PEAKS2IMAGE: RECONSTRUCTING FMRI STATISTICAL MAPS FROM PEAKS

Anonymous authors
Paper under double-blind review

ABSTRACT

Neuroscience strives to overcome the lack of power due to the small sample size of imaging-based studies. An important step forward has been the creation of large-scale public image repositories, such as NeuroVault. Such repositories allow images to be compared across studies and automatically associated with cognitive terms. Yet, this type of meta-analysis faces a major roadblock: the scarcity and inconsistency of image annotations and metadata. Another resource containing rich annotations is the neuroscientific literature. However it only yields a handful of brain-space coordinates per publication, those of the main activity peaks reported in each study. This has led the community to mostly perform meta-analysis based on these reported coordinates. In this work, we propose Peaks2Image, a neural-network approach to reconstruct continuous spatial representations of brain activity from peak activation tables. Peaks2Image thus associates rich annotations from the neuroscientific literature with dense brain reconstructions. Using those reconstructions, we train a decoder using tf-idf features as labels, leading to a much broader set of decoded terms than current image-based studies. We validate the decoder on 43,000 NeuroVault images, successfully decoding 65 out of 81 concepts in a zero-shot setting.

1 INTRODUCTION

Cognitive neuroscience aims to map cognitive processes onto brain regions. Functional Magnetic Resonance Imaging (fMRI) is one of the most powerful techniques available to identify such associations. This approach measures brain activity while subjects perform cognitive tasks in an MRI scanner, and then contrasts brain signals associated with different mental conditions. Statistical testing of these contrasts then identifies brain regions where the neural activity elicited by the conditions probed is significantly different. Unfortunately, the high cost of data acquisition limits the number of participants and tasks involved in each study (Poldrack et al., 2017). A small number of participants causes low statistical power and a high proportion of false discoveries (Ioannidis, 2005; Button et al., 2013). Analyzing a restricted set of cognitive tasks introduces the risk of over-interpreting statistical effects that are not specific to the mental functions under study (Poldrack, 2011).

Meta-analysis consists in aggregating the results of several studies to find effects that are reported consistently (Wager et al., 2007). It helps overcome the challenge of small sample sizes and uncovers more reliable associations between brain activity and mental function. When feasible, meta-analysis should use the full statistical brain maps produced by the original studies (Salimi-Khorshidi et al., 2009). Unfortunately, the vast majority of studies does not share the actual brain images. Instead, only the locations of the peaks of activation are reported. These are communicated in the form of tables, in scientific publications, containing *stereotactic* coordinates – 3D coordinates in a standard spatial referential for the brain. This results in a poor representation of brain activity – most of the information contained in the original statistical maps is lost. Meta-analyses that rely on peak activation coordinates reported in publications are called *Coordinate-Based Meta-Analyses* (CBMA).

Recent efforts to openly share full brain images and statistical maps such as NeuroVault (Gorgolewski et al., 2015) or OpenNeuro (Gorgolewski et al., 2017), could facilitate *Image-Based Meta-Analysis* (IBMA) – relying on the brain images rather than coordinates. However, annotations available on those large databases are scarce and inconsistent, leading to difficulties to capture the semantics of the cognitive processes associated with images (Menuet et al., 2022). There is no standard,

agreed-upon ontology or vocabulary of mental functions. Moreover, all that is known is the task that the participant is performing, and ascribing mental functions to tasks is difficult (Poldrack et al., 2011). Therefore, formally describing the mental states that underlie a brain image remains an open problem.

Scientific literature and image repositories are therefore complementary: the literature provides rich descriptions of the studied cognitive tasks but poor spatial information, whereas image repositories contain complete representations of brain activity but lack useful annotations. In the present study, we introduce a meta-analysis method that combines the strengths of both.

Beyond the challenge of sample size, effects uncovered by individual neuroimaging studies suffer from a lack of specificity. When a cognitive task is studied in isolation, there is no way of knowing if the observed brain activations are specific to the mental functions of interest, or associated with a broader set of mental processes. Authors often over-interpret the observed associations, resulting in the fallacy of unwarranted *reverse inference* (Poldrack, 2011). One way of identifying more specific associations between brain activity and mental function is *decoding*: inferring the mental processes at play, given a brain image of neural activity. Indeed, to discriminate between a wide variety of cognitive states, a decoding model must identify brain regions that *characterize* each state, rather than brain regions that are merely activated consistently. For the resulting associations to be specific, it is crucial that many and diverse cognitive states are decoded jointly. Due to the difficulty of formalizing the mental states associated with a brain image, high-quality labels are lacking for this supervised task. Performing such large-scale (sometimes called "open-ended") decoding is therefore challenging, and in practice most studies that claim to perform "open-ended" decoding only discriminate a restricted set of cognitive concepts.

In this work, Peaks2Image learns to discriminate dozens of cognitive terms on the largest available image repository, NeuroVault. To do so, it leverages both the rich descriptions of mental processes found in the literature and the high-quality neural activity data found in full-brain statistical maps. Peaks2Image reduces the gap between IBMA and CBMA, by reconstructing brain maps from peaks coordinates contained in neuroscientific publications. We leverage brain images from an unlabeled dataset to extract peaks, and train Peaks2Image to reconstruct images from the extracted peaks. We use Peaks2Image to obtain for the first time images associated with neuroscientific studies that only provide stereotactic coordinates. We evaluate whether those reconstructions are relevant by using them for brain image decoding. We associate labels with the studies using some criteria on the term-frequency inverse-document-frequency of the text. We use the decoding architecture from Neural Networks on Dictionaries (NNoD) (Menuet et al., 2022) trained on the neuroscientific corpus. We evaluate the decoding performance against 81 terms from NeuroVault. Peaks2Image succesfully decodes 65 of them on thousands of brain images from the NeuroVault database, without using any supervision from NeuroVault samples during training. While the evaluation has been performed on a limited set of terms, Peaks2Image can decode in a zero-shot setting any term from its vocabulary.

2 Related Work

Automated meta-analysis (Laird et al., 2005; Yarkoni et al., 2011) has risen over the last few years to handle the growth of published neuroscientific studies. Handling neuroscientific concepts properly has emerged as a challenge for CBMA, leading to the use of more complex textual features and models. Dockès et al. (2020) broadened the spectrum of terms analyzed by mapping rare concepts to more common cognitive terms. Ngo et al. (2021) leveraged language models (Beltagy et al., 2019) to capture term relationships semantically, leading to the encoding of any query. However, CBMA suffers from the drastic information reduction inherent to peak reporting.

Using dense images yields more information for meta-analysis (Salimi-Khorshidi et al., 2009). Thanks to the rise of large-scale databases of fMRI brain images such as NeuroVault (Gorgolewski et al., 2015) or OpenNeuro (Gorgolewski et al., 2017), decoding can now be performed across multiple studies (Mensch et al., 2017; Walters et al., 2022). Nonetheless, annotations associated with those images are often of low quality. Automatic strategies to improve this labeling are necessary to benefit properly from the scale of the data (Poldrack and Yarkoni, 2016). Menuet et al. (2022) leverages the Cognitive Atlas (Poldrack et al., 2011) to improve the label quality of NeuroVault images, enabling the decoding of a large set of concepts. Overall, neuroscience meta-analyses have

increasingly relied on data-driven approaches, learning more efficiently across a large set of data (Beam et al., 2021).

Although there has been interest in enriching brain images to increase the amount of data (Zhuang et al., 2019), to generate or reconstruct fMRI images (Manning et al., 2018; Cai et al., 2016), there has been limited interest in generating reconstructions of statistical brain maps from peaks (Gorgolewski et al., 2019), probably due to the daunting data dimensionality. In particular, we found no attempt at merging rich textual information from neuroscientific studies with the dense spatial information of brain images.

We focus on this particular aspect, proving that obtaining image representations for neuroscientific studies helps overcome the limitations of CBMA by generating a corpus with both rich textual annotations and dense spatial representations.

3 METHOD

Neuroscientists most frequently extract peaks (local maxima) above a statistical threshold from brain images. An image containing hundreds of thousands voxels is thus reduced to a few datapoints – less than 100 for most studies. We aim to reconstruct brain images from those few datapoints. This reconstruction problem is an inverse problem between the set of peaks and the image from which they were extracted. From a given set of l peaks $\mathbf{P_y} \in \mathbb{R}^{3\times l}$ extracted from an image $\mathbf{y} \in \mathbb{R}^m$ with m voxels, we learn the transformation $f : \mathbb{R}^{3\times l} \mapsto \mathbb{R}^m$ by minimization of the $\mathbf{L_2}$ error over the original image:

$$f^* = \arg\min_f \mathbf{E_y} \| f(\mathbf{P_y}) - \mathbf{y} \|_2^2 \tag{1}$$

To solve the minimization problem in eq 1, we leverage a set of statistical maps taken from NeuroVault, an online repository of fMRI maps, from which we extract peak coordinates. This leads to a training set of peaks and the corresponding brain images to be reconstructed. Once trained, Peaks2Image can be applied to any set of peak coordinates from neuroscientific publications. This leads to a combination of both continuous brain images and extensively annotated studies, which opens the possibility to decode a larger set of cognitive concepts.

We show in Figure 1 a summary of Peaks2Image. The figure is divided between the learning of the reconstruction process from peaks (A), the application of the reconstruction to neuroscientific studies (B), and the decoding of brain images with a model trained from those reconstructions in a zero-shot setting.

3.1 RECONSTRUCTION

Peaks2Image consists of a neural network that reconstructs brain images from activity peaks. We tame the effect of outliers by rescaling image values to a standard range, thus losing the statistical scale of the effect but keeping the overall statistical distribution. Images are also resampled at a fixed resolution and masked to keep the brain volume only.

Instead of leveraging the peak coordinates directly as a sets problem (Zaheer et al., 2017), Peaks2Image builds an internal sparse Gaussian representation similar to the literature as by Laird et al. (2005); Dockès et al. (2020); Ngo et al. (2021). It performs a transformation between the Gaussian representation space and the image space. Formally, we consider $\mathbf{P} = (p_1, .., \mathbf{p}_l)$ peaks extracted from image $\mathbf{y}$ where $\forall i \in [1..l], \mathbf{p}_i \in \mathbb{R}^3$. Given a fixed sampling grid $\mathbf{G} = (\mathbf{g}_1, .., \mathbf{g}_m)$ representing the voxels, where $\forall j \in [1..m], \mathbf{g}_j \in \mathbb{R}^3$, and a Gaussian kernel κ we define the vector $\mathbf{x_y}$ that samples $\mathbf{P}$ on the grid $\mathbf{G}$: $\mathbf{x_y} = (x_1, .., x_m)$ with $\forall j \in [1..m] : x_j = \sum_{i=1}^{l} \kappa(p_i, g_j)$.

High-resolution is not needed here, as population-level images, that represent the average of individual data with variable shapes, are intrinsically smooth. We can thus perform a drastic dimension reduction of the data, which is welcome given that the sample size is inherently limited. We use the Dictionary of Functional Modes (DiFuMo) probabilistic atlas (Dadi et al., 2020), that efficiently reduces the $m = 50,000$ voxel values to $k = 1024$ components. DiFuMo has been trained on thousands of resting-state and task fMRI with non-negativity and sparsity constraints, making this atlas

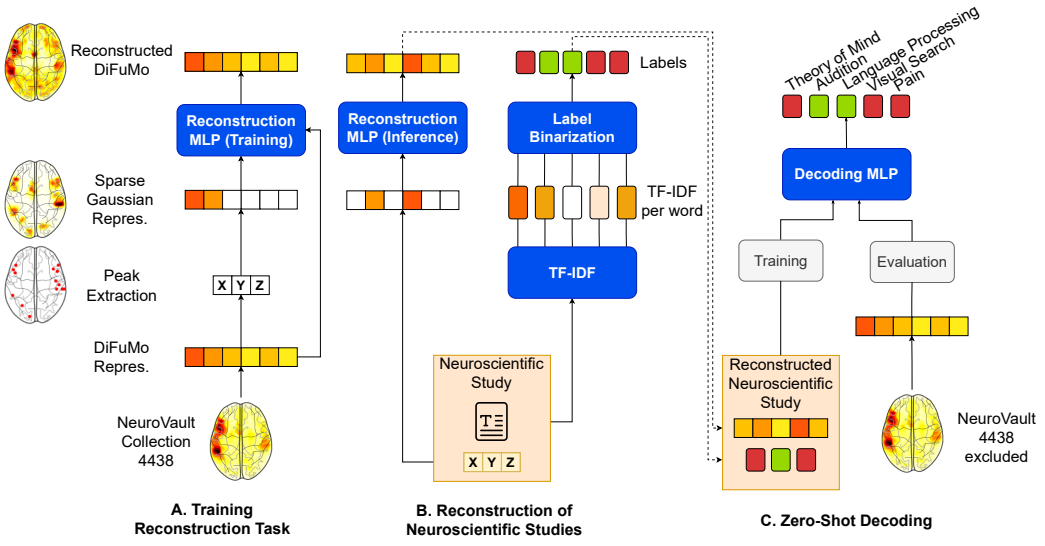

Figure 1: **Reconstructing Brain Activation Maps from Activation Peaks** **(A)** We learn the reconstruction process from NeuroVault collection #4438. After selection and preprocessing of the images, we extract the peaks that are above a statistical threshold. From those peaks, we build an intermediate representation by placing a Gaussian kernel at each peak as in Dockès et al. (2020); Ngo et al. (2021). We compress the representation using the DiFuMo atlas. We regress the original DiFuMo components of each image by leveraging a 3-layer multi-layer perceptron (MLP). **(B)** Once the reconstruction model is trained, we build intermediate Gaussian-kernel representations for the scraped studies, and reconstruct a brain map for each study using the above model, supporting further image-level computations such as decoding. From the text, we extract the tf-idf values for each word, and infer labels on a percentile-based threshold. **(C)** We perform decoding of NeuroVault images by training on the newly acquired representations of neuroscientific studies. We decode a broad set of cognitive processes benefiting from the large vocabulary.

relevant to represent population-level information. In the rest of the paper, the term *brain image* refers to its 1024-dimensional DiFuMo representation.

Mapping voxels $\mathbf{x} \in \mathbb{R}^m$ to coefficients $\alpha \in R^k$ along DiFuMo dictionary $\mathbf{D} \in \mathbb{R}^{m \times k}$ of dimension k is a mere linear regression problem following equation 2. All the following steps are performed in the k-dimensional DiFuMo space.

$$\alpha(\mathbf{x}) = \underset{\alpha \in R^k}{\arg\min} \|\mathbf{x} - \mathbf{D}\alpha\|_2^2 \tag{2}$$

The DiFuMo-encoded sparse Gaussian representation is turned into a dense representation with a 3-layer neural network where the input and output are DiFuMo coefficients. Using the DiFuMo atlas instead of the raw voxels helps us keep a relatively low number of parameters for this model. This transforms Eq 1 into:

$$f^* = \underset{f}{\arg\min}\, \mathbf{E}_{\mathbf{y}} \|f(\alpha(\mathbf{x}_\mathbf{y})) - \alpha(\mathbf{y})\|_2^2 \;\text{ where }\; \mathbf{x} = \sum_i \kappa(\mathbf{P}_{\mathbf{y}_i}, \mathbf{G}) \tag{3}$$

To train this reconstruction process, we use around 6k images from the #4438 NeuroVault collection (`https://neurovault.org/collections/4438`).

The reconstructed images are obtained from the predicted DiFuMo components by linear combinations of the dictionary components: $\hat{\mathbf{y}} = \mathbf{D} f^*(\alpha(\mathbf{x}_\mathbf{y}))$.

3.2 DECODING FROM PEAKS

Peaks2Image outputs images from any set of positions. We apply it to neuroscientific publications to yield a set of dense brain maps accompanied by rich textual annotations, reducing the technical gap between CBMA and IBMA. By doing this on the most comprehensive public literature corpus, we reconstruct 13,000 brain images from the peak activation tables.

We use these representations to decode cognitive processes in a zero-shot multi-label setting. Indeed, the ability to decode cognitive concepts from the produced database ensures its validity and circumvents the biases inherent to image reconstruction quality metrics, such as the confounding effect of smoothness. We evaluate the decoding performance on the NeuroVault database (#4438 excluded).

We extract the term frequency-inverse document frequency (tf-idf) (Salton and Buckley, 1988) from the studies' text. The tf-idf reflects the relevance of a specific term to a study by giving higher power to terms that occurs repetitively in few studies. Based on it, we assign labels to studies, by setting a threshold at the 95th percentile of the tf-idf. This ensures that the few studies that are most relevant to each term get positive labels. The computed percentile can be null for some extremely rare terms. When the percentile corresponds to a null tf-idf value, we only consider strictly positive tf-idf values.

We leverage a 3-layer dense neural network to predict the set of binary labels $\mathbf{L} = \{0,1\}_1 \times \cdots \times \{0,1\}_l$, l denoting the total number of classes, from the input image $\mathbf{X} \in \mathbb{R}^d$ compressed in DiFuMo space. The model is trained from the sole neuroscientific studies, without supervision from any NeuroVault sample or its labels. We were limited for the evaluation by the annotations available in NeuroVault. In the following experiments, we predict only l=81 terms from the 6308 in the tf-idf vocabulary. Peaks2Image could extend to a set of cognitive terms as large as the vocabulary size without any additional cost.

4 EXPERIMENTS

4.1 DATASET DESCRIPTION

We leverage the NeuroVault database. We exclude a subset of images according to a set of rules such as having incoherent range of values (indicating that an image does not contain Z or T statistics), duplicated images or images with missing metadata to keep around 50k brain maps for our task. We split those maps into two different sets in a collection-wise manner. We take collection #4438, without labels (6k maps), to train the reconstruction task. We consider the remaining 43k brain maps from 2376 different collections to evaluate the decoding performances.

We use NeuroQuery Data Collection (Dockès et al., 2020) to fetch around 13k neuroscientific studies from PubMed Central. The tool helps us parse the activation tables of those studies, totalling 400k peaks. It also transforms each study text into tf-idf features across a vocabulary of 6308 terms.

For validation, we restrict the analysis to shared terms between the tags of NeuroVault images and the vocabulary from the tf-idf. Among the 240 tags from the NeuroVault test collections, 189 are present in the vocabulary. We focus on those 189 tags without adding any additional mappings between the left-out tags and words from the vocabulary. Additionally, we filter out terms that contain less than 50 positive examples in the test set to prevent any non-significant effects, keeping 81 terms to decode.

4.2 DATA PREPROCESSING

Image preprocessing Images were fetched from NeuroVault using the Nilearn library (Abraham et al., 2014). We resample all images to the standard MNI152 template and apply the MNI152 brain mask. We use the DiFuMo atlas in dimension 1024 available in Nilearn. As the coefficients of the DiFuMo representation can range to large values, we scale them by a constant to facilitate the convergence of the neural network.

Peak extraction We generate pairs of matching (peaks, image) pairs from the collection (#4438) from NeuroVault images only. Similarly, we apply masking, resampling and DiFuMo compression without the normalization step. We use Nilearn's method for peak extraction, which first separates

the images into statistically active clusters and retrieves peak coordinates from each cluster. As these images represent statistical maps, equivalent to p-values of voxel-level tests, we keep peaks with a p-value lower than 10^{-3}, and with a minimal distance between peaks within the same cluster of 8mm. These are standard parameters from peak extraction in brain imaging contexts.

Sparse Gaussian Representations From a given set of peaks, we build a sparse Gaussian representation by placing Gaussian kernels at each peak position with a 9mm full-width at half-maximum, following what is done in Dockès et al. (2020). Similarly to the NeuroVault images, we extract the DiFuMo representations of the sparse Gaussian maps using Nilearn.

4.3 Training details

Reconstruction The image reconstruction model is an MLP built from 3 dense layers with 1024 units and sigmoid activation implemented using PyTorch (Paszke et al., 2019). It is trained for 50 epochs using Adam optimizer and a learning rate of 1e-3. The loss is the mean-squared error (MSE) over the DiFuMo components. We select the best set of weights over the validation set, which is composed of 20% of collection #4438.

We compare Peaks2Image to linear models (linear regression, kernel Ridge). We use implementations from scikit-learn (Pedregosa et al., 2011) with default parameters.

Decoding The decoder model is an MLP composed of one hidden layer of 300 units and ReLU activation. It is trained for 200 epochs on the reconstructed neuroscientific studies using the Adam optimizer with a learning rate of 3e-3 on the binary cross-entropy loss. During evaluation, we apply an additional sigmoid operation to the model outputs. We use the Area Under the Receiver Operating Characteristic Curve (ROC AUC) over each label of the testing set to evaluate the decoding performance. Chance level is 0.5.

We performed 20 different runs of the experiments with different splits to obtain confidence intervals.

5 Results

5.1 Reconstruction

We evaluate the reconstruction produced by Peaks2Image on all NeuroVault except collection #4438 which represents a testing set of around 43k images. As Peaks2Image reconstructs a normalized image, we use as metrics both the mean-squared error that has been used at training time, and the correlation between the original and its reconstruction in the voxel space. Even though the model has been trained with respect to mean-squared error in DiFuMo space, we find correlation more appropriate to measure the effects in those maps.

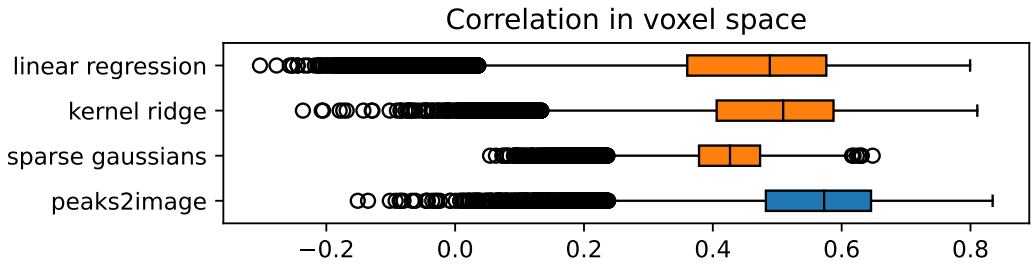

Figure 2: **Correlation between original images and reconstructions**: We evaluate the reconstruction for different models (Linear regression, kernel Ridge regression, sparse Gaussian representation and Peaks2Image) using correlation in voxel space. Evaluation is performed on collection 4337 from NeuroVault (18k images).

We show the reconstruction performance of each model in Fig. 2. Peaks2Image significantly outperforms linear models. As shown in the next section, although the correlations obtained with the

different methods are comparable, the Peaks2Image reconstruction yields a much better performance on the downstream decoding task. This suggests that some aspects of the image reconstruction are not captured by simple metrics such as the MSE or correlation. We plot examples of reconstructions for different labels in Fig. 7.

5.2 Decoding performance

We validate the relevance of the reconstructions by decoding NeuroVault images using reconstructed neuroscientific studies as training set. The problem is a multi-label classification, which we evaluate by computing the ROC AUC over each label. We exclude collection #4438 that has been used as training set in the reconstruction task, and evaluate the decoding of all remaining images from NeuroVault. We apply label enrichment to the original image labels. This includes regular expressions to map different terminologies for a same concept, and a hierarchy-based inference of labels based on the Cognitive Atlas ontology. For a given term, we assign parent terms from the ontology used in (Menuet et al., 2022).

This results in 43k images, totalling around 200 different cognitive processes. We evaluate only cognitive processes with at least 50 positive examples to avoid results with a too small sample. This reduces the evaluation to 81 terms. The limitation of terms is induced by the evaluation against NeuroVault, but Peaks2Image could extend the prediction to any word included in the tf-idf vocabulary.

In Fig. 3, we summarize the decoding performance by aggregating over Cognitive Atlas categories. Peaks2Image performs well over all the main concept families of Cognitive Atlas.

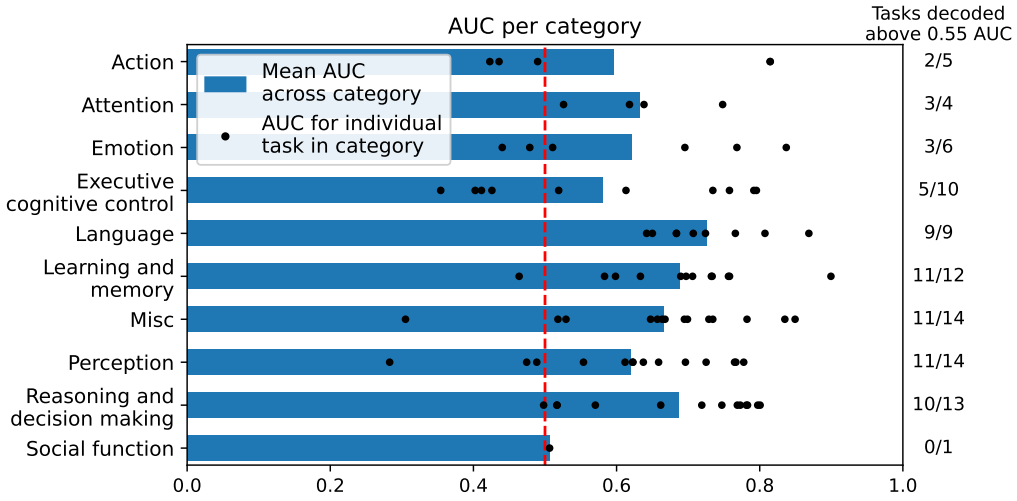

Figure 3: **Zero-shot decoding performance on NeuroVault**: evaluation is performed on all NeuroVault except collection #4438. We exclude categories with less than 50 positive examples for stability. Peaks2Image decodes better than chance 65 out of 81 cognitive processes over 43k images (AUC > 0.55).

We show the importance of the reconstruction process in the decoding performance by running the same decoding task over different representations of the input. In particular, we compare Peaks2Image reconstructions to the sparse Gaussian representations used in the literature. We show that, even though the correlation of sparse Gaussian representations remains close to Peaks2Image reconstruction, Peaks2Image largely outperforms the sparse Gaussian model on the decoding task. We observe a statistically significant gain of 0.1 on the mean AUC across labels between Peaks2Image and sparse Gaussian representations. We report the performance of each method in table 1.

We also compare Peaks2Image to approaches from the literature. On terms that are shared with NeuroSynth (Yarkoni et al., 2011), we show that Peaks2Image outperforms the former, while largely expanding the number of cognitive processes decoded.

Reconstruction Architecture	Decoding Model	Mean AUC across tasks
Peaks2Image	Logistic Regression	0.503 ($\pm$ 0.001)
Sparse Gaussian	NNoD	0.613 ($\pm$ 0.004)
Kernel Ridge	NNoD	0.638 ($\pm$ 0.004)
Peaks2Image	NNoD	**0.698 ($\pm$ 0.002)**

Table 1: **AUC performance of Peaks2Image and baselines**: decoding from Peaks2Image reconstructions outperforms the simple sparse Gaussian representation, as well as a kernel Ridge learner. Using the NNoD model with a hidden layer brings strong power over mere Logistic Regression.

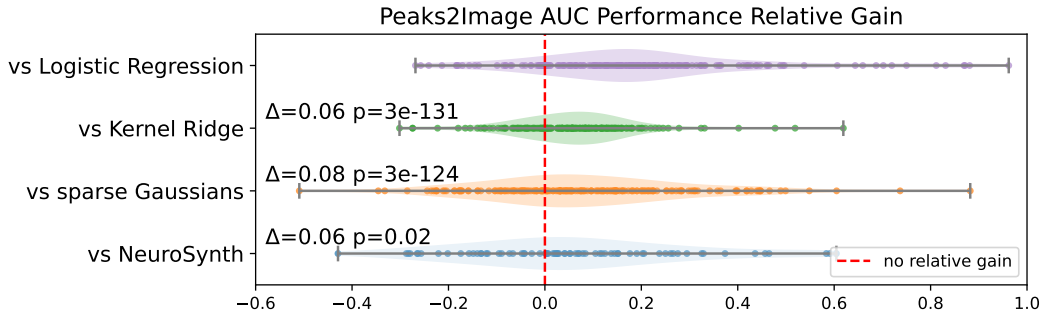

Figure 4: **Distribution of relative AUC performance across tasks of Peaks2Image compared to other representations**: We show the relative AUC gain of Peaks2Image for each task, along with the difference of mean AUC across labels vs different methods. We perform a t-test to probe the statistical significance of the difference between Peaks2Image and other methods. We also display the average AUC difference (Δ). For NeuroSynth, we compute the difference on a subset of shared terms.

We plot in figure 5 the difference between NNoD performance and Peaks2Image performance. Particularly, we add to this graph the performance of Peaks2Image on unseen terms, showing that Peaks2Image enables the successful decoding of a broader set of terms. Peaks2Image cannot compete with NNoD on terms that are in the training set. NNoD has been trained specifically on NeuroVault images, leading logically to better decoding performances. However, Peaks2Image's value is that it makes it possible to decode unseen terms, leading to a larger set of decoded terms while keeping relatively good performances where NNoD outperforms it.

We noted an effect of NNoD being able to decode some terms unseen during training. We made sure that no data leakage occurred and compared results with other linear models which reached an expected 0.5 AUC. Our intuition is that the multi-label setting during training and the multi-layer structure lead to a compression effect akin to PCA, leading to above chance decoding levels for certain tasks. This effect is negatively correlated with the number of positive samples in the testing set.

6 Discussion

We introduced Peaks2Image, a model for reconstructing brain images from peak activity coordinates. By reconstructing dense spatial representations, Peaks2Image bridges the gap between neuroscientific literature and brain image databases. Compared to images-based models, it allows to decode a much broader set of cognitive processes by leveraging the rich textual information of neuroscientific studies, along with the reconstructions. Most importantly, the decoding performance is reached without any labeling requirements on the images side, avoiding a time-consuming and error-prone task.

A limitation of Peaks2Image is our strategy regarding the studies text. It hinders comparisons by limiting to exact matches between vocabulary and tags from NeuroVault. While Peaks2Image could extend to any word contained in the neuroscientific literature, evaluation was reduced to a limited number of concepts due to the lack of extensively annotated brain images. Exploration of more

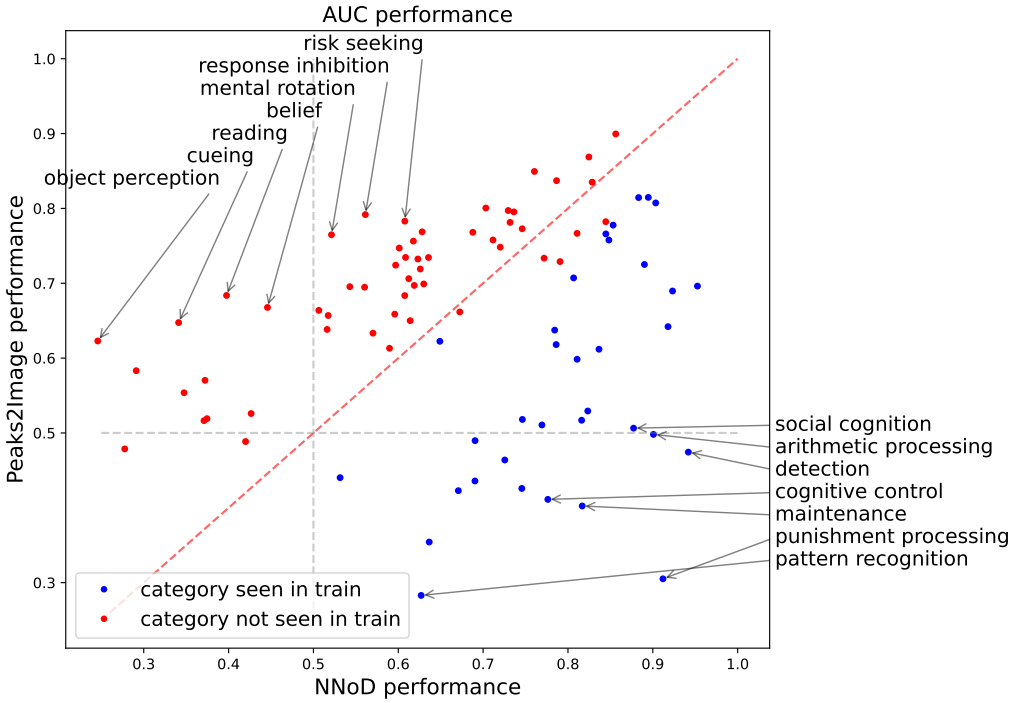

Figure 5: **AUC performance of our reconstructions vs NeuroVault samples**: we compare the performance of the NNoD architecture on NeuroVault (except collection #4438 used for the reconstruction training) in two settings: training on NeuroVault data (collection 4438) or training on reconstructions with percentile-inferred labels. NeuroVault training outperforms when it contains the term considered, but Peaks2Image provides a better guess when no such training data are available.

powerful language models as in Ngo et al. (2021) is a future direction to better benefit from the continuous brain images generated for the neuroscientific studies. Though, the poor quality of NeuroVault annotations would still limit our ability to evaluate those models.

Further investigations are still necessary to study the impact of the dataset used for training. A promising avenue consists in relying on abundant resting-state data, that do not carry cognitive labels, yet display a rich repertoire of topographical maps, that can be extracted e.g. with Independent components analysis Smith et al. (2013).

Peaks2Image paves the way towards better generation of large sets of images for neuroscientific studies. It shows that one can outrun the limitations of large-scale poor labeling quality by performing zero-shot decoding. In particular, neuroscientific studies could become a challenging application for recent methods in text-to-image representation learning such as Radford et al. (2021).

7 Reproducibility Statement

Data was collected from public sources using either the NQDC module from NeuroQuery or the Nilearn library. All scripts to reproduce the experiments and figures will be released publicly after review, along with the downloading scripts and preprocessed data. In particular, we will share the DiFuMo components of all images from our current version of the NeuroVault dataset. For neuroscientific studies, we leveraged the data available at `https://github.com/neuroquery/neuroquery_data/tree/main/data`.

REFERENCES

A. Abraham, F. Pedregosa, M. Eickenberg, P. Gervais, A. Mueller, J. Kossaifi, A. Gramfort, B. Thirion, and G. Varoquaux. Machine learning for neuroimaging with scikit-learn. *Frontiers in neuroinformatics*, 8:14, 2014.

E. Beam, C. Potts, R. A. Poldrack, and A. Etkin. A data-driven framework for mapping domains of human neurobiology. *Nature neuroscience*, 24(12):1733–1744, 2021.

I. Beltagy, K. Lo, and A. Cohan. Scibert: A pretrained language model for scientific text. *arXiv preprint arXiv:1903.10676*, 2019.

K. S. Button, J. P. Ioannidis, C. Mokrysz, B. A. Nosek, J. Flint, E. S. Robinson, and M. R. Munafò. Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5):365, 2013.

M. Cai, N. W. Schuck, J. W. Pillow, and Y. Niv. A bayesian method for reducing bias in neural representational similarity analysis. *Advances in Neural Information Processing Systems*, 29, 2016.

K. Dadi, G. Varoquaux, A. Machlouzarides-Shalit, K. J. Gorgolewski, D. Wassermann, B. Thirion, and A. Mensch. Fine-grain atlases of functional modes for fmri analysis. *NeuroImage*, 221: 117126, 2020.

J. Dockès, R. A. Poldrack, R. Primet, H. Gözükan, T. Yarkoni, F. Suchanek, B. Thirion, and G. Varoquaux. NeuroQuery, comprehensive meta-analysis of human brain mapping. *Elife*, 9, Mar. 2020.

K. Gorgolewski, O. Esteban, G. Schaefer, B. Wandell, and R. Poldrack. Openneuro—a free online platform for sharing and analysis of neuroimaging data. *Organization for human brain mapping. Vancouver, Canada*, 1677(2), 2017.

K. J. Gorgolewski, G. Varoquaux, G. Rivera, Y. Schwarz, S. S. Ghosh, C. Maumet, V. V. Sochat, T. E. Nichols, R. A. Poldrack, J.-B. Poline, et al. Neurovault. org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Frontiers in neuroinformatics*, 9: 8, 2015.

K. J. Gorgolewski, T. Yarkoni, and R. A. Poldrack. peaks2maps: reconstructing unthresholded statistical maps from peak coordinates using deep neural networks. *F1000Research*, 8, 2019.

J. P. Ioannidis. Why most published research findings are false. *PLoS medicine*, 2(8):e124, 2005.

A. R. Laird, J. J. Lancaster, and P. T. Fox. Brainmap. *Neuroinformatics*, 3(1):65–77, 2005.

J. R. Manning, X. Zhu, T. L. Willke, R. Ranganath, K. Stachenfeld, U. Hasson, D. M. Blei, and K. A. Norman. A probabilistic approach to discovering dynamic full-brain functional connectivity patterns. *NeuroImage*, 180:243–252, 2018. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2018.01.071. URL `https://www.sciencedirect.com/science/article/pii/S1053811918300715`. New advances in encoding and decoding of brain signals.

A. Mensch, J. Mairal, D. Bzdok, B. Thirion, and G. Varoquaux. Learning neural representations of human cognition across many fmri studies. *Advances in neural information processing systems*, 30, 2017.

R. Menuet, R. Meudec, J. Dockès, G. Varoquaux, and B. Thirion. Comprehensive decoding mental processes from web repositories of functional brain images. *Sci. Rep.*, 12(1):7050, Apr. 2022.

G. H. Ngo, M. Nguyen, N. F. Chen, and M. R. Sabuncu. Text2brain: Synthesis of brain activation maps from free-form text query. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 605–614. Springer, 2021.

A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. URL http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

R. A. Poldrack. Inferring mental states from neuroimaging data: from reverse inference to large-scale decoding. *Neuron*, 72(5):692–697, 2011.

R. A. Poldrack and T. Yarkoni. From brain maps to cognitive ontologies: informatics and the search for mental structure. *Annual review of psychology*, 67:587, 2016.

R. A. Poldrack, A. Kittur, D. Kalar, E. Miller, C. Seppa, Y. Gil, D. S. Parker, F. W. Sabb, and R. M. Bilder. The cognitive atlas: toward a knowledge foundation for cognitive neuroscience. *Frontiers in neuroinformatics*, 5:17, 2011.

R. A. Poldrack, C. I. Baker, J. Durnez, K. J. Gorgolewski, P. M. Matthews, M. R. Munafò, T. E. Nichols, J.-B. Poline, E. Vul, and T. Yarkoni. Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nature reviews neuroscience*, 18(2):115–126, 2017.

A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, et al. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, pages 8748–8763. PMLR, 2021.

G. Salimi-Khorshidi, S. M. Smith, J. R. Keltner, T. D. Wager, and T. E. Nichols. Meta-analysis of neuroimaging data: a comparison of image-based and coordinate-based pooling of studies. *Neuroimage*, 45(3):810–823, 2009.

G. Salton and C. Buckley. Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24(5):513–523, 1988.

S. M. Smith, C. F. Beckmann, J. Andersson, E. J. Auerbach, J. Bijsterbosch, G. Douaud, E. Duff, D. A. Feinberg, L. Griffanti, M. P. Harms, M. Kelly, T. Laumann, K. L. Miller, S. Moeller, S. Petersen, J. Power, G. Salimi-Khorshidi, A. Z. Snyder, A. T. Vu, M. W. Woolrich, J. Xu, E. Yacoub, K. Uğurbil, D. C. Van Essen, and M. F. Glasser. Resting-state fMRI in the Human Connectome Project. *Neuroimage*, 80:144–168, Oct 2013.

T. D. Wager, M. Lindquist, and L. Kaplan. Meta-analysis of functional neuroimaging data: current and future directions. *Social cognitive and affective neuroscience*, 2(2):150–158, 2007.

J. Walters, M. King, P. G. Bissett, R. B. Ivry, J. Diedrichsen, and R. A. Poldrack. Predicting brain activation maps for arbitrary tasks with cognitive encoding models. *NeuroImage*, page 119610, 2022.

T. Yarkoni, R. A. Poldrack, T. E. Nichols, D. C. Van Essen, and T. D. Wager. Large-scale automated synthesis of human functional neuroimaging data. *Nature methods*, 8(8):665–670, 2011.

M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. R. Salakhutdinov, and A. J. Smola. Deep sets. *Advances in neural information processing systems*, 30, 2017.

P. Zhuang, A. G. Schwing, and O. Koyejo. Fmri data augmentation via synthesis. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 1783–1787, 2019. doi: 10.1109/ISBI.2019.8759585.

A APPENDIX

A.1 RECONSTRUCTION OF NEUROVAULT IMAGES

We show an example of the representation at different steps of the Peaks2Image model in figure 6.

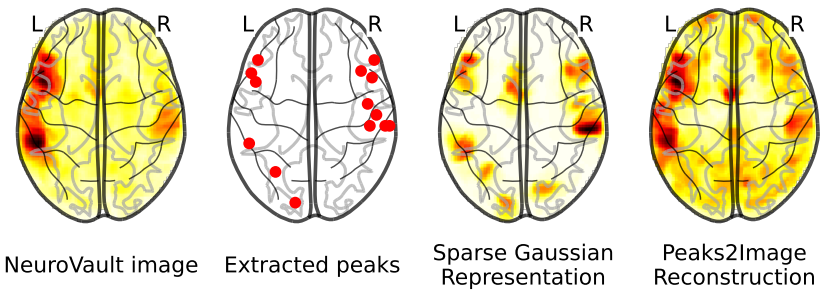

Figure 6: **Successive steps of Peaks2Image**: we show the successive steps of Peaks2Image representations. From left to right, we collect brain images from NeuroVault from which we extract peaks above a statistical threshold, we transform the peaks into a sparse Gaussian representation from which we predict the Peaks2Image reconstruction.

We also apply Peaks2Image to NeuroVault images that are associated to a set of cognitive terms. We compute the average of those reconstructions to show that Peaks2Image generates brain images that are consistent with expected patterns from the literature (fig 7).

A.2 PER-LABEL DECODING PERFORMANCE

We ran 20 runs of the experiment. In figure 8, we report the per-label decoding performance along with the variance of the results. Most terms are decoded above chance.

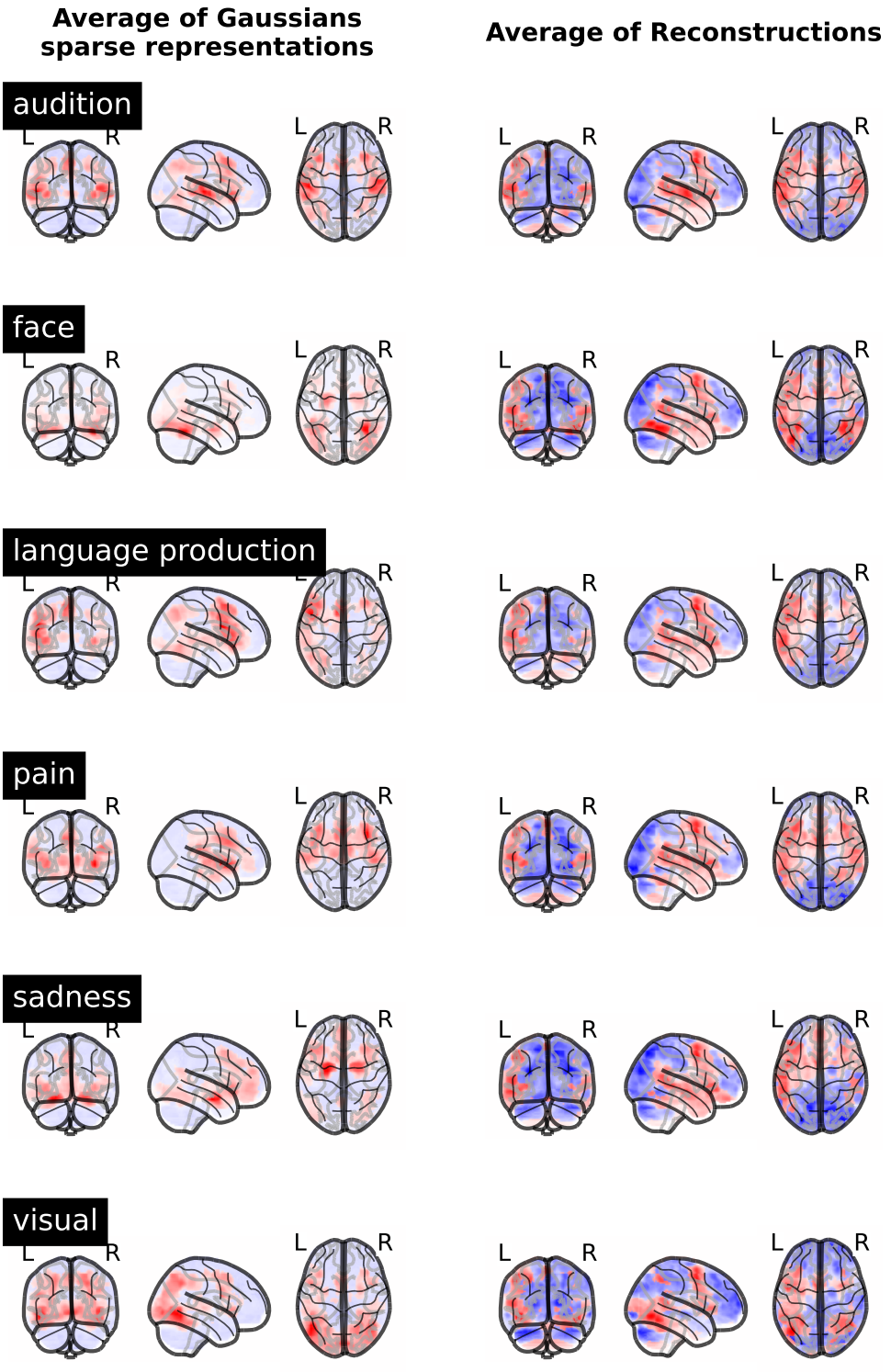

Figure 7: **Reconstructions of NeuroVault samples**: we show the average sparse Gaussian representation used in standard meta-analysis procedures (left) along with the Peaks2Image reconstruction (right) for NeuroVault samples containing certain terms. The reconstruction yields more extended networks (audition, face, language production) and sometimes biases the image away from the initial map (sadness). In some rare cases (visual) it focuses the activity patterns.

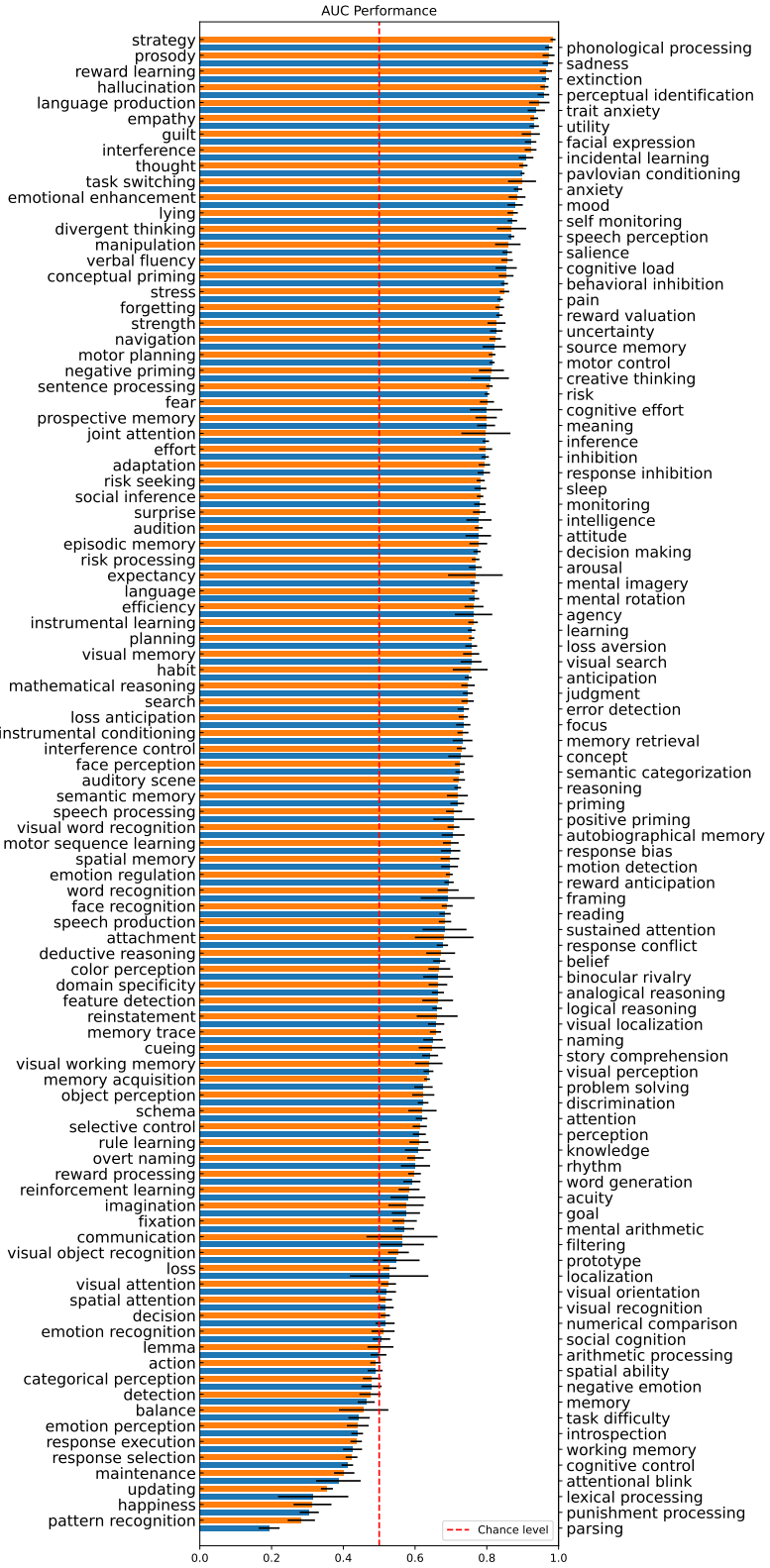

Figure 8: **Zero-shot decoding performance on NeuroVault**: we use Peaks2Image to produce dense brain images for neuroscientific studies. We train a decoder from the generated data on a broad set of cognitive terms. We evaluate the decoding performance on NeuroVault data. Peaks2Image succesfully decodes a large part of those terms in a zero-shot setting. Peaks2Image could extend to any word from the studies' vocabulary, but could be evaluated on the NeuroVault annotations only.