

EEG2fMRI

Paul Bricman, Jelmer Borst

Tackled Problem

Current techniques for recording neural activity are limited. EEG (based on electrodes placed on the subject’s scalp) is a neuroimaging technique with high temporal resolution (sampling frequency), yet poor spatial resolution (due to the signal being obstructed and filtered by the brain and the skull). In contrast, fMRI is a technique with high spatial resolution (due to its reliance on electromagnetic waves which penetrate the brain), yet poor temporal resolution. However, both EEG and fMRI are used to record the same thing: neural activity. It’s plausible that EEG signal could be used to somehow “triangulate” neural activity at the precise locations captured by fMRI, at least in brain regions located near the scalp. Essentially, a machine learning model could be used to approximately reconstruct fMRI samples based on EEG samples. Multi-modal neuroimaging studies provide a unique opportunity for exploring this mapping. Those studies are based on simultaneous recordings with both EEG and fMRI, providing a perfectly aligned cross-modal signal which a machine learning model could use for training. A model capable of supporting this mapping would greatly augment the spatial resolution which can be achieved with EEG.

A metaphor might be useful for understanding the problem being tackled. Neural activity is akin to a subject being photographed. In this analogy, EEG is like an action camera with extremely poor resolution, yet which manages to record hundreds of frames per second. Conversely, fMRI is like a vintage film-based camera. It takes a long time for the subject figure to get imprinted on the film, and recording multiple frames is very tedious, yet the resolution of the individual photograph is superb. Training a model to map EEG to fMRI data is like training a model to map the action camera recording to the sequence of film-based photographs.

Implementation

We found a dataset from a multi-modal neuroimaging study from which we derived data points with the following structure. The input part consists of a 30-second-long EEG recording across 34 electrodes at 1000Hz. This data resembles 34 parallel time series unfolding across 30 seconds. Each time series comes from a certain electrode placed in a certain place on the subject’s scalp. The output data consists of a three-dimensional tensor of shape 53x63x52 which depicts the 3D “picture” of the subject’s brain obtained through fMRI.

We trained models with various architectures, including: fully-connected, convolutional, transformer-based, and combinations of the previous. We tweaked

activation functions, loss functions, optimizers, learning rates, batch size, layer sizes, learning rate schedules, and others.

Results & Obstacles

In the case in which data from all subjects is shuffled randomly into training and testing data, testing performance eventually becomes very good. However, if the testing data is configured to only contain data from a subject which has not been part of the training phase at all, then testing performance doesn't reach a useful level. In the one-subject-left-out setup, the model fails to generalize to the unseen subject. To tackle this, we tried: dropout, weight decay, batch normalization, and reducing model complexity. Still, generalization to an unseen subject remains challenging.