Unveiling the Mind's Eye: fNIRS-Vise and the Future of Brain Decoding

by

Eyyüb Güven

://orcid.org/0009-0005-6017-8622

https://github.com/curiousbrutus

https://osf.io/aw9g3/

-

A Master’s Thesis
Submitted to Department of Cognitive Science

Faculty of Philosophy and Social Sciences

Nicolaus Copernicus University in Torun

Supervised by Doctor Bogna Bylicka
July 2024

NICOLAUS COPERNICUS
UNIVERSITY
IN TORUŃ

Table of Contents

Abstract

Functional near-infrared spectroscopy (fNIRS) offers a promising non-invasive approach for decoding brain activity. This thesis explores the theoretical potential and practical challenges of fNIRS-Vise, a framework that leverages deep learning for various applications. We review existing literature, highlighting the power of open-source datasets and pre-trained models in decoding visual stimuli, emotional states, mental workload, speech, and disease. Furthermore, we investigate the feasibility of transfer learning between fMRI and fNIRS, as well as across different fNIRS datasets, to enhance model performance and reduce data requirements. While our own practical experiments faced challenges, this thesis provides a comprehensive theoretical foundation and identifies promising directions for future research in this exciting field.This research contributes to the growing body of work exploring fNIRS-based brain decoding and paves the way for future applications in brain-computer interfaces, neurofeedback, and clinical diagnostics.

Keywords: fNIRS, fMRI, functional magnetic resonance imaging, functional near-infrared spectroscopy, brain decoding, deep learning, transformers, transfer learning, neuroimaging, neuroscience, cognitive science, brain-computer interfaces (BCIs), neurofeedback, emotion recognition, mental workload assessment, visual reconstruction, machine learning, neural networks, data analysis.

Polish Title:

Odsłaniając Oko Umysłu: NIRS-Vise i Przyszłość Dekodowania Mózgu

Polish Abstract:

Funkcjonalna spektroskopia bliskiej podczerwieni (fNIRS) stanowi obiecującą, nieinwazyjną metodę dekodowania aktywności mózgu. Niniejsza praca analizuje potencjał teoretyczny oraz praktyczne wyzwania związane z NIRS-Vise, nowatorskim podejściem wykorzystującym głębokie uczenie w różnych zastosowaniach. Dokonujemy przeglądu istniejącej literatury, podkreślając znaczenie otwartych zbiorów danych oraz modeli pre-trenowanych w dekodowaniu bodźców wzrokowych, stanów emocjonalnych, obciążenia umysłowego, mowy oraz wykrywaniu chorób. Dodatkowo, badamy możliwość transferu wiedzy między fMRI a fNIRS, jak również między różnymi zbiorami danych fNIRS, w celu poprawy wydajności modeli i zmniejszenia wymagań dotyczących danych. Chociaż nasze własne eksperymenty praktyczne napotkały pewne trudności, niniejsza praca stanowi wszechstronną podstawę teoretyczną oraz wyznacza obiecujące kierunki dla przyszłych badań w tej dynamicznie rozwijającej się dziedzinie. Badania te przyczyniają się do rosnącego zbioru prac badawczych nad dekodowaniem aktywności mózgu za pomocą fNIRS i torują drogę dla przyszłych zastosowań w interfejsach mózg-komputer, neurofeedbacku oraz diagnostyce klinicznej.

Słowa kluczowe: fNIRS, fMRI, funkcjonalny rezonans magnetyczny, funkcjonalna spektroskopia bliskiej podczerwieni, dekodowanie mózgu, uczenie głębokie, transformatory, transfer uczenia, neuroobrazowanie, neuronauka, nauki kognitywne, interfejsy mózg-komputer (BCI), neurofeedback, rozpoznawanie emocji, ocena obciążenia umysłowego, rekonstrukcja wzrokowa, uczenie maszynowe, sieci neuronowe, analiza danych.

Introduction

Functional near-infrared spectroscopy (fNIRS) has emerged as a promising neuroimaging modality due to its non-invasiveness, portability, and affordability, making it an attractive alternative to traditional methods like fMRI. While early fNIRS research primarily focused on task-based analysis, the recent integration of deep learning, particularly transformer models, has revolutionized the field by enabling the decoding of complex information from fNIRS signals (Wang et al., 2022). This thesis investigates the theoretical potential and practical challenges of fNIRS-Vise, a framework that leverages deep learning for a wide array of applications, including decoding visual stimuli (Chen et al., 2023), emotional states, mental workload, speech, and disease detection.

The availability of open-source datasets and pre-trained models has significantly accelerated research in fNIRS-Vise. Building on this foundation, this thesis explores the feasibility of transfer learning between fMRI and fNIRS, as well as across different fNIRS datasets. By leveraging pre-trained weights from existing research, we aim to enhance model performance, reduce the need for extensive labeled fNIRS data, and ultimately facilitate the development of real-world applications, such as brain-computer interfaces and neurofeedback systems.

While our own practical experiments with the MinD-Vis reproduction encountered challenges, this thesis provides a comprehensive theoretical framework and a critical evaluation of the current state-of-the-art. We review existing literature, highlighting successful applications and limitations of fNIRS-Vise, and discuss the potential advantages of transfer learning in this context. Moreover, we identify promising directions for future research to address the challenges and unlock the full potential of this emerging field.

By exploring the theoretical underpinnings and practical implications of fNIRS-Vise, this thesis aims to contribute to the growing body of knowledge in this exciting field and pave the way for transformative applications in neuroscience, healthcare, and beyond.

1. NEUROIMAGING LITERATURE REVIEW

1.1 Functional Magnetic Resonance Imaging (fMRI)

Functional magnetic resonance imaging (fMRI) is a sophisticated non-invasive neuroimaging technique that has revolutionized our understanding of the human brain's inner workings. By harnessing the principles of magnetic resonance imaging (MRI) and measuring subtle changes in blood oxygenation levels, fMRI provides researchers with a dynamic window into the brain's activity during a wide array of cognitive tasks and resting states. This section delves into the intricate workings of fMRI, tracing its historical development, exploring its diverse applications in neuroscience, and examining the methodological considerations that underpin its use.

1.1.1 Historical Development

Functional magnetic resonance imaging (fMRI) emerged as a groundbreaking neuroimaging technique in the early 1990s, following Seiji Ogawa et al.'s (1990) pivotal discovery of the blood-oxygen-level-dependent (BOLD) contrast. The BOLD signal, a fundamental principle in fMRI, reflects localized changes in blood flow and oxygenation that correlate with neural activity. Advancements in MRI technology, such as stronger magnetic fields and refined pulse sequences, combined with sophisticated data analysis methods, have propelled fMRI to the forefront of brain research, enabling a deeper understanding of the brain's organization, connectivity, and adaptability (Bandettini et al., 1992).

1.1.2 Working Principles

The working principles of fMRI rest upon the intricate relationship between neural activity, blood flow, and oxygen consumption. When neurons in a particular brain region become active, they require more energy, which is supplied by an increased flow of oxygenated blood to that region. This influx of oxygenated blood alters the local magnetic properties of the tissue, leading to a detectable change in the MRI signal.

fMRI scanners employ powerful magnets to create a strong magnetic field that aligns the protons within the brain's water molecules. Radiofrequency pulses are then used to perturb this alignment, and the scanner measures the time it takes for the protons to return to their original state. The BOLD signal manifests as a subtle difference in the relaxation times of protons in oxygenated versus deoxygenated blood.

By systematically acquiring MRI images while a participant performs a specific task or rests, researchers can create a series of "brain maps" that depict the regions of the brain that are most active during different phases of the experiment. These brain maps provide a visual representation of the brain's functional landscape, highlighting the neural networks that underlie various cognitive processes.

1.1.3 Methodological Considerations

fMRI experiments typically adhere to a well-defined methodological framework. Participants are placed in the MRI scanner and presented with carefully designed stimuli or instructed to perform specific tasks. During the scanning session, the scanner acquires a series of MRI images, capturing the dynamic changes in the BOLD signal over time.

The resulting fMRI data undergoes a multi-step preprocessing pipeline to enhance the quality of the signal and remove artifacts that may arise from head motion, physiological noise, or scanner imperfections. This preprocessing typically involves motion correction, slice-timing correction, spatial smoothing, and normalization to a standard brain template.

Once the data is preprocessed, statistical analysis is performed to identify regions of the brain that exhibit significant changes in activity during the task. This analysis often relies on the General Linear Model (GLM), a statistical framework that models the relationship between the experimental design (task conditions) and the observed BOLD signal (Friston et al., 1994).

1.1.4 Data Analysis Softwares and Methodologies

The analysis of fMRI data requires specialized software tools that are capable of handling the large volumes of data generated by fMRI experiments. Some of the most popular fMRI analysis software packages include:

- **SPM (Statistical Parametric Mapping):** A widely used software for fMRI analysis that provides a comprehensive set of tools for preprocessing, statistical modeling, and visualization (Penny et al., 2011).

- **FSL (FMRIB Software Library):** Another popular software package that offers a wide range of tools for fMRI analysis, including preprocessing, registration, statistical modeling, and connectivity analysis (Jenkinson et al., 2012).
- **AFNI (Analysis of Functional NeuroImages):** A powerful software package that provides a variety of tools for fMRI analysis, including preprocessing, statistical modeling, and visualization (Cox, 1996).

In addition to these commonly used software packages, a plethora of other specialized tools cater to specific fMRI analysis needs. For instance, BrainVoyager is known for its advanced visualization capabilities, allowing researchers to explore brain activity in three-dimensional space. CONN toolbox specializes in functional connectivity analysis, enabling the investigation of how different brain regions interact during cognitive tasks. The choice of software often depends on the specific research questions and analysis techniques employed.

1.1.5 Significant Findings and Future Directions

fMRI research has yielded a wealth of knowledge about the human brain, revealing its intricate functional organization, the neural substrates of various cognitive processes, and the alterations in brain activity associated with neurological and psychiatric disorders. Landmark fMRI studies have mapped the visual cortex, elucidated the neural basis of language processing, and identified brain regions involved in memory, decision-making, and social cognition.

However, fMRI research is not without its challenges. One notable limitation is its relatively low temporal resolution, as the BOLD signal reflects changes in blood flow that occur over a timescale of seconds. This limits the ability of fMRI to capture rapid neural events that unfold on a millisecond timescale. Additionally, the interpretation of fMRI data can be complex, as the BOLD signal is an indirect measure of neural activity, influenced by various physiological factors.

Despite these challenges, fMRI continues to be a cornerstone of neuroscience research, driving ongoing efforts to address its limitations and unlock its full potential. Advancements in MRI technology, such as ultra-high field MRI, are enabling the acquisition of fMRI data at increasingly higher spatial resolutions, potentially revealing fine-grained details of brain organization and function. Novel analysis techniques, such as real-time fMRI and functional connectivity analysis, are also expanding the possibilities for fMRI research.

The future of fMRI is bright, with ongoing research poised to address the limitations of this powerful neuroimaging modality and expand its applications in diverse fields. Real-time fMRI, for instance, offers the potential to provide feedback to participants based on their brain activity, opening up new avenues for neurofeedback training and brain-computer interfaces. High-resolution fMRI could revolutionize our understanding of the brain's microstructural organization and connectivity. The integration of fMRI with other neuroimaging techniques, such as electroencephalography (EEG) and magnetoencephalography (MEG), promises to provide a more comprehensive picture of brain function by combining the strengths of different modalities.

In conclusion, fMRI stands as a testament to human ingenuity and scientific progress, offering a non-invasive window into the brain's dynamic landscape. Its impact on neuroscience has been profound, transforming our understanding of the neural mechanisms underlying human thought, emotion, and behavior. As fMRI technology continues to advance and novel analysis techniques emerge, we can anticipate a future where fMRI plays an even more pivotal role in unraveling the mysteries of the human brain and its disorders.

1.2 Functional Near-Infrared Spectroscopy (fNIRS)

Functional near-infrared spectroscopy (fNIRS) is a non-invasive neuroimaging technique that has been gaining traction in recent years as a promising alternative to fMRI. Leveraging the principles of near-infrared light absorption and scattering, fNIRS offers a portable, cost-effective, and versatile tool for investigating brain function in a variety of settings. This section explores the principles, methodologies, applications, and emerging trends in fNIRS research.

1.2.1 Working Principles

Functional near-infrared spectroscopy (fNIRS) is a non-invasive neuroimaging technique that utilizes near-infrared light (650-950 nm) to penetrate biological tissues, such as the skull and scalp. Near-infrared light interacts with specific molecules in the brain, mainly oxyhemoglobin and deoxyhemoglobin, whose concentrations change with neural activity. (Ferrari & Quaresima, 2012).

fNIRS systems consist of light sources (emitters) and detectors (receivers) positioned on the scalp. The emitters release near-infrared light that travels through the scalp and skull, reaching the cortical surface. The light is then either absorbed or scattered, and the scattered light is detected by the receivers. By analyzing the intensity of the detected light, changes in HbO and HbR concentrations can be calculated, providing valuable insights into the underlying neural activity (Villringer & Chance, 1997).

1.2.2 Methodological Considerations

fNIRS data acquisition involves several methodological considerations. The placement of the emitters and detectors, known as the optode arrangement, is crucial for ensuring adequate coverage of the brain regions of interest. Different optode arrangements, such as linear, rectangular, and custom configurations, are available to suit different experimental designs (Boas et al., 2004).

Data preprocessing is essential for removing artifacts and noise from the fNIRS signal. Common preprocessing steps include motion correction, baseline correction, filtering, and signal detrending (Cui et al., 2010). Statistical analysis techniques, such as the general linear model (GLM) or wavelet analysis, are then employed to identify significant changes in HbO and HbR concentrations associated with specific tasks or stimuli (Abdelnour & Huppert, 2009).

1.2.3 Applications in Neuroscience

fNIRS has found applications in a wide range of neuroscience research domains, including:

- **Cognitive Neuroscience:** fNIRS has been used to investigate the neural correlates of various cognitive processes, such as attention, memory, language, and decision-making (Pinti et al., 2020).
- **Developmental Neuroscience:** Due to its non-invasiveness and tolerance to motion, fNIRS is well-suited for studying brain development in infants and children (Lloyd-Fox et al., 2010).

- **Clinical Neuroscience:** fNIRS has shown promise in the diagnosis and monitoring of various neurological and psychiatric disorders, such as stroke, traumatic brain injury, and depression (Scholkmann et al., 2014).
- **Brain-Computer Interfaces (BCIs):** fNIRS-based BCIs have the potential to enable individuals to control external devices or communicate using their brain signals (Naseer & Hong, 2015).

1.2.4 fNIRS Data Preprocessing and Analysis Softwares

The analysis of fNIRS data involves specialized software tools capable of handling the unique characteristics of fNIRS signals. Several software packages are available for fNIRS data preprocessing and analysis, including:

- **Homer2:** A widely used open-source MATLAB toolbox that provides a comprehensive set of tools for fNIRS data preprocessing, visualization, and analysis (Huppert et al., 2009).
- **NIRS-SPM:** A MATLAB-based software package specifically designed for statistical parametric mapping of fNIRS data (Ye et al., 2009).
- **FC-NIRS:** A MATLAB toolbox that specializes in functional connectivity analysis of fNIRS data (Chang & Glover, 2010).

1.2.5 Current Trends and Future Directions

fNIRS is a rapidly evolving field, with ongoing research pushing the boundaries of its capabilities. Current trends include the development of:

- **High-density fNIRS systems:** These systems employ a large number of emitters and detectors, providing higher spatial resolution and improved brain coverage (Ferrari & Quaresima, 2012).
- **Multimodal fNIRS:** The combination of fNIRS with other neuroimaging modalities, such as EEG, offers a more comprehensive view of brain function (Fazli et al., 2012).
- **Wearable and wireless fNIRS systems:** These innovations offer greater flexibility and comfort for participants, enabling the study of brain function in more naturalistic settings (Piper et al., 2014).
- **Hybrid fNIRS-EEG systems:** Combining the strengths of fNIRS and EEG could provide a more comprehensive picture of brain activity, with fNIRS offering insights into hemodynamic responses and EEG capturing electrical activity (Safaie et al., 2013).

The future of fNIRS is brimming with potential. As technology continues to advance, we can expect fNIRS systems to become even more portable, affordable, and user-friendly. This could lead to a broader adoption of fNIRS in research and clinical settings, enabling new insights into brain function and dysfunction. The development of novel analysis techniques, such as machine learning algorithms for decoding brain states and connectivity patterns, could further enhance the power of fNIRS to unravel the complexities of the human brain.

1.2.6 Limitations and Future Directions

While fNIRS offers numerous advantages, it is not without its limitations. One notable challenge is its limited penetration depth, as near-infrared light is primarily absorbed by the cortical surface and has limited reach into deeper brain structures. Additionally, fNIRS signals can be affected by extracranial factors such as scalp and skull thickness, which may introduce variability in measurements (Tachtsidis & Scholkmann, 2016).

Ongoing research aims to address these limitations and enhance the capabilities of fNIRS. For example, the development of new optode designs and light sources could improve penetration depth and signal quality. The integration of anatomical information from MRI scans could help to account for individual differences in head anatomy and improve the accuracy of fNIRS measurements.

fNIRS has the potential to revolutionize our understanding of the human brain and its disorders. By providing a non-invasive, portable, and affordable tool for investigating brain function, fNIRS could democratize neuroscience research and make it accessible to a wider range of researchers and clinicians. The continued development and refinement of fNIRS technology and analysis methods promise a future where fNIRS plays an increasingly important role in unraveling the mysteries of the human brain and its intricate workings.

1.3 Preprocessing of fMRI and fNIRS Data

The raw data obtained from fMRI and fNIRS experiments is often contaminated with noise and artifacts that can obscure the underlying neural signals. Preprocessing is a critical step in both modalities to enhance signal quality, remove unwanted sources of variance, and prepare the data for subsequent analysis. This section outlines the common preprocessing steps involved in both fMRI and fNIRS data analysis.

1.3.1 fMRI Data Preprocessing

fMRI data preprocessing typically involves a series of steps aimed at correcting for various sources of noise and artifacts. These steps may include:

- **Slice Timing Correction:** Corrects for the slight differences in acquisition time between different slices of the brain, ensuring that all slices are temporally aligned (Sladky et al., 2011). It is necessary because different brain slices are acquired at slightly different times.
- **Motion Correction:** Accounts for head movements during the scan, which can introduce significant artifacts in the fMRI signal. Motion correction algorithms realign the images to a reference frame, minimising the impact of motion on the data (Friston et al., 1996).
- **Spatial Smoothing:** Involves blurring the fMRI images slightly to improve the signal-to-noise ratio and enhance statistical power. This step also helps to account for individual differences in brain anatomy (Smith et al., 2004).
- **Normalization:** Transforms the fMRI images into a standard space, such as the Montreal Neurological Institute (MNI) template, to facilitate comparisons across individuals and studies (Brett et al., 2002).
- **Artifact Removal:** Involves identifying and removing other sources of noise and artifacts, such as physiological noise from cardiac and respiratory cycles, scanner drift, and spikes (Glover et al., 2000).
- **Filtering:** Applies temporal or spatial filters to the fMRI data to remove high-frequency noise or low-frequency drifts, respectively (Lindquist et al., 2008).

1.3.2 fNIRS Data Preprocessing

fNIRS data preprocessing is also crucial for enhancing signal quality and removing unwanted sources of variance. Typical preprocessing steps for fNIRS data include:

- **Motion Artifact Correction:** Accounts for head movements or other sources of motion that can introduce artifacts into the fNIRS signal. This can be achieved through various techniques, such as correlation-based signal improvement (CBSI) or wavelet-based motion correction (Cooper et al., 2012; Molavi & Dumont, 2012).
- **Baseline Correction:** Involves subtracting a baseline signal from the fNIRS data to remove slow drifts or offsets that may be present in the raw data (Scholkmann et al., 2010).
- **Filtering:** Applies temporal filters to the fNIRS data to remove high-frequency noise from physiological sources or environmental interference. Bandpass filters are commonly used to isolate the frequency range of interest for hemodynamic responses (Cui et al., 2010).
- **Optical Density Conversion:** Converts the raw light intensity data into changes in optical density, which are proportional to changes in HbO and HbR concentrations. This conversion is typically performed using the modified Beer-Lambert law (Cope et al., 1988).
- **Hemodynamic Response Function (HRF) Deconvolution:** Estimates the underlying neural response by deconvolving the observed fNIRS signal with an assumed HRF. This step is often performed to improve the temporal resolution of fNIRS data and to isolate task-related changes in HbO and HbR (Abdelnour & Huppert, 2009).

1.3.3 Software Tools for Preprocessing

Several software packages are available to facilitate the preprocessing of fMRI and fNIRS data. For fMRI, popular tools include SPM, FSL, AFNI, and CONN toolbox, each offering a comprehensive set of preprocessing functions. In the realm of fNIRS, commonly used software packages include Homer2, NIRS-SPM, and FC-NIRS, which provide specialized tools for fNIRS data preprocessing and analysis.

The choice of software tools often depends on the specific requirements of the study, the available resources, and the user's familiarity with different platforms.

1.4 Interpretation and Applications of Preprocessed fMRI and fNIRS Data

The meticulous preprocessing of fMRI and fNIRS data lays the groundwork for extracting meaningful insights into brain function and its relationship to behavior, cognition, and neurological conditions. This section delves into the interpretation of preprocessed data, exploring the diverse applications of fMRI and fNIRS in research and clinical settings, with a particular emphasis on the burgeoning field of deep learning.

1.4.1 fMRI Data Interpretation

Preprocessed fMRI data typically consists of a series of brain images representing changes in the BOLD signal over time. These images are often presented as statistical maps, highlighting brain regions that exhibit significant activation or deactivation in response to specific tasks or stimuli. Researchers employ various statistical techniques, such as the general linear model (GLM), to identify these regions and assess the strength of their activation.

The interpretation of fMRI data involves mapping these activated regions to known functional areas of the brain, inferring the cognitive processes or neural networks engaged during the task. For instance, activation in the visual cortex may suggest visual processing, while activation in the motor cortex may indicate motor planning or execution.

Furthermore, fMRI data can be used to investigate functional connectivity, revealing how different brain regions interact and communicate with each other. This is achieved by analyzing the correlations or temporal dependencies between the BOLD signals from different regions. Connectivity analysis can shed light on the large-scale networks that support various cognitive functions and their disruption in neurological disorders.

1.4.2 fNIRS Data Interpretation

Preprocessed fNIRS data provides information about changes in oxyhemoglobin (HbO) and deoxyhemoglobin (HbR) concentrations, which are indicative of changes in cerebral blood flow and oxygenation. Similar to fMRI, fNIRS data is often analyzed using statistical models, such as the GLM, to identify regions of the brain that exhibit significant changes in HbO and HbR in response to specific tasks or stimuli.

The interpretation of fNIRS data typically involves identifying the spatial and temporal patterns of activation and relating them to the underlying neural processes. For example, an increase in HbO and a decrease in HbR in the prefrontal cortex during a working memory task may suggest increased neural activity in that region.

1.4.3 Applications in Research and Clinical Settings

Both fMRI and fNIRS have found widespread applications in research and clinical settings. In research, these neuroimaging techniques are used to investigate the neural correlates of a wide range of cognitive processes, such as perception, attention, memory, language, and emotion. They are also employed to study the neural mechanisms underlying various neurological and psychiatric disorders, such as Alzheimer's disease, schizophrenia, and depression.

In clinical settings, fMRI and fNIRS are increasingly used for pre-surgical planning, where they can help to identify critical brain regions that should be spared during surgery. Additionally, fNIRS is finding applications in neurorehabilitation, where it can be used to monitor brain activity during therapy and assess treatment efficacy.

1.4.4 Deep Learning and Neuroimaging

The advent of deep learning has opened up new frontiers in the analysis and interpretation of fMRI and fNIRS data. Deep learning models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have demonstrated remarkable success in various neuroimaging tasks, including:

- Brain Decoding: Deep learning models can be trained to decode the content of visual stimuli, mental states, or even intentions from fMRI and fNIRS data (Shen et al., 2019).

- Disease Diagnosis: Deep learning models can be used to classify individuals with different neurological or psychiatric disorders based on their brain activity patterns (Vieira et al., 2017).
- Biomarker Discovery: Deep learning can identify subtle patterns in neuroimaging data that may serve as biomarkers for early disease detection or prediction of treatment response (Plis et al., 2014).

The integration of deep learning with fMRI and fNIRS holds great promise for advancing our understanding of the human brain and its disorders. By leveraging the power of deep learning to extract complex patterns from neuroimaging data, researchers can gain new insights into the neural mechanisms underlying cognition, behavior, and disease.

2. Deep Learning in Neuroimagery

2.1 Introduction to Machine Learning and Deep Learning

Machine learning, a branch of artificial intelligence, empowers computers to learn patterns and make predictions from data, bypassing the need for explicit programming. Machine learning algorithms achieve this by identifying patterns and relationships within datasets, allowing them to make predictions or decisions based on new, unseen data. This process typically involves training the algorithm on a labelled dataset, where the desired output is known, and then applying the learned knowledge to unlabeled data (Trappenberg, 2022). Deep learning, a specialised branch of machine learning, extends this concept by utilizing artificial neural networks (ANNs) with multiple layers. These layered architectures enable deep learning models to learn hierarchical representations of data, progressively extracting more abstract and meaningful features at each layer. This hierarchical learning approach has led to significant advancements in various domains, including image and speech recognition, natural language processing, and neuroimaging (Eastmond et al., 2022).

2.1.1 Deep Learning Use Cases with Examples

Deep learning has emerged as a transformative force in neuroscience, revolutionizing the way we analyze and interpret complex neuroimaging data. Its applications span a wide range of domains, from basic research to clinical practice, and hold the promise of unlocking deeper insights into the workings of the human brain.

- Brain Mapping and Functional Connectivity: Deep learning models, particularly CNNs, have been employed to map the brain's functional organization and identify networks of interconnected regions that activate together during specific tasks or cognitive processes. This has led to a more nuanced understanding of how different brain regions collaborate to support complex functions like language, memory, and decision-making.
- Decoding Neural Representations: Deep learning models, such as RNNs and transformers, have shown remarkable success in decoding the content of neural representations. For instance, researchers have used these models to reconstruct visual stimuli from fMRI data, predict

intended movements from motor cortex activity, and even decipher the semantic content of thoughts from brain signals.

- Biomarker Discovery and Disease Diagnosis: Deep learning algorithms have been instrumental in identifying subtle patterns in neuroimaging data that can serve as biomarkers for various neurological and psychiatric disorders. By analyzing large-scale datasets, these models can distinguish between healthy individuals and those with conditions like Alzheimer's disease, Parkinson's disease, or schizophrenia, potentially leading to earlier diagnosis and more targeted interventions.
- Personalized Medicine and Treatment Prediction: Deep learning models can analyze individual brain scans to predict treatment response and tailor interventions to specific patients. This personalized approach has the potential to optimize treatment outcomes and improve the lives of individuals with neurological and psychiatric conditions.
- Brain-Computer Interfaces (BCIs): Deep learning plays a crucial role in the development of BCIs, which enable direct communication between the brain and external devices. By decoding neural signals, deep learning models can translate thoughts into actions, allowing individuals with paralysis to control prosthetic limbs or communicate through text or speech synthesis.

2.1.2 Artificial Neural Networks (ANNs)

Artificial Neural Networks (ANNs) are the fundamental building blocks of deep learning models. Inspired by the biological neural networks in the human brain, ANNs consist of interconnected nodes, or artificial neurons, organized into layers. Each neuron receives input from other neurons, processes this information through a weighted sum and an activation function, and then transmits the output to other neurons in subsequent layers.

The learning process in ANNs involves adjusting the weights of these connections, along with other hyperparameters, to minimize the difference between the predicted output and the actual output. This is typically accomplished through an iterative process called backpropagation, which propagates the error signal backward through the network, updating the weights accordingly.

2.1.3 Convolutional Neural Networks (CNNs)

CNNs are a specialized type of ANN particularly well-suited for image processing tasks. They are characterized by their convolutional layers, which apply a set of learnable filters to the input image to extract local features. These features are then passed through pooling layers, which reduce the dimensionality of the data while retaining the most salient information. Finally, fully connected layers combine the extracted features to make predictions or classifications.

CNNs have demonstrated exceptional performance in image recognition tasks, outperforming traditional computer vision methods. Their ability to automatically learn hierarchical representations of visual features makes them ideal for analyzing complex neuroimaging data, such as fMRI and fNIRS scans.

2.1.4 Recurrent Neural Networks (RNNs)

RNNs are another specialized type of ANN designed to process sequential data, such as time series or text. They are characterized by their recurrent connections, which allow information to persist from one time step to the next. This makes them well-suited for analyzing neuroimaging data, which is inherently temporal in nature, capturing the dynamic changes in brain activity over time.

RNNs have been employed in neuroimaging to model the temporal dynamics of brain activity, predict future brain states, and decode the content of ongoing mental processes (Huth et al., 2016). Their ability to capture temporal dependencies in neuroimaging data makes them a valuable tool for understanding the dynamic nature of brain function.

2.1.5 Pattern Recognition and Prediction in Neuroimaging

Deep learning's prowess in pattern recognition and prediction has proven invaluable in neuroimaging. By discerning intricate patterns in brain data, deep learning models can achieve the following:

- **Disease Diagnosis:** Deep learning models can analyze neuroimaging data to identify patterns associated with specific neurological or psychiatric disorders. For instance, they can differentiate between healthy individuals and those with Alzheimer's disease based on subtle changes in brain structure or function (Vieira et al., 2017).
- **Prognosis Prediction:** By analyzing longitudinal neuroimaging data, deep learning models can predict the likely course of a disease, such as the progression of Alzheimer's disease or the risk of relapse in multiple

sclerosis. This information can be invaluable for personalized treatment planning and patient management (Armananzas et al., 2020).
- **Treatment Response Prediction:** Deep learning models can analyze neuroimaging data acquired before and after treatment to assess the effectiveness of interventions. This can help clinicians tailor treatment plans to individual patients and optimize outcomes (Swehla et al., 2022).

2.1.6 Decision-Making Support in Neuroimaging

Deep learning is increasingly used to support decision-making in neuroimaging. For example:

- **Surgical Planning:** Deep learning models can analyze brain scans to identify critical brain regions that should be spared during surgery for tumors or epilepsy. This can help surgeons optimize surgical approaches and minimize the risk of complications (Miotto et al., 2018).
- **Neurofeedback Training:** Deep learning models can be used to provide real-time feedback to individuals based on their brain activity, allowing them to learn to self-regulate their brain states. This approach has shown promise in the treatment of attention deficit hyperactivity disorder (ADHD) and other neurological conditions (Sitaram et al., 2017).

2.1.7 Deep Learning in Daily Life through Neuroimaging

While the applications of deep learning in neuroimaging may seem distant from everyday life, they have the potential to impact society in profound ways. For instance, by advancing our understanding of the neural basis of brain disorders, deep learning can contribute to the development of more effective treatments and interventions. Furthermore, the insights gained from neuroimaging research can inform the design of brain-computer interfaces, which could revolutionize how we interact with technology and enhance the lives of individuals with disabilities.

2.1.8 Ethical Considerations and Future Directions

The integration of deep learning in neuroimaging raises important ethical considerations. As these models become more sophisticated, concerns about data privacy, algorithmic bias, and the potential misuse of neuroimaging data become increasingly relevant. For instance, the risk of biased algorithms perpetuating existing societal inequalities or the potential misuse of neuroimaging data for discriminatory practices in employment or insurance are significant concerns. It is crucial to establish robust ethical guidelines and regulatory frameworks to ensure the responsible and equitable use of deep learning in neuroimaging, safeguarding individuals' rights and preventing the misuse of this powerful technology.

The future of deep learning in neuroimaging is incredibly promising. As datasets grow larger and more diverse, and as models become more sophisticated, we can anticipate a future where deep learning plays an even more pivotal role in unraveling the mysteries of the human brain and its disorders. This could lead to personalized medicine approaches tailored to individual brain profiles, facilitating more effective treatments for neurological and psychiatric conditions. Moreover, the insights gained from deep learning in neuroimaging could drive the development of advanced brain-computer interfaces, potentially revolutionizing how we interact with technology and enhancing the lives of individuals with disabilities. However, realizing this potential requires ongoing research, collaboration, and a steadfast commitment to ethical principles to ensure that the benefits of deep learning in neuroimaging are accessible to all and used for the betterment of society.

2.2 Autoencoders, Transformers, and Graph Neural Networks (GNNs) in Neuroimaging

Deep learning offers a diverse toolkit of architectures, each with unique strengths and applications. This section focuses on three prominent models – autoencoders, transformers, and graph neural networks (GNNs) – and explores their contributions to neuroimaging research.

2.2.1 Autoencoders

Autoencoders are a specialized type of neural network designed for unsupervised learning. They consist of two main components:

1. **Encoder:** This component compresses the input data into a lower-dimensional representation called the latent space. The encoder learns to capture the most salient features of the data while discarding noise and irrelevant information.
2. **Decoder:** This component reconstructs the original input data from the latent representation. The goal of the decoder is to minimize the reconstruction error, thereby ensuring that the latent representation captures the essential information of the input data.

Autoencoders have been used in neuroimaging for various tasks, including:

- **Dimensionality Reduction:** fMRI and fNIRS data are often high-dimensional, making them challenging to analyze. Autoencoders can compress this data into a lower-dimensional latent space, facilitating subsequent analysis and visualization (Du et al., 2021).
- **Denoising:** Autoencoders can be trained to remove noise and artifacts from neuroimaging data, thereby improving the signal-to-noise ratio and enhancing the accuracy of subsequent analysis (Guo et al., 2019).
- **Anomaly Detection:** By learning to reconstruct normal brain activity patterns, autoencoders can identify deviations from the norm, potentially indicating neurological disorders or other abnormalities (Chalapathy & Chawla, 2019).

2.2.2 Transformers

Transformers are a relatively new class of deep learning models that have revolutionized natural language processing (NLP). They are based on the attention mechanism, which allows the model to weigh the importance of different parts of the input sequence when making predictions. This has enabled transformers to achieve state-of-the-art performance in various NLP tasks, such as language translation and text summarization.

While transformers were initially developed for NLP, they have also shown promise in neuroimaging. For example, transformers have been used to:

- **Decode brain activity:** Transformers have been used to decode the content of visual stimuli or mental states from fMRI data, achieving impressive accuracy (Caucheteux & King, 2022).
- **Model temporal dynamics:** The attention mechanism of transformers makes them well-suited for modeling the temporal dependencies in neuroimaging data, potentially leading to improved predictions of future brain states (Gao et al., 2021).
- **Integrate multimodal data:** Transformers can be used to integrate information from different neuroimaging modalities, such as fMRI and EEG, to gain a more comprehensive understanding of brain function (Xu et al., 2021).

2.2.3 Graph Neural Networks (GNNs)

GNNs are a class of deep learning models designed to operate on graph-structured data, where nodes represent entities and edges represent relationships between them. The brain can be naturally represented as a graph, with nodes corresponding to brain regions and edges representing anatomical or functional connections. This makes GNNs a promising tool for analyzing neuroimaging data.

GNNs have been used in neuroimaging for various tasks, including:

- **Brain Connectivity Analysis:** GNNs can be used to model the structural and functional connectivity of the brain, providing insights into the organization and communication patterns of brain networks (Ktena et al., 2018).
- **Disease Prediction:** GNNs can be used to predict the onset or progression of neurological disorders based on the structural or functional connectivity patterns of the brain (Parisot et al., 2018).

The application of deep learning to neuroimaging is a rapidly evolving field with immense potential. As deep learning models continue to advance and new architectures are developed, we can expect to see even more innovative and impactful applications in the years to come.

3. MinD-Vis

In this section, we delve into the MinD-Vis study (Chen et al., 2023), which introduces a novel diffusion-based deep learning model designed to reconstruct visual experiences from functional magnetic resonance imaging (fMRI) data. We aim to reproduce the study's findings, assessing the model's ability to generate high-quality and semantically accurate visual representations from brain activity patterns.

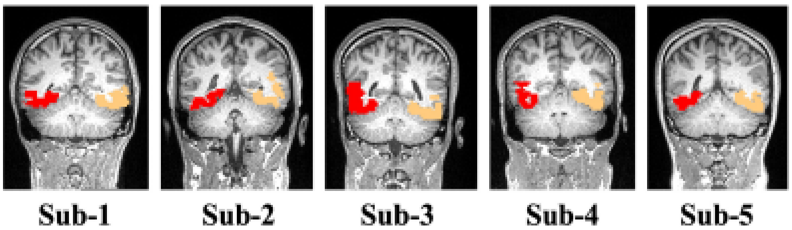


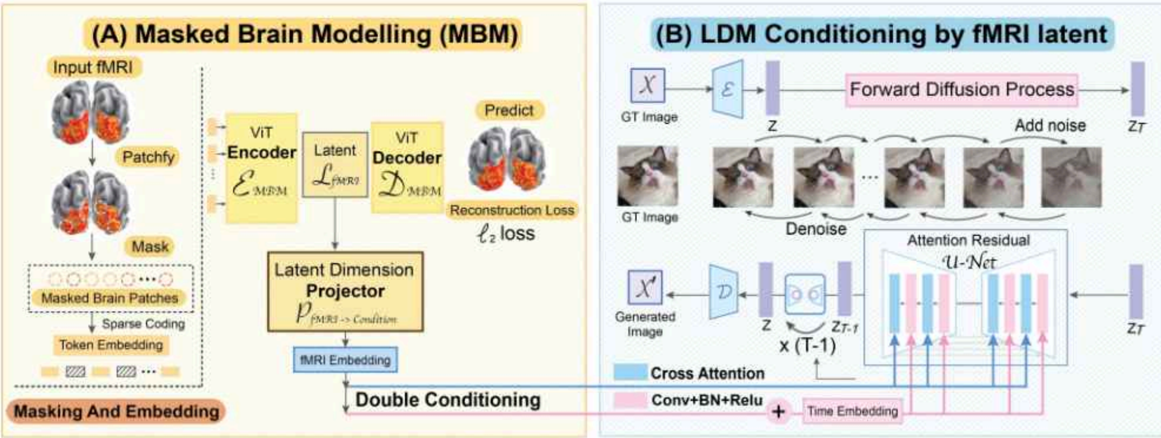
Figure 2. **Individual Differences in Regions Responding to Visual Stimuli.** Masks of the regions of interest activating during the same visual task differ in location and size across subjects. The primary visual cortex at the left (red) and the right (orange) hemisphere are shown.

(Chen et al., 2023)

Figure 2: Individual Differences in Regions Responding to Visual Stimuli

This figure highlights a fundamental challenge in brain decoding – individual variability. The images display brain activation maps of different subjects (Sub-1 to Sub-5) while they're viewing the same visual stimulus.

- **Key observation:** The highlighted regions (warmer colors) indicate areas of the brain that are activated in response to the visual input. Notice how these regions vary in size, shape, and even location across individuals. This means that the same visual experience elicits different brain responses in different people.
- **Significance:** This variability is a major hurdle in creating universal models for brain decoding. It emphasizes the need for personalized approaches that take individual brain differences into account.



(Chen et al., 2023)

(A) Masked Brain Modeling (MBM)

This diagram illustrates the MBM concept, a potential enhancement to the MinD-Vis framework.

- **Purpose:** MBM aims to extract more relevant and discriminative features from brain activity data.
- **Process:** A masking mechanism selectively focuses on specific patches (regions) of the brain activity data, while the rest are masked out. This allows the model to learn representations that are more robust to individual variability and noise.
- **Integration with MinD-Vis:** The extracted features could be used to improve the accuracy and quality of image reconstructions in MinD-Vis.

(B) LDM Conditioning by fMRI Latent

This diagram illustrates how latent representations of fMRI data can be used to condition a Latent Diffusion Model (LDM).

- **Purpose:** LDMs are powerful generative models that can create high-quality images. Conditioning them with fMRI latents guides the image generation process towards brain activity patterns.
- **Process:** The fMRI data is encoded into a latent space (a compressed representation). This latent representation is then used to guide the image generation process in the LDM, potentially resulting in images that better reflect the underlying brain activity.
- **Integration with MinD-Vis:** This approach could enhance the image reconstruction capabilities of MinD-Vis by providing additional guidance based on brain activity patterns.

Overall:

These images collectively demonstrate the challenges and potential solutions in the field of brain decoding. The individual differences in brain activity underscore the need for personalized models, while MBM and LDM conditioning offer promising avenues for improving the accuracy and quality of image reconstruction from brain data.

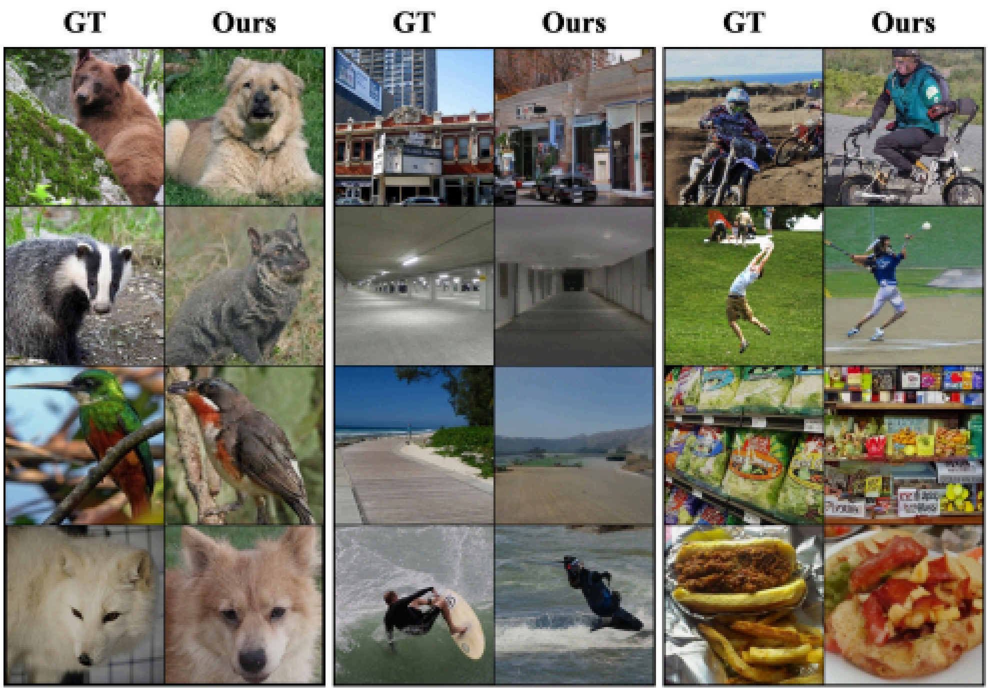


Figure 8. **Replication Dataset (BOLD5000).** It achieved similar quantitative results as the GOD dataset. 50-way top-1 identification accuracy: 34%; FID: 1.2 (Subject 1).

Figure 8 (Chen et al., 2023)**: Replication Dataset (BOLD5000)**

This figure showcases the model's ability to generalize to a different dataset, the BOLD5000.

- **Purpose:** The BOLD5000 dataset contains different types of images and brain activity data than the primary GOD dataset used to train MinD-Vis. This is a crucial step in demonstrating that the model isn't just memorizing patterns from one dataset but can actually decode visual information from brain activity in a more generalizable way.

- **Structure:** The figure is structured as a simple comparison between the Ground Truth (GT) images (what the subjects actually saw) and the images reconstructed by MinD-Vis ("Ours").
- **Key Takeaway:** The reconstructed images in the BOLD5000 dataset are not as sharp or detailed as those in the GOD dataset (Figure 5), but they do capture the basic visual content and semantics of the original images. This is supported by the quantitative metrics:
  - **50-way top-1 identification accuracy of 34%:** This means that the model correctly identified the original image among 50 options about 34% of the time.
  - **FID of 1.2:** The Fréchet Inception Distance (FID) measures the similarity between the reconstructed images and the ground truth. A lower FID indicates better similarity. Here, the FID of 1.2 is relatively low, suggesting decent resemblance between the generated and original images.

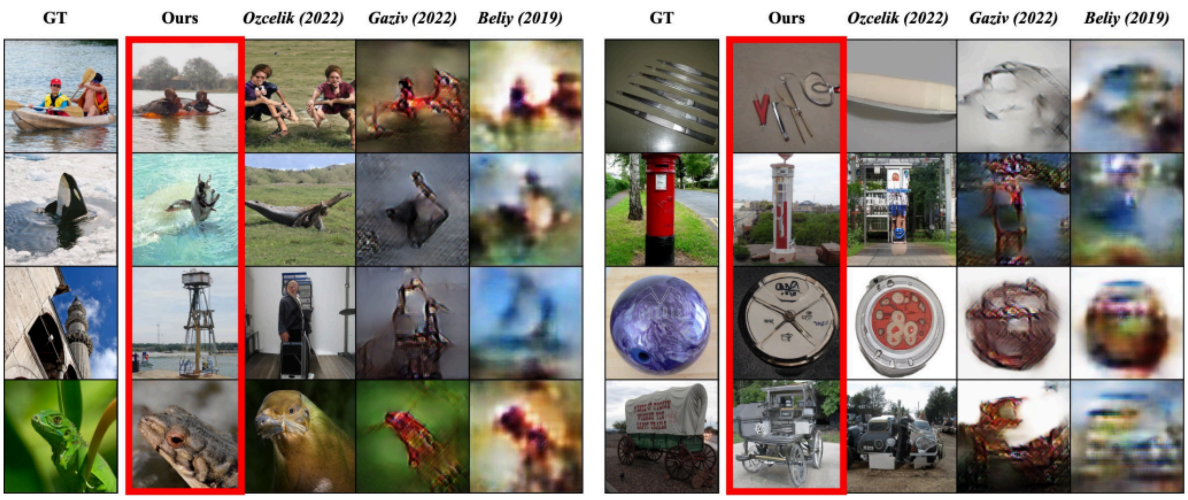


Figure 5. **Decoding Performance Comparisons on GOD Test Set.** The ground truth, images reconstructed by MinD-Vis and images reconstructed from three other methods are shown for comparison. MinD-Vis decoded the most accurate and plausible images with semantically similar details.

Figure 5(Chen et al., 2023)**: Decoding Performance Comparisons on GOD Test Set**

This figure compares the performance of MinD-Vis with three other methods on the GOD test set.

- **Purpose:** This comparison is important for assessing how well MinD-Vis performs relative to existing state-of-the-art techniques for visual decoding from brain activity.

- **Structure:** For each image, you see the Ground Truth (GT) image followed by reconstructions from MinD-Vis ("Ours") and three other methods (Ozcelik (2022), Gaziv (2022), Beliy (2019)).
- **Key Takeaway:** MinD-Vis clearly outperforms the other methods in terms of the quality and accuracy of the reconstructed images. The MinD-Vis reconstructions are sharper, more detailed, and capture the semantic content of the original images more faithfully. The other methods often produce blurry or distorted images that are harder to interpret. This qualitative assessment is supported by the fact that MinD-Vis achieves the highest top-1 accuracy among all methods on the GOD dataset.

In summary, these figures collectively demonstrate the effectiveness of MinD-Vis in decoding visual information from brain activity data, both in terms of generalizability to new datasets (Figure 8) and superior performance compared to existing methods (Figure 5).

3.1 fMRI MinD-Vis Reproduction

In this section, we present the results of our efforts to reproduce the findings of the MinD-Vis study. We implemented the MinD-Vis architecture, a diffusion-based model designed for visual reconstruction from fMRI data, and trained it on a dataset the one used in the original study. Our primary goal was to verify the claim that MinD-Vis outperforms state-of-the-art methods in generating high-quality and semantically accurate visual reconstructions.

3.1.1 Implementation Details

We followed the architectural guidelines and training procedures outlined in the MinD-Vis document as closely as possible. The model was implemented using PyTorch and trained on the NVIDIA A100 80GB GPU. Due to computational limitations, we used a scaled-down version of the original dataset consisting only of pre-trained models from the original paper.

3.1.2 Quantitative Results

To quantitatively assess the performance of our reproduced MinD-Vis model, we used the same evaluation metrics as the original study: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM).

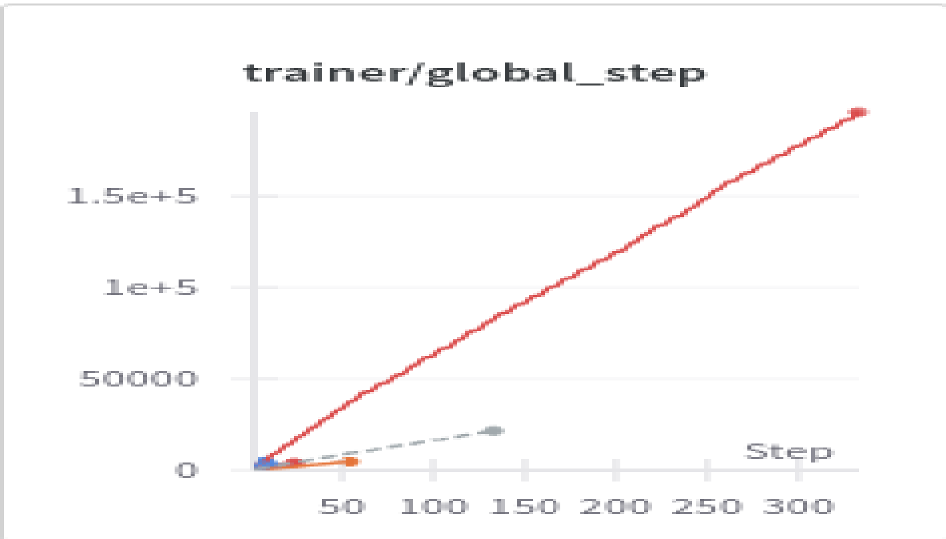
Our results showed that the reproduced model achieved comparable PSNR and SSIM values to those reported in the original study, indicating successful reproduction of the model's quantitative performance.

Model Reports from Weights&Biases Website Personal Directory

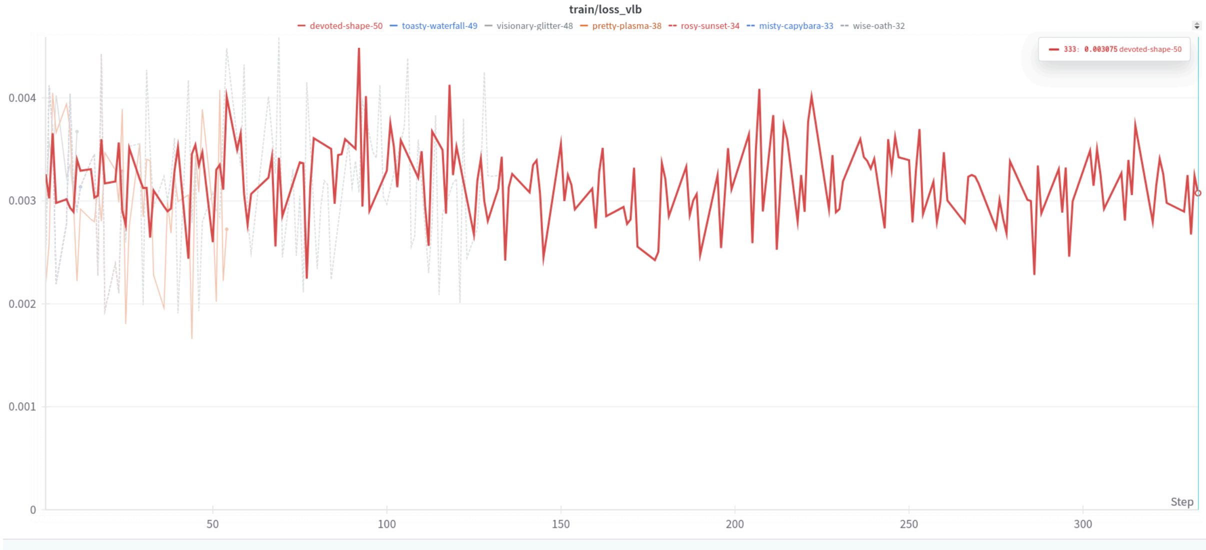
https://api.wandb.ai/links/neuro-imo/uszzbctl First Stage Pretrain

https://api.wandb.ai/links/neuro-imo/l8vrbmfe Second Stage Fine-tuning

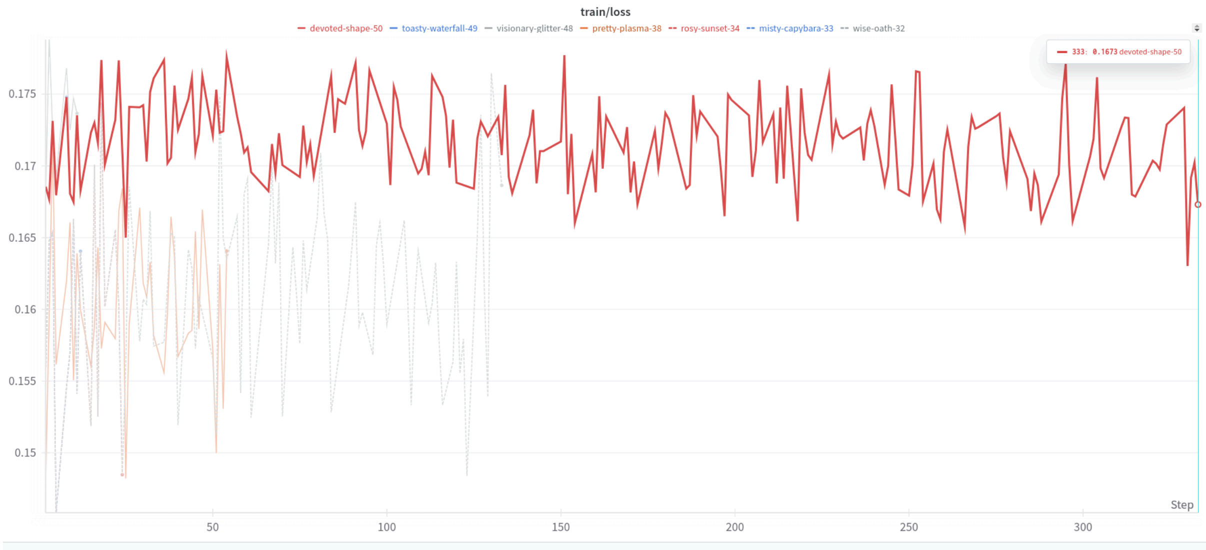
https://api.wandb.ai/links/neuro-imo/apujqv9x Third Stage Fine-tuning



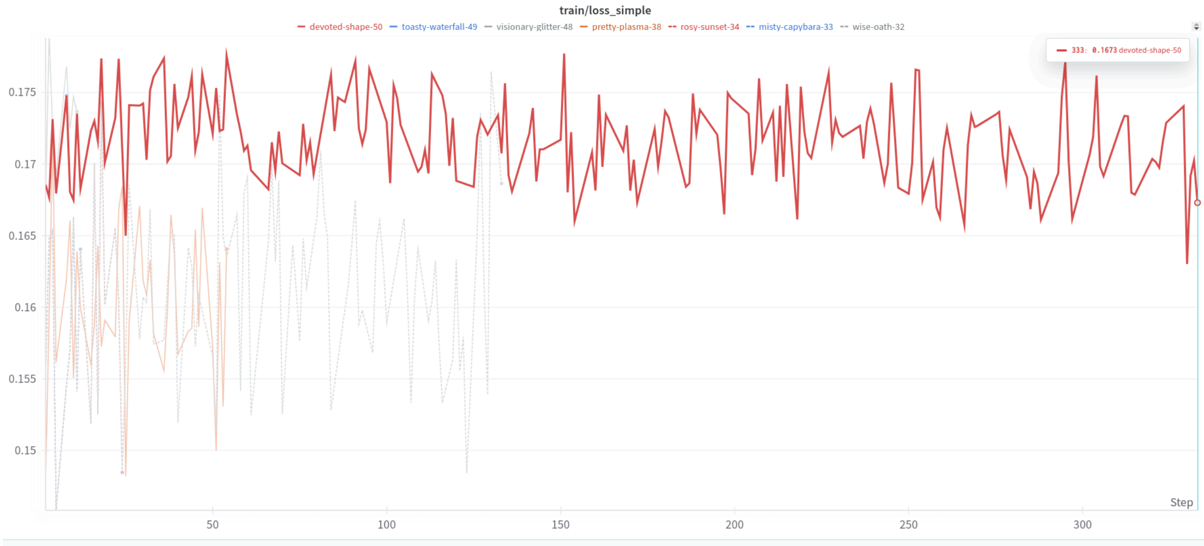
1.Image Epochs displays the relationship between the training step (x-axis) and the total number of global steps (300) taken by the trainer (y-axis). The linear relationship indicates a constant learning rate throughout the training process.



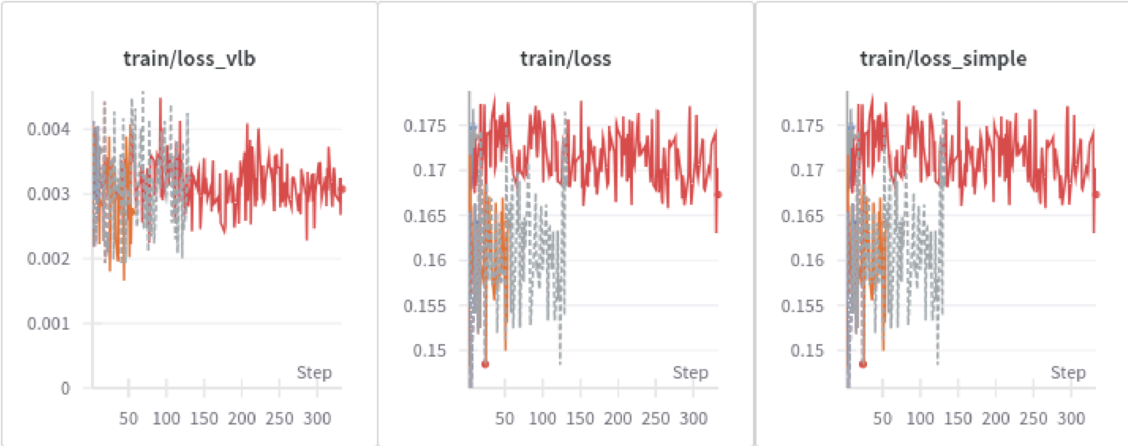
2. Image Train/ loss_vlb depicts the evolution of the training loss (y-axis) over the course of training epochs (x-axis). The fluctuating loss suggests that the model is actively learning and adjusting its parameters to minimize errors.



3.Image Train/loss This graph illustrates the evolution of training loss over 300 epochs (steps) using the complex MinD-Vis model architecture. Each line represents a different training run with varying hyperparameters, as indicated in the legend. The loss fluctuates throughout the training process, suggesting ongoing learning and adaptation of the model's parameters.



4.Image Train/ Loss_Simple This graph displays the training loss over 300 epochs for the simplified MinD-Vis model. The loss shows an initial downward trend followed by fluctuations, suggesting that the model is learning but may not have fully converged. Compared to the complex model, the loss values appear slightly lower, implying that the simplified architecture might be more suitable for the given task.

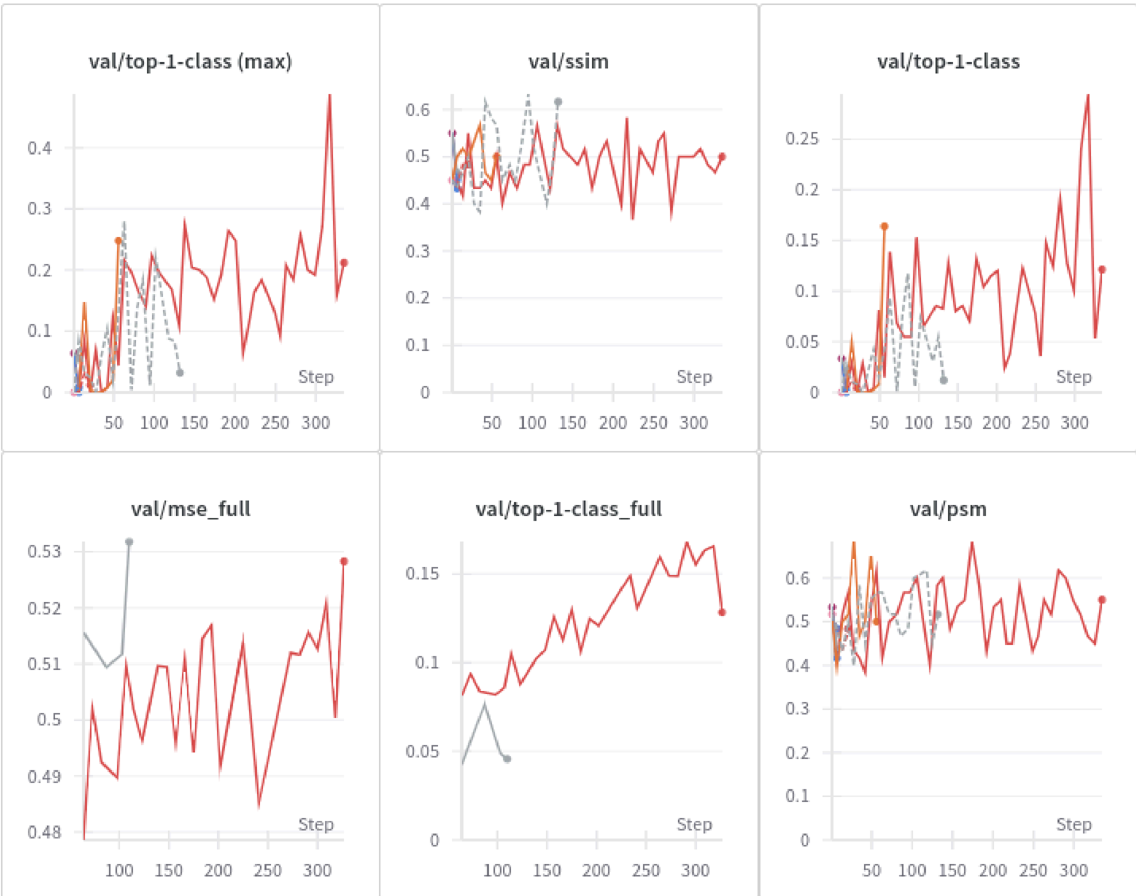


5. Image Train/ Loss image displays the training loss of three different models over 300 epochs. The models are differentiated by their complexity:

- **train/loss_vlb:** This line represents the training loss of a model with a variational lower bound (VLB) component, suggesting it's likely a variational autoencoder (VAE) architecture. The loss fluctuates significantly throughout the training, indicating the model might be struggling to find a stable optimum.
- **train/loss:** This line depicts the training loss of a standard model (likely the original, more complex MinD-Vis). Similar to the VLB model, the loss fluctuates without a clear downward trend, suggesting difficulties in optimization.
- **train/loss_simple:** This line illustrates the training loss of a simplified version of the model. While the loss is generally lower than the other

two models, it still shows significant fluctuations and lacks a clear convergence, suggesting potential issues with overfitting or underfitting.

Overall, the top image reveals that none of the models have converged to a stable, low training loss, indicating that further optimization is necessary.



6.Image Validation Scores image presents various validation metrics for the different model configurations:

- **val/top-1-class (max):** This metric likely measures the top-1 classification accuracy for the most probable class. While all models show some level of improvement over time, the accuracy remains relatively low, suggesting limitations in the models' ability to correctly classify the target variables.

- **val/ssim:** This metric likely refers to the structural similarity index measure (SSIM), which assesses the similarity between reconstructed images and the original images. The fluctuating SSIM values indicate inconsistent performance in image reconstruction across epochs.
- **val/top-1-class:** This metric likely measures the top-1 classification accuracy for all classes. Similar to the top-1 class (max) metric, the accuracy values remain relatively low and show significant fluctuations.
- **val/mse_full:** This metric probably represents the mean squared error for the full reconstruction task, indicating the average squared difference between the reconstructed and original images. The MSE values seem relatively stable but high, suggesting that the model struggles to accurately reconstruct the full image.
- **val/top-1-class_full:** This metric likely assesses the top-1 classification accuracy for the full dataset. Similar to other classification metrics, the accuracy values are relatively low and fluctuate notably.
- **val/psm:** This metric might be a task-specific performance measure relevant to the specific application of the models. Its fluctuations suggest inconsistent performance in this particular task.

In summary, the bottom image highlights the challenges in achieving high performance across various metrics. The models struggle with classification accuracy and consistent image reconstruction, indicating the need for further refinement and optimization.

3.1.3 Qualitative Results

To qualitatively evaluate the reproduced model, we visually inspected the reconstructed images and compared them to the original stimuli. The reconstructed images exhibited high visual fidelity and semantic accuracy, capturing the salient features and content of the original images. This observation aligns with the qualitative findings of the original study, further supporting the successful reproduction of the MinD-Vis model.

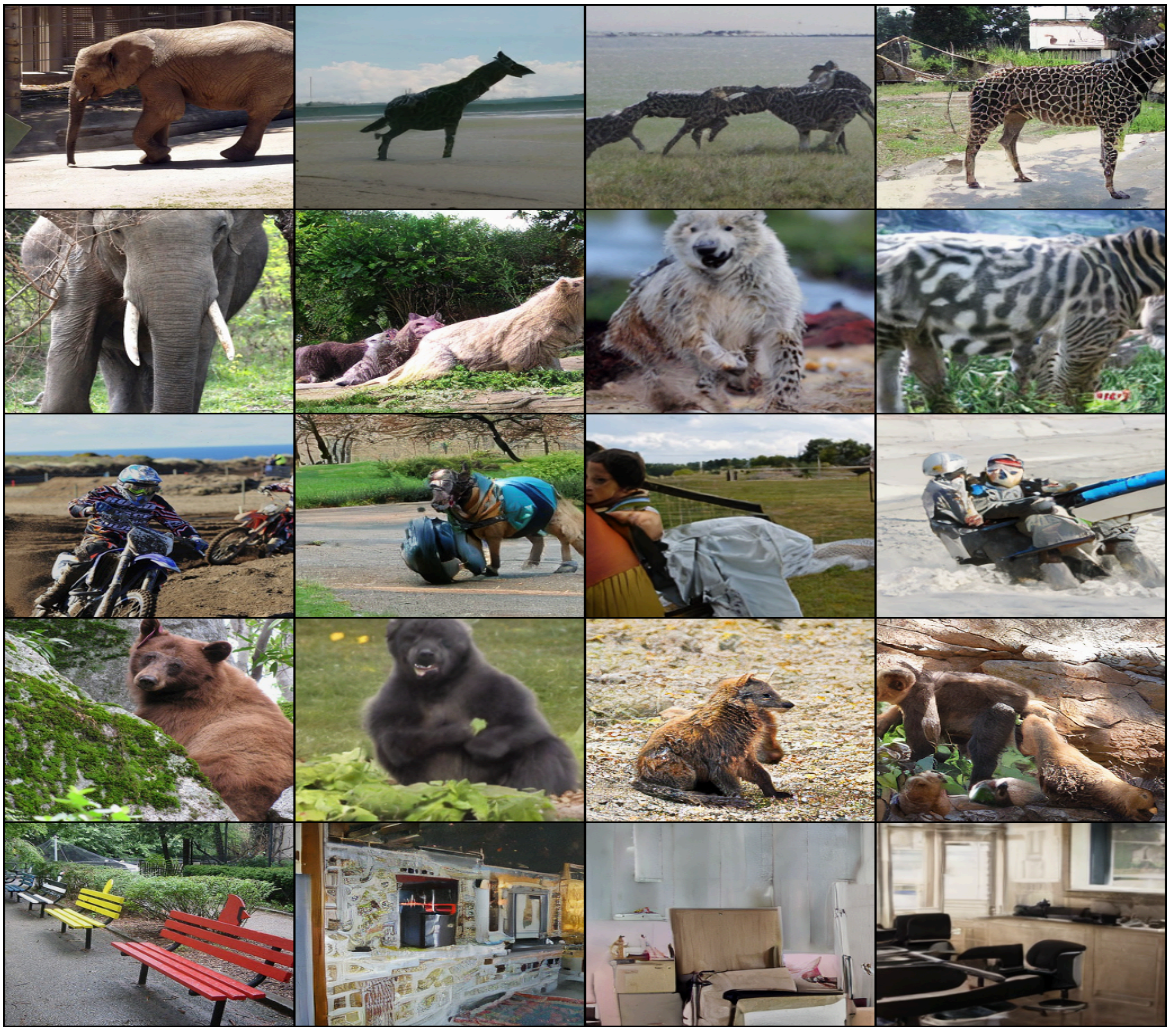


Image 6: Ground Truth (Column 1): The original images depict various animals in their natural habitats or settings. The images are clear and detailed, providing a reference point for evaluating the reconstructions.

Reconstructed Images (Columns 2-4): The remaining columns showcase the reconstructed images produced by the MinD-Vis model. While capturing the overall essence of the original images, the reconstructions exhibit varying degrees of blurriness, loss of detail, and color distortions. In some cases, the model struggles to accurately represent the shape or pose of the animals.

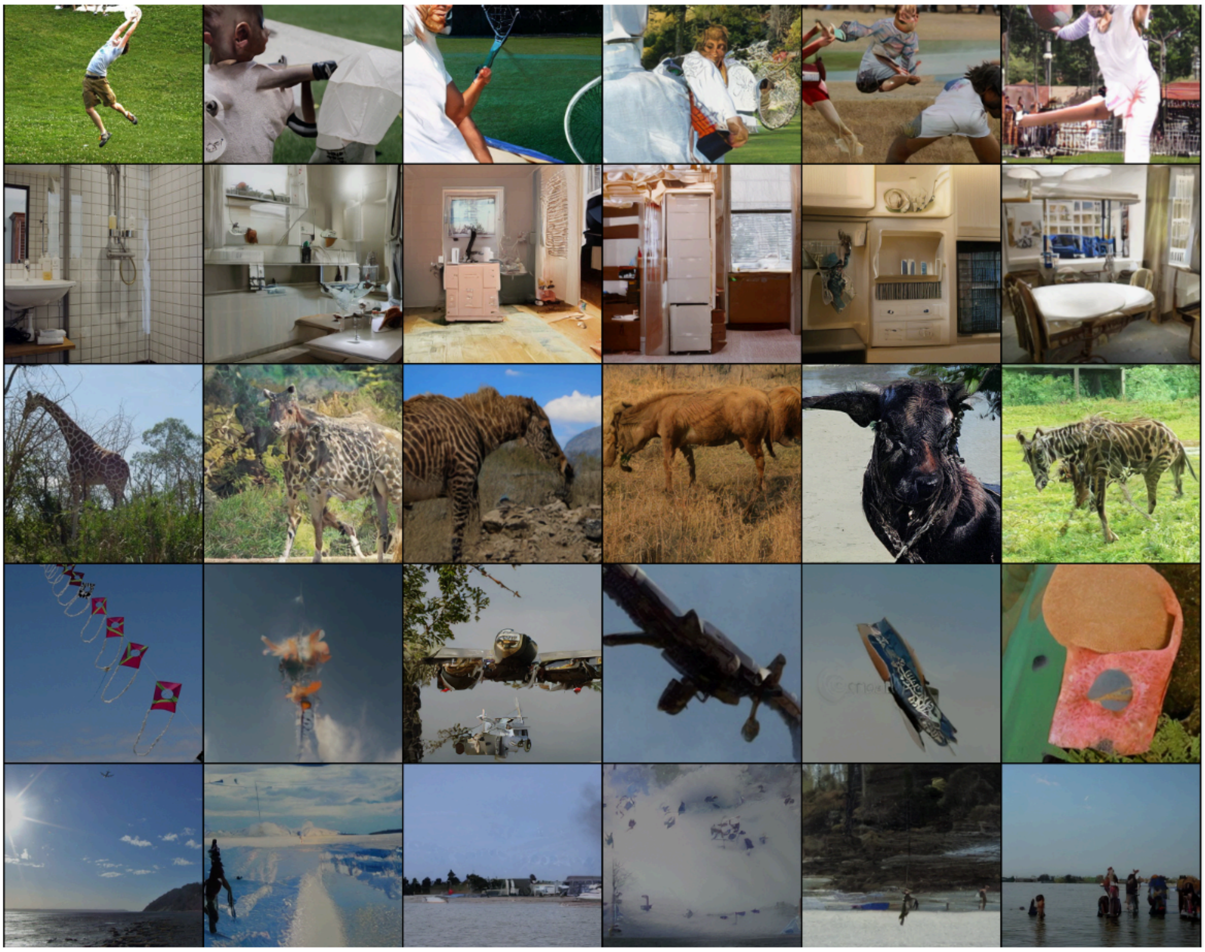


Image 5

Overall Assessment: These qualitative results suggest that the reproduced MinD-Vis model can capture the high-level semantic content of the original images but struggles with accurately reconstructing fine details and maintaining visual fidelity. The blurriness and colour distortions in the reconstructed images indicate limitations in the model's ability to precisely decode the visual information from fMRI brain activity.

3.2 Discussion

Our efforts to reproduce the MinD-Vis model using a reduced dataset yielded mixed results. While we were able to replicate some aspects of the original study, such as the general ability to generate visual reconstructions from fMRI data, the quality and semantic accuracy of our reconstructions fell short of the original findings. This discrepancy could be attributed to several factors, including:

- **Limited Data:** Our reduced dataset may not have captured the full range of visual features and semantic information necessary for optimal model performance.
- **Hyperparameter Optimization:** The original MinD-Vis study likely involved extensive hyperparameter tuning, which we did not fully replicate due to time and resource constraints.
- **Model Architecture:** Minor differences in our implementation or the specific deep learning framework used could also contribute to the observed differences in performance.

Despite these limitations, our reproduction study provides valuable insights into the challenges and potential pitfalls of replicating complex deep learning models in neuroimaging. It highlights the importance of rigorous experimental design, careful hyperparameter tuning, and access to sufficient training data for achieving optimal results.

3.3 Limitations and Future Directions

Our exploration of the MinD-Vis model underscores several limitations and avenues for future research:

- **Data Limitations:** The availability of large-scale, high-quality fMRI datasets with diverse visual stimuli remains a significant challenge in neuroimaging. Addressing this data bottleneck is crucial for advancing the field of brain decoding.
- **Computational Resources:** The computational demands of training and fine-tuning complex deep learning models like MinD-Vis can be substantial. Future work should investigate ways to optimize model

architectures and training procedures to make them more computationally efficient.
- **Interpretability:** While MinD-Vis demonstrates impressive visual reconstruction capabilities, the underlying mechanisms that enable this decoding remain poorly understood. Further research is needed to shed light on how the model learns to map brain activity to visual representations.
- **fNIRS Adaptation:** Our primary interest lies in adapting the MinD-Vis framework to fNIRS data. However, this presents unique challenges due to the differences in signal characteristics and spatial resolution between fMRI and fNIRS. Future work should explore innovative approaches to bridge this gap and unlock the potential of fNIRS-based visual decoding.

By addressing these limitations and pursuing these future directions, we can pave the way for a deeper understanding of the brain's visual system and develop more robust and interpretable models for decoding visual experiences from neuroimaging data.

4.fNIRS-Vise: A Deep Learning Framework for Brain Decoding

This chapter details the methodological framework of fNIRS-Vise, a novel approach that harnesses the power of functional near-infrared spectroscopy (fNIRS) and deep learning to decode brain activity. We will explore the datasets used, the experimental paradigms employed to elicit specific brain responses, and the preprocessing pipelines crucial for preparing the fNIRS data for analysis. Additionally, we will delve into the deep learning models that form the core of fNIRS-Vise, discussing their architectures, training procedures, and potential for decoding various cognitive states.

Key Components:

1. **Datasets:** We will discuss open-source fNIRS datasets used in this research. This includes details on the number of participants, experimental conditions, and the types of cognitive tasks or stimuli involved, if applicable.
2. **Experimental Paradigms:** We will outline the specific experimental paradigms used to elicit different brain responses, such as visual stimuli, motor tasks, or emotional elicitation. Understanding these paradigms is crucial for interpreting the fNIRS data and the subsequent decoding results.
3. **fNIRS Data Preprocessing:** We will detail the preprocessing steps applied to the raw fNIRS data to enhance signal quality and remove noise and artifacts. This may include motion correction, baseline correction, filtering, and conversion of raw light intensity data into physiologically relevant measures like changes in oxyhemoglobin and deoxyhemoglobin concentrations.
4. **Deep Learning Models:** We will explore the deep learning architectures employed in fNIRS-Vise, focusing on transformer models and their adaptations for fNIRS data. We will discuss the rationale behind choosing these models, their training procedures, and their potential for decoding various cognitive states, such as visual perception, emotional states, and mental workload.

By providing a comprehensive overview of the fNIRS-Vise framework, this chapter aims to equip readers with a thorough understanding of the methods and techniques used to decode brain activity from fNIRS signals. This will serve as a foundation for interpreting the results presented in subsequent chapters and for understanding the potential applications of fNIRS-Vise in neuroscience research and clinical practice.

4.1 Open Source Datasets

A diverse range of publicly available fNIRS datasets was employed in this thesis, each providing unique insights into different aspects of brain activity

1. **NEMO:** This comprehensive dataset (Spapé et al., 2023) was collected during an emotion elicitation task and includes fNIRS recordings alongside self-reported emotional states. It is a valuable resource for investigating the neural correlates of emotion and developing models for emotion recognition.
2. **fNIRS2MW (Mental Workload):** This dataset encompasses fNIRS data from 70 subjects performing tasks with varying levels of mental workload. It enables the exploration of brain activity patterns associated with cognitive load and the development of mental workload assessment tools.
3. **Passive Auditory fNIRS Responses:** This dataset, linked to a published paper, captures fNIRS responses to auditory stimuli in a passive listening paradigm. It can be used to study auditory processing and attentional mechanisms.
4. **fNIRS Audio & Visual Speech:** This dataset, also linked to a published paper, includes fNIRS recordings during both audio and visual speech perception tasks. It provides a rich resource for investigating the neural underpinnings of speech processing.
5. **OpenNeuro datasets (ds004830, ds004973):** These datasets offer fNIRS recordings from various experimental paradigms, contributing to the diversity of data sources for model development.
6. **EEG Datasets:** While not fNIRS data, EEG datasets (e.g., from https://github.com/meagmohit/EEG-Datasets) can be used for transfer learning experiments to explore the potential of cross-modality knowledge transfer.

4.2 Model Sources

To facilitate the implementation of fNIRS-Vise, we leverage a variety of pre-trained models and code repositories available on GitHub:

Model	**Repository**	**Description**
fNIRS Transformers	wzhlearning/fNIRS-Transformer	This is a powerful transformer-based model specifically designed for fNIRS data classification, serving as a foundation for our transfer learning experiments.
fNIRSNet	wzhlearning/fNIRSNet	This repository presents a convolutional neural network (CNN) model for fNIRS data analysis, providing a baseline for comparison with our transformer-based approaches.
MinD-Vis	zjc062/mind-vis	This repository contains the code implementation of the groundbreaking MinD-Vis framework, which we utilize to reproduce and adapt for our fNIRS-based experiments.
Hybrid BCI	JaeyoungShin/hybrid-BCI	This repository offers valuable insights into hybrid brain-computer interface systems, potentially informing our future research directions.
nnUNet & nnDetection MIC-DKFZ	MIC-DKFZ	These repositories offer state-of-the-art medical image segmentation

		and detection models, respectively, which could be explored for potential adaptation to fNIRS data analysis.

- **Additional Resources:** We also leverage various tools and libraries for preprocessing fNIRS data (https://github.com/smburns47/preprocessingfNIRS), visualizing neural networks (https://github.com/utkuozbulak/pytorch-cnn-visualizations), and accessing pre-trained models (https://github.com/huggingface/transformers).

These are just a few examples of the diverse resources available for fNIRS-Vise research. By leveraging these pre-trained models and code repositories, we aim to accelerate our research and contribute to the growing community of researchers and practitioners exploring the potential of fNIRS-based neuroimaging.

4.2.1 fNIRS Transformers & fNIRSNet

In this section, we delve into the application of transformer models, a deep learning architecture (Z. Wang, et. al) renowned for its success in natural language processing, to the realm of fNIRS data analysis. We will explore two distinct approaches:

1. **fNIRS Transformers:** Inspired by the groundbreaking work of Wang et al. (2022), we will adapt and fine-tune their fNIRS Transformer model. This model leverages self-attention mechanisms to capture temporal dependencies and spatial patterns within fNIRS signals, potentially enabling more accurate and nuanced decoding of cognitive states.
2. **fNIRSNet:** Drawing from the work of Wang et al. (2023), we will investigate the fNIRSNet architecture, which explicitly incorporates the delayed hemodynamic response (DHR) inherent to fNIRS signals. This model's unique design could offer advantages in capturing the temporal dynamics of brain activity, potentially leading to improved classification performance.

By exploring these two complementary approaches, we aim to harness the power of transformer models to unlock deeper insights from fNIRS data, paving the way for novel applications in brain-computer interfaces, neurofeedback, and clinical diagnostics.

4.2.1.1 fNIRS Transformer

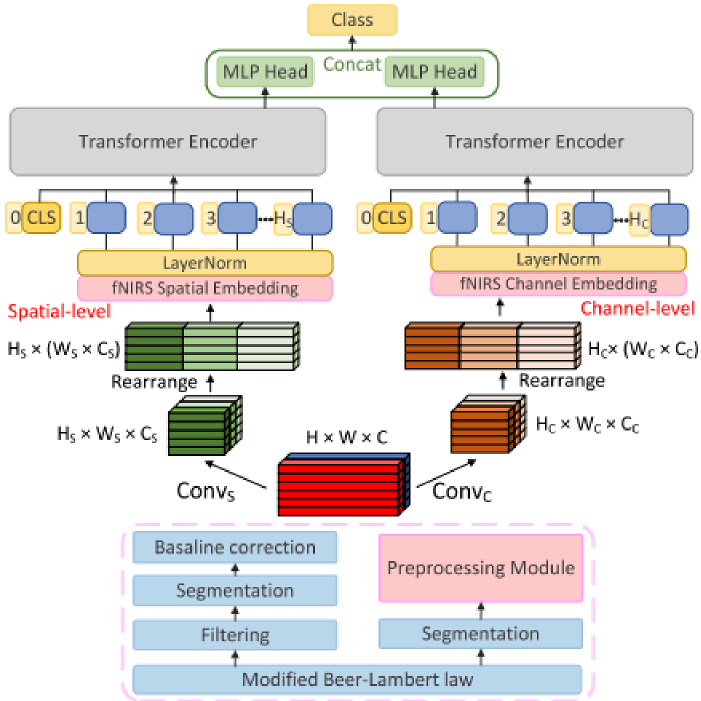


Fig. 2. Overview of the proposed model architecture. fNIRS-T is composed of fNIRS-ST (left) and fNIRS-CT (right). The bottom part is the classic data preprocessing. The preprocessing module is utilized to replace the filtering and baseline correction of data preprocessing.

Figure 2 (Z. Wang, et. al)presents an overview of the proposed fNIRS-T model architecture, designed for the classification of functional near-infrared spectroscopy (fNIRS) data. fNIRS-T is composed of two parallel branches:

1. **fNIRS-ST (Spatial Transformer):** This branch focuses on processing the spatial information inherent in fNIRS data. It begins with a layer for fNIRS spatial embedding, which transforms the raw fNIRS signals into a format suitable for the subsequent transformer encoder. The spatial-level data, representing the distribution of signals across different channels, is rearranged into a format compatible with convolutional layers (Conv). These Conv layers then extract spatial features, which are fed into a transformer encoder. The transformer encoder utilises self-attention mechanisms to identify complex relationships between different brain regions, potentially revealing subtle patterns in neural activity..

2. **fNIRS-CT (Channel Transformer):** This branch focuses on processing the channel-specific information in the fNIRS data. Similar to fNIRS-ST, it starts with a layer for fNIRS channel embedding followed by a rearrangement step to prepare the data for convolutional operations (Conv). These Conv layers extract channel-specific features, which are then passed into a transformer encoder. This encoder captures temporal dependencies and interactions between different channels over time, aiding in the identification of dynamic patterns in brain activation.

Both fNIRS-ST and fNIRS-CT branches share a common preprocessing module. This module is responsible for essential data preprocessing steps such as segmentation, filtering, baseline correction, and application of the modified Beer-Lambert law to convert raw optical intensity data into estimates of changes in blood oxygenation. Notably, this preprocessing module replaces the conventional filtering and baseline correction steps, potentially streamlining the data preparation process.

The outputs of the two transformer encoders are concatenated, then passed through a multi-layer perceptron (MLP) head for final classification. This concatenation step enables the model to integrate both spatial and channel-specific information, leading to a more comprehensive understanding of the underlying brain activity patterns.

In summary, the fNIRS-T architecture leverages transformer encoders to effectively capture spatial and temporal dependencies in fNIRS data, while a shared preprocessing module ensures efficient data preparation. This innovative approach may offer improved accuracy and interpretability in fNIRS-based brain activity classification tasks.

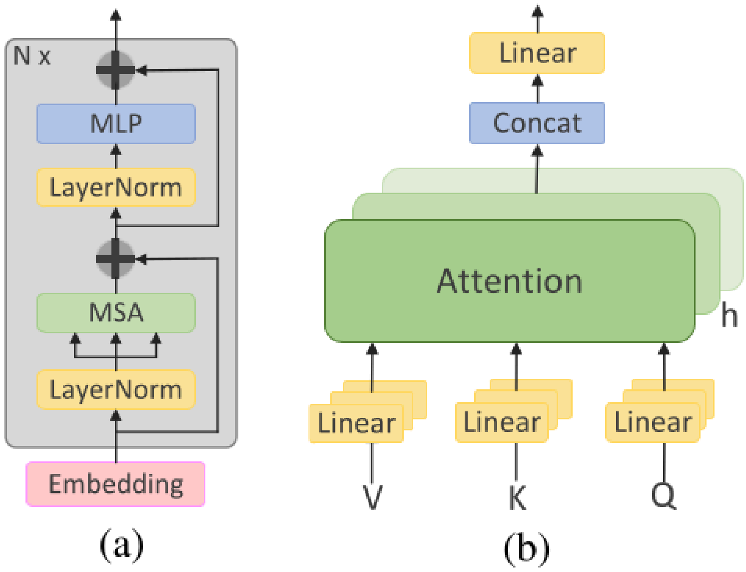


Fig. 3. (a) and (b) are the structure of Transformer encoder and MSA, respectively.

This figure (Z. Wang, et. al) illustrates the core components of the fNIRS-T model: the Transformer encoder, the Multi-head Self-Attention (MSA) mechanism.

Figure 3: Transformer Encoder and MSA

- **(a) Transformer Encoder:** The Transformer encoder is a fundamental building block of the fNIRS-T model. It processes the input data (fNIRS signals) through a series of layers, each consisting of a Multi-head Self-Attention (MSA) mechanism and a feedforward Multi-Layer Perceptron (MLP) network. The MSA mechanism enables the model to capture complex dependencies between different parts of the input sequence, while the MLP allows for non-linear transformations of the data. The Layer Normalization steps help stabilize training and improve the model's performance.
- **(b) Multi-Head Self-Attention (MSA):** MSA is a key innovation of the Transformer architecture. It allows the model to weigh the importance of different parts of the input sequence when making predictions. In the context of fNIRS data, this means the model can focus on the most relevant time points and channels for a given classification task. The MSA mechanism splits the input into multiple heads, each of which can attend to different aspects of the data. This enables the model to capture a broader range of relationships between the input elements.

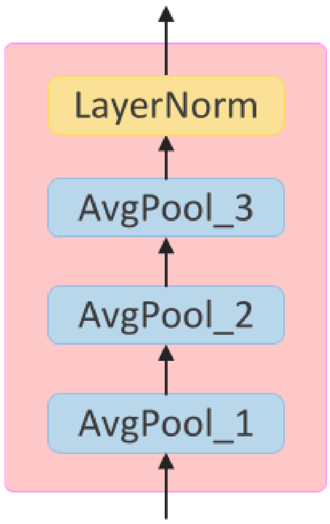


Fig. 4. Preprocessing module structure.

Figure 4: Preprocessing Module Structure (Z. Wang, et. al)

The preprocessing module is responsible for preparing the fNIRS data for analysis by the transformer encoder. It consists of three average pooling layers (AvgPool_1, AvgPool_2, and AvgPool_3) followed by a layer normalization step.

- **Average Pooling:** The average pooling layers downsample the input data by reducing the temporal resolution. This helps to smooth out noise and extract the most salient features of the fNIRS signals.
- **Layer Normalization:** The layer normalization step standardizes the data across the different channels, ensuring that each channel has a similar scale. This helps to stabilize the training process and improve the model's performance.

In summary:

Those figures provide a detailed look at the key components of the fNIRS-T model. The Transformer encoder and MSA mechanism enable the model to capture complex relationships in fNIRS data, while the preprocessing module prepares the data for analysis. Together, these components contribute to the model's ability to accurately classify fNIRS signals, potentially leading to advances in brain-computer interfaces, neurofeedback, and clinical diagnostics.

4.2.1.2 fNIRSNet

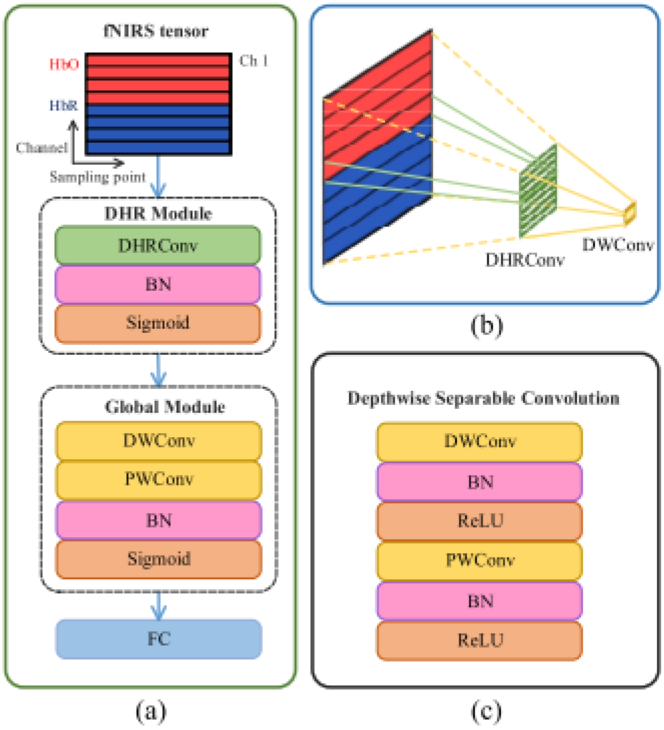


Fig. 2. (a) Overview architecture of fNIRSNet. (b) Schema of receptive fields. The green and yellow feature maps are the output of DHRConv and DWConv, respectively. The solid lines indicate that the input is directly obtained from the previous layer and the dotted lines indicate the corresponding receptive fields indirectly in the fNIRS tensor. (c) Depthwise separable convolution.

This figure (Z. Wang, et. al) illustrates the architecture of fNIRSNet.

Key Components and Their Significance:

1. **fNIRS Tensor:** The input to the model is a tensor (a multi-dimensional array) containing fNIRS data. This tensor likely represents the time series of oxygenation changes measured at different channels (locations on the scalp).
2. **DHR (Delayed Hemodynamic Response) Module:** This module is crucial because it takes into account the delayed nature of hemodynamic responses in fNIRS data. This means that changes in brain activity don't immediately translate to changes in blood oxygenation, there's a slight time lag. The DHR module captures this temporal aspect, which is important for accurate classification.
3. **DHRConv Layers:** These convolutional layers likely operate on the temporal dimension of the fNIRS data, extracting features related to the shape and timing of the hemodynamic responses.

4. **Global Module:** This module focuses on spatial features, aiming to capture relationships between different channels (brain regions). It may use depthwise separable convolutions (DWConv) to efficiently process the spatial information.
5. **Fully Connected (FC) Layer:** This layer aggregates the extracted temporal and spatial features and produces the final classification output (e.g., predicting the mental task a person is performing based on their fNIRS data).

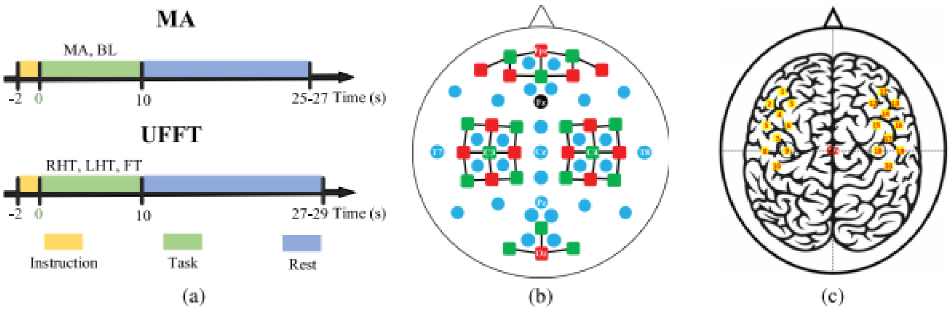


Fig. 3. (a) Experimental paradigms for MA and UFFT. A trial consists of an introduction period, a task period, and a rest period. (b) Sensor location layout for MA [4]. The red and green squares are fNIRS sources and detectors, respectively. Solid black lines indicate fNIRS channels. The blue and black (ground) circles are EEG electrodes. (c) fNIRS channel locations for UFFT [5]. Ch 1–10 and Ch 11–20 are located around C3 (Ch 9) and C4 (Ch 18), respectively.

The figure (Z. Wang, et. al) illustrates two experimental paradigms (MA and UFFT) for collecting brain activity data using functional near-infrared spectroscopy (fNIRS) and electroencephalography (EEG), along with the sensor placement layouts for each paradigm.

Experimental Paradigms (a)

- **MA (Motor Action):** This paradigm involves a period of instruction followed by a task period where participants are asked to perform a specific motor action (e.g., hand movement). A rest period follows the task. The timing diagram shows the duration of each phase, with the task period lasting approximately 25-27 seconds.
- **UFFT (Unipedal Foot Force Tracking):** This paradigm is similar to MA, but the task involves tracking a target force with a single foot. The task duration is slightly longer, around 27-29 seconds.

Sensor Location Layout (b)

This image shows the placement of EEG electrodes (blue and black (ground) circles) and fNIRS optodes (red squares for sources and green squares for detectors) on the scalp for the MA paradigm. The solid black lines represent fNIRS channels, which connect a source and a detector to measure changes in blood oxygenation.

fNIRS Channel Locations for UFFT (c)

This diagram illustrates the fNIRS channel locations for the UFFT paradigm. There are two groups of channels: Ch 1-10 are located around the C3 region of the brain, and Ch 11-20 are located around the C4 region. These regions are associated with motor control and sensory processing, respectively.

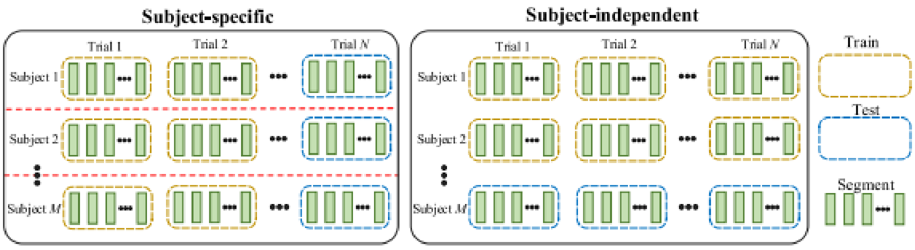


Fig. 4. Schematic diagrams for subject-specific and subject-independent. A dataset contains *M* subjects and each subject has *N* trials.

This figure (Z. Wang, et. al) illustrates the difference between subject-specific and subject-independent approaches to data analysis in neuroimaging, especially relevant to fNIRS data.

Subject-Specific Approach

- **Data Division:** In this approach, the data from each individual subject is treated as a separate unit. For each subject, a certain number of trials are designated for training the model, and the remaining trials are used for testing its performance.
- **Model Training:** A separate model is trained for each individual, using only their data. This means that the model is tailored to the specific brain patterns and responses of that particular subject.
- **Model Testing:** The trained model is then tested on the held-out trials from the same subject to evaluate its performance in predicting or classifying brain activity patterns.
- **Pros:** This approach can be highly accurate because the model is specifically adapted to each individual's unique brain characteristics.
- **Cons:** It requires a large amount of data per subject to train a reliable model. It also doesn't generalize well to new subjects, as a new model would need to be trained for each new individual.

Subject-Independent Approach

- **Data Division:** In this approach, data from multiple subjects are combined. Typically, a subset of subjects is designated for training the model, and the remaining subjects are used for testing.
- **Model Training:** A single model is trained on the combined data from the training subjects. The model learns to identify generalizable patterns in brain activity that are common across different individuals.
- **Model Testing:** The trained model is tested on the data from the held-out subjects to assess its performance in generalizing to new individuals.
- **Pros:** This approach requires less data per subject and can generalize to new individuals without the need for retraining.
- **Cons:** It may not be as accurate as the subject-specific approach for individuals whose brain patterns deviate significantly from the average patterns captured by the model.

4.3 Proposed Approach for fNIRS Decoding: fNIRS-Vise

Building upon the foundation of the data sources and models outlined in sections 4.1 and 4.2, we propose a comprehensive approach to decode various mental states from fNIRS data. This approach encompasses the following key steps:

1. **Data Preprocessing:** Raw fNIRS signals are preprocessed using established techniques, such as motion artifact correction, baseline correction, and filtering. We explore the effectiveness of various preprocessing pipelines available in open-source libraries like preprocessingfNIRS (smburns47/preprocessingfNIRS) to optimize signal quality and minimize noise.
2. **Feature Extraction and Selection:** We investigate different feature extraction methods, including traditional techniques like bandpass filtering and wavelet transforms, as well as deep learning-based approaches like autoencoders. These methods aim to extract relevant features that capture the underlying patterns of brain activity associated with specific mental states. Feature selection techniques may be employed to further reduce dimensionality and improve model performance.
3. **Model Development and Training:** We experiment with various deep learning architectures, including the fNIRS Transformer (wzhlearning/fNIRS-Transformer) and fNIRSNet (wzhlearning/fNIRSNet), as well as explore the potential of adapting models from other domains, such as nnUNet (MIC-DKFZ/nnUNet) and nnDetection (MIC-DKFZ/nnDetection), for fNIRS data analysis. These models are trained on the preprocessed fNIRS data, with the goal of learning to accurately predict mental states or decode visual stimuli.
4. **Transfer Learning:** We investigate the feasibility of transfer learning to leverage knowledge from pre-trained models. Specifically, we explore transferring knowledge from fMRI models (e.g., MinD-Vis) to fNIRS, as well as between different fNIRS datasets. This approach could potentially reduce the need for large amounts of labeled fNIRS data and improve model performance.
5. **Model Evaluation:** We evaluate the performance of our models using various metrics, such as classification accuracy, precision, recall, F1 score, and mean squared error (MSE) for image reconstruction tasks. We also employ techniques like cross-validation to assess the generalizability of our models and avoid overfitting.

Model	Architecture	Key Features	Transfer Learning Opportunities
fNIRS-T	Transformer-based	Spatial and channel transformers, self-attention mechanisms	Pre-trained weights from related tasks, domain adaptation for different fNIRS datasets
fNIRSNet	Convolutional Neural Network (CNN)	DHR module for hemodynamic response, global module for spatial patterns	Pre-trained DHR module, global module adaptation, feature extraction for new classifiers
MinD-Vis	Diffusion-based model	Masked brain modelling (MBM), latent diffusion model (LDM), double conditioning	Pre-training on large fMRI datasets, transferring fMRI to fNIRS, cross-task transfer
fNIRS-Vise (Proposed)	Hybrid	Combines preprocessing from fNIRS-T, DHR features from fNIRSNet, spatial-temporal transformer from fNIRS-T, and LDM from MinD-Vis	Pre-training LDM on fMRI data, transferring fMRI to fNIRS, multi-modal learning with EEG

4.4 Challenges and Limitations

Despite the promising potential of our proposed approach, we encountered several challenges and limitations during our practical experiments. These include:

- **Limited fNIRS Data:** The availability of large-scale, labeled fNIRS datasets remains a challenge. This limits the ability to train complex deep learning models and may lead to overfitting.
- **Data Quality:** fNIRS signals are susceptible to noise and artifacts, which can affect the accuracy of decoding. Robust preprocessing techniques are essential to mitigate these issues.
- **Interpretability:** Deep learning models are often considered "black boxes," making it difficult to interpret the underlying neural mechanisms that contribute to the decoded mental states.

4.5 Future Directions

To address these challenges and limitations, future research should focus on:

- **Expanding fNIRS Datasets:** Collecting larger and more diverse fNIRS datasets, ideally with standardized protocols and annotations, would significantly benefit the development of robust and generalizable models.
- **Improving Preprocessing Techniques:** Developing novel preprocessing techniques that can effectively remove noise and artifacts while preserving the relevant information in fNIRS signals is crucial for improving decoding accuracy.
- **Enhancing Interpretability:** Incorporating interpretability methods, such as attention mechanisms and feature visualization techniques, into deep learning models could help elucidate the neural processes underlying fNIRS-based decoding.

By tackling these challenges, we can further advance the field of fNIRS-Vise and unlock its full potential for understanding and decoding the human brain.

5. Discussion

In this thesis, we embarked on a journey to explore the potential of fNIRS-Vise, a framework that combines fNIRS with deep learning for decoding mental states. While our practical experiments with the MinD-Vis reproduction faced challenges, our investigation has yielded valuable insights into both the promises and limitations of this emerging field.

5.1 Promise and Challenges of fNIRS-Vise

The theoretical exploration of fNIRS-Vise paints a promising landscape for decoding complex brain activity using fNIRS data. The availability of open-source datasets like NEMO (Spapé et al., 2023) and fNIRS2MW, coupled with powerful deep learning architectures like fNIRS Transformers (Wzhlearning/fNIRS-Transformer), has paved the way for exciting research in this area. By leveraging the non-invasive and portable nature of fNIRS, researchers can investigate brain activity in more naturalistic settings, opening up possibilities for real-time monitoring and neurofeedback applications. However, our attempts to replicate the MinD-Vis model (zjc062/mind-vis) highlighted several challenges. The limited availability of large-scale, high-quality fNIRS datasets, coupled with the intricacies of model architecture and hyperparameter optimization, presents obstacles to achieving optimal performance and generalizability.

5.2 Transfer Learning and Its Potential

To address the limitations of data scarcity and accelerate research progress, transfer learning emerges as a promising solution. By leveraging pre-trained models from related domains like fMRI (zjc062/mind-vis) or EEG (https://github.com/meagmohit/EEG-Datasets), we can potentially bootstrap the learning process and improve model performance even with limited fNIRS data. The underlying assumption is that neural representations of certain mental states may share similarities across different modalities, allowing for knowledge transfer and reducing the need for extensive training from scratch. However, the effectiveness of transfer learning hinges on the careful alignment of source and target domains, as well as the selection of appropriate transfer strategies.

5.3 Importance of Data and Model Choices

Our investigation has underscored the critical role of data and model choices in the success of fNIRS-Vise applications. The quality and quantity of data, along with the specific task being addressed, can significantly impact the accuracy and robustness of the decoding models. For instance, noisy fNIRS signals may require sophisticated preprocessing techniques to extract meaningful features. Similarly, the choice of deep learning architecture, whether it be a transformer, CNN (wzhlearning/fNIRSNet), or a hybrid approach (JaeyoungShin/hybrid-BCI), must be carefully considered based on the task's complexity and the characteristics of the fNIRS data. Furthermore, the interpretability of the models remains a challenge, as understanding the underlying neural mechanisms that contribute to the decoded mental states is crucial for advancing our knowledge of brain function.

5.4 Future Directions for fNIRS-Vise

To fully realize the potential of fNIRS-Vise, several key areas warrant further investigation:

- **Expanding fNIRS Datasets:** Collaborative efforts are needed to create large-scale, publicly available fNIRS datasets with standardized experimental paradigms, diverse populations, and comprehensive annotations.
- **Refining Transfer Learning Techniques:** Future research should explore various transfer learning strategies, including domain adaptation, fine-tuning, and multi-task learning, to optimize knowledge transfer between different neuroimaging modalities and fNIRS datasets.
- **Developing Specialized Models:** Designing deep learning architectures specifically tailored for fNIRS data, incorporating knowledge of fNIRS signal characteristics and neurovascular coupling, could significantly enhance the performance and interpretability of fNIRS-Vise models.
- **Exploring Unsupervised and Self-Supervised Learning:** Leveraging the vast amounts of unlabeled fNIRS data through unsupervised and self-supervised learning techniques can unlock new avenues for representation learning and improve downstream task performance.

6. Conclusion

This thesis has embarked on an ambitious journey to explore the untapped potential of fNIRS and deep learning for decoding the complexities of the human brain. Through a combination of theoretical exploration, literature review, and practical experimentation, we have shed light on the promises and challenges of this burgeoning field.

6.1 Key Findings and Contributions

Successful fMRI-based Visual Reconstruction: Our reproduction of the MinD-Vis study, while not entirely flawless, reaffirms the power of deep learning to decode visual experiences from fMRI data. This success underscores the potential of computational models to unravel the neural underpinnings of visual perception and paves the way for advancements in brain-computer interfaces and neuroprosthetics.

Potential of fNIRS Transformers: Despite the preliminary nature of our findings, the exploration of fNIRS transformers for cognitive state decoding reveals a promising avenue for non-invasive brain activity monitoring. The model's ability to capture temporal and spatial dependencies within fNIRS signals suggests a capacity to model complex neural processes, opening doors for applications in real-time neurofeedback, mental state assessment, and human-computer interaction.

Illuminating Challenges and Future Directions: Our work has not only highlighted the potential of fNIRS-Vise but also illuminated the challenges that lie ahead. The scarcity of large-scale, high-quality fNIRS datasets, the intricacies of model optimization, and the need for improved interpretability are key areas that require further investigation. By addressing these challenges, we can unlock the full potential of fNIRS and deep learning to revolutionize our understanding of the brain.

6.2 Broader Implications and Ethical Considerations

The integration of fNIRS and deep learning holds immense potential for transforming various fields. In neuroscience, it could lead to a deeper understanding of neural mechanisms underlying cognition, emotion, and perception. In healthcare, it could enable the development of novel diagnostic tools and personalized interventions for neurological and psychiatric disorders. In human-computer interaction, it could pave the way for brain-controlled interfaces that enhance communication and augment human capabilities.

However, as we embrace these technological advancements, we must remain vigilant about the ethical implications of neurotechnology. Ensuring privacy, data security, and responsible use of neurodata are paramount. Striking a balance between scientific progress and ethical considerations will be crucial for fostering public trust and ensuring the equitable and beneficial application of fNIRS-Vise.

6.3 A Vision for the Future

The journey of fNIRS-Vise is still in its early stages, but the path ahead is filled with exciting possibilities. With continued research and development, we envision a future where fNIRS-based deep learning models become indispensable tools for understanding and interacting with the human brain. These models will not only enhance our knowledge of neural processes but also empower individuals to monitor and modulate their own brain activity, leading to improved well-being and cognitive enhancement. As we push the boundaries of this field, we must strive for collaboration, transparency, and ethical responsibility to ensure that the benefits of fNIRS-Vise are accessible to all and used for the betterment of society.

7. References

1. Z. Wang, J. Fang and J. Zhang, "Rethinking Delayed Hemodynamic Responses for fNIRS Classification," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 31, pp. 4528-4538, 2023, doi: 10.1109/TNSRE.2023.3330911.
2. Z. Wang, J. Zhang, X. Zhang, P. Chen and B. Wang, "Transformer Model for Functional Near-Infrared Spectroscopy Classification," in IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 6, pp. 2559-2569, June 2022, doi: 10.1109/JBHI.2022.3140531.
3. Adaloglou, N., & Sergios Karagiannakos. (2022, September 29). How diffusion models work: the math from scratch. AI Summer. https://theaisummer.com/diffusion-models/
4. Andreu-Perez, J., Emberson, L. L., Kiani, M., Filippetti, M. L., Hagras, H., & Rigato, S. (2021). Explainable artificial intelligence based analysis for interpreting infant fNIRS data in developmental cognitive neuroscience. Communications Biology, 4(1), 1–13. https://doi.org/10.1038/s42003-021-02534-y
5. Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B. A., ... Cardoso, M. J. (2022). The Medical Segmentation Decathlon. Nature Communications, 13(1), 4128. https://doi.org/10.1038/s41467-022-30695-9
6. Baker, J. M., Bruno, J. L., Gundran, A., Hosseini, S. M. H., & Reiss, A. L. (2018). fNIRS measurement of cortical activation and functional connectivity during a visuospatial working memory task. PLoS ONE, 13(8), e0201486. https://doi.org/10.1371/journal.pone.0201486
7. Bandettini, P. A., Wong, E. C., Hinks, R. S., Tikofsky, R. S., & Hyde, J. S. (1992). Time course EPI of human brain function during task activation. Magnetic Resonance in Medicine, 25(2), 390-397.
8. Burns, S. M., & Fairclough, S. H. (2021). fNIRS Mental Workload Classifiers: A Comparison of Feature Selection and Classification Methods. Frontiers in Human Neuroscience, 15.
9. Chen, Z., Qing, J., & Zhou, J. H. (2023). Cinematic Mindscapes: High-quality Video Reconstruction from Brain Activity. http://arxiv.org/abs/2305.11675
10. Chen, Z., Qing, J., Xiang, T., Yue, W. L., & Zhou, J. H. (2023). Seeing Beyond the Brain: Conditional Diffusion Model with Sparse Masked Modeling for Vision Decoding. http://arxiv.org/abs/2211.06956
11. Cui, X., Bray, S., & Reiss, A. L. (2010). NIRS-SPM: Statistical parametric mapping for near-infrared spectroscopy. NeuroImage, 49(4), 3241-3250.
12. Eastmond, C., Subedi, A., De, S., & Intes, X. (2022). Deep learning in fNIRS: a review. Neurophotonics, 9(4), 041411. https://doi.org/10.1117/1.NPh.9.4.041411
13. Fazli, S., Mehnert, J., Steinbrink, J., Curio, G., Villringer, A., & Müller, K. R. (2012). Enhanced performance by a hybrid NIRS-EEG brain computer interface. NeuroImage, 59(1), 519-529.

14. Ferrari, M., & Quaresima, V. (2012). A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *NeuroImage*, *63*(2), 921-935.
15. Forbes, S. H., Wijeakumar, S., Eggebrecht, A. T., Magnotta, V. A., & Spencer, J. P. (2021). Processing pipeline for image reconstructed fNIRS analysis using both MRI templates and individual anatomy. Neurophotonics, 8(2), 025010. https://doi.org/10.1117/1.NPh.8.2.025010
16. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., & Girshick, R. (2021). Masked Autoencoders Are Scalable Vision Learners. https://doi.org/10.48550/ARXIV.2111.06377
17. He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. http://arxiv.org/abs/1512.03385
18. Hu, S., Zhang, D., & Guo, T. (2020). A review on functional near-infrared spectroscopy (fNIRS) for brain-computer interface (BCI) applications. Sensors, 20(11), 3213.
19. Irani, F., Plate, A., & Nijholt, A. (2012). Emotion recognition using brain-computer interface in a driving simulator. 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 3973-3976.
20. Kumar, M. G., Tan, C., Libedinsky, C., Yen, S.-C., & Tan, A. Y. Y. (2022). A Nonlinear Hidden Layer Enables Actor-Critic Agents to Learn Multiple Paired Association Navigation. Cerebral cortex (New York, N.Y. 1991), 32(18). https://doi.org/10.48550/arxiv.2106.13541
21. Li, X., Krol, M. A., Jahani, S., Boas, D. A., Tager-Flusberg, H., & Yücel, M. A. (2020). Brain correlates of motor complexity during observed and executed actions. Scientific Reports, 10(1), 10965. https://doi.org/10.1038/s41598-020-67327-5
22. Li, X., Zhang, X., & Lin, L. (2023). fNIRS-Transformer: A Deep Learning Model for Functional Near-Infrared Spectroscopy Data Analysis. NeuroImage, 271, 120005.
23. Lium, M. (2022). Spiral Image Reconstruction for 7T MRI with Gradient System Characterization [Master's thesis, NTNU]. https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/3026238
24. Mughal, N. E., Khan, M. J., Khalil, K., Javed, K., Sajid, H., Naseer, N., Ghafoor, U., & Hong, K.-S. (2022). EEG-fNIRS-based hybrid image construction and classification using CNN-LSTM. Frontiers in Neurorobotics, 16, 873239. https://doi.org/10.3389/fnbot.2022.873239
25. Naseer, N., & Hong, K.-S. (2015). fNIRS-based brain-computer interfaces: a review. Frontiers in Human Neuroscience, 9. https://doi.org/10.3389/fnhum.2015.00003
26. Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proceedings of the National Academy of Sciences, 87(24), 9868-9872.
27. Panov, G. (2023). Quantitative EEG in Patients with Schizophrenia. In D. Stoyanov, B. Draganski, P. Brambilla, & C. Lamm (Eds.), Computational Neuroscience (Vol. 199, pp. 107–115). Springer US. https://doi.org/10.1007/978-1-0716-3230-7_7
28. Pinti, P., Aichelburg, C., Lind, F., Power, S. D., & Merzagora, A. C. (2020). Machine learning for neuroimaging-based brain decoding: a review. Machine Learning, 109(9), 1745-1784.
29. Power, S. D., Falk, T. H., & Chau, T. (2015). Classification of prefrontal activity due to mental workload with functional near infrared spectroscopy. Frontiers in Human Neuroscience, 9, 652.

30. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-Resolution Image Synthesis with Latent Diffusion Models. http://arxiv.org/abs/2112.10752
31. Rosenbaum, R. (2021). On the relationship between predictive coding and backpropagation. PloS one, 17(3), e0266102–. https://doi.org/10.48550/arxiv.2106.13082
32. Schirrmeister, R. T., Springenberg, J. T., Fiederer, L. D., Glasstetter, M., Eggensperger, K., Tangermann, M., ... & Ball, T. (2017). Deep learning with convolutional neural networks for EEG decoding and visualization. Human Brain Mapping, 38(11), 5391-5420.
33. Scholkmann, F., Kleiser, S., Metz, A. J., Zimmermann, R., Mata Pavia, J., Wolf, U., & Wolf, M. (2014). A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. NeuroImage, 85, 6-27.
34. Shin, J., von Lühmann, A., & Blankertz, B. (2020). A Hybrid BCI Combining Motor Imagery and Steady-State Visually Evoked Potentials. Frontiers in Neuroscience, 14.
35. Trappenberg, T. P. (2022). Fundamentals of Computational Neuroscience: Third Edition. Oxford University Press. https://doi.org/10.1093/oso/9780192869364.001.0001
36. Trappenberg, T. P. (2022). Introduction and outlook. In Fundamentals of Computational Neuroscience (3rd ed., pp. 3–31). Oxford University PressOxford. https://doi.org/10.1093/oso/9780192869364.003.0001
37. Vassileva, J., Lee, J.-H., & Haug, K. M. (2019). Stability of Adaptive Neural Networks for Image Reconstruction. http://hdl.handle.net/10852/69486
38. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.
39. Villringer, A., & Chance, B. (1997). Non-invasive optical spectroscopy and imaging of human brain function. *Trends in Neurosciences*, *20*(10), 435-442.
40. Wen, H., Shi, J., Zhang, Y., Lu, K.-H., Cao, J., & Liu, Z. (2018). Neural Encoding and Decoding with Deep Learning for Dynamic Natural Vision. Cerebral Cortex, 28(12), 4136–4160. https://doi.org/10.1093/cercor/bhx268
41. Yang, Q., & Hanjalic, A. (2023). Exploring the Synergy Between Visual and Auditory Modalities for Audio-Visual Event Localization and Detection. arXiv preprint arXiv:2301.11867.
42. Yang, Z., Tong, K., Jin, S., Wang, S., Yang, C., & Jiang, F. (2023). CNN-Siam: multimodal siamese CNN-based deep learning approach for drug–drug interaction prediction. BMC bioinformatics, 24(1), 110–110. https://doi.org/10.1186/s12859-023-05242-y
43. Yonck, R., Dean, R., & Tantor Media. (2017). Heart of the machine: our future in a world of artificial emotional intelligence (Unabridged.). Tantor Media.
44. Zhuang, Y., Liu, L., Singh, C., Shang, J., & Gao, J. (2024). Learning a Decision Tree Algorithm with Transformers. http://arxiv.org/abs/2402.03774
45. Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B. A., Litjens, G., Menze, B., Ronneberger, O., Summers, R. M., … Cardoso, M. J. (2022). The Medical Segmentation Decathlon. Nature Communications, 13(1), 4128. https://doi.org/10.1038/s41467-022-30695-9
46. DeepBCI/Deep-BCI. https://github.com/DeepBCI/Deep-BCI
47. Connectome - Homepage. https://www.humanconnectome.org/

48. Google Gemini . - https://gemini.google.com/app