OUKA
The University of Osaka
Institutional Knowledge Archive

Title	Brain-aligning of semantic vectors improves neural decoding of visual stimuli
Author(s)	Vafaei, Shirin
Citation	大阪大学, 2026, 博士論文
Version Type	
URL	https://hdl.handle.net/11094/106969
rights	
Note	やむを得ない事由があると学位審査研究科が承認したため、全文に代えてその内容の要約を公開しています。全文のご利用をご希望の場合は、大阪大学の博士論文についてをご参照ください。

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

論　文　内　容　の　要　旨

Synopsis of Thesis

氏　　名 Name	Vafaei Shirin
論文題名 Title	Brain-aligning of semantic vectors improves neural decoding of visual stimuli (AIの特徴空間を脳表現にアラインメントすることで視覚刺激の神経デコーディング性能が向上する)

論文内容の要旨

〔目　的(Objective)〕

Accurate brain decoding requires mapping neural data onto feature vectors derived from image- or text-based spaces. However, the intrinsic structure of these pretrained vectors may fundamentally differ from how the brain organizes visual information, limiting decoder performance. To address this, we propose brain-aligning of semantic vectors, a framework that fine-tunes pretrained feature vectors to better align with the brain's internal representational geometry. The model was trained on fMRI data and evaluated under zero-shot conditions across fMRI, MEG, and ECoG datasets, demonstrating that reshaping the geometry of semantic spaces toward neural organization can meaningfully improve decoding across diverse recording modalities.

方法ならびに成績(Methods/Results)〕

A multimodal autoencoder was developed to restructure pretrained CLIP or GloVe vectors using a two-term loss function: an MSE term for faithful reconstruction of the original vectors, and a penalty term that minimizes divergence between the autoencoder's representational similarity matrix (RSM) and the RSM derived from real fMRI brain activity. A hyperparameter α controls the trade-off between these objectives, with lower values imposing stronger brain alignment. Training used the publicly available Generic Object Decoding (GOD) fMRI dataset, comprising recordings from five participants viewing 1,200 images across 200 ImageNet categories. Training and test categories were non-overlapping, enabling true zero-shot generalization to entirely novel visual categories. Neural signals from the lateral occipital complex (LOC) and V4 were used for alignment, given their established roles in object and shape processing. Decoding performance was assessed via samplewise and dimensionwise Pearson correlation, as well as stimulus identification accuracy. Brain-aligned vectors were subsequently applied to MEG data (source-localized to the ventral visual stream) and ECoG data (231 subdural electrodes from 4 epilepsy patients covering ventral visual cortex), both from participants entirely independent of the fMRI training set, providing a rigorous test of cross-modal and cross-individual generalization.

〔総　括(Conclusion)〕

Brain-aligned vectors consistently and substantially outperformed standard pretrained vectors across all three neuroimaging modalities. In fMRI decoding, stronger alignment levels ($\alpha = 10^{-3}$–10^{-4}) yielded the largest gains for both CLIP and GloVe feature spaces, significantly outperforming both original vectors and reconstruction-only baselines for perceived and imagined stimuli ($p < 0.001$, Bonferroni corrected), without degrading stimulus identification accuracy. Brain-aligned CLIP vectors also demonstrated greater category discriminability across tens of thousands of ImageNet classes, suggesting the restructured semantic space more cleanly separates distinct object categories. MEG decoding showed successful transfer of fMRI-derived brain-aligned vectors to different participants, with significant improvements in samplewise accuracy ($p < 0.001$) and identification performance at intermediate-to-strong alignment levels. ECoG results were particularly striking: original CLIP vectors failed entirely to exceed chance-level identification accuracy ($p = 0.9991$), whereas brain-aligned vectors achieved clear and significant performance ($p < 0.001$), despite being derived from a completely different neuroimaging modality and participant group. Taken together, these findings demonstrate that aligning feature vector geometry to the brain's own representational structure produces more decodable semantic spaces, and that this alignment generalizes robustly across modalities, individuals, and visual categories unseen during training.

論文審査の結果の要旨及び担当者

（申請者氏名）Vafaei　Shirin

論文審査担当者	（職）	氏　名	
	主　査　大阪大学教授	柳澤琢史	署名
	副　査　大阪大学教授	西田幸二	署名
	副　査　大阪大学教授	北澤茂	署名

論文審査の結果の要旨

脳活動から視覚情報を読み取る脳情報解読は、重度運動障害患者向けの意思伝達支援や視覚情報処理の解明に向けて重要な技術である。従来手法では、事前学習済みモデルから得た特徴ベクトルに脳活動を対応付けてきたが、これらのベクトル構造が脳内表現と必ずしも一致しないことが精度向上の障壁となっていた。本研究では、fMRIデータを用いて事前学習済み特徴ベクトルを脳の表現類似性構造に合わせてファインチューニングするアライメント手法を新たに提案した。アライメントされたベクトルを用いてゼロショット脳情報解読を行った結果、fMRI・MEG・ECoGの3種類の計測モダリティ全てで脳情報解読精度が向上し、手法の汎用性を示した。また、刺激カテゴリに依存した精度向上も確認され、脳の表現空間とAIモデルと関係性が示唆された。本研究はAIと神経科学を融合した独創的な取り組みであり、博士（医学）の学位授与に値する.