MindSimulator

Exploring Brain Concept Localization via Synthetic fMRI

Guangyin Bao[1], Qi Zhang[1], Zixuan Gong[1], Zhuojia Wu[1], Duoqian Miao[1*]

[1] Tongji University

Introduction

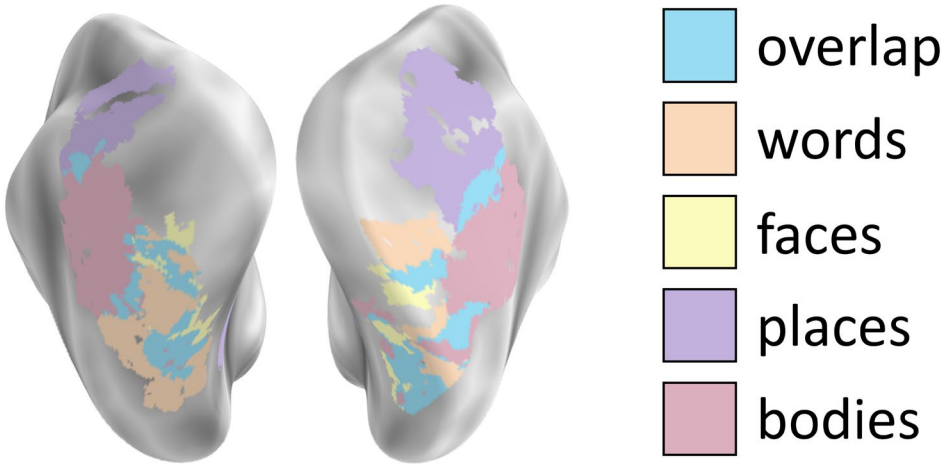
Goal of Neuroscience

The human brain's visual cortex plays a decisive role in processing and perceiving visual information.
Neuroscience has long been dedicated to uncovering the mechanisms of vision in the brain.

Neuroscience
&
Cognitive Science

Brain's Concept Selectivity

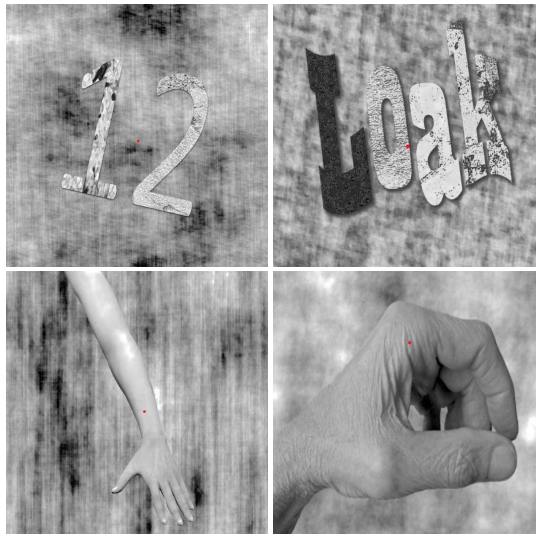
In the higher visual cortex, specific regions demonstrate selectivity for broad concepts such as words, faces, places, and bodies.
When visual stimuli corresponding to these concepts are received, these regions become activated.
Identifying these regions helps unravel the mysteries of visual cognition formation.



Introduction

How to Identify Concept-Specific Regions?
—— Functional Localizer Experiment (fLoc)

- Manually construct a set of visual stimuli associated with specific visual concepts (see right image).
- Use advanced and costly equipment to collect corresponding fMRI data.
- Perform significance testing and statistical analysis.

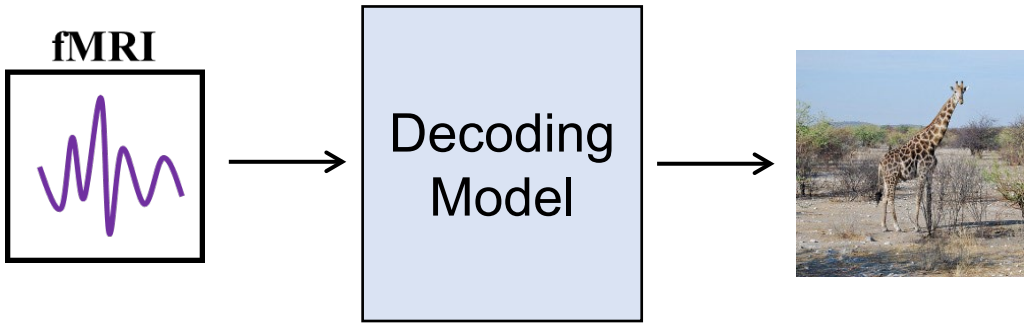


Limitations of the fLoc Experiment

- Scarcity of real fMRI data: the diversity of concept localization is limited.
- Human bias in stimulus set: manually chosen visual stimuli may not reflect the universality of visual concepts.
- Effectiveness: the effectiveness of concept localization remains controversial.

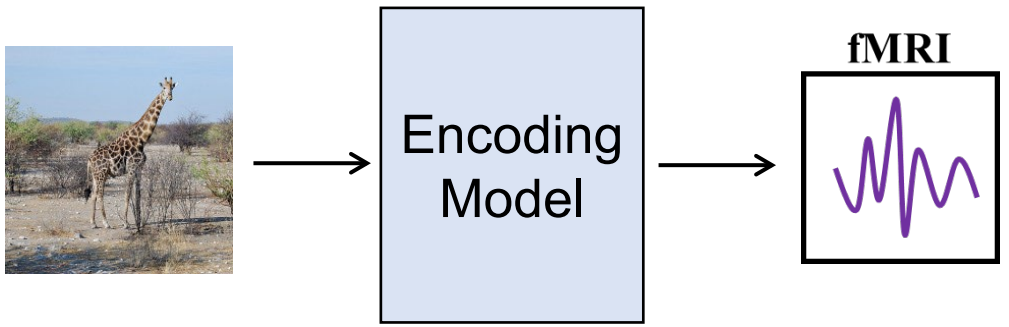
So, exploring Concept Localization with Synthetic fMRI!

Brain Visual Decoding and Encoding



fMRI Decoding ✓

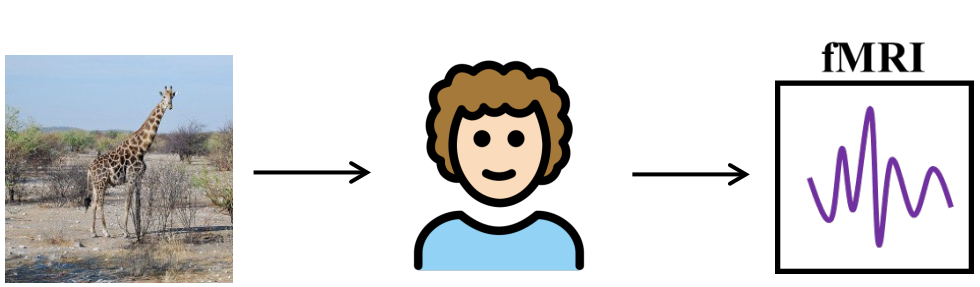
(With the help of deep learning, great success has been achieved.)



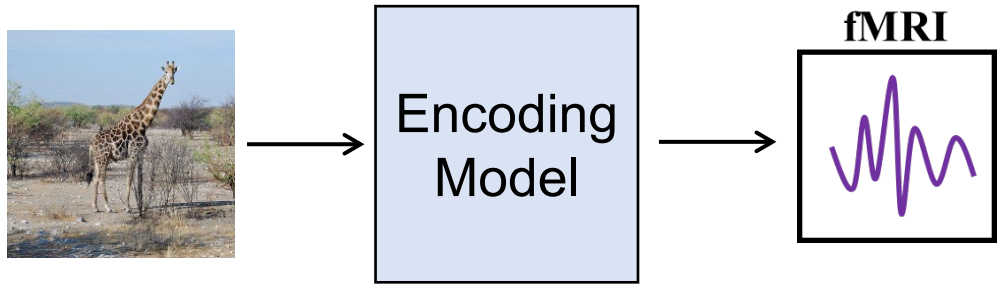
Inverse process: **fMRI Encoding** ?

- fMRI decoding and encoding are dual processes.
- With the help of deep learning, fMRI decoding models have made significant progress.
- Similarly, with multimodal learning and generative models, it is possible to develop powerful fMRI encoding models.

fMRI Encoding Models Simulate Human Brain



Human Visual Process



Simulated Visual Process

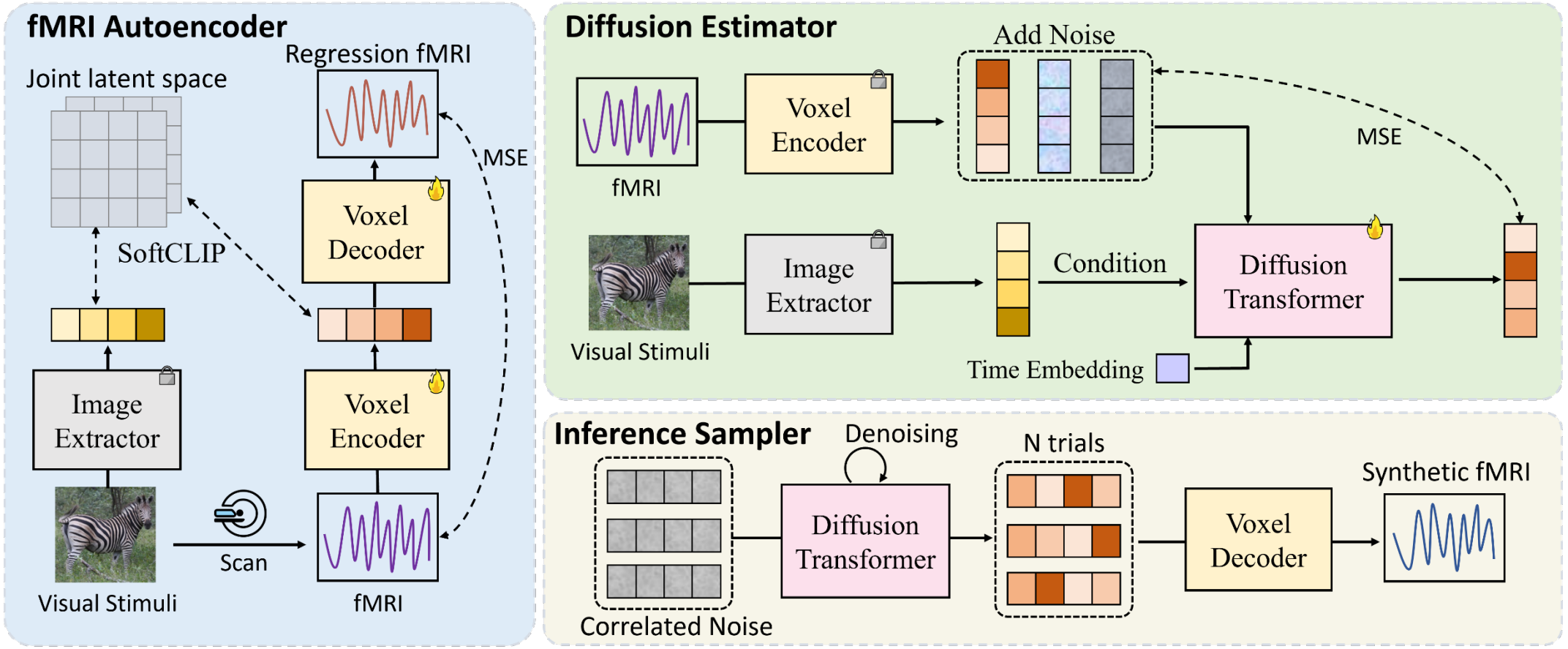
- Once a high-precision fMRI encoding model is obtained, it will simulate the human visual process.
- Given any visual stimulus (image), the model can generate the corresponding fMRI.
- The role of synthetic data:

  1) To augment rare fMRI data.

  2) To facilitate exploration in neuroscience and break through the limitations of conceptual categories.

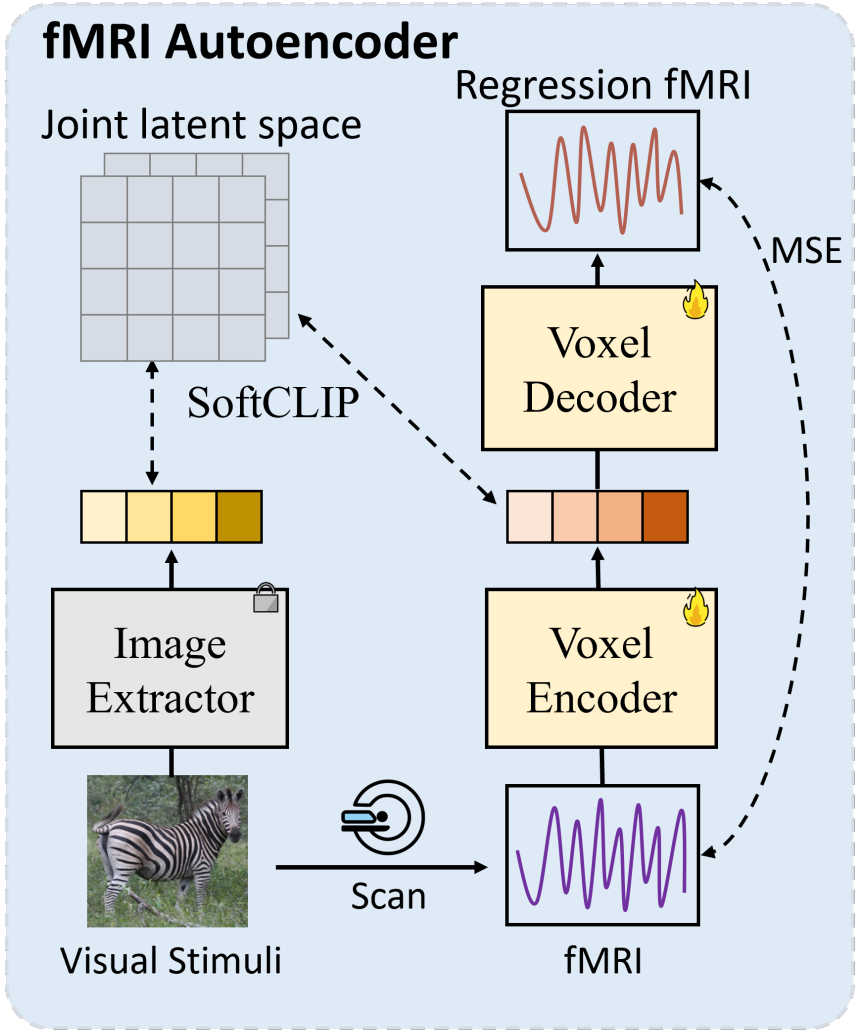
  3) To provide hypothesis constraints for experimental research in neuroscience.

Generative fMRI Encoding Model



Composition: 1) fMRI Autoencoder; 2) Diffusion Estimator; 3) Inference Sampler

MindSimulator: fMRI Autoencoder



Motivation

fMRI data with low signal-to-noise ratio and low dimensionality are difficult to generate directly, therefore, an autoencoder is used to construct the fMRI representation space.

Methods

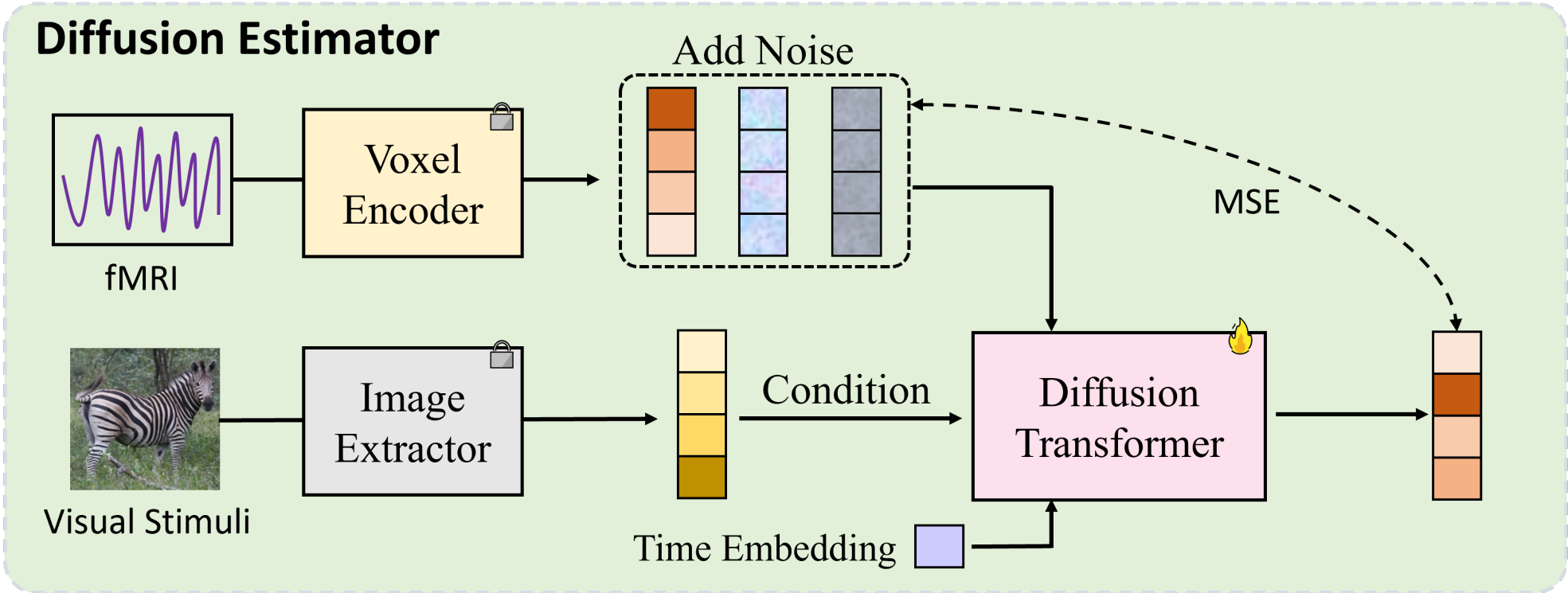
1.Voxel Autoencoder

$$\mathcal{L}_{\text{mse}} = \mathbb{E}_{x\sim\mathcal{S}}||x,\hat{x}||_2^2$$

2.fMRI-Stimuli Joint Latent Space

$$\mathcal{L}_{\text{softclip}} = -\frac{1}{|\mathcal{S}|}\sum_{i=1}^{|\mathcal{S}|}\sum_{j=1}^{|\mathcal{S}|}\left[\frac{\exp(\mathcal{X}_i\cdot\mathcal{X}_j/\tau)}{\sum_{k=1}^{|\mathcal{S}|}\exp(\mathcal{X}_i\cdot\mathcal{X}_k/\tau)}\cdot\log\left(\frac{\exp(\mathcal{Y}_i\cdot\mathcal{X}_j/\tau)}{\sum_{k=1}^{|\mathcal{S}|}\exp(\mathcal{Y}_i\cdot\mathcal{X}_k/\tau)}\right)\right]$$
$$-\frac{1}{|\mathcal{S}|}\sum_{i=1}^{|\mathcal{S}|}\sum_{j=1}^{|\mathcal{S}|}\left[\frac{\exp(\mathcal{Y}_i\cdot\mathcal{Y}_j/\tau)}{\sum_{k=1}^{|\mathcal{S}|}\exp(\mathcal{Y}_i\cdot\mathcal{Y}_k/\tau)}\cdot\log\left(\frac{\exp(\mathcal{X}_i\cdot\mathcal{Y}_j/\tau)}{\sum_{k=1}^{|\mathcal{S}|}\exp(\mathcal{X}_i\cdot\mathcal{Y}_k/\tau)}\right)\right].$$

MindSimulator: Diffusion Estimator

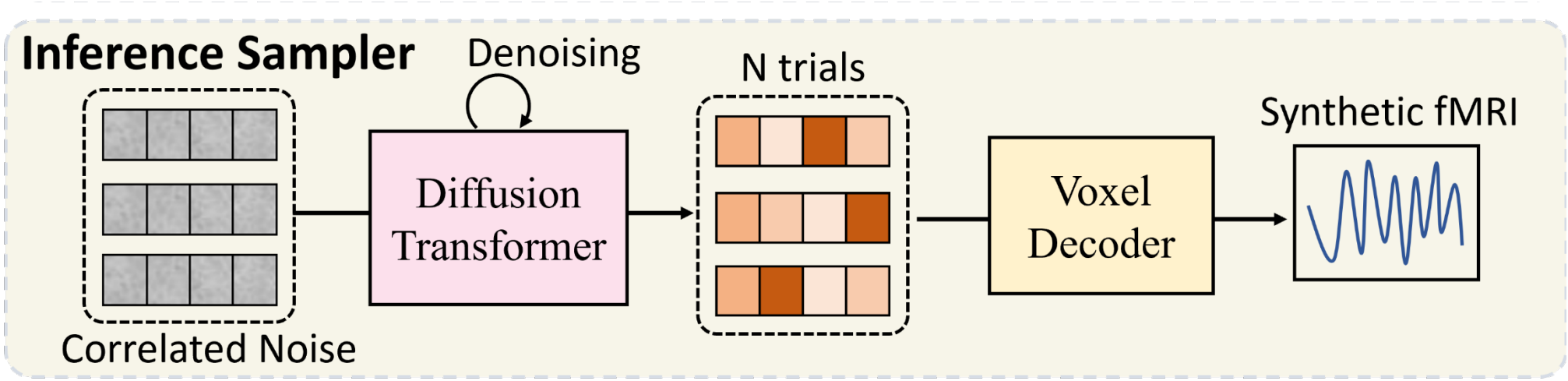


Motivation Learning the conditional distribution of fMRI representations given a visual stimulus.

Methods In the fMRI-Stimuli joint latent space, the diffusion model (DiT) is used to fit the conditional distribution:

$$\mathcal{L}_{\text{diffusion}} = \mathbb{E}_{\epsilon,t,(x,y)\sim\mathcal{S}}[\ ||\mathcal{P}(\mathcal{Z}_t^{\mathcal{X}}, \mathcal{Y}, \mathcal{T}_t) - \mathcal{X}||_2^2\ ]$$

MindSimulator: Inference Sampler



Motivation

Synthesize high-fidelity fMRI signals.

Methods

Multiple enhancements: fully considering the reproducibility of each voxel, integrate multiple generated fMRI results to improve prediction accuracy.

Correlated noise: Use correlated noise to increase the similarity of each generated result.

$$\epsilon_n = \sqrt{\beta_n} \cdot \epsilon_1 + \sqrt{1-\beta_n} \cdot \epsilon_2, \quad n \in [1, 2, \cdots, N]$$

Evaluation Metrics

What makes a good synthetic fMRI?

1. Existing evaluation metrics focus on voxel-level accuracy, such as MSE, Cosine Similarity, Pearson correlation, etc.
2. Fine-grained metrics often fail to account for global accuracy (for example, as illustrated with the following images).

When generating two reconstructions for an image:

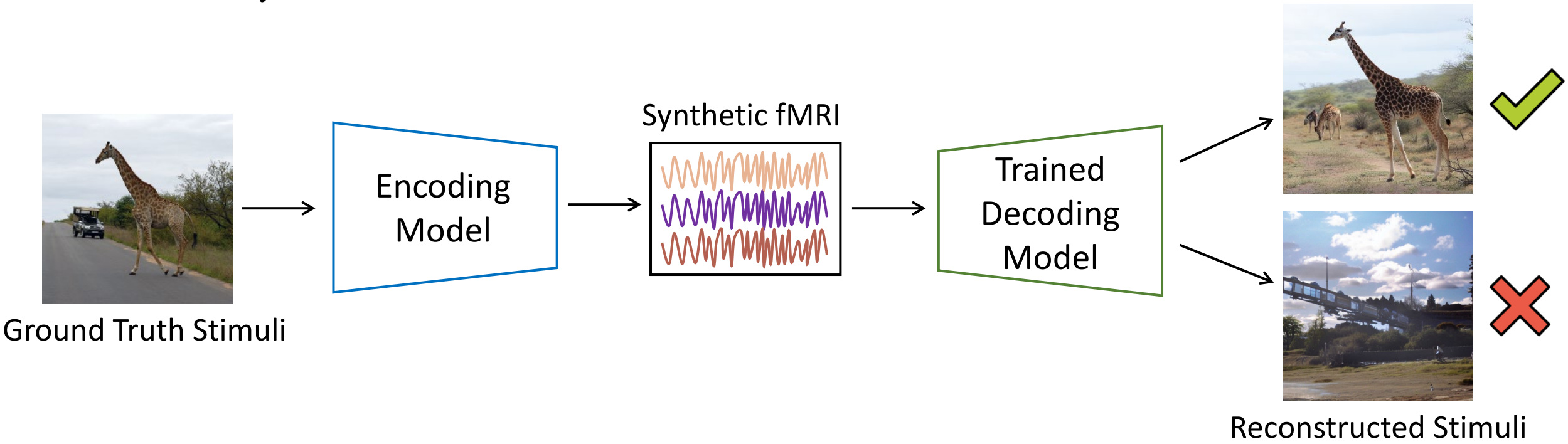
- P1 has better pixel-level MSE but lacks semantic meaning.
- P2, on the other hand, has better semantic meaning but might not perform as well on pixel-level MSE.

3. Neuroscience research requires us to focus more on the accuracy of synthetic fMRI in terms of global neural response patterns (i.e., the fMRI semantics), which embody the essence of human visual phenomena.

Evaluation Metrics

fMRI semantic-level evaluation metric:

Using a trained fMRI-to-Image decoding model (e.g., MindEye), decode the synthetic fMRI to assess the accuracy at the semantic level.



More accurate image reconstruction means more accurate neural response patterns.

Therefore, image reconstruction metrics are used to compare the GT Stimuli and Reconstructed Stimuli.

Evaluation

Method	Voxel-Level		Semantic-Level							
	Pearson↑	MSE↓	PixCorr↑	SSIM↑	Alex(2)↑	Alex(5)↑	Incep↑	CLIP↑	Eff↓	SwAV↓
GT fMRI (upper bound)	-	-	0.278	0.328	95.2%	99.0%	96.4%	94.5%	0.622	0.343
Linear Regressive	0.334	0.394	0.174	0.266	85.4%	94.2%	90.1%	87.2%	0.728	0.432
Transformer Encoding	0.337	0.387	0.166	0.286	83.5%	93.0%	89.8%	85.5%	0.759	0.440
MindSimulator (Trials=1)	0.346	0.403	0.197	0.297	88.9%	96.5%	92.1%	90.4%	0.701	0.396
MindSimulator (Trials=5)	**0.357**	**0.385**	**0.202**	**0.298**	**89.7%**	**97.0%**	**93.1%**	**91.2%**	**0.689**	**0.391**

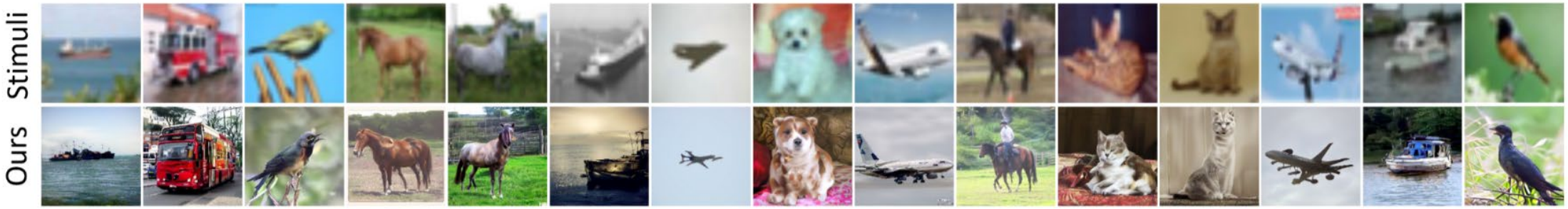
Results:

- Quantitative results show that the synthesized fMRI outperforms the baseline method and is close to the GT fMRI.
- Visual results also indicate that the synthesized fMRI is quite similar to the real fMRI at the semantic level.

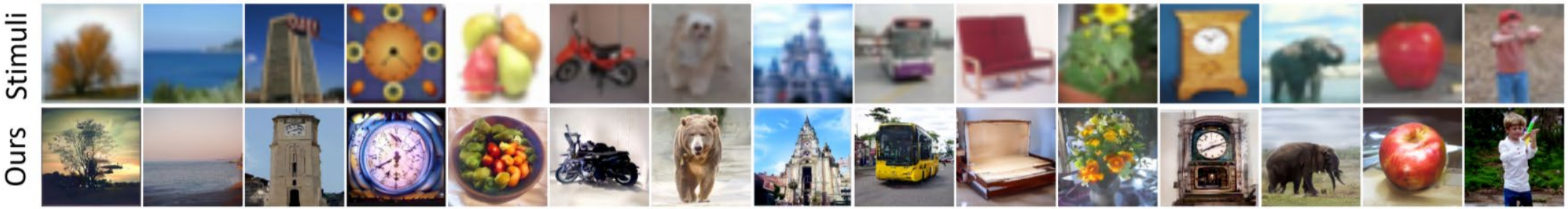
Evaluation

Datasets	Semantic-Level							
	PixCorr↑	SSIM↑	Alex(2)↑	Alex(5)↑	Incep↑	CLIP↑	Eff↓	SwAV↓
MSCOCO	0.201	0.302	**89.5%**	**96.8%**	**91.5%**	88.7%	**0.724**	**0.409**
CIFAR-10	**0.269**	0.406	88.5%	94.3%	84.3%	**90.5%**	0.898	0.645
CIFAR-100	0.260	**0.420**	86.6%	93.0%	82.8%	86.3%	0.916	0.659

Datasets: CIFAR-10



Datasets: CIFAR-100

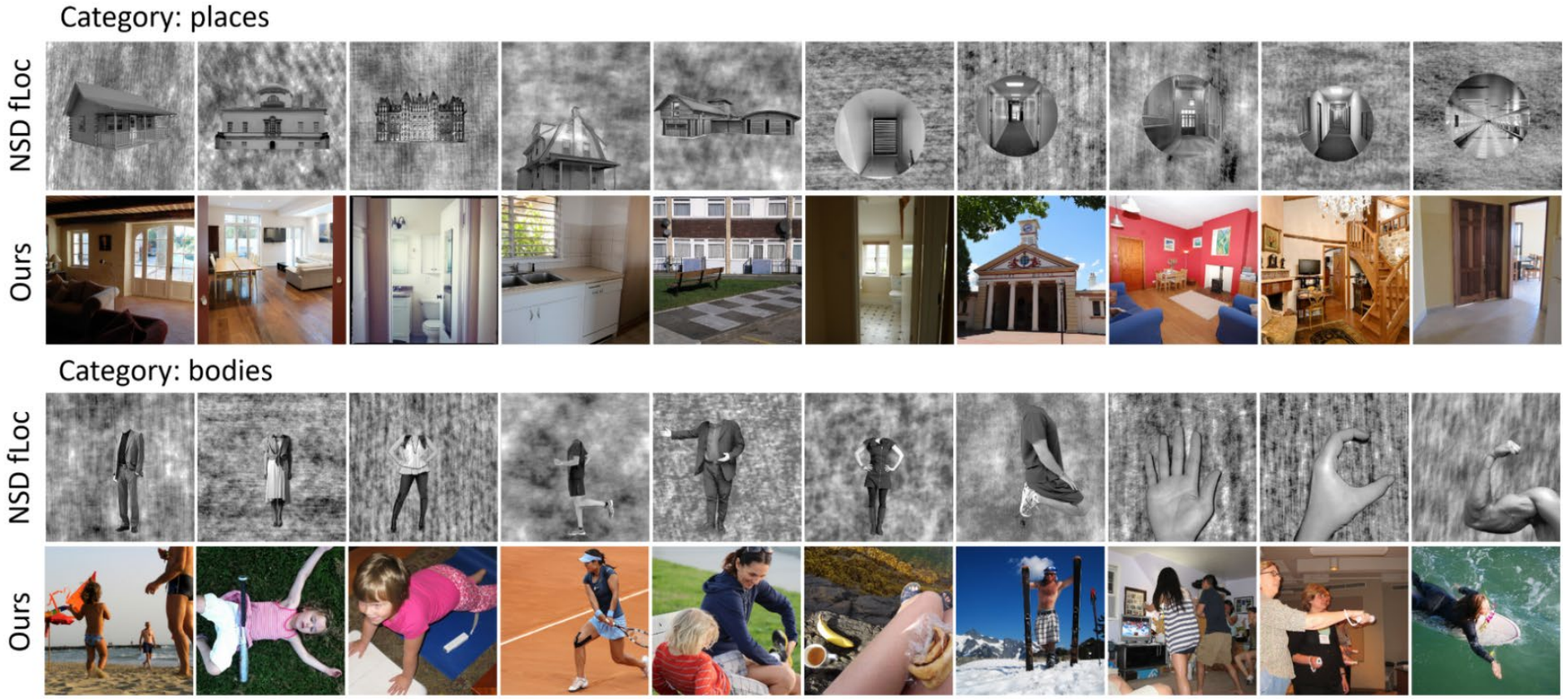


MindSimulator demonstrates generalization capability across mainstream image datasets.

Concept-selective Region Localization: Validation

Using synthesized fMRI for neuroscience exploration, locate **existing** concept-selective regions and **validate** them based on empirical findings.

Step1: Construct a **visual stimulus set for real-world scenarios** (e.g., concepts like places and bodies). Use CLIP for **zero-shot classification** to select visual stimuli with high semantic relevance.



Concept-selective Region Localization: Validation

Step2: Use MindSimulator to generate synthetic fMRI.

Step3: Perform voxel-wise single-sample t-tests to calculate the significance of each voxel, and select the highly significant voxels as the localized concept-selective regions. Evaluate the localization results using **accuracy (Acc) and F1 score**.

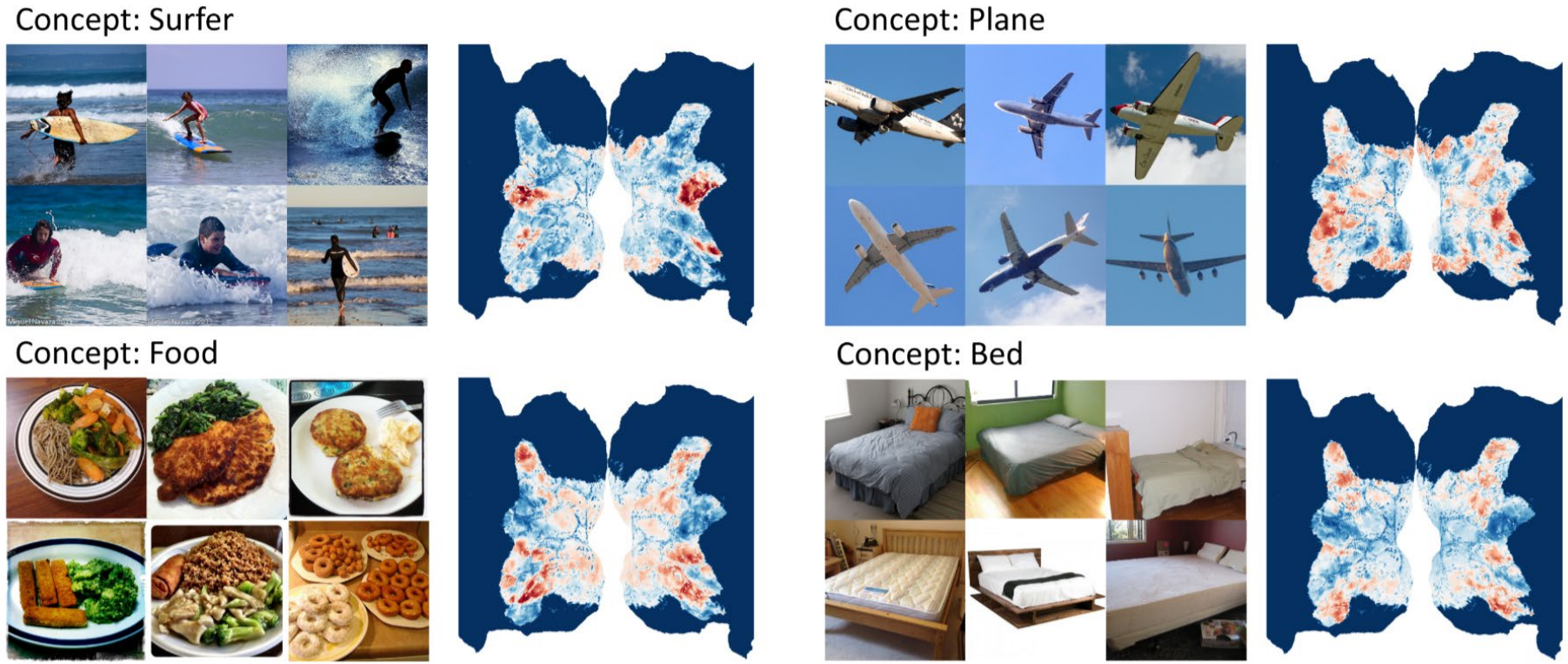
# Images	places-Specificity	places-Acc↑		places-F1↑		bodies-Specificity	bodies-Acc↑		bodies-F1↑	
		Linear	Ours	Linear	Ours		Linear	Ours	Linear	Ours
Top 100	0.9608	36.0%	**64.4%**	0.498	**0.517**	0.9988	51.1%	**96.2%**	**0.577**	0.493
Top 200	0.9391	33.0%	**56.2%**	0.470	**0.570**	0.9977	45.9%	**92.3%**	0.562	**0.628**
Top 300	0.9189	31.5%	**51.3%**	0.458	**0.581**	0.9968	43.8%	**90.4%**	0.556	**0.683**
Top 500	0.8834	30.4%	**46.3%**	0.449	**0.570**	0.9953	41.6%	**87.1%**	0.545	**0.728**
Top 1000	0.8084	29.1%	**39.7%**	0.437	**0.531**	0.9918	40.0%	**78.9%**	0.535	**0.737**

- The localization results are **satisfactory**, and are **positively correlated with** the semantic specificity of the concepts.
- The localization results outperform commonly used linear encoding models.

Concept-selective Region Localization: Exploring

Use synthetic fMRI for neuroscience exploration to discover new concept-selective regions

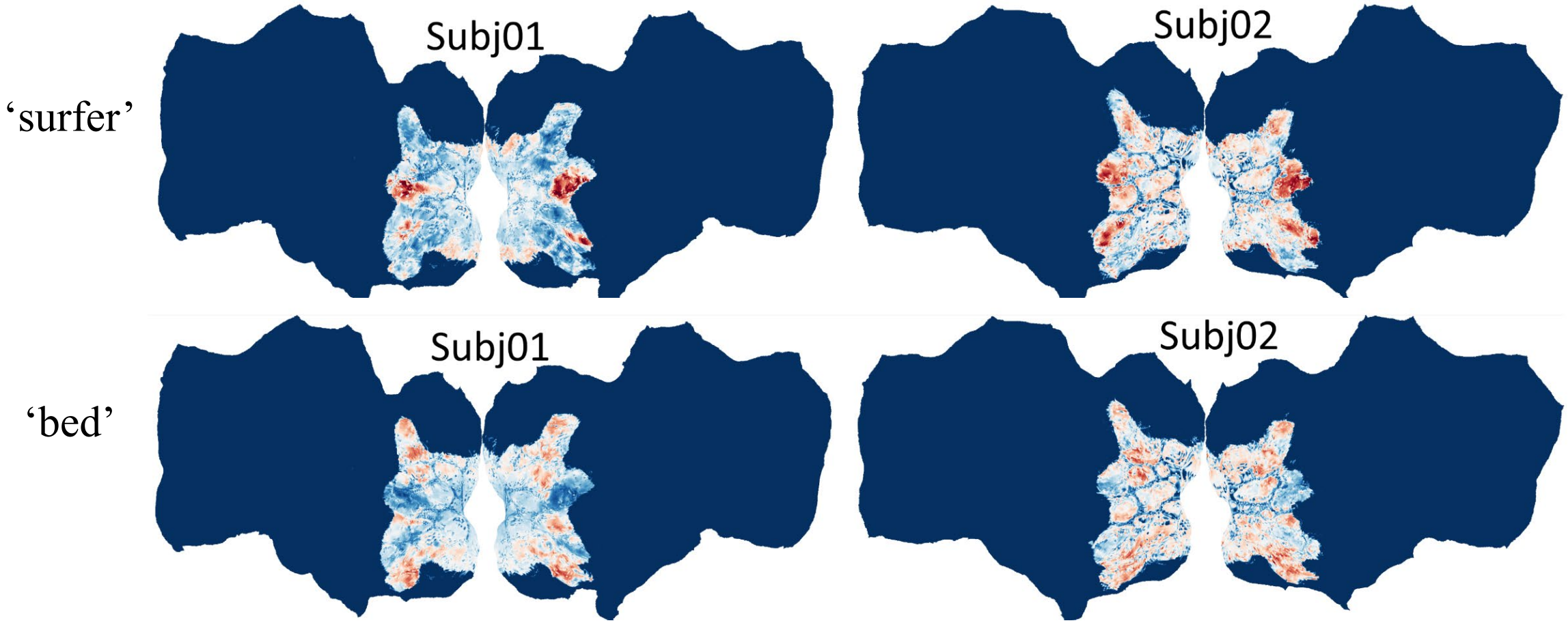
(e.g., for concepts such as surfing, airplanes, food, and beds).



We use synthetic data to identify the concept-selective regions corresponding to these concepts.

概念区域定位：探索

Use synthetic fMRI for neuroscience exploration to discover new concept-selective regions and perform cross-subject comparisons.



The same concept-selective regions across different subjects exhibit both similarities and differences.

概念区域定位：消融

Use synthetic fMRI for neuroscience exploration to discover new concept-selective regions, and validate them through voxel ablation.

For example, mask the surfer-selective region's voxels (6.5%) and reconstruct using a decoding model.

After voxel masking, the reconstructed visual stimulus loses concepts such as "surfing," "surfer," and "ocean," indicating that the localized concept-selective region is correct.

Thanks !