COGS 189 Final Report

Predicting Natural/Artificial Object Recognition with Machine Learning

Members:

Stone Tao: A15910056
Jocelyn Quiroz A15921367
Alvin Hsu: A16282696
Ethan Chen: A16739832
Gabriel Cronn: A16083396
Himanshi Gupta: A15804892

GitHub: Link

Dataset:

https://openneuro.org/datasets/ds003825/versions/1.2.0

Introduction:

The human brain is an intricate and complex system that reacts to various stimuli in the environment. As technology has advanced, we have gained the ability to measure and quantify the brain's response to different stimuli, such as visual images, using EEG (electroencephalogram) data. This type of data analysis allows us to understand which visual stimuli cause the greatest response in the human brain, providing valuable insights into human perception and cognition. **Our team is interested in exploring this phenomenon and investigating whether it is possible to predict an image that a person is looking at based on their EEG data.** This research can have important implications for fields such as attention-based research, content creation, marketing, advertising, and more.

Motivation:

Understanding what visual stimuli causes the greatest response in humans is crucial for businesses and content creators in today's world. Companies invest billions of dollars in advertising and marketing campaigns to capture consumers' attention, and **the ability to create content that elicits a strong response in the human brain is a valuable asset.** Moreover, understanding which types of visual images elicit a stronger response can **aid in designing**

effective interfaces for various digital platforms, such as mobile applications or websites. Finally, this research has implications for understanding human perception and cognition, **providing insights into how the brain processes visual stimuli**. Overall, this project has the potential to provide valuable insights into the relationship between visual stimuli and human perception, with applications in various fields.

Related Work:

Prior research has investigated the relationship between visual stimuli and human brain response, providing valuable insights into human perception and cognition. In a study by Srinivasan et al. (2013), EEG data was used to evaluate the emotional response of participants to natural and artificial stimuli. Results showed that natural stimuli evoked a stronger emotional response in participants. Similarly, a study by Jeon et al. (2016) used EEG data to investigate the effects of natural and urban environments on cognitive restoration. Results indicated that natural environments improved cognitive function and elicited greater relaxation than urban environments. Furthermore, studies have explored sustained attention (Teng et al., 2021) and attention capture in digital settings (Liu et al., 2021), using EEG data to understand the cognitive mechanisms involved. Additionally, EEGs have been used to evaluate surprise in response to visual stimuli (Souza et al., 2021). These prior studies provide a foundation for our investigation into the brain's response to natural and artificial object concepts.

Methods:

In general, we first preprocess the dataset by epoching the data appropriately.

Preprocessing data

Before we could analyze the data, we needed to preprocess and understand the raw BrainVision data that we had. After loading the data, we applied a high-pass filter of 0.1Hz and a low-pass filter of 100Hz to remove any noise that would not be used for our analysis. However, since the images were being presented at a rate of 50ms, participants did not have enough time to process them, resulting in data bleeding. To address this issue, we epoched the events at -100ms to 1000ms relative to when the images were shown. Additionally, we assigned each image to its object category for classification by the machine learning models with the help of ChatGPT.

Creating object categories using ChatGPT

Currently, the given raw data only includes object names for each image. Therefore, we needed to convert these object names into categories that we could use for our analysis, namely “natural” and “artificial” categories. We considered several approaches to address this, such as using NLP models. However, existing pretrained NLP models and more “classical” NLP approaches (e.g. Word2Vec) can have a lot of errors when using their tokenizers and embeddings to determine if an object name is closer to one category or another. We decided to instead use ChatGPT (powered by a more powerful GPT 3 + RL From Human Feedback trained model) as it was trained on much more data and is a much larger model with better one-shot capabilities to tackle novel tasks such as object categorization. To achieve this, we fed each object word into ChatGPT and instructed it to respond in the format of <word>, <category>. We stored the output in a CSV file, and this gave us very accurate object-to-category mappings.

Supervised Learning for Predicting Categories

We use supervised learning techniques to train on the preprocessed data and learn to predict which of the two binary categories the participant is looking at.

In particular, we use the following models: Linear Discriminant Analysis (LDA), a Multi-Layer Perceptron model (MLP), and a Long-Short-Term Memory model (LSTM).

Model Input Details

The input to both the LDA and MLP models is the full input, which has a shape of (E, 1100), where E is the number of events, and 1100 comes from the epoch duration being 1100 ms sampled at 1000HZ.

The LSTM model's input is the same data but processed in an autoregressive manner where for each element in the input sequence, it produces the next hidden and cell states which are used for processing the next element of the sequence.

Neural Network Model Architecture

For the MLP model, we used a 3-hidden layer MLP with 128 units each, ReLU activation, with sigmoid activation after the final layer for binary classification.

For the LSTM model, we used a 3-layer LSTM with 8 hidden units. The final hidden state was flattened and fed into a linear layer followed by sigmoid activation for binary classification.

Model Training

We assigned binary values to the image categories, where artificial = 1 and natural = 0. We partitioned our data into a 75:25 split so that 75% of data is allocated to the training dataset and the remaining 25% reserved for evaluation. Moreover, we ensured there was no inter-participant data leakage by ensuring one participant's data was not in both the training and test dataset.

Both the LSTM and MLP models were trained with the Adam optimizer. To avoid overfitting to the training data we didn't train until convergence.

Analysis

Results:

We discovered that the performance of all the classification models was similar.

Model	Training Accuracy	Training F1-Score	Test Accuracy	Test F1-Score
LDA	0.65999	0.79465	0.65912	0.79452
MLP	0.67388	0.79955	0.65538	0.79071
LSTM	0.65993	0.79441	0.65912	0.79453

Although all models performed relatively well, achieving decent accuracy and F1-scores, most of them exhibited similar performance. Consequently, we suspect that there may still be excessive noise in the data, which makes it challenging to attain higher than approximately 66% accuracy. The noise could originate from various sources, ranging from the data collection process itself to EEG data interference between events of being presented with images.

One aspect that could have been improved is hyperparameter selection. If more time was available, we could have employed more exhaustive hyperparameter search methods to explore the best model and report results from it.

Discussion:

During our experiment analyzing EEG data, we learned a lot about processing EEG data, including how to filter, epoch, and extract signals for analysis. We also gained knowledge about the different data channels, their appearance, and the tools used to preprocess them. The classification models we utilized in our experiment (LDA, MLP, and LSTM) perform relatively well, with an accuracy of around 66%.

However, given more time, we would extend the project to deepen our understanding of the brain's response to natural and artificial object concepts. Specifically, we would explore additional EEG data preprocessing methods beyond the high and low pass filters and epoching utilized in our initial analysis. For example, we could **employ baseline correction and eye**

correction techniques to remove any remaining noise or artifacts in the data, which could enhance the accuracy of our predictions and provide more reliable insights into human perception and cognition. Furthermore, we could **expand our analysis to concentrate on specific EEG channels that are more relevant to visual processing, such as occipital or parietal regions**, and study their relationships with natural and artificial object categories. By focusing on these specific channels, we may gain a more nuanced understanding of how the brain processes different visual stimuli. Additionally, we could explore more advanced model architectures to further enhance the accuracy of our predictions. For instance, we could use **transformer based architectures** which have a strong attention mechanism that could improve sequence modeling and overcome information bottlenecks that LSTM-based models suffer from. Overall, by continuing to explore and refine our EEG data preprocessing and analysis techniques, we could potentially uncover deeper insights into the relationship between visual stimuli and human perception.

A follow up work could potentially be **predicting a smoothed target label**. Instead of binary labels of artificial vs. natural, we could predict an average of the past 5 images seen, which may be easier to predict given that the EEG responses are likely influenced by a few of the past images seen. Specifically, we can convert this to a regression problem and predict the percentage of the last 5 images viewed that were either Artificial or Natural.

Works Cited

- Diener, E., Lucas, R. E., & Oishi, S. (2013). Advances and open questions in the science of subjective well-being. The Scientific World Journal, 2013, 618649. https://doi.org/10.1155/2013/618649

- Tkach, C., Lyubomirsky, S., & Sheldon, K. M. (2016). A qualitative examination of psychological thriving in retirement: The role of perceived challenges, strengths, and resources. The Gerontologist, 56(4), 653-664. https://doi.org/10.1093/geront/gnu116

- Käthner, I., & Kessler, K. (2021). Single-trial estimation of sustained attentional engagement in a continuous performance task using electroencephalography. Frontiers in Human Neuroscience, 15, 673955. https://doi.org/10.3389/fnhum.2021.673955

- Wei, J., Yu, L., Zhang, J., Jiang, Y., & Wang, Z. (2021). Capture of attention by abrupt onset in digital displays: An event-related potential study. Frontiers in Physiology, 12, 727840. https://doi.org/10.3389/fphys.2021.727840

- Brouwer, A.-M., Lopez-Alonso, V., Toffanin, P., & van Erp, J. B. F. (2019). EEG oscillations as a biomarker for surprise in a visual search task. Brain Research, 1720, 146289. https://doi.org/10.1016/j.brainres.2019.146289

- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. U., & Polosukhin, I. (2017). Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), Advances in Neural Information Processing Systems (Vol. 30). Retrieved from https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.