Investigating Effects of Participant Variation on Performance of Visual Stimuli Reconstruction From fMRI Signals Using Machine Learning

Quan Zheng[1]

Supervisor: Xucong Zhang[1]

[1]EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 25, 2023

Name of the student: Quan Zheng
Contact: q.zheng-7@student.tudelft.nl
Final project course: CSE3000 Research Project
Thesis committee: Xucong Zhang, Nergis Tömen

Abstract

Image reconstruction from neural activation data is a field that has been growing in popularity with developments such as neuralink in the brain-machine interface space. To make better decisions when collecting data for this purpose, it is important to know what qualities to optimize for. The present paper investigates the relation between participant variation and visual stimulus reconstruction performance from functional magnetic resonance imaging (fMRI) data, which can guide decisions on whether resources should be spent collecting more data from fewer individuals or vice versa. We conducted performance evaluation on the Self-Supervised Image Reconstruction machine learning architecture proposed by Gaziv et al. using three pixel-wise and two structural image similarity measures. Our results show that reconstructions from one subject's fMRI data consistently performed best across all five performance metrics. However, statistically significant variance in reconstruction performance across subjects was found for only the feature-based similarity index. While the present paper found statistically significant results, we recommend future research to further investigate this notion by employing similar evaluation on other models.

1 Introduction

The field concerning reconstruction of visual stimuli from neural activity has made great leaps in recent years with the advent of highly performant machine learning (ML) algorithms and improvements in functional magnetic resonance imaging (fMRI) data acquisition techniques. The goal of this discipline is to reconstruct images shown to a human subject as accurately as possible based on blood oxygen levels across their brain as measured using fMRI scans, which serve as a proxy for the subject's neural firing patterns that result from viewing the visual stimulus [1].

Data sets with more samples at higher resolutions and novel machine learning architectures are continuing to bring improved results in both semantic and visual reconstruction [2] [3]. However, certain contingencies between the data used, and the performance of resulting reconstruction models are yet to be investigated. One of said contingencies is that of a given participant's data quality on reconstruction performance. In our context, data quality can depend on multiple factors, including noise occurring from the scanning device [4], the participant's head motion, or even their heart beat [5].

The present paper investigates the question: "What is the predictive ability of participant selection on image reconstruction from fMRI signals using machine learning?". In order to answer this, we ask the following sub-questions:

1. Does any one participant's fMRI data consistently result in better reconstruction performance?
2. Is there a significant difference in reconstruction performance between participants?
3. What pros and cons do differing image similarity metrics bring to the table?

It has been shown that no two individuals have the same brain anatomy [6]. Our research can indicate the generalizability of machine learning models to the neural firing patterns of differing individuals. It is thus relevant for future research regarding machine learning models operating on brain imaging data, as well as brain-machine interfaces such as Neuralink.

2 Methodology

The following section goes into detail on the methodology used in the present paper. We explain the reconstruction method used, as well as how data was acquired. In addition, we elaborate on the methods employed for reconstruction performance evaluation, and finally discuss how statistical analysis was conducted.

2.1 Reconstruction Method

For the investigative purposes of this paper, the Self-Supervised Image Reconstruction machine learning architecture proposed by Gaziv et al. was used. This approach makes use of a neural-network based encoder decoder architecture, where an encoder E is trained to predict the fMRI responses to a visual stimulus, with a decoder D being trained to reconstruct the original stimulus from said responses [7]. An example of reconstructions achieved using this method can be seen in Figure 1.

Figure 1: Visual reconstructions achieved using Gaziv et al.'s self-supervised method [7]

The amount of publicly available "paired images" (images with corresponding fMRI responses) is in the thousands. This represents a quite limited amount of data when compared to the hundreds of millions that computer vision or image generation algorithms are trained on [8] [9]. The self-supervised approach proposed by Gaziv et al. attempts to mitigate the issue of data scarcity using the encoder decoder model, allowing for derivation of synthetic fMRI signals from "unpaired images" (images with no corresponing fMRI response) using the encoder. From this, the decoder attempts to reconstruct the original image. This enables both artificial expansion of the data set, and the use of an unsupervised approach to training of the reconstruction model.

The paper, website[1], and repository[2] for the self-supervised reconstruction model showcase only a subset of reconstructions that have been achieved with the data set. In order to gain access to a sample size that is as large as possible, the training process for the model was re-run on a local machine. This yielded 250 image reconstructions in total, being 50 image reconstructions for each of the five subjects in the Generic Object Decoding (GOD) data set.

2.2 Data Acquisition

The self-supervised reconstruction approach by Gaziv is trained using the Generic Object Decoding data set, consisting of 1250 paired images from ImageNet [10] with their corresponding fMRI responses across five subjects [11]. In addition, unsupervised training is conducted using 49 thousand images, also taken from courtesy of ImageNet. 1200 paired images of the GOD data set were used to train the model, while the remaining 50 were used for reconstruction of images and performance evaluation of the model after training.

2.3 Performance Evaluation

The accuracy of image reconstructions was evaluated using the ”image similarity measures” python library [12]. More specifically, evaluations were made using the Root Mean Square Error (RMSE), Peak Signal to Noise Ratio (PSNR), Signal to Reconstruction Error ratio (SRE), Structural Similarity Index Measure (SSIM), as well as the Feature-based Similarity Index (FSIM). These are explained in further detail in the Evaluation Metrics section.

2.4 Statistical Analysis

Performance evaluations analyzed by aggregating results using mean, standard deviation, maximum and minimum values, as well as the amount of best performances evaluation by subject out of the 50 images. For further details and interpretation we ask the reader to refer to the Results and Discussion section respectively.

To answer our first research question ”is there a significant difference in reconstruction performance between participants?”, we made use of Analysis Of Variance (ANOVA). ANOVA is a method in statistical analysis used to gauge whether the means of subsets in a sample demonstrate a significant variance between said subsets [13]. In the case of the present paper, a full sample can be the reconstruction accuracies as measured by RMSE across all participants, while a subset of this sample is given by the RMSE accuracies for only the first subject.

3 Evaluation Metrics

In the following section, we explain the five image similarity metrics used for performance evaluation and how our chosen model performed on each. We will first discuss the root mean squared error and the peak signal to noise ratio. Then, we elaborate on the signal reconstruction error ratio, as well as the feature based similarity index. Finally, we take a closer look at the structural similarity index.

[1]https://www.wisdom.weizmann.ac.il/ vision/SSReconstnClass/
[2]https://github.com/WeizmannVision/SelfSuperReconst

3.1 Root Mean Squared Error

The Root Mean Squared Error (RMSE) is a metric that is widely used in machine learning to judge error of the model's predicted value to the ground truth during training. It is calculated by first taking the Mean Squared Error (MSE), given by

$$MSE = \frac{\sum_{M,N}[I_1(m,n) - I_2(m,n)]^2}{M * N}$$

for a picture of M x N pixels, where $I_1(m,n)$ and $I_2(m,n)$, are the respective ground truth and reconstruction RGB values for the pixel at row m and column n

The RMSE is given by the root of the MSE. Thus we have

$$RMSE = \sqrt{MSE}$$

with higher values indicating worse reconstruction quality.

Due to the fact that error values are squared in the MSE, the RMSE is always non-negative, and greater error values are penalized more. Taking the root of the summed value yields error in units of the response variable (as opposed to squared units), for more intuitive interpretation of results.

While the RMSE is an objective measure that is fast and intuitive to calculate - it shows disadvantages when it is used in the context of image similarity. The vector proximity of color values does show how similar two pictures are based on individual pixels. However, the measure does not capture higher similarity in the sense of higher level shapes and structures depicted in the image.

3.2 Peak Signal to Noise Ratio

The peak signal to noise ratio (PSNR) indicates the ratio between the peak signal of the ground truth image and the corrupting noise in the reconstructed image. It is commonly used to assess effects of compression on image quality. The PSNR is expressed in decibels (dB), with higher values indicating higher reconstruction quality.

Calculating the PSNR is done by the following formula:

$$PSN = 10 * log_{10}(\frac{R^2}{MSE})$$

where R is the highest possible value that a pixel can take [14]. For a single color channel in 8-bit space, this is 255.

Similarly to the RMSE, the PSNR is a pixel-wise comparison technique that provides a quantitative measure of accuracy of colors in the reconstructed image, while not taking into account the structural components of the overall image.

3.3 Signal to Reconstruction Error ratio

The Signal-to-Reconstruction Error ratio (SRE) is another pixel-wise comparison metric. It is compared to the PSNR in the literature, with a the key difference being that the SRE measures reconstruction error in relation to the mean intensity of the original image, which is variable, rather than its peak possible intensity, which is constant. This makes the SRE preferrable when comparing errors between pictures of varying brightness [15].

The SRE is calculated by:

$$SRE = 10 * log_{10} \frac{\mu_x^2}{(\hat{x} - x)^2/n}$$

with the average intensity of the picture x being denoted as μ_x, the reconstruction as $\hat{x}$, and n being the number of pixels in x. As with the PSNR, SRE values are given dB, with higher values indicating better reconstruction performance.

3.4 Structural Similarity Index Measure

The Structural Similarity Index Measure (SSIM) assesses the similarity between two images by their luminance (intensity), contrast, and structural patterns. Its values can range between -1 and 1, with higher values indicating higher similarity, and 1 expressing an exact match.

We calculate the SSIM for two pictures x and y by:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + x_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

where μ_x and μ_y are the average values of x and y respectively, σ_x σ_y are their variances, and σ_{xy} is the covariance between x and y. c_1 and c_2 are given by $c1 = (k_1 * L)^2$ and $c1 = (k_1 * L)^2$, with L being the highest possible value for a pixel (255 for 8-bit color space), and $k_1 = 0.01$, $k_2 = 0.03$ as default values [16].

The SSIM is commonly employed in image processing, and image quality assessment to evaluate compression, denoising, and enhancement algorithm performance. It provides a comprehensive measure of image similarity that goes beyond simple pixel-wise comparison by taking into account the interdependence of pixels that are in close spacial proximity [17]. Despite this, the SSIM has been shown to yield non-intuitive, and undefined results at times [18].

3.5 Feature-based Similarity Index

The Feature-based Similarity Index (FSIM) measures similarity between two images based on their phase congruency and gradient magnitude. Here, phase congruency gives measure of the significance of local structures, while the gradient magnitude factors visual contrast into the evaluation [19].

Computation of the FSIM is done in a two-stage process, first evaluating a local similatity map, then aggregating said map into a single score. The mathematical details of the FSIM are long winded and thus out of scope for the present paper. For an exact explanation we ask readers to refer to the original paper.

FSIM is meant as an alternative to the SSIM and is generally applied in the same use cases. Values of the measure can range from 0 to 1 with higher scores indicating greater similarity between pictures, and 1 indicating a perfect match.

4 Contribution

The present work contributes to the field of reconstruction of visual stimuli from neural imaging data using machine learning by investigating the predictive ability of participant selection on reconstruction performance. Results of this study can give researchers guidance deciding on making trade-offs between inviting more participants while taking fewer fMRI scans to increase chances of finding well suited participants, or vice versa to gain more participant specific data to train models on. Reconstruction performance was assessed using five objective image similarity measures (RMSE, PSNR, SSIM, FSIM, SRE), in an effort to give a more holistic view on the topic. The performance metrics give a quantitative measure on the fidelity of reconstructions, providing a basis to compare further reconstruction techniques.

5 Experimental Setup

Model training and image reconstruction was conducted on a Windows 11 PC running Windows Subsystem for Linux, equipped with an RTX 3080 GPU, Ryzen 7 3700X CPU, and 64GB of DDR4 RAM. The code for training can be found in the the Self-Supervised RGBD Reconstruction From Brain Activity repository. [3]

Reconstruction performance evaluation was conducted on a base model 14" MacBook Pro 2021 (M1 Pro, 16GB RAM) running macOS Ventura 13.4. Code for this purpose was run in Python 3.9.17 using the split image library to separate reconstruction images from their original, the "image similarity measures" library to evaluate reconstruction performance, pandas for data manipulation, as well as matplotlib for visualization purposes.

6 Results

The following section shows the results from statistical aggregation of performance evaluations of the reconstructions that were achieved. We present the mean, standard deviation, as well as the minimum and maximum score achieved by subject for each measure rounded to five decimals. We also list the number of times that a given subject's image reconstruction was rated best for each metric (out of the 50 reconstructed images). For further detail on performance evaluation, please refer to the Evaluation Metrics section.

To determine whether there is a significant difference in reconstruction performance when varying between participants, we conducted Analysis Of Variance (ANOVA) with the subject as the independent variable, and reconstruction performance as the dependent variable. We employed a significance level of $p < 0.05$, our null hypothesis being that there is no difference in image reconstruction performance resulting from differing participants, while the alternative hypothesis is that there is indeed such a difference present.

6.1 Root Mean Squared Error

Aggregated statistics for RMSE evaluation are shown in Table 1. Lower values indicate higher reconstruction quality. RMSE scores range from 0.009 to 0.044 between subjects. Mean values by subject range from 0.019 for subject 3 to 0.020 for subject 4, with standard deviation ranging from 0.005 to 0.007. With 18 lowest RMSE scores by picture, subject 3 is significantly ahead of subject 2 with 12 best performances, while the remaining subjects scored best 7 or fewer times.

[3] https://github.com/WeizmannVision/SelfSuperReconst

Table 1: RMSE Aggregations by Subject

sub	mean	std	min	max	# best
sub1	0.0202	0.0059	0.0090	0.0373	7
sub2	0.0197	0.0059	0.0091	0.0353	12
sub3	0.0187	0.0055	0.0075	0.0324	18
sub4	0.0205	0.0069	0.0096	0.0437	7
sub5	0.0199	0.0045	0.0105	0.0304	6

Observing the boxplot for RMSE shown in Figure **??** shows that there are one or more outliers for four out of five subjects towards the high (worse performing) end.

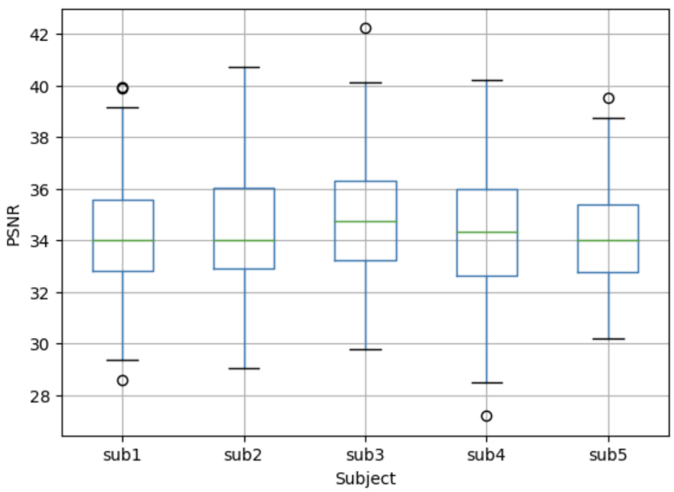

Figure 2: SSIM boxplots grouped by subject

6.2 Peak Signal to Noise Ratio

Evaluation aggregations using the PSNR are listed in Table 2, with higher values indicating higher reconstruction quality. The highest peak signal achieved in our reconstruction is 42.238 dB, with the lowest one being 27.191 dB. Subject 3 achieved the highest mean PSNR at 34.882 dB, while subject 4 has the lowest with 34.165 dB. PSNR variance by subject ranges between 2.053 dB to 2.769 dB.
The PSNR yields the same amount of highest PSNR scores by picture per participant as the RMSE, subject 3 is significantly ahead of subject 2 with 18 and 12 best performances respectively, and remaining subjects scoring best 7 or fewer times.

Table 2: PSNR Aggregations by Subject

sub	mean	std	min	max	# best
sub1	34.1764	2.5043	28.5723	39.9585	7
sub2	34.3978	2.6221	29.0467	40.7438	12
sub3	34.8819	2.6194	29.7855	42.2375	18
sub4	34.1649	2.7694	27.1911	40.2398	7
sub5	34.1855	2.0530	30.1917	39.5482	6

Looking to the boxplot in Figure 3, the reconstruction performance as measured by the PSNR shows statistical outliers for four out of five subjects, although fewer than

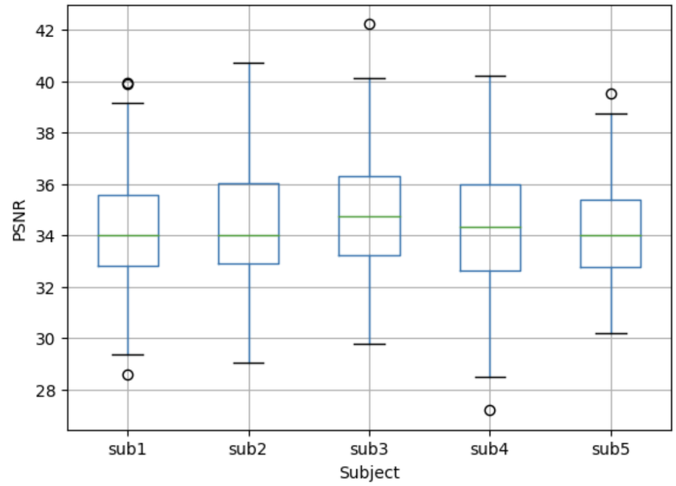

Figure 3: PSNR boxplots grouped by subject

6.3 Signal to Reconstruction Error ratio

Table 3 shows aggregated SRE values from our evaluation. Higher values indicate higher reconstruction quality. In our sample, the highest SRE was 51.905 dB with the lowest at 32.100 dB. Mean values range between 42.056 dB for subject 1 and 42.399 dB for subject 3. The standard deviation ranges between 3.554 dB to 3.885 dB. As with the PSNR, reconstructions from fMRI signals of subject 3 as evaluated by SRE score best by a significant amount of the time, with 18 times, while the remaining subjects performed best 10 or fewer times.

Table 3: SRE Aggregations by Subject

sub	mean	std	min	max	# best
sub1	42.0562	3.8313	32.2476	50.9959	8
sub2	42.1409	3.7663	33.5379	51.9050	10
sub3	42.3985	3.8109	33.4778	50.6304	18
sub4	42.0798	3.8849	32.0999	50.0271	8
sub5	42.0801	3.5543	34.7736	49.6416	6

When shown as a boxplot as seen in Figure 4, we can see that the mean SRE between subjects is relatively close, with some outliers on the low and high end for three out of five subjects.

6.4 Structural Similarity Index Measure

SSIM scores aggregated by subject are listed in Table 4, with larger values indicating better reconstruction performance. Scores for our reconstructions ranged between 0.329 to 0.957, with mean scores by subject ranging from 0.741 to 0.772. Standard deviation by subject was found to be between 0.100 to 0.114. As for the number of best reconstructions, subject 3 is again in the lead with 18 best scores, with remaining subjects performing best 9 or fewer times.

Table 4: SSIM Aggregations by Subject

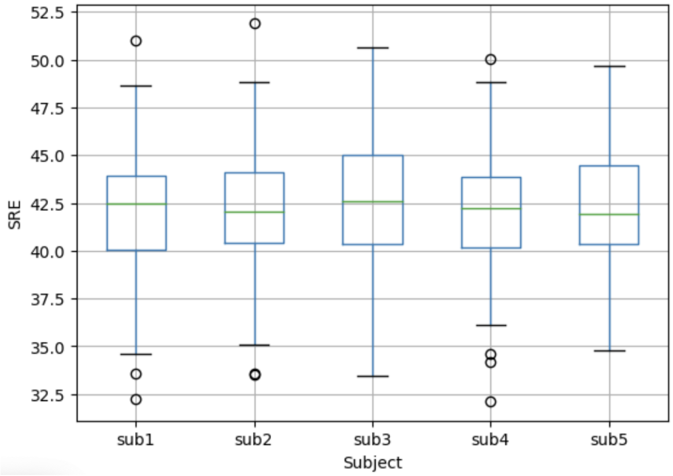

Figure 4: SRE boxplots grouped by subject

sub	mean	std	min	max	# best
sub1	0.7546	0.1110	0.3930	0.9362	9
sub2	0.7553	0.1075	0.3932	0.9555	9
sub3	0.7721	0.1000	0.4117	0.9571	18
sub4	0.7493	0.1149	0.3295	0.9366	7
sub5	0.7413	0.1004	0.4717	0.9309	7

The boxplot in Figure 5 for SSIM scores by subject shows means that are quite close to each other, with outliers mostly on the worse performing end.

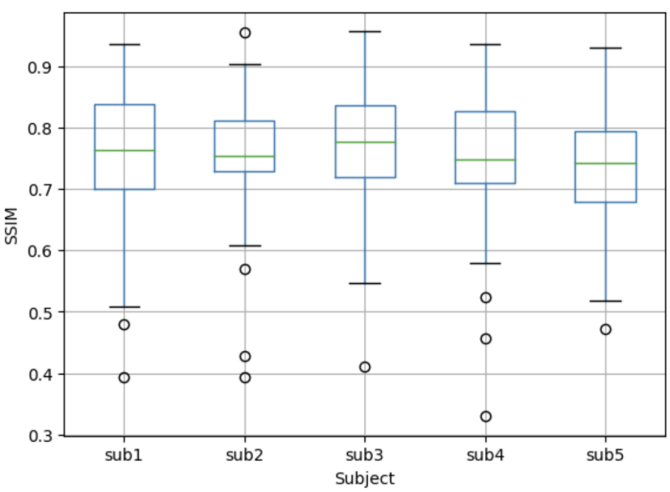

Figure 5: SSIM boxplots grouped by subject

6.5 Feature-based Similarity Index

Aggregated FSIM scores by subject are shown in Table 5. Higher values indicate better reconstruction performance. The lowest and highest scores are 0.273, and 0.471 respectively. Overall reconstruction scores were found to be between 0.273 to 0.471. Mean scores by subject are between 0.345 and 0.363, with standard deviation ranging from 0.024 to 0.034. Subject 3 performed best on 21 out of the 50 pictures, while subject 4 did this 14 times.

Table 5: FSIM Aggregations by Subject

sub	mean	std	min	max	# best
sub1	0.34703	0.02990	0.27298	0.42097	7
sub2	0.34867	0.02716	0.30908	0.44404	4
sub3	0.36255	0.02694	0.28770	0.44948	21
sub4	0.35757	0.03425	0.28244	0.47116	14
sub5	0.34534	0.02392	0.30092	0.42178	4

When visualized as a boxplot as in Figure 6, the FSIM shows comparatively little variance in reconstruction performance by subject for the bulk of the data. However, there are many more outliars when compared to the other performance evaluation metrics employed in the present paper.

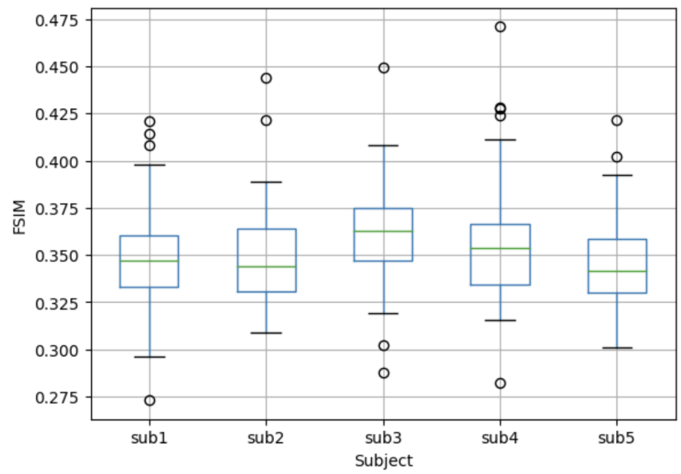

Figure 6: FSIM boxplots grouped by subject

6.6 Analysis Of Variance

ANOVA statistics by performance measure are listed in Table 6. With our selected p-value of $p < 0.05$, ANOVA analysis yields that variation in image reconstruction performance between subjects is non-significant for the RMSE, PSNR, SRE, and SSIM. This is with the exception of the Feature-based Similarity Index, where a p-value of p = 0.010 was found.

Table 6: ANOVA statistics by performance measure

metric	F-value	p-value
RMSE	0.66159	0.61924
PSNR	0.73673	0.56770
SRE	0.07071	0.99084
SSIM	0.55854	0.69298
FSIM	3.38634	0.01017

7 Responsible Research

Our research is conducted using the Generic Object Decoding (GOD) data set [11]. The GOD is a collection of functional magnetic resonance imaging (fMRI) responses to visual stimuli depicting natural scenes across five human subjects. All data in the data set is de-identified in order to protect the participants' identities and further sensitive information. When requesting access to the data set, individuals must pledge to not attempt to retrieve protected health information (PHI) of the study's participants and to inform their principal investigator if this should happen accidentally. This includes information such as names, addresses, phone numbers, physical or mental health condition, or similar.

Despite these pre-cautions in place, there are ethical questions that arise in the context of our research. Technology that allows for the reconstruction of visual stimuli based on an individual's brain activity can potentially be misused by malicious parties to extract information from non-consenting subjects. Current methods require extensive participant pre-screening and data pre-processing, as well as bulky expensive machinery which needs to be run under the supervision of experts. However, future developments may bring about miniaturized devices which allow for acquisition of neural activity data in an open setting, or even from afar. This brings about the question what potential misuses or misunderstandings these development can bring about.

Privacy of thoughts is a major concern in connection with any research in the area of decoding brain signals. While the domain of one's own thoughts are currently regarded as something that is untouchable by other parties, this kind of technology at a mature stage would lead to unprecedented scenarios, many that we cannot imagine at this stage.

The technology can be used for both good and bad. In justice enforcement, it can enable eye witnesses to give account by recalling visual stimuli such as the face of a criminal, or the scene of a crime. However, it can also be misused by malicious parties attempting to gain information from an individual's brain activity patterns without their consent.

Another potential risk of this technology lies in its use before maturity. Misreadings can lead to misinterpretation of an individual's thoughts. In domains such as justice enforcement, this can lead to mis-attribution of fines all the way to conviction of innocent people. It is thus of utmost importance to set precautions in place to avoid the interpretation of brain signals in stark consequence scenarios before the technology has been shown to be highly reliable.

8 Discussion

Looking at our results, it is interesting to note that the RMSE and PSNR evaluations yield the same outcome when it comes to finding the best scoring participant. This can be explained by the fact that by definition, both the RMSE and PSNR are fixed proportional functions of the MSE, which in our context is a pixel wise color space distance measure. Since lower RMSE scores indicate better performance, while the opposite is true for the PSNR they can loosely be interpreted as inverses of each other.

We can see that the SRE follows the same sentiment as the RMSE and PSNR evaluating best performance, though with not exactly the same results. Both PSNR and SRE evaluate the preserved signal from the original image source. However, a possible explanation for the small deviance in results is that the SRE is calculated with a dynamic ceiling value which varies by image, rather than the static ceiling of the PSNR, resulting in favoring different pictures for certain reconstructions.

The number of best performances for the RMSE, PSNR, and SRE by particpant are relatively consistent with the best mean values for subjects 2 and 3. However, for all three measures, subject 5 scores higher on average than subjects 1 and 4, though with fewer best performances than either. This may be due to subject 5 giving better reconstruction on average for pictures where subject 2 and 3 performed best, but not on others.

Subject 3 has been found to perform best a significant amount of the time on all five of our chosen metrics, leading in number of best performances by least 50% in every case. This points towards participant selection as a relevant factor for performance when reconstructing visual stimuli from fMRI data using machine learning.

However, when conducting ANOVA analysis, no significant differences are found in the mean values of performance evaluation for the RMSE, PSNR, SRE, and SSIM. Only the FSIM is shown to have a significant variation in performances between subjects, with a p-value of 0.01. This points to the FSIM being able to distinguish details in image reconstructions that the remaining measures do not. As the sample size of our reconstructions is limited (5 participants with 50 reconstructions each), this finding is not definitive. It does however point the way for future research to further validate it.

9 Conclusions and Future Work

Our results found that one of the participants, namely subject 3, in the Generic Object Decoding data set to consistently perform best in terms of image reconstruction performance across all five employed measures (RMSE, PSNR, SER, SSIM, FSIM). However, analysis of variance using $p < 0.05$ yielded statistically significant results for only the FSIM at $p = 0.01$.

This points to FSIM being a more discriminate performance metric, being more sensitive to changes from the original image than the remaining four. However, as our sample size for reconstructions is small, we recommend future research to investigate further to strengthen this finding.

The MSE, PSNR, and SER were found to rank participants similarly, as evaluation metrics that compare images by the pixel, with the SER's ranking deviating only slightly from the other two. These pixel-wise evaluation metrics are fast and intuitive to calculate, but have been shown to not represent similarity as perceived by the human visual system well.

In contrast, the SSIM and FSIM are comparison methods that model the structural retainment of a reconstructed image to its original, by taking into account factors such as luminance, contrast, and structural patterns. These measures showed differing rankings of subjects by reconstruction performance, though subject 3 was still ranked highest.

The present paper investigated the effect of participant variation on the reconstruction performance on the model proposed by Gaziv et al. However, differing models may favor differing brain anatomies or have varying susceptibility to noise. Future research should thus examine weather the results found in this paper apply to other reconstruction models as well, in order to allow for a more comprehensive understanding of predictive effect of participant selection on reconstruction performance.

Acknowledgements

We would like give special thanks to Xucong Zhang for acting as our supervisor and responsible professor for this project, Nikita Smolin for running model training and reconstruction on their PC, as well as Natalia Bryła and Natalia Pietnoczko for advice on formatting and spaghetti.

References

[1] Zarina Rakhimberdina, Quentin Jodelet, Xin Liu, and Tsuyoshi Murata. Natural image reconstruction from fmri using deep learning: A survey. *Frontiers in Neuroscience*, 15, 2021.

[2] Yu Takagi and Shinji Nishimoto. *High-resolution image reconstruction with latent diffusion models from human brain activity*, 2022.

[3] Sikun Lin, Thomas Sprague, and Ambuj K Singh. Mind reader: Reconstructing complex images from brain activities, 2022.

[4] D. Tomasi, E.C. Caparelli, L. Chang, and T. Ernst. Fmri-acoustic noise alters brain activation during working memory tasks. *NeuroImage*, 27(2):377–386, 2005.

[5] Thomas T. Liu. Noise contributions to the fmri signal: An overview. *NeuroImage*, 143:141–151, 2016.

[6] Seyed Abolfazl Valizadeh, Franziskus Liem, Susan Mérillat, Jürgen Hänggi, and Lutz Jäncke. Identification of individual subjects on the basis of their brain anatomical features. *Scientific Reports*, 8(1), 2018.

[7] Guy Gaziv, Roman Beliy, Niv Granot, Assaf Hoogi, Francesca Strappini, Tal Golan, and Michal Irani. Self-supervised natural image reconstruction and large-scale semantic classification from brain activity. *NeuroImage*, 254:119121, 2022.

[8] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models, 2022.

[9] Peng Wang, Shijie Wang, Junyang Lin, Shuai Bai, Xiaohuan Zhou, Jingren Zhou, Xinggang Wang, and Chang Zhou. One-peace: Exploring one general representation model toward unlimited modalities, 2023.

[10] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.

[11] Tomoyasu Horikawa and Yukiyasu Kamitani. "generic object decoding (fmri on imagenet)", 2019.

[12] M. U. Müller, N. Ekhtiari, R. M. Almeida, and C. Rieke. Super-resolution of multispectral satellite images using convolutional neural networks. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-1-2020:33–40, 2020.

[13] Lars Sthle and Svante Wold. Analysis of variance (anova). *Chemometrics and Intelligent Laboratory Systems*, 6(4):259–272, 1989.

[14]

[15] Charis Lanaras, José Bioucas-Dias, Silvano Galliani, Emmanuel Baltsavias, and Konrad Schindler. Super-resolution of sentinel-2 images: Learning a globally applicable deep neural network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146:305–319, 2018.

[16] Ph.D. Niku Ekhtiari. Comparing ground truth with predictions using image similarity measures, 2021.

[17] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[18] Jim Nilsson and Tomas Akenine-Möller. Understanding ssim, 2020.

[19] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. Fsim: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 20(8):2378–2386, 2011.