Mind-to-Image: Projecting Visual Mental Imagination of the Brain from fMRI

Hugo Caselles-Dupré [1] **Charles Mellerio** [2 3 4] **Paul Hérent** [5] **Alizée Lopez-Persem** [6] **Benoit Béranger** [7]
Mathieu Soularue [1] **Pierre Fautrel** [1] **Gauthier Vernier** [1] **Matthieu Cord** [8]

Abstract

The reconstruction of images observed by subjects from fMRI data collected during visual stimuli has made strong progress in the past decade, thanks to the availability of extensive fMRI datasets and advancements in generative models for image generation. However, the application of visual reconstruction has remained limited. Reconstructing visual imagination presents a greater challenge, with potentially revolutionary applications ranging from aiding individuals with disabilities to verifying witness accounts in court. The primary hurdles in this field are the absence of data collection protocols for visual imagery and the lack of datasets on the subject. Traditionally, fMRI-to-image relies on data collected from subjects exposed to visual stimuli, which poses issues for generating visual imagery based on the difference of brain activity between visual stimulation and visual imagery. For the first time, we have compiled a substantial dataset (around 6h of scans) on visual imagery along with a proposed data collection protocol. We then train a modified version of an fMRI-to-image model and demonstrate the feasibility of reconstructing images from two modes of imagination: from memory and from pure imagination. The resulting pipeline we call Mind-to-Image marks a step towards creating a technology that allow direct reconstruction of visual imagery.

[1]Obvious Research, Paris, France [2]Neuroradiology Department, GHU Paris Psychiatrie et Neurosciences, Hôpital Sainte Anne, F-75014 Paris, France [3]Université Paris Cité, Institute of Psychiatry and Neuroscience of Paris (IPNP), INSERM U1266, 75014 Paris, France [4]Centre Imagerie du Nord, Saint Denis, France [5]Raidium [6]FrontLab, Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, INSERM, CNRS, AP-HP, Hôpital de la Pitié Salpêtrière, Paris, France [7]Centre de NeuroImagerie de Recherche (CENIR), Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, INSERM, CNRS, AP-HP, Hôpital de la Pitié Salpêtrière, Paris, France [8]Institut des Systèmes Intelligents et de la Robotique, Sorbonne Université. Correspondence to: <research.obvious@gmail.com>.

Accepted at ICML 2024 AI for Science workshop.

1. Introduction

In the rapidly evolving field of neuroscience and artificial intelligence, the quest to decode the complex mechanisms of the human brain has led to groundbreaking developments, particularly in the area of functional magnetic resonance imaging (fMRI). Among the myriad applications of fMRI, one of the most intriguing has been the reconstruction of visual experiences. Traditional methods have successfully reconstructed images based on fMRI BOLD (blood-oxygen-level-dependent) data captured while subjects view specific images (Chen et al., 2022; 2023; Scotti et al., 2023). This process involves preprocessing the BOLD data to isolate voxels associated with the brain's visual areas, generating clip embeddings from these voxels, and then employing image generation algorithms that utilize these embeddings to recreate the observed images. Such methodologies have shown promising results, especially when employing datasets like the NSD (Natural Scenes Dataset) (Allen et al., 2022), a substantial 7T fMRI dataset designed to bridge cognitive neuroscience and artificial intelligence. This dataset, featuring extensive scans from eight subjects, has had tremendous impact for research in this domain.

Progress in cognitive neuroscience and artificial intelligence has not only allowed to advance our understanding of brain mechanisms but also to holds the promise to develop practical applications, such as real-time clinical tools and brain-computer interfaces. However, despite the progress in visual reconstruction from observed images, a challenge remains in the domain of visual imagination. Visual imagination encompasses both the recall of previously seen images (which we refer to as weak imagination in the paper) and the generation of entirely new images within the mind's eye (which we refer to as strong imagination). The ability to decode and reconstruct these imagined visuals offers a wide range of promising possibilities from aiding individuals with disabilities, to enhancing creative processes.

While using models based on reconstruction of images based on visual stimuli to generate images from imagination is theoretically possible, we believe that we need novel approaches to tackle the issue of visualizing imagination in the brain, particularly in data collection and model devel-

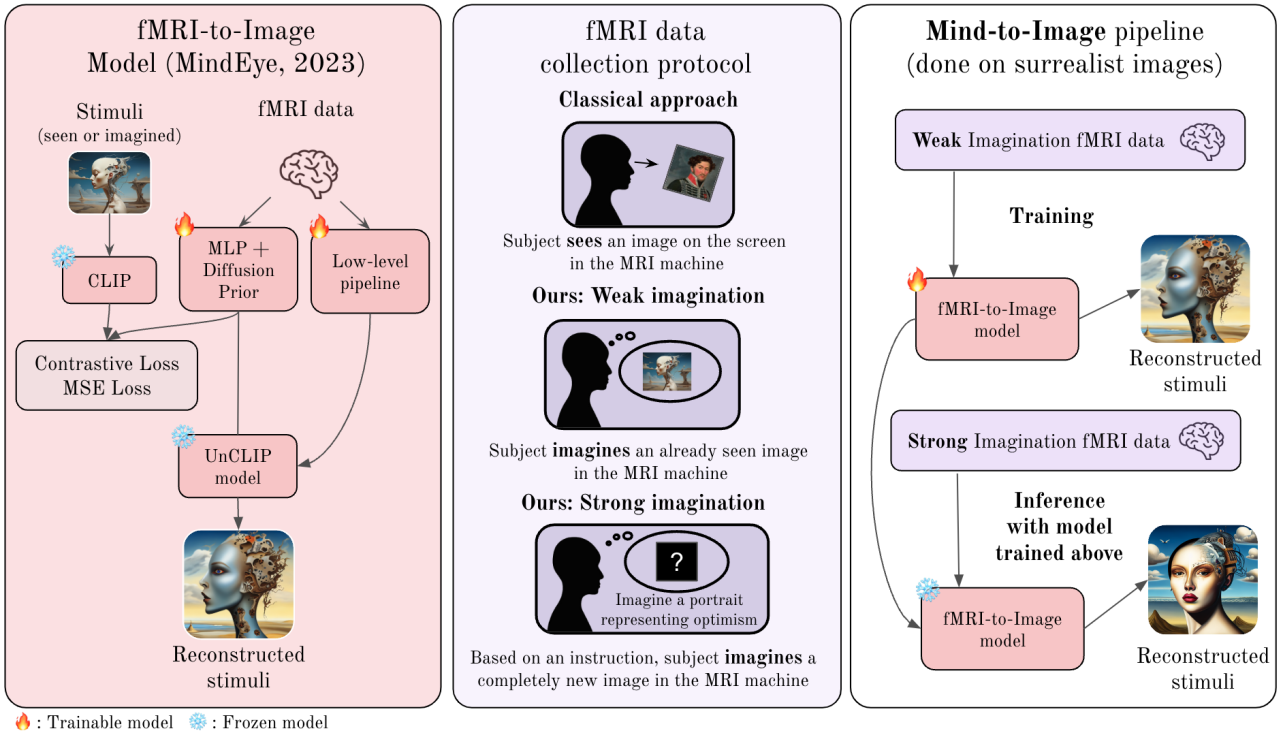

Figure 1. Overview of our Mind-to-Image pipeline which allows to reconstruct images from mental imagery. **Left**: overview of MindEye, the fMRI-to-Image model used in our Mind-to-Image approach. The stimuli (an image seen or in our case imagined) and the associated fMRI data is fed to a high-level pipeline (composed of an MLP and Diffusion Prior) and a low-level pipeline to create aligned CLIP embeddings of the fMRI data. **Middle**: fMRI data collection protocols. In the classic fMRI-to-Image approach, a subject sees images in an MRI scanner, BOLD (Blood-Oxygen Level Dependent) data are recorded then used to train an fMRI-to-Image model that convert brain data to match the seen images. In our approach, we devise two mental imagery protocols: weak and strong imagination. For weak imagination, we let the subject imagine previously seen images from a dataset of surrealist images (face portraits and nature landscapes) which we create. The brain data associated to the recollection of such images is gathered. For strong imagination, we collect brain data where the subject completely imagines new images based on instructions. **Right**: our Mind-to-Image pipeline. We use the weak imagination fMRI data to train an fMRI-to-Image model to reconstruct recollected images from visual imagination brain data. At inference time we use strong imagination fMRI data and this trained model to generate the imagined images, thanks to transfer learning.

opment. To this end, we introduce a protocol inspired by the NSD dataset but tailored to capture the nuances of visual imagination. Our approach distinguishes between weak and strong imagination, developing specific protocols for each type of visual imagery. We restrict our analysis on one theme of images: surrealism, and two distinct modalities of images: portraits and landscapes. This is done by collecting a dataset of surrealist face portraits and nature landscapes, using real and generated images. This choice is not only methodological but also thematic, aligning with the project's broader artistic aspirations. The intersection of art and science, particularly through the lens of surrealism and its exploration of automatic processes, provides a rich context for our research.

Leveraging the MindEye model (Scotti et al., 2023), a state-of-the-art fMRI-to-image framework, we adapt its architecture to adapt to the larger dimensional complexity inherent in visual imagination data. Our approach is summarised in Fig.1. Our results, while preliminary, indicate the model's capability to, at least to some degree, generate portraits and landscapes based on the nature of the imagination. Although the fidelity of the reconstructions to the imagined content varies, which is measured quantitatively through usual metrics in the field, our findings underscore the potential of this approach. With further refinement and expansion of the dataset, this research could pave the way for more precise reconstructions of visual imagination, with the promise of a better understanding of the brain's creative and cognitive processes.

2. Related work

In the domain of mental imagery, several studies have sought to decode and reconstruct images from the brain's visual cortex when subjects engage in imagining specific scenes or objects. (Stokes et al., 2009) provided early evidence

of the brain's capacity to activate shape-specific population codes during mental imagery, highlighting the top-down influence of visual imagination on neural activity in the visual cortex. (Reddy et al., 2010) explored the decoding of category information during mental imagery, providing an understanding on how different types of visual information are processed and represented in the brain.

More recently, (Naselaris et al., 2015) developed a voxel-wise encoding model capable of decoding mental images of remembered scenes, demonstrating the relationship between memory, imagination, and visual perception. (VanRullen & Reddy, 2019) employs a combination of variational auto-encoders (VAE) and Generative Adversarial Networks (GAN) to perform pairwise decoding of faces with a high degree of accuracy but also to classify gender and decode which face was imagined by subjects. (Goebel et al., 2022) showed the feasibility of reading imagined letter shapes from the mind's eye using real-time 7 Tesla fMRI, presenting a novel approach to understanding how specific visual information can be decoded from brain activity.

These studies collectively show the potential of fMRI data in decoding not just seen images but also those that are imagined or remembered.

More close to our work, a recent paper (Koide-Majima et al., 2024) provides preliminary results on reconstructing images from memory. While their results are encouraging, their reconstruction pipeline backbone is not state-of-the-art like MindEye is, which partly explains the results obtained.

3. fMRI data collection for visual imagery

The development of a comprehensive data collection protocol is one the main contributions of our study, aimed at capturing the complex neural correlates of visual imagination. Drawing inspiration from the NSD paper (Allen et al., 2022), our protocol is designed to explore two distinct modes of visual imagination: weak imagination (imagination from memory) and strong imagination (pure imagination). This approach enables us to study how the brain navigates between recalling visual information and generating novel visual constructs.

3.1. Image dataset used: Surrealism

We do not use the images from the COCO dataset (Lin et al., 2014) as it is done in the NSD dataset. Because our research is also associated to an artistic research, we decided to narrow the dataset to one theme (surrealism) and two modes (face portraits and nature landscapes). This choice narrows the visual space of the fMRI-to-Image model, which simplify the task and allows us to have results without having to do 40 hours of scans as in the NSD dataset.

Thus, we created our own dataset of images for our experiments. The images come from three sources: real images and generated images using Versatile Diffusion (Xu et al., 2023) and Midjourney. The dataset comprises 1200 images.

This dataset is used in the weak imagination protocol, where the images are recollected from memory and imagined. Then, the images are used to train the fMRI-to-Image model. Finally, only brain data is given to the fMRI-to-Image model to produce the reconstructions, both in the weak and strong imagination cases.

3.2. fMRI parameters

Data were collected on Siemens 3T PrismaFit using a 64 channels headcoil. BOLD data were acquired using the CMRR multiband EPI sequence. The main parameters are: 2mm isotropic voxels with full brain coverage, TR/TE = 1300ms/27ms, MultiBand=4, partialFourier=7/8, FlipAngle=68°, P>>A phase encoding direction, EchoSpacing at minimum.

3.3. Weak imagination: data collection protocol

The protocol for weak imagination aims at recording brain activity corresponding to the recall of previously seen images. In this phase, subjects are first shown an image for 3 seconds, followed by a 1-second rest period. This sequence is repeated with three different images, each selected from our curated set of 600 surrealist portraits and 600 surrealist landscapes. To engage the brain's imagination, the same images are then flashed for 0.1 seconds, prompting the subject to imagine the image for 5 seconds before resting for another second. This process is repeated over a span of 6 hours, encompassing a total of 1200 images. Out of the data collected, 75 data points (comprising brain activity and the corresponding image) are reserved for validation and evaluation, with the remainder serving as the training set for our fMRI-to-image model. The evaluation focuses on the model's ability to generate images that not only match the correct modality (face portrait or nature landscape) but also reflect the contents of the original image.

3.4. Strong imagination: data collection protocol

For strong imagination, the subject is now asked to generate entirely new visual content based on verbal prompts. The subject is instructed to imagine a portrait or landscape evoking one of ten primal emotions (such as fear, happiness, love, etc.). This task is designed to stimulate the pure imagination process, encouraging the brain to construct novel visual imagery without the reference of previously seen images. Each imagination session lasts for 6 seconds, followed by a 1-second rest, cycling through combinations of landscape/portrait and emotional prompts. This cycle is repeated

10 times. Unlike the data collected for weak imagination, this dataset is solely used during the model's inference stage. The evaluation here is similar to the weak imagination evaluation: quantitively assess if the model matches the correct modality (face portrait or nature landscape) and qualitatively assess if the contents of the generated image matches the visual imagery of the subject. However we cannot compute all the quantitative metrics as the ground truth image does not actually exists outside the subject's mind. To ensure that we recall what the subject has imagined, the subject provides an oral description of the visual contents of the imagined image.

3.5. Pre-processing of raw BOLD data and masking of the regions of interest

After the raw BOLD (blood-oxygen-level-dependent) data is collected, we apply a preprocessing routine to provide the model with data that decorrelates all the noise that impair the following steps. We first co-register all BOLD data with a T1 anatomical scan. Then, using a General Linear Model (GLM) with the Glover hemodynamic response function (HRF) (implemented via the Nilearn library (Abraham et al., 2014)), we extracted beta values associated with each voxel for every distinct event, including both the displayed images and the periods of imagination. This allows us to isolate the specific brain activity corresponding to each event, which is common practice for such tasks.

Mask selection. In the subsequent data pre-processing phase, we identify the brain regions most important to visual imagination and visual perception, using a GLM model to manually extract these areas of interest. Unlike the NSD dataset, which predominantly focuses on visual areas, our study required the creation of a custom mask to englobe the regions identified as crucial in our context of weak and strong imagination.

It is a well known fact that the generation of the perceptive or imagined visual experience (Dijkstra et al., 2019) involves the ventral stream cortical areas (in occipital and temporal lobe). In these visual regions, various aspects of visual information contribute to creating a mental image. This includes features like shape, color, motion, depth, and texture. These features are processed and integrated across low-level (V1,V2), associative (V3, V4, V5) and high-level (including fusiform gyrus in temporal lobe) cortical areas to form a coherent mental representation of what we see or what we think of.

This is the reason why we designed the brain mask to collect the cortical responses during the imagination conditions using the mean responses of the general linear model applied to the event vs. rest condition, in the occipital and temporal lobe. This method aimed to consider only the cortical activity involved in creating a mental image in one subject during the imagination phase.

We thus create and compare three masks manually curated based on the GLM analysis results, ensuring that only voxels corresponding to our defined regions of interest were included in further analyses. The delineation and analysis of these regions, and how they diverge from the predominantly visual areas highlighted in the NSD study, is studied in the results section.

Finally, we create our training, validation (weak imagination) and test (strong imagination) datasets by flattening the pre-processed betas corresponding the selected voxels and pair them with the images or prompts related to their events.

4. Mind-to-Image pipeline

We know describe the fMRI-to-Image model used for our experiments and then outline our Mind-to-Image pipeline by connecting the dots between the data collection protocol and the fMRI-to-Image model.

4.1. fMRI-to-Image model

We start from the state-of-the-art MindEye model which reconstructs images viewed by subjects from fMRI data by combining two pipelines: a high-level (semantic) pipeline and a low-level (perceptual) pipeline. The high-level pipeline maps fMRI voxels to the CLIP Vision Transformer L/14 image space, capturing the semantics of the image. The low-level pipeline targets the perceptual fidelity of reconstructed images by mapping voxels to Stable Diffusion's Variational Autoencoder (VAE) embedding space. They leverage contrastive and mean squared error (MSE) losses for training of its projection and reconstruction capabilities. At inference time, fMRI data associated to an image is fed through the model which outputs a low-level reconstructed image and aligned CLIP embeddings. These embeddings are fed to an UnCLIP models, which is a model able to reconstruct an image based on its CLIP embeddings. In this case, the embeddings as well as the low-level reconstructed image is fed to a Versatile Diffusion Image Variations pipeline to obtain the final reconstruction, as it is done in the original MindEye paper. The whole fMRI-to-Image pipeline is illustrated in Fig.1.

Adapting fMRI-to-Image to larger masks for visual imagery. To accommodate the increased dimensionality inherent in our fMRI BOLD voxel data, which covers extensive brain areas notably larger than only the visual areas of the brain, we adapted MindEye's architecture. This adjustment aimed to manage the growth in input dimensions without exponentially increasing the model's parameter count. We thus simply refine the architecture of the MLP so that the model efficiently processes the data while maintaining a reasonable cost in GPU memory.

4.2. Mind-to-Image pipeline

Our Mind-to-Image experimental procedures are as outlined below:

- We first collect a dataset of weak and strong imagination using the protocol defined in Sec.3, with a total of 6 hours of scans.
- Then, we train our adapted fMRI-to-Image model using the dataset obtained from the weak imagination data collection protocol. Then, this model is evaluated using the validation set to assess its performance.
- We finally freeze the parameters of the fMRI-to-Image model. This frozen model is then employed to perform inference on the dataset derived from the strong imagination data collection protocol, with the goal of visualising the purely imagined images thanks to transfer learning.

5. Results

We first analyze the most activate regions of the brain during visual perception, weak imagination and strong imagination.

We then go over the results obtained for training reconstruction fMRI-to-Image models for weak imagination before looking at strong imagination.

5.1. What happens in the brain during visualisation, weak and strong imagination?

In this section we describe the brain regions involved during visual perception, weak imagination, and strong imagination. This is done though the visualisations presented in Fig.2.

The results of the first test (A) show that the brain regions involved in normal visual perception are, as expected, the primary and secondary visual areas in the occipital lobe, the superior parietal lobule and the bottom of the temporal lobe.

Weak imagination (B) involves occipital primary visual areas in a much weaker way. Instead, there is an activation of the basal temporal associative visual areas (fusiform gyrus) involved in mental imagery. The latter are partially overlapped with brain areas involved in visual perception.

Strong imagination (C) involves even more brain areas distinct from visual cortex, with a strong activation in prefrontal regions, involved in the strategy of generating and controlling mental imagery. There is also an incomplete overlap of activations between visual perception and imagination in the temporal lobe, mainly limited to the fusiform gyrus, especially in the right hemisphere, more specialized in the face processing (as this result was obtained during the portrait imagination test).

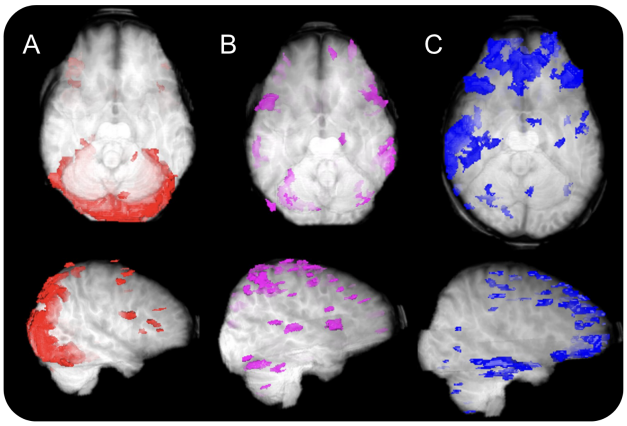

Figure 2. Functional MRI results of the three conditions, respectively: simple visual perception (A- in red), weak imagination (B- in pink) and strong imagination (C- in blue) presented through a lower (top row) and right lateral (bottom row) view. Activation maps were generated by comparing the "test" and "rest" conditions, using the same statistical threshold for each test (T-score =5).

These results justify the need for a specialised dataset for imagination relative to visual perception, given that the regions most active during the different tests are not quite the same.

We thus devise three masks that will be used to select the voxels of the brain that are given as input to the fMRI-to-Image model. **Mask A** (based on the activations during visual stimuli), **Mask B** (based on the activations during weak imagination), and **Mask A∪B** (union of A and B, englobing areas associated to visual perception and weak imagination).

5.2. Can we reconstruct images from weak imagination / memory?

We present curated results of the images generated by our model from weak imagination in Fig.3.

In the exploration of weak imagination, where the subject is asked to recall previously seen images, our results show the capacity of the model to identify to the original class of the images. Specifically, if a subject recalled a face portrait, the reconstructed image is almost always a face portrait, and similarly for nature landscapes. This observation suggests a the ability of the model to grasp the category of the imagined content accurately within the BOLD data.

First, as a quick assessment of the performance, we fine-tune a pretrained ResNet50 model (He et al., 2016) to classify images as either face portraits or nature landscapes using the images in our dataset. We apply this classifier to the

Stimuli & Dataset	Low-Level				High-Level			
Model	PixCorr ↑	SSIM ↑	Alex(2) ↑	Alex(5) ↑	Incep ↑	CLIP ↑	Eff ↓	SwAV ↓
Vision results on NSD dataset								
MindEye (Scotti et al., 2023)	.309	.323	94.7%	97.8%	93.8%	94.1%	.645	.367
Weak imagination on Surrealism dataset								
Mind-to-Image (Mask A)	**.177**	.051	57.8%	53.0%	58.4%	57.9%	**.718**	.079
Mind-to-Image (Mask B)	.082	.023	54.8%	51.6%	55.1%	54.9%	.797	.098
Mind-to-Image (Mask A∪B)	.165	**.052**	**65.1%**	**66.4%**	**61.4%**	**68.5%**	.764	**.065**

Table 1. Quantitative comparison of Mind-to-Image reconstruction performance against other models trained on different datasets, computed on usual metrics in the field. The results are not directly comparable, but are given to provide an element of comparison on the ranges of the values of the metrics. PixCorr is the pixel-level correlation of reconstructed and groundtruth images. SSIM (Wang et al., 2004) is the structural similarity index metric. AlexNet(2) and AlexNet(5) are the 2-way comparisons of the second and fifth layers of AlexNet (Krizhevsky et al., 2012), respectively. Inception is the 2-way comparison of the last pooling layer of InceptionV3 (Szegedy et al., 2016). CLIP is the 2-way comparison of the output layer of the CLIP-Vision model (Radford et al., 2021). EffNet-B and SwAV are distance metrics gathered from EfficientNet-B (Tan & Le, 2019) and SwAV-ResNet (Caron et al., 2020) models, respectively.

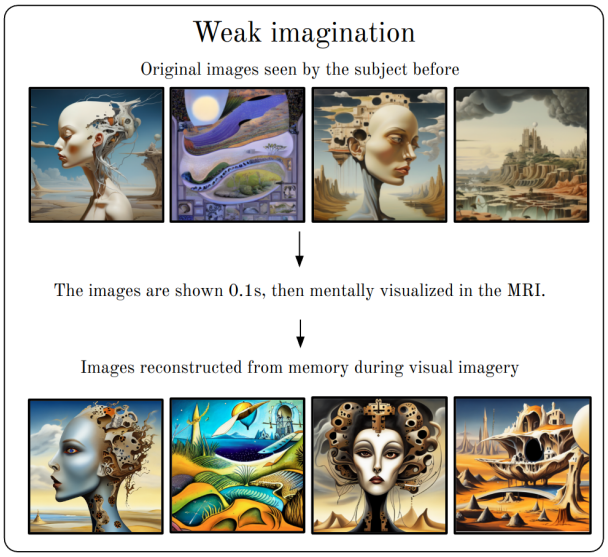

Figure 3. Results for weak imagination on the validation dataset. Top row: images from the validation set of our surrealism image dataset, which were recollected from memory by the subject in the fMRI scanner. Bottom row: reconstructed images from the fMRI-to-Image model based on the brain data associated to the recollection of the images in the top row.

reconstructed images from the weak imagination tasks of the validation set. We obtain, on our best performing model trained with **Mask A∪B**, a classification accuracy of 91% in correctly identifying the reconstructed images as belonging to their original categories of either portraits or landscapes. This level of accuracy demonstrates the model's ability in capturing the category information within the neural activity during weak imagination.

Then, we provide in Tab.1 the usual metrics to quantitatively assess the performance of reconstruction from weak imagination. We compare our results with MindEye (Scotti et al., 2023) Note that the comparisons are not apples-to-apples: (Scotti et al., 2023) provide results for visual stimuli experiments (not for weak imagination) on a large-scale (approx. 10x size) dataset that uses another image domain (realistic images), explaining the big gap in performance. Our results, while inferior to those of reconstructions from visual stimuli, show that we are able to have significantly better than chance results on reconstruction from weak imagination. The best results are obtained when using **Mask A∪B**, showing the usefulness of considering both the visual areas and weak imagination areas of the brain. Specifically, our results indicate that the importance of visual area is superior relative to the imagination areas, as the model trained with **Mask A** outperforms the one trained with **Mask B**. We noted two main outliers on our quantitative results: we get poor results on the SSIM metric, and excellent results on the SwAV metric. We posit that the difference comes from the nature and diversity of images we used (narrow surrealist images compared to more various natural images), but further investigation remain needed to better understand this surprising result.

Qualitatively, while the model demonstrated a robust ability to regenerate the broad category of the original images, the fidelity of content reconstruction within these images presented a more varied outcome. Some reconstructions were visually semantically close to the remembered images. However, others diverged more significantly, indicating a variability in the model's ability to precisely capture and reconstruct the nuanced details of each imagined scene. We provide uncurated reconstructions in Fig.5 in Appendix. This variance in reconstruction fidelity can be attributed, in part, to the limitations imposed by the dataset's size and the duration of the fMRI scanning sessions. With only 6

hours of scans and 1200 images constituting the dataset, the goal for training a model from scratch to achieve perfect reconstruction is out of reach. Nevertheless, these initial results are promising, suggesting a capacity for accurate content reproduction. It is reasonable to anticipate that with an expanded dataset, the fidelity of the reconstructed images to their original counterparts would see important improvement, enhancing both the precision and consistency of the content reproduction in future iterations.

5.3. Can we reconstruct images from strong imagination / pure imagination?

We present curated results of the images generated by our model from strong imagination in Fig.4.

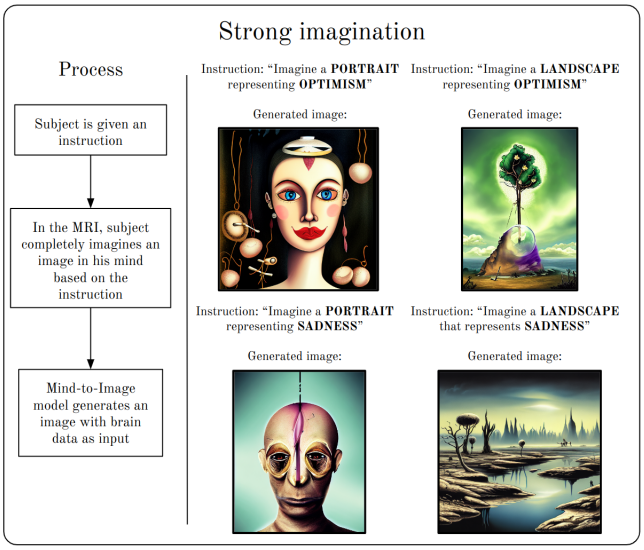

Figure 4. Results from strong imagination. The subject is given an instruction on the fMRI screen, such as "Imagine a portrait representing optimism". Then, the subject purely imagines such image, giving an oral description. Then, the associated brain data is fed to the fMRI-to-Image model to produce a reconstruction, which is then evaluated compared to the oral description.

In the phase dedicated to strong imagination, we employed the previously trained and then frozen fMRI-to-Image model to process BOLD data obtained while subjects engaged in pure imagination tasks, such as imagining a portrait or landscape related to a broad concept like optimism. We thus rely on transfer learning, leveraging a model trained exclusively on weak imagination data to interpret and generate visual outputs from the distinct context of strong imagination, without fine-tuning. Indeed, we do not have tangible images to serve as direct training data for strong imagination, since these images exist solely within the mind of the subject. Hence we employ transfer learning from the weak imagination process to generate from strong imagination. Remarkably, the application of the fine-tuned ResNet50 classifier to the generated images resulted in an 88% accuracy rate in categorizing the images according to the instructions provided to the subjects. This accuracy in category generation demonstrates the effectiveness of transfer learning from weak to strong imagination, which represent our main contribution.

Determining the content accuracy of images generated from strong imagination poses a unique challenge, as the 'ground truth' images solely exist in the subject's mental imagery, inaccessible for direct comparison. To navigate this, we based our evaluation on subjects' descriptions of their imagined scenes, collected during or following the scanning sessions. The variability in the reconstructed content ranges from highly accurate renditions of specific imagined elements, like a tree, dark-haired woman, or flowers, to more abstract interpretations or plain errors. This also reflects the subjective nature of this task. Given the reliance on subjective descriptions and the inherent open-endedness of strong imagination, the results should be interpreted with caution. While it's challenging to draw definitive conclusions about the model's capacity for precise content reconstruction in this context, the instances of recognisable elements emerging in the reconstructions indicate that what could be achievable with increased data volume and further model refinement.

6. Discussion

6.1. Ethical considerations

As we explore further into the capabilities of decoding and reconstructing visual content from brain activity, we get closer to the boundaries of personal thought and imagination, a concept known as "mind privacy". The ability to visualise an individual's mental imagery raises questions about consent, mental autonomy, and the potential for misuse of such technologies. On the one hand, such technology provides potential for good, opening avenues for novel communication methods for those unable to speak or write, enhancing creative expression, and deepening our understanding of the human brain and its processes. On the other hand, it introduces the possibility of invasive breaches of personal thought, where the most private mental images and ideas could be exposed without consent. This requires a careful, deliberate approach to ensure that the development and application of these technologies are governed by ethical principles that prioritise individual privacy and autonomy. Currently, the barriers to entry, including the need for explicit consent and the use of expensive and sophisticated MRI machinery, provide a layer of protection.

6.2. Future work

Future directions for this line of research include the acquisition of more extensive data to improve the accuracy and specificity of content match in the reconstructed images. A refined experimental protocol for both weak and strong imagination tasks could lead to insights into the brain's imaging processes, while improved evaluation methods for strong imagination would enhance our understanding of the model's effectiveness in capturing the depth and detail of imagined content.

Additionally, the question of whether MRI technology can be replaced or complemented by EEG is promising. Recent studies suggest the feasibility of reconstructing visual stimuli from brain activity using EEG (Benchetrit et al., 2023), which offers a more accessible, and cost-effective alternative to MRI.

Finally, in our experiments, the subject clearly expressed that the imagined images were getting more clearer and more precise as he repeatedly attempted to imagine the same image. This opens a potentially promising avenue for "training" one's own mental imagery, which could be applied to subjects presenting various degrees of aphantasia (Zeman et al., 2015), i.e. the degree of their inability to create mental imagery. A set of experiments directed toward this question is thus a potential future work.

7. Conclusion

In our work, we introduced and applied two novel data collection protocols: weak and strong imagination. These protocols were designed to study visual imagery and create training datasets for fMRI-to-Images models. We implemented them during a 6-hour scanning session to create a comprehensive dataset. This dataset then served as the basis for training our fMRI-to-image model. Initially, the model was trained using data from the weak imagination protocol, which involved reconstructing images that the subject had previously seen. Subsequently, the same model, now frozen, was applied to data from the strong imagination protocol, where the subject was asked to imagine new images based on a written instruction. Our study demonstrates that the model successfully distinguishes between broad categories such as portraits and landscapes, effectively grasping the category of the content imagined by the subject. Accurately capturing the detailed contents of these imagined images was partly successful, but proved to be more challenging. This research provides a promising path towards the generation of visual representations directly from human thought.

Acknowledgements

The author acknowledges the support of the French Agence Nationale de la Recherche (ANR) under reference ANR-23-CE23-0023.

References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., and Varoquaux, G. Machine learning for neuroimaging with scikit-learn. Frontiers in Neuroinformatics, 8:14, 2014. doi: 10.3389/fninf.2014.00014.

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. A massive 7t fmri dataset to bridge cognitive neuroscience and artificial intelligence. Nature Neuroscience, 25:116–126, 2022. doi: https://doi.org/10.1038/s41593-021-00997-w.

Benchetrit, Y., Banville, H., and King, J.-R. Brain decoding: toward real-time reconstruction of visual perception. arXiv preprint arXiv:2310.19812, 2023.

Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., and Joulin, A. Unsupervised learning of visual features by contrasting cluster assignments. Advances in neural information processing systems, 33:9912–9924, 2020.

Chen, Z., Qing, J., Xiang, T., Yue, W. L., and Zhou, J. H. Seeing beyond the brain: Conditional diffusion model with sparse masked modeling for vision decoding. arXiv preprint arXiv:2211.06956, 2022. URL https://doi.org/10.48550/arXiv.2211.06956.

Chen, Z., Qing, J., and Zhou, J. H. Cinematic mindscapes: High-quality video reconstruction from brain activity. arXiv preprint arXiv:2305.11675, 2023. URL https://doi.org/10.48550/arXiv.2305.11675.

Dijkstra, N., Bosch, S. E., and van Gerven, M. A. Shared neural mechanisms of visual perception and imagery. Trends in cognitive sciences, 23(5):423–434, 2019.

Goebel, R., van Hoof, R., Bhat, S., Lührs, M., and Senden, M. Reading imagined letter shapes from the mind's eye using real-time 7 tesla fmri. In 2022 10th International Winter Conference on Brain-Computer Interface (BCI), pp. 1–3, 2022. doi: 10.1109/BCI53720.2022.9735031.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778, 2016.

Koide-Majima, N., Nishimoto, S., and Majima, K. Mental image reconstruction from human brain activity: Neural decoding of mental imagery via deep neural network-based bayesian estimation. Neural Networks, 170:349–363, 2024.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25, 2012.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. Microsoft coco: Common objects in context. In Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13, pp. 740–755. Springer, 2014.

Naselaris, T., Olman, C. A., Stansbury, D. E., Ugurbil, K., and Gallant, J. L. A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. NeuroImage, 105: 215–228, 2015. doi: 10.1016/j.neuroimage.2014.10.018. URL https://www.sciencedirect.com/science/article/pii/S1053811914008428.

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

Reddy, L., Tsuchiya, N., and Serre, T. Reading the mind's eye: Decoding category information during mental imagery. NeuroImage, 50(2): 818–825, 2010. doi: 10.1016/j.neuroimage.2009.11.084. URL https://www.sciencedirect.com/science/article/pii/S1053811909012701.

Scotti, P. S., Banerjee, A., Goode, J., Shabalin, S., Nguyen, A., Cohen, E., Dempster, A. J., Verlinde, N., Yundler, E., Weisberg, D., Norman, K. A., and Abraham, T. M. Reconstructing the mind's eye: fmri-to-image with contrastive learning and diffusion priors. arXiv preprint arXiv:2305.18274, 2023. URL https://doi.org/10.48550/arXiv.2305.18274.

Stokes, M., Thompson, R., Cusack, R., and Duncan, J. Top-down activation of shape-specific population codes in visual cortex during mental imagery. Journal of Neuroscience, 29(5):1565–1572, 2009. doi: 10.1523/JNEUROSCI.4657-08.2009. URL https://www.jneurosci.org/content/29/5/1565.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818–2826, 2016.

Tan, M. and Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning, pp. 6105–6114. PMLR, 2019.

VanRullen, R. and Reddy, L. Reconstructing faces from fmri patterns using deep generative neural networks. Communications Biology, 2, 2019. doi: https://doi.org/10.1038/s42003-019-0438-y. URL https://www.nature.com/articles/s42003-019-0438-y.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing, 13(4):600–612, 2004.

Xu, X., Wang, Z., Zhang, E., Wang, K., and Shi, H. Versatile diffusion: Text, images and variations all in one diffusion model. arXiv preprint arXiv:2211.08332, 2023.

Zeman, A., Dewar, M., and Della Sala, S. Lives without imagery–congenital aphantasia. Cortex, 73:378–380, 2015.

A. Uncurated samples for weak imagination

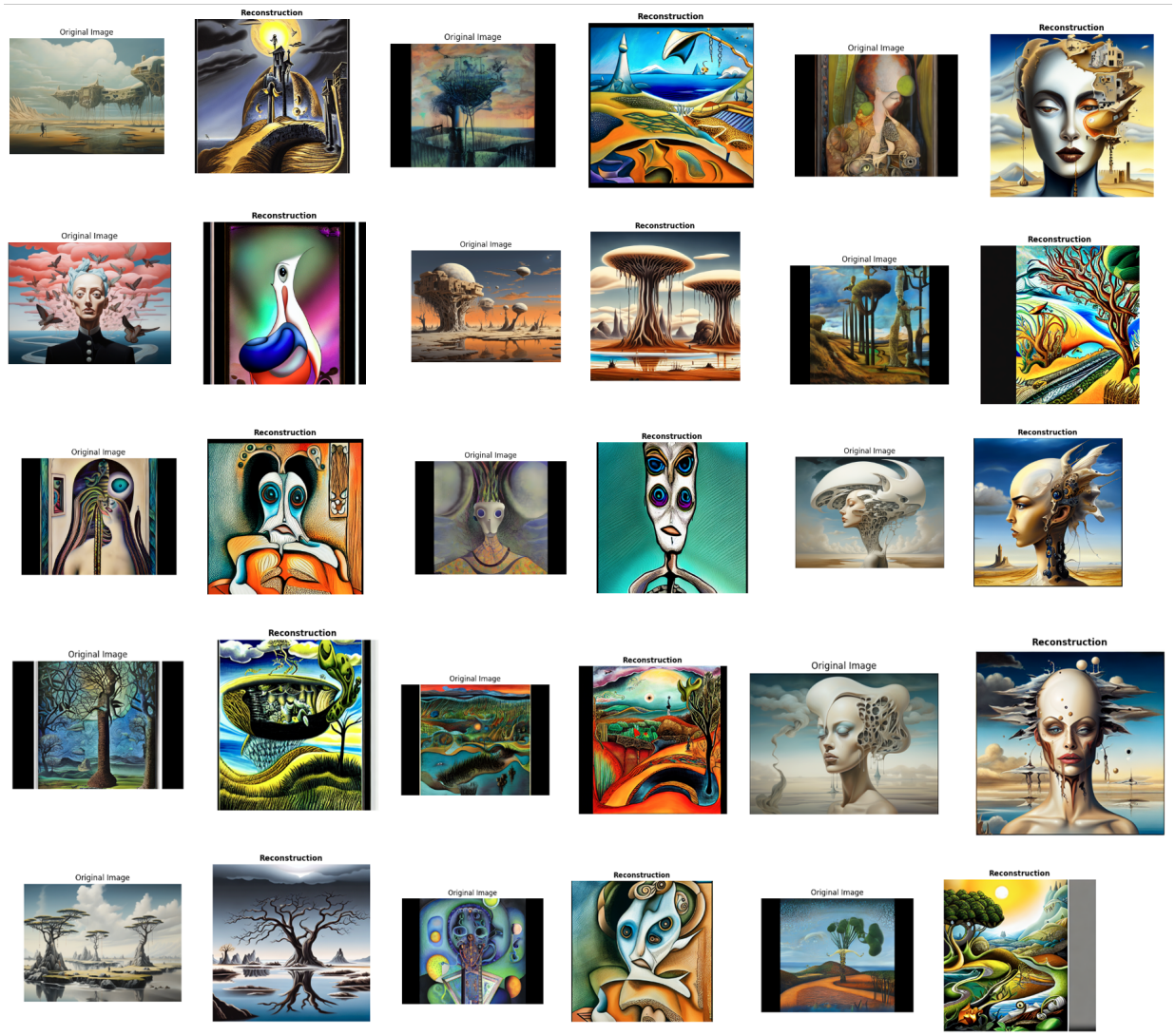

Figure 5. Uncurated results for weak imagination on the validation dataset. In each example, we present an image from the validation set of our surrealism image dataset, which is recollected from memory by the subject in the fMRI scanner, and the reconstructed image from the fMRI-to-Image model based on the brain data associated to the recollection of the original image.

In Fig.5, we present uncurated results from the validation dataset.