NEUROCIFE - Implementation of Generative Models on Brain Signals in the context of Civil Engineering

Guilherme Simioni Bonfim
Stanford University
gsbonfim@stanford.edu

Guilherme Simioni Bonfim
Stanford University
gsbonfim@stanford.edu

Abstract

The processing and application of neural signals in the context of image generation are still in their nascent stages, with significant advancements emerging in recent years, particularly within the specific realm of Civil Engineering. However, recent academic developments have paved the way for exploring such applications by connecting generative models with the processing of electroencephalograph (EEG) data, tailored to this particular focus.

This project endeavors to capitalize on recent advancements in academia by integrating research-based architectures with gathered and processed neural data, in consideration of the limited availability of resources in the field. Despite incomplete access to reference works, this project has partially replicated their outcomes and achieved comprehensive results in generating Civil Engineering images from proprietary data.

The project acknowledges and expresses gratitude for the valuable orientation and guidance provided by PhD candidate Alberto Tono.

1. Introduction

1.1. Problem

The introductory state of research on the application of EEG signals in generative models has, so far, generally limited the scope of its applicability. Despite the latest endeavors on the application of EEG-collected data for image generation, which have been able to visually produce sufficiently satisfactory outcomes, it has still been presented far from a specific orientation or industry specialization. These factors would be significant in order to configure viable employment in multiple fields. Such imminency is considerably more noticeable in the field of Civil Engineering, which presents a vast scope of possible applications of such models, despite the still limited solutions derived from the field.

Recognizing this critical gap in research and the myriad solutions that could stem from its resolution, this project endeavors to address these challenges head-on. By navigating the uncharted territory of EEG-based generative models within the context of Civil Engineering, it seeks to not only advance scientific understanding but also unlock transformative opportunities for innovation and problem-solving within the industry.

1.2. Motivation

In subsequent consideration of the presented problem, the project has been motivated by advances in research fields connecting EEG signal processing and generative models, most noticeably by "DreamDiffusion" (Bai et al., 2023). Regarding its applicability, Neurocife was also influenced by the possible future implementation of generative models in the context of architecture and engineering 3D models, which is further discussed in section 7.1.

1.3. Inputs

Neurocife utilized EEG tensors obtained from 129 signals in the laboratory as input data for the pre-training and generation sets. Similarly, future applications are expected to likely involve EEG signals and their reconstructions.

1.4. Outputs

There are two different outcomes arising from Neurocife's architecture. Initially, in relation to its pre-training, the project generated checkpoint models that could be used for fine-tuning in equivalent scope in future projects. Additionally, and most significantly, the generative model utilized in the project resulted in the outputs of generative-based images (.png).

2. Related Work

The utilization of neural signals has been approached initially through functional magnetic resonance imaging (fMRI), in which case the use of generative adversarial networks (GAN) has been widely applied, including Ozcelik

et al. (2022) and Shen et al. (2019), with the later applying it to the training of a deep neural networks.

Despite the relative success in the generative capacity of fMRI models, the extensive cost of data acquisition in such methods has led researchers to explore alternative approaches, such as EEG. One of the pioneering endeavors in utilizing EEG data for generative models is "Brain2Image," developed by Kavasidis et al. (2017). Brain2Image significantly revolutionized the understanding of neural signals obtained from patient stimulation and their consequent reproduction. The project utilized a wide and internet-available dataset, ImageNet, to obtain training images, which were subsequently exposed to individuals for a period of 0.5 seconds concurrently with the measurement of their respective signals.

Architecture-wise, Brain2Image relied on Auto-Encoding Variational Bayes, or Variational Autoencoders (VAE), proposed by Kingma and Welling (2022), instead of the more common application of GANs. VAE introduced the utilization of a stochastic variational inference algorithm with advancements in determining a lower bound estimator. In Brain2Image, this involved the utilization of an encoder by feeding the obtained signals into an LSTM network, whose output was processed by a fully connected neural network using ReLU activation. The decoding architecture of the model utilized fully connected neural networks followed by deconvolution layers. The accuracy of the final outcomes, measured by the accuracy of labeling the generated image, reached 0.35.

Sequentially, one of the most recent progresses on the field has been demosntrated by "Seeing Through the Brain" (Lan et al., 2023), which introduced its own architecture, "Neuroimagen", a cohesive pipeline which proposed the incorporation of a multi-level semantics extraction module. Similarly, Zeng et al. (2023) ahve proposed DM-RE2I, a framework which aimed to introduce the extraction temporal and spatial information of the signals, whose evaluated metric, the Inception score, was superior to models using GAN architectures. Still, many models have continued to apply GAN structures, including NeuroVision (Kahre et a., 2022) and ThoughtViz (Tirupattur et al., 2018), with the later presenting a limited dataset of 230 EEG signals, which correlate with the conditions of the current project. As in Brain2Image, LSTM has been applied with visible success in Bozal, which reproduced part of a similar architecture used by Spanizato et al. (2019), which by itself was pioneer on automated classification using visual descriptors from directly measured EEG signals, and also used ImageNet for image data.

Finally, and most importantly for the current project, the advancements of DreamDiffsuion (Bai et al., 2023) have brought similar methodology in comparison to "Brain2Image", in which it was inspired, but with the development of architecture modellings that constituted in a more robust system on the accuracy of its outcome. Although the metrics of the result evaluation configuring an essentially qualitative analysis, DreamDiffusion demonstrated a visually significant improvement over its predecessor. Due to the improvements presented in comparison to similar academic works, DreamDiffusion served as the base structure and architecture of Neurocife. Importantly, most of its positive outcomes can be traced to the extensive dataset utilized, which consisted on 120,000 EEG data samples gathered from a open-source repository entitled MOABB, and similarly to Brain2Image used ImageNet as the image source for the gathering of data through the exposition of figures for laboratoriy voluntaries. The processing includes a masked signal pre-training and the fine-tuning, through the reference images from ImageNet, into pre-trained stable diffusion.

3. Methods

The methodology applied within the training architecture for this application has been based on the model from DreamDiffusion.

3.1. Signal Reconstruction

In its initial state, the architecture targets the reconstruction of EEG signals to establish a reliable and robust generative input. After the initial data gathering, the resulting EEG signals are often inefficient for any generative outcome due to several layers of noise involved. For instance, measurements occurring at sequential time intervals frequently present non-continuous measurements, which are largely influenced by the complexity of identifying such patterns as well as by the imprecision of the data acquisition methods.

Consequently, the inaccuracies found in the original data gathering necessitate the use of architectural structures to correct and reconstruct the original signals. In this direction, the project has employed Masked Autoencoders (MAE), as proposed by He et al. (2021). This model utilizes asymmetric encoding of randomly assigned mask tokens.

The initial stage is named "random sampling," which involves the random selection of non-overlapping samples from the original data (in this case, a two-dimensional signal) with the removal of the remaining information.

The subsequent encoding proceeds with the use of Vision Transformers (ViT) following the approach outlined by Dosovitskiy et al. (2021). ViT is applied exclusively to the selected samples, initially through a linear projection with positional embeddings, followed by the application of transformers in blocks. The basis of this transformer application lies in the use of multi-headed self-attention (MSA) and Multilayer perceptron (MLP) through blocks alternating layers, with the application of LayerNorm (LN) after each block.

MAE then progresses to its decoder, where the inputs consist of both the set of tokens resulting from the encoded process described above and the originally masked (removed) tokens. All tokens once again receive positional embeddings.

The relevant mathematics in this aspect are mostly relate to the transformers within the ViT encoding, in which:

1. $z_0 = [x_{class}; x_p^1 E; ...; x_p^n E] + E_{pos}$

2. $z_l' = MSA(LN(z_{l-1})) + z_{l-1}$

3. $z_l = MLP(LN(z_l')) + z_{'l}$

4. $y = LN(z_L^0)$

3.2. Fine Tuning through owned EEG data

After obtaining the reconstructed EEG signals in organized checkpoints, the generation of images is followed by the fine-tuning of Stable Diffusion (SD) in pretrained models.

Like all diffusion models, Stable Diffusion is essentially based on probability, which involves learning a data distribution ($p(x)$) with the structured implementation applied to a normally distributed variable to gradually denoise it (Rombach et al., 2022). Similarly, it also presents an extension to the application of UNet, an architecture designed, as in the pretraining, to obtain context through contraction and precise localization by the application of decontraction (Ronneberger et al., 2022). Stable Diffusion applies a methodology class entitled "Latent Diffusion Models", which is an advance in relation to purely transofer-based models. Initially, it proposes the encoding of the input image x through ϵ into $z = \epsilon(x)$, further decoded by $\delta(z)$. This consists essentially on the application application of a vector quantization (VQ) GAN (whose details are out of scope for this project), with the difference that the quantization layer is in fact absorbed by σ.

Sequentially, the model presented is the introduction of a domain specific encoder, τ_θ, which projects the already processed input y to $\tau(y)$.

Attention is subsequently calculated in a similar methodology in relation to the well-known transformer models:

$$Y_i(Q, K, V) = softmax((\frac{(XQ_i)(XK_i)^T}{\sqrt{d/h}}) \cdot V)$$

Given the values of:

$$Q = W_Q^{(i)} \cdot \phi_i$$

$$K = W_k^{(i)} \cdot \tau_0(y)$$

$$V = W_V^{(i)}$$

For intermediate values ϕ_i, or a representation of U-NET implementing ϵ_θ and $W_V^{(i)}$.

Another mathematical definition whose details are out of escope for thesis explanation is:

$$\epsilon_\theta := \epsilon_\theta(t)_{t=1}^T$$

which is a set of T functions, each $\epsilon_\theta(t) : \chi \rightarrow \chi$ (indexed by t) is a function with trainable parameters $\theta(t)$ (Song, 2021).

Thus, the loss function is finally:

$$L_{SD} = \mathrm{E}_{x,\epsilon}(||\epsilon - \epsilon_\theta(x_t, t, \tau_\theta(y)||))^2$$

The optimization of which is done in terms of the variables ϵ_θ and τ_θ.

As the project aimed to develop an application of the previous findings of DreaamDiffusion to its own processed data, the same architecture structure and code base has been maintained, following the GitHub repository "Reproduce DreamDiffusion". The code changes were made on the configuration of the original code to the newly processed outputs and differences in datasets and training machines.

4. Dataset and Features

4.1. Data Gathering

Considering the utilization of EEG signals on the pre-training, the project required the acquisition of data through laboratory reproduction of similar methodologies applied by the related projects described before. Therefore, the student was the voluntary for the data gathering through Stanford's Wu Tsai Neurosciences Institute laboratory.

4.2. Images

The methodology used in the previous projects required that the voluntary was exposed to images in different classes in order for the EEG signals to be obtained. As described, the reference projects used ImageNet, which was also applied on the first stage of the project, which consisted on the reproduction of the results obtained by DreamDiffusion.

For the adaptation to the owned dataset, through the assistance of PhD candidate Alberto Tono from the Department of Computer Science, 50 images from each of the three classes in the scope of the project, ”Firmitas,” ”Utilitas,”, and ”Aesthetic”, were found virtually and shaped to an adequate format. The images can be found at https://drive.google.com/drive/folders/1iaeg_wODnHFiysnQBgMe75mieSfcjaJO?usp=sharing.

4.3. EEG lab collection

After following the guidelines for such operations in accordance with Koret HNCL, the EEG stimulus was configured through the software xDiva. The name of the files followed the structure ”cife-001[name of the participants]”

4 total stimulus were ran, including 2 vertical with and without Vision Pro at slow and high frequency. Among those, only one referred to the gathering of EEG data directly focused on processed, with the other being in respect of the gathering of information of validation and possible future applications, especially on 3D contexts.

There were in total frequency ranges: 14-70Hz, 5-95Hz and 55-95Hz, which were created through Milena.

Following the general structure of dataset detail for EEG configurations, the description consists on:

The recording protocol involved 3 object classes with 50 images each, sourced from a new dataset, resulting in a total of 150 images. Visual stimuli were presented to users in a block-based setting, with images of each class shown consecutively in a single sequence, each displayed for 0.5 seconds. A 10-second black screen (during which EEG data were recorded) was presented between class blocks. The collected dataset comprises 150 segments (time intervals recording the response to each image); each EEG segment contains 128 channels, recorded for 0.5 seconds at a 1 kHz sampling rate, represented as a 128×L matrix, with L approximately 500, indicating the number of samples in each segment on each channel. The exact duration of each signal may vary, so the first 20 samples (20 ms) were discarded to reduce interference from the previous image, and the signal was then cut to a common length of 440 samples (to accommodate signals with L ¡ 500). The dataset includes data already filtered into three frequency ranges: 14-70Hz, 5-95Hz, and 55-95Hz.

In our case, the 5-95Hz, due to its use on DreamDiffusion, was employed for pre-training.

Importantly, the generated output was converted to .mff, due to the necessity of compatibility with MNE-Python and MATLAB.

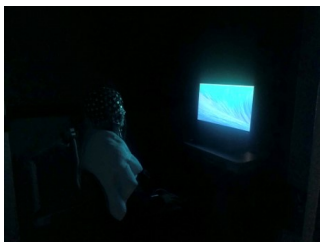

Figure 1. Data Gathering

4.4. Data Tranformation

In order to use to the gathered EEG signals for the pre-training, it was necessary to convert them to a .pth file, wtih the following dictionary structure:

’eeg’: tensor, ’image’: int, ’label’: int, ’subject’: int ’label’: string ’image’: string

Which was accomplished by using the library Python MNE for data processing. In total, therefore, each of the tensors was of the shape (129, 500) and were later converted to the shape (128, 500) for matching the expected output.

5. Experiments/Results/Discussion

5.1. Reproduction

During the reproduction, importantly, the definition of the hyperparmeters was done by the utilization of the same values of the original implementation, in order to guarantee fidelity of results.

On the reproduction of the initial outcomes by DreamDiffusion, the project was able to obtain equivalent results in comparison to the original project. Initially, the project aimed to obtain, using the same dataset, which consists, by its own description:

”This dataset includes EEG data from 6 subjects. The recording protocol included 40 object classes with 50 images each, taken from the ImageNet dataset, giving a total of 2,000 images. Visual stimuli were presented to the users in a block-based setting, with images of each class shown consecutively in a single sequence. Each image was shown for 0.5 seconds. A 10-second black screen (during which we kept recording EEG data) was presented between class blocks. The collected dataset contains in total 11,964 segments (time intervals recording the response to each image); 36 have been excluded from the expected 6×2,000 = 12,000 segments due to low recording quality or subjects not looking at the screen, checked by

using the eye movement data. Each EEG segment contains 128 channels, recorded for 0.5 seconds at 1 kHz sampling rate, represented as a 128×L matrix, with L about 500 being the number of samples contained in each segment on each channel. The exact duration of each signal may vary, so we discarded the first 20 samples (20 ms) to reduce interference from the previous image and then cut the signal to a common length of 440 samples (to account for signals with L ¡ 500). The dataset includes data already filtered in three frequency ranges: 14-70Hz, 5-95Hz and 55-95Hz.”

The obtained signals can visualized as follows:

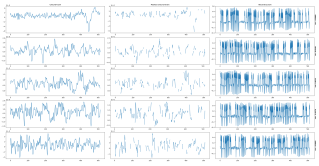

Figure 2.

Quantitatively, the procedure obtained a considerably small loss after 14 epochs:

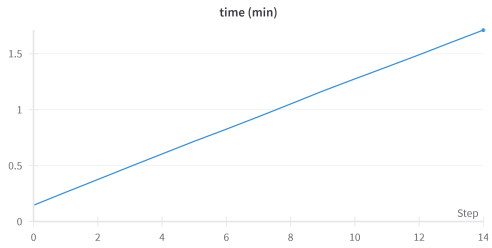

Figure 3.

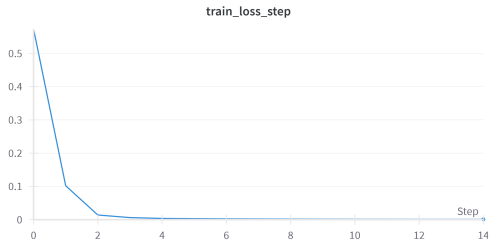

Figure 4.

In sequence, in the iamge generation, the results were very similar to those found by DreamDiffusion in image quality. Some of the samples can be seen below:

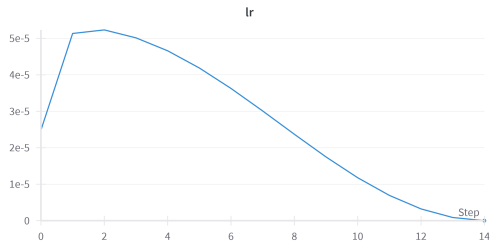

Figure 5.

Figure 6. Enter Caption

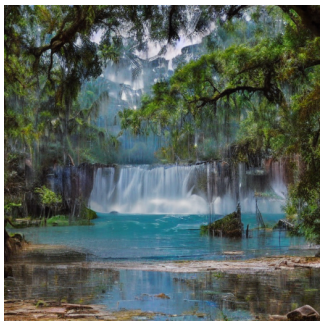

Figure 7. Enter Caption

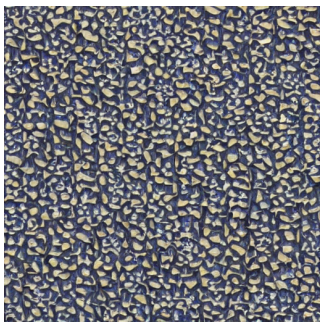

Figure 8. Enter Caption

6. Neurocife

As an extension to the original project of DreamDiffusion, Neurocife was also effective on the reconstruction of the obtained signals in lab, which was also reproduced un-

der a short amount of time.

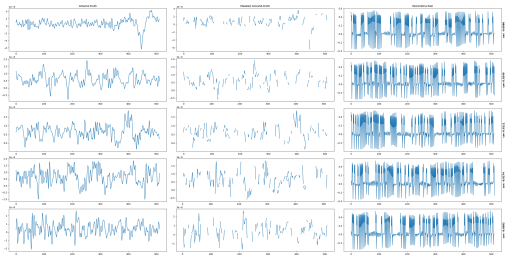

Figure 9.

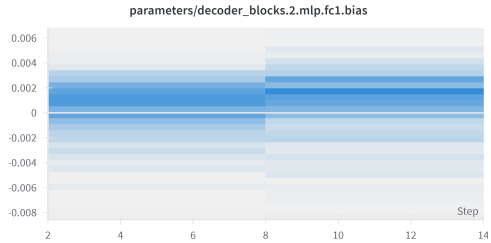

Figure 10.

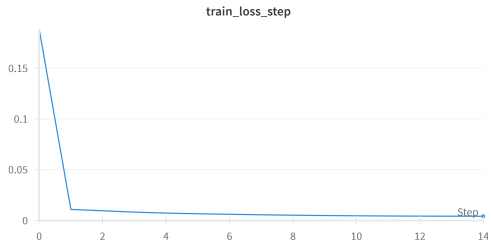

Figure 11.

In addition, the generated images qualitatively presented a similar quality to those of DreamDiffusion. Nonetheless, to incoehrences were found: i. some of the images presented unrelated content in relation to that present in training, e.g., people or objects not presented in the image dataset; ii. abstract content, with non-clarity of representation, was more common than on the original project.

Those issues could be explained by some primordial factors:

- The selected image dataset, as well as the obtained signals from training, are still limited in scope and comprehensiveness

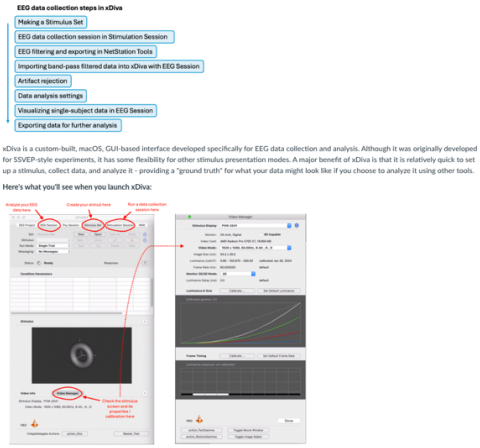

Figure 12.

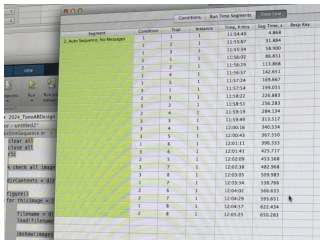

Figure 13.

Figure 14. Enter Caption

- The brain processing of more abstract concepts, such as the structure of a building, may be harder to identify in Diffusion Models
- The vast majority of the data in the pre-trained diffusion model is not correlated to Civil Engineering

Some of the samples are:

7. Conclusion/Future Work

In general, Neuro-cife was succesful on reproducing the original outcomes from DreamDiffusion and on obtain-

Figure 15. Enter Caption

Figure 16. Enter Caption

Figure 17. Enter Caption

ing visually equivalent images for the own obtained data. This demonstrates that, despite the non-extensive focus presented by brain signal analysis generative models in the field of civil engineering up to this point, there is a great possibility of application on the field.

In future works, a important factor would be obtaining more comprehensive training data, especially from more than one voluntary. In addition, a possible future extension is the implementation of similar generative models for 3D structures or models based on scatches instead of purely signals.

8. References

Bai et al. (2023), *DreamDiffusion: Generating High-Quality Images from Brain EEG Signals*

C. Spampinato, S. Palazzo, I. Kavasidis, D. Giordano, N. Souly, M. Shah, Deep Learning Human Mind for Automated Visual Classification, International Conference on Computer Vision and Pattern Recognition, CVPR 2017

Chen et al. (2016), Improved techniques for training gains. *Advances in Neural Information Processing Systems, 29.*

Kavasidis, Isaak, et al. "brain2image." Proceedings of the 25th ACM International Conference on Multimedia, 23 Oct. 2017, https://doi.org/10.1145/3123266.3127907.

Lan et al. (2023), *Seeing through the Brain: Image Reconstruction of Visual Perception from Human Brain Signals*

S. Palazzo, C. Spampinato, I. Kavasidis, D. Giordano, J. Schmidt, M. Shah, Decoding Brain Representations by Multimodal Learning of Neural Activity and Visual Features, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, 2020, doi: 10.1109/TPAMI.2020.2995909

Song et al. (2022), *DENOISING DIFFUSION IMPLICIT MODELS*

Rombach. (2022), *High-Resolution Image Synthesis with Latent Diffusion Models*