What You Think Is What You See - A Survey On Evolution Of Electroencephalography For Image Reconstruction

1st Aleena Lifiya
dept. Computer Science and Engineering
American University Of Sharjah
Sharjah, UAE
g00105151@aus.edu

2nd Seba Al Mokdad
dept. Computer Science and Engineering
American University Of Sharjah
Sharjah, UAE
g00105436@aus.edu

3rd Siva Adduri
dept. Computer Science and Engineering
American University Of Sharjah
Sharjah, UAE
g00095118@aus.edu

Abstract—This review explores the evolution of electroencephalography (EEG) as a viable modality for reconstructing images from brain activity evoked by visual stimuli. We begin by outlining the motivation behind such models, followed by a detailed overview of publicly available datasets, signal preprocessing methods, feature extraction techniques, and image generation strategies employed in the field. We then analyze the results across different dimensions, including dataset choice, model architecture, and across different modalities, highlighting how each factor influences reconstruction performance. Finally, we discuss current and emerging applications of EEG-based image reconstruction and propose promising directions for future research.

Index Terms—Brain signals , Image , Feature extraction , GAN , Diffusion.

INTRODUCTION

Retinotopy refers to the forward mapping from visual stimuli in the environment to the activation patterns within the primary visual cortex. Conversely, inverse retinotopy deals with the reconstruction of visual stimuli based on recorded activations in visual brain areas, effectively reversing the encoding process [1]. This reverse mapping has attracted significant attention, especially with advancements in neuroimaging techniques capable of capturing brain activity during visual perception.

One commonly used approach to inverse retinotopy involves blood-oxygen-level-dependent (BOLD) signals captured via functional magnetic resonance imaging (fMRI). These signals provide high spatial resolution, enabling precise localization of brain activations. However, fMRI suffers from poor temporal resolution, rendering it unsuitable for reconstructing dynamic or rapidly changing visual scenes [2]. Furthermore, the high cost and lack of portability of fMRI equipment limit its application to controlled laboratory environments, making it impractical for real-time brain-computer interface (BCI) systems.

Electroencephalography (EEG), on the other hand, is an accessible and portable neuroimaging modality. It is significantly more affordable than fMRI and provides excellent temporal resolution in the millisecond range. These characteristics make EEG a compelling alternative for real-time applications, including dynamic image generation and interactive BCI systems [3]. Despite its advantages, EEG is inherently noisy and exhibits considerable inter-subject variability, presenting significant challenges for decoding visual content from brain activity. The spatial resolution of EEG is also limited, which affects the granularity of image reconstruction tasks. Additionally , aligning brain signals to corresponding images is non-trivial. This requires us to align these modalities in their latent space and then use them to generate the images. This calls for the use of Generative Neural Networks (GNN).

Given these limitations and advantages, it is essential to contextualize EEG within the broader spectrum of non-invasive brain signal modalities. By comparing EEG to other technologies such as fMRI, MEG, and fNIRS, we can identify paradigms and methodologies that have proven successful in those domains and explore their adaptation to EEG-based systems. Such cross-modal insights can inform the development of more robust and generalizable EEG-to-image reconstruction models.

The motivations for pursuing brain signal-to-image reconstruction are multifaceted:

- Advancing Neuroscientific Understanding: Decoding visual experiences from brain signals can provide insights into the neural mechanisms underlying visual perception [4].

- Exploring Animal Perception: Investigating how animals perceive their environment through neural decoding can enhance our understanding of cross-species visual processing [5].
- Personalized Perception Analysis: Studying individual differences in visual perception, such as those highlighted by phenomena like the "dress" illusion, can shed light on subjective visual experiences [6].

The objective of this review is to provide a comprehensive overview of brain signal-based image reconstruction methods, with a particular emphasis on EEG-based techniques. We aim to chart the landscape of current research and highlight the methodologies, datasets, and neural decoding strategies that underpin this emerging field. To achieve this, the paper is structured as follows:

Section I outlines the systematic approach used in selecting and reviewing the literature, including search strategies and inclusion criteria. Section II provides an in-depth explanation of the image reconstruction pipeline from EEG signals, covering signal acquisition, preprocessing, feature extraction, and image generation. In Section III, we evaluate existing studies along key dimensions such as signal modality, dataset used, and image generation technique; this section also includes comparative tables and highlights notable trends. Section IV highlights the application of reconstructing image from brain signals highlighting the importance of more research in this field. Section V presents a comprehensive survey of publicly available datasets across multiple modalities, categorized by signal type and intended task. Section VI discusses the results across various dimensions such as the signal types, datasets and different type of GANS. Section VII summarizes key insights and outlines potential directions for advancing the field of brain signal-based image generation. Section VIII discusses the primary challenges and points of concern in this field. Section IX provides a summary on the various topics covered in this review

Through this review, we hope to foster a deeper understanding of EEG's potential in image reconstruction and inspire future interdisciplinary research that leverages insights across modalities to build more accurate and responsive brain-to-image systems.

I. METHODOLOGY

This review focuses primarily on the task of generating images from EEG signals. However, EEG alone does not provide a comprehensive view of the brain-signal-to-image generation landscape. To situate EEG within this broader context, we also examined studies that utilized non-EEG signals. This allowed us to explore the general frameworks and datasets used across modalities, and to better understand the strengths and limitations of EEG compared to other input signals.

We followed the PRISMA methodology for conducting this systematic review [7], [8]. An initial pool of 210 papers was reviewed, from which 47 were selected based on the following inclusion criteria:

1) The publication year must be 2019 or later;
2) The study must use an open-source dataset;
3) The work must involve image generation from brain signals.

Studies focusing solely on changes in brain activity during vision tasks—without an image generation component—as well as those that addressed only classification tasks, were excluded.

We observed that a significant proportion of selected studies (∼57%) employed diffusion models for image generation. This trend may be attributed to the known challenges associated with training GANs, particularly issues related to convergence [9]. Among the brain signal modalities, fMRI was the most frequently used for image generation tasks, followed by EEG and other signals.

TABLE I
Systematic Review Methodology Overview

Method	Description
Papers Surveyed	210 papers were surveyed overall.
Selected Papers	47 papers were selected for this review.
Eligibility Criteria	1) Publication year must be 2019 or later. 2) Study must use an open-source dataset. 3) Work must involve image generation from brain signals.
Exclusion Criteria	• Studies focusing solely on brain activity changes during vision tasks (no image generation). • Works addressing only classification tasks.
Information Sources	IEEE Xplore, arXiv, Taylor & Francis, PLOS, Elsevier, ScienceDirect, ResearchGate were used in the collection of papers for this review.

II. BACKGROUND

This section discusses the pipeline from a theoretical point of view. We discuss the subjects on which the research is conducted on and its purpose in order to highlight the importance of the EEG to image paradigm . Then we discuss the type of signals used , the preprocessing techniques, the different forms of the input signal. The Models subsection discusses both feature extraction techniques and image generation techniques. Following this is the loss functions and the metrics used for evaluating the models. This section is intended to equip the reader with different terminologies and techniques that will be discussed throughout this review.

A. Subject

There is widespread interest in understanding the visual processing mechanisms of both human and animal

brains. As a result, several experiments have attempted to reconstruct visual stimuli from brain signals in both groups. Within the time frame considered in this survey, approximately 3.8% of the studies focused on non-human subjects, while 96.2% focused on humans, as shown in Figure 1.

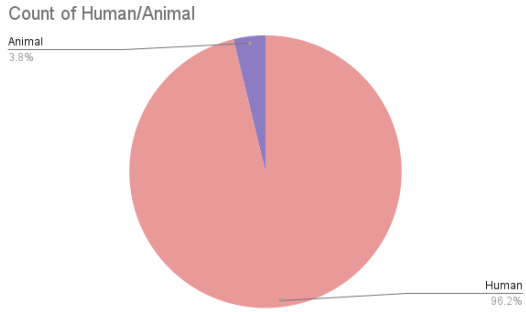

Fig. 1. Proportion of studies using human vs. animal brain signals for image reconstruction.

1) Image Reconstruction from Animal Brain Signals: Table II summarizes key studies that have investigated image reconstruction from animal brain signals, along with the instrumentation methods used. The primary objective of these studies is to understand how different areas of the primary visual cortex respond to visual stimuli. To achieve this, researchers often employ invasive recording techniques, wherein microelectrode arrays are placed directly on the surface of the brain.

These invasive methods enable the acquisition of neural signals with a significantly higher signal-to-noise ratio (SNR) compared to non-invasive techniques, thereby providing more precise information about the neural correlates of visual perception. Figure 2 illustrates the primary visual areas studied in these datasets.

TABLE II
Summary of Image Reconstruction Studies Using Animal Brain Signals

Study	Brain Region / Instrumentation
Chang et al. (2021) [10]	Tungsten microelectrodes inserted 3–5 mm below dura surface
Le et al. (2024) [11]	Electrodes in V1, V4, and IT cortex
Dado et al. (2024) [12]	Electrodes in V1, V4, and IT cortex
Yamashiro et al (2024) [13]	Electrodes in primary somatosensory cortex (S1)
Li et al. (2023) [14]	Publicly available macaque V1 dataset from [15]

2) Image Reconstruction from Human Brain signals: Table III outlines a generalized framework commonly

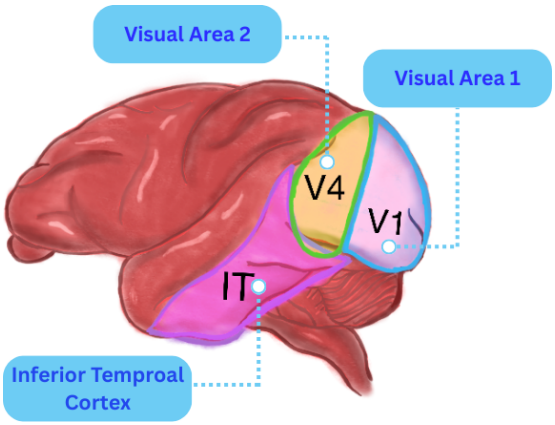

Fig. 2. The figure is a schematic diagram of a macaque brain showing the visual area 1 (V1) , visual area 4 (V4) and Inferior Temporal Cortex (IT)

followed for collecting brain signals from human subjects using various non-invasive methods. Unlike invasive studies, the goal here is typically to learn mappings between brain activity and visual stimuli without surgical intervention.

These studies aim to develop models capable of decoding and reconstructing images based on brain signals, ultimately paving the way for real-world applications such as brain-computer interfaces (BCIs) and neuroprosthetics. The non-invasive nature of these approaches also allows for broader scalability and usability outside research laboratory settings.

B. Type of Signals

1) Non-Invasive Methods:

a) Electroencephalography (EEG): Electroencephalography (EEG) measures the brain's electrical activity, primarily arising from the postsynaptic potentials of cortical pyramidal neurons. It utilizes a cap fitted with electrodes that are placed on the scalp. These electrodes detect voltage fluctuations caused by ionic current flows resulting from neural activity. EEG provides high temporal resolution, making it suitable for tracking rapid brain dynamics.

b) Functional Magnetic Resonance Imaging (fMRI): Functional magnetic resonance imaging (fMRI) measures changes in blood oxygenation levels, which serve as an indirect indicator of neural activity. It employs an MRI scanner that uses strong magnetic fields and radio waves to detect variations in the blood-oxygen-level-dependent (BOLD) signal. Although fMRI offers excellent spatial resolution, its temporal resolution is relatively low due to the slow nature of hemodynamic responses.

c) Magnetoencephalography (MEG): Magnetoencephalography (MEG) records the magnetic fields generated by synchronized neuronal activity, particularly from pyramidal neurons in the cortex. The technique uses a helmet containing superconducting quantum interference

TABLE III
Generalized Frameworks for Non-Invasive Brain Signal Acquisition Methods

Type of Signal	Framework
EEG	Participants view a rapid sequence of category-specific images using the RSVP paradigm. The experiment cycles through category labels, fixation periods, image presentations, and engagement tests, while EEG is recorded throughout [16].
fMRI	High-resolution whole-brain scans are taken while participants view tens of thousands of natural scenes over multiple sessions, allowing high-quality modeling of visual representations [17].
MRI	Participants wear individualized head casts to reduce motion. The framework includes localizer scans for functional regions followed by multiple sessions of visual stimulus presentation using naturalistic object images [18].
MEG	MEG data is collected across several sessions while participants view thousands of object images. Head stabilization is ensured using custom head casts, and eye-tracking is used. Stimulus timing includes jitter to reduce neural synchronization [18].
fNIRS	Participants perform multiple cognitive tasks while seated and interacting with a monitor. Optical signals are recorded using optodes fixed on a cap also used for EEG. Tasks are sequenced to reduce attention fatigue [19].

devices (SQUIDs) to detect extremely weak magnetic fields. MEG provides high temporal resolution and moderate to high spatial resolution, although it requires a magnetically shielded room and is associated with high costs.

d) Functional Near-Infrared Spectroscopy (fNIRS): Functional near-infrared spectroscopy (fNIRS) monitors changes in blood oxygenation and hemodynamics by measuring the absorption of near-infrared light by oxygenated and deoxygenated hemoglobin. The instrument consists of a wearable cap with light sources and detectors that emit and detect near-infrared light. fNIRS offers moderate temporal and spatial resolution and is comparatively more portable and affordable than fMRI or MEG.

Comparison of Non-Invasive Methods: Table IV provides comparison between non-invasive methods in terms of their temporal , spatial resolution , cost and mobility. In order to be useful to general population , the method needs to be accessible and feasible . Compared to the other methods , EEG is one of the most promising techniques to enable the out of laboratory applications of EEG to image systems.

2) Invasive Methods:

a) Electrocorticography (ECoG): Electrocorticography (ECoG) records electrical activity directly from the cortical surface by placing electrodes subdurally on the brain. This technique is typically performed during neurosurgical procedures and provides high temporal and spatial resolution. Since the electrodes are in direct contact with the cortex, signal quality is significantly better than that of non-invasive methods.

TABLE IV
Comparison of Non-Invasive Brain Signal Acquisition Methods

Method	Temporal Resolution	Spatial Resolution	Cost	Mobility
EEG	High (ms)	Low	Low	High
fMRI	Low (s)	High (mm)	High	Low
MEG	High (ms)	Moderate to High	Very High	Low
fNIRS	Moderate (s)	Moderate	Low	High

b) Intracranial Electroencephalography (iEEG): Intracranial EEG (iEEG) involves the implantation of depth electrodes into the brain to monitor electrical activity from deep and cortical structures. It is used primarily in clinical settings, such as pre-surgical epilepsy evaluation. iEEG offers precise spatial localization and excellent temporal resolution, albeit with increased surgical risk.

c) Multi-Unit Activity (MUA): Multi-unit activity (MUA) captures the combined electrical signals of multiple neurons in the vicinity of an implanted electrode. It is often recorded using microelectrode arrays inserted into the brain. MUA provides detailed temporal and spatial resolution of population-level neuronal dynamics but requires complex surgical implantation.

d) Single-Unit Activity (SUA): Single-unit activity (SUA) records the firing patterns of individual neurons using fine microelectrodes implanted in the brain. SUA is the most spatially precise method for monitoring neural activity and is commonly used in neuroscience research. Despite its high resolution, SUA is highly invasive and presents challenges for long-term, large-scale application in humans.

Comparison of Invasive vs Non-Invasive Methods: Table V provides a comparison between invasive methods and non invasive methods based on the following dimensions: signal quality, temporal resolution , spatial resolution , potential usage risk , cost and mobility.

C. Pre-processing

This section deals with the common pre-processing steps for EEG signals. The diagrams if provided for a pre-processing step could be considered to be the result of applying that pre-processing step on Figure 3

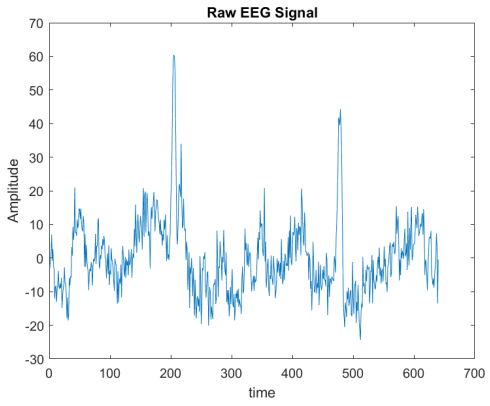

Fig. 3. Raw eeg signal

1) Filtering: Filtering of EEG signals is crucial for removing noise, focusing on relevant frequency bands, and enhancing signal quality. Below are the common types of filtering techniques applied in EEG preprocessing.

2) Removing Frequencies Above the Band of Interest: The typical frequency range of interest for EEG signals lies between 0.5–100 Hz. Frequencies above this range are generally considered noise (e.g., muscle artifacts or hardware interference). A low-pass filter is commonly applied to remove these high-frequency components and retain only the neural activity of interest.

3) Removing Frequencies Below the Band of Interest: Very low-frequency components, such as DC shifts, can distort EEG signals and interfere with analysis. These may arise due to sensor drift or slow physiological processes. A high-pass filter is used to eliminate these low-frequency components, typically removing frequencies below 0.5 Hz.

4) Isolating Frequencies Based on Specific Bands: EEG activity is often studied within specific frequency bands:

- Delta (0.5–4 Hz)
- Theta (4–8 Hz)
- Alpha (8–13 Hz)
- Beta (13–30 Hz)
- Gamma (30–100 Hz)

A band-pass filter allows for the isolation of activity within these frequency ranges, enabling detailed analysis of oscillatory brain dynamics.

5) Removing Line Noise: Power-line interference (50 Hz or 60 Hz depending on region) can significantly degrade EEG signal quality. A notch filter (also known as a band-stop filter) is used to remove this narrowband interference while preserving nearby EEG frequencies.

6) Adaptive Noise Removal: In cases where noise overlaps with EEG frequencies (e.g., muscle artifacts), traditional filters may not suffice. Adaptive filtering techniques dynamically model and subtract non-stationary noise, allowing better preservation of underlying EEG signals.

7) Multi-Resolution Wavelet Analysis: Wavelet decomposition enables time-frequency analysis at multiple resolutions. It is especially effective for removing transient artifacts such as eye blinks or movement artifacts. This method decomposes the signal into wavelets and selectively removes or reconstructs components corresponding to noise. Figure 4 shows MRWA applied on the signal given in Figure 3

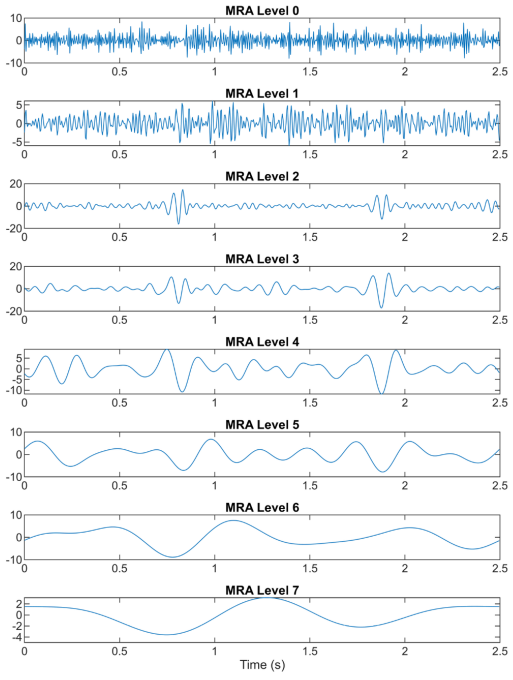

Fig. 4. Multi Resolution Wavelet Analysis of eeg signal

8) Independent Component Analysis (ICA): ICA is a statistical technique used to separate the EEG signal into independent components. It is widely used for artifact rejection (e.g., removing eye blinks, heartbeat artifacts). Once identified, the components associated with noise can

TABLE V
Comparison of Invasive and Non-Invasive Methods

Aspect	Non-Invasive Methods	Invasive Methods
Signal Quality	Susceptible to noise and artifacts due to distance from neural sources	High fidelity signals with minimal noise due to direct contact with brain tissue
Temporal Resolution	Moderate to high; EEG and MEG offer millisecond-level resolution	Very high; capable of capturing precise timing of individual neuron spikes
Spatial Resolution	Low to moderate; limited localization, especially for deep brain activity	High spatial accuracy, including access to deep and cortical brain structures
Risk	Safe and non-surgical; suitable for healthy participants and repeated sessions	High surgical risk; typically used in clinical or research settings only
Cost	Relatively affordable for most methods (especially EEG, fNIRS)	Expensive due to surgical procedures, equipment, and clinical care
Mobility	Portable and adaptable for real-world use (e.g., EEG, fNIRS headsets)	Non-portable; requires controlled clinical or lab environments

be excluded or modified before reconstructing the clean signal.Figure 5 shows ICA of eeg signal.

Fig. 5. Figure shows components in an EEG signal after independent component analysis (ICA) is performed on it . We will be remove the components that are noise or not of interest and reconstruct the signal

9) Hilbert Transform: The Hilbert transform is used to extract the instantaneous amplitude and phase of EEG signals. This is useful for analyzing phase synchronization, phase-amplitude coupling, and other dynamic features of brain activity.

D. Form of Input Signal

The form in which EEG signals are represented can significantly affect the type of analysis or model applied downstream.

1) Raw Signal: The raw time-domain signal is the most direct representation of EEG data. It captures voltage changes over time across multiple electrodes and is often the starting point for preprocessing pipelines.

2) Oscillatory Power: Oscillatory power refers to the signal energy within a particular frequency band over time. It can be obtained via Fourier or wavelet transforms and is useful for studying rhythmic brain activity, such as alpha suppression or beta bursts.

3) Spectrogram: A spectrogram provides a time-frequency representation of the signal by displaying power across frequency bands over time. It is particularly useful for non-stationary EEG analysis and can serve as input to convolutional neural networks for classification tasks. Figure 6 shows spectrogram of eeg given in Figure 3

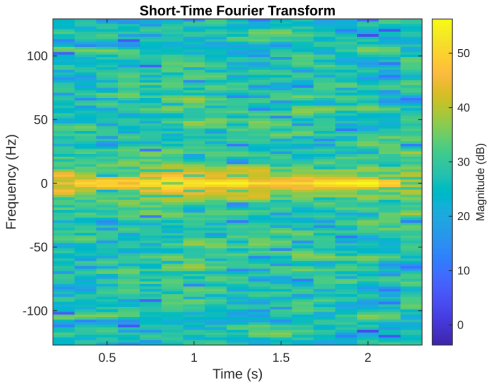

Fig. 6. Spectrogram

E. Models

1) Feature extraction techniques:

a) Masked Auto Encoder: A Masked Autoencoder (MAE) [20] is a self-supervised learning model designed for image reconstruction. It follows an encoder-decoder architecture and is particularly effective for pretraining vision transformers (ViTs) [21].

Working Principle The MAE framework consists of the following steps:

1) Masking: A significant portion (e.g., 75%) of image patches is randomly removed before feeding into the encoder.
2) Encoding: The encoder processes only the visible patches, making the model efficient.
3) Decoding: The decoder reconstructs the missing patches using encoded representations and learnable mask tokens.

4) Loss Calculation: The model is trained by minimizing the difference between reconstructed and original patches.

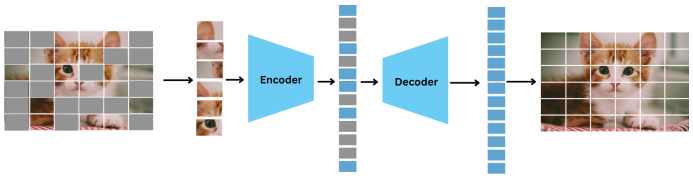

Fig. 7. Overview of the Masked Autoencoder (MAE) process.

b) Long Short-Term Memory (LSTM): Long Short-Term Memory (LSTM) [22] networks are a type of recurrent neural network (RNN) [23] designed to model temporal sequences and capture long-range dependencies. Unlike traditional RNNs, LSTMs include a memory cell and a set of gating mechanisms — the input gate, forget gate, and output gate — that regulate information flow and mitigate vanishing gradients.

Given an input vector x_t at time step t, a previous hidden state h_{t-1}, and previous cell state C_{t-1}, the LSTM computes the following:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad \text{(forget gate)} \tag{1}$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad \text{(input gate)} \tag{2}$$

$$\tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C) \quad \text{(candidate cell state)} \tag{3}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad \text{(new cell state)} \tag{4}$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad \text{(output gate)} \tag{5}$$

$$h_t = o_t \odot \tanh(C_t) \quad \text{(new hidden state)} \tag{6}$$

Here, σ denotes the sigmoid activation, tanh is the hyperbolic tangent, and $\odot$ is element-wise multiplication. In EEG-based applications, LSTMs are valuable for learning temporal dependencies across the signal sequence, enabling more effective decoding of cognitive or perceptual states over time.

c) Geometric Deep Network (GDN): Geometric Deep Networks (GDNs) [24] are designed to leverage the non-Euclidean structure of EEG data by modeling it as a graph. In this framework, EEG electrodes are treated as nodes, and edges are defined based on functional connectivity measures such as Pearson correlation, mutual information, or phase-locking value (PLV). This graph-based representation captures spatial and functional relationships across brain regions.

Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of nodes (electrodes) and $\mathcal{E}$ is the set of edges, and a signal matrix $X \in \mathbb{R}^{N \times F}$ representing F-dimensional features over N nodes, a Graph Convolutional Layer (GCL) performs the following operation:

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}H^{(l)}W^{(l)}\right) \tag{7}$$

Here, $\tilde{A} = A + I$ is the adjacency matrix with added self-loops, $\tilde{D}$ is the degree matrix of $\tilde{A}$, $W^{(l)}$ are trainable weights, and $\sigma(\cdot)$ is a non-linear activation function. This graph convolution enables the model to learn spatially-aware and functionally-informed features from EEG signals. By stacking multiple GCLs, GDNs extract hierarchical, discriminative representations that can enhance downstream decoding tasks. A visualization of this process is shown in Fig.8 .

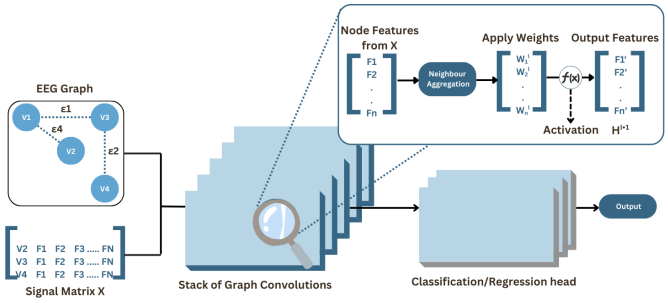

Fig. 8. Overview of Geometric Deep Network (GDN) with EEGs

2) Image generation techniques :

a) Stable Diffusion: Stable diffusion [25] is a generative text to image model. In order to achieve this , the EEG signals should be aligned with text signals before passing it into the stable diffusion model

From EEG to Latents

Let $\mathbf{e}$ represent the EEG embedding after alignment into the image-text latent space, typically via a learned projection network or contrastive training. This vector $\mathbf{e} \in \mathbb{R}^d$ is semantically similar to the CLIP image embedding for the target concept. Our goal is to generate an image $\mathbf{x}_0$ that visually matches this latent concept.

Diffusion as a Generative Process

Stable Diffusion is based on a generative process that learns to reverse the gradual addition of noise to a sample. This can be interpreted as learning to travel backward through a sequence of noisy images:

$$\mathbf{z}_T \to \mathbf{z}_{T-1} \to \cdots \to \mathbf{z}_0$$

Here, $\mathbf{z}_T$ is a sample of pure Gaussian noise, and $\mathbf{z}_0$ is the final latent vector that is decoded into an image. The reverse steps are parameterized by a neural network $\boldsymbol{\epsilon}_\theta$ that learns to denoise each step.

Conditioning on EEG Embeddings

To guide the denoising process using EEG input, we condition the model at each timestep using the EEG embedding $\mathbf{e}$. This is analogous to text conditioning in standard Stable Diffusion. At every timestep t, the network predicts the noise in the latent $\mathbf{z}_t$ as:

$$\boldsymbol{\epsilon}_\theta(\mathbf{z}_t, t, \mathbf{e})$$

Latent Space Generation

Rather than performing diffusion directly in pixel space, Stable Diffusion operates in the latent space of a pretrained autoencoder. The final denoised latent $\mathbf{z}_0$ is passed through a decoder to obtain the output image:

$$\mathbf{x}_0 = \text{Decoder}(\mathbf{z}_0)$$

This improves both efficiency and fidelity, as the model focuses on learning high-level structure rather than low-level pixels.

To summarize, given an EEG signal:

1) The signal is projected into a shared embedding space as vector $\mathbf{e}$.
2) A random Gaussian noise vector $\mathbf{z}_T$ is generated.
3) A denoising U-Net $\boldsymbol{\epsilon}_\theta$ iteratively refines $\mathbf{z}_T$ into $\mathbf{z}_0$, guided by the EEG embedding.
4) The decoder transforms $\mathbf{z}_0$ into a final image $\mathbf{x}_0$.

This pipeline allows the generation of visual interpretations of brain activity, leveraging the power of pretrained diffusion models while grounding them in EEG-derived latent concepts. An outline of this process is shown in Fig. 9.

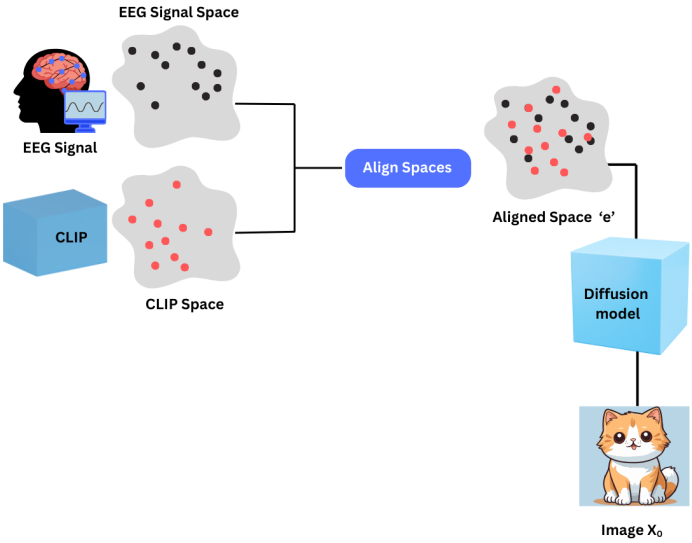

Fig. 9. Overview of the stable diffusion image generation process.

b) Conditional Generative Adversarial Network (Conditional GAN): A Conditional Generative Adversarial Network (cGAN) is an extension of the original GAN framework, designed to generate data conditioned on auxiliary information. In the context of EEG-to-Image generation, the model is conditioned on EEG signals, which may already be aligned with corresponding image embeddings. The generator G receives a random noise vector z along with the EEG feature vector x_{EEG} as input, and produces a synthetic image embedding $G(z, x_{\text{EEG}})$. The discriminator D then takes both the real or generated image embedding and the EEG vector x_{EEG}, and learns to distinguish between real and generated pairs.

$$\min_G \max_D \; \mathbb{E}_{x,y}[\log D(y,x)] + \mathbb{E}_z[\log(1 - D(G(z,x),x))] \quad (8)$$

Here, x represents the EEG signal, y is the real text embedding, and z is a latent noise vector. This formulation encourages the generator to produce outputs that are not only realistic but also conditionally consistent with the EEG input.

F. Loss Function

In this section, we describe the key loss functions that are widely used. Each component targets a specific objective critical to the model's performance.As illustrated in Fig. 10, the loss functions are organized into four categories: reconstruction loss, which ensures accurate recovery of masked or corrupted inputs; diffusion loss, which supervises the denoising process through score-matching; semantic consistency loss, which aligns similarity relationships between EEG and image features; and geometric consistency loss, which encourages class-level structure preservation across modalities. Together, these objectives form a cohesive optimization strategy that enhances both reconstruction fidelity and cross-modal alignment.

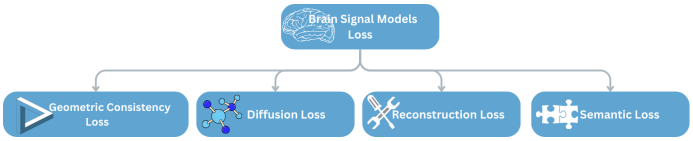

Fig. 10. Overview of Loss Functions

1) Reconstruction loss: The reconstruction loss used in MAE is the Mean Squared Error (MSE):

$$\mathcal{L} = \frac{1}{N} \sum_{i \in \mathcal{M}} (x_i - \hat{x}_i)^2 \quad (9)$$

where:

- x_i is the original pixel value of the masked patches.
- $\hat{x}_i$ is the reconstructed pixel value.
- $\mathcal{M}$ is the set of masked patches.
- N is the total number of masked patches.

Since the loss is computed only on the masked patches, the model learns to infer missing information efficiently.

2) Stable diffusion loss: The training objective of the diffusion model is to minimize the difference between the predicted noise and the true noise $\boldsymbol{\epsilon}$ that was added during the forward process shown in Fig. 11:

$$\mathcal{L}_{\text{diffusion}} = \mathbb{E}_{\mathbf{z}_0,t,\boldsymbol{\epsilon}}\left[\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\mathbf{z}_t, t, \mathbf{e})\|^2\right]$$

where:

- $\mathbf{z}_0$ is the original clean sample (e.g., an image or latent vector).
- t is the diffusion time step.
- $\boldsymbol{\epsilon}$ is the true noise sampled from a Gaussian distribution, $\mathcal{N}(0, I)$.
- $\mathbf{z}_t$ is the noisy version of $\mathbf{z}_0$ at step t.
- $\mathbf{e}$ is the conditioning input (e.g., text prompt, EEG feature, etc.).
- $\boldsymbol{\epsilon}_\theta(\mathbf{z}_t, t, \mathbf{e})$ is the predicted noise by the neural network with parameters θ.
- $\|\cdot\|^2$ denotes the squared L_2 norm.

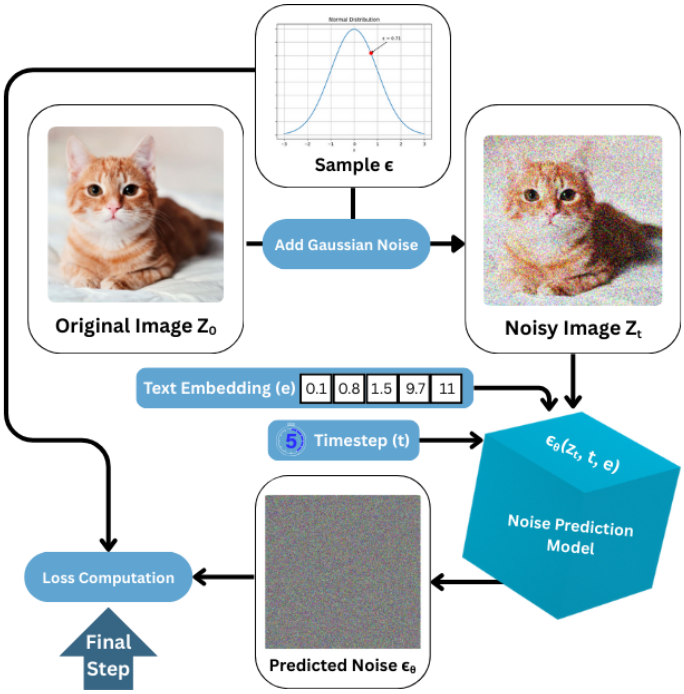

Fig. 11. Overview of the stable diffusion loss.

3) Triplet loss: A triplet loss is used for contrastive learning [26] which has three main aims :

1) Minimize the distance between data with the same label
2) Maximize the distance between data with different labels while maintaining stability
3) Create stable training.

The equation is below satisfies these conditions:

$$\min_\theta \mathbb{E}\left[\|f_\theta(x^a) - f_\theta(x^p)\|_2^2 - \|f_\theta(x^a) - f_\theta(x^n)\|_2^2 + \beta\right], \tag{10}$$

where:

- f_θ : A parameterized function (typically a neural network) that maps EEG signals to a feature space,

$$f_\theta : \mathbb{R}^{C\times T} \to \mathbb{R}^k.$$

- x^a : Anchor EEG signal.
- x^p : Positive EEG signal (same class as anchor).
- x^n : Negative EEG signal (different class from anchor).
- $\|\cdot\|_2^2$: Squared Euclidean (L2) norm, used to measure distance between feature embeddings.
- β : A margin hyperparameter

The process of computing triplet loss is described in Fig. 12

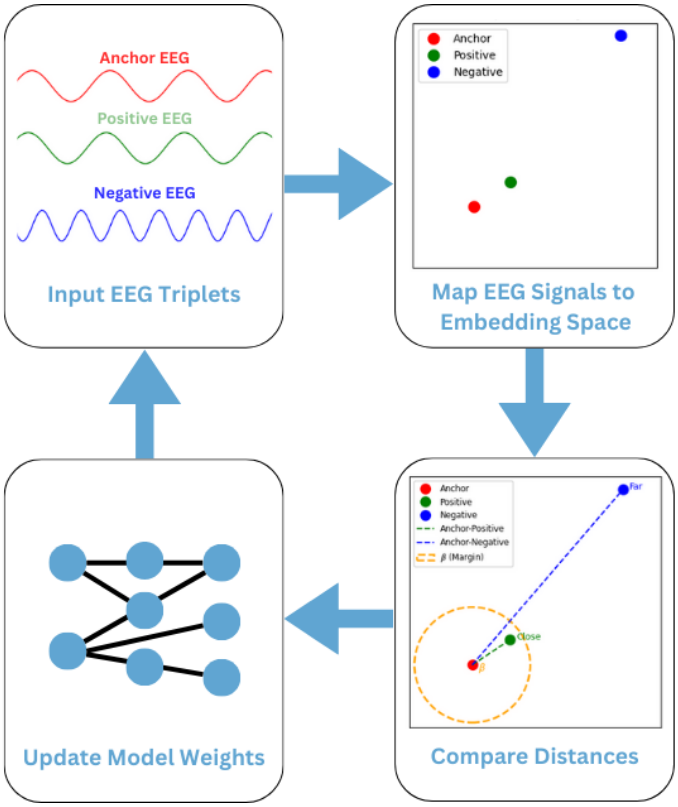

Fig. 12. Overview of the triplet loss computation.

4) Semantic Consistency loss: The objective of the semantic consistency loss is to ensure that the similarity relationships between samples in the EEG feature space reflect those in the image feature space. It is defined as:

$$\mathcal{L}_{\text{Semantic}}(\theta) = \frac{\left\|M^I - M^X\right\|_F^2}{B^2} \tag{11}$$

In this equation, $M^I \in \mathbb{R}^{B\times B}$ denotes the cosine similarity matrix of the image features extracted by CLIP, where each element $M^I[i,j]$ is the cosine similarity between image features z_i and z_j in a batch. Similarly, $M^X \in \mathbb{R}^{B\times B}$ is the cosine similarity matrix of the EEG features $\hat{z}_i$, with $M^X[i,j]$ indicating the cosine similarity

between EEG features $\hat{z}_i$ and $\hat{z}_j$. B is the batch size, and $\|\cdot\|_F$ denotes the Frobenius norm.

5) Geometric Consistency Loss: To promote intra-class similarity between EEG and image features, the Gaussian potential energy between an EEG feature and an image feature is defined as:

$$\mathcal{V}_{\text{Energy}} = \exp\left(-\frac{\|\hat{z}_i - z_k\|^2}{2\sigma^2}\right) \tag{12}$$

where $\|\hat{z}_i - z_k\|$ is the Euclidean distance between the EEG feature $\hat{z}_i$ and the image feature z_k, and σ is the standard deviation of the Gaussian kernel, controlling the decay of the potential energy.

To enforce intra-class consistency during training, for each EEG feature $\hat{z}_i$, n images from the same class are randomly selected. The Gaussian potential energy is then computed between $\hat{z}_i$ and each selected image feature z_k, where $k \in \{1, 2, \ldots, n\}$ and $n \in [1, 10]$. The geometric consistency loss is defined as the average negative potential energy:

$$\mathcal{L}_{\text{Geometric}} = \frac{1}{n}\sum_{k=1}^{n} -\mathcal{V}_{\text{Energy}} \tag{13}$$

G. Metrics

1) Mean Absolute Error (MAE): Mean Absolute Error MAE measures the distance between the embeddings of the real images $E_l(x)$ and that of the synthetic images $E_l(G(z))$. In general , the loss function is given by :

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |E_l(x) - E_l(G(z))| \tag{14}$$

2) Frechet Inception Distance (FID): FID [27] measures the similarity of the feature distribution between real and generated image that belong to the same class. The range of this score is between 0 and infinity , where lower the value , the better.

3) Kernel Inception Distance (KID): KID [28] is similar to FID, it evaluates the quality of generated images but uses Maximum Mean Discrepancy MMD. The scale of this score is between 0 and infinity where the lower the value the better.

4) Structural Similarity Index (SSIM): The Structural Similarity Index (SSIM) [29] is a perceptual metric used to measure the similarity between two images. It is designed to model the human visual system's sensitivity to structural information in an image.

The SSIM value ranges between -1 and 1, where:

- $SSIM = 1$ indicates perfect structural similarity.
- $SSIM = 0$ means no structural similarity.
- $SSIM < 0$ indicates that the images are structurally different.

The SSIM index between two images x and y is computed as:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{15}$$

where:

- μ_x and μ_y are the mean pixel intensities of images x and y.
- σ_x^2 and σ_y^2 are the variances of x and y.
- σ_{xy} is the covariance between x and y.
- C_1 and C_2 are small constants to prevent division by zero, defined as:

$$C_1 = (K_1 L)^2, \quad C_2 = (K_2 L)^2 \tag{16}$$

where L is the dynamic range of pixel values (e.g., 255 for 8-bit images), and K_1, K_2 are typically small values (e.g., 0.01 and 0.03).

5) Top N accuracy: Top-N Accuracy is a performance metric used in image classification tasks to measure whether the correct class label is among the top N predicted classes. It is formally defined as the ratio of correct Top-N predictions to the total number of samples:

$$\text{Top-}N\text{ Accuracy} = \frac{\text{Correct Top-}N\text{ predictions}}{\text{Total samples}} \tag{17}$$

Here, the model ranks all possible class labels based on their predicted probabilities, and a classification is considered correct if the ground-truth label appears within the top N predictions.

6) Inception Score: Inception score [30] access the quality of an image created by a generative model based the output of a pre-trained inceptionv3 image classifier. To achieve a high inception score , the probability distribution of the predicted labels should have a low entropy (high confidence) and the classifier predicts all possible classes that we train the model to generate. A high inception score is desirable and the scale goes from 0 to infinity .

7) Class Diversity: To assess the diversity of generated images for each class, we use a pre-trained image classifier. The classifier outputs predictions in the form of one-hot encoded vectors. The class diversity of a given set of generated images is computed by taking the entropy of the average of these one-hot vectors.

The class diversity score is defined as:

$$\text{Score} = \frac{1}{\log(N)} H\left(\frac{1}{|X|}\sum_{G(z)\in X} C(G(z))\right). \tag{18}$$

Here, N represents the total number of classes, $|X|$ denotes the number of generated samples from EEG signals corresponding to a particular class, C is the classifier that outputs a one-hot prediction vector, and H is the entropy function.

The class diversity score ranges between 0 and 1. A lower score is preferable, indicating that the generated images are relevant to the correct class. A high diversity score suggests that the model is synthesizing images corresponding to multiple classes from EEG signals of a single class, implying poor class relevance.

8) Peak signal-to-noise ratio (PSNR): Peak Signal-to-Noise Ratio (PSNR) [31] quantifies the ratio between the maximum possible signal value and the distortion introduced by noise. It is expressed in decibels (dB) and is given by:

$$PSNR = 10 \log_{10} \left(\frac{MAX^2}{MSE} \right)$$

where MAX is the maximum pixel intensity. Higher PSNR values indicate better image fidelity, with typical values ranging from 20 to 40 dB for natural images.

9) Feature Similarity Index Measure (FSIM): Feature Similarity Index Measure (FSIM) [32] assesses image quality based on low-level perceptual features, particularly phase congruency and gradient magnitude. It is computed as:

$$FSIM = \frac{\sum_i PC_i \cdot S(i)}{\sum_i PC_i}$$

where PC_i is the phase congruency map and $S(i)$ represents structural similarity. FSIM values range from 0 to 1, with higher values indicating better perceptual quality.

10) Signal to Reconstruction Error Ratio (SRE): Signal to Reconstruction Error Ratio (SRE) [33] quantifies the ratio between the original signal strength and the reconstruction error. It is given by:

$$SRE = 10 \log_{10} \left(\frac{\sum x_i^2}{\sum (x_i - y_i)^2} \right)$$

where x_i and y_i are pixel values of the original and generated images, respectively. Higher SRE values indicate lower reconstruction error and better image quality.

11) Universal Image Quality Index (UIQ): Universal Image Quality Index (UIQ) [34] measures image similarity based on loss of correlation, luminance distortion, and contrast distortion. It is defined as:

$$UIQ = \frac{4\sigma_{xy}\mu_x\mu_y}{(\sigma_x^2 + \sigma_y^2)(\mu_x^2 + \mu_y^2)}$$

where μ_x, μ_y are the mean intensities, and $\sigma_x^2, \sigma_y^2, \sigma_{xy}$ represent variance and covariance. UIQ values range from -1 to 1, with 1 indicating perfect similarity.

12) CLIP similarity: Measures the semantic similarity between the generated images and ground truth [35]. The higher the score the better.

Table VI summarizes the metrics in terms of what they measure , their domain values and if they are expected to be higher or lower for a good model.

III. RESULTS

In this section we analyze the input of the papers reviewed across various dimensions- type of input of signals ,datasets and models.

A. Results based on input to the model

1) Raw EEG Signal as Input: Several studies have explored the direct use of raw EEG signals as input for brain-to-image reconstruction pipelines, despite the inherent noise and variability associated with such data. Chen et al. [36] and Li et al. [37] fed unprocessed EEG signals into their respective deep neural architectures, while Singh et al. [38] employed raw EEG signals as input to an LSTM network. In a subsequent work, Singh et al. [39] proposed a stacked LSTM-based encoder, aiming to capture more complex temporal dependencies in raw EEG data.

Beyond recurrent networks, several studies introduced alternative architectures tailored for raw EEG signals. Wang et al. [40] and Khare et al. [41] explored customized EEG encoders, whereas Yang et al. [42] proposed EEG-ConvNet, a convolutional network designed specifically for spatial and temporal feature extraction. Zeng et al. [43] leveraged EVRNet [44], originally designed for video restoration, to process raw EEG data for image generation.

Performance across studies indicates that specialized temporal and convolutional models significantly enhance reconstruction quality. Singh et al. [39] achieved an Inception Score (IS) of 10.82 using a stacked LSTM architecture, while Yang et al. [42] reported the highest IS of 12.38 with EEGConvNet. In comparison, Khare et al. [41] and Zeng et al. [43] achieved lower IS scores of 5.15 and 7.46, respectively. Notably, Li et al. [37] obtained a relatively modest IS of 0.734, highlighting the challenges associated with relying solely on raw EEG without extensive architectural adaptations. Overall, these results underscore that while raw EEG signals present challenges, appropriate neural architectures—particularly those capable of capturing fine-grained spatio-temporal dynamics—can lead to competitive or even superior image reconstruction performance. Table VII summarizes the results of this section.

2) Spectral Maps: An alternative approach in EEG-to-image reconstruction involves representing EEG signals as spectral maps, such as spectrograms, to better capture the time-frequency characteristics inherent in brain signals. Ferrante et al. [45] utilized spectrograms as input to their neural network, demonstrating that frequency-domain features can offer a more structured and discriminative representation for decoding visual information. Kumari et al. [46] similarly employed spectrograms within a Capsule Network-enhanced GAN (CapsGAN) framework, aiming to leverage both spatial hierarchies and frequency patterns to enhance image generation fidelity.

TABLE VI
Evaluation Metrics Summary

Metric Name	What It Measures	Value Range	Better If
Mean Absolute Error (MAE)	Average absolute difference between embeddings	0 to ∞	Lower
Fréchet Inception Distance (FID)	Similarity between feature distributions	0 to ∞	Lower
Kernel Inception Distance (KID)	Quality via Maximum Mean Discrepancy	0 to ∞	Lower
Structural Similarity Index (SSIM)	Perceptual similarity between images	-1 to 1	Higher
Top-N Accuracy	Correct label in top N predictions	0 to 1	Higher
Inception Score (IS)	Image quality and diversity	>0 to ∞	Higher
Class Diversity	Entropy of predicted class distribution	0 to 1	Lower
Peak Signal-to-Noise Ratio (PSNR)	Ratio of signal to noise	0 to ∞ (dB)	Higher
Feature Similarity Index Measure (FSIM)	Image quality based on features	0 to 1	Higher
Signal to Reconstruction Error Ratio (SRE)	Ratio of signal to reconstruction error	0 to ∞ (dB)	Higher
Universal Image Quality Index (UIQ)	Image similarity considering various distortions	-1 to 1	Higher
CLIP Similarity	Semantic similarity between images and text	0 to 100	Higher

TABLE VII
Comparison of Inception Score (IS ↑) across EEG-to-Image Models when input is raw eeg

Paper	IS (↑)
Li et al. (2024) [37]	0.734
Singh et al. (2024) [39]	10.82
Khare et al. (2022) [41]	5.15
Yang et al. (2024) [42]	12.38
Zeng et al. (2023) [43]	7.46

Performance results suggest that spectral representations can meaningfully improve reconstruction outcomes. Ferrante et al. [45] reported a Top-1 accuracy of 41.2% in their spectrogram-based decoding pipeline, while Kumari et al. [46] achieved a high Structural Similarity Index Measure (SSIM) of 0.9203 using CapsGAN. These results highlight the potential of time-frequency domain representations to serve as an effective bridge between raw EEG signals and visual reconstruction models, particularly when coupled with architectures designed to exploit spatial and spectral dependencies.

3) Raw Signal and Spectrogram Combined: Fu et al. [47] proposed a hybrid strategy where both raw EEG signals and their spectrogram representations were used in parallel. In their framework, masked segmentation was first performed on the raw EEG signal to enhance relevant temporal segments, while the corresponding spectrograms were input into an LSTM-based model to capture frequency-based dynamics.

This dual-input design is based on the intuition that raw signals retain fine-grained temporal information, whereas spectrograms capture global frequency trends, making the combination particularly powerful for reconstructing complex visual scenes.

B. Results across different datasets

1) MOABB Dataset: Bai et al. [48] utilized the MOABB dataset for pre-training their model. The MOABB dataset provides a large-scale collection of standardized EEG recordings, allowing robust feature learning across diverse tasks. The authors evaluated their model's performance based on classification accuracy, achieving a score of 45.8%.

2) EEG-ImageNet Dataset: Bai et al. [48] also employed the EEG-ImageNet dataset, a collection of paired EEG recordings and natural images, for EEG-to-image generation tasks. Their model was trained on these EEG-image pairs and evaluated using classification accuracy, achieving a result of 45.8%.

3) Kumar Dataset: The Kumar dataset [49] is one of the most widely adopted benchmarks for EEG-based image reconstruction, owing to its extensive and high-quality EEG recordings captured during visual stimulus experiments. Several studies have explored different strategies on this dataset, yielding a wide range of performance outcomes. Singh et al. [38] combined LSTM-based feature extraction with a conditional GAN, achieving an Inception Score (IS) of 6.78, highlighting the potential of sequential modeling in EEG-to-image tasks. In contrast, Fu et al. [47] extracted both temporal and spectral features and aligned them prior to generation using Stable Diffusion, significantly improving performance with a top-1 accuracy of 45.46%, an IS of 30.9985, and an FID of 126.6576. Lopez et al. [50] further advanced the performance by encoding EEG and image features into a shared latent space using CLIP and conditioning generation via ControlNet [51], resulting in a lower FID of 66.33, an IS of 13.76, and an accuracy of 0.78. Additionally, Singh et al. [39] employed LSTM-based encoders with a StyleGAN-ADA [52] generator, reporting an FID of 174.13 and a KID of 0.065. Overall, methods incorporating feature alignment

and diffusion-based models tend to outperform traditional GAN-based approaches on the Kumar dataset, as reflected in higher IS scores and lower FID values.

TABLE VIII
Comparison of Inception Scores (IS ↑) across Kumar Dataset

Paper	IS (↑)
Singh et al. (2023) [38]	6.78
Fu et al. (2025) [47]	30.99
López et al. (2025) [50]	13.76
Singh et al. (2024) [39] (Kumar's dataset)	9.23

4) EEGCVPR40 (Brain2Image): The EEGCVPR40 dataset [53]–[55] has become a standard benchmark for EEG-based image reconstruction studies, supporting a wide variety of modeling strategies. Singh et al. [39] utilized stacked LSTM encoders in combination with StyleGAN-ADA [52], achieving an Inception Score (IS) of 10.82, a Frechet Inception Distance (FID) of 174.13, and a Kernel Inception Distance (KID) of 0.065. In contrast, Zeng et al. [43] leveraged EVRNet with a denoising diffusion model [56] to achieve a slightly higher IS of 12.55. More recently, Lopez et al. [50] proposed an architecture combining a VAE with CLIP-aligned EEG embeddings and ControlNet guidance, significantly improving generation quality with an IS of 33.87, an FID of 78.11, and a top-1 classification accuracy of 0.91.

Other notable approaches include Khare et al. [41], who employed sequential networks followed by ProGAN [57] to achieve an IS of 5.15, and Yang et al. [42], whose EEGConvNet model reached an IS of 12.38 and an FID of 46.37. Ferrante et al. [58] incorporated knowledge distillation for EEG representation learning prior to Stable Diffusion generation, obtaining a top-1 classification accuracy of 41.2%. Meanwhile, Wang et al. [40] aligned EEG and CLIP embeddings through triple contrastive learning and used FiLM modulation [59] with Stable Diffusion, achieving a CS score of 68.1.

Overall, results on EEGCVPR40 highlight that while traditional architectures like LSTM and ProGAN achieve moderate reconstruction quality, more advanced pipelines incorporating contrastive learning, ControlNet, and diffusion models consistently outperform earlier methods, pushing the boundaries of EEG-to-image synthesis.

TABLE IX
Comparison of Inception Score (IS ↑) and Fréchet Inception Distance (FID ↓) for EEGCVPR40 dataset across EEG-to-Image Models

Paper	IS (↑)	FID (↓)
López et al. (2025) [50]	33.87	78.11
Singh et al. (2024) [39]	10.82	174.13
Khare et al. (2022) [41]	5.1519	–
Yang et al. (2024) [42]	12.38	41.62

5) MindBigData: The MindBigData dataset [60] offers a rich collection of EEG signals recorded during visual stimulus tasks. Kumari et al. [46] utilized spectrograms as input to a CapsGAN framework built upon CapsNet [61], achieving a PSNR of 38.87, SSIM of 0.92, FSIM of 0.62, and UIQ of 0.079.

TABLE X
Comparison of CLIP Accuracy (↑) across EEG-to-Image Models for THINGS-EEG dataset

Paper	CLIP Accuracy (↑)
Chen et al. (2024) [36]	0.278
Li et al. (2024) [37]	0.786
Ferrante et al. (2024) [45]	0.58

6) THINGSEEG dataset: The THINGS [62] and THINGSEEG datasets have become popular in recent EEG-to-image generation research, with several approaches focusing on improving performance through novel model architectures. Li et al. [37] proposed an adaptive temporal model combined with Stable Diffusion XL (SDXL) [63], achieving a Structural Similarity Index Measure (SSIM) of 0.345. In contrast, Chen et al. [36] introduced the NERV model, a multi-attention EEG encoder feeding into SDXL-Turbo with an IP adapter, and reported a Classification Accuracy Transfer (CAT) score of 439.7. Li et al. [64] enhanced feature extraction using channel-wise attention and spatio-temporal convolutions, followed by Stable Diffusion, which resulted in a much higher SSIM of 0.884. Ferrante et al. [45] employed a K-means clustering strategy on CLIP embeddings before conducting text-to-image generation with Stable Diffusion, yielding a top-1 classification accuracy of 58%.

When comparing across these studies, it is clear that specialized models and attention mechanisms significantly improve reconstruction accuracy. For instance, the RealMind model [64] demonstrated superior SSIM performance compared to Li et al.'s earlier work [37]. Additionally, CLIP accuracy benchmarks further emphasize the value of embedding-based methods, with RealMind achieving the highest CLIP accuracy of 78.6% [37], followed by Ferrante et al. [45] with 58% and Chen et al. [36] at 27.9%.

C. Results across different models

1) Attention-based GAN: Shimizu et al. [65] proposed a model using their own dataset, leveraging SincEEGNet [66] for feature extraction and an attention-based GAN for image generation. Their approach achieved a top-1 classification accuracy of 18.3%. Similarly, Mishra et al. [67] employed the Kumar EEG dataset [49] and proposed an attention-based GAN architecture for both feature extraction and image generation. Their model reported an Inception Score (IS) of 6.02 and a mean class diversity score of 0.4051 with a standard deviation of 0.0645.

2) GAN: Jiao et al. [68] utilized the Kumar EEG dataset and implemented a CNN for feature extraction followed by a standard GAN for visual reconstruction, resulting in an IS of 6.33. Mishra et al. (2024) [69] also used the Kumar dataset, extracting features with

a C-former comprising convolutional and self-attention modules, followed by GAN-based image reconstruction. Their model achieved an Inception Score of 4.62 and a class diversity score of 0.7897. Khaleghi et al. [70] employed a Geometric Deep Network (GDN) for feature extraction along with a GAN for image synthesis, and reported a Structural Similarity Index Measure (SSIM) of 0.89.

3) Conditional GAN and StyleGAN: Singh et al. [38] used the Kumar dataset [49], extracting features via LSTM and generating images using a conditional GAN. Their model achieved an Inception Score of 6.78. Singh et al. [39], who also utilized both the Kumar dataset and EEGCVPR40 [53], adopted an LSTM-based EEG encoder and used StyleGAN-ADA [52] for image generation. Their model attained an FID of 174.13 on EEGCVPR40 and 109.49 on the Kumar dataset, along with KID scores of 0.065 and 0.039 respectively.

4) Progressive and Capsule GANs: Khare et al. [41] also used the Kumar dataset and passed EEG time series through sequential neural networks to create feature vectors, which were input into ProGAN [57] for reconstruction, yielding an Inception Score of 5.15. Kumari et al. [46] employed the MindBigData dataset [54], where EEG signals were first converted into spectrograms and then processed through a CapsGAN, which combines CapsNet [61], [71], [72] with a GAN. The model achieved PSNR of 38.8734, SSIM of 0.9203, FSIM of 0.6198, and UIQ of 0.0791.

5) Stable Diffusion: Bai et al. [48] used the datasets from Aristimuña [73] and Zhu [16], applying masked signal modeling for feature extraction and Stable Diffusion for image generation, reporting a top-1 accuracy of 45.8%. Li et al. [37] utilized the THINGSEEG dataset, applying an Adaptive Temporal Model and Stable Diffusion XL (SDXL) [63], achieving an SSIM of 0.345. Chen et al. [36] used the THINGS dataset [62], proposed a novel EEG encoder (NERV) using multi-attention, and passed representations through an IP adapter into SDXL-Turbo, reporting a CAT score of 439.7. Fu et al. [47] used the Kumar dataset, extracting both time and frequency domain features which were aligned and fed into Stable Diffusion. Their model reached a top-1 accuracy of 45.46%, IS of 30.9985, and FID of 126.6576. Wang et al. [40] used MindBigData and aligned EEG embeddings with CLIP using mask-based triple contrast learning, followed by Feature-wise Linear Modulation (FiLM) [59] into Stable Diffusion, achieving a CS score of 68.1. Ferrante et al. [58] used EEGCVPR40 and applied knowledge distillation for EEG feature extraction before generating images with Stable Diffusion, obtaining a top-1 accuracy of 41.2%. Ferrante et al. (2024) [45] also used both THINGSEEG and Kumar datasets, creating 8 pseudo-label clusters via CLIP and K-means, followed by text-to-image generation via Stable Diffusion from predicted class prompts, achieving a top-1 accuracy of 58%. Li et al. [64] utilized the THINGS dataset with a channel-wise attention model followed by temporal-spatial convolutions, feeding learned EEG embeddings into Stable Diffusion and achieving an SSIM score of 0.373 on low level features and 0.884 on high level. Yang et al. [42] used Kumar's dataset and proposed EEGConvNet for representation extraction and classification, reporting a mean top-1 accuracy of 25.21%, top-5 accuracy of 34.08%, IS of 12.38, and FID of 46.37. Lopez et al. [50] used EEGCVPR40 and Kumar datasets, employing VAE-based image encodings and EEG-CLIP alignment, integrating into ControlNet [51] to condition the Stable Diffusion model. On EEGCVPR40, their model achieved FID of 78.11, IS of 33.87, and accuracy of 0.91; on Kumar's dataset, the FID was 66.33, IS was 13.76, and accuracy was 0.78.

6) Denoising Diffusion Probabilistic Models (DDPM): Zeng et al. [43] used the EEGCVPR40 dataset and proposed EVRNet to extract temporal and spatial information from EEG, generating images using a Denoising Diffusion Probabilistic Model (DDPM) [56] in combination with ConvNet [74], achieving a high Inception Score of 12.55.

7) Axial Attention-based GAN: Zhao et al. [75] used MindBigData and combined CNN and axial attention [76] for capturing fine and long-term EEG dependencies. They proposed a dual-axial GAN integrated with the WGAN framework [77], achieving a classification accuracy of 84.65%, MSE of 0.72, and PSNR of 15.91 dB.

Table XI summarizes the key values for different across various research papers.

IV. APPLICATION OF RECONSTRUCTING IMAGE FROM BRAIN SIGNALS

Reconstructing visual stimuli from brain signals can help us understand the visual pathway of humans and animals alike. For animals , it would provide us insights with how they perceive their surroundings and can be useful in fields like nature inspired robotics and also helps in understanding how information is encoded in animal brains better which can subsequently lead to insights to human brain. For example, [81] showed that in macaque , the image of a face is encoded in around 200 neurons and can be linearly reconstructed by their response.

For humans, we might be able to understand dreams which of interest to the neuroscience community.

It can be used in facial composites when the victim or the eyewitness may not have the ability to communicate verbally. It may help us understand mental illnesses better when standardized questionnaires and behavioral assessments maybe sub optimal [82]

V. DATASETS

This section discusses the major open source datasets available across various types of brain signals . FMRI is the most prevalent type of signal used for image generation and hence has has the largest number of literature covered

TABLE XI
EEG-to-Image Models Grouped by Image Generator Type

Image Generator	Paper	Architecture	Acc (%)	IS	FID	SSIM
Attention-based GAN	Shimizu et al. (2022) [65]	Sinc-EEGNet	18.3	-	-	-
	Mishra (2023) [67]	Attention-based GAN	-	6.02	-	-
GAN	Jiao et al. (2019) [68]	CNN	-	6.33	-	-
	Mishra et al. (2024) [69]	C-former	-	4.62	-	-
	Khaleghi (2022) [70]	GDN	-	-	-	0.89
WGAN	Zhao et al. (2023) [75]	Axial WGAN Modules	84.65	-	-	-
Conditional GAN	Liu et al. (2023) [78]	Seq2Seq LSTM	-	6.78	-	-
StyleGAN-ADA	Singh et al. (2023) [38]	LSTM Encoder + Classifier	-	-	109.49	-
Stable Diffusion	Li et al. (2024) [64]	Channel-Aware Conv	-	-	-	0.884
	Fu et al. (2025) [47]	Multi-path Embedding	-	30.99	-	-
	Bai et al. (2023) [48]	Masked Signal Modeling	45.8	-	-	-
	Ferrante (2024a) [45]	Distillation (CLIP + CNN)	41.2	-	-	-
	Ferrante (2024b) [79]	Distilled Residual CNN	58	-	-	-
Latent Diffusion	Lan et al. (2023) [80]	Coarse + Fine	85.6	33.50	-	-
CapsGAN	Kumari et al. (2023) [46]	Spectrogram + CapsGAN	-	-	-	0.92
ControlNet + SD	Lopez et al. (2025) [50]	Cross-latent Mapping	-	13.76	66.33	-
SDXL-Turbo	Chen et al. (2024) [36]	NERV + IP Adapter	27.9	-	-	-
DDPM + ConvNet	Zeng et al. (2023) [43]	Temporal-Spatial Net	-	12.55	-	-

. However , due to the factors mentioned in section II-B , EEG is gaining popularity as a input signal for this task. Figure 13 shows the proportion of signals in the papers reviewed.

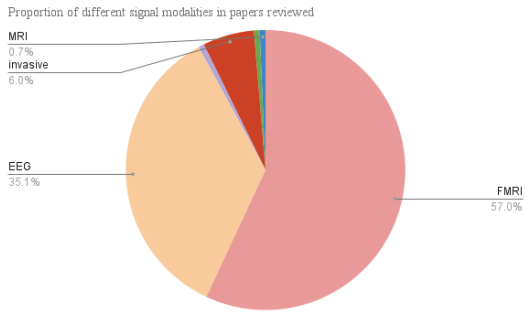

Fig. 13. Proportion of different types of signals used in the research papers that are reviewed.

A. Signals obtained from Non Invasive

1) Electroencephalography (EEG):

a) MOABB dataset: [1] [73] is an online platform that allows researchers to build benchmarks of various algorithms on publicly available dataset. This platform provides data for various BCI paradigms such as motor imagery , P300/ERP ,SSVEP , c-VEP , Resting States and other compound dataset. These dataset can be leveraged for pre-training feature extraction models or used to generate images (example , from motor imagery dataset one can create images of arrows pretaining to different directions)

b) ImageEEGNet dataset: [2] The ImageEEGNet [16] dataset is constructed from the ImageNet dataset [83]. The dataset comprises of recordings from 16 subjects exposed to 4000 images.

c) Kumar's EEG dataset: [3] [49] used a 14 channel EEG setup and 23 participants to get the EEG signals from viewing the images belonging to three categories : Char , digit image. There were total of 690 images (230 per category).

d) EEGCVPR40 dataset: [4] [53] dataset was created using 6 subjects . A total of 2000 images from ImageNet dataset were shown as the visual stimuli [55]

[1]MOABB dataset is available at https://moabb.neurotechx.com/docs/dataset_summary.html

[2]ImageEEGNet dataset is available at https://github.com/Promise-Z5Q2SQ/EEG-ImageNet-Dataset?tab=readme-ov-file

[3]This dataset is available at https://www.kaggle.com/datasets/ignazio/kumars-eeg-imagined-speech

[4]This dataset is available at https://github.com/perceivelab/eeg_visual_classification

e) Hirokatsu dataset: [5] [65] dataset is created using 40 classes of the ImageNet dataset just like [53]. The authors of the dataset created 50 images per class summing to a total of 2000 EEG-image pairs. The eeg was collected from 4 subjects.

f) ThingsEEG dataset: [84] collected EEG-image pairs from 10 participants . The visual stimuli was the THINGS dataset [62], [85] . The dataset has 12 or more images of objects on a natural background for each of 1854 object concepts sub categories, where each concept belongs to one of 27 higher-level categories .

g) MindBigData "IMAGENET" of The Brain: [6] [54] dataset contains 70,060 brain signals of 3 seconds each, captured with the stimulus of seeing a random image (14,012 so far) from the Imagenet ILSVRC2013 train dataset from a single participant

2) Functional magnetic resonance imaging (FMRI):

a) Shen's Deep Image Reconstruction: The dataset is created from three participants . The visual stimuli belonged to three categories natural images, artificial shapes, and alphabetical letters from the ImageNet dataset.

b) Natural Scene Dataset: [17] dataset was created from 8 participants being shown the visual stimuli of 73,000 color natural scenes from the Microsoft Common Objects in Context(COCO) image dataset [86].

c) Faces dataset: [87] dataset was created from 4 participants with the visual stimuli being the Celeb A dataset [88].

d) Visual image reconstruction: [89] was created from 4 participants were shown random images.

e) Human Connectome Project: Similar to [73] for EEG , the [90] is a collection of FMRI datasets pretaining to various tasks. This also allows researches to leverage the data for vision related tasks.

f) Generic Object Decoding (fMRI on ImageNet) (GOD) : [91] created the dataset from five participants with images that were collected from an online image database ImageNet dataset.

g) Brain, Object, Landscape Dataset (BOLD5000): [92] dataset was created from 4 participants with 4,916 unique scenes

h) Vim-1 dataset: [93]: This dataset was created from two subjects who were shown natural images across 70 runs. Table XIII summarises the datasets and the papers that uses it , along with the links to their sources.

3) Magnetoencephalography(MEG):

a) THINGS-MEG dataset dataset: [18] dataset is created from 4 participants who underwent 12 MEG sessions during which they were presented with 22,000+ unique images belonging to 1,800+ categories from the THINGS database

[5]Hirokatsu dataset is available at https://osf.io/2fgks/

[6]MindBigData dataset is available at https://mindbigdata.com/opendb/imagenet.html

4) FNIRS:

a) Shin et.al open access dataset: [19] dataset was created from 26 participants performaning three different tasks. Three datasets -dataset A , dataset B and dataset C - are derived from these three tasks namely n-back , discrimination/selection response and word generation which can be used generate images.

B. Signals obtained from Invasive Methods

a) Macaque V1 dataset: Micro electrode arrays were used to obtain MUA from macaque when they were presented with natural images and gratings on a screen [157]

b) Brain2gan: The MUA from the primate was collected when it was shown images of faces and natural images [12].

The table XIV shows the summary of these datasets , the papers that use them and the link to the sources. However , for non-invasive , we havent directly provided the link but cited the paper that describes it for further reference.

VI. DISCUSSION

A. Efficacy of different signal types

FMRI has an overall better performance when compared to other modalities. However , Geometric Deep Network [70] (Figure 14) and Multi-level Semantics Extraction [80] (Figure 15) have shown to produce better results compared to their FMRI counterparts while using the same image generation techniques. Hence, we have fair reasons to believe that better feature extraction techniques and encoder designs can help eeg signals outperform FMRI signals , hence improving its utility,

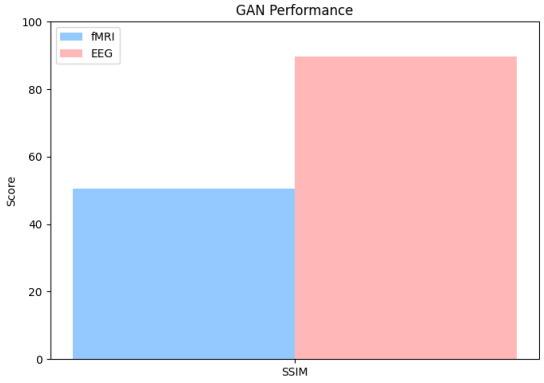

Fig. 14. grouped bar chart for performance of EEG and FMRI using the GAN

B. Efficacy of different datasets

[50] tested their model on both Kumar's datatset (ThoughtViz) [49] and EEGCVPR40 [53]. Their model performed better on EEGCVPR40 compared to Kumar's as shown in Figure 16. This is an important point to notice while comparing the results between models if they dont

TABLE XII
Summary of EEG Datasets Used in Image-EEG Research

Dataset	Subjects	Images	Source	EEG Setup	Procedure / Description
MOABB [73]	Varies	Varies	Multiple BCI tasks	Varies	Benchmarking platform with EEG data for motor imagery, P300/ERP, SSVEP, c-VEP, resting state, and compound tasks. Useful for pre-training or generating task-specific stimuli.
ImageEEGNet [16]	16	4000	ImageNet [83]	62	EEG collected while subjects viewed 4000 natural images. Dataset derived from ImageNet for vision-based decoding research.
Kumar's EEG Dataset [49]	23	690	Characters, Digits, Objects	14 channels	EEG signals recorded during viewing of 690 images (letters and digits). 230 images per class.
EEGCVPR40 [53]	6	2000	ImageNet [55]	128	2000 ImageNet images used as stimuli. Collected for visual classification using EEG signals.
Hirokatsu Dataset [65]	4	2000	ImageNet	128	EEG recorded while participants viewed 50 images per class for 40 ImageNet classes. 2000 EEG-image pairs.
ThingsEEG [84]	10	22,248+	THINGS [62], [85]	64	EEG collected from viewing objects in natural scenes. Includes 12+ images per concept for 1854 object concepts across 27 categories.
MindBigData [54]	1	14,012+	ImageNet (ILSVRC2013)	128	Contains 70,060 EEG signals (3 sec each), recorded while viewing random ImageNet images.

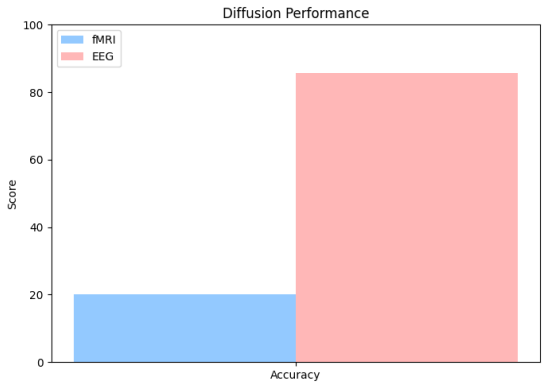

Fig. 15. grouped bar chart for performance of EEG and FMRI using the GAN

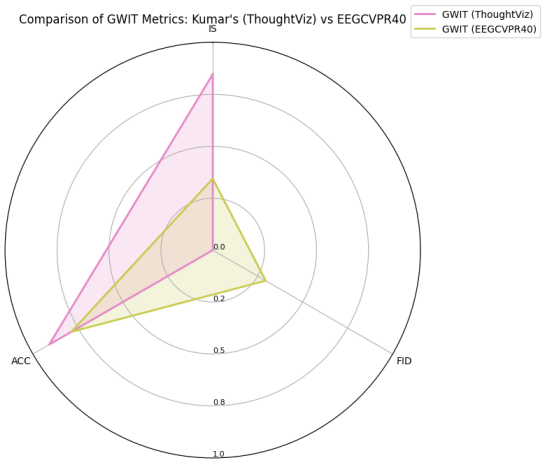

Fig. 16. Radar chart showing the performance of different GANs on mindbigdata in spectrogram form [53] and [49]

test on the same dataset. This may also indicate that dataset used in training might influence the performance of the model and consequently the need for pretraining.

C. Efficacy of different types of GANS

According to the experiments conducted by [46], which evaluated various GAN architectures for image generation from spectrogram-based EEG representations obtained from the [60] dataset, CapsGAN [162] consistently outperformed vanilla GANs, CGAN [163], DCGAN [164], BiGAN [165], and ACGAN [166] across multiple evaluation metrics, as shown in Figure 17.

The superior performance of CapsGAN can be attributed to its use of capsule networks with dynamic routing, which enables the model to better preserve hierarchical spatial relationships and encode richer feature representations. Compared to traditional convolutional architectures, capsule networks are more effective at modeling part–whole relationships and maintaining spatial pose information, which is particularly beneficial when mapping complex, structured inputs like EEG spectrograms into coherent visual outputs.

TABLE XIII
FMRI Datasets for Visual Image Reconstruction

Dataset (Citation)	Papers Using It	Link
Shen's Deep Image Reconstruction (Shen et al., 2019 [94])	[94]–[100]	https://github.com/KamitaniLab/DeepImageReconstruction
Natural Scene Dataset (Allen et al., 2022 [17])	[78], [79], [101]–[128]	https://openneuro.org/datasets/ds003701
Faces Dataset (VanRullen et al., 2019 [87])	[87], [129], [130]	https://openneuro.org/datasets/ds001761/versions/2.0.1
Visual Image Reconstruction (Miyawaki et al., 2008 [89])	[129]	https://brainlife.io/dataset/OpenNeuro/ds000255/00002
Human Connectome Project (Van Essen et al., 2013 [90])	[121], [131], [132]	https://www.humanconnectome.org/
Generic Object Decoding (GOD) (Horikawa et al., 2017 [91])	[78], [79], [103], [131]–[152]	https://openneuro.org/datasets/ds001246/versions/1.2.1
BOLD5000 (Chang et al., 2019 [92])	[79], [131], [132], [142], [144], [146], [150], [153]	https://bold5000-dataset.github.io/website/
Vim-1 (Kay et al., 2008 [93])	[137], [139], [154]–[156]	https://crcns.org/data-sets/vc/vim-1/about-vim-1

TABLE XIV
Other Brain Recording Modalities Used for Image Generation

Modality	Dataset (Citation)	Papers Using It	Link
MEG	THINGS-MEG Dataset (Hebart et al., 2023 [18])	[158]	https://openneuro.org/datasets/ds004212/versions/2.0.1
fNIRS	Shin et al. Open Access Dataset (Shin et al., 2018 [19])	[159]	https://doc.ml.tu-berlin.de/simultaneous_EEG_NIRS/
Invasive (MUA)	Macaque V1 Dataset (Coen-Cagli et al., 2015 [157])	[157], [160]	-
Invasive (MUA)	Brain2GAN Dataset (Dado et al., 2024 [12])	[12], [161]	-

VII. FUTURE WORK

Most of the work that has been done is on generating a single object as an image without necessarily considering the background information. In order to create more realistic looking images we need background information , texture , lighting and shadows which makes the generation process more complex, but is essential, for most practical application of image generation from brain signals.

If eeg to image models needs to be used for generating novel images then there needs to be a way to measure how successful it has been in doing so . This is going to be a challenging task because its often not easy to quantitatively compare an idea that is non-materialised to an actual image.

plethora of research is needed on the potential of using these models as a way to safely use these models in the realm of psychiatry

VIII. CHALLENGES

One of the primary challenges in the field is the lack of standardized evaluation metrics. Different studies report results using varied benchmarks such as accuracy, IS, FID, or SSIM, making it difficult to draw a direct and fair comparison between models. Beyond technical hurdles, there are also significant ethical, security, and privacy concerns associated with EEG-to-image generation, as highlighted by [167]. These risks become even more critical as models grow more capable. In order to generate more sophisticated and realistic images, models must be trained on high-quality datasets; however, this raises intellectual property (IP) concerns and questions around the ethical sourcing of images, as discussed in [168]. Furthermore, if

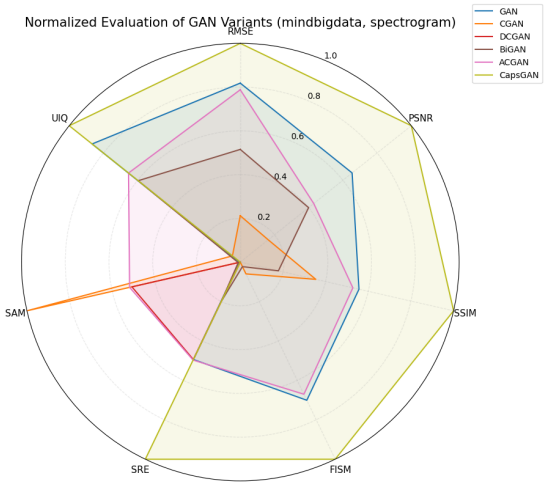

Fig. 17. Radar chart showing the performance of different GANs on mindbigdata in spectrogram form [46]

these models are to be deployed on a larger scale, the environmental impact associated with the computational resources required for training and image generation must be carefully considered. Prioritizing medical and rehabilitative applications, where the societal benefit is more pronounced, could help justify the resource expenditure more responsibly compared to uses in purely entertainment contexts in the hands of giant media conglomerates.

IX. CONCLUSION

In this review, we examined the diverse methodologies employed in reconstructing images from brain signals, highlighting the unique challenges that make this a non-trivial and still-maturing area of research. We explored the full reconstruction pipeline, from signal acquisition and preprocessing to image generation and evaluation. Additionally, we compiled a comprehensive overview of publicly available datasets spanning multiple paradigms to support future benchmarking and development. Our analysis underscores the need for continued research, particularly in improving cross-subject generalization, integrating multimodal inputs, and achieving real-time, high-fidelity reconstructions.

References

[1] B. Thirion, E. Duchesnay, E. Hubbard, J. Dubois, J.-B. Poline, D. Lebihan, and S. Dehaene, "Inverse retinotopy: inferring the visual content of images from brain activation patterns," Neuroimage, vol. 33, no. 4, pp. 1104–1116, 2006.

[2] S. Nishimoto, A. T. Vu, T. Naselaris, Y. Benjamini, B. Yu, and J. L. Gallant, "Reconstructing visual experiences from brain activity evoked by natural movies," Current biology, vol. 21, no. 19, pp. 1641–1646, 2011.

[3] X.-H. Liu, Y.-K. Liu, Y. Wang, K. Ren, H. Shi, Z. Wang, D. Li, B.-L. Lu, and W.-L. Zheng, "Eeg2video: Towards decoding dynamic visual perception from eeg signals," in Advances in Neural Information Processing Systems, A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, Eds., vol. 37. Curran Associates, Inc., 2024, pp. 72 245–72 273. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2024/file/84bad835faaf48f24d990072bb5b80ee-Paper-Conference.pdf

[4] Y. Kamitani and F. Tong, "Decoding the visual and subjective contents of the human brain," Nature neuroscience, vol. 8, no. 5, pp. 679–685, 2005.

[5] E. I. Knudsen, "Evolution of neural processing for visual perception in vertebrates," Journal of Comparative Neurology, vol. 528, no. 17, pp. 2888–2901, 2020.

[6] L. Schlaffke, A. Golisch, L. M. Haag, M. Lenz, S. Heba, S. Lissek, T. Schmidt-Wilcke, U. T. Eysel, and M. Tegenthoff, "The brain's dress code: How the dress allows to decode the neuronal pathway of an optical illusion," Cortex, vol. 73, pp. 271–275, 2015.

[7] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan et al., "The prisma 2020 statement: an updated guideline for reporting systematic reviews," bmj, vol. 372, 2021.

[8] M. J. Page, D. Moher, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan et al., "Prisma 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews," bmj, vol. 372, 2021.

[9] L. Mescheder, A. Geiger, and S. Nowozin, "Which training methods for gans do actually converge?" in International conference on machine learning. PMLR, 2018, pp. 3481–3490.

[10] L. Chang, B. Egger, T. Vetter, and D. Y. Tsao, "Explaining face representation in the primate brain using different computational models," Current Biology, vol. 31, no. 13, pp. 2785–2795, 2021.

[11] L. Le, P. Papale, K. Seeliger, A. Lozano, T. Dado, F. Wang, P. Roelfsema, M. A. van Gerven, Y. Güçlütürk, and U. Güçlü, "Monkeysee: Space-time-resolved reconstructions of natural images from macaque multi-unit activity," Advances in Neural Information Processing Systems, vol. 37, pp. 93 826–93 848, 2024.

[12] T. Dado, P. Papale, A. Lozano, L. Le, F. Wang, M. van Gerven, P. Roelfsema, Y. Güçlütürk, and U. Güçlü, "Brain2gan: Feature-disentangled neural encoding and decoding of visual perception in the primate brain," PLoS computational biology, vol. 20, no. 5, p. e1012058, 2024.

[13] K. Yamashiro, N. Matsumoto, and Y. Ikegaya, "Diffusion model-based image generation from rat brain activity," Plos one, vol. 19, no. 9, p. e0309709, 2024.

[14] W. Li, S. Zheng, Y. Liao, R. Hong, C. He, W. Chen, C. Deng, and X. Li, "The brain-inspired decoder for natural visual image reconstruction," Frontiers in Neuroscience, vol. 17, p. 1130606, 2023.

[15] A. Kohn and M. A. Smith, "Utah array extracellular recordings of spontaneous and visually evoked activity from anesthetized macaque primary visual cortex (v1)," http://dx.doi.org/10.6080/K0NC5Z4X, 2016, available from CRCNS.org.

[16] S. Zhu, Z. Ye, Q. Ai, and Y. Liu, "Eeg-imagenet: An electroencephalogram dataset and benchmarks with image visual stimuli of multi-granularity labels," arXiv preprint arXiv:2406.07151, 2024.

[17] E. J. Allen, G. St-Yves, Y. Wu, J. L. Breedlove, J. S. Prince, L. T. Dowdle, M. Nau, B. Caron, F. Pestilli, I. Charest et al., "A massive 7t fmri dataset to bridge cognitive neuroscience and artificial intelligence," Nature neuroscience, vol. 25, no. 1, pp. 116–126, 2022.

[18] M. N. Hebart, O. Contier, L. Teichmann, A. H. Rockter, C. Y. Zheng, A. Kidder, A. Corriveau, M. Vaziri-Pashkam, and C. I. Baker, "Things-data, a multimodal collection of large-scale datasets for investigating object representations in human brain and behavior," Elife, vol. 12, p. e82580, 2023.

[19] J. Shin, A. Von Lühmann, D.-W. Kim, J. Mehnert, H.-J. Hwang, and K.-R. Müller, "Simultaneous acquisition of eeg and nirs during cognitive tasks for an open access dataset," Scientific data, vol. 5, no. 1, pp. 1–16, 2018.

[20] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," 2021. [Online]. Available: https://arxiv.org/abs/2111.06377

[21] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," 2021. [Online]. Available: https://arxiv.org/abs/2010.11929

[22] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[23] J. L. Elman, "Finding structure in time," Cognitive science, vol. 14, no. 2, pp. 179–211, 1990.

[24] M. M. Bronstein, J. Bruna, T. Cohen, and P. Veličković, "Geometric deep learning: Grids, groups, graphs, geodesics, and gauges," arXiv preprint arXiv:2104.13478, 2021.

[25] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 10 684–10 695.

[26] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in International conference on machine learning. PmLR, 2020, pp. 1597–1607.

[27] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," 2018. [Online]. Available: https://arxiv.org/abs/1706.08500

[28] M. Bińkowski, D. J. Sutherland, M. Arbel, and A. Gretton, "Demystifying mmd gans," 2021. [Online]. Available: https://arxiv.org/abs/1801.01401

[29] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600–612, 2004.

[30] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," 2016. [Online]. Available: https://arxiv.org/abs/1606.03498

[31] R. C. Gonzalez and R. E. Woods, Digital Image Processing. Prentice Hall, 2002.

[32] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "Feature similarity index for image quality assessment," IEEE Transactions on Image Processing, vol. 20, no. 8, pp. 2378–2386, 2011.

[33] S. Mallat, A Wavelet Tour of Signal Processing: The Sparse Way. Academic press, 2008.

[34] Z. Wang and A. C. Bovik, "A universal image quality index," IEEE Signal Processing Letters, vol. 9, no. 3, pp. 81–84, 2002.

[35] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, "Learning transferable visual models from natural language supervision," 2021. [Online]. Available: https://arxiv.org/abs/2103.00020

[36] C.-S. Chen, "Necomimi: Neural-cognitive multimodal eeg-informed image generation with diffusion models," arXiv preprint arXiv:2410.00712, 2024.

[37] D. Li, C. Wei, S. Li, J. Zou, H. Qin, and Q. Liu, "Visual decoding and reconstruction via eeg embeddings with guided diffusion," 2024. [Online]. Available: https://arxiv.org/abs/2403.07721

[38] P. Singh, P. Pandey, K. Miyapuram, and S. Raman, "Eeg2image: image reconstruction from eeg brain signals," in ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023, pp. 1–5.

[39] P. Singh, D. Dalal, G. Vashishtha, K. Miyapuram, and S. Raman, "Learning robust deep visual representations from eeg brain recordings," in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2024, pp. 7553–7562.

[40] L. Wang, C. Wu, and L. Wang, "Braindreamer: Reasoning-coherent and controllable image generation from eeg brain signals via language guidance," arXiv preprint arXiv:2409.14021, 2024.

[41] S. Khare, R. N. Choubey, L. Amar, and V. Udutalapalli, "Neurovision: perceived image regeneration using cprogan," Neural Computing and Applications, vol. 34, no. 8, pp. 5979–5991, 2022.

[42] G. Yang and J. Liu, "A new framework combining diffusion models and the convolution classifier for generating images from eeg signals," Brain Sciences, vol. 14, no. 5, p. 478, 2024.

[43] H. Zeng, N. Xia, D. Qian, M. Hattori, C. Wang, and W. Kong, "Dm-re2i: A framework based on diffusion model for the reconstruction from eeg to image," Biomedical Signal Processing and Control, vol. 86, p. 105125, 2023.

[44] S. Mehta, A. Kumar, F. Reda, V. Nasery, V. Mulukutla, R. Ranjan, and V. Chandra, "Evrnet: Efficient video restoration on edge devices," 2020. [Online]. Available: https://arxiv.org/abs/2012.02228

[45] M. Ferrante, T. Boccato, S. Bargione, and N. Toschi, "Decoding eeg signals of visual brain representations with a clip based knowledge distillation," in ICLR 2024 Workshop on Learning from Time Series For Health, 2024.

[46] N. Kumari, S. Anwar, V. Bhattacharjee, and S. K. Sahana, "Visually evoked brain signals guided image regeneration using gan variants," Multimedia Tools and Applications, vol. 82, no. 21, pp. 32 259–32 279, 2023.

[47] H. Fu, H. Wang, J. J. Chin, and Z. Shen, "Brainvis: Exploring the bridge between brain and visual signals via image reconstruction," in ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2025, pp. 1–5.

[48] Y. Bai, X. Wang, Y.-p. Cao, Y. Ge, C. Yuan, and Y. Shan, "Dreamdiffusion: Generating high-quality images from brain eeg signals," arXiv preprint arXiv:2306.16934, 2023.

[49] P. Kumar, R. Saini, P. P. Roy, P. K. Sahu, and D. P. Dogra, "Envisioned speech recognition using eeg sensors," Personal and Ubiquitous Computing, vol. 22, pp. 185–199, 2018.

[50] E. Lopez, L. Sigillo, F. Colonnese, M. Panella, and D. Comminiello, "Guess what i think: Streamlined eeg-to-image generation with latent diffusion models," in ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2025, pp. 1–5.

[51] L. Zhang, A. Rao, and M. Agrawala, "Adding conditional control to text-to-image diffusion models," in Proceedings of the IEEE/CVF international conference on computer vision, 2023, pp. 3836–3847.

[52] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila, "Training generative adversarial networks with limited data," Advances in neural information processing systems, vol. 33, pp. 12 104–12 114, 2020.

[53] S. Palazzo, C. Spampinato, I. Kavasidis, D. Giordano, J. Schmidt, and M. Shah, "Decoding brain representations by multimodal learning of neural activity and visual features," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 11, pp. 3833–3849, 2020.

[54] I. Kavasidis, S. Palazzo, C. Spampinato, D. Giordano, and M. Shah, "Brain2image: Converting brain signals into images," in Proceedings of the 25th ACM international conference on Multimedia, 2017, pp. 1809–1817.

[55] C. Spampinato, S. Palazzo, I. Kavasidis, D. Giordano, N. Souly, and M. Shah, "Deep learning human mind for

automated visual classification," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 6809–6817.
[56] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," Advances in neural information processing systems, vol. 33, pp. 6840–6851, 2020.
[57] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," arXiv preprint arXiv:1710.10196, 2017.
[58] M. Ferrante, T. Boccato, S. Bargione, and N. Toschi, "Decoding visual brain representations from electroencephalography through knowledge distillation and latent diffusion models," Computers in Biology and Medicine, vol. 178, p. 108701, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482524007868
[59] E. Perez, F. Strub, H. De Vries, V. Dumoulin, and A. Courville, "Film: Visual reasoning with a general conditioning layer," in Proceedings of the AAAI conference on artificial intelligence, vol. 32, no. 1, 2018.
[60] D. Vivancos and F. Cuesta, "Mindbigdata 2022 a large dataset of brain signals," 2022. [Online]. Available: https://arxiv.org/abs/2212.14746
[61] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," Advances in neural information processing systems, vol. 30, 2017.
[62] M. N. Hebart, A. H. Dickter, A. Kidder, W. Y. Kwok, A. Corriveau, C. Van Wicklin, and C. I. Baker, "Things: A database of 1,854 object concepts and more than 26,000 naturalistic object images," PloS one, vol. 14, no. 10, p. e0223792, 2019.
[63] D. Podell, Z. English, K. Lacey, A. Blattmann, T. Dockhorn, J. Müller, J. Penna, and R. Rombach, "Sdxl: Improving latent diffusion models for high-resolution image synthesis," arXiv preprint arXiv:2307.01952, 2023.
[64] D. Li, H. Qin, M. Wu, J. Tang, Y. Cao, C. Wei, and Q. Liu, "Realmind: Advancing visual decoding and language interaction via eeg signals," arXiv preprint arXiv:2410.23754, 2024.
[65] H. Shimizu and R. Srinivasan, "Improving classification and reconstruction of imagined images from eeg signals," Plos one, vol. 17, no. 9, p. e0274847, 2022.
[66] T. A. Izzuddin, N. M. Safri, and M. A. Othman, "Compact convolutional neural network (cnn) based on sincnet for end-to-end motor imagery decoding and analysis," Biocybernetics and Biomedical Engineering, vol. 41, no. 4, pp. 1629–1645, 2021.
[67] R. Mishra, K. Sharma, R. R. Jha, and A. Bhavsar, "Neurogan: image reconstruction from eeg signals via an attention-based gan," Neural Computing and Applications, vol. 35, no. 12, pp. 9181–9192, 2023.
[68] Z. Jiao, H. You, F. Yang, X. Li, H. Zhang, and D. Shen, "Decoding eeg by visual-guided deep neural networks." in IJCAI, vol. 28. Macao, 2019, pp. 1387–1393.
[69] R. Mishra and A. Bhavsar, "Generating visual stimuli from eeg recordings using transformer-encoder based eeg encoder and gan," arXiv preprint arXiv:2402.10115, 2024.
[70] N. Khaleghi, T. Y. Rezaii, S. Beheshti, S. Meshgini, S. Sheykhivand, and S. Danishvar, "Visual saliency and image reconstruction from eeg signals via an effective geometric deep network-based generative adversarial network," Electronics, vol. 11, no. 21, p. 3637, 2022.
[71] G. E. Hinton, Z. Ghahramani, and Y. W. Teh, "Learning to parse images," Advances in neural information processing systems, vol. 12, 1999.
[72] G. E. Hinton, A. Krizhevsky, and S. D. Wang, "Transforming auto-encoders," in Artificial Neural Networks and Machine Learning–ICANN 2011: 21st International Conference on Artificial Neural Networks, Espoo, Finland, June 14-17, 2011, Proceedings, Part I 21. Springer, 2011, pp. 44–51.
[73] B. Aristimunha, I. Carrara, P. Guetschel, S. Sedlar, P. Rodrigues, J. Sosulski, D. Narayanan, E. Bjareholt, Q. Barthelemy, R. T. Schirrmeister, R. Kobler, E. Kalunga, L. Darmet, C. Gregoire, A. Abdul Hussain, R. Gatti, V. Goncharenko, J. Thielen, T. Moreau, Y. Roy, V. Jayaram, A. Barachant, and S. Chevallier, "Mother of all bci benchmarks," 2025. [Online]. Available: https://github.com/NeuroTechX/moabb
[74] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 11 976–11 986.
[75] X. Zhao and Y. Guo, "Dual axatgan: A feature intregrate bci model for image reconstruction," in 2023 13th International Conference on Information Technology in Medicine and Education (ITME). IEEE, 2023, pp. 135–139.
[76] J. Ho, N. Kalchbrenner, D. Weissenborn, and T. Salimans, "Axial attention in multidimensional transformers," arXiv preprint arXiv:1912.12180, 2019.
[77] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in International conference on machine learning. PMLR, 2017, pp. 214–223.
[78] Y. Liu, Y. Ma, W. Zhou, G. Zhu, and N. Zheng, "Brainclip: Bridging brain and visual-linguistic representation via clip for generic natural visual stimulus decoding," arXiv preprint arXiv:2302.12971, 2023.
[79] M. Ferrante, T. Boccato, L. Passamonti, and N. Toschi, "Retrieving and reconstructing conceptually similar images from fmri with latent diffusion models and a neuro-inspired brain decoding model," Journal of Neural Engineering, vol. 21, no. 4, p. 046001, 2024.
[80] Y.-T. Lan, K. Ren, Y. Wang, W.-L. Zheng, D. Li, B.-L. Lu, and L. Qiu, "Seeing through the brain: image reconstruction of visual perception from human brain signals," arXiv preprint arXiv:2308.02510, 2023.
[81] L. Chang and D. Y. Tsao, "The code for facial identity in the primate brain," Cell, vol. 169, no. 6, pp. 1013–1028, 2017.
[82] Z. Lu, "Visualizing the mind's eye: a future perspective on applications of image reconstruction from brain signals to psychiatry," Psychoradiology, vol. 3, p. kkad022, 2023.
[83] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009, pp. 248–255.
[84] A. T. Gifford, K. Dwivedi, G. Roig, and R. M. Cichy, "A large and rich eeg dataset for modeling human visual object recognition," NeuroImage, vol. 264, p. 119754, 2022.
[85] T. Grootswagers, I. Zhou, A. Robinson, M. Hebart, and T. Carlson, "Human eeg recordings for 1,854 concepts presented in rapid serial visual presentation streams. sci. data 9, 3," 2022.
[86] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in Computer vision–ECCV 2014: 13th European conference, zurich, Switzerland, September 6-12, 2014, proceedings, part v 13. Springer, 2014, pp. 740–755.
[87] R. VanRullen and L. Reddy, "Reconstructing faces from fmri patterns using deep generative neural networks," Communications biology, vol. 2, no. 1, p. 193, 2019.
[88] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in Proceedings of the IEEE international conference on computer vision, 2015, pp. 3730–3738.
[89] Y. Miyawaki, H. Uchida, O. Yamashita, M.-a. Sato, Y. Morito, H. C. Tanabe, N. Sadato, and Y. Kamitani, "Visual image reconstruction from human brain activity using a combination of multiscale local image decoders," Neuron, vol. 60, no. 5, pp. 915–929, 2008.
[90] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. Behrens, E. Yacoub, K. Ugurbil, W.-M. H. Consortium et al., "The wu-minn human connectome project: an overview," Neuroimage, vol. 80, pp. 62–79, 2013.
[91] T. Horikawa and Y. Kamitani, "Generic decoding of seen and imagined objects using hierarchical visual features," Nature communications, vol. 8, no. 1, p. 15037, 2017.
[92] N. Chang, J. A. Pyles, A. Marcus, A. Gupta, M. J. Tarr, and E. M. Aminoff, "Bold5000, a public fmri dataset while viewing 5000 visual images," Scientific data, vol. 6, no. 1, p. 49, 2019.

[93] K. N. Kay, T. Naselaris, R. J. Prenger, and J. L. Gallant, "Identifying natural images from human brain activity," Nature, vol. 452, no. 7185, pp. 352–355, 2008.

[94] G. Shen, T. Horikawa, K. Majima, and Y. Kamitani, "Deep image reconstruction from human brain activity," PLoS computational biology, vol. 15, no. 1, p. e1006633, 2019.

[95] N. Koide-Majima, S. Nishimoto, and K. Majima, "Mental image reconstruction from human brain activity," BiorXiv, pp. 2023–01, 2023.

[96] G. Shen, K. Dwivedi, K. Majima, T. Horikawa, and Y. Kamitani, "End-to-end deep image reconstruction from human brain activity," Frontiers in computational neuroscience, vol. 13, p. 21, 2019.

[97] T. Fang, Y. Qi, and G. Pan, "Reconstructing perceptive images from brain activity by shape-semantic gan," in Advances in Neural Information Processing Systems, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., vol. 33. Curran Associates, Inc., 2020, pp. 13 038–13 048. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2020/file/9813b270ed0288e7c0388f0fd4ec68f5-Paper.pdf

[98] L. Yang, H. Zhen, L. Li, Y. Li, H. Zhang, X. Xie, and R.-Y. Zhang, "Functional diversity of visual cortex improves constraint-free natural image reconstruction from human brain activity," Fundamental Research, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2667325823003059

[99] A. Lad and R. Patel, "Decoding with purpose: Improving image reconstruction from fmri with multitask learning," in 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON), Sep. 2021, pp. 1–6.

[100] F. Kalantari, K. Faez, H. Amindavar, and S. Nazari, "Improved image reconstruction from brain activity through automatic image captioning," Scientific Reports, vol. 15, no. 1, p. 4907, 2025.

[101] Y. Takagi and S. Nishimoto, "High-resolution image reconstruction with latent diffusion models from human brain activity," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 14 453–14 463.

[102] J. Huo, Y. Wang, Y. Wang, X. Qian, C. Li, Y. Fu, and J. Feng, "Neuropictor: Refining fmri-to-image reconstruction via multi-individual pretraining and multi-level modulation," in European Conference on Computer Vision. Springer, 2024, pp. 56–73.

[103] T. Fang, Q. Zheng, and G. Pan, "Alleviating the semantic gap for generalized fmri-to-image reconstruction," Advances in Neural Information Processing Systems, vol. 36, pp. 15 096–15 107, 2023.

[104] Z. Gu, K. W. Jamison, M. Khosla, E. J. Allen, Y. Wu, G. St-Yves, T. Naselaris, K. Kay, M. R. Sabuncu, and A. Kuceyeski, "Neurogen: Activation optimized image synthesis for discovery neuroscience," NeuroImage, vol. 247, p. 118812, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811921010831

[105] Y. Lu, C. Du, Q. Zhou, D. Wang, and H. He, "Minddiffuser: Controlled image reconstruction from human brain activity with semantic and structural diffusion," in Proceedings of the 31st ACM International Conference on Multimedia, 2023, pp. 5899–5908.

[106] M. Ferrante, T. Boccato, F. Ozcelik, R. VanRullen, and N. Toschi, "Multimodal decoding of human brain activity into images and text," in UniReps: the First Workshop on Unifying Representations in Neural Models, 2023.

[107] G. Shen, D. Zhao, X. He, L. Feng, Y. Dong, J. Wang, Q. Zhang, and Y. Zeng, "Neuro-vision to language: Enhancing brain recording-based visual reconstruction and language interaction," Advances in Neural Information Processing Systems, vol. 37, pp. 98 083–98 110, 2024.

[108] P. Scotti, A. Banerjee, J. Goode, S. Shabalin, A. Nguyen, A. Dempster, N. Verlinde, E. Yundler, D. Weisberg, K. Norman et al., "Reconstructing the mind's eye: fmri-to-image with contrastive learning and diffusion priors," Advances in Neural Information Processing Systems, vol. 36, pp. 24 705–24 728, 2023.

[109] H. Li, H. Wu, and B. Chen, "Neuraldiffuser: Neuroscience-inspired diffusion guidance for fmri visual reconstruction," IEEE Transactions on Image Processing, 2025.

[110] W. Mai and Z. Zhang, "Unibrain: Unify image reconstruction and captioning all in one diffusion model from human brain activity," arXiv preprint arXiv:2308.07428, 2023.

[111] W. Xia, R. De Charette, C. Oztireli, and J.-H. Xue, "Dream: Visual decoding from reversing human visual system," in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2024, pp. 8226–8235.

[112] J. Guo, C. Yi, F. Li, P. Xu, and Y. Tian, "Mindldm: Reconstruct visual stimuli from fmri using latent diffusion model," in 2024 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), June 2024, pp. 1–6.

[113] G. M. Balisacan and A. T. A. Paulo, "Neuro-vis: Guided complex image reconstruction from brain signals using multiple semantic and perceptual controls," in Proceedings of the International Conference on Computing, Machine Learning and Data Science, 2024, pp. 1–8.

[114] D. Xie, P. Zhao, J. Zhang, K. Wei, X. Ni, and J. Xia, "Brainram: Cross-modality retrieval-augmented image reconstruction from human brain activity," in Proceedings of the 32nd ACM International Conference on Multimedia, ser. MM '24. New York, NY, USA: Association for Computing Machinery, 2024, p. 3994–4003. [Online]. Available: https://doi-org.aus.idm.oclc.org/10.1145/3664647.3681296

[115] Z. Gong, Q. Zhang, G. Bao, L. Zhu, K. Liu, L. Hu, and D. Miao, "Mindtuner: Cross-subject visual decoding with visual fingerprint and semantic correction," 2024. [Online]. Available: https://arxiv.org/abs/2404.12630

[116] J. Wang, X. Han, J. Tang, S. Zhang, and X. Ji, "Rethinking brain-to-image reconstruction: What should we decode from fMRI signals?" 2024. [Online]. Available: https://openreview.net/forum?id=UUNTAwJIIn

[117] P. S. Scotti, M. Tripathy, C. K. T. Villanueva, R. Kneeland, T. Chen, A. Narang, C. Santhirasegaran, J. Xu, T. Naselaris, K. A. Norman, and T. M. Abraham, "Mindeye2: Shared-subject models enable fmri-to-image with 1 hour of data," 2024. [Online]. Available: https://arxiv.org/abs/2403.11207

[118] Z. Gu, K. Jamison, A. Kuceyeski, and M. Sabuncu, "Decoding natural image stimuli from fmri data with a surface-based convolutional network," 2023. [Online]. Available: https://arxiv.org/abs/2212.02409

[119] C. Zangos, D. Ebadulla, T. C. Sprague, and A. Singh, "Efficient multi subject visual reconstruction from fMRI using aligned representations," 2025. [Online]. Available: https://openreview.net/forum?id=z2QdVmhtAP

[120] Y. Xiong, W. Zhu, Z.-L. Lu, and Y. Wang, "Reconstructing retinal visual images from 3t fmri data enhanced by unsupervised learning," in 2024 IEEE International Symposium on Biomedical Imaging (ISBI), May 2024, pp. 1–5.

[121] W. B. Zhu, R. Z. Gui, Y. Q. Wang, Y. M. Yin, and M. S. Tong, "Pre-training deep neural network for decoding multiple brain regions can enhance image reconstruction," in 2024 Photonics Electromagnetics Research Symposium (PIERS), April 2024, pp. 1–6.

[122] Z. Ye, L. Yao, Y. Zhang, and S. Gustin, "See what you see: Self-supervised cross-modal retrieval of visual stimuli from brain activity," 2022. [Online]. Available: https://arxiv.org/abs/2208.03666

[123] Y. Takagi and S. Nishimoto, "Improving visual image reconstruction from human brain activity using latent diffusion models via multiple decoded inputs," 2023. [Online]. Available: https://arxiv.org/abs/2306.11536

[124] S. Wang, S. Liu, Z. Tan, and X. Wang, "Mindbridge: A cross-subject brain decoding framework," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2024, pp. 11 333–11 342.

[125] D. Zongxin and L. Lin, "Braintransformer: Subject-wise patch embed transformer for cross-subject brain visual information decoding," in 2024 21st International Computer Conference on

Wavelet Active Media Technology and Information Processing (ICCWAMTIP), Dec 2024, pp. 1–4.
[126] T. Fei, A. Uppal, I. Jackson, S. Ravishankar, D. Wang, and V. R. de Sa, "Perceptogram: Reconstructing visual percepts from eeg," 2025. [Online]. Available: https://arxiv.org/abs/2404.01250
[127] M. Ferrante, T. Boccato, F. Ozcelik, R. VanRullen, and N. Toschi, "Generative multimodal decoding: Reconstructing images and text from human fMRI," in Deep Generative Models for Health Workshop NeurIPS 2023, 2023. [Online]. Available: https://openreview.net/forum?id=XUgIZQvxg4
[128] J. Park, P. Y. Kim, J. Cha, S. Yoo, and T. Moon, "Seed: Towards more accurate semantic evaluation for visual brain decoding," 2025. [Online]. Available: https://arxiv.org/abs/2503.06437
[129] Z. Ren, J. Li, X. Xue, X. Li, F. Yang, Z. Jiao, and X. Gao, "Reconstructing seen image from brain activity by visually-guided cognitive representation and adversarial learning," NeuroImage, vol. 228, p. 117602, 2021.
[130] P.-C. Chang, Y.-Y. Tien, C.-L. Chen, L.-F. Chen, Y.-S. Chen, and H.-L. Chan, "Facial image reconstruction from functional magnetic resonance imaging via gan inversion with improved attribute consistency," in 2022 International Joint Conference on Neural Networks (IJCNN), July 2022, pp. 1–8.
[131] Z. Chen, J. Qing, T. Xiang, W. L. Yue, and J. H. Zhou, "Seeing beyond the brain: Conditional diffusion model with sparse masked modeling for vision decoding," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 22 710–22 720.
[132] J. Sun, M. Li, and M.-F. Moens, "Decoding realistic images from brain activity with contrastive self-supervision and latent diffusion," 2023. [Online]. Available: https://arxiv.org/abs/2310.00318
[133] M. Mozafari, L. Reddy, and R. VanRullen, "Reconstructing natural scenes from fmri patterns using bigbigan," in 2020 International joint conference on neural networks (IJCNN). IEEE, 2020, pp. 1–8.
[134] F. Ozcelik and R. VanRullen, "Natural scene reconstruction from fmri signals using generative latent diffusion," Scientific Reports, vol. 13, no. 1, p. 15666, 2023.
[135] M. Ferrante, T. Boccato, and N. Toschi, "Semantic brain decoding: from fmri to conceptually similar image reconstruction of visual stimuli," arXiv preprint arXiv:2212.06726, 2022.
[136] L. Meng and C. Yang, "Dual-guided brain diffusion model: Natural image reconstruction from human visual stimulus fmri," Bioengineering, vol. 10, no. 10, 2023. [Online]. Available: https://www.mdpi.com/2306-5354/10/10/1117
[137] R. Beliy, G. Gaziv, A. Hoogi, F. Strappini, T. Golan, and M. Irani, "From voxels to pixels and back: Self-supervision in natural-image reconstruction from fmri," in Advances in Neural Information Processing Systems, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32. Curran Associates, Inc., 2019. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2019/file/7d2be41b1bde6ff8fe45150c37488ebb-Paper.pdf
[138] L. Meng and C. Yang, "Semantics-guided hierarchical feature encoding generative adversarial network for visual image reconstruction from brain activity," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 32, pp. 1267–1283, 2024.
[139] G. Gaziv, R. Beliy, N. Granot, A. Hoogi, F. Strappini, T. Golan, and M. Irani, "Self-supervised natural image reconstruction and large-scale semantic classification from brain activity," NeuroImage, vol. 254, p. 119121, 2022.
[140] F. Ozcelik, B. Choksi, M. Mozafari, L. Reddy, and R. VanRullen, "Reconstruction of perceived images from fmri patterns and semantic brain exploration using instance-conditioned gans," in 2022 International Joint Conference on Neural Networks (IJCNN), 2022, pp. 1–8.
[141] M. Kuang, Z. Zhan, and S. Gao, "Natural image reconstruction from fmri based on node–edge interaction and multi–scale constraint," Brain Sciences, vol. 14, no. 3, 2024. [Online]. Available: https://www.mdpi.com/2076-3425/14/3/234
[142] P. Ni and Y. Zhang, "Natural image reconstruction from fmri based on self-supervised representation learning and latent diffusion model," in Proceedings of the 15th International Conference on Digital Image Processing, 2023, pp. 1–9.
[143] K. Qiao, J. Chen, L. Wang, C. Zhang, L. Tong, and B. Yan, "Reconstructing natural images from human fmri by alternating encoding and decoding with shared autoencoder regularization," Biomedical Signal Processing and Control, vol. 73, p. 103397, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809421009940
[144] J. Sun, M. Li, Z. Chen, Y. Zhang, S. Wang, and M.-F. Moens, "Contrast, attend and diffuse to decode high-resolution images from brain activities," in Advances in Neural Information Processing Systems, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, Eds., vol. 36. Curran Associates, Inc., 2023, pp. 12 332–12 348. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2023/file/28dad4a70f748a2980998d3ed0f1b8d2-Paper-Conference.pdf
[145] L. Meng and C. Yang, "Semantics-guided hierarchical feature encoding generative adversarial network for natural image reconstruction from brain activities," in 2023 International Joint Conference on Neural Networks (IJCNN), 2023, pp. 1–9.
[146] B. Zeng, S. Li, X. Liu, S. Gao, X. Jiang, X. Tang, Y. Hu, J. Liu, and B. Zhang, "Controllable mind visual diffusion model," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, no. 7, 2024, pp. 6935–6943.
[147] Z. Zhao, H. Jing, J. Wang, W. Wu, and Y. Ma, "Images structure reconstruction from fmri by unsupervised learning based on vae," in International Conference on Artificial Neural Networks. Springer, 2022, pp. 137–148.
[148] J. Chen, Y. Qi, and G. Pan, "Rethinking visual reconstruction: Experience-based content completion guided by visual cues," in Proceedings of the 40th International Conference on Machine Learning, ser. Proceedings of Machine Learning Research, A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, and J. Scarlett, Eds., vol. 202. PMLR, 23–29 Jul 2023, pp. 4856–4866. [Online]. Available: https://proceedings.mlr.press/v202/chen23v.html
[149] K. Chen, Y. Ma, M. Sheng, and N. Zheng, "Foreground-attention in neural decoding: Guiding loop-enc-dec to reconstruct visual stimulus images from fmri," in 2022 International Joint Conference on Neural Networks (IJCNN), July 2022, pp. 1–8.
[150] X. Qian, Y. Wang, X. Sun, Y. Fu, X. Xue, and J. Feng, "LEA: Learning latent embedding alignment model for fMRI decoding and encoding," Transactions on Machine Learning Research, 2024. [Online]. Available: https://openreview.net/forum?id=89QT2DsKyj
[151] Z. Yu, K. Qiao, C. Zhang, L. Wang, and B. Yan, "End-to-end image reconstruction of image from human functional magnetic resonance imaging based on the "language" of visual cortex," in Proceedings of the 2020 6th International Conference on Computing and Artificial Intelligence, ser. ICCAI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 176–181. [Online]. Available: https://doi-org.aus.idm.oclc.org/10.1145/3404555.3404593
[152] Y. Lin, J. Li, and H. Wang, "Dcnn-gan: Reconstructing realistic image from fmri," in 2019 16th International Conference on Machine Vision Applications (MVA), May 2019, pp. 1–6.
[153] Q. Li, "Visual image reconstructed without semantics from human brain activity using linear image decoders and nonlinear noise suppression," Cognitive Neurodynamics, vol. 19, no. 1, p. 20, 2025.
[154] C. Zhang, K. Qiao, L. Wang, L. Tong, Y. Zeng, and B. Yan, "Constraint-free natural image reconstruction from fmri signals based on convolutional neural network," Frontiers in Human Neuroscience, vol. Volume 12 - 2018, 2018. [Online]. Available: https://www.frontiersin.org/journals/human-neuroscience/articles/10.3389/fnhum.2018.00242
[155] K. Qiao, J. Chen, L. Wang, C. Zhang, L. Tong, and B. Yan, "Biggan-based bayesian reconstruction of natural images from human brain activity," Neuroscience, vol. 444, pp. 92–105, 2020.

[156] K. Bhargav, S. Ambika, S. Deepak, and S. Sudha, "Imagenation-a dcgan based method for image reconstruction from fmri," in 2020 fifth international conference on research in computational intelligence and communication networks (ICRCICN). IEEE, 2020, pp. 112–119.
[157] R. Coen-Cagli, A. Kohn, and O. Schwartz, "Flexible gating of contextual influences in natural vision," Nature neuroscience, vol. 18, no. 11, pp. 1648–1655, 2015.
[158] Y. Benchetrit, H. Banville, and J.-R. King, "Brain decoding: toward real-time reconstruction of visual perception," arXiv preprint arXiv:2310.19812, 2023.
[159] N. E. Mughal, M. J. Khan, K. Khalil, K. Javed, H. Sajid, N. Naseer, U. Ghafoor, and K.-S. Hong, "Eeg-fnirs-based hybrid image construction and classification using cnn-lstm," Frontiers in Neurorobotics, vol. 16, p. 873239, 2022.
[160] S. A. Cadena, G. H. Denfield, E. Y. Walker, L. A. Gatys, A. S. Tolias, M. Bethge, and A. S. Ecker, "Deep convolutional models improve predictions of macaque v1 responses to natural images," PLoS computational biology, vol. 15, no. 4, p. e1006897, 2019.
[161] S. Safarani, A. Nix, K. Willeke, S. Cadena, K. Restivo, G. Denfield, A. Tolias, and F. Sinz, "Towards robust vision by multi-task learning on monkey visual cortex," Advances in Neural Information Processing Systems, vol. 34, pp. 739–751, 2021.
[162] R. Saqur and S. Vivona, "Capsgan: Using dynamic routing for generative adversarial networks," 2018. [Online]. Available: https://arxiv.org/abs/1806.03968
[163] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014. [Online]. Available: https://arxiv.org/abs/1411.1784
[164] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2016. [Online]. Available: https://arxiv.org/abs/1511.06434
[165] J. Donahue, P. Krähenbühl, and T. Darrell, "Adversarial feature learning," 2017. [Online]. Available: https://arxiv.org/abs/1605.09782
[166] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier gans," 2017. [Online]. Available: https://arxiv.org/abs/1610.09585
[167] Z. Tian, R. Quan, F. Ma, K. Zhan, and Y. Yang, "Brainguard: Privacy-preserving multisubject image reconstructions from brain activities," arXiv preprint arXiv:2501.14309, 2025.
[168] T. Šarčević, A. Karlowicz, R. Mayer, R. Baeza-Yates, and A. Rauber, "U can't gen this? a survey of intellectual property protection methods for data in generative ai," 2024. [Online]. Available: https://arxiv.org/abs/2406.15386