HEMIPATCHGATE: PROBING CONTENT- AND LOCATION-DEPENDENT HEMISPHERIC DIFFERENCES WITH FMRI-BASED VISUAL DECODING

Anonymous authors
Paper under double-blind review

ABSTRACT

How visual information from natural scenes is represented across the cerebral hemispheres is a key question about the organization of the human visual system. A fundamental feature of this organization is contralateral preference: each hemisphere preferentially represents the opposite visual field. Beyond this contralateral preference, do visual representations show systematic hemispheric differences associated with scene content and spatial location? To address this question, we introduce HemiPatchGate, which separately encodes left- and right-hemisphere functional magnetic resonance imaging (fMRI) responses and uses gates to assign relative weights to their representations at each image-patch location for visual feature prediction. Across four subjects in the Natural Scenes Dataset, the gating weights recover a contralateral pattern that is consistent across natural images without explicit supervision of the visual-field–hemisphere relationship. Beyond this consistent contralateral pattern, greater image coverage by people and animals is associated with whole-image gate shifts toward the right-hemisphere branch. At the local level, increasing foreground coverage is associated with a greater tendency for weighting to shift toward the ipsilateral branch near the image's vertical midline than elsewhere. Changes in prediction error after zeroing either hemispheric input further support the contralateral pattern and the more pronounced foreground-related tendency toward ipsilateral shifts near the midline.

1 INTRODUCTION

The visual system exhibits a well-established contralateral organization: under central fixation, the left hemisphere preferentially represents the right visual field, and the right hemisphere preferentially represents the left visual field (Wandell et al., 2007; Fig. 1a). However, visual representation in natural scenes is not determined by spatial position alone: different visual content can occur at the same visual-field location, and similar content can occur at different locations. This raises a further question: beyond this stable contralateral spatial organization, do visual representations show systematic hemispheric differences associated with stimulus content and spatial location?

The Natural Scenes Dataset (NSD) provides large-scale fMRI responses to natural images (Allen et al., 2022). Recent visual decoders increasingly allow these responses to be analyzed in spatially resolved and interpretable ways: they recover high-level representations for retrieval and reconstruction (Scotti et al., 2023), preserve correspondence between predictions and image locations (Beliy et al., 2026), and support analyses linking cortical inputs to predictions (Feng et al., 2026). Together, these capabilities make it possible to compare predictions at matched image locations and relate them to cortical inputs. However, existing methods do not explicitly compare information derived from the two hemispheres within the same local prediction. Addressing our question therefore requires a decoder that keeps the two hemispheric representations distinguishable before integration, allowing their relative weighting to be compared across stimuli.

HemiPatchGate addresses this requirement by separately encoding left- and right-hemisphere fMRI responses and weighting the resulting representations at each image-patch location before joint CLIP feature prediction (Fig. 1b,c). We refer to these relative gating weights as hemispheric weighting and summarize them spatially to obtain a model-based hemispheric readout. To assess whether

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

this model-based readout reflects hemispheric representational differences in the recorded fMRI responses, we first validate its spatial organization against the established contralateral reference and use hemisphere-input zeroing to test whether image-specific gating variations correspond to differences in predictive information available from the two hemispheric fMRI inputs. Retraining controls then assess whether the directional consistency depends on the original input organization and correct fMRI–target pairing. With this validation in place, we then examine whether visual content in natural scenes is associated with changes in the relative weighting of the two hemispheric branches. We first test whether whole-image content is associated with an overall hemispheric shift, and then whether local content-related shifts vary across image locations, with particular attention to the vertical midline that separates the visual hemifields under central fixation.

Our main contributions are threefold:

- A framework for hemisphere-specific visual decoding. HemiPatchGate provides an explicit image-patch-level comparison of hemispheric weighting before the prediction head integrates the two branch representations.
- A hemispheric gating readout validated using established contralateral organization and hemisphere-input zeroing. In all four NSD subjects, the readout consistently follows the expected contralateral direction across natural images without explicit supervision of the visual-field–hemisphere relationship. Hemisphere-input zeroing further indicates that both the spatial pattern and image-specific variations in the readout reflect differences in predictive information available from the two hemispheric fMRI inputs.
- Hemispheric representational differences associated with visual content and spatial location. Our results show that greater animate coverage (the image fraction occupied by people and animals) is associated with an overall right-branch shift after controlling for total foreground coverage and scene category. Local foreground coverage is more positively associated with ipsilateral gate shifts near the vertical midline than elsewhere, with a corresponding spatial pattern in prediction-error changes after ipsilateral input zeroing.

2 Related work

Early visual decoding studies showed that viewed natural images could be identified from fMRI responses (Kay et al., 2008), while subsequent work reconstructed visual content from brain activity (Naselaris et al., 2009; Takagi & Nishimoto, 2023; Scotti et al., 2023; 2024). Alongside advances in image-level identification and reconstruction, local decoding methods preserve correspondence between model outputs and image space. Miyawaki et al. (2008) combined multiple local decoders to reconstruct visual patterns, whereas Brain-IT predicts visual features at individual image-patch locations (Beliy et al., 2026). This position-aligned formulation provides the basis for comparing information derived from the two hemispheres within matched local predictions.

Beyond image reconstruction, visual decoding models have also been used to examine how brain signals relate to model predictions. Generated-image analyses can probe the visual preferences of cortical regions (Luo et al., 2023; Ozcelik & VanRullen, 2023), while analyses of attention, internal weights, and information routing relate cortical inputs to predicted visual representations (Ozcelik & VanRullen, 2023; Wei et al., 2025; Feng et al., 2026). For example, supplementary attention analyses in Brain-IT report a lateralized pattern linking the left hemisphere more strongly to the right side of the image and the right hemisphere to the left (Beliy et al., 2026). HemiPatchGate differs by making the comparison between the two hemispheric sources an explicit part of the decoding architecture: representations derived from the left- and right-hemisphere fMRI inputs remain separate until a gating stage, where their relative weighting is read out at matched image-patch positions before integration. This design makes hemisphere-resolved weighting a directly testable quantity for subsequent analyses of stimulus content and spatial location.

Retinotopic mapping and population receptive field studies reveal systematic spatial correspondences between visual-field locations and cortical representations (Engel et al., 1997; Wandell et al., 2007; Dumoulin & Wandell, 2008). A fundamental feature of this organization is contralaterality: each hemisphere generally prefers stimuli in the opposite visual field. This contralateral preference extends beyond early visual cortex to object- and face-selective cortex and some scene-selective regions (Hemond et al., 2007; Arcaro et al., 2009). Beyond spatial organization, visual process-

ing also shows content-related hemispheric biases, including a left-hemisphere bias for visual word processing and a right-hemisphere bias for face processing (Behrmann & Plaut, 2015). Together, these findings suggest that hemispheric differences in visual representation may reflect both spatial organization and stimulus content. The former provides an established reference for evaluating the hemispheric gating readout, while the latter motivates examining whether natural-scene content is systematically associated with hemispheric representational differences.

3 METHODS

Figure 1 presents the overall framework of HemiPatchGate, with panel b showing the three-stage prediction of CLIP patch features from fMRI responses. Separate left- and right-hemisphere branches first encode their respective responses. A gating module then applies complementary channel-wise weights to the two branch representations at each image-patch location. The weighted representations are concatenated and passed to a prediction head that estimates the target features. These weights also define a model-based readout of the relative weighting of the two branches during prediction.

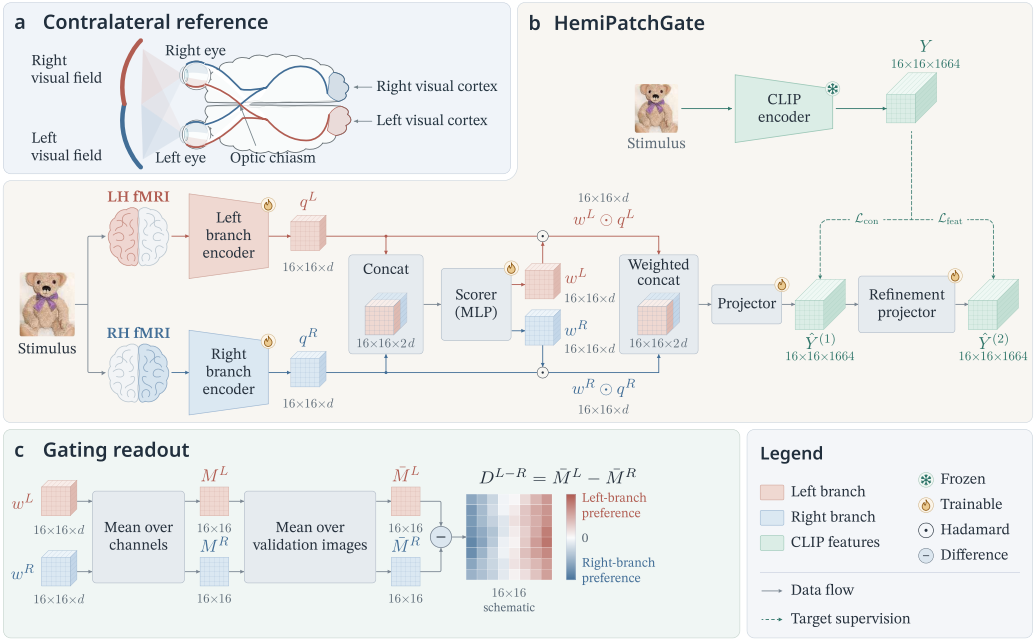

Figure 1: Contralateral reference and overview of HemiPatchGate. a, Under central fixation, information from the right visual field is represented predominantly in the left hemisphere, whereas information from the left visual field is represented predominantly in the right hemisphere. This relationship provides the spatial reference used to interpret the hemispheric gating readout. b, Separate left-hemisphere (LH) and right-hemisphere (RH) branches encode fMRI responses into 256 prediction slots, each trained against the CLIP token at the corresponding image-patch location. Complementary channel-wise weights w^L and w^R scale the normalized branch representations q^L and q^R. The weighted representations are concatenated in left-then-right order before entering the prediction head; $d = D_h = 512$ is the branch width. CLIP target features are extracted offline for training. c, Averaging w^L and w^R across channels and validation images yields subject-level gating maps $\bar{M}^L$ and $\bar{M}^R$. Their difference is $D^{L-R} = \bar{M}^L - \bar{M}^R$; positive values indicate a preference for the left branch.

3.1 IMAGE FEATURE GRID

We precompute target CLIP patch features for each stimulus image using OpenCLIP, an open-source implementation of CLIP (Radford et al., 2021; Cherti et al., 2023). The resulting matrix $Y \in \mathbb{R}^{P \times D_y}$

contains P = 256 tokens with D_y = 1664 channels each, arranged from left to right and top to bottom in a 16 × 16 grid (Dosovitskiy et al., 2021; Vaswani et al., 2017). Each token Y_p encodes content at image-patch location p while incorporating information from other locations through ViT self-attention (Amir et al., 2023; Zhou et al., 2022; Wang et al., 2025); this same patch index defines the corresponding prediction and gating-weight location. A re-extraction occlusion analysis found that the token corresponding to a masked location showed the largest change in 53.8% of trials (Appendix D).

3.2 HEMISPHERIC BRANCH ENCODING

HemiPatchGate takes stimulus-evoked fMRI response estimates from selected visual cortical regions in each hemisphere. The inputs X^L and X^R pass through branches with identical architectures and separate parameters. Following CSMSF (Zhang et al., 2026), each branch groups neighboring cortical vertices (sampling points on the cortical surface) at three scales. Each group's responses form one position in the input sequence. The three response streams are encoded into sequences of lengths 128, 64, and 64 and concatenated into 256 prediction slots. A hemisphere-specific Transformer allows each slot to integrate information from the entire branch while preserving sequence length and order. The resulting representations are h^L and $h^R \in \mathbb{R}^{B\times P\times D_h}$, where B is the batch size, P = 256, and D_h is the branch width (implementation details are given in Appendix A.2).

Slot p is supervised against the CLIP token at the same index p in the 16 × 16 grid. This correspondence is defined at the prediction level: each slot can integrate information from the full hemispheric branch through self-attention, while the gate and prediction head operate on the two branch slots with the same index p (Fig. 1b).

3.3 CHANNEL-WISE HEMISPHERIC GATING

Channel-wise hemispheric gating allows the relative weighting of the two branches to vary across channels at each image-patch location. We first normalize the branch representations, obtaining $q^L = \mathrm{Norm}_L(h^L)$ and $q^R = \mathrm{Norm}_R(h^R)$. At each location, the normalized representations are concatenated along the channel dimension and passed through a two-layer MLP shared across locations (with GELU activation). Its $2D_h$-dimensional output is split into left and right halves, giving gating scores ℓ^L and ℓ^R.

For each image-patch location p and channel c, the complementary weights are $w^L_{p,c} = \sigma((\ell^L_{p,c} - \ell^R_{p,c})/\tau)$ and $w^R_{p,c} = 1 - w^L_{p,c}$, so they sum to 1. Here, σ is the sigmoid function and τ is a learnable temperature (Appendix A.3). The normalized representations are weighted element-wise and concatenated along the channel dimension in left-then-right order to form $z = [w^L \odot q^L,\ w^R \odot q^R]$, which is passed to the prediction head. Concatenation retains the two weighted representations in distinct channel blocks; the subsequent prediction head integrates both blocks, while the same channel-wise gating weights are later summarized spatially to form the readout in Section 3.5.

3.4 FEATURE PREDICTION

The prediction head has two stages. The projector maps z to $\hat{Y}^{(1)}$ and is supervised by an image-level soft contrastive loss computed from sample-level similarities between predicted and target feature grids within the minibatch. The refinement projector maps $\hat{Y}^{(1)}$ to the final prediction $\hat{Y}^{(2)}$ and is supervised by a position-wise Smooth L1 feature-regression loss. The sum of these two losses, $\mathcal{L} = \mathcal{L}_{\mathrm{feat}} + \mathcal{L}_{\mathrm{con}}$, jointly trains the branch encoders, gate, and prediction head because z depends differentiably on the gating weights (loss definitions are given in Appendix A.3).

3.5 HEMISPHERIC GATING READOUT

To obtain an image-level hemispheric weighting map, we average the left-branch channel-wise gating weights $w^L_{p,c}$ over channels c at each image-patch location p, yielding M^L_p and a 16 × 16 map M^L for each validation image. Because the left- and right-branch weights are complementary within each channel, the corresponding right-branch map is $M^R_p = 1 - M^L_p$. Averaging these image-level maps across validation images at corresponding locations gives the subject-level maps $\bar{M}^L$ and $\bar{M}^R$

(Fig. 1c). We summarize their spatial contrast as $D_p^{L-R} = \bar{M}_p^L - \bar{M}_p^R = 2\bar{M}_p^L - 1$, where positive values indicate greater left-branch weighting and negative values greater right-branch weighting.

To summarize the left–right asymmetry with a scalar, we divide the image grid into left-half columns j = 1,...,8 and right-half columns j = 9,...,16. For each image, the gating difference Δ^L is the mean of M^L over the right half minus its mean over the left half. Averaging Δ^L across validation images gives the subject-level gating difference $\bar{\Delta}^L$; under the central-fixation reference, positive values indicate the expected contralateral direction. We use D^{L-R} to visualize the spatial pattern and $\bar{\Delta}^L$ to summarize its left–right direction (Section 4.3; Fig. 2).

The resulting gating maps summarize the relative weighting of the left- and right-hemisphere branches before the prediction head integrates the two representations. We use these maps as a model-based hemispheric readout in the analyses that follow.

4 EXPERIMENTS

We next evaluate HemiPatchGate in three stages. First, we assess prediction performance and test whether the hemispheric gating readout follows the expected contralateral organization and corresponds to hemispheric zeroing sensitivity. Second, we distinguish the spatial pattern shared across images from image-specific gating variation and examine the predictive association of the latter. Finally, we examine whether this image-specific variation is associated with whole-image content and local foreground coverage near the vertical midline.

4.1 EXPERIMENTAL SETUP

We train and evaluate HemiPatchGate on sub01, sub02, sub05, and sub07, the four subjects in the Natural Scenes Dataset (NSD; Allen et al., 2022) who completed all 40 sessions. NSD contains fMRI responses to natural scenes from the Common Objects in Context (COCO; Lin et al., 2014) dataset. We train an independent model for each subject using 27,000 single-presentation samples. The validation set contains the same 1,000 shared stimuli for all four subjects, with repeated fMRI responses averaged by stimulus within each subject.

Following the CSMSF preprocessing pipeline (Zhang et al., 2026), we group cortical vertices at three scales within each hemisphere and use the corresponding NSD beta estimates in fsaverage surface space as model inputs (Appendix A.2). Stimulus-related analyses use image-level gating readouts from the shared validation images for each subject-specific model. Input zeroing and left–right input swapping are applied only during evaluation; training settings are given in Appendix A.4 (Table 4). During NSD acquisition, stimuli were presented centrally and subjects were instructed to fixate the center of the screen. Under this fixation assumption, the image's vertical midline corresponds to the vertical meridian of the visual field. The right and left image halves therefore correspond to the right and left visual fields, which are represented predominantly by the left and right hemispheres, respectively (Fig. 1a). This relationship defines the contralateral reference used in the following analyses.

4.2 PREDICTION PERFORMANCE

We evaluate each subject's final checkpoint after 400 training epochs using 1,000-way retrieval. For each validation image, we flatten the predicted features $\hat{Y}^{(2)}$ and rank the 1,000 candidates by cosine similarity between the prediction and each candidate's target features. Subject-averaged Top-1 and Top-10 accuracies are 54.0% and 86.9%, respectively, well above the chance levels of 0.1% and 1% (Table 1). Per-subject validation losses are reported in Appendix B.1 (Table 5).

Table 1 also compares HemiPatchGate with a direct-concatenation reference (labeled No-gate) that concatenates the normalized left- and right-branch representations without gating weights. The prediction head still integrates both representations. Mean Top-1 accuracy is 54.0% for HemiPatchGate and 53.3% for the No-gate reference. The corresponding mean Top-10 accuracies are 86.9% and 87.2%, respectively. HemiPatchGate thus provides an explicit readout of hemispheric weighting at each patch while retaining retrieval accuracy close to that of the No-gate reference on this validation set.

Table 1: Prediction-to-target retrieval for HemiPatchGate (HPG) and the direct-concatenation reference (No-gate) on 1,000 validation images. Chance levels are 0.1% for Top-1 and 1% for Top-10.

Subject	**Top-1**		**Top-10**	
	HPG	**No-gate**	**HPG**	**No-gate**
sub01	66.6%	66.7%	92.2%	93.3%
sub02	60.5%	61.5%	90.4%	91.4%
sub05	47.3%	46.9%	85.8%	87.7%
sub07	41.5%	38.1%	79.1%	76.4%
Mean	54.0%	53.3%	86.9%	87.2%

4.3 CONTRALATERAL GATING

We test whether the hemispheric gating readout follows the expected contralateral direction under the central-fixation reference. The subject-level gating difference $\bar{\Delta}^L$ is positive for all four subjects, with a mean of 0.0716 and a range of 0.0231–0.1113 (Table 2). Left-branch gating is therefore higher on average in the right image half than in the left half, while the complementary right-branch gating is higher in the left half. Figure 2 shows the corresponding spatial pattern. The image-level gating difference Δ^L is also positive for all 1,000 validation images in each subject. Thus, the readout consistently follows the contralateral direction without supervision of the visual-field–hemisphere relationship.

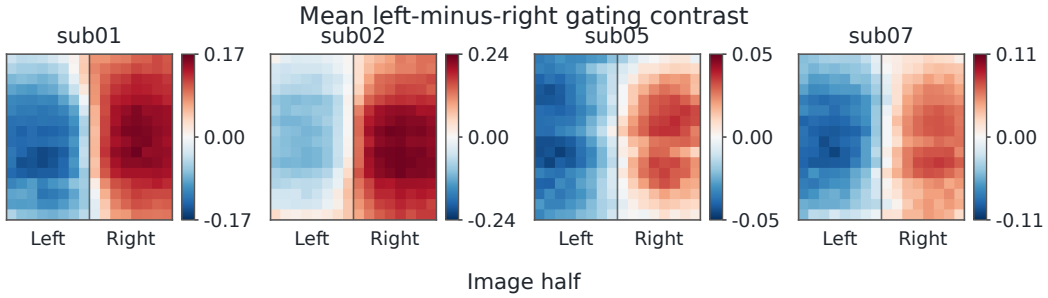

Figure 2: Subject-level gating contrast maps. Each panel shows $D^{L-R} = \bar{M}^L - \bar{M}^R$ for one independently trained subject model, where $\bar{M}^L$ and $\bar{M}^R$ are the left- and right-branch gating maps averaged across channels and validation images. Each panel uses an independently scaled color bar whose range matches the contrast values for that subject. Red indicates higher left-branch gating and blue higher right-branch gating; the vertical line marks the boundary between the two image halves. In all four subjects, gating favors the left branch in the right image half and the right branch in the left image half.

To assess the predictive relevance of this weighting pattern, we separately zero the left- and right-hemisphere fMRI inputs at inference and measure the increase in Smooth L1 feature-regression loss relative to the original-input condition. In this analysis, gates are recomputed from the perturbed inputs. Zeroing the left input increases loss more in the right image half, whereas zeroing the right input increases loss more in the left half. Averaged across subjects, the contralateral-minus-ipsilateral difference in loss increase is 0.0064 for left-input zeroing and 0.0046 for right-input zeroing. Thus, gating and zeroing sensitivity agree in their contralateral direction at the half-image level (Table 2).

As a complementary control, we further held the gates obtained from the full-input condition fixed and repeated hemisphere-input zeroing, asking whether the contralateral pattern in zeroing sensitivity would remain when post-perturbation gate reassignment was prevented. The contralateral-minus-ipsilateral loss contrast remained positive for both zeroing directions in every subject, with means of 0.0072 for left-input and 0.0049 for right-input zeroing (Table 2). This indicates that the contralateral pattern in zeroing sensitivity does not depend on gate reassignment after zeroing and

further supports that the hemispheric gating readout corresponds to differences in predictive information available from the two hemispheric fMRI inputs. Implementation details are provided in Appendix B.5.

Table 2: Contralateral gating and hemispheric zeroing sensitivity. Gating difference is the mean left-branch weight in the right image half minus that in the left half. Zeroing contrasts are the contralateral-minus-ipsilateral differences in Smooth L1 loss increase; positive values indicate the contralateral direction. Recomputed gates are obtained from perturbed inputs, whereas fixed gates are cached from the corresponding intact inputs. Means are arithmetic means across the four subjects.

Subject	**Gating difference**	**Left zeroed**		**Right zeroed**	
	$\bar{\Delta}^L$	**Recomputed**	**Fixed**	**Recomputed**	**Fixed**
sub01	0.1022	0.008796	0.010282	0.006066	0.006547
sub02	0.1113	0.009128	0.010823	0.006250	0.005617
sub05	0.0231	0.003617	0.003886	0.002726	0.002924
sub07	0.0497	0.004116	0.003995	0.003541	0.004394
Mean	0.0716	0.006414	0.007247	0.004646	0.004870

sub05 has the smallest gating difference and the smallest difference between the two image halves in zeroing-induced loss increase. Allen et al. (2022, Extended Data Fig. 4) reported greater eye movements in this subject, which may partly account for the weaker left–right separation under the central-fixation reference.

We use two retraining controls to examine whether this consistent gating direction can arise from the architecture alone. After responses from the two hemispheres are randomly regrouped, gating differences no longer show a consistent contralateral direction across all four subjects, although validation losses remain close to those of the main model. Permuting fMRI–target pairings sharply increases validation loss and likewise disrupts this cross-subject directional consistency. Together, these controls show that the two-branch architecture and positional encoding alone do not guarantee a consistent contralateral readout; the original organization of the fMRI inputs and correct fMRI–target pairing also matter. Detailed settings and per-subject results are provided in Appendices A.4 and B.3.

4.4 Spatial stability and predictive associations of gating variation

We first characterize the spatial pattern shared across natural images. Correlations between each image's gating map and the leave-one-out mean map range from 0.930 to 0.988 when averaged within subjects (Appendix B.4; Fig. 4; Table 10). Bilateral input zeroing retains 27.3%–68.0% of the original gating difference, whereas swapping hemispheric inputs reduces it to below 10% in every subject. These observations motivate separating the shared spatial pattern from image-specific gating variation. We therefore subtract each location's leave-one-out mean from both the left-minus-right gating contrast and the difference between left- and right-input zeroing losses, and test their correspondence under fixed gates. Correlations pooled over matched image–location pairs are 0.234, 0.354, 0.126, and 0.202 for sub01, sub02, sub05, and sub07, respectively; all image-level bootstrap 95% confidence intervals exclude zero (Table S1). Thus, after removing the shared location-wise pattern, image-specific gating variation remains positively associated with relative hemispheric zeroing sensitivity, supporting its use in the subsequent stimulus-content analyses. A stricter spatial-pattern control is reported in Appendix B.5.

4.5 Whole-image gate shifts

We quantify content-related variation using the whole-image gate shift. For each image, we average the left-branch gating map over all patches and subtract the mean of this quantity across all validation images for the same subject. Because the branch weights are complementary, the left-branch

mean alone is sufficient to summarize the whole-image shift. Positive values indicate higher left-branch weighting than the subject's average, whereas negative values indicate a relative shift toward the right branch. We relate this shift to animate coverage, defined as the fraction of the image occupied by person and animal regions in COCO panoptic segmentation (Appendix C.1; Fig. 5), while controlling for total foreground coverage and scene category.

The partial correlations are negative in all four subjects, ranging from -0.218 to -0.184 (Appendix C.2; Table 11). Thus, images with greater animate coverage tend to show a whole-image shift toward the right branch relative to the subject average. In 50,000 within-category permutations per subject, none of the permuted correlations has an absolute value as large as that of the observed correlation (Appendix C.1), supporting the reliability of this association within the analyzed image set. Together with the stable contralateral spatial pattern described above, this result shows that hemispheric weighting exhibits both a consistent spatial organization and systematic image-to-image variation related to stimulus content. Figure 3 illustrates this association using selected examples from sub01.

To examine sensitivity to the horizontal spatial distribution of animate regions, we further performed a spatial control analysis based on the same animate masks. We extracted two spatial covariates, including the left–right imbalance and horizontal centroid of animate regions, and incorporated them into the regression model together with the original control variables. After adding these two measured spatial summaries, the standardized animate-coverage coefficient changed by -3.4% to $+5.3\%$ across subjects (Appendix C.1). This shows robustness to these controls; other aspects of horizontal placement were not tested.

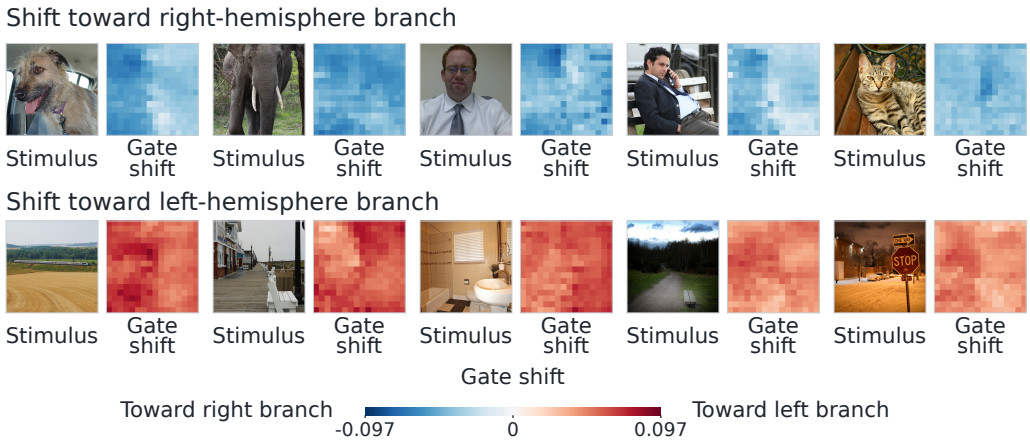

Figure 3: Selected examples of content-related gate shifts in sub01. Each pair shows a stimulus (left) and its left-branch gating map minus the mean map of the other 999 validation images (right). Blue and red indicate shifts toward the right- and left-hemisphere branches, respectively, with one shared color scale. From the 25 most strongly shifted images in each direction, we selected the five with the highest animate coverage for right-branch shifts and the five with the lowest for left-branch shifts. These purposefully selected examples illustrate, but do not independently test, the full-sample association (Appendix C.1; Fig. 5).

4.6 Variation near the image's vertical midline

The whole-image analysis does not show where stimulus-related changes occur within an image, so we examine their spatial distribution. For each image and patch location, we subtract the mean weight over the other 999 validation images from each left-branch weight at the same channel and location. We then take the absolute differences and average them across the 512 channels. This gating-change magnitude is 3.9%–5.5% higher in the vertical-midline band (columns 7–10) than in peripheral regions (columns 1–4 and 13–16) in all four subjects, with a mean increase of 4.6% (Appendix C.2; Table 11; Fig. 6b). Because the band is also centrally located, we additionally regress gating-change magnitude on proximity to the image's vertical and horizontal midlines. Vertical-midline proximity has a positive coefficient in all four subjects (0.105–0.115), whereas horizontal-

midline proximity has a negative coefficient (-0.163 to -0.030), with both effects significant in every subject. The lack of a corresponding increase toward the horizontal midline argues against a simple general center-proximity explanation.

The preceding analysis shows where gating varies most across images, but not which local image properties are associated with this variation. We therefore screened 52 candidate measures using separate sets of 500 discovery and 500 confirmation images. Local foreground coverage—the fraction of each patch occupied by foreground objects—had the smallest discovery-stage p-value among the 52 candidates and passed confirmation for both the gating contrast and the difference in loss increases between left- and right-hemisphere input zeroing (Appendix C.3).

Having identified local foreground coverage as the candidate content variable, we next ask whether the stronger gating variation observed near the vertical midline has a systematic direction as foreground coverage increases. Because gating-change magnitude captures only the size of a change, we use the ipsilateral gate shift to describe its direction relative to the average weighting at each location. We use left-branch weights for patches in the left image half and right-branch weights for patches in the right half. At each patch location, the ipsilateral gate shift is the current image's channel-averaged weight minus the mean of the corresponding weights across the other validation images. Positive values therefore indicate a shift toward the ipsilateral branch. As local foreground coverage increases, the associated gate shift is directed more toward the ipsilateral branch inside the vertical-midline band than outside it. This is reflected in a more positive foreground-coverage slope inside the band in all four subjects (outcome-standardized slope differences 0.080–0.231; all $p \leq 4.5 \times 10^{-4}$; Table 12a), whereas the corresponding horizontal-midline differences show no consistent pattern.

We next ask whether this location-dependent gate shift is accompanied by a corresponding pattern in sensitivity to the ipsilateral hemispheric input. Zeroing that input produces loss increases that are more positively associated with local foreground coverage inside the vertical band than outside it in every subject (outcome-standardized slope differences 0.095–0.381; all significant; Table 12b). Each vertical-band difference also exceeds its horizontal-band counterpart. Continuous proximity analyses yield the same qualitative pattern (Appendix C.5.2).

Repeating the same regression with the original gates held fixed yields positive vertical-band slope differences of 0.174, 0.159, 0.049, and 0.288 for sub01, sub02, sub05, and sub07, respectively (all $p \leq 0.013$). The vertical-minus-horizontal contrasts are also positive in all four subjects (0.176–0.389; Appendix C.5.3). Thus, the more positive association between foreground coverage and ipsilateral zeroing sensitivity near the vertical midline persists without gate reassignment. Together with the gating results, this supports a location-dependent relationship between local foreground content and predictive information available from the two hemispheric fMRI inputs.

5 Conclusion

Across four NSD subjects, hemispheric weighting shows a contralateral spatial pattern that remains consistent across natural images. After removing location-wise mean patterns, image-specific gating contrasts are positively associated with relative hemispheric zeroing sensitivity, including when the original gates are held fixed. Against this spatial background, greater animate coverage is associated with whole-image shifts toward the right branch, and local foreground coverage is more positively associated with ipsilateral gate shifts near the vertical midline than elsewhere. The corresponding foreground-by-location pattern in ipsilateral zeroing sensitivity persists with fixed gates. These convergent readouts support content- and location-dependent hemispheric representational differences as assessed through visual decoding. Further evaluation with larger populations and additional cortical measurements may help characterize the generality and neural basis of these hemispheric differences.

AI USE STATEMENT

Generative AI was used to assist with manuscript writing and language editing. The authors reviewed the AI-assisted content and take responsibility for the final paper.

REPRODUCIBILITY STATEMENT

Details of data preprocessing, target extraction, model architecture, training, evaluation, and statistical analyses are provided in the main text and appendices. The implementation code is provided in the supplementary material.

ETHICS STATEMENT

This study analyzes the publicly released Natural Scenes Dataset; the original acquisition protocol, informed consent, and ethical approval were handled by the NSD data providers. No new human-subject data were collected in this study.

REFERENCES

Emily J. Allen, Ghislain St-Yves, Yihan Wu, Jesse L. Breedlove, Jacob S. Prince, Logan T. Dowdle, Matthias Nau, Brad Caron, Franco Pestilli, Ian Charest, J. Benjamin Hutchinson, Thomas Naselaris, and Kendrick Kay. A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nature Neuroscience*, 25:116–126, 2022. doi: 10.1038/s41593-021-00962-x.

Shir Amir, Yossi Gandelsman, Shai Bagon, and Tali Dekel. On the effectiveness of ViT features as local semantic descriptors. In *Computer Vision – ECCV 2022 Workshops*, Lecture Notes in Computer Science, pp. 39–55. Springer Nature Switzerland, Cham, 2023. doi: 10.1007/978-3-031-25069-9_3.

Michael J. Arcaro, Stephanie A. McMains, Benjamin D. Singer, and Sabine Kastner. Retinotopic organization of human ventral visual cortex. *Journal of Neuroscience*, 29(34):10638–10652, 2009. doi: 10.1523/JNEUROSCI.2807-09.2009.

Marlene Behrmann and David C. Plaut. A vision of graded hemispheric specialization. *Annals of the New York Academy of Sciences*, 1359(1):30–46, 2015. doi: 10.1111/nyas.12833.

Roman Beliy, Amit Zalcher, Jonathan Kogman, Navve Wasserman, and Michal Irani. Brain-IT: Image reconstruction from fMRI via brain-interaction transformer. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=9KjXqkfbPw.

Mehdi Cherti, Romain Beaumont, Ross Wightman, Mitchell Wortsman, Gabriel Ilharco, Cade Gordon, Christoph Schuhmann, Ludwig Schmidt, and Jenia Jitsev. Reproducible scaling laws for contrastive language-image learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2818–2829, 2023.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=YicbFdNTTy.

Serge O. Dumoulin and Brian A. Wandell. Population receptive field estimates in human visual cortex. *NeuroImage*, 39(2):647–660, 2008. doi: 10.1016/j.neuroimage.2007.09.034.

Stephen A. Engel, Gary H. Glover, and Brian A. Wandell. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral Cortex*, 7(2):181–192, 1997. doi: 10.1093/cercor/7.2.181.

Pinyuan Feng, Hossein Adeli, Wenxuan Guo, Fan Cheng, Ethan Hwang, and Nikolaus Kriegeskorte. Towards interpretable visual decoding with attention to brain representations. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=YWlYITAhMC.

Christopher C. Hemond, Nancy G. Kanwisher, and Hans P. Op de Beeck. A preference for contralateral stimuli in human object- and face-selective cortex. *PLOS ONE*, 2(6):e574, 2007. doi: 10.1371/journal.pone.0000574.

Kendrick N. Kay, Thomas Naselaris, Ryan J. Prenger, and Jack L. Gallant. Identifying natural images from human brain activity. *Nature*, 452:352–355, 2008. doi: 10.1038/nature06713.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common objects in context. In *Computer Vision – ECCV 2014*, volume 8693 of *Lecture Notes in Computer Science*, pp. 740–755. Springer, Cham, 2014. doi: 10.1007/978-3-319-10602-1_48.

Andrew Luo, Maggie Henderson, Leila Wehbe, and Michael Tarr. Brain diffusion for visual exploration: Cortical discovery using large scale generative models. In *Advances in Neural Information Processing Systems*, volume 36, pp. 75740–75781. Curran Associates, Inc., 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/ef0c0a23a1a8219c4fc381614664df3e-Abstract-Conference.html.

Yoichi Miyawaki, Hajime Uchida, Okito Yamashita, Masa-aki Sato, Yusuke Morito, Hiroki C. Tanabe, Norihiro Sadato, and Yukiyasu Kamitani. Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron*, 60(5):915–929, 2008. doi: 10.1016/j.neuron.2008.11.004.

Thomas Naselaris, Ryan J. Prenger, Kendrick N. Kay, Michael Oliver, and Jack L. Gallant. Bayesian reconstruction of natural images from human brain activity. *Neuron*, 63(6):902–915, 2009. doi: 10.1016/j.neuron.2009.09.006.

Furkan Ozcelik and Rufin VanRullen. Natural scene reconstruction from fMRI signals using generative latent diffusion. *Scientific Reports*, 13:15666, 2023. doi: 10.1038/s41598-023-42891-8.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, pp. 8748–8763, 2021.

Paul S. Scotti, Atmadeep Banerjee, Jimmie Goode, Stepan Shabalin, Alex Nguyen, Ethan Cohen, Aidan J. Dempster, Nathalie Verlinde, Elad Yundler, David Weisberg, Kenneth A. Norman, and Tanishq Mathew Abraham. Reconstructing the Mind's Eye: fMRI-to-image with contrastive learning and diffusion priors. In *Advances in Neural Information Processing Systems*, volume 36, pp. 24705–24728. Curran Associates, Inc., 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/4ddab70bf41ffe5d423840644d3357f4-Abstract-Conference.html.

Paul S. Scotti, Mihir Tripathy, Cesar Kadir Torrico Villanueva, Reese Kneeland, Tong Chen, Ashutosh Narang, Charan Santhirasegaran, Jonathan Xu, Thomas Naselaris, Kenneth A. Norman, and Tanishq Mathew Abraham. MindEye2: Shared-subject models enable fMRI-to-image with 1 hour of data. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 44038–44059. PMLR, 2024. URL https://proceedings.mlr.press/v235/scotti24a.html.

Yu Takagi and Shinji Nishimoto. High-resolution image reconstruction with latent diffusion models from human brain activity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14453–14463, 2023.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, pp. 5998–6008, 2017.

Brian A. Wandell, Serge O. Dumoulin, and Alyssa A. Brewer. Visual field maps in human cortex. *Neuron*, 56(2):366–383, 2007. doi: 10.1016/j.neuron.2007.10.012.

Feng Wang, Jieru Mei, and Alan Yuille. SCLIP: Rethinking self-attention for dense vision-language inference. In *Computer Vision – ECCV 2024*, Lecture Notes in Computer Science, pp. 315–332. Springer Nature Switzerland, Cham, 2025. doi: 10.1007/978-3-031-72664-4_18.

Yuxiang Wei, Yanteng Zhang, Xi Xiao, Tianyang Wang, Xiao Wang, and Vince D. Calhoun. MoRE-Brain: Routed mixture of experts for interpretable and generalizable cross-subject fMRI visual decoding. In *Advances in Neural Information Processing Systems*, volume 38, pp. 55870–55902. Curran Associates, Inc., 2025. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/50ea4dbd1cff6bd3daef939eff10c092-Abstract-Conference.html.

Baogang Zhang, Liyun Dai, Wutao Lou, Yongnu Jin, Sihan Chen, Runlin Wen, and Jiancheng Sun. fMRI-based cross-subject visual reconstruction via structural consistency and multi-scale cortical modeling. *Information Fusion*, 133:104352, 2026. doi: 10.1016/j.inffus.2026.104352.

Chong Zhou, Chen Change Loy, and Bo Dai. Extract free dense labels from CLIP. In *Computer Vision – ECCV 2022*, Lecture Notes in Computer Science, pp. 696–712. Springer Nature Switzerland, Cham, 2022. doi: 10.1007/978-3-031-19815-1_40.

A IMPLEMENTATION DETAILS

A.1 TARGET IMAGE FEATURES

Target image features are extracted with a pretrained OpenCLIP ViT-bigG-14 image encoder using the laion2b_s39b_b160k weights. NSD RGB images are resized to 224 pixels and center-cropped to 224×224. The encoder uses a patch size of 14 and produces 16×16 patch tokens from each input image. The extraction script uses FrozenOpenCLIPImageEmbedder with output_tokens=True and only_tokens=True, returning 256 visual Transformer patch tokens per image. We remove the batch dimension and save the resulting 256×1664 tensor directly as the training target. The OpenCLIP image encoder is used only offline to generate target tensors and does not participate in HemiPatch-Gate training. Target extraction preserves the original token scale and disables l2_norm_tokens.

A.2 MULTISCALE HEMISPHERIC BRANCH ENCODING

Following CSMSF (Zhang et al., 2026), each hemispheric branch receives three scales of fMRI vertex-group representations. We use NSD surface beta estimates in fsaverage/betas_fithrf_GLMdenoise_RR, loaded separately from the left- and right-hemisphere session files. Before training, beta values are z-scored element-wise using training-set statistics at each scale; the mean and standard deviation are computed separately for the two hemispheres. Sample organization follows the NSD experimental design: responses to non-shared training stimuli are retained as single-presentation samples, whereas repeated fMRI responses to each shared validation stimulus are averaged before model inputs are constructed. Because the beta estimates are already aligned to the fsaverage surface, the predefined vertex groups can directly index the corresponding beta values. Vertex groups are constructed from the NSD fsaverage surface and ROI probability maps thresholded at 0.5. We use seven ROI categories (EarlyVisPlus, FaceBody, OPA, PPA, RSC, Word, and EBA) and allocate group counts in proportion to the number of labeled vertices from each category within each hemisphere. Group centers are selected by farthest-point sampling based on geodesic distance along the cortical surface; each group contains a fixed number of vertices nearest to its center within the same ROI category. The three scale configurations are 256×256, 512×128, and 1024×64, where the first number in each configuration is the total number of groups across both hemispheres and the second is the number of vertices per group. After splitting by hemisphere, each branch receives tensors with shapes 128×256, 256×128, and 512×64.

The center coordinates of each vertex group are taken from the 3D coordinates of the selected center vertex on the fsaverage cortical surface and are fixed after preprocessing. At each scale, a two-layer one-dimensional convolutional encoder first produces the local response representation. To construct positional encoding, the model first computes a pairwise Euclidean distance matrix between the 3D coordinates of group centers within the same hemisphere and scale; this distance matrix is linearly projected, normalized, passed through a nonlinear activation, and added to the fMRI response representation. Thus, geodesic distance along the cortical surface determines vertex-group construction, while Euclidean distance between cortical-surface coordinates provides relative positional encoding to the model.

Each scale preserves the ROI concatenation order established during preprocessing and the original group order within each ROI. Sequence downsampling uses a separate set of layers. The length-128 scale is not downsampled. The length-256 scale is reduced to 64 positions by two Conv1d layers with kernel size 2 and stride 2, and the length-512 scale is reduced to 64 positions by three layers with the same kernel size and stride. These strided convolutions aggregate adjacent positions in the fixed sequence, while cortical spatial relationships between group centers are represented separately by the relative positional encoder. Each scale is then projected to 512 channels with a Conv1d layer of kernel size 3 and padding 1. Concatenating the three scale sequences yields a branch sequence with 128 + 64 + 64 = 256 prediction slots.

The left and right hemispheric branches use separate 6-layer Transformer encoders with 2 attention heads per layer; parameters are not shared between branches. Global self-attention allows each prediction slot to integrate information from all slots in the same branch while preserving sequence length and order. The encoder therefore produces a tensor of shape $B \times P \times D_h$, where P = 256 and D_h = 512. Its output is passed to the gating module and prediction head described above.

The hemispheric gating readout compares the relative gating weights assigned to the left- and right-hemisphere branch representations. Hemisphere-input zeroing removes the modeled fMRI input from one hemisphere at inference. These hemisphere-level analyses do not isolate differences among individual visual areas, functional ROIs, or pRF subregions within a hemisphere.

A.3 GATING, PREDICTION HEAD, AND TRAINING OBJECTIVE

Before gating, each hemispheric branch representation is normalized by its own layer-normalization module. The gating module is an MLP shared across patch locations. At each location, it takes $[q^L_{p,:}, q^R_{p,:}] \in \mathbb{R}^{2D_h}$ and applies a $2D_h \to D_h \to 2D_h$ linear–GELU–linear mapping, producing $\ell^L_{p,:}, \ell^R_{p,:} \in \mathbb{R}^{D_h}$. Here D_h = 512. The temperature parameter τ is initialized to 1.0 and clipped to $[10^{-3}, 10]$ during the forward pass.

The prediction head consists of two MLP stages. The projector maps the fused representation from B $\times$ 256 $\times$ 1024 to $\hat{Y}^{(1)} \in \mathbb{R}^{B\times 256\times 1664}$ for the soft contrastive loss. The refinement projector takes $\hat{Y}^{(1)}$ as input and produces the final prediction $\hat{Y}^{(2)} \in \mathbb{R}^{B\times 256\times 1664}$ for the Smooth L1 regression loss. Table 3 summarizes the implementation configuration.

Table 3: Implementation configuration of the branch encoder, gate, and prediction head. Here V_g is the number of vertices per group, and N_g is the number of group positions for one hemisphere at a given scale.

Module	**Configuration**
Local response encoder	Two Conv1d layers, $V_g \to 256 \to 256$, kernel 3, padding 1; BatchNorm and SiLU after each layer; dropout 0.1.
Relative positional encoder	Pairwise Euclidean-distance row; Linear $N_g \to 256$; LayerNorm; SiLU; added to the local response representation.
Scale downsampling	Length 128: identity; length 256: two Conv1d layers with kernel size 2 and stride 2; length 512: three layers. Each layer uses 256 channels, BatchNorm, and SiLU.
Scale projection	Conv1d $256 \to 512$, kernel 3, padding 1; BatchNorm; SiLU; dropout 0.1.
Branch Transformer	6 encoder layers; width 512; 2 attention heads; feed-forward width 2048; ReLU; dropout 0.1.
Gate scorer	Linear $1024 \to 512$; GELU; Linear $512 \to 1024$, split into two 512-channel score tensors.
Projector	$1024 \to 1024 \to 1024 \to 1024 \to 1664$; LayerNorm and GELU after the first three hidden layers; dropout 0.1 after the first two.
Refinement projector	$1664 \to 1024 \to 1024 \to 1024 \to 1664$; LayerNorm and GELU after the three 1024-dimensional layers; dropout 0.1 after the first two.

The training objective combines a local feature-regression loss and an image-level soft contrastive loss:

$$\mathcal{L} = \mathcal{L}_{\text{feat}} + \mathcal{L}_{\text{con}}. \tag{1}$$

The local feature-regression loss uses Smooth L1 with $\beta = 1$, averaged over all elements:

$$\mathcal{L}_{\text{feat}} = \text{SmoothL1}\left(\hat{Y}^{(2)}, Y\right). \tag{2}$$

At each sample and patch index, $\hat{Y}^{(2)}_p$ is aligned with the OpenCLIP target feature Y_p. For the soft contrastive loss, Y and $\hat{Y}^{(1)}$ are reshaped from $\mathbb{R}^{B\times P\times D_y}$ to $\mathbb{R}^{B\times(PD_y)}$ and L2-normalized along

the final dimension, yielding $\bar{Y}$ and $\overline{\hat{Y}^{(1)}}$. These normalized copies are used for the contrastive term; Smooth L1 compares $\hat{Y}^{(2)}$ with the original target Y. The target–target and prediction–target similarity matrices are

$$S_{YY} = \frac{\bar{Y}\bar{Y}^\top}{t}, \qquad S_{\hat{Y}Y} = \frac{\overline{\hat{Y}^{(1)}}\bar{Y}^\top}{t}, \tag{3}$$

where $t = 0.006$ is the contrastive temperature. The soft target distribution is constructed from the precomputed target features with gradients stopped. We define the contrastive loss using row-wise softmax as follows:

$$\begin{aligned}\mathcal{L}_{\text{con}} = -\frac{1}{2B}\sum_{i=1}^{B}\Bigg[\sum_{j=1}^{B}\big[\operatorname{softmax}(S_{YY})\big]_{ij}\log\big[\operatorname{softmax}(S_{\hat{Y}Y})\big]_{ij} \\ +\sum_{j=1}^{B}\big[\operatorname{softmax}(S_{YY}^\top)\big]_{ij}\log\big[\operatorname{softmax}(S_{\hat{Y}Y}^\top)\big]_{ij}\Bigg].\end{aligned} \tag{4}$$

In Eq. (4), the first term constrains the within-batch similarity distribution when predicted features are used as queries; the second imposes the same constraint in the reverse direction with target features as queries. The feature-regression and contrastive losses jointly train the model to predict local image features while preserving image-level structure.

A.4 Training run configuration

HemiPatchGate and the retraining controls use the same training configuration, listed in Table 4: 400 epochs with OneCycleLR, a two-epoch warm-up, a final learning rate of one tenth of the initial value, and dropout 0.1.

Table 4: Training settings shared by HemiPatchGate and the retraining controls. Optimization uses AdamW; the OneCycleLR schedule uses a peak learning rate of 1.0×10^{-4}.

Item	**Setting**
Maximum training epochs	400
Batch size	64
Optimizer	AdamW
Learning rate / weight decay	1.0×10^{-4} / 0.1
Random seed	42
Contrastive temperature	0.006 for both prediction–target and target–target similarities
Loss weights	Smooth L1 loss: 1.0; soft contrastive loss: 1.0
Evaluation checkpoint	Model weights saved after 400 training epochs

For response regrouping, left- and right-hemisphere response rows are pooled within each scale, permuted once per subject and scale, and split back using the original branch sizes. The permutation is shared across samples and fixed throughout training; positional encodings remain at the original branch positions. Because this operation changes both hemispheric assignment and response-to-position correspondence, it tests the original input organization rather than hemispheric identity in isolation.

For target-pairing permutation, training target-feature tensors are permuted across samples using one fixed permutation per subject. Validation retains the correct fMRI–target pairing.

B SUPPLEMENTARY PREDICTION AND CONTROL RESULTS

These analyses supplement the prediction, contralateral-gating, spatial-stability, and gating–zeroing correspondence results in Sections 4.2–4.4. Training settings and control construction are specified in Appendix A; fixed-gate evaluation and image-level inference are detailed in Appendix B.5.

B.1 PREDICTION PERFORMANCE

Table 5 decomposes the final-model validation loss, and Table 6 compares prediction performance with the direct-concatenation reference (Section 4.2).

Table 5: Validation-loss decomposition of the final HemiPatchGate checkpoints on the shared validation set.

Subject	Epochs	Smooth L1	Soft contrastive	Total loss
sub01	400	0.315884	0.000538	0.316422
sub02	400	0.317913	0.000609	0.318523
sub05	400	0.319475	0.012283	0.331758
sub07	400	0.320473	0.049737	0.370211

Table 6: Prediction performance of HemiPatchGate (HPG) and the direct-concatenation reference (No-gate). The two models use the same data split, hemispheric encoders, normalization, prediction head, loss, optimizer, and training schedule. Top-1 and Top-10 are prediction-to-target retrieval accuracies among the same 1,000 validation images.

Subject	Validation loss		Top-1		Top-10	
	HPG	No-gate	HPG	No-gate	HPG	No-gate
sub01	0.316422	0.315602	66.6%	66.7%	92.2%	93.3%
sub02	0.318523	0.318137	60.5%	61.5%	90.4%	91.4%
sub05	0.331758	0.325090	47.3%	46.9%	85.8%	87.7%
sub07	0.370211	0.335365	41.5%	38.1%	79.1%	76.4%
Mean across subjects	0.334228	0.323549	54.0%	53.3%	86.9%	87.2%

B.2 CONTRALATERAL INPUT-ZEROING EFFECTS

Table 7 expands the recomputed-gate input-zeroing results summarized in Table 2, reporting the loss increase separately for each image half and each zeroed hemisphere. With the original gates held fixed, the contralateral-minus-ipsilateral contrasts are 0.010282, 0.010823, 0.003886, and 0.003995 for left-input zeroing, and 0.006547, 0.005617, 0.002924, and 0.004394 for right-input zeroing, in subject order sub01, sub02, sub05, sub07. All eight fixed-gate contrasts are positive. The full-input baseline, target features, and patchwise loss definition are unchanged; the fixed-gate procedure is described in Appendix B.5.1.

Table 7: Complete per-image-half results for hemisphere-specific fMRI input zeroing. Values are increases in Smooth L1 loss relative to the original input. When the left-hemisphere input is zeroed, the contralateral contrast is defined as right-half minus left-half; when the right-hemisphere input is zeroed, it is defined as left-half minus right-half.

Subject	**Left input zeroed**			**Right input zeroed**		
	Left half	**Right half**	**Contrast**	**Left half**	**Right half**	**Contrast**
sub01	0.020993	0.029790	0.008796	0.028766	0.022700	0.006066
sub02	0.020811	0.029939	0.009128	0.022518	0.016268	0.006250
sub05	0.022234	0.025851	0.003617	0.017840	0.015114	0.002726
sub07	0.018351	0.022467	0.004116	0.018772	0.015231	0.003541
Mean across subjects	0.020597	0.027012	0.006414	0.021974	0.017328	0.004646

B.3 RETRAINING CONTROLS

Tables 8 and 9 report the per-subject results and cross-subject summary for the two retraining controls in Section 4.3. Control procedures and training settings are given in Appendix A.4.

Table 8: Per-subject validation losses and gating differences for the two retraining controls after 400 retraining epochs. Gating difference is defined as the mean left-branch gating in the right image half minus that in the left image half. Each $\pm$ term is the image-level normal-approximation 95% CI half-width $1.96s/\sqrt{1000}$, where s is the sample standard deviation of the 1,000 validation-image gating differences within the trained subject model. Mean r is the arithmetic mean of the Pearson spatial correlations between the 1,000 image-level M^L maps and the corresponding subject-level gating map.

Subject	**Total loss**	**Smooth L1**	**Soft contrastive**	**Gating difference**	**Mean r**
Response-regrouping control					
sub01	0.320891	0.317671	0.003220	0.005807 ± 0.000250	0.960
sub02	0.321481	0.317990	0.003491	-0.028136 ± 0.000296	0.956
sub05	0.354472	0.319751	0.034721	0.000754 ± 0.000183	0.562
sub07	0.426916	0.321836	0.105080	-0.003235 ± 0.000234	0.664
Target-pairing permutation					
sub01	4.491305	0.342806	4.148499	-0.000021 ± 0.000008	0.666
sub02	4.498462	0.342849	4.155614	-0.000078 ± 0.000005	0.626
sub05	4.497491	0.342703	4.154788	0.000101 ± 0.000010	0.746
sub07	4.497198	0.342781	4.154418	0.000123 ± 0.000004	0.660

Table 9: Cross-subject summary of the two retraining controls after 400 retraining epochs. Mean total validation loss and mean gating difference are arithmetic means of the four subject-level values. The $\bar{\Delta}^L > 0$ column counts subjects with positive gating differences. Mean r is the arithmetic mean across the four subjects of the per-subject mean spatial correlations defined in Table 8.

Control	**Total loss**	**Gating difference**	$\bar{\Delta}^L > 0$	**Mean r**
Response-regrouping control	0.355940	-0.006202	2/4	0.785
Target-pairing permutation	4.496114	0.000031	2/4	0.674

B.4 SPATIAL STABILITY AND INPUT MANIPULATIONS

Figure 4 and Table 10 supplement Section 4.4. They distinguish stability across natural images from sensitivity to simultaneous input zeroing and left–right input swapping. The residual gating-difference ratio under bilateral zeroing describes the remaining readout under that perturbation condition and is not interpreted as a decomposition of the sources of the original gating pattern.

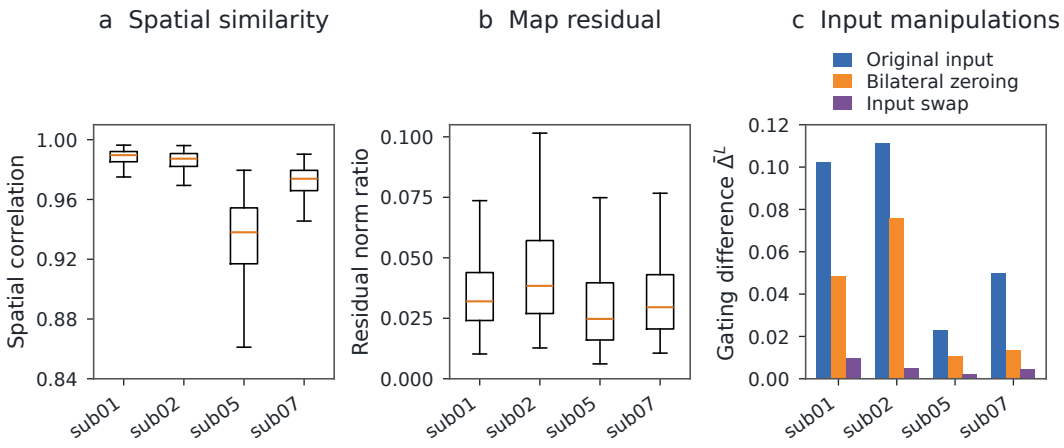

Figure 4: Gating-map stability and changes under input zeroing and swapping. a, Pearson spatial correlation between each image-level left-branch gating map under the original validation input and the leave-one-out mean of the remaining 999 validation-image gating maps. b, Residual-norm ratio $\|M_i^L - \bar{M}_{-i}^L\|_F / \max(\|M_i^L\|_F, 10^{-12})$, using the Frobenius norm. Here M_i^L is the image's channel-averaged gating map, and $\bar{M}_{-i}^L$ averages the other 999 maps from the same subject. c, Subject-level left-branch gating difference under the original input, simultaneous zeroing of both fMRI inputs, and left–right hemisphere input swapping. Box plots in a and b summarize 1,000 images per subject: lines mark medians, boxes span the quartiles, and whiskers reach the most extreme values within 1.5 interquartile ranges of the box; outlier markers are hidden. Zeroing and swapping reduce gating difference, while original-input spatial correlations remain high.

Table 10: Numerical summary of the gating-map diagnostics in Fig. 4. Left-branch gating difference $\bar{\Delta}^L$ is the mean image-level gating difference across the 1,000 validation images for each subject under the corresponding input condition. Ratio is the gating difference after each manipulation divided by that subject's gating difference under the original input. Mean r is computed from the original validation input as the arithmetic mean of the Pearson spatial correlations between the 1,000 image-level M^L maps and the leave-one-out mean of the remaining 999 validation-image gating maps.

Subject	**Original**	**Bilateral zeroing**		**Input swap**		**Mean r**
	$\bar{\Delta}^L$	$\bar{\Delta}^L$	**Ratio**	$\bar{\Delta}^L$	**Ratio**	
sub01	0.102177	0.048438	47.4%	0.009798	9.6%	0.988
sub02	0.111280	0.075671	68.0%	0.005090	4.6%	0.985
sub05	0.023119	0.010624	46.0%	0.001950	8.4%	0.930
sub07	0.049743	0.013570	27.3%	0.004634	9.3%	0.971

B.5 GATING READOUT AND HEMISPHERIC ZEROING SENSITIVITY

B.5.1 FIXED-GATE INPUT ZEROING

For each validation image, we first run the model with both hemispheric inputs intact and cache the complete channel-wise left and right gates in float32. We then zero one hemispheric input at

a time and recompute the branch representations, while reusing the cached gates before weighted concatenation. The prediction head and all other computation remain unchanged. The archived fixed-gate predictions are evaluated against the same targets using Smooth L1 with $\beta = 1$, averaged over feature channels separately at each image–location pair.

Let $e^{0}_{s,i,p}$ denote the patchwise loss under intact inputs and $e^{(h,\mathrm{fixed})}_{s,i,p}$ the loss after zeroing input h while retaining the intact-input gates. We define the loss increase as $\Delta e^{(h,\mathrm{fixed})}_{s,i,p} = e^{(h,\mathrm{fixed})}_{s,i,p} - e^{0}_{s,i,p}$, for $h \in \{L, R\}$; the recomputed-gate condition is defined analogously. As a pipeline check, the recomputed-gate analysis reproduces the mean half-image contrasts of 0.006414 and 0.004646 and the Table 12b vertical-band coefficients of 0.2147, 0.2398, 0.0950, and 0.3810.

B.5.2 Location-demeaned correspondence

Using the notation of Appendix C.3, the intact-input gating contrast is $G_{s,i,p} = \mathrm{mean}_c(w^{L}_{s,i,p,c} - w^{R}_{s,i,p,c}) = 2M^{L}_{s,i,p} - 1$. For fixed-gate zeroing, $S^{\mathrm{fixed}}_{s,i,p} = \Delta e^{(L,\mathrm{fixed})}_{s,i,p} - \Delta e^{(R,\mathrm{fixed})}_{s,i,p}$. Positive S indicates greater loss after zeroing the left input than after zeroing the right input. We subtract the mean over the other 999 images at the same location from each quantity, obtaining $\delta G_{s,i,p} = G_{s,i,p} - \mathrm{mean}_{j \neq i}\, G_{s,j,p}$ and $\delta S^{\mathrm{fixed}}_{s,i,p} = S^{\mathrm{fixed}}_{s,i,p} - \mathrm{mean}_{j \neq i}\, S^{\mathrm{fixed}}_{s,j,p}$. For each subject, we compute Pearson correlation over all matched image–location pairs. This is a pooled association, not a separate significance test at every patch.

Confidence intervals use 2,000 image-level paired bootstrap resamples. Each sampled image carries all 256 locations and both zeroing conditions together; location-wise means are re-estimated within each resample. This preserves the image as the resampling unit rather than treating 256,000 image–location pairs as independent observations. The same sampled images are used to estimate the fixed-minus-recomputed correlation difference. Table S1 reports fixed- and recomputed-gate correlations and their paired differences.

The fixed-gate and recomputed-gate correlations are positive in all four subjects. Paired fixed-minus-recomputed differences are 0.001 [–0.006, 0.009], 0.037 [0.030, 0.045], 0.023 [0.012, 0.034], and 0.027 [0.019, 0.035] for sub01, sub02, sub05, and sub07, respectively (95% confidence intervals). Thus, the correspondence persists without post-zeroing gate reassignment; the paired difference excludes zero in sub02, sub05, and sub07 but not in sub01.

Table S1: Paired comparison of gating–zeroing correspondence. Brackets denote image-level bootstrap 95% confidence intervals.

Subject	**Fixed r [95% CI]**	**Recomputed r [95% CI]**	**Fixed minus recomputed [95% CI]**
sub01	0.234 [0.203, 0.264]	0.233 [0.203, 0.262]	0.001 [–0.006, 0.009]
sub02	0.354 [0.325, 0.382]	0.318 [0.287, 0.346]	0.037 [0.030, 0.045]
sub05	0.126 [0.089, 0.160]	0.103 [0.067, 0.138]	0.023 [0.012, 0.034]
sub07	0.202 [0.169, 0.235]	0.175 [0.139, 0.211]	0.027 [0.019, 0.035]

B.5.3 Additional spatial-template removal

A stricter analysis additionally removes each image's overall offset and left–right hemifield pattern. Under this specification, fixed-gate correlations range from 0.064 to 0.164, with positive point estimates for all four subjects. The reduction indicates sensitivity to the spatial components retained in the analysis. We therefore distinguish correspondence after location-wise centering from correspondence after additional removal of image-level spatial patterns, rather than treating them as interchangeable estimates.

B.5.4 Scope of the perturbation check

Holding the gates fixed prevents their reassignment after zeroing, while the cached gates are computed from both intact hemispheric inputs. Input zeroing replaces natural responses with an artificial

baseline, so this analysis evaluates the correspondence between the gating readout and perturbation sensitivity within the fitted decoder. The same 1,000 validation images are reused, making this a supplementary within-model analysis rather than an independent-sample confirmation.

C STIMULUS-RELATED MODULATION ANALYSES

This appendix details the whole-image analysis in Section 4.5 (C.1) and the local analyses in Section 4.6 (C.2–C.6), with primary results preceding sensitivity analyses.

C.1 WHOLE-IMAGE GATE SHIFTS

We relate animate coverage (the image fraction occupied by people and animals in COCO panoptic segmentation) to whole-image gate shift: the mean left-branch weight across patches minus its subject-specific validation-set mean. Positive and negative values indicate shifts toward the left and right branches, respectively, relative to this subject-specific mean (Section 4.5).

Scene-category assignment. We assign scene-category labels using OpenAI CLIP ViT-L/14 with its pre-trained weights and standard image preprocessing. The 11 labels are architecture/building, indoor room, person/people, animal, vehicle/transport, food/drink, sport/action, nature/landscape, water/snow, plant/flower, and miscellaneous object. Each label has four text prompts. After L2-normalizing the image and text embeddings, we scale their cosine similarities by 100 and apply softmax across the 44 prompts. We average the four prompt probabilities for each label and assign the label with the highest mean probability. The resulting image labels are shared across subjects. Exact prompts and the screened candidate list will be provided with the released code.

Association test. We compute partial correlations after controlling for total foreground coverage and scene category. Significance is assessed with a two-sided within-category permutation test using 50,000 permutations per subject (all subjects $p < 0.00002$). Figure 5 shows the image-level relationship; Figure 6a and Table 11 in Appendix C.2 summarize the partial correlations across subjects.

Spatial control analysis. To assess sensitivity to the horizontal spatial distribution of animate regions, we performed an additional control analysis using the same animate masks. We extracted two spatial covariates: left–right imbalance of animate pixels, defined as $(A_{\text{right}} - A_{\text{left}})/(A_{\text{right}} + A_{\text{left}})$, and the horizontal centroid of animate regions. We added these variables to the regression model together with the original controls. The standardized animate-coverage coefficient changed by -3.4% to +5.3% across subjects after this adjustment (Table S2), showing robustness to these two measured spatial summaries; other aspects of horizontal placement were not tested.

Table S2: Spatial controls for the animate-coverage association with whole-image gate shift. M0: animate coverage plus total foreground and scene-category controls. M1: M0 plus left–right animate-mask imbalance and normalized horizontal centroid. Values are standardized animate-coverage beta and normal-approximation p (1,000 images per subject).

Subject	M0 β	M0 p	M1 β	M1 p	Change in $\lvert\beta\rvert$ (%)
sub01	-0.235	4.0×10^{-9}	-0.248	1.2×10^{-10}	+5.3%
sub02	-0.272	2.2×10^{-12}	-0.285	1.1×10^{-14}	+4.5%
sub05	-0.259	3.7×10^{-11}	-0.251	1.1×10^{-10}	-3.4%
sub07	-0.244	1.6×10^{-9}	-0.247	1.1×10^{-9}	+1.0%

For a sensitivity analysis, we excluded the 348 images without animate pixels (n=652). The changes in $|\beta|$ after adding the two spatial covariates were +17.8%, +18.4%, -5.7%, and +3.7% (sub01/02/05/07); corresponding M1 normal-approximation p-values were 4.33×10^{-4}, 5.26×10^{-5}, 5.12×10^{-6}, and 3.92×10^{-3}, respectively. Change = $(|\beta_{M1}| - |\beta_{M0}|)/|\beta_{M0}| \times 100\%$.

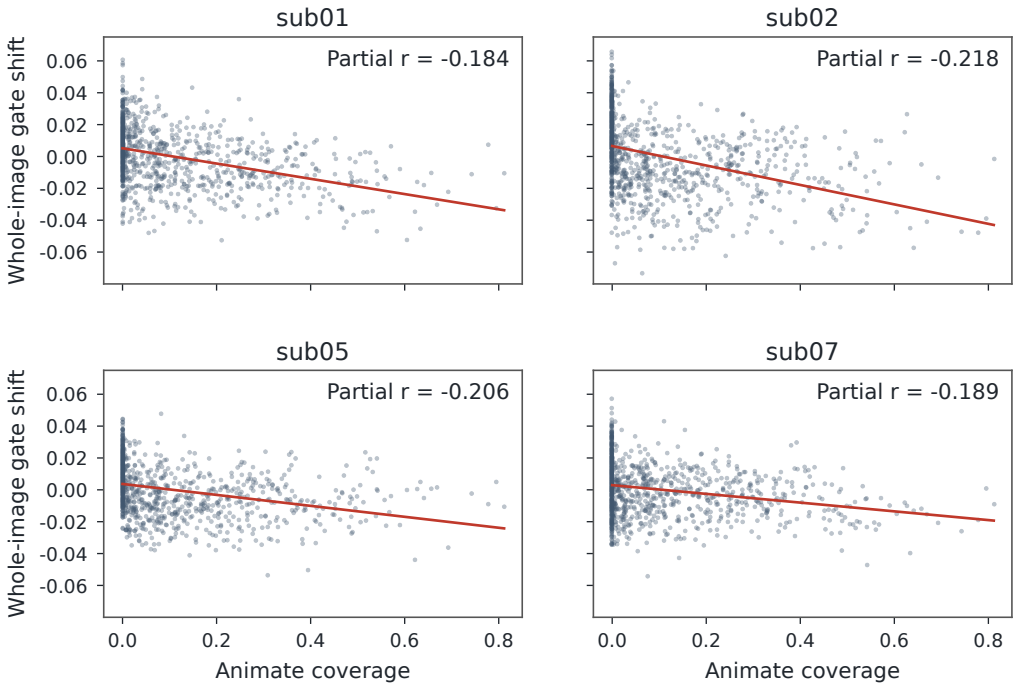

Figure 5: Whole-image gate shift versus animate coverage across 1,000 validation images. Each panel corresponds to one independently trained subject model; each point is one validation image. The x-axis shows animate coverage and the y-axis whole-image gate shift. The line is an ordinary least-squares fit; the annotated value is the partial correlation after controlling for total foreground coverage and scene category (Table 11). Whole-image gate shift decreases with increasing animate coverage in all four subjects.

C.2 SPATIAL VARIATION IN GATING-CHANGE MAGNITUDE

The unsigned analysis in Section 4.6 asks where gates vary most across images. At each patch and channel, we subtract the mean left-branch weight over the other 999 validation images from the current weight, take the absolute difference, and average across the 512 channels. The vertical-midline band comprises columns 7–10; the peripheral reference comprises columns 1–4 and 13–16. Figure 6b and Table 11 report the percentage increase in this gating-change magnitude inside the band relative to the peripheral reference.

To distinguish proximity to the vertical midline from general center proximity, we jointly regress gating-change magnitude on vertical-midline proximity, horizontal-midline proximity, and local foreground coverage. Each image–location pair is one observation. Before fitting, all predictors and the outcome are z-scored across image–location observations. Each coefficient therefore represents the change in the outcome, measured in standard deviations, associated with a one-standard-deviation increase in the predictor, holding the other predictors fixed. Standard errors are cluster-robust with image as the clustering unit.

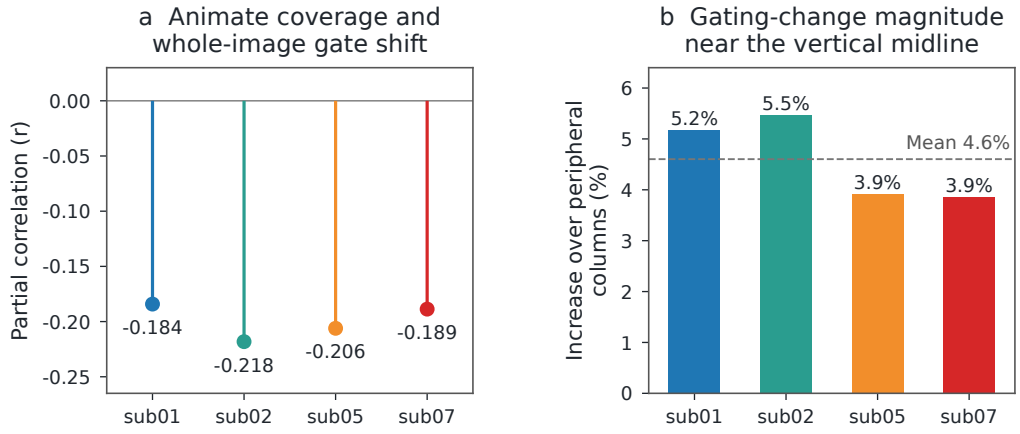

Figure 6: Per-subject statistics for stimulus-related modulation analyses. a, Partial correlation between animate coverage and whole-image gate shift for each subject, controlling for total foreground coverage and scene category. b, Percentage increase in gating-change magnitude (defined in Section 4.6) in the vertical-midline band (columns 7–10) relative to peripheral columns 1–4 and 13–16; the dashed line marks the mean across the four subjects (4.6%). Both effects have consistent directions across all four subjects.

Table 11: Whole-image content associations and spatial variation (Appendices C.1–C.2). Animate: partial correlation between animate coverage and whole-image gate shift, controlling for total foreground coverage and scene category. Band increase: percentage increase in gating-change magnitude in columns 7–10 relative to peripheral columns 1–4 and 13–16. Vertical and horizontal proximity: standardized coefficients and p-values from the joint spatial regression, with image-clustered standard errors.

Subject	**Animate**	**Band**	**Vertical proximity**		**Horizontal proximity**	
	Partial r	**increase**	β	p	β	p
sub01	-0.184	5.2%	0.115	1.0×10^{-222}	-0.128	8×10^{-113}
sub02	-0.218	5.5%	0.110	6.1×10^{-188}	-0.163	3×10^{-136}
sub05	-0.206	3.9%	0.108	1.1×10^{-151}	-0.030	4×10^{-7}
sub07	-0.189	3.9%	0.105	1.2×10^{-175}	-0.040	1×10^{-12}

C.3 LOCAL-CONTENT SCREENING

This analysis identifies the foreground predictor used in the signed local regressions of Appendix C.4.

C.3.1 CANDIDATES AND SCREENING OUTCOMES

We split the 1,000 images into 500 discovery and 500 confirmation images (seed 20260811). Candidates comprise COCO category and supercategory coverage plus four aggregates: all foreground objects, all background regions, animate entities (person and animal), and natural background (plant, sky, ground, and water). To ensure sufficient coverage in both splits, we retain candidates that meet two criteria in each split: at least 25 images with summed patch coverage above 0.25, and total coverage of at least 40 patch-equivalents. One patch-equivalent is the area of one fully covered patch. This yields 48 category or supercategory measures and four aggregates.

For subject s, image i, and patch p, we use two screening outcomes. The first is the left-minus-right gating contrast $G_{s,i,p} = 2M^L_{s,i,p} - 1$. The second is the difference between the loss increases after left- and right-hemisphere input zeroing, $S_{s,i,p} = \Delta\ell^L_{s,i,p} - \Delta\ell^R_{s,i,p}$, where $\Delta\ell^h$ is the patch-level loss increase after zeroing the fMRI input from hemisphere h. Positive values of S indicate a larger loss increase for left-input zeroing.

For each candidate coverage map F, we construct the predictor $X_{i,p} = r_p F_{i,p}$, where $r_p = -1$ in the left image half and +1 in the right half. With both outcomes expressed as left minus right, this sign convention gives the two image halves a common interpretation: positive and negative screening slopes indicate contralateral and ipsilateral associations, respectively.

Before fitting, we subtract the discovery-set mean at each patch from both predictors and outcomes. We then remove their projections onto the spatial basis $\{1, r_p, x_p, y_p, x_p y_p, x_p^2, y_p^2\}$ within each image; row and column coordinates span [-1, 1]. The transformed variables are scaled by their discovery-set standard deviations. For the group analysis, we average the four scaled subject outcomes.

C.3.2 DISCOVERY AND CONFIRMATION

For each candidate, we regress the transformed outcome $\tilde{Y}$ on the transformed coverage predictor $\tilde{X}$. The screening slope is

$$\hat{\beta} = \frac{\sum_{i,p} \tilde{X}_{i,p}\tilde{Y}_{i,p}}{\sum_{i,p} \tilde{X}^2_{i,p}}.$$

We assess the association at the image level using the contribution of each image to the numerator:

$$q_i = \sum_p \tilde{X}_{i,p}\tilde{Y}_{i,p}.$$

For the gating outcome, discovery used a two-sided normal-approximation test based on these image-level contributions. The 22 candidates with p < 0.01 and matching effect directions in at least three subjects advanced to confirmation.

For each outcome, confirmation used 20,000 image-level sign flips of q_i (seed 20261802), with Holm correction across the 22 candidates. A candidate passed confirmation if the gating-effect sign matched across the two splits and the slope signs matched between the two outcomes. Each outcome also had to show the same effect direction in at least three subjects and meet the adjusted p < 0.05 threshold. Nineteen candidates met these criteria.

C.3.3 SELECTED PREDICTOR

Local foreground coverage had the smallest discovery-stage p-value among all 52 candidates (β = 0.159, p = 2.1×10^{-104}). It passed confirmation for both the gating contrast (β = 0.171, Holm-adjusted p = 0.0011) and the loss-increase difference (β = 0.200, Holm-adjusted p = 0.0011), with

all four subjects showing the same effect direction. The broader screening results distinguished foreground from background measures, motivating the selection of local foreground coverage as the general foreground predictor for the vertical-midline analysis. We kept this predictor fixed when refitting the regressions on all 1,000 validation images. Figure 7 provides a descriptive comparison of image content (Appendix C.6).

C.4 Foreground-related ipsilateral shifts and input zeroing

The primary analysis compares the foreground-coverage slopes inside and outside the vertical-midline band for two outcomes: ipsilateral gate shift and the loss increase after ipsilateral input zeroing (Section 4.6). Foreground comprises independently segmentable object categories in COCO panoptic segmentation.

C.4.1 Regression model

For each subject s, we fit the following model across images i and patch locations p:

$$\begin{aligned} U_{s,i,p} = a_{s,i} + d_{s,p} + b_{V,s} F_{i,p} V_p + b_{H,s} F_{i,p} H_p \\ + \eta_s F_{i,p} + \boldsymbol{\gamma}_s^\top \boldsymbol{C}_{i,p} + \boldsymbol{\theta}_s^\top \boldsymbol{Q}_{i,p} + \kappa_s B_{i,p} + \epsilon_{s,i,p}. \end{aligned} \tag{5}$$

Here U is ipsilateral gate shift or the loss increase after zeroing the ipsilateral fMRI input, and F is local foreground coverage. V indicates whether a patch lies in the vertical band (columns 7–10), and H indicates whether it lies in the horizontal band (rows 7–10). Unlike the screening contrast, this zeroing outcome uses only the ipsilateral input condition at each patch. Image and patch location effects are represented by a and d. C contains the 80 foreground-category coverage fractions. Q contains instance-mask coverage and the coverage-weighted mean area, width, and height of instances overlapping each patch. B is original-input prediction loss averaged across subjects and is included only for the zeroing outcome. Geometric controls and B are standardized.

We demean all variables within each image, remove least-squares projections onto the control space, and estimate the two interaction coefficients from the residuals, using generalized inverses for redundant control columns. The coefficient b_V measures the foreground slope difference inside versus outside the vertical band, and $b_V - b_H$ compares the vertical and horizontal differences. Standard errors are clustered by image.

C.4.2 Coefficient scaling and group-mean fits

For the foreground-coverage interaction analyses, each reported outcome-standardized coefficient is the raw coefficient divided by the standard deviation of the within-image-demeaned outcome. The interaction predictors retain their original units.

For the group-mean local-regression fits, we first average the outcome across the four subjects at each image–location pair and then refit the model. The resulting p-values quantify image-level evidence for this averaged outcome; they do not support population-level inference across subjects.

C.4.3 Primary results

Table 12a reports the ipsilateral gate-shift results. The foreground-slope difference inside minus outside the vertical band is positive and statistically significant in all four subjects and in the group-mean fit. The vertical-minus-horizontal contrast is also positive and significant, whereas the horizontal inside-minus-outside slope difference is not significant. For ipsilateral input zeroing (Table 12b), the vertical inside-minus-outside slope difference and the vertical-minus-horizontal contrast are positive and statistically significant in all four subjects and in the group-mean fit.

Table 12: Foreground-related ipsilateral gate shifts and input-zeroing sensitivity.

(a) Ipsilateral gate shift

Subject	Vertical		Horizontal		Difference	
	β	p	β	p	β	p
sub01	0.211	4.3×10^{-21}	0.001	0.98	0.210	1.1×10^{-12}
sub02	0.231	2.8×10^{-26}	-0.024	0.25	0.255	6.8×10^{-17}
sub05	0.129	1.7×10^{-10}	0.033	0.098	0.096	4.4×10^{-4}
sub07	0.080	4.5×10^{-4}	-0.034	0.090	0.114	1.6×10^{-4}
Group mean	0.256	1.0×10^{-28}	-0.012	0.60	0.268	7.2×10^{-17}

(b) Loss increase after ipsilateral input zeroing

Subject	Vertical β	Horizontal β	Difference
sub01	0.215	-0.021	0.236
sub02	0.240	-0.041	0.281
sub05	0.095	-0.095	0.190
sub07	0.381	-0.045	0.426
Group mean	0.343	-0.076	0.419

Vertical and horizontal β are the foreground-coverage slope differences inside minus outside columns 7–10 and rows 7–10, respectively; Difference is vertical minus horizontal β. Coefficients are divided by the standard deviation of each within-image-demeaned outcome and are not directly comparable in magnitude between panels. Tests use 1,000 validation images per subject and image-clustered standard errors. In (b), all vertical differences have $p < 10^{-5}$, and all vertical-minus-horizontal contrasts have $p < 10^{-9}$. The group-mean fit averages outcomes across subjects before refitting; it is an image-level analysis, not a population-level test across subjects. Regression controls are specified in Appendix C.4; the zeroing model additionally controls for original-input prediction loss.

C.5 Sensitivity analyses

These analyses examine slopes within the vertical-midline band, sensitivity to the binary-band definition, and the effect of holding the original gates fixed during input zeroing. They complement the primary inside-minus-outside slope comparisons in Appendix C.4.

C.5.1 Within-band slopes

We examine the foreground-coverage slope within the vertical-midline band itself, rather than the difference between the inside- and outside-band slopes. For this analysis, we fit a model without the collinear category-coverage main effects, retaining the other controls. For ipsilateral gate shift, the within-band slope is positive in all four subjects (raw slopes 0.0020–0.0063), significant for sub01 (p = 0.039) and sub02 (p = 2.8×10^{-4}), and positive and significant in the group-mean fit (p = 7.7×10^{-4}). For the loss increase after zeroing the ipsilateral input, the within-band slope is positive and significant in all four subjects (raw slopes 0.0059–0.0102; all p $\leq 3.8 \times 10^{-4}$) and in the group-mean fit (p = 5.0×10^{-12}). The vertical-midline interaction estimates remain close to those from the primary specification.

C.5.2 Continuous midline proximity

To test sensitivity to the fixed-band definition, we refit the zeroing analysis with the same controls but replace the binary vertical-midline and horizontal-midline bands with continuous proximity measures, defined for a 0-based row or column coordinate u as 1 - |u - 7.5|/7.5. The interaction between local foreground coverage and vertical-midline proximity is positive and significant in all four subjects (outcome-standardized coefficients 0.317, 0.329, 0.158, and 0.578 for sub01, sub02, sub05,

and sub07; all p $\leq 1.1 \times 10^{-5}$) and in the group-mean fit (0.509, p = 4.4×10^{-41}). The corresponding vertical-minus-horizontal interaction contrasts are also positive and significant in all four subjects (0.234–0.735; all p $\leq 5.6 \times 10^{-6}$) and in the group-mean fit (0.625, p = 7.9×10^{-30}).

C.5.3 Fixed-gate zeroing near the vertical midline

We repeat the ipsilateral zeroing regression of Appendix C.4 using fixed-gate loss increases as the outcome. The image and location effects, 80 foreground-category coverage fractions, total foreground coverage, instance-geometry controls, group-averaged original-input loss, and image-clustered standard errors remain unchanged. The vertical band comprises columns 7–10 and the horizontal band rows 7–10. The only change is whether gates are recomputed after zeroing or retained from the intact-input forward pass.

The outcome-standardized vertical inside-minus-outside slope differences are 0.174 (p = 3×10^{-17}), 0.159 (p = 2×10^{-16}), 0.049 (p = 0.013), and 0.288 (p = 9×10^{-39}), in subject order sub01, sub02, sub05, and sub07. The corresponding vertical-minus-horizontal contrasts are 0.224, 0.176, 0.189, and 0.389, all positive (Table S3). The group-mean outcome is formed before refitting, as in Appendix C.4.2; its vertical slope difference is 0.252 (p = 2×10^{-33}) and its vertical-minus-horizontal contrast is 0.367 (p = 5×10^{-33}). These group-mean coefficients describe image-level evidence for the averaged outcome, not population-level inference across subjects.

Table S3: Fixed-gate ipsilateral zeroing: foreground slope contrasts near the midlines. Coefficients are outcome-standardized as in Appendix C.4.2; p-values use image-clustered standard errors. Group-mean outcomes are averaged before refitting.

Subject	**Vertical**	**p**	**Vertical – horizontal**	**p**
sub01	0.174	3×10^{-17}	0.224	3×10^{-14}
sub02	0.159	2×10^{-16}	0.176	1×10^{-9}
sub05	0.049	0.013	0.189	1×10^{-10}
sub07	0.288	9×10^{-39}	0.389	2×10^{-36}
Group mean	0.252	2×10^{-33}	0.367	5×10^{-33}

The fixed-gate analysis preserves the direction of the vertical inside-minus-outside and vertical-minus-horizontal contrasts. These contrasts quantify differences in foreground-related associations across locations; they do not represent a decomposition of the original effect into contributions from gate reassignment.

C.6 Descriptive content comparison

Figure 7 compares image content between stimuli with larger and smaller increases in unsigned gating-change magnitude near the vertical midline (Appendix C.2). This descriptive comparison complements the local-content screening in Appendix C.3.

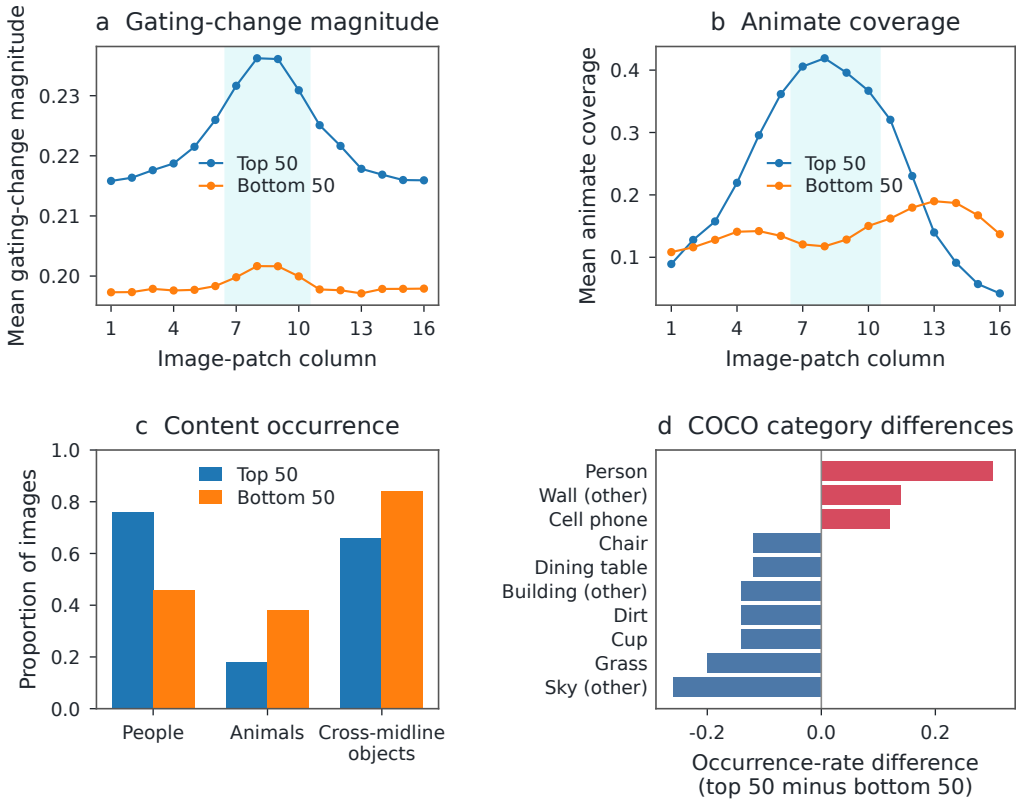

Figure 7: Descriptive content comparison of the top 50 and bottom 50 images ranked by the increase in gating-change magnitude near the image's vertical midline. For each image, this increase is the mean gating-change magnitude (Section 4.6) in columns 7–10 minus that in peripheral columns 1–4 and 13–16, averaged across the four subjects before ranking. a, Gating-change magnitude across image columns; b, animate coverage (person and animal regions) across image columns. Curves in a and b average each column over its 16 rows and the 50 images in each group; a additionally averages across the four subjects. Light blue shading marks columns 7–10. c, Occurrence rates of people, animals, and cross-midline objects; d, occurrence-rate differences (top 50 minus bottom 50) for selected COCO categories.

D SPATIAL SPECIFICITY OF THE CLIP TARGET TOKENS

Although CLIP patch tokens incorporate contextual information through ViT self-attention, their spatial specificity is important for interpreting position-aligned prediction. We assessed this sensitivity by re-extracting tokens for the 1,000 shared validation images using the same OpenCLIP ViT-bigG-14 model configuration and preprocessing used for target-feature extraction. For each image, we selected 16 patch locations at random and independently zeroed one 14×14-pixel patch per occlusion trial (16,000 trials total), then measured cosine-distance changes at all 256 tokens. The token at the occluded location showed the largest change in 53.8% of trials (chance: 0.39%) and ranked among the 10 most changed tokens in 99.5% of trials. These results support measurable spatial sensitivity at the corresponding token location under this occlusion analysis. They do not establish a spatial-distance decay profile or imply that tokens are purely local.

These results do not imply that CLIP tokens are purely local representations, as self-attention introduces interactions among image locations. Instead, they support the use of spatially aligned CLIP targets for position-level analyses in HemiPatchGate.