Dual-Stream EEG Decoding for 3D Visual Perception

Ninon Lizé Masclef NINON@MIT.EDU
Taisija Demcenko TAISIJA@MIT.EDU
Antonella Catanzaro ANTOCAT@MIT.EDU
Nataliya Kosmyna NKOSMYNA@MIT.EDU
Massachusetts Institute of Technology

Editors: List of editors' names

Abstract

This paper explores a novel brain decoding model for 3D shape perception through a dual pathway architecture mirroring biological vision. Our bio-inspired approach implements separate decoding modules for object identity and spatial orientation, inspired by ventral and dorsal pathways, during continuous rotations. We employ circular regression for angle prediction and develop EEG-conditioned multiview diffusion for 3D reconstruction. Our approach successfully decodes both object identity and spatial orientation from EEG signals and enables 3D reconstruction from neural activity, with interpretability analyses revealing temporally structured involvement of ventral, dorsal, and motor-related channels rather than a static ventral dominance in supporting object and angle decoding.

1. Introduction

Neural networks for visual perception have drawn fundamental inspiration from biological vision systems, particularly their hierarchical organization (Hubel and Wiesel, 1962; Fukushima, 1980; Lehky and Sejnowski, 1988). While convolutional neural networks show correspondence with the ventral visual pathway — the brain's "what" system for object recognition (DiCarlo et al., 2012; Cichy et al., 2016) — they struggle with 3D spatial and geometric understanding (Hinton et al., 2018; Bronstein et al., 2017). This limitation is particularly evident in 3D tasks: standard CNNs are biased toward texture over shape (Geirhos et al., 2018), and humans significantly outperform CNNs in cross-viewpoint 3D shape matching (O'Connell et al., 2023). A similar challenge arises in brain decoding, where models struggle to differentiate viewpoints within object classes and extract orientation features from EEG signals (Li et al., 2024).

We address this limitation with a dual-stream decoding architecture inspired by ventral and dorsal visual pathways, implementing separate modules for object identity and spatial orientation. Our approach reconstructs rotating 3D objects from EEG by explicitly modeling viewpoint-tolerant object recognition and viewpoint-dependent spatial transformations, decoding object identity and angular positioning to provide novel insight into geometric representation in biological and artificial systems. Our results demonstrate that EEG signals preserve both object identity (up to 68% accuracy across six semantic categories) and spatial orientation (10-11° mean absolute error) during 3D perception, with interpretability analysis revealing time-resolved integration across ventral, dorsal, and motor-related pathways, consistent with evidence for dynamic dorsal–ventral interactions and visuo-motor coupling in 3D object perception.

2. Related Works

2.1. 3D Shape Perception

Human 3D object perception involves hierarchical processing in specialized areas, beginning in early visual areas (V1-V3) which extract basic 2D features such as edges, orientation, and contrast (DiCarlo et al., 2012). 3D perception differs from 2D perception by its complexity integrating spatial, depth and structural information, engaging stereo cues such as binocular disparity, leveraging the slight differences in images received by each eye to provide robust depth information (Julesz, 1971). Current evidence indicates this information then diverges into two distinct but interconnected pathways: the ventral stream ("what" pathway) projects from V1 through V2 and V4 to the inferior temporal cortex, and is crucial for object recognition, form representation, and high-level semantic processing (Mishkin et al., 1983). In parallel, the dorsal stream (initially "where", later "how" pathway (Goodale and Milner, 1992)) projects from V1 through V2 and regions like MT/V5 to the posterior parietal cortex. This pathway specializes in spatial relationships, motion, and action-relevant information, and critically transforms visual input into egocentric coordinate systems necessary for motor planning and execution (Gallivan and Culham, 2015; Goodale and Westwood, 2004). Despite their distinct specializations, dorsal and ventral streams interact extensively in attention (Chica et al., 2013) and visual mental imagery (Spagna et al., 2023). Neurons in dorsal regions can respond selectively to 3D contours and surfaces (Theys et al., 2015), and neuroimaging and lesion work shows substantial cross-talk, with object and spatial information represented in both streams (Kravitz et al., 2011; Vaziri-Pashkam and Xu, 2017; Bartolomeo, 2022). This motivates models that treat ventral-dorsal processing as interacting components rather than a strict dichotomy.

2.2. Bio-inspired 3D Brain Decoding

Lehky and Sejnowski (1988) demonstrated that neural networks trained to extract 3D shape from shading develop cortex-like receptive fields, suggesting that the computational demands of 3D vision may drive similar representational solutions across artificial and biological neural networks. This observation has motivated extensive research into bio-inspired architectures for visual processing (Lehky and Sejnowski, 1988). Convolutional neural networks (CNNs) trained on large naturalistic image collections show correspondence with primate ventral visual stream activity patterns (Güçlü and Gerven, 2015; Yamins et al., 2014), with deeper layers responding to increasingly complex features mirroring the V1-to-IT progression. Performance-optimized hierarchical models accurately predict neural responses in IT and V4 without explicit neural constraints, suggesting that categorization objectives alone may drive brain-like representations (Khaligh-Razavi and Kriegeskorte, 2014). However, divergences emerge when examining transformation tolerance mechanisms. While CNNs match human accuracy on object categorization, they differ from biological systems in how they handle spatial transformations (Xu and Vaziri-Pashkam, 2022). Human visual areas show progressively more transformation-tolerant and consistent representations from V1 to higher regions, whereas CNNs often lose representational consistency at deeper layers, suggesting a reliance on 'brute-force' memorization rather than geometric structure. This algorithmic divergence matters for brain decoding: human-aligned encoders that better

match cortical representational geometry can improve EEG/MEG decoding performance (Rajabi et al., 2025).

Effective bio-inspired architectures require explicit integration of geometric and perceptual principles, particularly for dorsal-ventral pathway modeling where spatial processing and transformation tolerance are computational requirements. Neural networks for 3D processing span several paradigms: volumetric CNNs (Maturana and Scherer, 2015), geometric deep learning on non-Euclidean domains (Bronstein et al., 2017), attention-based point-cloud models such as Point Transformer (Zhao et al., 2021), and implicit neural representations like NeRF, Occupancy Networks, and DeepSDF (Mildenhall et al., 2020; Mescheder et al., 2019; Park et al., 2019). Recent approaches incorporate explicit geometric symmetries, for example capsules (Sabour et al.; Hinton et al., 2018), SO(3)-equivariant networks (Thomas et al., 2018), and steerable CNNs (Cohen and Welling, 2016; Weiler et al.). Understanding how such inductive biases translate to brain decoding requires examining the geometric structure preserved in neural signals.

2.3. 3D Geometric Inductive Bias

Human 3D shape perception demonstrates the brain's intrinsic geometric processing capabilities through mechanisms that achieve multiview consistency while revealing systematic inductive biases. The visual system's foundation lies in its topographic organization, e.g. retinotopic maps in early visual areas (V1-V3) maintain precise spatial correspondence between functional and cytoarchitectonic borders, creating an inherent geometric framework for processing orientation, disparity, and spatial features (Tsao et al., 2008; Tsutsui et al., 2001). This spatial organization enables hierarchical geometric processing and depth cue combination, where low-level shape features are decoded within 60 ms (Foxe and Simpson, 2002), building toward complex 3D representations through integration of binocular disparity, perspective, shading, and motion information in higher areas like LOC and hMT+/V5 (Lehky and Sejnowski, 1988; Beeck et al., 2008). This processing is so fundamental that we automatically perceive 2D sketches as 3D objects, unable to suppress our geometric interpretation even when consciously aware of the flat medium (Torralba et al., 2024).

However, human 3D perception is not perfectly viewpoint-invariant, as object recognition speed correlates with rotation angle from canonical poses (Shepard and Metzler, 1971; Palmer et al., 1981), revealing innate inductive biases encoded in domain-specific connectivity patterns that facilitate rapid object recognition while predisposing the brain to extract stable 3D representations from ambiguous inputs (Welchman et al., 2005). These biases reflect equivariant neural representations where transformations in visual input produce corresponding transformations in neural activity patterns, enabling consistent object recognition while maintaining sensitivity to spatial relationships crucial for action and navigation.

These intrinsic geometric processing capabilities are reflected in EEG signals, which capture the synchronized activity of geometrically aligned pyramidal cells across visual cortex. Despite EEG's limited spatial resolution for detailed reconstruction (Halliday and Michael, 1970), its high temporal resolution enables decoding of visual features through visual-evoked potential components, such as N1, P2, and N2 components, specifically involved in processing 3D shape and depth, showing hemispheric dissociation during the N1/N2 complex for 3D shape (left hemisphere) and depth (right hemisphere) (Kasai and Morotomi, 2001). These

components also provide insights into low-level visual features including visual field position (Halliday and Michael, 1970; Jeffreys and Axford, 1972). The geometric sensitivity of EEG extends to complex spatial cognition: Kashihara and Nakahara (2011) demonstrated that EEG effectively captures neural signatures during both 2D and 3D mental imagery tasks, suggesting preserved information for distinguishing geometrical shapes across dimension. This spatial sensitivity follows the brain's hierarchical organization: ablation experiments in visual brain decoding confirmed that occipital regions achieve highest accuracy and reconstruction performance compared to the temporal, parietal and frontal areas, directly reflecting the visual processing hierarchy (Li et al., 2024).

Recent advances demonstrate the feasibility of geometric decoding from brainwave activity: primitive 3D shapes (Esfahani and Sundararajan, 2012), colored geometric symbols (Bang et al., 2021), EEG-based 3D object reconstruction as colored point cloud (Guo et al., 2024), 3D reconstruction from fMRI (Gao et al., 2024), and EEG-driven 3D object generation using latent diffusion models (Xiang et al., 2024). However, current EEG-based 3D decoding approaches treat perception as a single pathway problem, failing to leverage the geometric structure underlying biological vision. From a geometric perspective, 3D shape understanding requires disentangling invariant object properties from equivariant spatial transformations; this decomposition is embodied in the distinct but interacting functional roles of the ventral "what" and dorsal "where/how" pathways. We develop a dual-stream architecture that explicitly models this geometric decomposition, enabling principled 3D reconstruction from EEG through separate object identity classification and angular transformation regression.

3. Methods

To validate this bio-inspired architecture, we decompose 3D visual decoding into three complementary tasks that aim to mirror, as closely as possible, proposed biological processing stages: object classification (ventral stream), spatial orientation regression (dorsal stream), and their integration for 3D reconstruction.

3.1. Dataset and Experimental Design

Eleven (11) participants were recruited for this study. Neural activity was recorded using a 64-channel ANT Neuro wet electrode system positioned according to the international 10-20 system at 512 Hz sampling frequency, with impedances maintained below 20 kΩ. With 11 subjects, the main group-level effect we report corresponds to a very large effect size (Cohen's $d \approx 2.8$), indicating that the study is highly powered at $\alpha = 0.05$. Participants viewed rotating 3D object stimuli in an immersive virtual reality environment rendered using Unity and presented via Meta Quest 2 head-mounted display. The VR paradigm was specifically chosen to leverage binocular disparity cues and enable natural depth perception mechanisms unavailable in conventional 2D presentation modalities. The stimulus set comprised 78 distinct 3D objects uniformly distributed across six semantic categories (13 examples per category): organic natural objects (banana, strawberry), manufactured objects (basketball), and biological entities (human face, panda, tiger). Each object underwent continuous 360° rotation over an 8-second presentation window with angular velocity $\frac{d\theta}{dt} = \frac{\pi}{4}$ rad/s, followed by a 2 s inter-stimulus interval to mitigate potential neural adaptation effects. Example of

stimuli is displayed in Figure 1. To control for temporal confounds and sequence-dependent neural responses, stimulus presentation order was fully randomized across participants and recording sessions. Each participant completed one ∼65-minute session with 78 objects × 6 repetitions (468 trials per subject, 5,148 trials total). This experimental design enables us to separately analyze object identity across rotations and spatial transformations from EEG, providing well-defined labels for both geometric invariance (object identity) and equivariance (rotation angle).

3.2. Model Architectures and Training

We benchmarked three state-of-the-art architectures: EEGNet (Lawhern et al., 2018), a compact CNN for visual evoked potential decoding; EEGConformer (Song et al., 2023), a hybrid CNN-transformer for long-range temporal dependencies; and CTNet (Zhao et al., 2024), a convolutional transformer combining spatial feature extraction with multi-head attention. These architectures offer complementary strengths for processing spatiotemporal EEG features across extended time windows. Hyperparameters were optimized per subject using Bayesian optimization (Snoek et al., 2012). For EEGNet, we increased the depthwise kernel length to 256 samples to better capture visual event-related potentials (Luck, 2014). Model performance was assessed via mean accuracy across 5-fold cross-validation.

3.2.1. Circular Regression Architecture

To enable rotation angle estimation, EEGNet was adapted for circular regression by replacing the classification head with a 2-dimensional output layer predicting unit vectors $[\sin(\theta), \cos(\theta)]$. This geometric representation naturally handles periodic boundary conditions and avoids discontinuities near $\pm\pi$, ensuring angles differing by 2π are treated as identical. The circular regression loss computes squared chord length between predicted and target unit vectors:

$$L_{\text{circular}} = \frac{1}{B} \sum_{i=1}^{B} \left[1 - \cos(\hat{\theta}_i - \theta_i) \right] \tag{1}$$

where B is batch size, equivalent to mean squared error in circular space. Classification and angle-regression models were instantiated as separate EEGNet networks with no weight sharing across tasks.

3.2.2. EEG-Conditioned 3D Reconstruction

We developed a dual-stream 3D reconstruction pipeline that integrates object classification and angle regression to condition a multiview diffusion model, mirroring biological integration of object identity ("what") and spatial orientation ("where") pathways for 3D shape generation from neural signals. We adapt MVDream (Shi et al., 2024), a text-to-multiview diffusion model that employs an inflated 3D self-attention mechanism to extend pretrained 2D self-attention into multi-view 3D attention, promoting consistent generation across multiple camera viewpoints. A pretrained, frozen EEGNet object classifier is repurposed as a feature extractor: its classification head is replaced by a 512-dimensional embedding layer whose output is projected to 1024 dimensions via a fixed random linear

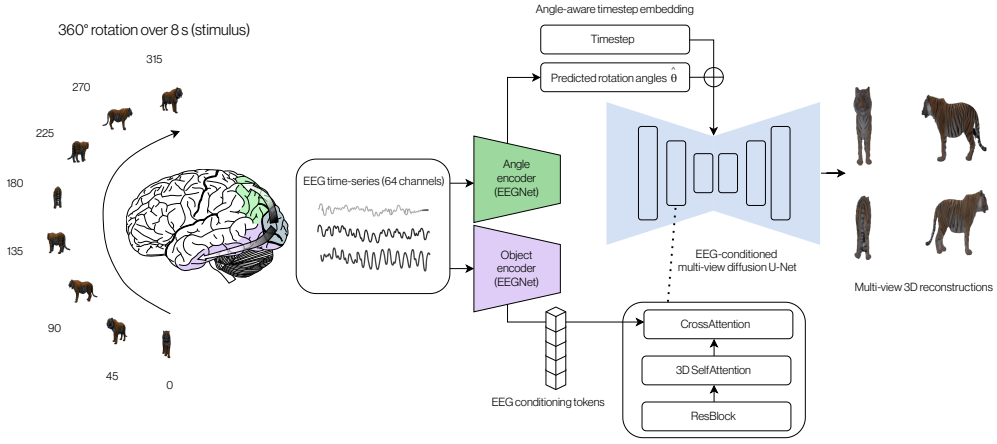


Figure 1: Proposed dual-stream architecture for 3D brain decoding from EEG.

layer matching CLIP's text-embedding space. In the diffusion U-Net, multi-view latents provide the queries for cross-attention, while the projected EEG embedding supplies the keys and values, enabling EEG-conditioned 3D-aware denoising. An angle encoder predicts EEG-derived rotation angles that are injected through an angle-aware timestep embedding. The diffusion model is fine-tuned for EEG-conditional generation using Low-Rank Adaptation (LoRA (Hu et al., 2021); $r = 24$ $\alpha = 48$) applied to attention and feed-forward layers, preserving the pretrained multi-view prior while enabling efficient adaptation. The multi-view diffusion loss is:

$$\mathcal{L}_{\text{MV}}(\theta) = \mathbb{E}_{t,\epsilon}\left[\|\epsilon - \epsilon_\theta(x_t; f_{\text{obj}}, \hat{\theta}, t)\|_2^2\right] \tag{2}$$

where x_t are noisy multi-view latents at diffusion step t, f_{obj} denotes the EEG-derived object identity embedding used as conditioning tokens, and $\hat{\theta}$ are the predicted rotation angles injected through an angle-aware timestep embedding. We fine-tune for up to 2,000 epochs on a single A100 GPU with early stopping based on validation loss, using batch size 32 and learning rate 10^{-4}; generation uses 50 denoising steps with the same scheduler as during training.

3.3. Channel Activation Analysis

To investigate whether EEGNet captures the functional specialization of dorsal (spatial processing) and ventral (object identity) visual pathways in 3D shape perception, we applied Gradient-weighted Class Activation Mapping (Selvaraju et al., 2020) to the convolutional layers. GradCAM computes spatiotemporal importance maps as:

$$\text{GradCAM}(i,j) = \text{ReLU}\left(\sum_k \alpha_k \cdot A_{k,i,j}\right) \tag{3}$$

where $\alpha_k = \frac{1}{HW}\sum_{m,n}\frac{\partial y}{\partial A_{k,m,n}}$ represents the average gradient of the target output y (object class or rotation angle $\hat{\theta}$) with respect to feature map k, and $A_{k,i,j}$ denotes activations at spatial position (i,j).

To assess regional contributions, we defined three anatomically-motivated channel subsets of equal size (10 channels each): *dorsal* pathway (Cz, C1, C2, Pz, P1, P2, CP1, CP2, POz, Oz), *ventral* pathway (T7, T8, TP7, TP8, P7, P8, PO7, PO8, O1, O2), and *motor/premotor* regions (C3, C4, C5, C6, FC3, FC4, FC1, FC2, CP5, CP6). Channels were ranked by total GradCAM activation, and top bias scores were calculated as:

$$\text{Top Bias} = \frac{1}{|C|} \sum_{i \in C} 1_{p_i < N/2} \tag{4}$$

where C represents the channel subset, p_i is channel i's rank, and $N = 64$ total channels. We hypothesized that object classification would engage ventral-like patterns, while angle regression would reveal dorsal-like signatures, and explored whether these contributions would vary over time across ventral, dorsal, and motor-related subsets.

4. Results

3D Shape Classification: We evaluated whether EEG signals preserve object identity despite continuous spatial transformations by classifying rotating 3D objects into six semantic categories from 8-second epochs. All three architectures achieved above-chance performance (chance level: 16.7%), with EEGNet yielding the highest mean accuracy, followed by CTNet and EEGConformer (Table 1). Across subject-specific models, performance varied and some models showed instability in cross-validation (e.g., Sub5: SD = 0.03; Sub6: SD = 0.10), but a consistent ranking emerged: Sub5 achieved the highest accuracy, followed by Sub6 and Sub11 (for EEGConformer, Sub9 was third). The EEGNet model trained for Sub5 was used for subsequent 3D reconstruction, as it combined the best validation and test accuracy with the most stable cross-validation performance.

Table 1: 3D shape classification and angle regression performance from EEG signals (mean ± SD across 5-fold CV).

Subject	**EEGNet (acc)**	**EEGConformer (acc)**	**CTNet (acc)**	**Angle (MAE)**
Sub1	0.31 ± 0.05	0.30 ± 0.03	0.27 ± 0.03	11.1 ± 0.03
Sub2	0.36 ± 0.09	0.19 ± 0.02	0.40 ± 0.04	11.0 ± 0.08
Sub3	0.29 ± 0.03	0.29 ± 0.03	0.30 ± 0.05	10.8 ± 0.04
Sub4	0.30 ± 0.05	0.34 ± 0.04	0.26 ± 0.03	10.9 ± 0.09
Sub5	0.68 ± 0.03	0.56 ± 0.05	0.56 ± 0.03	10.7 ± 0.04
Sub6	0.46 ± 0.10	0.54 ± 0.07	0.54 ± 0.06	10.8 ± 0.06
Sub7	0.32 ± 0.06	0.29 ± 0.05	0.30 ± 0.04	11.1 ± 0.12
Sub8	0.32 ± 0.05	0.22 ± 0.02	0.39 ± 0.07	10.9 ± 0.12
Sub9	0.37 ± 0.08	0.39 ± 0.03	0.34 ± 0.04	11.1 ± 0.08
Sub10	0.34 ± 0.08	0.37 ± 0.08	0.39 ± 0.04	10.9 ± 0.06
Sub11	0.43 ± 0.07	0.32 ± 0.05	0.42 ± 0.08	11.1 ± 0.04

Rotation Angle Regression: We evaluated the dorsal stream component by regressing rotation angles from EEG windows of varying durations ($T_w \in \{512\}$ samples, corresponding to 1s epochs). The target angle was computed as $\theta_{\text{true}} = \frac{\pi}{4} \times t_{\text{window\_center}}$ where

$t_{\text{window\_center}}$ represents the temporal center of each analysis window. In contrast to object classification, all angle regression models exhibited comparable error rates, with a 1 s window size proving most effective for learning angular information. This finding is consistent with the expectation that larger temporal windows encompass a greater range of angular positions, thereby reducing prediction accuracy for specific angles. Within the 1 s window, the angular distance spans 45°. The achieved MAE values ($10 - 11°$) across all models demonstrate robust performance in angular prediction.

Interpretability Analysis: Subject-level top-bias scores (Figure 2) show that the top four object decoders exhibit equivalent or stronger ventral and motor activation compared to dorsal channels, although this pattern is heterogeneous across participants. Paired t-tests ($\alpha = 0.05$) confirmed that ventral, dorsal, and motor subsets contribute comparably at the group level across 1s windows, indicating that the observed motor dominance is not a trivial artifact of isolated channel noise but reflects a systematic pattern across subjects. Grad-CAM therefore indicates that decoding performance is not driven by a static binary 'ventral-only' versus 'dorsal-only' readout but by how models distribute attention across dorsal, ventral, and motor channels over time; mirroring human dorsal-ventral integration dynamics (Ayzenberg et al., 2023). Specifically, the dissociation between object and an-

Figure 2: Top bias scores of different channel subsets, with subjects sorted from best performing model (left) to worse performing (right).

gle models suggests a partial disentangling of invariant and equivariant components: for angle decoders, accuracy benefits from early, positive motor bias together with a relaxing dorsal penalty, as dorsal correlations move from negative values toward approximately zero ($p < 0.05$); whereas object decoders are penalized when they overweight ventral channels early and dorsal channels late, consistent with more invariant object information arising from a balanced, temporally structured integration across dorsal, ventral, and motor-related subspaces rather than from a single ventral-like code (Farivar et al., 2009). Beyond subject-specific top-bias patterns, Grad-CAM maps show substantial between-subject variability (Appendix A), with a small subset of occipital–parietal and frontal channels displaying both high mean activation and high variance, while several midline posterior sites remain consistently low-importance. This suggests that dual-stream decoding relies on overlapping but individually idiosyncratic sensor configurations.

3D Reconstruction: Table 2 presents results on the held-out test set. Our method achieves SSIM of 0.856±0.038 and LPIPS of 0.275±0.061, slightly exceeding validation performance (SSIM: 0.833±0.035, LPIPS: 0.297±0.058) and indicating good generalization. Performance remains consistent across object categories (SSIM: 0.840–0.876; Appendix 3) and viewpoints (SSIM: 0.841–0.879; Appendix 4), suggesting that the EEG-conditioned multiview prior captures a viewpoint-dependent but geometrically coherent shape representation. Together with the channel-level Grad-CAM analysis, this supports a picture in

which ventral, dorsal, and motor-related channels jointly provide the invariant (identity) and equivariant (angle) information required for consistent 3D reconstruction. Figure 3 shows representative reconstructions, where generated objects preserve class-specific features (facial structure, organic forms) while maintaining geometric consistency across viewpoints, demonstrating the feasibility of EEG-conditioned neural-to-3D reconstruction from non-invasive signals.

Table 2: Quantitative evaluation of EEG-to-multiview reconstruction

Split	**PSNR ↑**	**SSIM ↑**	**LPIPS ↓**
Validation	15.82 ± 0.95	0.833 ± 0.035	0.297 ± 0.058
Test	$\mathbf{16.30 \pm 0.73}$	$\mathbf{0.856 \pm 0.038}$	$\mathbf{0.275 \pm 0.061}$

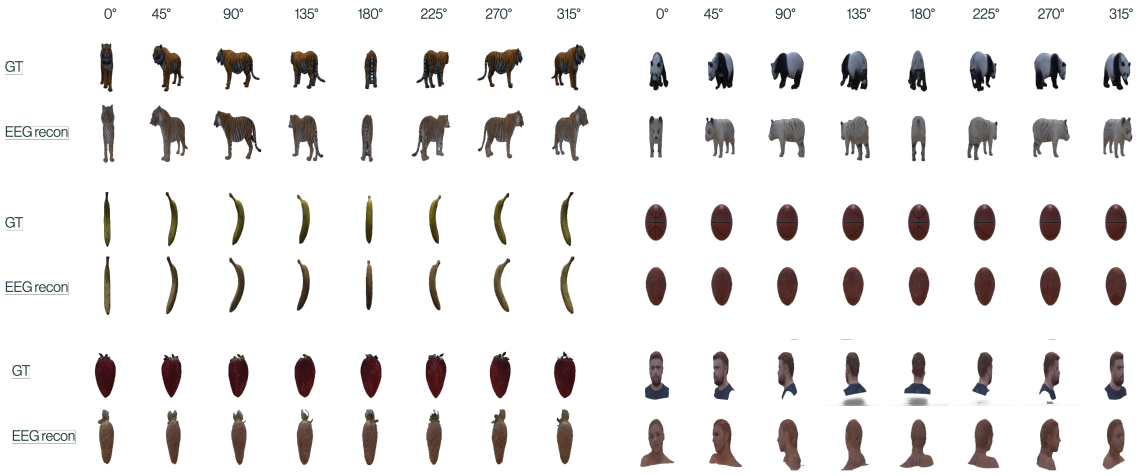


Figure 3: Multi-view test set reconstructions across 6 categories (ground truth: top, generated: bottom). Each shows 8 viewpoints (0°–315°).

5. Conclusion

This work makes three primary contributions to neural decoding and bio-inspired 3D vision: (1) We demonstrate the first dual-stream architecture for 3D brain decoding that uses separate modules for object identity (viewpoint-tolerant "what") and spatial transformation (viewpoint-dependent "where/how") processing, achieving robust performance on both classification (up to 68% accuracy) and angular regression (10-11° MAE); (2) We provide interpretable evidence for dynamic dorsal–ventral–motor involvement in EEG through Grad-CAM, showing that successful decoders avoid simple ventral dominance and instead rely on early motor engagement with time-varying dorsal-ventral contributions; and (3) We establish the feasibility of EEG-conditioned 3D reconstruction through multiview diffusion, enabling direct generation of 3D objects from neural signals. Further investigation of geometric priors through comparison with steerable CNNs and equivariant architectures would quantify the benefits of biological versus purely geometric inductive biases for 3D brain decoding. Quantitative evaluation through single-view versus multiview reconstruction metrics and analysis of rotational equivariance properties would further validate the geometric understanding captured by our dual-stream approach.

Acknowledgments

We thank Constanze Albrecht, Stephanie Chen, and Emmie Fitz-Gibbon for their assistance with data collection. Ninon was supported by an O'Shaughnessy Fellowship. We are grateful to Manuel Cherep and to the reviewers for feedback that improved this manuscript.

References

Vladislav Ayzenberg, Claire Simmons, and Marlene Behrmann. Temporal asymmetries and interactions between dorsal and ventral visual pathways during object recognition. *Cerebral Cortex Communications*, 4(1):tgad003, January 2023. ISSN 2632-7376. doi: 10.1093/texcom/tgad003. URL https://academic.oup.com/cercorcomms/article/doi/10.1093/texcom/tgad003/6987082.

Ji Seon Bang, Min Ho Lee, Siamac Fazli, Cuntai Guan, and Seong Whan Lee. Spatio-Spectral Feature Representation for Motor Imagery Classification Using Convolutional Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*, 33 (7):3038–3049, 2021. ISSN 2162-237X. doi: 10.1109/TNNLS.2020.3048385. URL http://www.scopus.com/inward/record.url?scp=85099724916&partnerID=8YFLogxK.

Paolo Bartolomeo. Chapter 10 - Visual objects and their colors. In Gabriele Miceli, Paolo Bartolomeo, and Vincent Navarro, editors, *Handbook of Clinical Neurology*, volume 187 of *The Temporal Lobe*, pages 179–189. Elsevier, January 2022. doi: 10.1016/B978-0-12-823493-8.00022-5. URL https://www.sciencedirect.com/science/article/pii/B9780128234938000225.

Hans P. Op de Beeck, Katrien Torfs, and Johan Wagemans. Perceived Shape Similarity among Unfamiliar Objects and the Organization of the Human Object Vision Pathway. *Journal of Neuroscience*, 28(40):10111–10123, October 2008. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.2511-08.2008. URL https://www.jneurosci.org/content/28/40/10111. Publisher: Society for Neuroscience Section: Articles.

Michael M. Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond Euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, July 2017. ISSN 1053-5888, 1558-0792. doi: 10.1109/MSP.2017.2693418. URL http://arxiv.org/abs/1611.08097. arXiv:1611.08097 [cs].

Ana B. Chica, Pedro M. Paz-Alonso, Antoni Valero-Cabré, and Paolo Bartolomeo. Neural bases of the interactions between spatial attention and conscious perception. *Cerebral Cortex (New York, N.Y.: 1991)*, 23(6):1269–1279, June 2013. ISSN 1460-2199. doi: 10.1093/cercor/bhs087.

Radoslaw Martin Cichy, Aditya Khosla, Dimitrios Pantazis, Antonio Torralba, and Aude Oliva. Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Scientific Reports*, 6(1): 27755, June 2016. ISSN 2045-2322. doi: 10.1038/srep27755. URL https://www.nature.com/articles/srep27755. Publisher: Nature Publishing Group.

Taco S. Cohen and Max Welling. Group Equivariant Convolutional Networks, June 2016. URL http://arxiv.org/abs/1602.07576. arXiv:1602.07576 [cs].

James J. DiCarlo, Davide Zoccolan, and Nicole C. Rust. How does the brain solve visual object recognition? *Neuron*, 73(3):415–434, February 2012. ISSN 1097-4199. doi: 10.1016/j.neuron.2012.01.010.

Ehsan Tarkesh Esfahani and V. Sundararajan. Classification of primitive shapes using brain–computer interfaces. *Computer-Aided Design*, 44(10):1011–1019, October 2012. ISSN 0010-4485. doi: 10.1016/j.cad.2011.04.008. URL https://www.sciencedirect.com/science/article/pii/S0010448511001035.

Reza Farivar, Olaf Blanke, and Avi Chaudhuri. Dorsal–Ventral Integration in the Recognition of Motion-Defined Unfamiliar Faces. *The Journal of Neuroscience*, 29(16):5336–5342, April 2009. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.4978-08.2009. URL https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.4978-08.2009.

John J. Foxe and Gregory V. Simpson. Flow of activation from V1 to frontal cortex in humans. A framework for defining "early" visual processing. *Experimental Brain Research*, 142(1):139–150, January 2002. ISSN 0014-4819. doi: 10.1007/s00221-001-0906-7.

Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36 (4):193–202, April 1980. ISSN 0340-1200, 1432-0770. doi: 10.1007/BF00344251. URL http://link.springer.com/10.1007/BF00344251.

Jason P Gallivan and Jody C Culham. Neural coding within human brain areas involved in actions. *Current Opinion in Neurobiology*, 33:141–149, August 2015. ISSN 0959-4388. doi: 10.1016/j.conb.2015.03.012. URL https://www.sciencedirect.com/science/article/pii/S0959438815000677.

Jianxiong Gao, Yuqian Fu, Yun Wang, Xuelin Qian, Jianfeng Feng, and Yanwei Fu. MinD-3D: Reconstruct High-quality 3D objects in Human Brain, July 2024. URL http://arxiv.org/abs/2312.07485. arXiv:2312.07485 [cs].

Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A. Wichmann, and Wieland Brendel. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness, November 2018. URL https://arxiv.org/abs/1811.12231v3.

M. A. Goodale and A. D. Milner. Separate visual pathways for perception and action. *Trends in Neurosciences*, 15(1):20–25, January 1992. ISSN 0166-2236. doi: 10.1016/0166-2236(92)90344-8.

Melvyn A. Goodale and David A. Westwood. An evolving view of duplex vision: Separate but interacting cortical pathways for perception and action. *Current Opinion in Neurobiology*, 14(2):203–211, 2004. ISSN 1873-6882. doi: 10.1016/j.conb.2004.03.002. Place: Netherlands Publisher: Elsevier Science.

Zhanqiang Guo, Jiamin Wu, Yonghao Song, Jiahui Bu, Weijian Mai, Qihao Zheng, Wanli Ouyang, and Chunfeng Song. Neuro-3D: Towards 3D Visual Decoding from EEG Signals, November 2024. URL http://arxiv.org/abs/2411.12248. arXiv:2411.12248 [cs].

Umut Güçlü and Marcel A. J. van Gerven. Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. *Journal of Neuroscience*, 35(27):10005–10014, July 2015. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.5023-14.2015. URL https://www.jneurosci.org/content/35/27/10005. Publisher: Society for Neuroscience Section: Articles.

A. M. Halliday and W. F. Michael. Changes in pattern-evoked responses in man associated with the vertical and horizontal meridians of the visual field. *The Journal of Physiology*, 208(2):499–513, June 1970. ISSN 0022-3751. doi: 10.1113/jphysiol.1970.sp009134. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1348763/.

Geoffrey Hinton, Sara Sabour, and Nicholas Frosst. MATRIX CAPSULES WITH EM ROUTING. 2018.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-Rank Adaptation of Large Language Models, October 2021. URL http://arxiv.org/abs/2106.09685. arXiv:2106.09685 [cs].

D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160(1):106–154, January 1962. ISSN 0022-3751. doi: 10.1113/jphysiol.1962.sp006837.

D. A. Jeffreys and J. G. Axford. Source locations of pattern-specific components of human visual evoked potentials. I. Component of striate cortical origin. *Experimental Brain Research*, 16(1):1–21, 1972. ISSN 0014-4819. doi: 10.1007/BF00233371.

Bela Julesz. *Foundations of cyclopean perception*. Foundations of cyclopean perception. U. Chicago Press, Oxford, England, 1971. Pages: xiv, 406.

Tetsuko Kasai and Takashi Morotomi. Event-related brain potentials during selective attention to depth and form in global stereopsis. *Vision Research*, 41(10-11):1379–1388, 2001. ISSN 0042-6989. doi: 10.1016/S0042-6989(01)00067-0. Place: Netherlands Publisher: Elsevier Science.

Koji Kashihara and Yoshibumi Nakahara. Evaluation of Task Performance during Mentally Imaging Three-Dimensional Shapes from Plane Figures. *Perceptual and Motor Skills*, 113 (1):188–200, August 2011. ISSN 0031-5125, 1558-688X. doi: 10.2466/03.04.22.PMS.113.4.188-200. URL https://journals.sagepub.com/doi/10.2466/03.04.22.PMS.113.4.188-200.

Seyed-Mahdi Khaligh-Razavi and Nikolaus Kriegeskorte. Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS computational biology*, 10 (11):e1003915, November 2014. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1003915.

Dwight J. Kravitz, Kadharbatcha S. Saleem, Chris I. Baker, and Mortimer Mishkin. A new neural framework for visuospatial processing. *Nature Reviews. Neuroscience*, 12(4): 217–230, April 2011. ISSN 1471-0048. doi: 10.1038/nrn3008.

Vernon J. Lawhern, Amelia J. Solon, Nicholas R. Waytowich, Stephen M. Gordon, Chou P. Hung, and Brent J. Lance. EEGNet: A Compact Convolutional Network for EEG-based Brain-Computer Interfaces. *Journal of Neural Engineering*, 15(5):056013, October 2018. ISSN 1741-2560, 1741-2552. doi: 10.1088/1741-2552/aace8c. URL http://arxiv.org/abs/1611.08024. arXiv:1611.08024 [cs].

Sidney R. Lehky and Terrence J. Sejnowski. Network model of shape-from-shading: neural function arises from both receptive and projective fields. *Nature*, 333(6172):452–454, June 1988. ISSN 1476-4687. doi: 10.1038/333452a0. URL https://www.nature.com/articles/333452a0. Publisher: Nature Publishing Group.

Dongyang Li, Chen Wei, Shiying Li, Jiachen Zou, Haoyang Qin, and Quanying Liu. Visual Decoding and Reconstruction via EEG Embeddings with Guided Diffusion, October 2024. URL http://arxiv.org/abs/2403.07721. arXiv:2403.07721 [cs].

Steven J. Luck. *An Introduction to the Event-Related Potential Technique*. MIT Press, Cambridge, MA, USA, 2 edition, May 2014. ISBN 978-0-262-52585-5.

Daniel Maturana and Sebastian Scherer. VoxNet: A 3D Convolutional Neural Network for real-time object recognition. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 922–928, Hamburg, Germany, September 2015. IEEE. ISBN 978-1-4799-9994-1. doi: 10.1109/IROS.2015.7353481. URL http://ieeexplore.ieee.org/document/7353481/.

Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy Networks: Learning 3D Reconstruction in Function Space, April 2019. URL http://arxiv.org/abs/1812.03828. arXiv:1812.03828.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis, August 2020. URL http://arxiv.org/abs/2003.08934. arXiv:2003.08934 [cs].

Mortimer Mishkin, Leslie G. Ungerleider, and Kathleen A. Macko. Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences*, 6:414–417, January 1983. ISSN 0166-2236. doi: 10.1016/0166-2236(83)90190-X. URL https://www.sciencedirect.com/science/article/pii/016622368390190X.

Thomas O'Connell, Tyler Bonnen, Yoni Friedman, Ayush Tewari, Josh Tenenbaum, Vincent Sitzmann, and Nancy Kanwisher. *Approaching human 3D shape perception with neurally mappable models*. August 2023. doi: 10.48550/arXiv.2308.11300.

S. Palmer, E. Rosch, and P. Chase. Canonical Perspective and the Perception of Objects. 1981.

Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation, January 2019. URL http://arxiv.org/abs/1901.05103. arXiv:1901.05103 [cs].

Nona Rajabi, Antônio H. Ribeiro, Miguel Vasco, Farzaneh Taleb, Mårten Björkman, and Danica Kragic. Human-Aligned Image Models Improve Visual Decoding from the Brain, June 2025. URL http://arxiv.org/abs/2502.03081. arXiv:2502.03081 [cs].

Sara Sabour, Nicholas Frosst, and Geoffrey E Hinton. Dynamic Routing Between Capsules.

Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *International Journal of Computer Vision*, 128(2):336–359, February 2020. ISSN 0920-5691, 1573-1405. doi: 10.1007/s11263-019-01228-7. URL http://arxiv.org/abs/1610.02391. arXiv:1610.02391 [cs].

Roger N. Shepard and Jacqueline Metzler. Mental rotation of three-dimensional objects. *Science*, 171(3972):701–703, 1971. ISSN 1095-9203. doi: 10.1126/science.171.3972.701. Place: US Publisher: American Assn for the Advancement of Science.

Yichun Shi, Peng Wang, Jianglong Ye, Mai Long, Kejie Li, and Xiao Yang. MVDream: Multi-view Diffusion for 3D Generation, April 2024. URL http://arxiv.org/abs/2308.16512. arXiv:2308.16512 [cs].

Jasper Snoek, Hugo Larochelle, and Ryan P. Adams. Practical Bayesian Optimization of Machine Learning Algorithms, August 2012. URL http://arxiv.org/abs/1206.2944. arXiv:1206.2944 [stat].

Yonghao Song, Qingqing Zheng, Bingchuan Liu, and Xiaorong Gao. EEG Conformer: Convolutional Transformer for EEG Decoding and Visualization. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:710–719, 2023. ISSN 1558-0210. doi: 10.1109/TNSRE.2022.3230250. URL https://ieeexplore.ieee.org/document/9991178. Conference Name: IEEE Transactions on Neural Systems and Rehabilitation Engineering.

Alfredo Spagna, Zoe Heidenry, Chloe Jeanne Lambert, Michelle Miselevich, Benjamin Eisenstadt, Laura Trembley, Zixin Liu, Jianghao Liu, and Paolo Bartolomeo. Visual mental imagery: evidence for a heterarchical neural architecture, June 2023. URL https://osf.io/vrqkh_v1.

Tom Theys, Maria C. Romero, Johannes van Loon, and Peter Janssen. Shape representations in the primate dorsal visual stream. *Frontiers in Computational Neuroscience*, 9:43, April 2015. ISSN 1662-5188. doi: 10.3389/fncom.2015.00043. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4406065/.

Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds, May 2018. URL http://arxiv.org/abs/1802.08219. arXiv:1802.08219 [cs].

Antonio Torralba, Phillip Isola, and William T. Freeman. *Foundations of computer vision*. Adaptive computation and machine learning series. The MIT Press, Cambridge, Massachusetts London, England, 2024. ISBN 978-0-262-04897-2.

Doris Y. Tsao, Nicole Schweers, Sebastian Moeller, and Winrich A. Freiwald. Patches of face-selective cortex in the macaque frontal lobe. *Nature Neuroscience*, 11(8):877–879, August 2008. ISSN 1546-1726. doi: 10.1038/nn.2158.

Ken-Ichiro Tsutsui, Min Jiang, Kazuo Yara, Hideo Sakata, and Masato Taira. Integration of Perspective and Disparity Cues in Surface-Orientation–Selective Neurons of Area CIP. *Journal of Neurophysiology*, 86(6):2856–2867, December 2001. ISSN 0022-3077, 1522-1598. doi: 10.1152/jn.2001.86.6.2856. URL https://www.physiology.org/doi/10.1152/jn.2001.86.6.2856.

Maryam Vaziri-Pashkam and Yaoda Xu. Goal-Directed Visual Processing Differentially Impacts Human Ventral and Dorsal Visual Representations. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 37(36):8767–8782, September 2017. ISSN 1529-2401. doi: 10.1523/JNEUROSCI.3392-16.2017.

Maurice Weiler, Mario Geiger, Max Welling, Wouter Boomsma, and Taco S Cohen. 3D Steerable CNNs: Learning Rotationally Equivariant Features in Volumetric Data.

Andrew E. Welchman, Arne Deubelius, Verena Conrad, Heinrich H. Bülthoff, and Zoe Kourtzi. 3D shape perception from combined depth cues in human visual cortex. *Nature Neuroscience*, 8(6):820–827, June 2005. ISSN 1097-6256. doi: 10.1038/nn1461.

Xin Xiang, Wenhui Zhou, and Guojun Dai. EEG-Driven 3D Object Reconstruction with Style Consistency and Diffusion Prior, November 2024. URL http://arxiv.org/abs/2410.20981. arXiv:2410.20981 [cs].

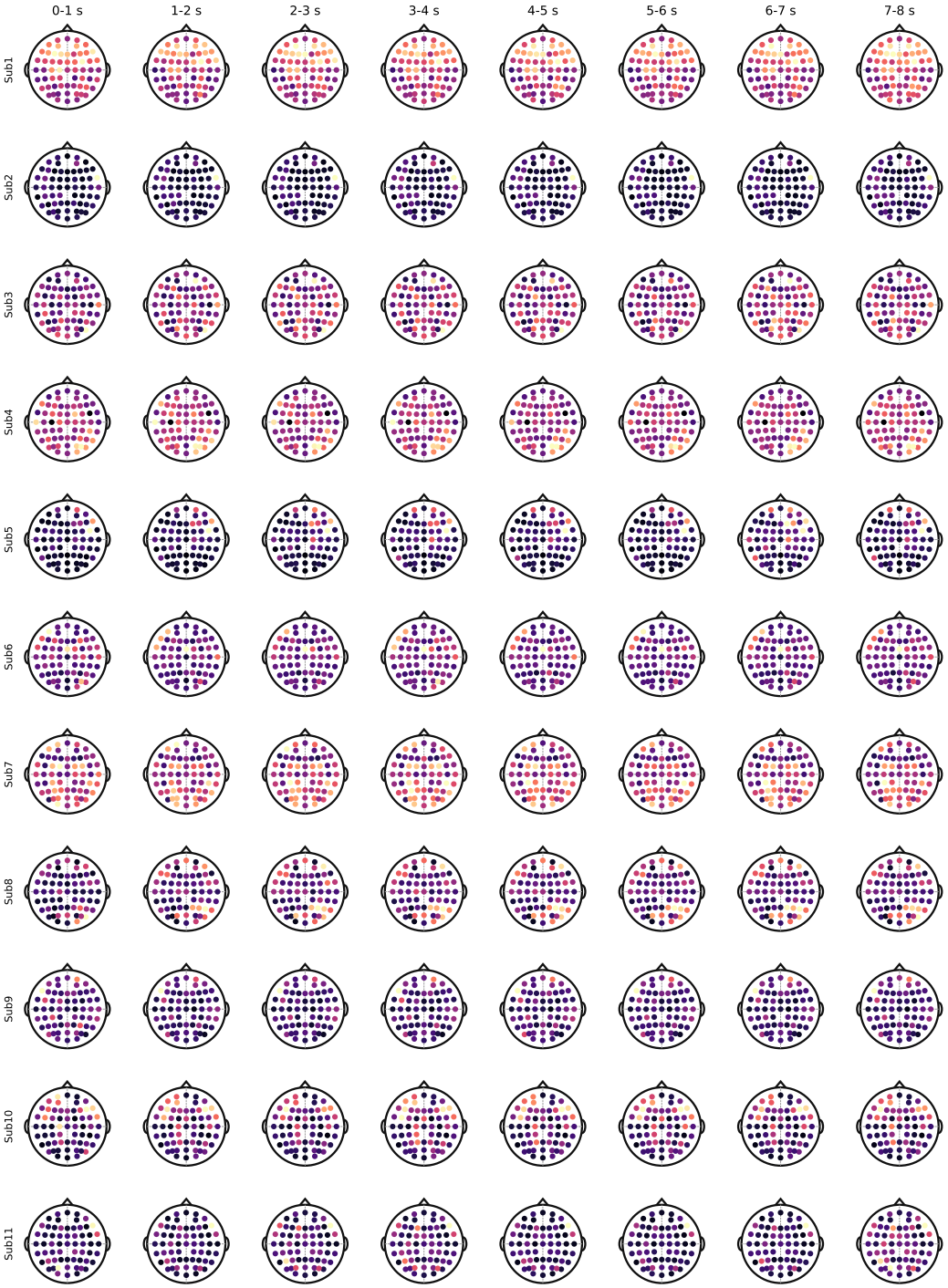
Yaoda Xu and Maryam Vaziri-Pashkam. Understanding transformation tolerant visual object representations in the human brain and convolutional neural networks. *NeuroImage*, 263:119635, November 2022. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2022.119635. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11283825/.

Daniel L. K. Yamins, Ha Hong, Charles F. Cadieu, Ethan A. Solomon, Darren Seibert, and James J. DiCarlo. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 111(23):8619–8624, June 2014. ISSN 1091-6490. doi: 10.1073/pnas.1403112111.

Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip Torr, and Vladlen Koltun. Point Transformer, September 2021. URL http://arxiv.org/abs/2012.09164. arXiv:2012.09164 [cs].

Wei Zhao, Xiaolu Jiang, Baocan Zhang, Shixiao Xiao, and Sujun Weng. CTNet: a convolutional transformer network for EEG-based motor imagery classification. *Scientific Reports*, 14(1):20237, August 2024. ISSN 2045-2322. doi: 10.1038/s41598-024-71118-7. URL https://www.nature.com/articles/s41598-024-71118-7. Publisher: Nature Publishing Group.

Appendix A. Gradient-weighted Class Activation Mapping 1-second windows for best 3D object recognition model per subject



Appendix B. Per-class reconstruction performance

Table 3: Per-class performance on test set.

Class	**PSNR ↑**	**SSIM ↑**	**LPIPS ↓**
Banana	16.50 ± 0.77	0.868 ± 0.041	0.249 ± 0.049
Basketball	15.77 ± 0.68	0.845 ± 0.028	0.292 ± 0.040
Face	16.38 ± 0.80	0.876 ± 0.027	0.282 ± 0.078
Panda	16.29 ± 0.99	0.868 ± 0.031	0.233 ± 0.040
Strawberry	16.01 ± 0.60	0.840 ± 0.030	0.306 ± 0.058
Tiger	16.52 ± 0.41	0.851 ± 0.041	0.280 ± 0.064
Overall	16.30 ± 0.73	0.856 ± 0.038	0.275 ± 0.061

Appendix C. Per-View Consistency

Table 4: Complete per-view metrics on validation and test sets.

View	**Angle**	**Validation**			**Test**		
		PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
0	0°	16.12 ± 1.19	0.862 ± 0.031	0.261 ± 0.063	16.46 ± 0.84	0.875 ± 0.026	0.251 ± 0.066
1	45°	15.84 ± 0.95	0.828 ± 0.040	0.303 ± 0.058	16.27 ± 0.78	0.852 ± 0.044	0.282 ± 0.065
2	90°	15.56 ± 0.96	0.811 ± 0.048	0.323 ± 0.062	16.12 ± 0.77	0.841 ± 0.053	0.295 ± 0.069
3	135°	15.71 ± 1.01	0.826 ± 0.042	0.299 ± 0.058	16.33 ± 0.70	0.853 ± 0.042	0.276 ± 0.060
4	180°	16.12 ± 1.19	0.866 ± 0.034	0.259 ± 0.070	16.55 ± 0.81	0.879 ± 0.027	0.244 ± 0.064
5	225°	15.71 ± 1.00	0.826 ± 0.040	0.307 ± 0.064	16.23 ± 0.76	0.852 ± 0.044	0.281 ± 0.063
6	270°	15.66 ± 0.89	0.816 ± 0.044	0.322 ± 0.062	16.13 ± 0.76	0.843 ± 0.049	0.292 ± 0.068
7	315°	15.79 ± 1.04	0.828 ± 0.039	0.306 ± 0.058	16.27 ± 0.75	0.853 ± 0.040	0.280 ± 0.063
Mean		15.82 ± 1.03	0.833 ± 0.040	0.297 ± 0.061	16.30 ± 0.77	0.856 ± 0.040	0.275 ± 0.063