Reading Thoughts Using GANs

Ahmed Mohammed
Machine Learning Institute
Johannes Kepler University Linz
MN: k12035954

Abstract

Have you ever considered how we can decode thoughts, particularly for individuals with disabilities or those who are unable to communicate verbally? This report introduces a solution to this challenge, exploring two distinct methodologies that leverage generative models to interpret EEG signals obtained from participants while they observe various objects displayed on a screen. We will analyze the intersecting outcomes and disparities between the two algorithms, which employ state-of-the-art GAN algorithms.

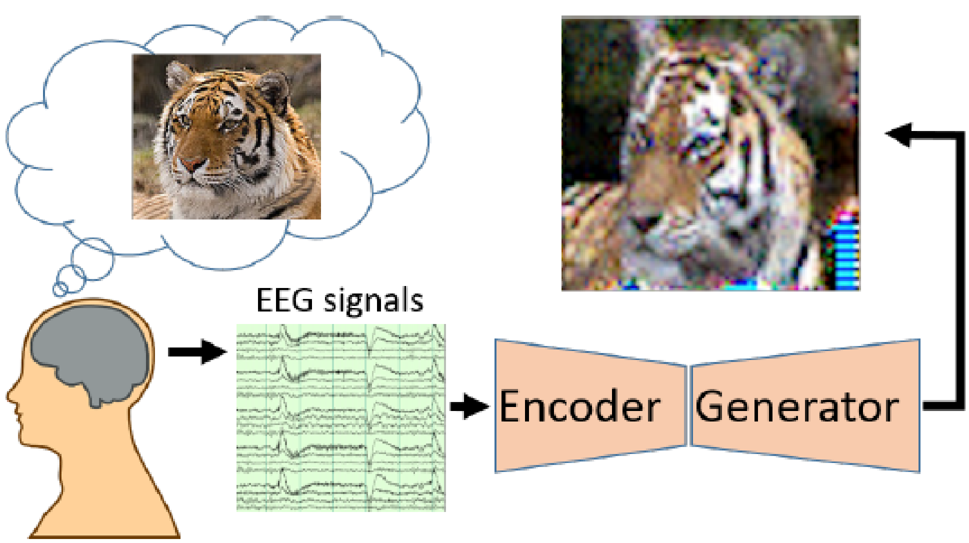

Figure 1: Overview of the proposed method: An EEG signal is fed into the encoder, which transforms it into an encoded format. This encoded signal is then used to generate a visualization that corresponds to the original EEG signal captured.

1 Introduction

1.1 Overview

Brain-computer interface (BCI) technology has significantly enhanced communication and control for individuals with disabilities or those unable to communicate verbally. Concurrently, this technology provides accessibility to reading brain activities through Electroencephalography (EEG) signals, serving as a practical and non-invasive method for recording brain activity.

EEG signals play a crucial role in decoding brain responses to imagination, visual stimuli, and various cognitive processes, thereby advancing the field of mind-reading technology. The utilization of different algorithms has facilitated the decoding of informative patterns from brain signals, particularly in deciphering visual and linguistic content.

The promising results achieved in decoding visualized images from brain signals highlight the potential of such techniques to better understand human cognition and facilitate communication through the interpretation of brain activity.

1.2 Challenging

Despite advancements in decoding brain signals, both papers acknowledge the significant challenge of training Generative Adversarial Networks (GANs) to construct images directly from brain activity by collecting EEG signals from different participants under specific conditions. This endeavor requires overcoming various obstacles, including accurately interpreting EEG signals to generate meaningful visual representations. Thus, innovative methodologies and computational techniques are essential in advancing the field of mind-reading technology.

To enhance result accuracy, both papers employ different techniques: EEG2IMAGE utilizes the "differentiable Data Augmentation" method, while the ThoughtViz paper incorporates an additional classifier to improve model accuracy.

1.3 Implementation

The collected EEG signals originate from participants imagining objects displayed on a screen. Participants are instructed to visualize each object for 10 seconds, followed by a 20-second rest period before repeating the process for subsequent objects.

The dataset is divided into two subsets: EEG data and image data. These subsets encompass three categories: digits, characters, and objects. This report focuses solely on characters and objects, as they yield common results across both papers.

The initial model is trained to encode EEG signals, which are then used to train GANs concurrently with image data. The generator receives encoded EEG inputs and attempts to deceive the discriminator. Meanwhile, the discriminator distinguishes between real and fake images generated by the generator.

After training the EEG classifier and utilizing it as an encoder, along with training the GAN model, we can employ them as depicted in Figure 1. This enables us to interpret thoughts and translate EEG signals into images.

2 Methodological Approaches

We will review the two different approaches implemented in both papers, discussing the differences between them and the loss functions they employed. Additionally, we'll review their architectures and training procedures, as well as the methods utilized to enhance accuracy.

2.1 EEG Encoder

The architecture of the EEG2IMAGE encoder comprises 2 LSTM layers, forming an LSTM network with 128 hidden units, aimed at transforming the EEG signal into a 128D feature vector as shown in Figure 2. The utilized data consist of [128 Hz x 10 seconds] with 128 channels recorded over 10 seconds. The encoder is utilized to extract features from the EEG and reduce the data dimension from [128, 10] to 128. Thus, the encoder is trained to classify the objects, and the formulation of the triplet loss used to train the encoder is as follows:

$$\min_{\theta} \mathbb{E}\left[\|f_\theta(x_a) - f_\theta(x_p)\|_2^2 - \|f_\theta(x_a) - f_\theta(x_n)\|_2^2 + \beta\right]$$

The architecture of the ThoughtViz encoder consists of 1D convolutions along the time axis with a kernel size of 4, followed by another 1D convolution along the channels axis with a kernel size of 14, representing the number of channels. These 1D convolutions are succeeded by two sets of alternating 2D convolution and 2D max pooling layers. Finally, two fully connected layers with 100 neurons each are added for classification as shown in Figure 2. The intermediate feature vector from the first fully connected layer is considered as the EEG encoding for a given input EEG signal.

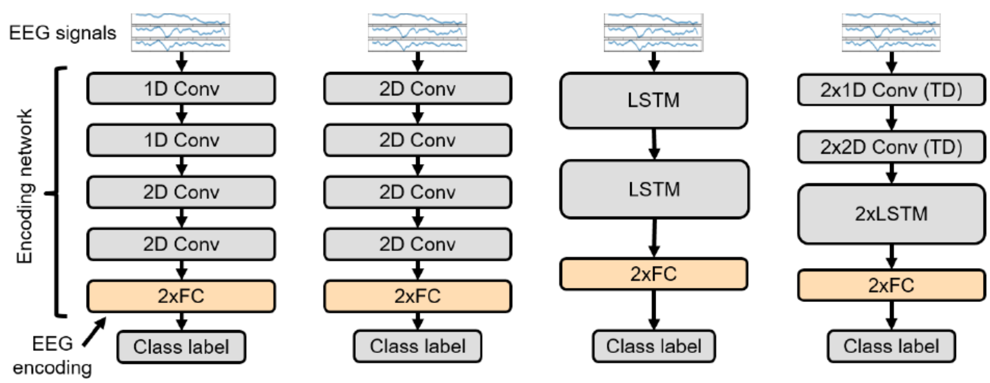

Figure 2: The architecture on the leftmost side is utilized as an encoder in ThoughtViz, while the architecture third from the left is employed as an encoder in EEG2Image.

2.2 EEG2IMAGE

The EEG2IMAGE approach employs a Conditional Generative Adversarial Network (cGAN), consisting of two networks: Generator G and Discriminator D. The conditional aspect of the cGAN allows for the generation of class-specific images based on specific brain activities. The Generator learns to transform a latent distribution p_Z into the real-world data distribution p_{data}, as shown in Figure 3. In this case, we assume the latent distribution as an isotropic Gaussian $\mathcal{N}(0, I)$, from which a noise vector $z \in \mathbb{R}^{128}$ is sampled. The Discriminator learns to distinguish between real images ($D(\mathbf{x})$) and generated images ($D(G(\mathbf{z}))$). The optimization process for the complete GAN is represented as:

$$\min_G \max_D \mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})}[\log(D(\mathbf{x}))] + \mathbb{E}_{\mathbf{z}\sim p_Z(\mathbf{z})}[\log(1 - D(G(\mathbf{z})))]$$

To address the issue of limited dataset size, differentiable data augmentation is employed as a block between the Generator (G) and the Discriminator (D), aiding in improving learning accuracy.

2.3 ThoughtViz

Traditional GAN architecture comprises a Generator (G) and a Discriminator (D). The Generator generates a sample image from a random noise input (z), while the Discriminator determines whether the input is a generated or real sample.

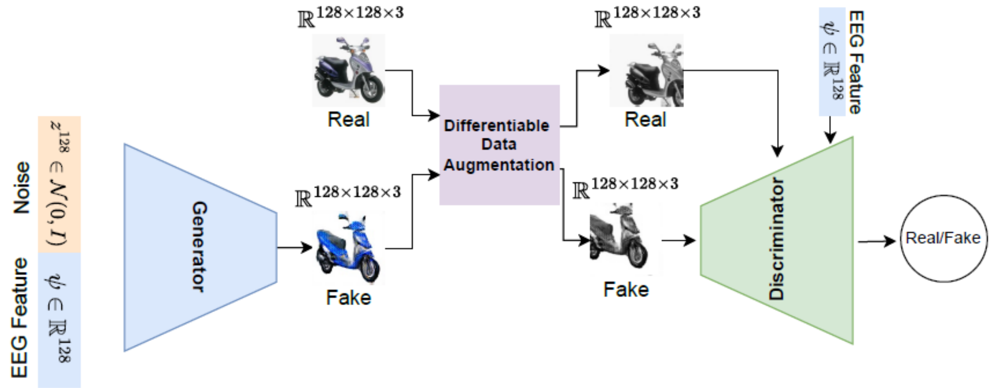

Figure 3: Representation of the GAN network featuring a data augmentation block, which mitigates the risk of the discriminator memorizing the limited dataset and facilitates the generation of high-quality images.

Similar to EEG2IMAGE, the ThoughtViz approach uses a Conditional Generative Adversarial Network (cGAN) with two networks: Generator G and Discriminator D, where the conditional aspect considers the class of a sample during generation. The objective function for the GAN network is:

$$\max_D \min_G V_D(D, G) = \max_D \min_G \mathbb{E}_{\mathbf{x}\sim p_{\text{data}}}[\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z}\sim p_z}[\log(1 - D(G(\mathbf{z})))]$$

Additionally, an architecture with an additional classifier is proposed in ThoughtViz, aiming to explicitly classify the generated samples instead of relying solely on the Discriminator, as shown in Figure 4. This enhancement has shown in experiments to lead to faster convergence of the GAN model. Moreover, GAN training presents challenges with limited training data.

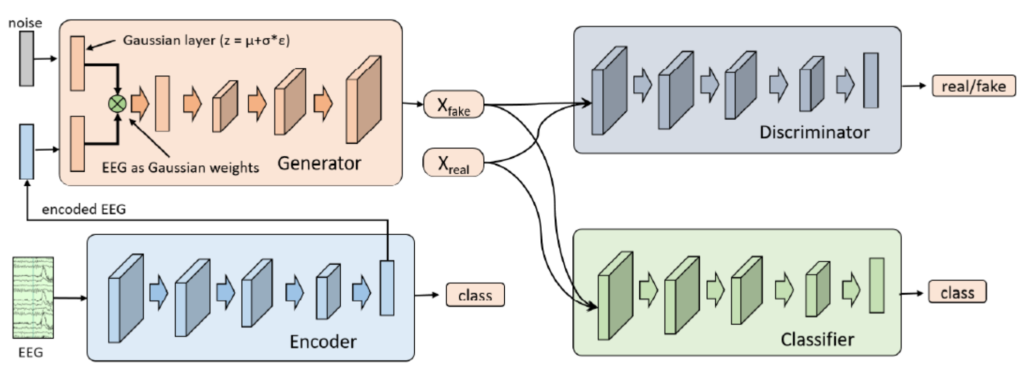

Figure 4: Illustration of the GAN architecture for generating images from EEG signals. The encoded EEG signal is used as conditioning for the generator. The resulting image is subjected to adversarial loss by the discriminator and classification loss by a classifier.

Table 1: Inception Score Comparison

Method	Inception Score
ThoughtViz	5.43
EEG2Image	6.78

3 Results

The results comparison between the two approaches reveals the outputs of the ThoughtViz outputs in (a) and the EEG2Image outputs in (b) characters data. Additionally, a red bounding box random sample is included, and the remaining images are the generated characters, as shown in Figure 5.

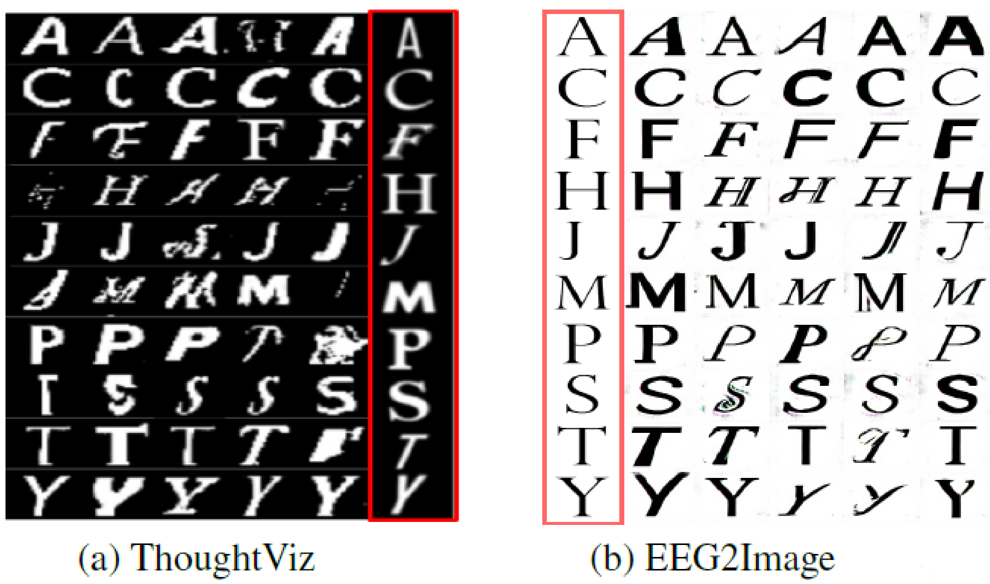

Figure 5: Comparison of generated character outputs between the ThoughtViz and EEG2Image approaches.

The results comparison between the two approaches reveals the outputs of the ThoughtViz approach in (a) and the EEG2Image outputs in (b) objects data. Additionally, a red bounding box random sample is included, and the remaining images are the generated objects, as shown in Figure 6.

The comparison of Inception Scores between ThoughtViz and EEG2Image shows that EEG2Image achieved a higher score of 6.78 compared to ThoughtViz's score of 5.43, as shown in Table 1. This suggests that EEG2Image generates images with greater diversity and quality than ThoughtViz. Despite this difference, both approaches demonstrate effectiveness in generating images from EEG signals, with EEG2Image showing a slight advantage in this evaluation metric.

(a) ThoughtViz (b) EEG2Image

Figure 6: Comparison of generated object outputs between ThoughtViz and EEG2Image approaches.

References

[1] Kumar, P., Saini, R., Roy, P.P., Sahu, P.K., & Dogra, D.P. (2018). Envisioned speech recognition using EEG sensors. Personal and Ubiquitous Computing, 22*(1), 185–199.

[2] Singh, P., Pandey, P., Miyapuram, K., & Raman, S. (2023). EEG2IMAGE: Image reconstruction from EEG brain signals. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-5). IEEE.

[3] Kumar, P., Saini, R., Roy, P.P., Sahu, P.K., & Dogra, D.P. (2018). Envisioned speech recognition using EEG sensors. *Personal and Ubiquitous Computing, 22, 185-199.