Towards Generalized Brain Decoding of visual stimuli: Cross-Subject and Dataset Perspectives with a simple alignment technique

Matteo Ferrante[1], **Tommaso Boccato**[1], **Furkan Ozcelik**[2], **Rufin VanRullen**[2], **Nicola Toschi**[1],

[1]Department of Biomedicine and Prevention, University of Rome Tor Vergata (IT)
[2]CerCo, CNRS UMR5549, Toulouse, France, Universite de Toulouse, Toulouse, France, ANITI, Toulouse, France
[3]Martinos Center For Biomedical Imaging MGH and Harvard Medical School (USA)

Abstract

To-date, brain decoding literature has focused on single-subject studies, i.e. reconstructing stimuli presented to a subject under fMRI acquisition from the fMRI activity of the same subject. The objective of this study is to introduce a generalization technique that enables the decoding of a subject's brain based on fMRI activity of another subject, i.e. cross-subject brain decoding. To this end, we also explore cross-subject data alignment techniques. Data alignment is the attempt to register different subjects in a common anatomical or functional space for further and more general analysis.

We worked with the Natural Scenes Dataset, a comprehensive 7T fMRI experiment focused on vision of natural images. The dataset contains fMRI data from multiple subjects exposed to 9841 images, where 982 images have been viewed by all subjects. Our method involved training a decoding model on one subject's data, aligning new data from other subjects to this space, and testing the decoding on the second subject based on information aligned to first subject. We found that cross-subject brain decoding is possible, even with a small subset of the dataset, specifically, using the common data, which are around 10% of the total data, namely 982 images, with performances in decoding compararble to the ones achieved by single subject decoding. Ridge regression emerged as the best method for functional alignment in fine-grained information decoding, outperforming all other techniques. By aligning multiple subjects, we achieved high-quality brain decoding and a potential reduction in scan time by 90%. This substantial decrease in scan time could open up unprecedented opportunities for more efficient experiment execution and further advancements in the field, which commonly requires prohibitive (20 hours) scan time per subject.

Introduction

Brain decoding involves reconstructing the original stimuli from neural data, such as recreating images that triggered specific brain activity, using measurements obtained from functional Magnetic Resonance Imaging (fMRI). Brain decoding faces the challenge of individual variability in brain anatomy and function. To overcome the need for extensive data collection from each subject, we propose a functional alignment of brain representation using ridge regression, aligning brain activity patterns across subjects. This enables the application of a fine-grained decoder trained on one subject to others, reducing the need for comprehensive data collection per individual (Du et al. 2022a; Zafar et al. 2015; Awangga, Mengko, and Utama 2020). The goal of this work is to decode the visual representation from brain activity measured through functional MRI across different subjects. Our approach demonstrates successful cross-subject decoding and visual stimulus reconstruction, transforming varied brain representations into a common aligned space. Unlike transfer learning, which requires re-training models with each new subject's data, our method aligns brain activity patterns without re-training the decoder, making it more efficient and scalable for cross-subject brain decoding. This facilitates applications with minimal data requirements and accommodates decoding under diverse conditions.

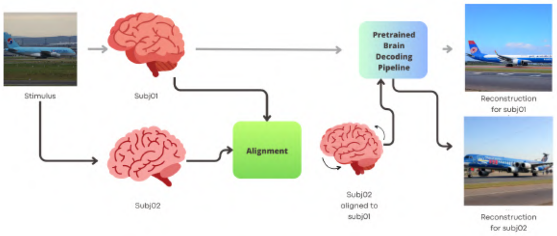
Stimulus
Subj01
Pretrained Brain Decoding Pipeline
Reconstruction for subj01
Subj02
Alignment
Subj02 aligned to subj01
Reconstruction for subj02

Figure 1: This scheme outlines our cross-subject decoding method using two subjects. First, a decoding model is trained with one subject's (Subj01) brain activity corresponding to 8859 unique images. Next, the model aligns and decodes images from a second subject (Subj02) based on their shared stimuli exposure (982 images), enabling image reconstruction from Subj02's brain activity without a separate model..

Related Work

In the rapidly advancing field of deep learning-based brain decoding, a variety of models have been employed to analyze preprocessed fMRI time series for decoding visual stimuli, focusing on reconstructing images from specific fMRI patterns. These include methods using variational autoencoders with generative adversarial components, sparse

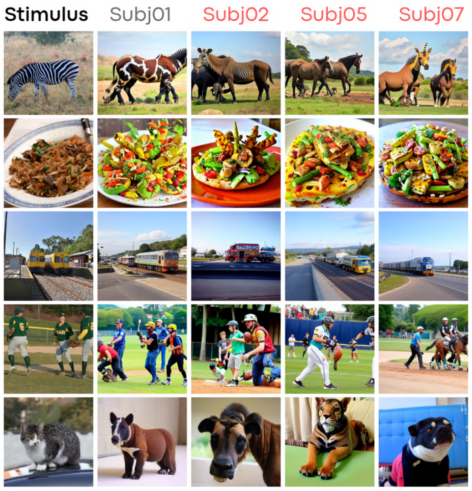

Figure 2: The ”Stimulus” column shows images from an fMRI experiment. The ”Subj01” column illustrates decoded results using a subject-specific model, serving as a performance baseline. Other columns demonstrate results from functional alignment using Ridge Regression with common data (952 images). Here, subjects are aligned to Subj01 and decoded with Subj01’s trained decoder. The images displayed were not used in the alignment process, highlighting the effectiveness of functional alignment on unseen data..

linear regression, unsupervised and adversarial strategies, as well as the application of pretrained architectures and diffusion models for enhanced image reconstruction (VanRullen and Reddy 2019; Horikawa and Kamitani 2017; Shen et al. 2019; Ren et al. 2019; Gaziv et al. 2022; Donahue and Simonyan 2019; Casanova et al. 2021; Takagi and Nishimoto 2023; Chen et al. 2022; Ferrante, Boccato, and Toschi 2023; Ozcelik and VanRullen 2023; Ferrante et al. 2023). For a detailed review of these methods, we remand to a recent literature review (Du et al. 2022b). Functional alignment techniques, including Hyperalignment, the Shared Response Model (SRM), and Independent Component Analysis (ICA), have been explored for aligning brain activity across individuals, each with its unique advantages and limitations (Haxby et al. 2011, 2020; Chen et al. 2015; Calhoun, Liu, and Adalı 2009; Bazeille et al. 2021). This paper introduces a simplified approach to align neural data across subjects for effective brain decoding of fMRI data, leveraging unified brain representations. This method contrasts traditional transfer learning by maintaining a constant model and aligning neural data, suitable for limited data and diverse experimental conditions.

Material and Methods

In this section, we describe the proposed method and the data we used. The data are publicly available and can be requested at https://naturalscenesdataset.org/. The study utilizes the Natural Scenes Dataset (NSD) (Allen et al. 2022), a vast fMRI data set from eight subjects exposed to images from the COCO21 dataset. We focused on four subjects, forming a unique training dataset of 8,859 images and 24,980 fMRI trials from Subj01, and a common dataset of 982 images and 2,770 trials for each one of the subjects. To reduce spatial dimensionality, we applied a mask to the fMRI signal (resolution of 1.8mm isotropic) using the NSDGeneral ROI, targeting various visual areas. This strategic ROI selection enhanced the signal-to-noise ratio and simplified data complexity, enabling exploration of both low-level and high-level visual features. Temporal dimensionality was reduced using precomputed betas from a general linear model (GLM) with a fitted hemodynamic response function (HRF) and a denoising process as detailed in the NSD paper. Data from Subj01, Subj02, Subj05, Subj07, warped into the Montreal Neurological Institute common space (MNI) and downsampled at 2mm, represented the brain activity of each subject and helped to decrease computational time and cost. We used the common dataset as alignment, keeping out 30 images for visual comparison, so there are 8859 unique images for each subject. We only used them for training the decoding model for Subj01. Then there are 952 common images across all subjects that were used to functionally align them to the activity of Subj01, and 30 common images kept out for visual comparison on images neither used in the training or in the alignment procedure. These 30 images were chosen because they’re used as visual qualitative evaluation of decoding results in other papers and could help the reader to compare results across different methods. Decoding metrics are evaluated on the 952 images which correspond to our test set for each one of the subjects, since these images are never seen by the decoder model, so the evaluation is still fair and on unseen images. When we refer to 100% of common data we are pointing to these 952 images.

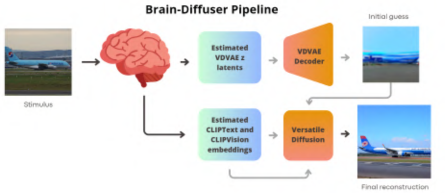

Figure 3: Brain-Diffuser pipeline, the decoder used in this study, begins with brain activity from viewing an image stimulus. A model is trained to estimate the latent representation of the VDVAE autoencoder as well as the text and visual embeddings of the CLIP model, using linear models. These estimated vectors are fed into Versatile Diffusion—a latent diffusion model—to reconstruct the final image.

Decoding model:The ”Brain-Diffuser” (Ozcelik and VanRullen 2023) model is a two-stage framework for reconstructing natural scenes from fMRI signals. Initially, a Very Deep Variational Autoencoder (VDVAE) provides an ”initial guess” of the reconstruction, focusing on low-level details. This guess is refined using high-level semantic features from CLIP-Text and CLIP-Vision models, employing a la-

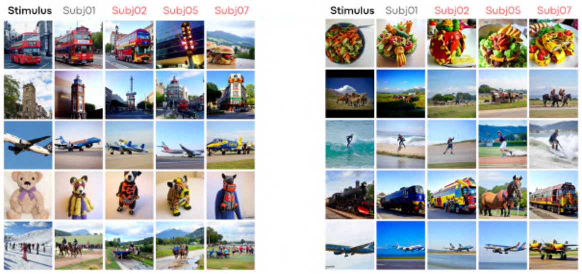

Figure 4: More example results. Format and conventions as in Figure 2

tent diffusion model (Versatile Diffusion) for final image generation. The model, represented in Fig. 3, takes fMRI signals as input and generates reconstructed images, capturing low-level properties and overall layout. As a state-of-the-art procedure, Brain-Diffuser was trained using data from Subj01 in the MNI space (cross-subject decoding requires a common space). Further details about the decoding model are available in the original paper.

Cross-subject decoding: This study investigates cross-subject brain decoding enabled by functional alignment. The pipeline involves two alignment steps: Anatomical alignment transforms data to a standard template (MNI space), facilitating structure comparison across subjects. This linear + nonlinear warping relies on software like FSL and ANTs. Functional alignment necessitates a more comprehensive approach. Consider the scenario where the brain activity of a source subject $\mathbf{S}$ needs to align with a target subject $\mathbf{T}$. These activities, responses to numerous stimuli, are matrices of shape *(# stimuli, # voxels)*. Given that subjects encounter several common stimuli (i.e., they view identical images in the fMRI scanner), we can divide the datasets into $\mathbf{T}common, \mathbf{T}different$ and $\mathbf{S}common, \mathbf{S}different$. Our goal is to leverage the $common$ dataset portion to learn a mapping from S to T, aligning the entire $\mathbf{S}$ dataset with the $\mathbf{T}$ functional space. The NSD experiment's structure, with separate training and test sets (the latter containing identical images for each subject), provides a common stimuli set for alignment purposes.

Our approach embraces a simple assumption: even in different subjects, all functional data contain the information for the same stimuli, albeit possibly spread across different voxels. This suggests that one subject's activity (source) might be expressed as a linear combination of the activity of another subject (target) for the same stimuli. By deriving a linear combination for each voxel of the target from all possible voxels of the source, we can create a linear map from the source to the target, facilitating functional alignment. The target subject activity can be expressed as $\mathbf{t}_i = \sum_j \mathbf{w}_j \mathbf{s}_j$ where $\mathbf{t}i$ is the i-th activity of the target voxel for each common dataset stimulus. Here, $\mathbf{t}i$ represents the i-th column of $\mathbf{T}common$, expressed as a linear combination of all $\mathbf{S}common$ columns. The challenge lies in finding the vector of $\mathbf{w}$ values. When extended to all the target subject voxels, the w vector morphs into a square matrix $\mathbf{W}$, each column of which contains weights to estimate one target subject voxel from a combination of source values. The objective can be redefined as minimizing $|\mathbf{S}common\mathbf{W}^T - \mathbf{T}common|^2$.

We employed Ridge Regression to determine the $\mathbf{W}$ matrix, conducting a 5-fold cross-validation to select the optimal hyper-parameter α. We computed these values to align all the sources Subj02, Subj05, and Subj07, to the functional space of Subj01 chosen as the target. This pipeline leverages both anatomical and functional considerations to enable precise cross-subject decoding, mitigating inter-individual variability. The functional alignment step is key for handling differences in how information is represented across subjects' brains.

Evaluation

Our research seeks to evaluate visual stimuli's detailed cross-subject brain decoding feasibility, scrutinizing the alignment methods and shared data ratio at play. Our shared dataset, or ”test dataset,” comprises 982 images, all viewed by every subject. In order to allow visual comparison, we excluded 30 images from the original Brain-Diffuser paper. Thus, these excluded images neither influenced the training of the decoding pipeline nor the alignment process. The remaining 952 images serve as the shared dataset. We computed transformations for each alignment method (anatomical, hyperalignment, ridge regression) and shared dataset proportion, applying the linear transformation to the complete dataset. This procedure aligns the images with Subject 01's functional space. We then used the pre-trained Brain-Diffuser pipeline for decoding the aligned fMRI activity and reconstructing the images. We assessed our image reconstruction process through both basic and advanced metrics, including PixCorr, SSIM, and 2-way accuracy in AlexNet, Inception, and CLIP latent spaces. This comprehensive evaluation approach allows us to benchmark our results against other brain decoding studies. Further studies in Appendix show the relation between the fraction of data used for alignment and performances, the impact of choosing one subject or another as ”target” and the comparison without functional alignment and using hyperalignment, a popular neuroimaging method as functional alignment approach.

Cross-Dataset decoding experiment

To assess our decoding pipeline's generalizability in hard conditions, we conducted cross-dataset decoding between BOLD5000(Chang et al. 2019) and the Natural Scenes Dataset (NSD). BOLD5000 includes fMRI data from five subjects viewing 5,000 images at 3T field strength, offering lower signal-to-noise ratio compared to NSD's 7T data. The semantic range in BOLD5000 is narrower than NSD's varied stimuli. Display protocols differ: NSD uses a rapid-event protocol with images shown for 2 seconds and a 1-second pause, while BOLD5000 shows images for 1 second followed by 9 seconds of cross fixation. Both datasets were processed identically within the visual cortex masks. Focusing on BOLD5000's subject CSI1, which shares 1,000 images with NSD subjects, we aimed for direct neural response comparison. The experiment involved training a de-

Subj	Pixcorr	SSIM	AlexNet2	AlexNet5	Inception	CLIP
subj01 (target)	0.287676	**0.268134**	0.847251	**0.96334**	0.89613	0.936864
subj02 (aligned)	**0.288028**	0.267577	0.839104	0.956212	0.893075	**0.955193**
subj05 (aligned)	0.283798	0.267467	0.836049	0.953157	0.904277	0.937882
subj07 (aligned)	0.283352	0.266303	**0.851324**	0.957230	**0.910387**	0.918534

	Pixcorr	SSIM	AlexNet2	AlexNet5	Inception	CLIP
within BOLD5000	0.102	0.103	**0.569**	0.596	0.532	0.705
Cross NSD	**0.1734**	**0.231**	0.562	**0.732**	**0.598**	**0.729**

Table 1: The upper section of the table shows quantitative metrics for cross-subject decoding experiments with bold values indicating better performance. Subjects 02, 05, and 07 are aligned and decoded using subject 01's data. The lower section presents metrics for cross-dataset decoding experiments, comparing within BOLD5000 and Cross NSD performances.

coder with NSD Subject 1, aligning common data from BOLD5000 CSI1, and conducting cross-dataset decoding. We compared this with within-dataset and within-subject decoding performances using the Brain-Diffuser pipeline on BOLD5000's training data (80% non-common data). Ridge Regression was used as the alignment matrix between BOLD5000 and NSD neural spaces. After transforming BOLD5000 test data, decoding was performed using the NSD Subj01-trained decoder. Metrics were calculated for decoded test sets within and across datasets, exploring the capability of high-quality cross-dataset and cross-field fMRI data decoding.

Results

Cross-Subject Decoding: Figures 2, 4 compare stimuli to decoded images for Subj01 (decoder training subject) and other aligned subjects using Ridge Regression. Table 1 shows quantitative metrics for aligned subjects. In supplementary material, more qualitative and quantitative comparisons and different baselines (only anatomical alignment, functional hyperalignment), as well as experiments with different amounts of data used for learning alignment matrices, can be found. Qualitatively and quantitatively, we found that functional alignment is critical for fine-grained decoding, precisely matching neural signals to brain regions. Ridge Regression achieved above-chance performance with just 10% of data, nearing qualitatively and quantitatively performances of within-subject decoding. This suggests reliable decoding with significantly reduced scan times is feasible.

Cross-Dataset Decoding: Despite protocol and field strength differences between datasets, cross-dataset decoding was successful. Qualitative analysis showed reconstructions from BOLD5000 subject surpassed within-dataset decoding when aligned to NSD subject and using its decoder (Figure 5). This demonstrates the feasibility of applying a decoder across distinct fMRI datasets with shared stimuli. Variations in protocols and field strengths pose challenges, but our pipeline achieved effective functional alignment despite these hurdles. Quantitative metrics (Table 1) further demonstrate enhanced performance across all parameters when using the cross-dataset decoder, highlighting the significant advantages of this approach. In summary, key findings emphasize functional alignment's pivotal role and the potential for reliable decoding with reduced data requirements. Cross-dataset decoding further highlights the possibilities of generalized models.

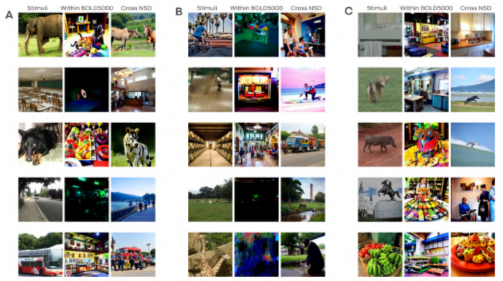

Figure 5: Cross-Dataset qualitative results. Columns A, B, C display random test set examples. "Stimuli" column presents original experimental stimuli. "Within BOLD5000" shows reconstructions by a BOLD5000-trained model. "Cross NSD" column depicts results from aligned NSD dataset activity decoded with an NSD-trained decoder, showing higher semantic similarity than the second column.

Discussion and Conclusion

Our study highlights the crucial role of functional alignment in brain decoding. This alignment enables accurate neural activity decoding from different individuals using a model trained on another subject. Our findings suggest that using a simple method like Ridge Regression can significantly reduce scan times, as reliable decoding is possible with just a portion of the dataset. This approach, along with the observation of qualitative similarities in decoded images across subjects, opens up new research directions to explore fine-grained inter-subject differences. We also discovered that despite individual brain structure and function differences, it is feasible to decode shared neural activity patterns, suggesting the potential for generalized brain decoding models. Our cross-dataset experiment between BOLD5000 and the Natural Scenes Dataset demonstrates the feasibility of this approach, even with disparities in acquisition protocols. Current research suggests that while certain brain activity aspects can be decoded across subjects, the process is not yet a comprehensive or intrusive 'mind-reading' tool. A key finding highlights the disruptive role of attention mechanisms (Çukur et al. 2013), suggesting that brain decoding is only possible with actively participating, aware subjects. While our methods currently prevent involuntary or covert 'mind reading', as the field advances, maintaining strong ethical frameworks for brain decoding research becomes even more critical. Informed consent, strict data privacy protocols, and potential societal implications consideration remain key aspects of the advance of science in this direction. In conclusion, our research underlines the importance of functional alignment in brain decoding of visual stimuli and reveals the possibility of reducing scanning durations while still achieving effective decoding.

References

Allen, E. J.; St-Yves, G.; Wu, Y.; Breedlove, J. L.; Prince, J. S.; Dowdle, L. T.; Nau, M.; Caron, B.; Pestilli, F.; Charest, I.; Hutchinson, J. B.; Naselaris, T.; and Kay, K. 2022. A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nature Neuroscience*, 25(1): 116–126.

Awangga, R. M.; Mengko, T. L. R.; and Utama, N. P. 2020. A literature review of brain decoding research. *IOP Conference Series: Materials Science and Engineering*, 830(3): 032049.

Bazeille, T.; DuPre, E.; Richard, H.; Poline, J.-B.; and Thirion, B. 2021. An empirical evaluation of functional alignment using inter-subject decoding. *NeuroImage*, 245: 118683.

Calhoun, V. D.; Liu, J.; and Adalı, T. 2009. A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. *NeuroImage*, 45(1, Supplement 1): S163–S172. Mathematics in Brain Imaging.

Casanova, A.; Careil, M.; Verbeek, J.; Drozdzal, M.; and Romero-Soriano, A. 2021. Instance-Conditioned GAN.

Chang, N.; Pyles, J. A.; Marcus, A.; Gupta, A.; Tarr, M. J.; and Aminoff, E. M. 2019. BOLD5000, a public fMRI dataset while viewing 5000 visual images. *Scientific Data*, 6(1): 49.

Chen, P.-H. C.; Chen, J.; Yeshurun, Y.; Hasson, U.; Haxby, J.; and Ramadge, P. J. 2015. A Reduced-Dimension fMRI Shared Response Model. In Cortes, C.; Lawrence, N.; Lee, D.; Sugiyama, M.; and Garnett, R., eds., *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc.

Chen, Z.; Qing, J.; Xiang, T.; Yue, W. L.; and Zhou, J. H. 2022. Seeing Beyond the Brain: Conditional Diffusion Model with Sparse Masked Modeling for Vision Decoding. arXiv:2211.06956.

Donahue, J.; and Simonyan, K. 2019. Large Scale Adversarial Representation Learning.

Du, B.; Cheng, X.; Duan, Y.; and Ning, H. 2022a. fMRI Brain Decoding and Its Applications in Brain and Computer Interface: A Survey. *Brain Sciences*, 12(2).

Du, B.; Cheng, X.; Duan, Y.; and Ning, H. 2022b. FMRI brain decoding and its applications in brain-computer interface: A survey. *Brain Sci.*, 12(2): 228.

Ferrante, M.; Boccato, T.; and Toschi, N. 2023. Semantic Brain Decoding: from fMRI to conceptually similar image reconstruction of visual stimuli. arXiv:2212.06726.

Ferrante, M.; Ozcelik, F.; Boccato, T.; VanRullen, R.; and Toschi, N. 2023. Brain Captioning: Decoding human brain activity into images and text. arXiv:2305.11560.

Gaziv, G.; Beliy, R.; Granot, N.; Hoogi, A.; Strappini, F.; Golan, T.; and Irani, M. 2022. Self-supervised Natural Image Reconstruction and Large-scale Semantic Classification from Brain Activity. *NeuroImage*, 254: 119121.

Gower, J. C. 1975. Generalized procrustes analysis. *Psychometrika*, 40(1): 33–51.

Haxby, J. V.; Guntupalli, J. S.; Connolly, A. C.; Halchenko, Y. O.; Conroy, B. R.; Gobbini, M. I.; Hanke, M.; and Ramadge, P. J. 2011. A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron*, 72(2): 404–416.

Haxby, J. V.; Guntupalli, J. S.; Nastase, S. A.; and Feilong, M. 2020. Hyperalignment: Modeling shared information encoded in idiosyncratic cortical topographies. *eLife*, 9: e56601.

Horikawa, T.; and Kamitani, Y. 2017. Generic decoding of seen and imagined objects using hierarchical visual features. *Nature Communications*, 8(1): 15037.

Ozcelik, F.; and VanRullen, R. 2023. Brain-Diffuser: Natural scene reconstruction from fMRI signals using generative latent diffusion. arXiv:2303.05334.

Ren, Z.; Li, J.; Xue, X.; Li, X.; Yang, F.; Jiao, Z.; and Gao, X. 2019. Reconstructing Perceived Images from Brain Activity by Visually-guided Cognitive Representation and Adversarial Learning. ArXiv:1906.12181 [cs].

Shen, G.; Dwivedi, K.; Majima, K.; Horikawa, T.; and Kamitani, Y. 2019. End-to-end deep image reconstruction from human brain activity. *Front. Comput. Neurosci.*, 13: 21.

Takagi, Y.; and Nishimoto, S. 2023. High-resolution image reconstruction with latent diffusion models from human brain activity. *bioRxiv*.

VanRullen, R.; and Reddy, L. 2019. Reconstructing faces from fMRI patterns using deep generative neural networks. *Communications Biology*, 2(1): 193.

Zafar, R.; Malik, A. S.; Kamel, N.; Dass, S. C.; Abdullah, J. M.; Reza, F.; and Abdul Karim, A. H. 2015. Decoding of visual information from human brain activity: A review of fMRI and EEG studies. *Journal of Integrative Neuroscience*, 14(02): 155–168.

Çukur, T.; Nishimoto, S.; Huth, A. G.; and Gallant, J. L. 2013. Attention during natural vision warps semantic representation across the human brain. *Nature Neuroscience*, 16(6): 763–770.

Appendix

In this section, we present additional results that support the main findings of the paper. In particular, we explored as baselines the use of only anatomical alignment and a popular functional alignment technique called hyperalignment. Then, we asked ourselves how the amount of data used for learning the alignment matrices will affect the decoding performances and finally what happens when changing the target subject. Fig A3 and A8 show additional qualitative results of our approach.

Baselines

This section investigates three alignment strategies to evaluate cross-subject fine-grained brain decoding's feasibility: anatomical alignment, functional alignment via hyperalignment, and functional alignment through ridge regression.

Anatomical Alignment Anatomical alignment, a common neuroscience method, aligns to a standard template, here, the MNI space, facilitating anatomical structure comparison. This alignment typically involves a linear coregistration of anatomical images between native and common spaces, followed by a nonlinear warping to match common brain structures. Several software options like FSL and ANTs can perform this task. The NSDData authors (Allen et al. 2022) elaborate on this process in their released code, providing betas (i.e. coefficients obtained by theressing the stimuls waveform against the fMRI data) for all subjects in the MNI common space (1mm). We downsampled these to 2mm to approximate the resolution used in the original Brain-Diffuser decoding paper (1.8mm) and to reduce spatial dimensionality.

HyperAlignment HyperAlignment (Haxby et al. 2011, 2020), a functional data alignment technique, models functional data as high-dimensional points, with each voxel representing a dimension with betas ranging in $\mathcal{R}$. This method, based on Procrustes Analysis (Gower 1975), presents a high-dimensional model of the representational space in the human ventral temporal (VT) cortex, wherein dimensions are response-tuning functions common across individuals.

To perform the Procrustes analysis for functional brain alignment, we aim to find a rotation matrix $\mathbf{R}$ and a scale factor c such that the difference $|c\mathbf{SR} - \mathbf{T}|^2$ is minimized.

This is achieved by computing the matrix product $\mathbf{P} = \mathbf{S}^T_{common}\mathbf{T}_{common}$, Performing the singular value decomposition of $\mathbf{P}$ to obtain left and right eigenvector matrices $\mathbf{U}$ and $\mathbf{V}$, Computing $\mathbf{R} = \mathbf{U}\mathbf{V}^T$ and the scaling factor $c = \frac{trace(\mathbf{T_{common}}^T(\mathbf{S_{common}}\mathbf{R}))}{trace(\mathbf{S}^T_{common}\mathbf{S}_{common})}$. Finally, we can apply the matrix $\mathbf{R}$ and the scaling c to both common and non-common source data to align them with the target subject. We computed these values for Subj02, Subj05, and Subj07 as source subjects, using Subj01 as the target, to align all subjects to the functional space of the first one. For detailed mathematical proofs and other insights, please refer to the original articles (Haxby et al. 2020, 2011; Gower 1975).

Impact of the amount of data used for alignment

We also examined how the alignment performance fluctuates when the shared data makes up 10%, 25%, 50%, and 100% of the total common data (952 images). However, the goal here is not merely comparison, but rather the examination of performance in relation to the shared data fraction and alignment method, given a fixed decoding pipeline, trained solely on Subj01 as a reference target.

Results

Fig A2 (A) shows qualitatively decoding performances of our apporach in function of the amount of data used for alignment. Fig A2 (B) shows qualitative results using other decoding baselines (namely only anatomical, hyperalignment and our approach based on Ridge regression). Anatomical and Hyperalignment methods fail to yield satisfactory results, demonstrating just above chance performance levels for 2-way classification accuracy and poor performance for low-level metrics such as SSIM and PixCorr. However, Ridge Regression exhibits an increasing performance based on the volume of data used for alignment mapping function learning. This method reaches performance levels comparable with the within-subject decoder in both low-level and high-level metrics, using all the common data (approximately 10% of the entire dataset).

Anatomical method's inefficacy: As corroborated by previous studies (Haxby et al. 2020), our research found that anatomical methods for brain decoding are ineffective. Relying on the physical structure of the brain for alignment and decoding does not deliver the requisite precision for fine-grained decoding tasks. This could be attributed to inherent anatomical variability across different individuals, which may not necessarily align with functional differences. The specialized areas in the brain with functional selectivity can sometimes yield performance above chance levels. However, in most cases, decoded images do not correlate with the stimulus, undermining the reliability of this method for cross-subject brain decoding.

Overfitting tendency of complex techniques: We noted that more sophisticated techniques, like hyperscanning for brain decoding, tend to overfit the data. This results in poor generalization to unseen data, with metrics measuring n-way accuracy reaching only chance levels. While these techniques might seem to offer superior decoding accuracy initially, their lack of generalizability limits their practical utility. Of course, room for improvement exists, perhaps through the incorporation of regularization techniques.

Our research reveals the limitations of anatomical methods for brain decoding, which rely on the physical brain structure for alignment and decoding. These methods underperformed due to inherent brain anatomical variability across individuals, which may not align with functional differences. Thus, this study emphasizes the need for functional, not merely anatomical, considerations in decoding studies. Fig A4 plot all the quantitative metrics in function for all the baselines in function of the fraction of data used for alignment (except for anatomical alignment which is independent given is only-structural nature).

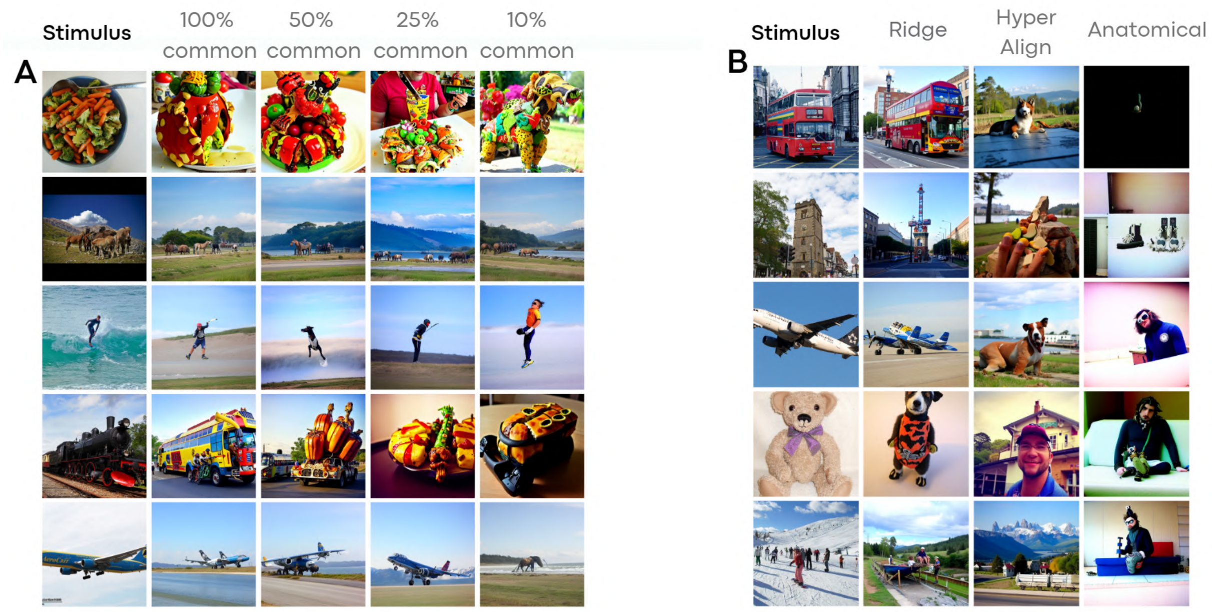

Figure A1: **A**: Functional alignment comparison using Ridge Regression across varying fractions of shared data. The ”Stimulus” column showcases experimental images, while subsequent columns depict the decoded and aligned activity of Subj02 based on Subj01. **B**: A comparison of distinct alignment techniques. The ”Stimulus” column again presents the experimental images, with the remaining columns illustrating the decoded activity of Subj02, aligned to Subj01 through various methodologies.

Figure A2: **A**: Functional alignment comparison using Ridge Regression across varying fractions of shared data. The ”Stimulus” column showcases experimental images, while subsequent columns depict the decoded and aligned activity of Subj02 based on Subj01. **B**: A comparison of distinct alignment techniques. The ”Stimulus” column again presents the experimental images, with the remaining columns illustrating the decoded activity of Subj02, aligned to Subj01 through various methodologies.

Impact of choice of the target subject

To systematically assess the influence of selecting distinct subjects for model training and alignment, we experimented with multiple combinations. For instance, the decoder was trained on Subject 1, followed by alignment of Subject 2 to this target, and subsequent decoding of Subject 2. This procedure was also executed with the decoder trained on Subject 2, alignment of Subject 1, and decoding of Subject 1. We extended this approach to encompass all potential combinations of our four subjects. As evidenced in figures A5 and A6, the qualitative nature of the decoded images remained consistent irrespective of the subjects chosen for training and alignment. These figures distinctly captured high-level content and foundational shapes across varying subject combinations, yielding analogous visual decoding results.

Within these figures, the diagonal cells represent within-subject decoding, wherein the model undergoes both training and testing on an identical subject. In contrast, off-diagonal cells signify cross-subject decoding, where distinct subjects are employed for training as opposed to alignment and decoding.

Despite the variations in quantitative metrics, the visual reconstructions derived from different combinations are qualitatively analogous. This underscores the decoder's proficiency in generalizing across diverse subjects. The presence of shared neural representations, even amidst individual disparities, facilitates precise cross-subject decoding across a spectrum of training and alignment configurations.

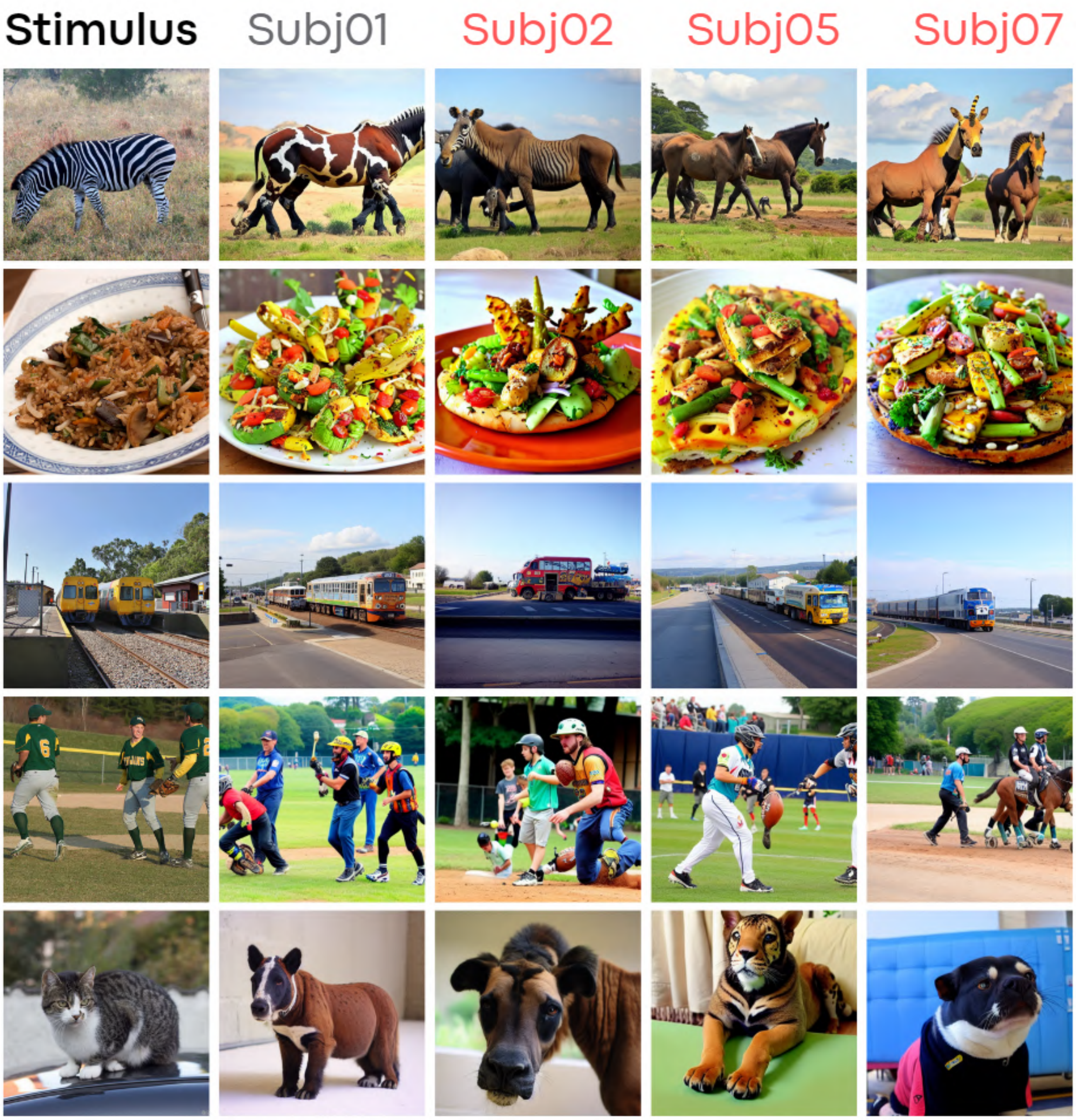

Figure A3: More example results. Format and conventions as in Figure 2

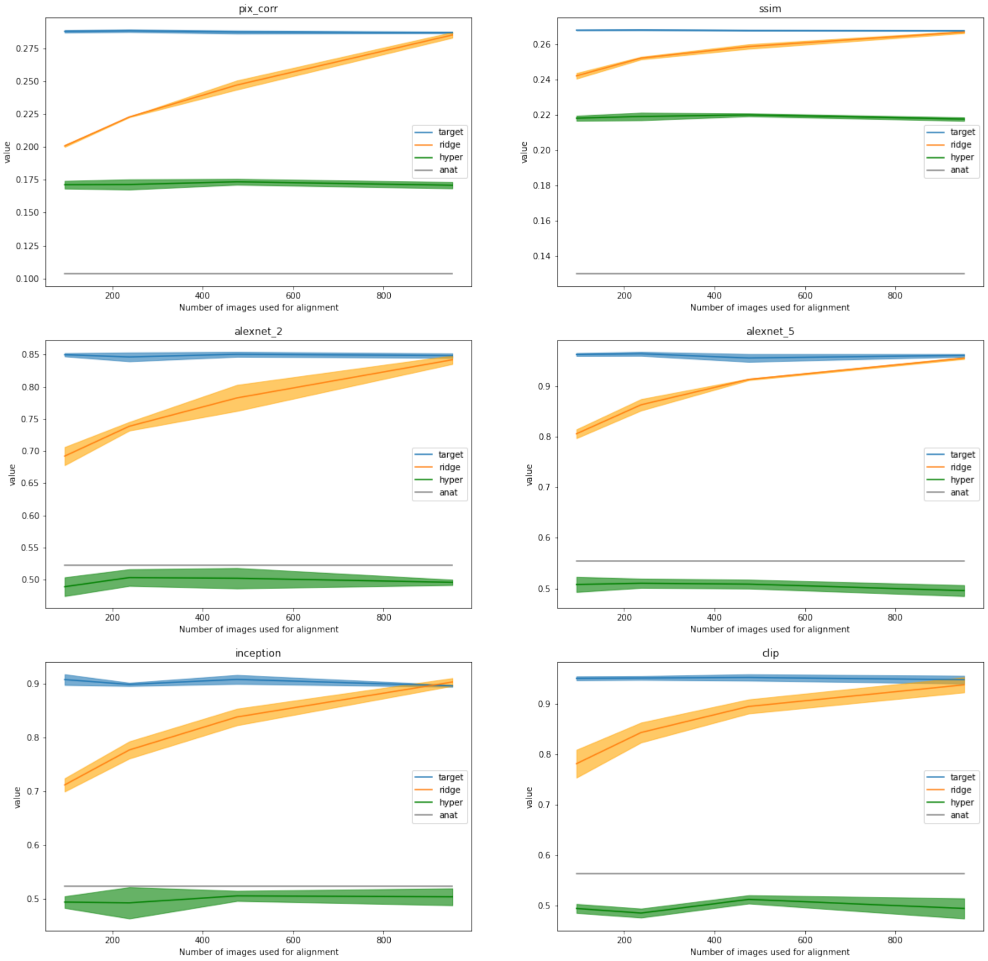

Figure A4: This figure illustrates the performance of various methods evaluated using different metrics. Blue lines represent metrics from the target subject's decoded images, derived from their test set brain activity. Green lines denote the mean and standard deviation (std) of performance on test sets from other subjects, aligned using hyperalignment. Gray lines present results achieved using anatomical alignment, while orange lines display outcomes using Ridge Regression. Remarkably, the Ridge Regression approach yields positive results even when using a tiny fraction of the entire dataset. Furthermore, as this fraction approaches roughly 10% of the total set, resulting in 952 images the performance becomes comparable with those obtained by the within-subject model.

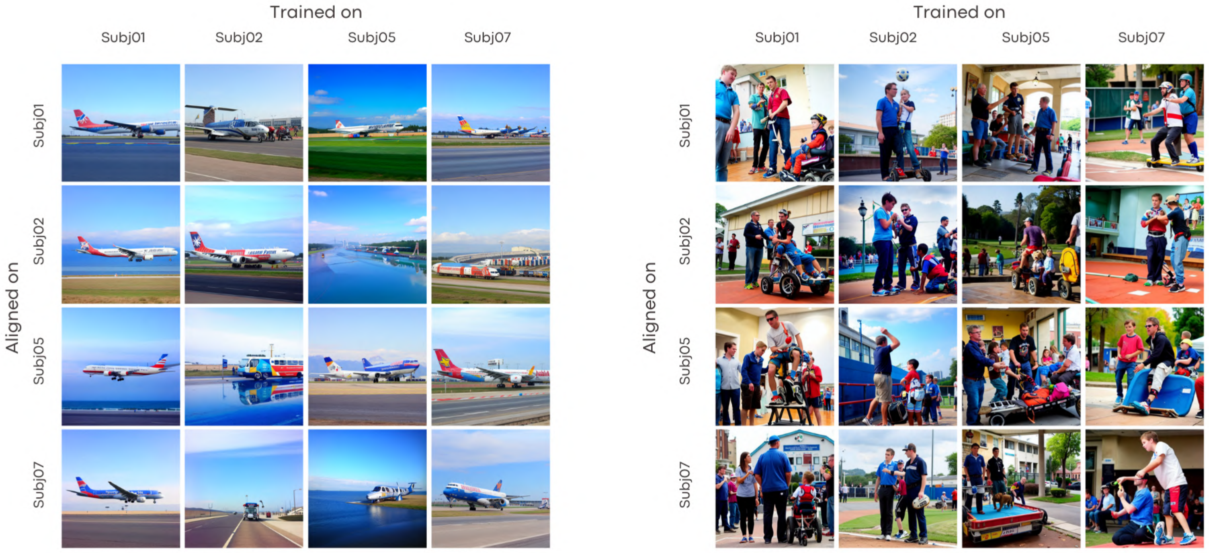

Figure A5: This figure illustrates decoding results from different combinations of subjects used for model training versus alignment. The columns represent decoders trained on individual target subjects. The rows show each remaining subject aligned to the target space of the column subject for decoding.

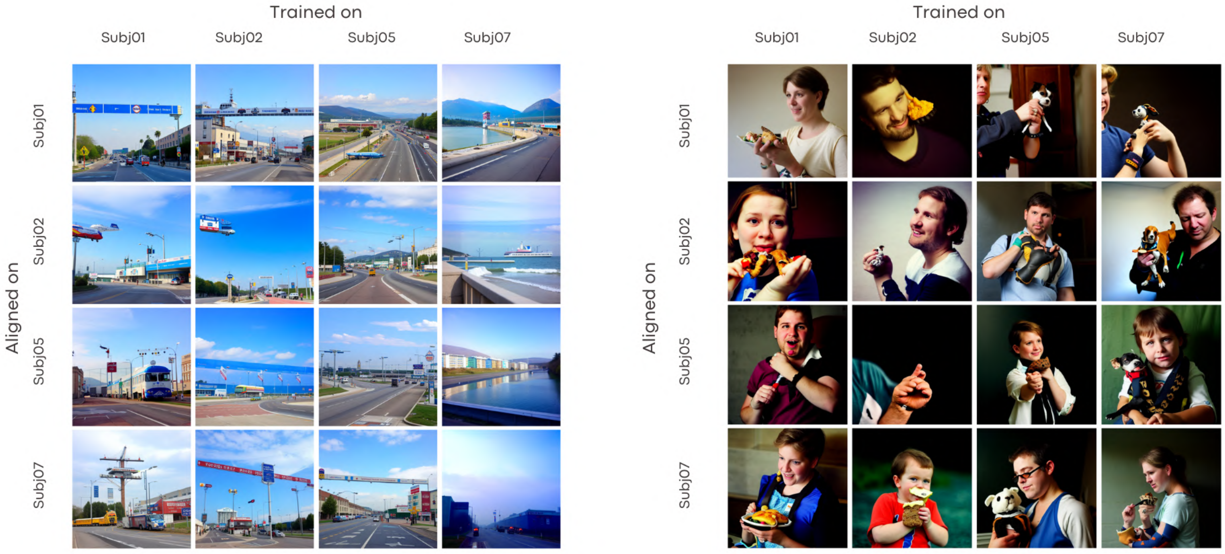

Figure A6: More examples of the impact of choosing a subject as a target. Same configuration as Fig A5

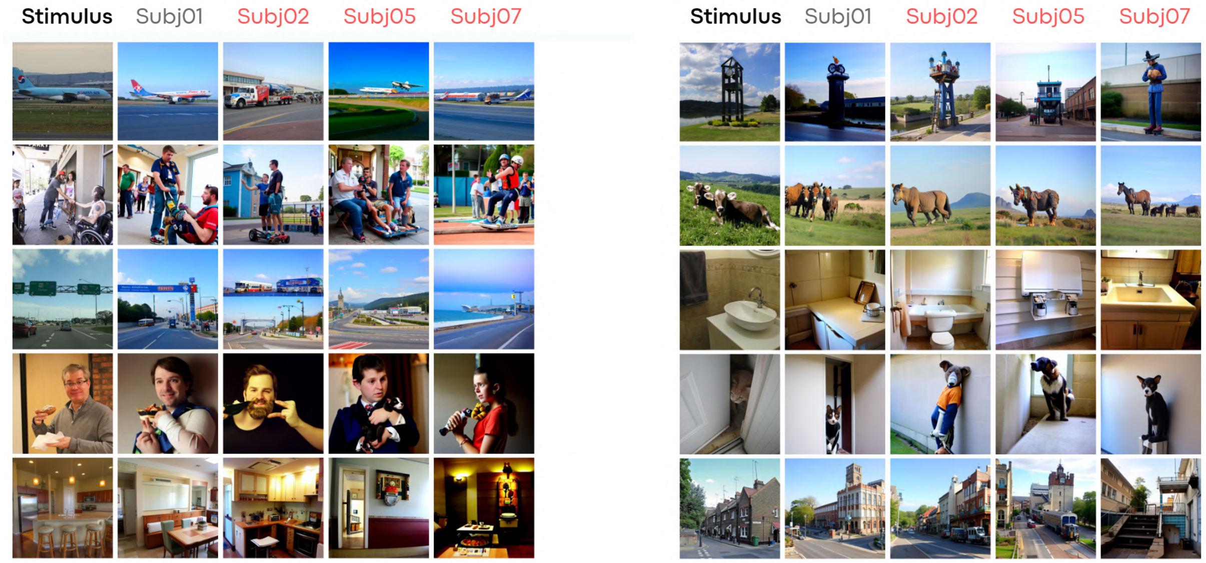

Figure A7: More example results. Format and conventions as in Figure 2

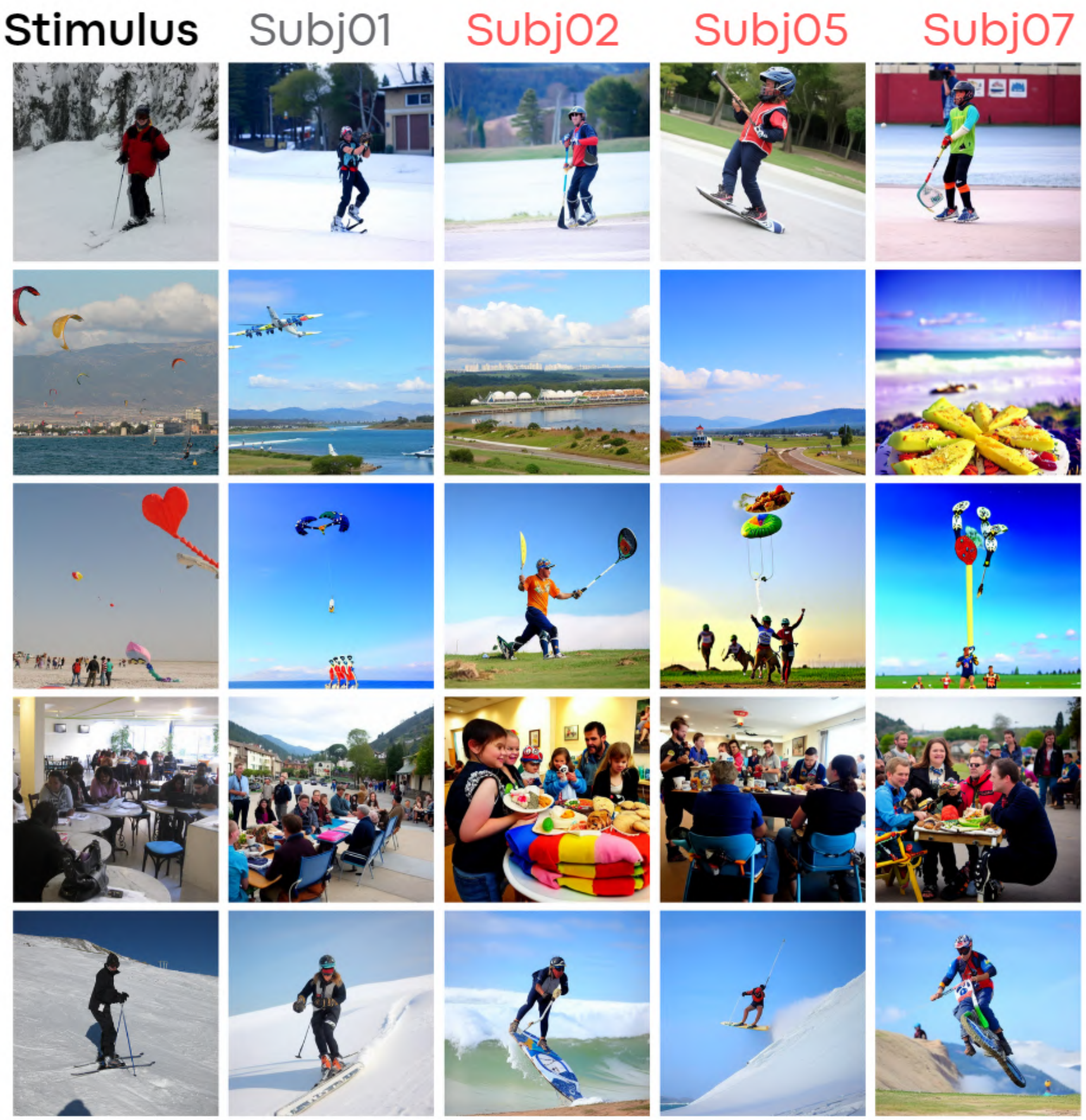

Figure A8: More example results. Format and conventions as in Figure 2