Semantic Alignment of fMRI Data Using CLIP

1. Introduction

The objective of this project is to explore and understand the fMRI dataset from the BOLD5000 project and implement a CLIP-based contrastive tuning approach. The goal is to align the fMRI data (brain activity) with semantic information derived from images. This alignment will allow us to investigate how brain activity patterns correspond to various stimuli, such as images from different datasets (COCO, ImageNet, Scene).

In this report, we document the following steps:

- Data Understanding: Conducted an initial analysis of the fMRI dataset to understand its structure and identify any challenges.
- Semantic Alignment: Implemented contrastive tuning using the CLIP model to align brain response data with image embeddings.

2. Data Understanding

2.1 Data Loading and Structure

We selected **Release 1.0** of the **BOLD5000** dataset, which provides extensive fMRI data along with corresponding visual stimuli. For our analysis, we chose the **BOLD5000_ROIs** subset, specifically focusing on a sample dataset from the **CSI1 session 1**.

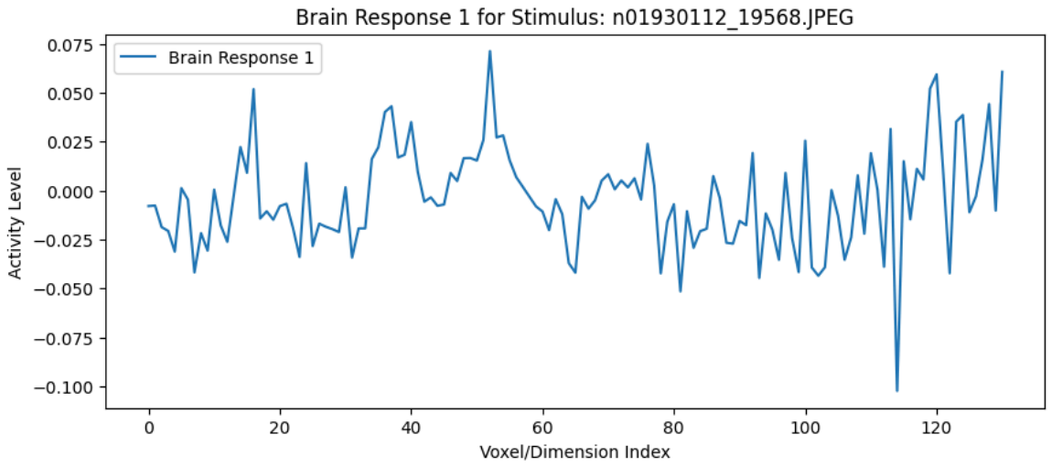
We began by loading the **CSI1_ROIs_TR1.mat** file, which contains fMRI brain response data for various regions of interest (ROIs). Specifically, we analyzed data from the **Left Hemisphere Parahippocampal Place Area (LHPPA)**, which is associated with scene recognition in the brain. The data matrix contains brain activity measurements (voxel data) for each stimulus presented during the fMRI session.

- **Shape of LHPPA data:** (5254, 131), indicating 5254 brain responses with 131 voxel measurements each.

We also loaded the corresponding stimulus list from **CSI01_stim_lists.txt**, which provides the filenames of the visual stimuli presented during the experiment. We confirmed that the number of stimuli matches the number of brain responses, ensuring correct alignment of the data for further analysis.

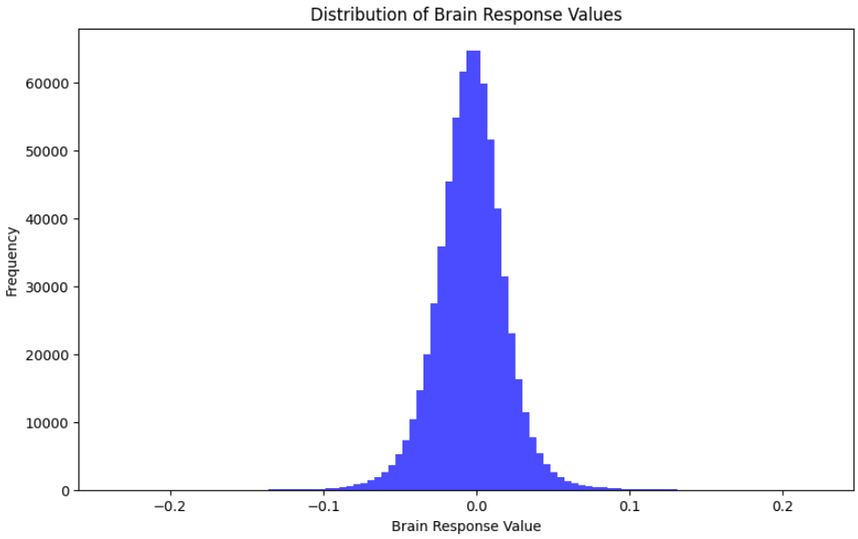
2.2 Visualization and Statistical Analysis

We visualized individual brain responses for the first few stimuli using line plots. These plots allowed us to compare the activity levels across different voxels for each stimulus.



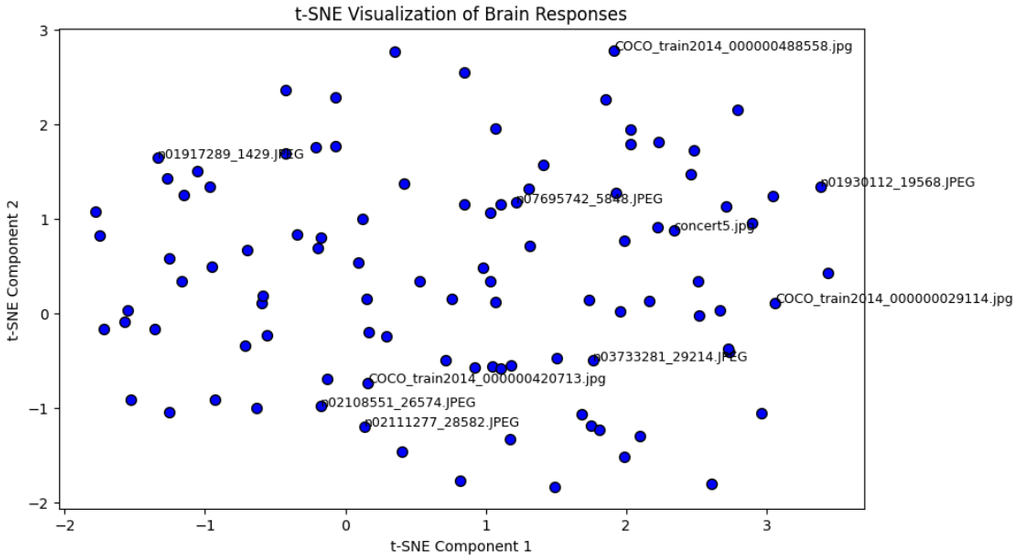
Additionally, we performed statistical analysis on the brain responses to understand their overall distribution:

- **Mean and standard deviation** were computed across all brain responses.
- We generated histograms to visualize the distribution of voxel activity levels.



2.3 Dimensionality Reduction

To explore the underlying structure of the brain response data, we applied **t-SNE** (t-distributed stochastic neighbor embedding) for dimensionality reduction. This reduced the brain response data to a 2D space, allowing us to visualize the high-dimensional brain activity in a more interpretable format.

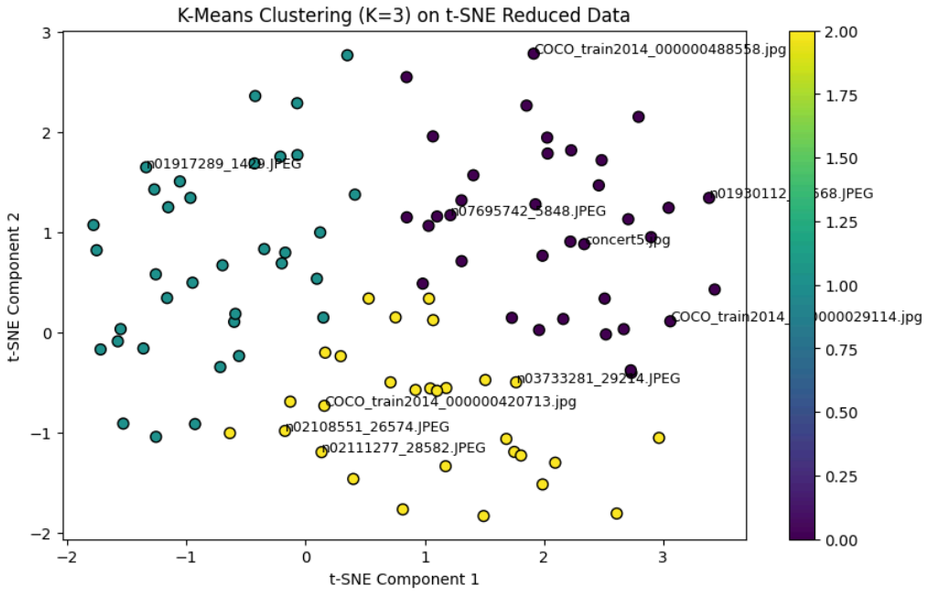


2.4 Clustering

We applied **K-Means clustering** to the t-SNE-reduced data to identify potential clusters in brain activity patterns. These clusters may correspond to different categories of stimuli (e.g., COCO dataset, Imagenet dataset,SUN dataset). We visualized the clusters and analyzed the distribution of stimuli across different clusters.

Key insights from this analysis:

- We observed distinct clusters of brain responses, suggesting that the brain may encode different types of stimuli in separate neural patterns.
- Some stimuli from the same dataset (e.g., COCO) tended to cluster together, indicating shared brain activity patterns for similar visual inputs.



3. Semantic Alignment

3.1 CLIP-Based Contrastive Tuning

After understanding the dataset, we proceeded to implement a CLIP-based contrastive tuning approach. The key idea was to align the brain response data with image embeddings generated using the pre-trained CLIP model. We focused on image data for this task.

- **Image Embeddings:** We precomputed image embeddings for the stimuli using the CLIP model.
- **Brain Response Projection:** We designed a **projection head** to map the brain responses into the same embedding space as the CLIP image embeddings.

3.2 Projection Head and Contrastive Loss

We experimented with two types of projection heads:

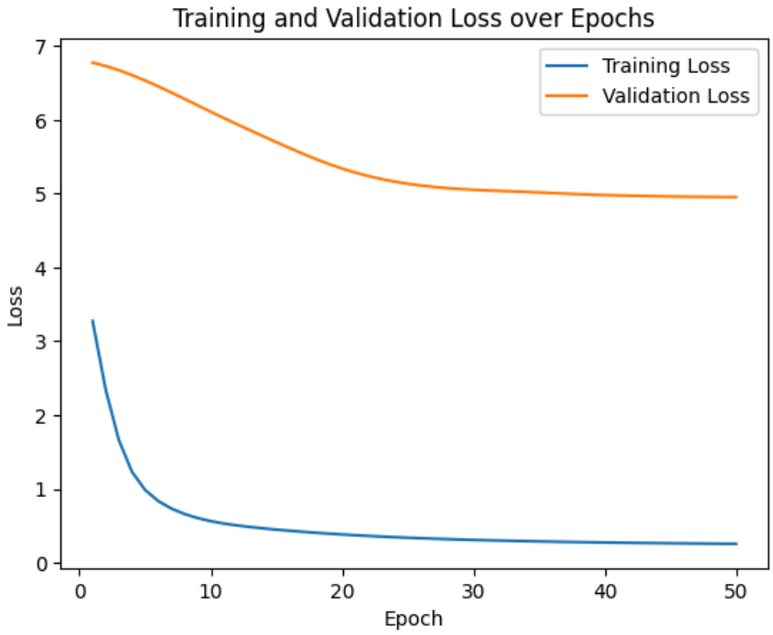
- **AdvancedProjectionHead:** A custom projection head with fully connected layers and attention mechanisms.
- **Attention-Based Projection Head:** We explored using attention blocks from CLIP-MUSED to improve the performance of the projection.

We used **InfoNCE Loss** for contrastive learning, aiming to minimize the distance between brain response projections and the corresponding image embeddings, while maximizing the distance to non-corresponding embeddings.

3.3 Training and Validation

We trained the model using an Adam optimizer with weight decay, and monitored the training and validation losses to ensure proper convergence. The training process was stopped early if no improvement was observed in validation loss for 20 consecutive epochs.

- **Number of epochs:** 50
- **Training loss:** The model converged to a low training loss, indicating that the brain responses were successfully mapped to the image embedding space.
- **Validation loss:** Although validation loss dropped drastically, it began to increase very slightly over each epoch, suggesting some degree of overfitting.



4. Challenges and Insights

Challenges Faced:

- **High Dimensionality:** The fMRI data is high-dimensional, which posed challenges during both the clustering and contrastive learning stages. Dimensionality reduction techniques like t-SNE were crucial for visualizing and understanding the data.
- **Alignment of Brain Responses and Stimuli:** Ensuring the correct alignment of brain responses and stimuli was critical for accurate training. We handled this by verifying the dataset structure and performing checks during data loading.
- **Overfitting:** As observed in the validation loss curve, the model showed signs of overfitting after a certain number of epochs. To address this, we introduced regularization techniques (e.g., weight decay, noise injection) and applied early stopping.

Novel Approaches:

- **Attention Mechanisms:** Incorporating attention blocks in the projection head architecture allowed us to capture complex interactions in the brain response data. This improved the model's ability to align brain responses with image embeddings.
- **Noise Injection:** Adding random noise to the brain data during training helped improve model robustness and reduce overfitting at certain levels.