Proceedings of the 4th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/47/20241135

Generating high-quality images from brain EEG signals

Daxiang Yang

The affiliated high school to jiangsu normal university, Xuzhou,JiangSu, 221000

Yangdaxiang93@gmail.com

Abstract. This study presents DreamDiffusion, an innovative approach to produce high-quality images straight from electroencephalogram (EEG) brain signals, eliminating the need for thought-to-text translation. By harnessing pre-trained text-to-image models, DreamDiffusion integrates temporal masked signal modeling to adeptly pre-train the EEG encoder, ensuring accurate and dependable EEG data representation. Moreover, by integrating the CLIP image encoder, this method fine-tunes the alignment of EEG, text, and image embeddings, even with a scant amount of EEG-image pairs. Effectively navigating the complexities inherent in EEG-based image creation, such as data noise, limited content, and personal variances, DreamDiffusion showcases promising outcomes. Both quantitative and qualitative assessments validate its efficacy, marking a considerable advancement in the realm of efficient, affordable "thought-to-image" conversions, with promising implications in both neuroscience and computer vision.

Keywords: High-Quality Images, Brain, EEG Signals, DreamDiffusion.

1. Introduction

The objective is to harness brain signals for efficient and user-friendly artistic creations. While fMRI tools are sophisticated and pricey, making them less accessible for widespread artistic use, EEG (electroencephalogram) offers an affordable and non-invasive means to record brain activity. With portable commercial EEG devices emerging, there's promising potential for art creation in the near future. This study explores the possibility of using potent pre-trained text-to-image models, like Stable Diffusion, to generate quality images directly from EEG brain signals. But this endeavor faces hurdles:

EEG recordings are non-invasive and inherently contain noise. With the variance in individual EEG data, how can we extract robust semantic interpretations amidst these challenges?

Stable Diffusion aligns text and image spaces effectively due to the integration of CLIP and extensive text-image pair training. But EEG signals have unique characteristics, making their alignment with text and image spaces challenging, especially given the noisy and limited EEG-image pair data[1].

To overcome the first challenge, we utilize vast EEG datasets for training, rather than just the scarce EEG-image pairs. Instead of the conventional approach that views input as two-dimensional images, we emphasize the temporal aspects of EEG signals. By masking random tokens and reconstructing them, the encoder gains a profound insight into varied EEG data across individuals and brain activities.

Addressing the second challenge, instead of directly fine-tuning Stable Diffusion models with minimal noisy data pairs, we propose the inclusion of additional CLIP guidance to aid the alignment of EEG, text, and image spaces. The text embeddings produced by SD differ considerably from pre-trained

DOI: 10.54254/2755-2721/47/20241135

EEG embeddings. By using CLIP's image encoder, we optimize EEG embedding representations to better align with CLIP's text and image embeddings. This optimizes the image generation quality in Stable Diffusion.

With these strategic approaches, our proposed DreamDiffusion method can efficiently transform EEG signals into high-quality, realistic images. In summary, our contributions are:

Introducing DreamDiffusion, an advanced method converting EEG signals into realistic images, pioneering in affordable and portable "thoughts-to-images".

Incorporating a temporal masked signal modeling to enhance the EEG encoder's effectiveness.

Utilizing CLIP's image encoder for enhanced alignment between EEG, text, and image spaces, even with limited EEG-image data.

Demonstrating DreamDiffusion's efficacy through both quantitative and qualitative outcomes [2].

2. Improvement Methods

2.1. Generating images from brain activity

This research field explores using brain signals like fMRI and EEG to generate images. Traditional fMRI-based methods use paired data to train models which then predict image features. Recent studies offer unsupervised approaches, for instance, reconfigurable autoencoder designs. The recent work, MinD-Vis, stands out by producing plausible images with rich semantic information. On the EEG side, deep learning has been employed for image generation. Works like Brain2image and ThoughtViz use LSTM and generative methods to produce images based on EEG data [3].

2.2. Model pre-training

Pre-training models have become a significant trend in computer vision. There are multiple methods, some of which focus on contrastive learning or autoencoding. A notable development is the CLIP method that creates a multi-modal embedding space by training on massive amounts of text-image pairs, achieving impressive results in zero-shot image classification [4].

2.3. Diffusion models

Diffusion models, popular as generative tools for high-quality content creation, are defined by a bi-directional Markov Chain. These models have shown great generative potential, especially with certain training objectives. However, they have computational challenges. Some methods, such as the LDMs, work on a lower-dimension compressed latent space to address these challenges and maintain synthesis quality.

3. Proposed Method

As shown in Figure1, the introduced method, "DreamDiffusion," aims to generate high-quality images from EEG signals through three main steps:

Masked signal pre-training: Given the challenges in EEG data like noise and variability, this step uses masked signal modeling techniques to capture meaningful information. The data is divided into tokens in the time domain, with a certain percentage randomly masked. These tokens are then transformed into embeddings to capture the deeper meaning of EEG signals.

Fine-tuning with Stable Diffusion: Using the pre-trained Stable Diffusion model, which was originally for text-to-image generation, the process is adapted to EEG data. Given the distinct characteristics of EEG signals, they are fine-tuned for better alignment with the pre-trained SD's text embeddings.

Aligning EEG, text, and image spaces with CLIP encoders: To make EEG representations more apt for generating images, the representations from pre-training are fine-tuned using CLIP. The CLIP model aligns the EEG, text, and image spaces, making EEG embedding more suited for SD image generation.

DOI: 10.54254/2755-2721/47/20241135

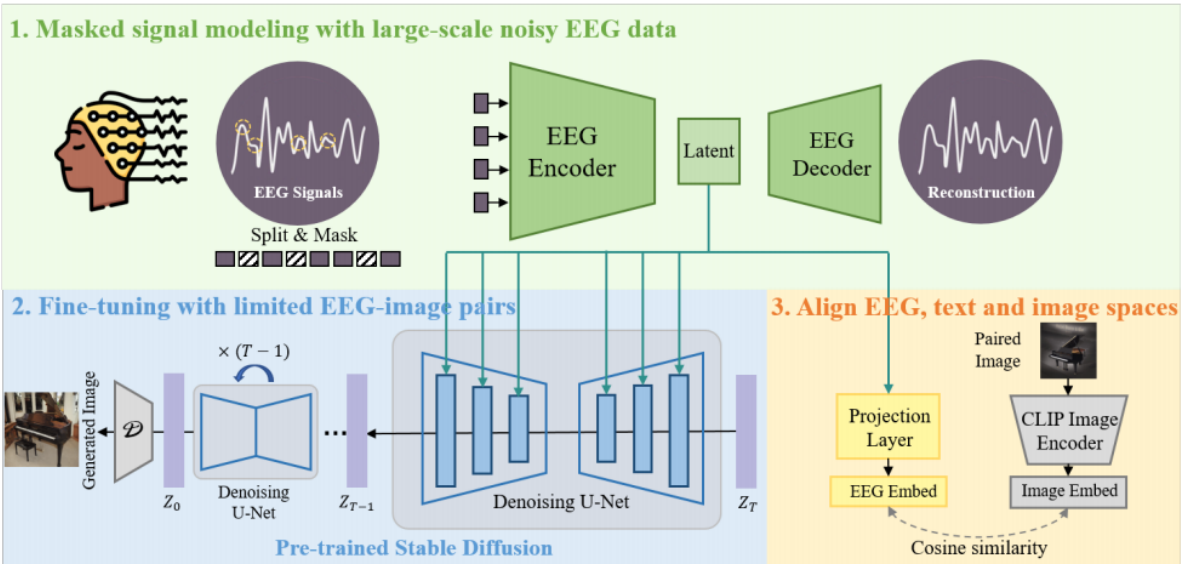

Figure 1. Masked signal modeling with large-scale noisy EEG data and fine-tuning with limited EEG-image pairs [5].

4. Discussion

The convergence of brain activity data and advanced deep learning techniques holds immense potential, especially in the realm of brain-computer interfaces. Techniques like fMRI and EEG provide unique insights into human brain activity. When combined with generative models, it opens doors to applications ranging from thought-to-image conversions to potentially aiding patients with speech or motor impairments.

However, there are evident challenges:

Data Complexity: Brain signals, especially EEG, are highly complex, often noisy, and influenced by numerous external factors. This complexity requires sophisticated pre-processing, understanding of the signal characteristics, and robust models to capture the inherent patterns.

Model Training and Computation: Advanced models, especially the likes of diffusion models, are computationally expensive. Despite the rewards, the resource requirements can be a limiting factor.

Alignment Challenges: Integrating EEG data with models primarily designed for textual or visual data, like the Stable Diffusion or CLIP, involves challenges in alignment. It's not just about the technicalities but also ensuring the semantic integrity of the generated outputs.

Despite these challenges, the progress made in the field, as outlined in the provided passages, showcases the advancements in harnessing brain signals for visual synthesis. If realized fully, these technologies could revolutionize human-computer interactions, medical interventions, and even entertainment. As shown in Figure2, the very idea of "thinking" and having AI generate visuals or actionable outputs from those thoughts is both exciting and a step closer to science fiction becoming reality [6-9].

DOI: 10.54254/2755-2721/47/20241135

Figure2. Failure cases of Dream Diffusion [9].

5. Conclusions

This study introduces "DreamDiffusion," an innovative approach to produce high-resolution images using EEG signals - a readily accessible and non-invasive measure of brain activity. By harnessing the insights from extensive EEG datasets combined with the robust generative power of image diffusion models, DreamDiffusion overcomes many challenges inherent to EEG-driven image creation. We employ a two-step process: pre-training followed by fine-tuning, which allows the EEG data to be aptly transformed into representations ideal for image generation through Stable Diffusion. This technique marks a pivotal progression in crafting images from brain signals.

However, it's crucial to note certain limitations. Presently, EEG data predominantly offers broad categoric insights, as evidenced by some anomalies in our experimental results depicted in Figure 2. In some instances, specific categories overlap with others sharing similar shape or color characteristics. This might stem from the brain's inclination to prioritize shape and color during object identification. Despite these challenges, DreamDiffusion holds immense promise across diverse fields, including neuroscience, psychology, and human-computer interfaces.

References

[1] Yunpeng Bai, Cairong Wang, Shuzhao Xie, Chao Dong, Chun Yuan, and Zhi Wang. Textir: A simple framework for text-based editable image restoration. arXiv preprint arXiv:2302.14736, 2023.

[2] Suzanna Becker and Geoffrey E Hinton. Self-organizing neural network that discovers surfaces in random-dot stereograms. Nature, 355(6356):161–163, 1992.

[3] Chris M Bird, Samuel C Berens, Aidan J Horner, and Anna Franklin. Categorical encoding of color in the brain. Proceedings of the National Academy of Sciences,111(12):4590–4595, 2014.

DOI: 10.54254/2755-2721/47/20241135

[4] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale gan training for high fidelity natural image synthesis. arXiv preprint arXiv:1809.11096, 2018.

[5] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale GAN training for high fidelity natural image synthesis. In 7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019. OpenReview.net, 2019.

[6] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. Advances in neural information processing systems, 33:1877–1901, 2020.

[7] Zijiao Chen, Jiaxin Qing, Tiange Xiang, Wan Lin Yue, and Juan Helen Zhou. Seeing beyond the brain: Conditional diffusion model with sparse masked modeling for vision decoding. arXiv preprint arXiv:2211.06956, 2022.

[8] Keith M Davis, Carlos de la Torre-Ortiz, and Tuukka Ruotsalo. Brain-supervised image editing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 18480–18489, 2022.

[9] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.