encourage the extraction of visually meaningful and semantically informative image representations before neural signal reconstruction. Unlike the end-to-end approach, where the CNN was optimized directly for ECoG prediction, this framework first trained the convolutional network as a supervised image classifier and subsequently reused the learned latent representations for the regression of cortical activity. The motivation for this strategy was the assumption that feature representations optimized for visual object recognition may also contain high-level semantic and structural information relevant to neural processing in visual cortical regions. The overall architecture of the proposed classification–regression framework is illustrated in Figure 4. In the first stage, the CNN was trained to categorize visual stimuli into predefined image classes using supervised learning. The network architecture consisted of multiple convolutional blocks followed by fully connected layers and a final softmax classification layer. Similar to the end-to-end framework, each convolutional block included convolutional operations, batch normalization, GELU activation, and max pooling. Through this hierarchical processing pipeline, the CNN progressively transformed raw image inputs into increasingly abstract feature representations that capture edges, textures, shapes, object structures, and category-specific visual patterns. The feature extraction process can be formulated as:

$$f_{img} = \mathrm{CNN}(I)$$

where I denotes the input image and f_{img} represents the latent visual feature representation learned by the convolutional backbone. These feature vectors encode the most informative visual characteristics required for object discrimination and semantic categorization. Following feature extraction, the latent

representations were passed to a fully connected classification layer that estimated the probability distribution over image categories using a softmax activation function:

$$\hat{y} = \text{Softmax}(W f_{img} + b)$$

where $\hat{y}$represents the predicted class probability distribution, while $W$and $b$denote the trainable weights and biases of the final classification layer, respectively. The classifier was optimized using the categorical cross-entropy loss function:

$$L_{\text{CE}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log\left(\hat{y}_{i,c}\right)$$

where $N$is the number of training samples, Cdenotes the total number of image classes, $y_{i,c}$is the ground-truth label for class c, and $\hat{y}_{i,c}$represents the predicted probability corresponding to that class. Minimizing this loss function encouraged the CNN to learn discriminative visual representations that could separate different semantic categories of visual stimuli. After completing the classification stage, the final softmax classification layer was removed, leaving the learned CNN backbone intact. The extracted latent feature vectors from the penultimate layer were then reused as input to the regression network responsible for reconstructing ECoG activity. In this second stage, the learned visual representations served as biologically informative embeddings that connected image content to cortical neural responses. The regression stage can be expressed as:

$$\hat{E} = \text{Regressor}(f_{img})$$

where $\hat{E}$denotes the reconstructed ECoG signal and f_{img}represents the classification-derived feature representation extracted from the CNN. The regression network architecture was identical to that used in the end-to-end framework and consisted of multiple fully connected layers with GELU activations, layer normalization, and dropout regularization. The rationale for this two-stage framework was that object classification forces the CNN to learn compact, semantically meaningful representations of visual stimuli before attempting neural reconstruction. Since human visual cortical regions are strongly involved in hierarchical object recognition and category-specific processing, features optimized for image classification may better approximate the representational structure underlying visually evoked cortical activity. Consequently, the regression network receives more informative feature embeddings than representations learned solely from direct signal reconstruction. This approach also provided an additional regularization effect during training. By constraining the CNN to first solve a structured visual recognition task, the extracted representations became more stable and less susceptible to overfitting to subject-specific neural noise. Experimental results demonstrated that the classification–regression framework consistently achieved higher reconstruction performance than the end-to-end approach across multiple visual cortical

regions, particularly in occipital and occipitotemporal areas associated with object perception and high-level visual processing.

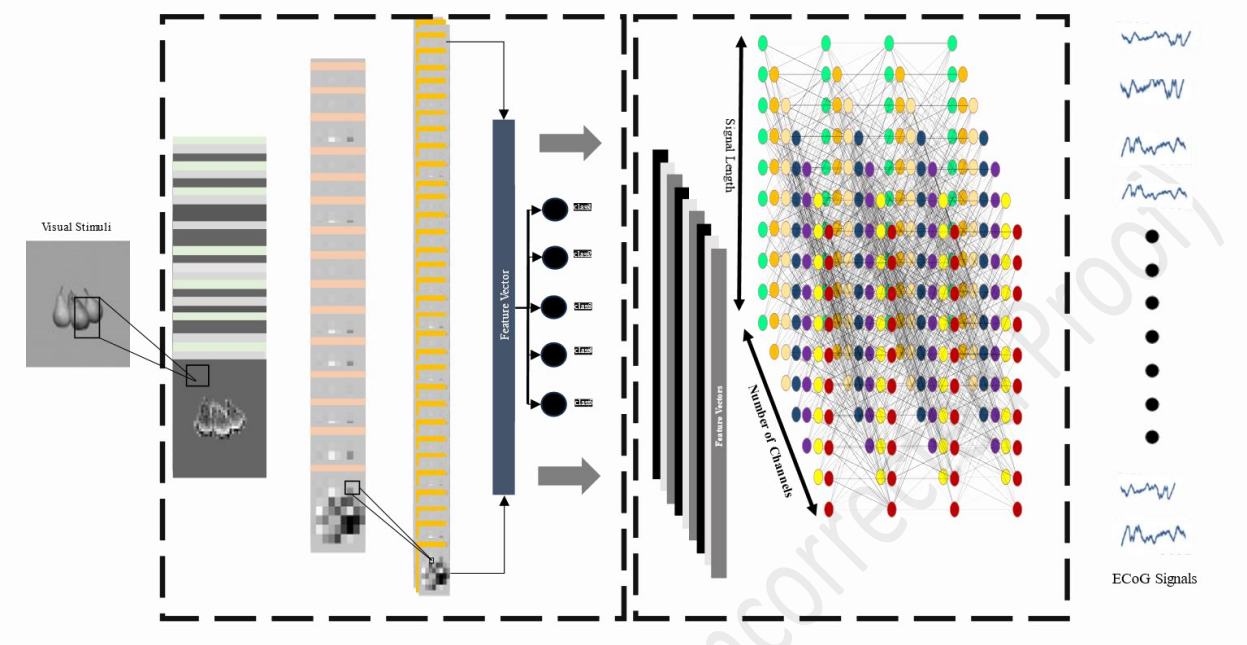

Figure 4 - Block diagram of the classification–regression framework illustrating the two-stage training procedure, including CNN-based image classification followed by ECoG signal reconstruction using extracted visual feature representations.

4. Results

To quantitatively evaluate reconstruction performance, we primarily used the Pearson correlation coefficient between reconstructed and recorded ECoG signals. Correlation analysis was selected as the primary evaluation metric because the proposed framework's primary objective was to preserve the temporal structure and dynamic behavior of neural activity rather than exact point-wise amplitude matching. In neural recordings such as ECoG, signal amplitudes may vary substantially across subjects, recording sessions, and electrode locations due to physiological variability and acquisition-related factors. Consequently, correlation-based evaluation provides a robust measure of whether reconstructed signals preserve the underlying temporal dynamics and waveform structure of cortical activity. In addition to correlation analysis, mean squared error (MSE) and explained variance (EV) metrics were also calculated to provide a complementary quantitative assessment of reconstruction quality. While MSE evaluates the average magnitude of reconstruction error, explained variance measures how much of the variability in the original neural signals can be captured by the model predictions. Together, these metrics provide a broader characterization of reconstruction performance across cortical regions. For each subject, reconstruction performance was calculated independently across all trials and recording channels. The average correlation, MSE, and explained variance were subsequently computed for each cortical region by aggregating across all valid subjects and electrodes associated with that anatomical area. Because electrode coverage differed

between participants, the number of valid channels contributing to each region varied across subjects. Regions with insufficient or unstable recordings were excluded from the final analysis.

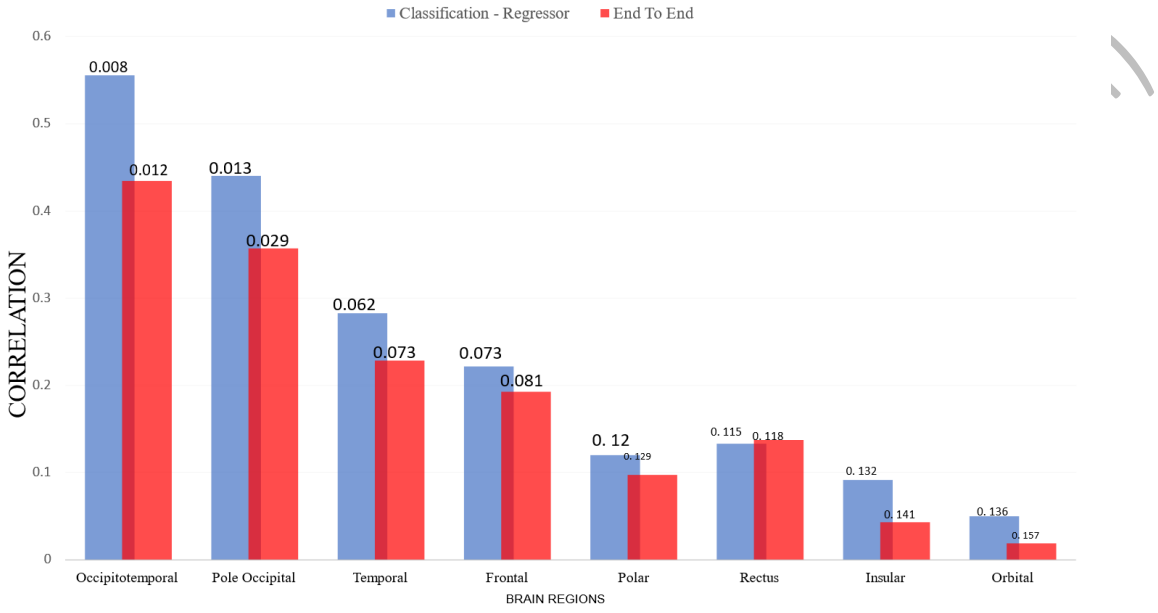

Figure 5 – Mean reconstruction performance across cortical regions for both the classification–regression and end-to-end frameworks. Bars represent the average Pearson correlation between reconstructed and recorded ECoG signals across all subjects. Corresponding statistical significance values are shown above each bar.

As illustrated in Figure 5, reconstruction performance varied substantially across cortical regions, with visual and occipitotemporal areas consistently demonstrating the strongest results in both reconstruction frameworks. The figure further shows that the classification–regression approach achieved higher correlation values, lower reconstruction error, and higher explained variance across most cortical regions compared with the end-to-end architecture. This overall trend suggests that classification-derived visual representations provide more informative features for predicting visually evoked neural activity. Among all analyzed regions, the occipitotemporal medial parahippocampal gyrus (G_oc-temp_med-Parahip) exhibited the highest reconstruction performance. As shown in Figure 5, the classification–regression framework achieved an average correlation of 0.5621 ($p = 0.0029$), while the end-to-end framework achieved a correlation of 0.4624 ($p = 0.019$). In addition to these correlation results, this region also demonstrated relatively low MSE values and high explained variance, indicating that the reconstructed signals closely matched both the temporal structure and variability of the recorded neural activity. The parahippocampal region is known to participate in visual scene representation, spatial perception, and memory-related visual integration, which may explain its strong responsiveness to image-based stimuli. The lateral occipitotemporal fusiform gyrus also demonstrated strong reconstruction performance in both frameworks, with correlations of 0.5367 ($p = 0.0061$) for the classification–regression approach and

0.4022 ($p = 0.0208$) for the end-to-end framework. As shown in Figure 5, this region also exhibited lower reconstruction error and higher explained variance than many frontal and temporal regions. Since the fusiform gyrus is strongly associated with high-level visual processing, including object and face recognition, these findings support the biological relevance of the proposed framework in modeling visually evoked cortical activity. The occipital pole likewise demonstrated comparatively strong reconstruction performance, with correlation values of 0.3191 ($p = 0.0492$) and 0.2241 ($p = 0.0935$) for the classification–regression and end-to-end approaches, respectively. Figure 5 further indicates that the classification–regression model achieved lower MSE and higher explained variance in this region than the end-to-end framework. Given the occipital cortex's role in early-stage visual processing, stronger reconstruction performance in this region was expected for image-driven neural responses. The posterior transverse collateral sulcus (S_collat_transv_post), located within the temporal lobe, also exhibited relatively high reconstruction performance. As shown in Figure 5, this region achieved one of the highest correlation values (0.5267; p = 0.0031), while simultaneously demonstrating favorable MSE and explained variance. This region is associated with higher-order visual integration and memory-related processing, both of which are likely to contribute to stable responses during visual stimulus presentation. In contrast, frontal cortical regions generally demonstrated weaker reconstruction performance. Regions such as G_front_inf-Triangul and G_front_middle exhibited lower correlation values, higher MSEs, and lower explained variance than occipital regions. As shown in Figure 5, the reconstruction results in these areas were less stable and often failed to achieve statistical significance. These findings likely reflect the functional specialization of frontal regions for executive control, language-related processing, and decision-making, which are less directly constrained by passive visual stimuli [18]. Similarly, orbitofrontal regions exhibited weak and unstable reconstruction performance across subjects. Figure 5 demonstrates that these regions were characterized by low correlation coefficients, elevated reconstruction error, and poor explained variance. Because orbitofrontal activity is more strongly linked to emotional evaluation, reward processing, and cognitive control, its neural dynamics may not be adequately predicted using visual information alone. Temporal regions not primarily specialized for visual processing, including G_temporal_middle and G_temporal_inf, showed moderate and variable reconstruction performance. As shown in Figure 5, these regions achieved intermediate correlation values, along with moderate explained variance and reconstruction error. The variability observed across temporal regions may reflect their involvement in auditory, semantic, or multimodal processing functions that are only indirectly associated with the visual stimuli used in this study.

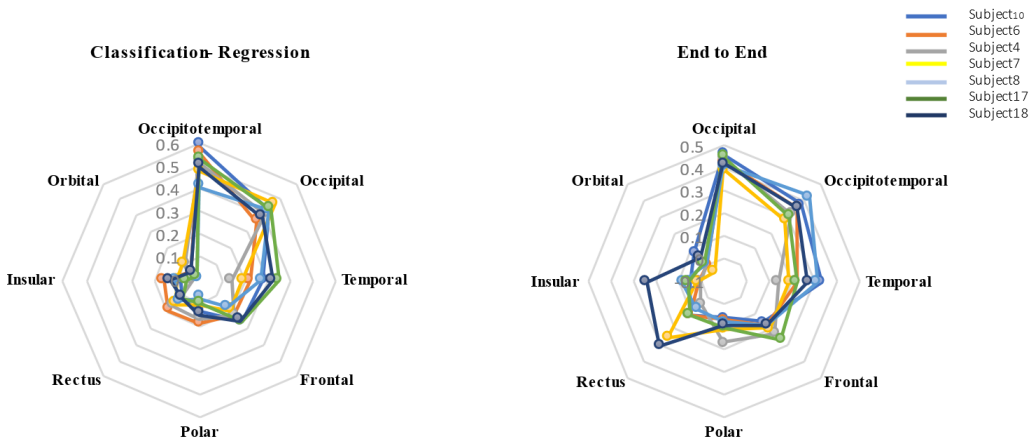

Figure 6 – Subject-wise reconstruction performance across cortical regions for both modeling approaches. Mean Pearson correlation values are shown for all subjects and anatomical regions.

Figure 6 illustrates reconstruction performance separately for each subject across all cortical regions. The figure indicates that regions with statistically significant reconstruction performance, particularly in occipital and occipitotemporal areas, showed relatively consistent behavior across subjects. In contrast, regions with low or non-significant correlations showed considerably greater inter-subject variability and lacked a stable regional pattern. This variability was also reflected in the corresponding MSE and explained variance values, where non-visual regions generally exhibited less stable reconstruction quality across participants. As shown in Figure 6, a limited number of isolated, high-correlation observations were detected in frontal regions, such as the rectus gyrus, for subjects 7 and 18. However, these findings were not consistently reproduced across the remaining participants and were therefore interpreted cautiously as potential outliers rather than reliable regional effects. The rectus gyrus has previously been associated with emotional and decision-related processing, which are unlikely to be strongly engaged during the present visual reconstruction task. Overall, Figures 5 and 6 collectively demonstrate that reconstruction performance was strongest and most stable in cortical regions associated with visual perception and visual-semantic integration. Across nearly all evaluated regions, the classification–regression framework consistently achieved higher correlation values, lower MSE, and higher explained variance than the end-to-end architecture. These findings suggest that semantically informed visual feature representations learned through supervised classification provide more robust embeddings for reconstructing visually evoked cortical activity than representations learned exclusively through direct signal regression [19], [20].

5. Discussion

The present study investigated whether visually evoked ECoG activity could be reconstructed from image stimuli using deep learning–based computational frameworks. Two distinct modeling strategies were evaluated: an end-to-end reconstruction architecture and a classification–regression framework that leveraged visual representations learned from supervised image classification to support neural reconstruction. Overall, the results demonstrated that cortical regions associated with visual perception exhibited the strongest reconstruction performance, supporting the biological plausibility of the proposed framework and highlighting the relationship between hierarchical visual representations and visually evoked neural activity. As demonstrated in Figures 5 and 6, occipital and occipitotemporal regions consistently achieved the highest reconstruction performance across subjects and evaluation metrics. In particular, the occipitotemporal medial parahippocampal gyrus and fusiform regions exhibited the strongest correlations, together with lower reconstruction error and higher explained variance. These findings are consistent with the established role of occipitotemporal pathways in object recognition and high-level visual processing [1] [4]. Previous neuroscience studies have shown that the ventral visual pathway is strongly involved in extracting semantic and structural information from visual stimuli, which likely explains the improved reconstruction performance observed in these cortical regions [1] [3] [4]. Furthermore, the fusiform and parahippocampal regions are known to contribute to visual scene analysis, object categorization, and memory-associated visual integration, making them highly responsive to image-based inputs [2]. The regional reconstruction patterns shown in Figure 7 further supports these observations. The heatmap illustrates that cortical regions associated with primary and higher-order visual processing exhibited the most stable and robust reconstruction performance across subjects. In particular, occipital and occipitotemporal regions showed consistently elevated correlation and explained variance, whereas frontal regions showed lower reconstruction quality and greater variability. These findings suggest that the proposed framework was more successful at capturing neural dynamics directly associated with visual information processing than at capturing activity related to abstract cognitive or executive functions.

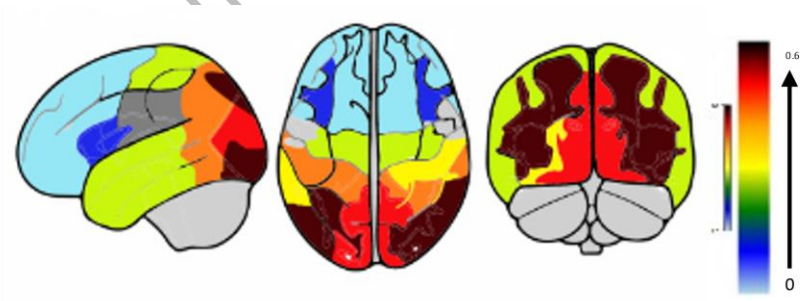

Figure 7 - Cortical heatmap illustrating regional reconstruction performance across the brain. Warmer colors indicate stronger reconstruction quality based on the average correlation between reconstructed and recorded ECoG signals across subjects. Regions associated with visual processing, particularly occipital and occipitotemporal cortices, demonstrated the highest reconstruction performance, whereas frontal and associative regions exhibited comparatively weaker results.

An important finding of the present study was the consistent superiority of the classification–regression framework over the end-to-end architecture. Across nearly all evaluated cortical regions, the classification–regression model achieved higher correlation values, lower mean squared reconstruction error, and higher explained variance. This suggests that visual feature representations learned through supervised classification provide more informative embeddings for neural reconstruction than representations learned solely through direct signal regression. One possible explanation for this improvement is that supervised classification encourages the CNN backbone to learn semantically meaningful visual representations that better approximate the hierarchical organization of biological visual processing. Hierarchical computational models and deep neural networks have previously been proposed as biologically inspired frameworks capable of approximating cortical visual representations [6] [7]. Prior studies have demonstrated that deep convolutional representations can capture important properties of neural coding and visual object recognition within the ventral visual stream [1] [6]. Consequently, transferring these learned visual embeddings to the regression stage may improve the model's ability to predict visually evoked cortical responses. Our findings are also consistent with previous work in neural decoding and brain representation analysis. Earlier studies demonstrated that neural activity patterns contain sufficient information to reconstruct or decode visual experiences from brain signals [8] [14]. For example, Naselaris et al. reconstructed natural images from human brain activity using Bayesian encoding approaches [10] while Nishimoto et al. demonstrated reconstruction of visual experiences from neural activity elicited by natural movies [11]. More recent deep learning–based approaches have further highlighted the ability of modern neural networks to model complex relationships between visual stimuli and neural responses [12] [14]. Although most previous studies primarily focused on fMRI or EEG modalities, the present work extends these concepts to ECoG-based signal reconstruction using CNN-driven feature representations. Importantly, the proposed framework should not be interpreted as a direct simulation of biological neural pathways or complete brain network dynamics. Rather, the model provides a computational approximation of statistical relationships between visual stimuli and recorded neural activity. While the reconstruction results demonstrate meaningful correspondence between image-derived representations and ECoG responses, neural processing in the human brain involves substantially more complex recurrent, multimodal, and context-dependent mechanisms that cannot be fully captured by the present framework. Therefore, the findings should be interpreted cautiously as evidence of predictive modeling rather than direct simulation of neural mechanisms. This distinction is particularly important when discussing neuroscience-inspired artificial intelligence systems and biologically motivated computational models [15] [21]. Figure 8 further illustrates the frequency-domain analysis performed on reconstructed ECoG signals under different visual frequency conditions. The results demonstrated relatively stable reconstruction performance across frequency reconstruction scenarios, although slightly improved reconstruction accuracy was observed for lower-frequency signal components. This observation may reflect the greater stability and lower noise characteristics of lower-frequency cortical activity compared with higher-frequency neural oscillations. Interestingly, the occipitotemporal medial parahippocampal region exhibited a distinct frequency-related pattern relative to other cortical areas, suggesting possible regional differences in frequency sensitivity during visual processing.

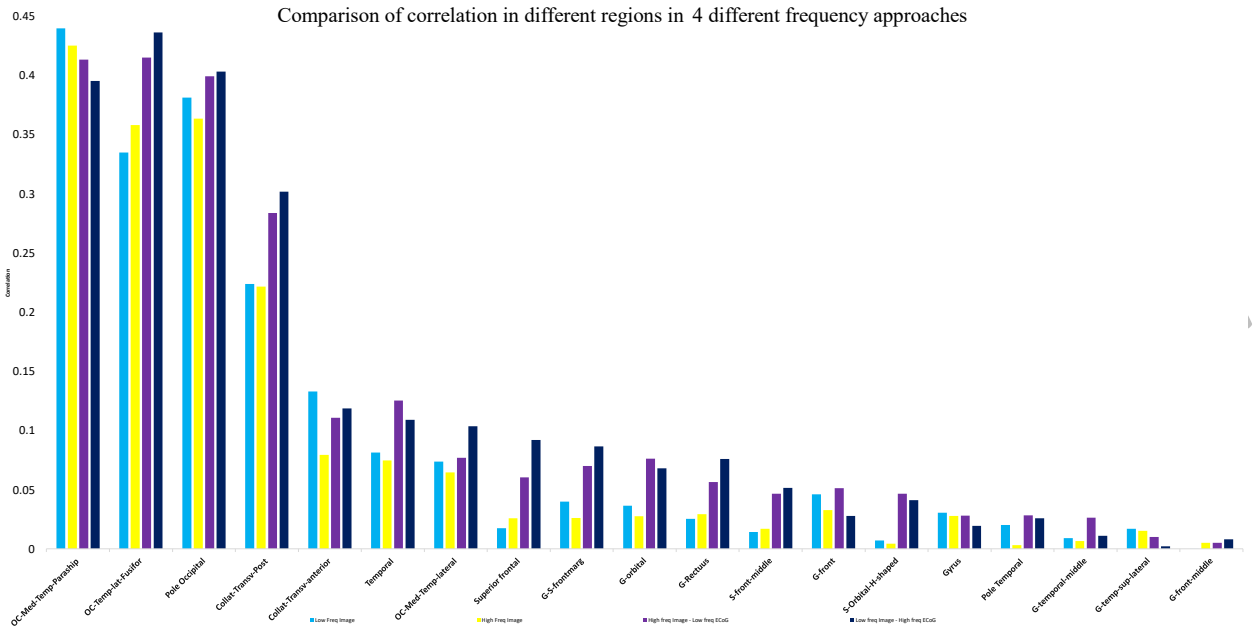

Figure 8 – Frequency-domain analysis of reconstructed ECoG signals under different image frequency conditions. The figure compares reconstruction performance across four experimental settings involving low-frequency and high-frequency image inputs and corresponding neural signal reconstruction targets. Results indicate relatively stable reconstruction performance across frequency conditions, with slightly improved performance observed for lower-frequency neural activity components.

The present findings may also have implications for future brain–computer interface (BCI) and neural prosthetic research. In particular, computational models capable of predicting cortical activity from visual inputs could contribute to the development of adaptive visual prosthetic systems and neural decoding technologies. Previous research has shown that neuroplasticity enables the brain to gradually adapt to artificial sensory feedback provided by prosthetic devices[22], [23], [25]. Over time, repeated exposure to artificial visual input may allow cortical networks to reorganize and improve the interpretation of externally generated neural stimulation patterns [24]. These adaptive processes are considered fundamental for successful sensory rehabilitation and neural prosthesis integration. However, the current study should be viewed as a preliminary computational modeling framework rather than a clinically deployable neural prosthetic system. Several limitations remain. First, the dataset size and electrode coverage were limited due to the inherent constraints of intracranial recordings. Second, the framework focused exclusively on static image stimuli and did not investigate temporally dynamic visual sequences or naturalistic video inputs. Third, although correlation and explained-variance analyses demonstrated meaningful reconstruction performance in visual regions, the overall reconstruction accuracy remains insufficient for practical neural-decoding applications. Future work could improve the generalizability and robustness of the framework by incorporating larger datasets, multimodal neural recordings, and temporally dynamic visual stimuli such as videos. In addition, more advanced architectures incorporating attention mechanisms, recurrent processing, or transformer-based visual representations may better capture the temporal and hierarchical dynamics of cortical activity. Further investigation of subject-independent modeling and cross-subject generalization may also improve the applicability of computational neural reconstruction frameworks in neuroscience and brain–computer interface research.

6. Conclusion

This study presented a CNN-based framework for reconstructing visually evoked ECoG signals from image stimuli using two different modeling strategies: an end-to-end architecture and a classification–regression framework. The results demonstrated that cortical regions associated with visual perception, particularly occipital and occipitotemporal areas, achieved the strongest reconstruction performance, supporting the biological relevance of the proposed approach. Among the two methods, the classification–regression framework consistently produced higher correlations, lower reconstruction error, and greater explained variance compared with the end-to-end model. These findings suggest that visual representations learned through supervised image classification provide informative features for predicting neural activity related to visual processing. The study also highlighted the variability of reconstruction performance across brain regions. Regions involved in higher-order cognitive and executive functions showed lower reconstruction accuracy, indicating that their activity is less directly determined by visual input alone. This underscores the complexity of large-scale neural processing and the challenges of modeling distributed cortical dynamics. Although the proposed framework remains a computational approximation rather than a direct simulation of neural mechanisms, the findings demonstrate the potential of deep learning approaches for studying relationships between sensory stimuli and cortical activity. Future work may improve model generalizability by incorporating larger datasets, dynamic visual stimuli such as videos, and more advanced neural architectures for modeling temporal brain activity.

Data Availability

The datasets that were analyzed during the present study are available to the public [https://klab.tch.harvard.edu/resources/liuetal_timing3.html], ensuring accessibility for further research. By following the provided link, one can access the datasheet, information regarding the data recording sessions, and supplementary files.

References

[1] J. J. DiCarlo, D. Zoccolan, and N. C. Rust, "How does the brain solve visual object recognition?," *Neuron*, vol. 73, no. 3, pp. 415–434, Feb. 2012, doi: 10.1016/J.NEURON.2012.01.010.

[2] K. Grill-Spector and R. Malach, "The human visual cortex," *Annu. Rev. Neurosci.*, vol. 27, pp. 649–677, 2004, doi: 10.1146/ANNUREV.NEURO.27.070203.144220.

[3] L. G. Ungerleider and J. V. Haxby, "'What' and 'where' in the human brain," *Curr. Opin. Neurobiol.*, vol. 4, no. 2, pp. 157–165, Jan. 1994, doi: 10.1016/0959-4388(94)90066-3.

[4] M. A. Goodale and A. D. Milner, "Separate visual pathways for perception and action," *Trends Neurosci.*, vol. 15, no. 1, pp. 20–25, 1992, doi: 10.1016/0166-2236(92)90344-8.

[5] E. M. Zion Golumbic *et al.*, "Mechanisms underlying selective neuronal tracking of attended speech at a 'cocktail party,'" *Neuron*, vol. 77, no. 5, pp. 980–991, 2013, doi: 10.1016/J.NEURON.2012.12.037.

[6] M. Riesenhuber and T. Poggio, "Hierarchical models of object recognition in cortex," *Nature Neuroscience 1999 2:11*, vol. 2, no. 11, pp. 1019–1025, Nov. 1999, doi: 10.1038/14819.

[7] N. Kriegeskorte, "Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing," *Annu. Rev. Vis. Sci.*, vol. 1, no. 1, pp. 417–446, Nov. 2015, doi: 10.1146/ANNUREV-VISION-082114-035447.

[8] J. V. Haxby, A. C. Connolly, and J. S. Guntupalli, "Decoding neural representational spaces using multivariate pattern analysis," *Annu. Rev. Neurosci.*, vol. 37, pp. 435–456, 2014, doi: 10.1146/ANNUREV-NEURO-062012-170325.

[9] F. Z. Jahromy and M. R. Daliri, "Semantic category-based decoding of human brain activity using a Gabor-based model by estimating intracranial field potential range in temporal cortex," *J. Integr. Neurosci.*, vol. 16, no. 4, pp. 419–428, Jan. 2017, doi: 10.3233/JIN-170028.

[10] T. Naselaris, R. J. Prenger, K. N. Kay, M. Oliver, and J. L. Gallant, "Bayesian reconstruction of natural images from human brain activity," *Neuron*, vol. 63, no. 6, pp. 902–915, Sep. 2009, doi: 10.1016/J.NEURON.2009.09.006.

[11] S. Nishimoto, A. T. Vu, T. Naselaris, Y. Benjamini, B. Yu, and J. L. Gallant, "Reconstructing visual experiences from brain activity evoked by natural movies," *Current Biology*, vol. 21, no. 19, p. 1641, Oct. 2011, doi: 10.1016/J.CUB.2011.08.031.

[12] Z. Rakhimberdina, Q. Jodelet, X. Liu, and T. Murata, "Natural Image Reconstruction From fMRI Using Deep Learning: A Survey," *Front. Neurosci.*, vol. 15, Dec. 2021, doi: 10.3389/FNINS.2021.795488/FULL.

[13] M. Ferrante, T. Boccato, S. Bargione, and N. Toschi, "Decoding visual brain representations from electroencephalography through knowledge distillation and latent diffusion models," *Comput. Biol. Med.*, vol. 178, p. 108701, Aug. 2024, doi: 10.1016/J.COMPBIOMED.2024.108701.

[14] G. Shen, T. Horikawa, K. Majima, and Y. Kamitani, "Deep image reconstruction from human brain activity," *PLoS Comput. Biol.*, vol. 15, no. 1, p. e1006633, 2019, doi: 10.1371/JOURNAL.PCBI.1006633.

[15] D. Hassabis, D. Kumaran, C. Summerfield, and M. Botvinick, "Neuroscience-Inspired Artificial Intelligence," *Neuron*, vol. 95, no. 2, pp. 245–258, Jul. 2017, doi: 10.1016/J.NEURON.2017.06.011.

[16] B. Kolb and I. Q. Whishaw, "Brain plasticity and behavior," *Annu. Rev. Psychol.*, vol. 49, pp. 43–64, 1998, doi: 10.1146/ANNUREV.PSYCH.49.1.43.

[17] H. Liu, Y. Agam, J. R. Madsen, and G. Kreiman, "Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex," *Neuron*, vol. 62, no. 2, pp. 281–290, Apr. 2009, doi: 10.1016/J.NEURON.2009.02.025.

[18] M. Lovstad *et al.*, "Executive functions after orbital or lateral prefrontal lesions: neuropsychological profiles and self-reported executive functions in everyday living," *Brain Inj.*, vol. 26, no. 13–14, pp. 1586–1598, 2012, doi: 10.3109/02699052.2012.698787.

[19] M. Ismail *et al.*, "Rectus gyrus hematoma: An overview," *Surg. Neurol. Int.*, vol. 13, no. 558, 2022, doi: 10.25259/SNI_1023_2022.

[20] W. Li, W. Lou, W. Zhang, R. K. Y. Tong, R. Jin, and W. Peng, "Gyrus rectus asymmetry predicts trait alexithymia, cognitive empathy, and social function in neurotypical adults," *Cerebral Cortex*, vol. 33, no. 5, pp. 1941–1954, Feb. 2023, doi: 10.1093/CERCOR/BHAC184.

[21] V. Koren, G. Bondanelli, and S. Panzeri, "Computational methods to study information processing in neural circuits," *Comput. Struct. Biotechnol. J.*, vol. 21, p. 910, Jan. 2023, doi: 10.1016/J.CSBJ.2023.01.009.

[22] D. R. Chebat, B. Heimler, S. Hofsetter, and A. Amedi, "The Implications of Brain Plasticity and Task Selectivity for Visual Rehabilitation of Blind and Visually Impaired Individuals," *Contemporary Clinical Neuroscience*, pp. 295–321, 2018, doi: 10.1007/978-3-319-78926-2_13.

[23] D. Caravaca-Rodriguez, S. P. Gaytan, G. J. Suaning, and A. Barriga-Rivera, "Implications of Neural Plasticity in Retinal Prosthesis," *Invest. Ophthalmol. Vis. Sci.*, vol. 63, no. 11, pp. 11–11, Oct. 2022, doi: 10.1167/IOVS.63.11.11.

[24] M. Beyeler, A. Rokem, G. M. Boynton, and I. Fine, "Learning to see again: biological constraints on cortical plasticity and the implications for sight restoration technologies," *J. Neural Eng.*, vol. 14, no. 5, p. 051003, Aug. 2017, doi: 10.1088/1741-2552/AA795E.

[25] S. Preißler, C. Dietrich, K. Blume, G. O. Hofmann, W. H. R. M. Miltner, and T. Weiss, "Plasticity in the visual system is associated with prosthesis use in phantom limb pain," *Front. Hum. Neurosci.*, vol. 7, no. JUN, p. 51665, Jun. 2013, doi: 10.3389/FNHUM.2013.00311/BIBTEX.