Project Proposal

Kevin Huang (zhuang60), Xu Lian (xlian3),
Zhuoxuan Li (zli256), Benjamin Shan(bshan1)

Reconstructing Images from fMRI Recordings

Problem Area

Understanding how the human brain processes visual stimuli is a fundamental challenge in neuroscience and artificial intelligence. Reconstructing images from functional Magnetic Resonance Imaging (fMRI) signals presents a unique opportunity to decode human visual perception, contributing to brain-computer interfaces (BCIs), medical imaging, and cognitive neuroscience. However, this task is highly complex due to the high dimensionality and noise in fMRI data, requiring sophisticated machine learning models to translate neural activity into interpretable visual representations. Recent advances in deep learning offer promising solutions, yet many existing models struggle with reconstructing high-fidelity images from fMRI recordings due to limited training data and inherent signal variability.

Context

Several recent studies have made progress in reconstructing images from fMRI data using deep learning techniques. One notable approach, **SelfSuperReconst** (WeizmannVision), leverages self-supervised learning to enhance feature extraction from brain signals, improving model generalization. Another study (Takagi & Nishimoto, 2023) demonstrates the use of **Stable Diffusion** for reconstructing high-quality images from fMRI data, showing that latent diffusion models outperform GANs in producing realistic outputs. Furthermore, recent work (Liu et al., 2023) explores Vision Transformers (ViTs) for fMRI decoding, highlighting their ability to capture long-range dependencies in brain signals. Additionally, research from IEEE (Ozcelik et al., 2023) investigates the use of **cross-subject generalization techniques**, addressing the challenge of model robustness across different individuals. While these methods have demonstrated promising results, challenges remain in improving reconstruction fidelity, reducing computational complexity, and ensuring better alignment between neural activity and visual perception.

Data

This project will utilize the **Natural Scenes Dataset (NSD)**, a large-scale dataset that contains high-resolution fMRI recordings from human participants viewing thousands of natural scene images. The NSD provides detailed voxel-wise brain activity measurements corresponding to specific visual stimuli, making it an ideal dataset for training deep learning models. More information on accessing the dataset is available on the official website: NSD Dataset. Preprocessing will involve normalizing fMRI signals, dimensionality reduction using Principal Component Analysis (PCA), and aligning brain responses across subjects to enhance model performance.

Proposed Solution

We propose a deep learning-based pipeline that integrates a **Hybrid Transformer-Diffusion Model** for effective fMRI-to-image reconstruction. Instead of treating feature extraction and image generation as separate stages, we aim to build a unified model that combines the advantages of Vision Transformers (ViTs) and Latent Diffusion Models (LDMs). The **ViT-based encoder** will first process volumetric fMRI data, capturing long-range dependencies and mapping neural signals to a compact latent space. This latent representation will then be fed into a **Latent Diffusion Model**, which iteratively refines and reconstructs the corresponding image using a learned denoising process. The diffusion model will be trained using a combination of **Mean Squared Error (MSE) loss**, **perceptual loss**, and **adversarial loss** to ensure both pixel accuracy and perceptual realism. Compared to traditional GANs, diffusion models offer higher fidelity and better capture fine-grained details, making them well-suited for this task. Evaluation metrics will include the **Structural Similarity Index (SSIM)**, **Peak Signal-to-Noise Ratio (PSNR)**, and **Learned Perceptual Image Patch Similarity (LPIPS)** to assess reconstruction quality.