Denoising task fMRI data for image reconstructions

Denoising of Functional Magnetic Resonance Imaging (fMRI) Data for Improved Visual Stimulus Reconstruction using Machine Learning

Nikita Smolin[1]

Supervisor(s): 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: Nikita Smolin
Final project course: CSE3000 Research Project
Thesis committee: Xucong Zhang, Nergis Tömen

Abstract

This study aims to investigate the impact of various denoising algorithms on the quality of visual stimulus reconstructions based on functional magnetic resonance imaging (fMRI) data. While fMRI provides a valuable, noninvasive method for assessing brain activity, the reliability of this data can be impaired by multiple noise types, including thermal, physiological, and scanner-related noise. Numerous denoising methods have been proposed, such as independent component analysis, confound regression and filtering, and GLM denoise. However, their efficiency, especially in the context of limited task fMRI data, remains largely unexplored.

Using the Generic Objects Dataset (GOD), our study explores three primary research subquestions: the effectiveness of different denoising algorithms in improving reconstruction quality; the impact of artificially induced noise on these algorithms and whether combining different denoising algorithms can further enhance image reconstruction quality.

The primary contributions of this research include the evaluation of various denoising methods with limited task fMRI data, determining the most effective denoising algorithms given a small dataset size, and analyzing how these algorithms perform in the presence of artificially introduced noise. The results of this investigation showed an improvement in performance of reconstruction models given multiple denoising algorithm, the best performer being kurtosis-based PCA used together with nuisanse regression with constant and linear terms bringing a 6.2% increase in score. The noise ceiling is the worst performer, bringing the score down by 4.4%.

Denoising algorithms fail to improve reconstructions poisoned with gaussian noise, however, ICA manages to achieve a minor improvement of the quality of reconstructed images given noise sampled from the uniform distribution.

Keywords: Visual stimulus reconstruction, fMRI, GOD dataset, fMRI on ImageNet, Denoising, Noise elimination, Artificial noise

1 Introduction

Functional magnetic resonance imaging (fMRI) has revolutionized our understanding of brain function by providing a noninvasive method for measuring neural activity over time.

Despite its immense potential, the accuracy and reliability of fMRI data can be severely compromised by various types of noise, such as thermal, physiological, and scanner-related noise. Denoising fMRI data is therefore a critical step in the process of reconstructing visual stimuli from brain activity, as it can enhance the signal-to-noise ratio of data used for image reconstructions.

Numerous denoising algorithms have been proposed in the literature, including independent component analysis [7], confound regression and filtering [8], and GLM denoise [5]. While these methods have been successful to some extent, there remains a significant knowledge gap in understanding the impact of denoising algorithms for fMRI data on the performance of reconstruction models. Furthermore, given limited availability of task fMRI data in some datasets, state of the art techniques, such as GLM denoise, are not always applicable due to requiring extensive time series data. This gap highlights the need for a systematic investigation of how different denoising algorithms impact the performance of visual stimulus reconstruction models.

In light of these unanswered questions, the primary research question of this study is: How does the denoising of task fMRI data impact the performance of visual stimulus reconstruction models? To address this question, we will investigate the following sub-questions:

1. Which denoising algorithms are most effective in improving reconstruction quality?
2. What is the impact of artificially induced noise on different denoising algorithms?
3. Can combining different denoising algorithms in a chain improve image reconstruction quality?

In order to answer these questions, the Generic Objects Dataset (GOD) will be employed as a testbed for our investigation. This dataset, which consists of fMRI data collected while participants observed generic objects, will provide a comprehensive platform for our exploration. By implementing various denoising algorithms and utilizing the self-supervised RGB reconstruction model from "Self-Supervised RGBD Reconstruction From Brain Activity" by Gaziv et al. [3] on the GOD dataset, this study aims to advance our understanding of the factors that contribute to the success or failure of fMRI data pre-processing and visual stimulus reconstruction.

The main contributions of this research are the evaluation of various denoising algorithms in the context of task fMRI data, the determination of the most effective denoising algorithms given limited data and a investigation of how denoising algorithms act in a presence of artificially added noise. The source code and the RGB-only model for the reconstruction model can be found at this GitHub repository[1].

The remainder of the paper is organized as follows: Section 2 provides a comprehensive review of the relevant literature; Section 3 outlines the methodology employed in the study, including the denoising algorithms, experimental design and evaluation metrics; Section 4 presents the results and analysis of the study; Section 5 includes a discussion of the findings; Section 6 discusses responsible research and Section 7 concludes the work and establishes its implications for future research.

2 Literature Review

The field of functional magnetic resonance imaging (fMRI) has seen significant advancements in recent years, particu-

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

larly in the area of denoising and visual stimulus reconstruction. A number of studies have contributed to our understanding of these processes.

The Human Connectome Project (HCP) has made significant contributions to the field of neuroimaging. Glasser et al. (2016) [4] presented an integrated approach to data acquisition, analysis, and sharing, building upon recent advances from the HCP. Their work emphasizes the importance of collecting multimodal imaging data from many subjects and accurately aligning corresponding brain areas across subjects and studies.

McKeown (2003) [7] discussed the use of independent component analysis (ICA) in fMRI data analysis and the differentiation between signal and noise. This study highlights the potential of data-driven methods, such as ICA, in identifying common features within fMRI data.

Tohka et al. (2008) [10] further expanded on the use of ICA in fMRI data analysis by introducing an automatic independent component labeling method for artifact removal. This work provided an effective strategy to enhance the quality of fMRI data and the accuracy of its interpretation.

St-Yves and Naselaris (2018) [9] presented a novel approach to visual stimulus reconstruction, introducing Generative Adversarial Networks (GANs) conditioned on brain activity. This study signified an advancement in the field, demonstrating the potential of deep learning techniques in fMRI data analysis and interpretation.

Gaziv et al. (2022) [3] made a significant stride in visual stimulus reconstruction with their work ”Self-Supervised RGBD Reconstruction From Brain Activity”. They presented a novel approach that uses self-supervised learning techniques for reconstructing RGB images from fMRI data using encoder-decoder model architecture. This study emphasized the power of high-level perceptual objectives with self-supervision to improve the quality of reconstructed images, providing a compelling case for the potential of self-supervised learning in advancing visual stimulus reconstruction.

Dubois and Adolphs (2016) [1] discussed the shift in focus to individual subjects in fMRI research. They emphasized the need for careful consideration of anatomical and vascular between-subject variability as well as sources of within-subject variability. Their work underscores the importance of individual differences in fMRI studies.

Erhardt et al. (2010) [2] provided extensive comparisons of four multi-subject ICA approaches in combination with data reduction methods for simulated and fMRI task data. Their work provides valuable insights into the performance of different ICA methods in the analysis of fMRI data.

Mascali et al. (2021) [6] have conducted an evaluation of denoising strategies for task-based functional connectivity, including pca and ica based methods, as well as global signal regression.

In summary, the literature provides a rich source of information on the denoising of fMRI data and visual stimulus reconstruction. The studies highlighted in this review have contributed to our understanding of these processes and have laid the groundwork for further research in this area.

3 Methodology

3.1 Dataset

The Generic Objects Dataset (GOD), otherwise known as fMRI on Imagenet, served as the primary dataset for this study. The GOD contained fMRI data collected while participants observed generic objects, as well as the images that the participants observed.. A version of the GOD dataset, pre-processed by by [3] with pre-calculated noise ceiling values was used in order to evaluate not only the efficiency of ICA/PCA based methods, but also directly compare them to processing methods used with the proposed visual reconstruction model.

3.2 Denoising Algorithms

Multiple denoising algorithms were implemented and evaluated to remove noise from the fMRI data. The selection of denoising algorithms was based on a literature review and established approaches such as independent component analysis (ICA), Principal Component Analysis (PCA), nuisance regression and noise ceiling. These algorithms had shown promising results in previous studies. The performance of each denoising algorithm was assessed by comparing the quality of the reconstructed images to the quality of reconstructed images without any denoising algorithms applied.

Component-based methods - ICA and PCA

Both ICA and PCA allowed us to split a fMRI signal into multiple components. Given such a decomposition, noise components can be identified using different measures. We decided to use kurtosis in order to detect non-gaussian distributed outlier components, as high kurtosis indicates outliers that could potentially be thermal, motion and physiological noise in a given fMRI response.

Nuisance Regression

Due to the lack of motion data, only two nuisance regressors were attempted - constant and linear terms - in order to possibly eliminate minor physiological trends over the course of experiments. The regressors were fit to fMRI data, the predictions of these regressors were later subtracted from the fMRI data.

Noise ceiling

Authors of [3] implemented noise ceiling, which was also used in evaluation and comparison of denoising algorithms with an exception of artificially induced noise - no noise ceiling was applied for such.

Combining methods

Besides standalone ICA, PCA, Nuisance Regression and Noise ceiling, a combination of pca and ica, as well as ica with regression and pca with regression and all of these combinations with and without noise ceiling were also employed in evaluation and comparisons, in order to answer the third research question.

3.3 One Subject

Due to possible differences in noise for different participants in different scanning sessions, the analysis was performed on

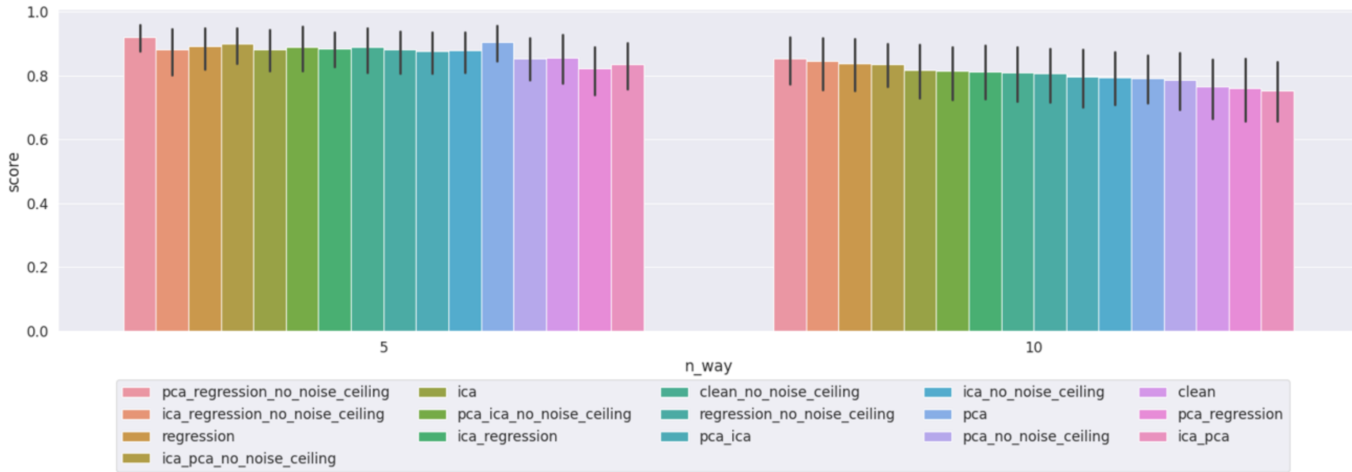

Figure 1: Accuracy(the higher the better)

exp_name n_way	clean	ica	ica_pca	ica_regression	pca_ica	pca	pca_regression	regression
5	0.888	0.878	0.900	0.880	0.888	0.854	**0.920**	0.880
10	0.810	0.794	0.836	0.844	0.814	0.786	**0.854**	0.806

Table 1: Accuracy for experiments without noise ceiling(the higher the better).

exp_name n_way	clean	ica	ica_pca	ica_regression	pca_ica	pca	pca_regression	regression
5	0.856	0.880	0.836	0.884	0.876	**0.904**	0.822	0.892
10	0.766	0.816	0.752	0.812	0.796	0.790	0.760	0.**838**

Table 2: Accuracy for experiments with noise ceiling(the higher the better).

one subject - subject number 3, to eliminate possible physiological and psychological differences between subjects that could introduce additional complications in both implementing and evaluating denoising algorithms.

3.4 Visual Stimulus Reconstruction Model

The RGB model utilized by Gaziv et al. is a part of their broader self-supervised approach to brain activity reconstruction. This model relies on both supervised and self-supervised learning to handle the limitations of available paired fMRI and image data.

The training process is divided into two stages: Encoder training and Decoder training. The Encoder encodes the RGB information into the corresponding fMRI responses, and the Decoder translates fMRI recordings back to their corresponding RGB information. Concatenating these two networks, Encoder-Decoder, forms a combined network where the input and output are the same RGB information. This structure enables unsupervised learning on unpaired data (i.e., RGB data without corresponding fMRI recordings), which is key to adapting the network to the statistical properties of previously unseen RGB data [3].

3.5 Experimental Procedure

Participants who contributed to the collection of the GOD dataset were not directly involved in this study. Instead, the focus was on the analysis of the pre-collected fMRI data. The denoising algorithms were applied to the full fMRI data, which was later split into a training and a testing set and the reconstructed stimuli were compared against the ground truth stimuli using relevant metrics as well as performing a manual inspection. The entire process was automated to ensure consistency and reproducibility, as well as providing human-centric details. The experiments were run on a single computer with a ryzen 3700x CPU, RTX 3080 with 10 gigabytes of virtual memory and 64 gigabytes of RAM.

The study followed a cross-validation design to assess the performance of different denoising algorithms and visual stimulus reconstruction models. The dataset was divided into training and testing sets. The training set was used for training the reconstruction models, while the testing set was used to evaluate their performance. The whole dataset was denoised using a variety of denoising methods. The original hyperparameters presented in the repository remain unchanged to ensure ease of experiment replication, with an exception of evaluating the effects of artificial noise, where the number of

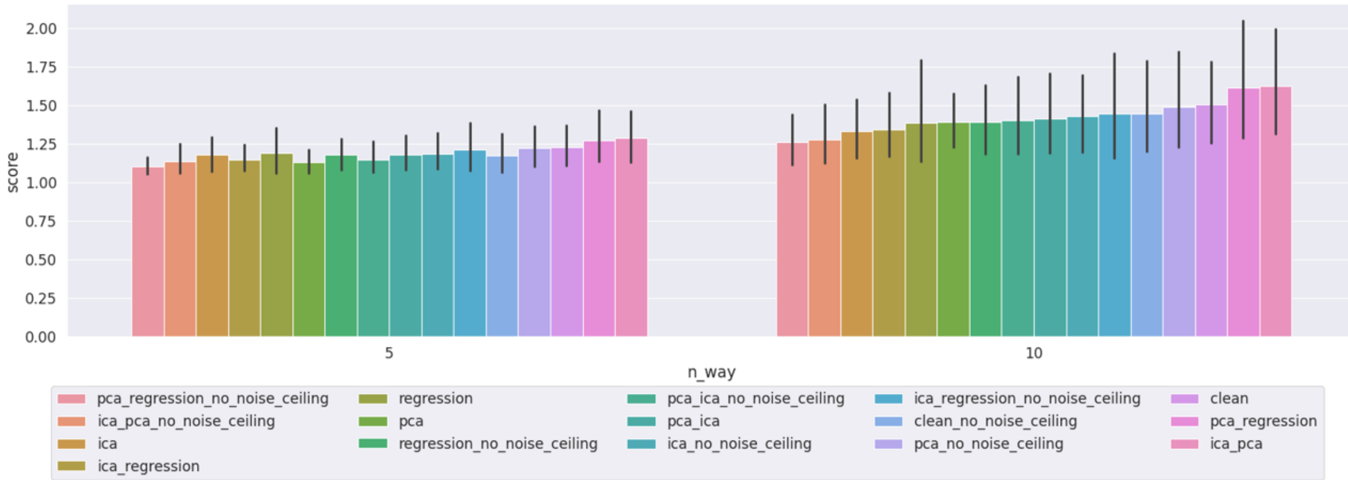

Figure 2: Score(the lower the better)

exp_name n_way	clean	ica	ica_pca	ica_regression	pca_ica	pca	pca_regression	regression
5	1.174	1.184	1.136	1.210	1.148	1.222	**1.102**	1.176
10	1.446	1.426	1.274	1.442	1.400	1.486	**1.260**	1.390

Table 3: Score for experiments without noise ceiling(the lower the better).

exp_name n_way	clean	ica	ica_pca	ica_regression	pca_ica	pca	pca_regression	regression
5	1.228	1.176	1.288	1.146	1.178	**1.128**	1.272	1.186
10	1.502	**1.332**	1.624	1.340	1.412	1.388	1.612	1.386

Table 4: Score for experiments with noise ceiling(the lower the better).

training epochs of a decoder was decreased from 150 to 30. Denoising algorithms were applied specifically to a complete dataset, before a cross-validation split, in order to ensure consistency of removed components.

3.6 Evaluation Metrics

To assess the performance of the denoising algorithms and visual stimulus reconstruction models, a ranking system with two evaluation metrics were employed, also used in [3]. The original image, as well as a reconstructed image and a set of 3 and 8 distraction images, to form a total of 5 and 10 images were ranked based on similarity to the original image. Later on, two evaluation metrics are calculated. The first metric is accuracy of correctly ranking the original image as the most similar to itelf, which is used as a measure of how accurate the ranking process is. The second metric is average rank, that we later refer to as score. The higher rank on average the reconstruction gets, the lower the score. This metric is used to evaluate the quality of reconstructions keeping human perception of results in mind. Furthermore, a manual reconstruction image inspection was conducted in order to infer the impact of noise reduction on high-level scene features such as shape, main subject color and background color.

In evaluating the effectiveness of our reconstruction models, a critical component is quantifying the similarity between the original and reconstructed images. To achieve this, we employ the Learned Perceptual Image Patch Similarity (LPIPS) metric [2]. LPIPS is a perceptual similarity metric that uses deep learning to approximate human visual perception. In essence, it measures the perceptual difference between two images as a human observer might perceive them, rather than simple pixel-to-pixel comparisons that traditional metrics rely upon. This metric allows us to simulate human judgement and evaluate results in a more interpretable way compared to traditional machine learning metrics, such as accuracy and loss.

Unlike some similarity metrics, the LPIPS score is not strictly bounded; higher scores indicate greater perceptual differences between images. As such, a successful model would aim to minimize the LPIPS score in its reconstructed images. This score provides an interpretable and human-centric measure to evaluate the success of our reconstruction models.

[2]https://github.com/richzhang/PerceptualSimilarity

3.7 Artificial Noise

To investigate the effects of artificial noise on quality of reconstructed images and study the possible negation of such effects with denoising algorithms, two different noise models were applied - gaussian noise sampled from a normal distrubution with mean 0 and standard deviation of 1, as well as noise sampled from a uniform distribution with a lower bound of -1 and a higher bound of 1.

4 Results

4.1 Reconstruction performance with different denoising pipelines

Figure 1 and tables 1 and 2 showcase a comparison of correct ranking accuracy of the original image with different denoising pipelines applied.

Figure 2, as well as tables 3 and 4 show an average ranking score for reconstructed images given a denoising pipeline.

In both cases with and without the noise ceiling, two versions of metric calculation are presented - a 5-way and a 10-way candidate comparison during ranking. Best performing algorithm is indicated in bold for both a 5-way comparison and a 10-way comparison

4.2 Reconstruction performance given artificially induced noise sampled from a gaussian distribution

Tables 5 and 6 show correct ranking accuracy of the original image and average ranking score of a reconstructed image respectfully in 5-way and 10-way variations given different denoising pipelines. All denoising algorithms clearly failed to improve reconstruction performance, which makes sense given that our comparison includes constant, linear and non-gaussian outlier noise identification techniques and no gaussian noise identification techniques.

4.3 Reconstruction performance given artificially induced noise sampled from a uniform distribution

Tables 7 and 8 show correct ranking accuracy of the original image and average ranking score of a reconstructed image respectfully in a 5-way and 10-way variations given different denoising pipelines. ICA performs the best, unlike pca, that performs the worst, even though the noise component selection method is the same for both component-based algorithms.

5 Discussion

5.1 Analysis of denoising algorithms

The results found in chapter 4 section 1 suggest that even with limited time-series data available, as well as a small amount of images, it is possible to achieve significant improvements in reconstruction performance by utilizing traditional machine-learning techniques such as linear regression, pca and ica. However, after manually inspecting the images, mixed results appear. For some images, denoising has a highly positive impact, for example for an umbrella image shown in Fig. 3, the color overall is closer to the original image after performing pca and nuisance regression.

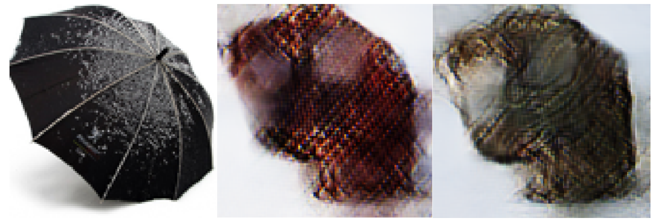

Figure 3: Original image of an umbrella next to reconstructed images without and with applying pca + nuisance regression

In contrast, the opposite situation happens with the picture of an airplane showcased in Fig. 4 - applying pca and regression results in a significant color distortion, as well as no visible shape improvement. A similar situation happens with other algorithms - for some pictures there is a visible improvement, for others - visible distortion.

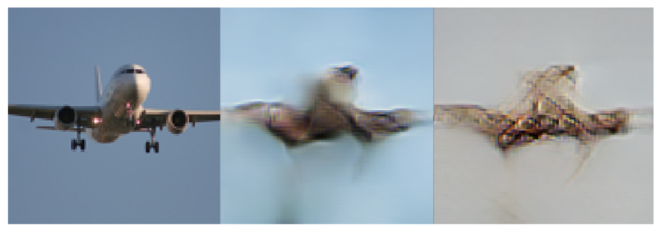

Figure 4: Original image of an airplane next to reconstructed images without and with applying pca + nuisance regression

Several causes of such behavior are possible. First of all, even though the size of the dataset is artificially increased during the training process, many images of objects appear only once in the dataset and there aren't many similar pictures in the training set, potentially causing very high or very low kurtosis leading to important components being removed from the dataset. One possible way to address this problem is to increase the dataset size, however it is difficult to do. Moreover, a new scanning session will likely introduce a different variation of thermal and physiological noise, as well as new head movement in a different machine. Another possible way to address this problem is to introduce a more complicated algorithm of identification of noise component, that not only considers kurtosis, but also other measures of outliers to prevent false positives from occurring, which causes the model to perform worse.

5.2 Artificially induced noise

Artificially induced noise creates a significant danger to the quality of reconstruction images, if specifics of noise are not considered. For instance, since kurtosis of uniform noise is significantly lower than kurtosis of a normally distributed variable, ica manages to mitigate some of the negative effects of the noise. However, with noise sampled from a gaus-

exp_name n_way	clean	ica	ica_regression	pca	pca_regression	regression
5	**0.626**	0.616	0.598	0.452	0.424	0.484
10	0.470	0.492	**0.492**	0.318	0.244	0.374

Table 5: Accuracy(the higher the better) for experiments with artificially induced noise sampled from a gaussian distribution.

exp_name n_way	clean	ica	ica_regression	pca	pca_regression	regression
5	**1.746**	1.812	1.778	2.206	2.154	2.164
10	**2.634**	2.844	2.842	3.752	3.500	3.682

Table 6: Score(the lower the better) for experiments with artificially induced noise sampled from a gaussian distribution

exp_name n_way	clean	ica	ica_regression	pca	pca_regression	regression
5	0.700	**0.742**	0.724	0.614	0.648	0.688
10	0.576	**0.656**	0.598	0.528	0.506	0.536

Table 7: Accuracy(the higher the better) for experiments with artificially induced noise sampled from a uniform distribution.

exp_name n_way	clean	ica	ica_regression	pca	pca_regression	regression
5	1.568	**1.404**	1.460	1.770	1.750	1.572
10	2.250	**1.852**	2.038	2.646	2.700	2.266

Table 8: Score(the lower the better) for experiments with artificially induced noise sampled from a uniform distribution

sian distribution, applying ica leads to a further model performance deterioration.

6 Responsible Research

Research ethics and reproducibility are fundamental aspects of any scientific investigation. These principles ensure that the research is conducted with integrity and that the findings can be trusted and built upon by future studies.

In the current study, ethical considerations were upheld throughout the research process. All used open-source data adhered to the principles of informed consent, confidentiality, and non-maleficence. This meant that all participants were fully informed about the purpose of the study, the procedures involved, and their right to withdraw at any point without any negative consequences. Their privacy was protected by anonymizing the data and securely storing all collected information.

In terms of reproducibility, the research methods were designed with transparency and replicability in mind. Detailed information about the experimental design, data collection and analysis procedures, and statistical methods were provided. This transparency will allow other researchers to understand the methods, reproduce the results, and extend the findings in their own studies.

Furthermore, the use of validated measures and techniques, such as the independent component analysis (ICA), principal component analysis (PCA), Linear Regression, as well as noise ceiling and a Encoder Decoder model architecture, contributes to the reliability and replicability of the findings. The use of a publicly available dataset and a model further enhances the reproducibility of this study.

In conclusion, the ethical considerations and reproducibility of the methods were thoroughly addressed in this research. Future studies are encouraged to maintain these high standards of responsible research to ensure the continued advancement of the field of fMRI data analysis and visual stimulus reconstruction.

7 Conclusions and Future Work

This study has provided new insights into the application of denoising algorithms on functional magnetic resonance imaging (fMRI) data given tight dataset size limits. Our findings underscore that denoising can indeed lead to an enhancement in visual stimulus reconstructions. Specifically, the kurtosis-based PCA with nuisance regression employing constant and linear terms emerged as the most effective algorithm, yielding a substantial 6.2% increase in score. However, performance was contingent on the image and noise type, with artificially induced noise posing a significant challenge to the reconstruction quality.

The results indicate that PCA and ICA, while effective in some instances, can benefit from further refinement. Future work should consider different noise component identification algorithms to enhance their performance, especially considering varying noise distributions. This conclusion is further reinforced by our finding that the noise component identification algorithm choice is highly dependent on the size of the dataset, indicating the need for tailored denoising strategies.

During model selection, our research has experienced poor replicability of existing algorithms and the state of open-source denoising repositories is lacking, emphasizing the need for more robust and consistent algorithm development and reporting practices.

Given these findings, there are several clear avenues for future work. More research is necessary to explore the efficiency of denoising algorithms given different dataset sizes, considering the implications of resource limitations and varying noise types. Furthermore, ongoing efforts should be made to improve the reproducibility of denoising methods and the robustness of open-source repositories to support the broader research community.

In conclusion, this study offers a crucial step forward in understanding the impact of denoising algorithms on the quality of visual stimulus reconstructions from fMRI data. However, it also illustrates that there are still many challenges to overcome, highlighting the need for ongoing investigation and development in the field of neuroimaging.

References

[1] Julien Dubois and Ralph Adolphs. Building a science of individual differences from fmri. *Trends in Cognitive Sciences*, 2016.

[2] Erik B Erhardt et al. Comparison of multi-subject ica methods for analysis of fmri data. *Human Brain Mapping*, 2010.

[3] 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.

[4] Matthew F Glasser et al. The human connectome project's neuroimaging approach. *Nature Neuroscience*, 2016.

[5] Kendrick Kay, Ariel Rokem, Jonathan Winawer, Robert Dougherty, and Brian Wandell. Glmdenoise: a fast, automated technique for denoising task-based fmri data. *Frontiers in Neuroscience*, 7, 2013.

[6] Daniele Mascali, Marta Moraschi, Mauro DiNuzzo, Silvia Tommasin, Michela Fratini, Tommaso Gili, Richard G. Wise, Silvia Mangia, Emiliano Macaluso, and Federico Giove. Evaluation of denoising strategies for task-based functional connectivity: Equalizing residual motion artifacts between rest and cognitively demanding tasks. *Human Brain Mapping*, 42(6):1805–1828, 2021.

[7] M McKeown. Independent component analysis of functional mri: what is signal and what is noise? *Current Opinion in Neurobiology*, 2003.

[8] T. D. Satterthwaite, M. A. Elliott, R. T. Gerraty, K. Ruparel, J. Loughead, M. E. Calkins, S. B. Eickhoff, H. Hakonarson, R. C. Gur, R. E. Gur, and D. H. Wolf. An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*, 64:240–256, 2013.

[9] Ghislain St-Yves and Thomas Naselaris. Generative adversarial networks conditioned on brain activity reconstruct seen images. *bioRxiv*, 2018.

[10] Jussi Tohka, Karin Foerde, Adam R Aron, Sabrina M Tom, Arthur W Toga, and Russell A Poldrack. Automatic independent component labeling for artifact removal in fmri. *Neuroimage*, 39(3):1227–1245, 2008.