Open Access

OBM Neurobiology

LIDSEN Publishing Inc.

Original Research

Image Generation Inspired by Electroencephalography for Neuromarketing Applications Using Extracted Features from Transformer-Based Models

Mahsa Zeynali, Hadi Seyedarabi *, Reza Afrouzian

Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran; E-Mails: m.zeynali@duke-nus.edu.sg; seyedarabi@tabrizu.ac.ir; afrouzian@tabrizu.ac.ir

* **Correspondence:** Hadi Seyedarabi; E-Mail: seyedarabi@tabrizu.ac.ir

Academic Editor: Maurizio Elia

Special Issue: Applications of Brain–Computer Interface (BCI) and EEG Signals Analysis

OBM Neurobiology
2026, volume 10, issue 1
doi:10.21926/obm.neurobiol.2601328

Received: September 25, 2025
Accepted: January 26, 2026
Published: March 10, 2026

Abstract

The design of products in Neuromarketing using machine learning methods has been a continuous challenge in Computer-aided design. Previously, deep learning techniques have been applied to generate random images for domains such as furniture, fashion, and product design. However, using deep generative methods requires a large amount of data and overlooks human aspects in the design process. This paper aims to extract human perceptual factors from brain signals using a Transformer-based model and involve them in artificial intelligence-based product design. To achieve this, Electroencephalography (EEG) signals are recorded while observing product images. In the first stage, an encoder based on the Transformer architecture extracts features from raw signals. In the second stage, an Auxiliary Classifier Generative Adversarial Network (ACGAN) is trained on the extracted features to generate product images. An accuracy of 92.8% obtained from the EEG encoder signifies that the features extracted by the Transformer-based model are well distinguishable as a condition of the generator. The generated images, with an inception score of 5.22 and 78% classification accuracy, exhibit features similar to those of the original images. This approach holds promise for enhancing designer-customer communication in Neuromarketing applications,

particularly in scenarios where customers may struggle to express their design preferences clearly.

Keywords
Neuromarketing; deep learning; Generative Adversarial Networks (GAN); Electroencephalography (EEG); image generation from EEG; transformer

1. Introduction

Neuromarketing is the application of neuroscience to marketing research that focuses on studying the cognitive and emotional reactions to services or products [1]. Neuromarketing not only concentrates on the impact of various market stimuli on sales but also elucidates the influence of changes made in presenting these stimuli on consumer choices [2]. Given the substantial costs involved in marketing, there is a strong incentive to achieve greater efficiency within market segments and customer groups. Traditional research methods primarily focus on post-purchase consumer attitudes toward products, typically assessed through post-purchase questionnaires. Thus, the received responses are delayed and simplified, unable to accurately reflect the real-time mental state of the customer during the purchase [3]. In contrast, the objective of neuromarketing is to capture immediate responses based on the biological and neural indicators of customers during the purchase. Another challenge in neuromarketing is the heterogeneity among consumers and its impact on consumer behavior. The observed heterogeneity can stem from diverse factors, including age, gender, various biological elements such as hormones and genes, and various physiological factors [4].

Researchers utilize a range of techniques, including Transcranial Magnetic Stimulation (TMS), Steady-State Topography (SST), Electroencephalography (EEG), and functional Magnetic Resonance Imaging (fMRI) to measure alterations in brain activity. Metrics such as respiratory rate, heart rate, facial expressions, skin response, and eye tracking are also utilized to gauge changes in the physical and emotional states of consumers. Subsequently, these metrics are used to understand how customers make product decisions and to establish the correlation between the metrics and those decisions [3].

The human brain is composed of neurons that communicate via electrical signals [5]. Measuring EEG signals is a practical way to detect changes in brain activity without temporal delay. The absence of this time delay is crucial for understanding unconscious reactions and sensory responses in individuals [4].

By utilizing neuromarketing, producers and sellers can determine the best strategies for advertising their products and prevent the waste of marketing resources. In this field, researchers focus on various marketing parameters, such as brand perception, brand evaluation decisions, brand relationships, brand preferences, pricing, product packaging, brand name, advertising, and the design and development of new products [6].

The automatic generation of designs based on personal preferences has always been a challenge for designers. In previous research, deep learning methods in the field of artificial intelligence have been employed to create diverse designs. For instance, generative Bionic design [7], using a

generative adversarial approach and combining features related to the design objective with biological sources, is utilized to generate images. However, these AI-based image generation methods do not consider human factors, so human cognition is not involved in the generated results. Considering human aspects in the design process is crucial. While a person's preferences regarding a design may sometimes be clear, there are occasions when they are not aware of their actual desires. Therefore, the ability to capture human preferences and integrate them into the production process can lead to significant improvements in AI-based designs [8].

EEG experiments, particularly those involving visual stimuli like films and images, often concentrate on examining the effects in the occipital region. This process combines perception, which affects how objects appear in terms of color, shape, and other attributes, and conception, which encompasses higher cognitive processes. Several studies in cognitive neuroscience [9-11] have examined parts of the visual cortex and the brain responsible for these mental processes.

Several studies have investigated the principles of psychological elements and principles of various Visual Design Elements and Principles (VDEPs) and their impact on observers [12]. VDEPs can evoke emotions and behavioral changes in viewers. For example, in color theory, empirical research has established that colors can elicit specific emotions in individuals viewing visual content [13-15]. Moreover, investigations in neurocognitive studies have revealed that identifiable patterns of brain activity correspond to distinct categories of visual stimuli. Recently, studies have indicated that EEG signals can be utilized to understand what an individual is seeing [16].

Recent advancements in neuroscience and Deep learning-based brain-decoding methods indicate the potential to reconstruct imagined or observed images from EEG signals. This has spurred increased research and motivation to develop a new design method based on artificial intelligence. Essentially, the goal of this method is to bridge the gap between product design and human brain activities. By incorporating the interests and preferences of individuals found in their brain signals into the realm of product design, it is possible to fill this gap and create a final product that aligns with their tastes and preferences.

Various machine learning methods have been employed for EEG to aid in understanding visual images. Bashivan et al. [17] introduced an approach to preserve the temporal, spectral, and spatial features of EEG. This approach identified features that are less sensitive to changes in each dimension, resulting in a notable increase in classification accuracy. Spampinato et al. [18] developed a classifier for visual objects guided by human brain signals. They used EEG data elicited by visual stimuli using an RNN. Zeynali et al. [19] introduced a novel Transformer-based model designed to extract temporal and spectral features from EEG signals based on visual stimuli, with a focus on classification purposes. In another study [20], they used Transformer-based models to extract features from EEG signals for biometric systems. Bagchi et al. [21] introduced the EEG-Conv Transformer model for classifying EEG signals based on visual stimuli. This model utilizes both self-attention layers and temporal convolutional layers.

Mishra et al. [22] proposed a dual-path deep learning model to classify EEG signals based on the image categories that triggered them. This framework combines CNNs applied to both the time and channel axes of the EEG signals. Additionally, a gradient reversal layer (GRL) is employed to learn subject-invariant features. Deng et al. [23] proposed a Reusable LSTM Network (RLN) model for the classification of visual-based EEG signals. This model combines a Long Short-Term Memory (LSTM) network with a one-dimensional CNN to extract features from each channel independently. Abbasi et al. [24] introduced a method for classifying EEG signals recorded in response to visual stimuli,

utilizing a combination of a Residual Network (ResNet) and an LSTM network. They employed a wavelet transform to convert the EEG signals into images, which were then used as model input.

The results of these studies indicate the potential of EEG signals for retrieving brain information. Researchers have also attempted to use the decoded information from neural signals to reconstruct relevant visual information. To generate visual images from EEG signals, Palazzo et al. [25] used features extracted from the EEG by an LSTM network. These features were then input into a GAN network to generate the expected images.

Kavasidis et al. [16] proposed the Brain2Image approach for generating images using recorded visual-evoked EEG signals. In their approach, the GAN network, despite producing fewer realistic images, demonstrated better image clarity than variational autoencoders (VAEs). Tirupattur et al. [26] also utilized ACGAN-based networks for generating images using EEG signals. The GAN model presented in this research provides learning distributions with limited training data.

Wang et al. [8] aimed to incorporate human aspects into the design process. In their proposed method, EEG signals were employed to bring individuals' preferences into the realm of product design. In their method, an LSTM network extracts features from EEG signals. Then, an ACGAN model was trained using the extracted features.

In another study by Wang et al. [27], a framework for better understanding cognitive brain function during mental imagery was developed using EEG and fMRI. Their proposed model involves decoding brain activities into images and is based on AlphaGAN. Wang et al. [28] also presented a framework for uncovering the impact of design on the brain by reconstructing mental imagery. Their method utilizes an RNN network as an EEG encoder and a conditional generative adversarial network (CGAN) to reconstruct mental imagery.

Deng et al. [29], aiming to reconstruct images from EEG signals, compared the capabilities of several EEG encoders, including the Transformer and the Capsule Network (CapsNet). They proposed a distribution-to-distribution mapping network to transform the extracted features from the encoder into a feature vector containing features of the observed images. Then they utilized a pre-trained model, IC-GAN, for image generation.

To regenerate high-quality perceived images from brain signals, Khare et al. [30] proposed an EEG encoder followed by a conditional Progressive Growing of GANs framework. Their EEG classifier was designed using RNN, LSTM, GRU, and a combination of LSTM and GRU architectures. Mishra et al. [31] presented a method for synthesizing visual stimuli for display during EEG acquisition. They developed an innovative GAN framework that incorporates attention mechanisms and an auxiliary classifier. Additionally, they utilized perceptual loss and attention modules to enhance the quality of the generated images.

1.1 System Overview

This paper aims to explore human involvement in the design of artificial intelligence to create a design that considers personal priorities. Human preferences and priorities can be inferred from EEG signals. This framework, illustrated in Figure 1, has two stages: training and testing.

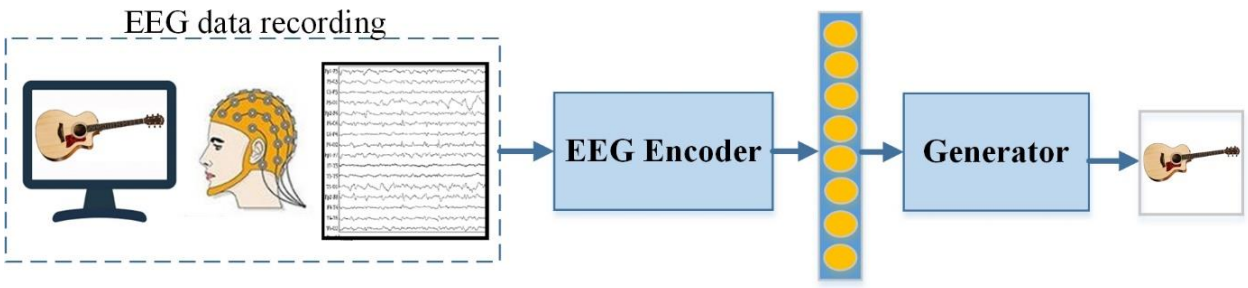

Figure 1 System Overview.

1.1.1 The Training Stage

This stage involves training the EEG encoder, followed by training the generator model using EEG signals and their corresponding images. In the training phase, EEG signals from individuals are recorded while they view photos of the product and then converted into feature vectors using an EEG encoder. The EEG encoder method can be based on different deep learning methods for extracting different features, such as frequency, time-frequency, or statistical feature extraction. EEG features are embedded as a condition for image generation in the GAN-based generator, forcing it to reconstruct images that possess the meaningful features of the original observed image.

1.1.2 The Testing Stage

This stage involves generating an image using the trained generator and the recorded EEG signal from the moment of imagining the product image. During the second stage, participants are instructed to mentally visualize the presented product images while their EEG signals are being recorded. These signals, potentially containing desired design features, are encoded and fed as input to the pre-trained generator. The generator then creates a design that incorporates the envisioned design features of the participants.

The key contributions of the proposed framework can be summarized as follows: a novel transformer-based approach for reconstructing images from EEG signals is introduced, utilizing only transformer blocks for extracting features from EEG signals. This transformer-based EEG encoder simultaneously extracts both time and frequency features from the signals. The high classification accuracy achieved from these features demonstrates their enhanced discriminability, leading to a more accurate generation of category-based images.

Unlike previous EEG-to-image reconstruction models that commonly rely on CNN-LSTM or CNN-RNN hybrids, the present study introduces a pure Temporal-Spectral Transformer encoder that jointly models temporal dependencies and spectral magnitudes extracted from multi-channel EEG. The Transformer's multi-head self-attention enables the model to capture long-range temporal correlations across 64 EEG channels while simultaneously integrating frequency-domain features (0.5-50 Hz), forming a unified and highly discriminative representation. To further enhance the performance of the ACGAN-based network and address challenges such as mode collapse and vanishing gradients, several performance-stabilization techniques previously proposed for GANs have been incorporated in this work. Additionally, a new EEG dataset based on visual stimuli has been recorded and explicitly introduced for EEG-based image reconstruction applications.

1.2 Limitations of the System

We recognize the critical role of large datasets in training deep learning models, as larger sample sizes improve the reliability and robustness of these systems. Unfortunately, due to time and location constraints and limited access to recording equipment, we were unable to collect data from a broader sample. Acknowledging this limitation, we recommend that future research aim to address it by including a more diverse and extensive participant pool. To mitigate this challenge, we applied several techniques, such as windowing EEG signals and augmenting the related images. Additionally, we employed a two-step training process during the image generation phase to further enhance model performance.

2. Materials and Methods

2.1 Experimental Data

In this paper, for recording a visual stimuli-based EEG dataset, stimuli consisted of five distinct product images (guitar, headphone, mug, backpack, and chair). These product categories were selected for their recognizability and commonality, ensuring that participants had comparable familiarity with each. Each category consisted of 50 images, resized to dimensions of 768 × 1024 pixels, and presented at the center of the display screen.

This human study received approval from the Research Ethics Committees of the University of Tabriz under the identification number IR.TABRIZ.REC.1402.084. All participants provided informed consent to take part in the experiment. The study involved 4 women and 4 men, right-handed, with ages ranging from 18 to 35 years, and normal or near-normal vision.

The ANT Event-Related Potential (ERP) recording system, accessible at the Cognitive Neuroscience Laboratory in Tabriz University's central laboratory, was used to record the signals at a sampling rate of 250 Hz. This amplifier's 64 channels were mounted on a wave guard cap according to the extended version of the international 10-20 system. Every electrode impedance was kept below 5 kΩ, and noise was removed from the signals by applying a band-pass filter with a frequency range of 0.5 to 50 Hz. Synchronization between the EEG and stimuli was achieved via event markers automatically sent from the presentation software at the onset of each image. Each 2-s epoch was timestamped based on these markers, ensuring millisecond-level alignment between EEG and visual stimuli.

EEG signal collection was performed in two stages, as shown in Figure 2. The first stage involved simultaneous signal collection while displaying an image, and the second stage involved signal collection after displaying the image. The data collected from the first stage were used to train the model, and the data from the second stage were used to evaluate the model. Throughout the experiment, participants were seated in a nearly silent room on a comfortable chair. Additionally, participants were given access to a push button to record their feedback throughout the experiment. Rest intervals were considered between each stage.

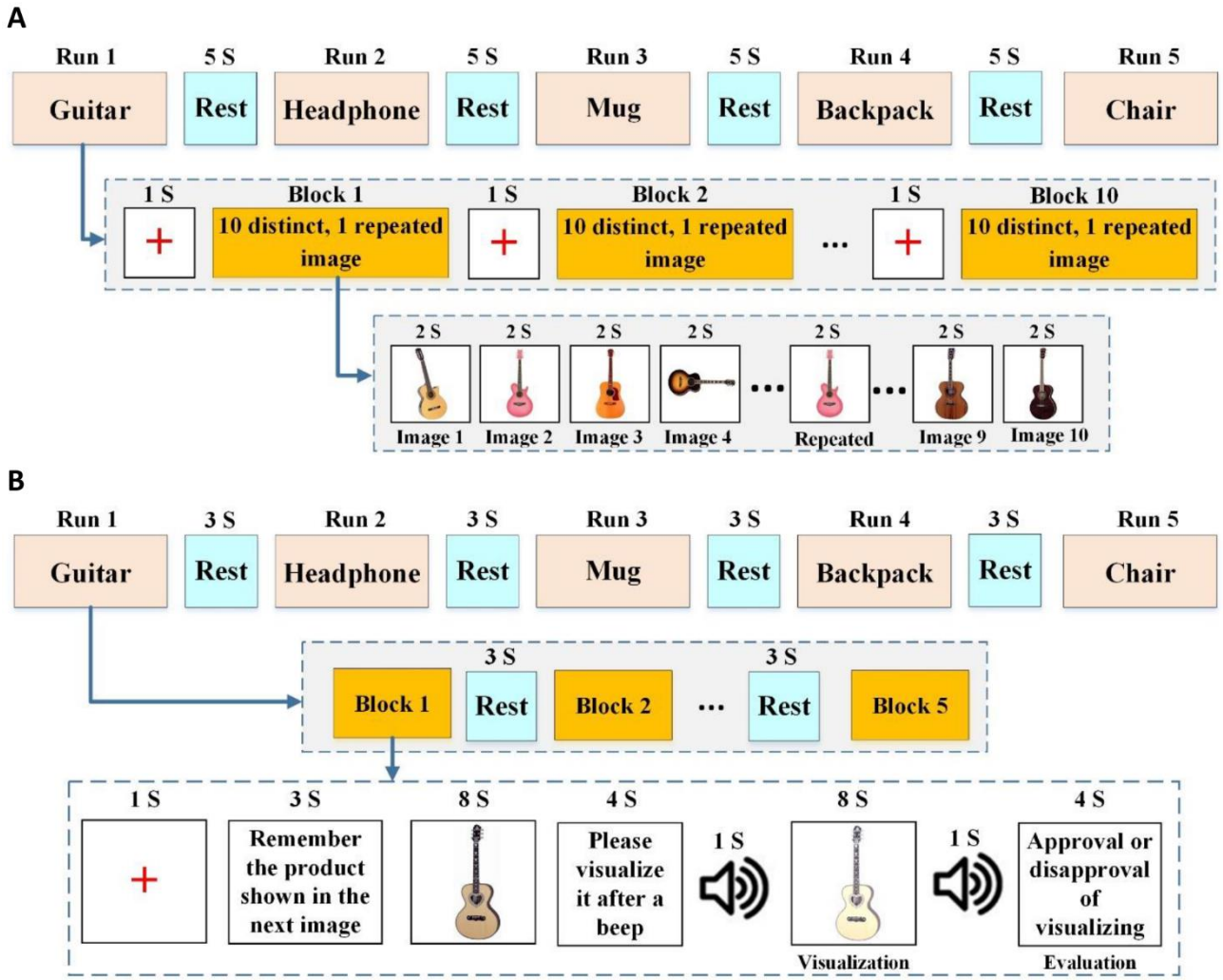

Figure 2 Stages of signal recording. (A) First stage: signal recording while viewing images, (B) Second stage: signal recording while visualizing images.

In the first stage, involving signal recording while viewing images, five image categories were presented across five runs. Each run comprised a set of 50 images arranged in five blocks. Each block contained 10 distinct images and one repeated image. A 5-second rest period followed the conclusion of each run. Participants were instructed to observe the images and press a designated button when identifying the repeated image, ensuring sustained attention throughout the experiment. A red fixation was displayed for 1 second at the start of each block, and an additional 5 seconds were allocated as rest time at the end of each run.

In the second stage, a single image of the selected products was shown to the participant. They were asked to visualize the same image with their eyes closed from the moment the first beep sounded until the next beep, for 8 seconds. Finally, in a survey, participants were asked to evaluate the experiment's accuracy and to assess their perception of the product by pressing one of two approval or disapproval buttons. Cases reported as disapproving responses were excluded from the dataset. This stage included 5 runs, each consisting of 5 blocks. Initially, red fixation was shown in the center of the display for 1 second, and after each block, 3 seconds were considered for rest. Following this, instructions appeared in the middle of the screen, and participants were asked to follow them.

2.2 Preprocessing

In this paper, the preprocessing method for EEG signals comprises two main stages. First, epochs are separated based on event markers. Subsequently, the extracted epochs are used as input to Independent Component Analysis (ICA) to reduce noise and artifacts, specifically eye-movement and muscle noise. Additionally, a band-pass filter ranging from 0.5 to 50 Hz is applied during data acquisition to reduce noise further.

The duration of each image viewing is two seconds with a sampling rate of 250 Hz. These 500 recorded samples for each image are divided into 5 segments, each containing 100 samples. All segments extracted from a given epoch inherited the product-category label of the corresponding image. The data are standardized before entering the model. Standardization, or Z-score transformation, is applied separately to each channel of the recorded data.

2.3 Transformer-Based Feature Extraction

Transformers [32] solved many problems in natural language processing (NLP) translation. Additionally, they have been applied in various domains, including text classification, image processing, and time series analysis. The Multi-Head Self-Attention Layer, Feed-forward blocks, Residual connections, Add and Layer Normalization constitute the structure of an encoder in a transformer-based feature extraction model.

A multi-head self-attention mechanism is employed to capture relational information among features. This mechanism segments the input into distinct portions and projects them onto multiple sets of weight matrices, including values (V), keys (K), and queries (Q). The Scaled Dot-Product Attention operates concurrently on each input projection, producing an output of the exact dimensions.

To expedite network training, residual connections and layer normalization are incorporated. The network takes a coded feature vector as input and produces the reconstructed feature vector that encapsulates the information regarding feature relationships in its output.

In this study, the Pre_LN Transformer models are employed. In this type of Transformer, the model input is normalized before each self-attention and feedforward layer. This normalization serves as a form of regularization, preventing overfitting [33].

According to Figure 3 in the Temporal-Spectral Transformer (TS-Transformer) model, to capture temporal dependencies, EEG data at the same time point are treated as features along the channel axis, and correlations across different time points are calculated. For extracting frequency dependencies, after transposing the input data, the next step involves computing the spectral magnitude of each channel. Then, the data is considered along the channel feature axis by applying another transpose; after that, the output is entered into a fully connected layer with 64 units to match the dimension of the attention layer. The output of this fully connected layer is fed into the encoding block. In the proposed architecture, the number of encoding blocks (N), the head size, and the number of heads in each multi-head attention layer are set to 4, 128, and 8, respectively. The number of units in the feedforward layers is chosen as 32 and 64. The outputs of the encoding blocks are fed into the global average pooling layer separately, then concatenated. After applying a batch normalization layer, the generated output is considered an EEG signal feature vector.

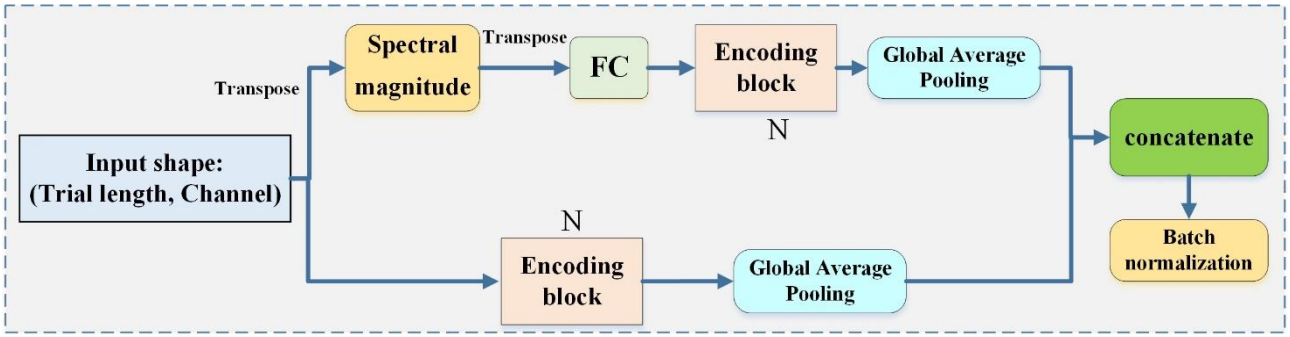

Figure 3 Proposed TS-Transformer-based feature extraction method.

A dropout layer with a 0.5 dropout rate is inserted after each layer in the encoding block. Additionally, each layer is equipped with L1 regularization, utilizing a rate of 0.0005. Throughout the training phase of the EEG encoder, a batch size of 250 is employed. The chosen loss function is cross-entropy. Training of the network involves the use of the Adam optimizer with a scheduled learning rate (initial learning rate = 0.0001, decay steps = 100, decay rate = 0.99). The network undergoes 1000 training epochs.

To evaluate the extracted features, the feature vector is fed into a fully connected layer with 100 units. The number of units in the output layer is equal to the number of classes, which is 5, and its activation function is defined as Softmax.

2.4 Generative Adversarial Network

The Generative Adversarial Network (GAN) was introduced in 2014 by Goodfellow et al. [34]. The GAN structure models a min–max game, with a generator (G) and a discriminator (D) competing during training. The generator model autonomously learns to produce patterns present in the input data, enabling it to generate new samples (outputs). GAN's capability to create new content has contributed to its popularity, particularly in the generation of realistic images.

Among the most critical challenges in training and optimizing GAN networks, mode collapse and vanishing gradients are notable. Mode collapse occurs when the generator consistently produces similar images, and the discriminator is unable to distinguish between them. In this scenario, the generator easily deceives the discriminator. This condition limits the learning of the generator, focusing on a restricted set of images instead of generating realistic ones. Vanishing gradients also occur in two situations. The first situation happens when the discriminator becomes overly strong and does not provide sufficient gradients as feedback to the generator. Consequently, the generator is unable to produce competitive samples. The second situation occurs when the discriminator becomes too weak, and the generator incurs no cost to create realistic samples [35].

Subsequently, a series of methods to address GAN network challenges and improve the network's output, which have been utilized in this paper, will be presented.

Minibatch Standard Deviation: ProGAN [36] proposed using minibatch standard deviation to address mode collapse. In this method, a feature map is added to the end of the discriminator, essentially to the end of each image in the minibatch. In the presented approach, first, the standard deviation is calculated for each feature on each feature map in the minibatch. Then, the estimates associated with each feature are averaged across all feature maps, resulting in a unit value. This unit value is replicated and appended to the dimensions of the feature map, creating an additional

(constant) feature map. This layer can be placed anywhere in the discriminator, but the suggested location for this feature map is the final layer of the discriminator.

- *Pixel Normalization:* Instead of using batch normalization, typically employed in the generator and discriminator architectures, the use of the pixel-wise normalization layer has been suggested in ProGAN [36]. This layer has no trainable weights and normalizes the feature vector at each pixel to unit length. It is applied after the convolutional layers in the generator to prevent signal magnification during training.
- *Spectral Normalization* [37]: It is a normalization technique used in GAN networks to stabilize discriminator training. It prevents vanishing gradients by continuously differentiating GAN gradients.
- *Label smoothing:* One-sided label smoothing is a technique that is used by Salimans et al. [38] for improving GANs. Label smoothing replaces the 0 and 1 labels in the discriminator with smoothed values, like 0.1 and 0.9. This technique acts as regularization and prevents the discriminator from giving a huge gradient signal to the generator.

As GAN cannot control the mode of the generated data, the Conditional Generative Adversarial Network (CGAN) [34] introduces additional information to both the generator and the discriminator to guide the data generation process.

Auxiliary Classifier Generative Adversarial Network (ACGAN) [39] is a kind of conditional GAN. In this architecture, in addition to distinguishing whether the generated images are real or fake, the discriminator is also responsible for classifying the images; Therefore, the ACGAN loss function is composed of two loss functions, L_S and L_C, defined as equations (1).

$$\begin{aligned} L_S &= E[\log P(S = \text{real} \mid X_{\text{real}})] + E[\log P(S = \text{fake} \mid X_{\text{fake}})] \\ L_C &= E[\log P(C = c \mid X_{\text{real}})] + E[\log P(C = c \mid X_{\text{fake}})] \end{aligned} \tag{1}$$

In equation (1), c represents the class label of the sample, and $\mathrm{P(c|x)}$ represents a probability over the class label that the classifier C has calculated. L_C is the log-likelihood of the correct class and, L_S is the log-likelihood of being real or fake. The discriminator is trained to maximize $L_S + L_C$, while the generator aims to maximize $L_C - L_S$.

Since CGAN and ACGAN are advanced GAN architectures, they use the Jensen-Shannon Divergence (JSD) [34] as the distance metric. Therefore, they exhibit similar challenging behaviors as GAN, such as mode collapse and unstable training, especially in highly imbalanced datasets, exacerbating the mode collapse issue [40].

In this study, an ACGAN-based architecture conditioned by EEG has been selected for image generation. The overall architecture is shown in Figure 4.

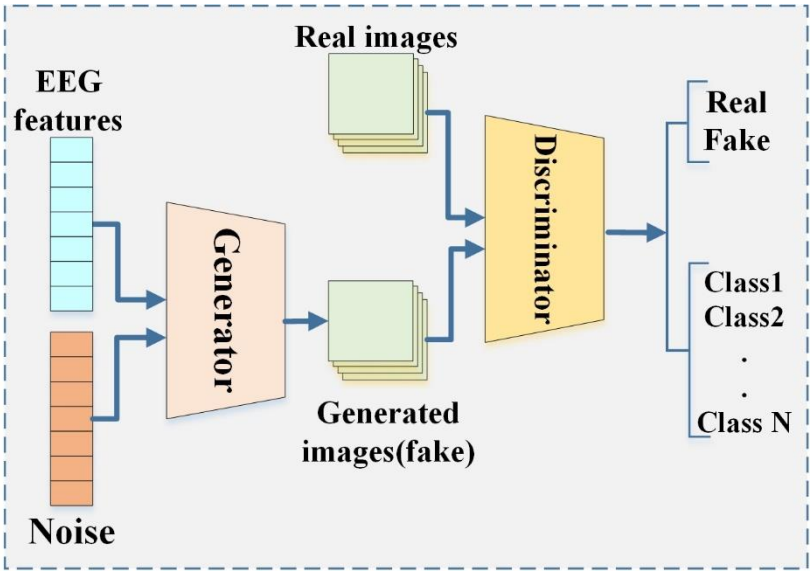

Figure 4 General view of ACGAN-based architecture conditioned by EEG.

Generator Architecture: According to Table 1, the generator of the proposed ACGAN-based model consists of two dense layers, followed by four 2D transposed convolutional layers with the LeakyReLU activation function, and a final transposed convolutional layer with the Tanh activation function. The input to the generator is composed of two vectors of dimension 200, including the feature vectors extracted from the EEG encoder and the noise vector. The EEG feature vector is fed into a dense layer, producing an output of 16 units. Similarly, the noise vector passes through a 8192-unit layer. The outputs of these two dense layers are concatenated and fed into a transposed convolutional layer after reshaping. The kernel size for each of these transposed convolutional layers is set to 4 × 4, with a stride size of 1 for the first layer and 2 for the rest. The number of filters is 512, 256, 64, 128, and 3, respectively. After each transposed convolutional layer, a pixel normalization layer is applied.

Table 1 The architecture of the generator.

operation	kernel	strides	padding	Number of filters	Output shape
Input1					(None, 200)
Input2					(None, 200)
Dense Layer (Input1)					(None, 8192)
Dense Layer (Input2)					(None, 16)
Reshape1					(None, 4, 4, 512)
Reshape2					(None, 4, 4, 1)
Concatenate					(None, 4, 4, 513)
Transposed convolution, pixel normalization, activation LeakyReLU (alpha = 0.2)	4 × 4	1	yes/same	512	(None, 4, 4, 512)
Transposed convolution, pixel normalization, activation LeakyReLU (alpha = 0.2)	4 × 4	2	yes/same	256	(None, 8, 8, 256)
Transposed convolution, pixel normalization, activation LeakyReLU (alpha = 0.2)	4 × 4	2	yes/same	128	(None, 16, 16, 128)
Transposed convolution, pixel normalization, activation LeakyReLU (alpha = 0.2)	4 × 4	2	yes/same	64	(None, 32, 32, 64)
Transposed convolution, pixel normalization, activation tanh	4 × 4	2	yes/same	3	(None, 64, 64, 3)

Discriminator Architecture: In the proposed ACGAN-based architecture, the discriminator plays a dual role: acting as a classifier to identify the image's class (a non-adversarial task) and discerning whether the generated image is fake or real (an adversarial task). Therefore, the cost incurred for updating the network backpropagates through both adversarial and non-adversarial pathways.

The discriminator in the ACGAN comprises five 2D convolutional layers with a LeakyReLU activation function. The kernel size for each convolutional layer is 4 × 4, and the stride size is 2 for all layers. The number of filters is set to 32, 64, 128, 256, and 512, respectively. Spectral normalization is applied to the weights of these layers. The output of the last convolutional layer is connected to a minibatch standard deviation (minibatch_stdev) layer, followed by a flatten layer. The flattened output enters a dense layer with 512 dimensions. For classification, the discriminator has an adversarial head to discern fake or real images generated by the generator. This head consists of a fully connected layer with 1 unit and a Sigmoid activation function. The binary cross-entropy loss function is employed to calculate the cost of the adversarial head, and label smoothing is

applied to the labels. The other head of the discriminator is connected to a fully connected layer with units equal to the number of classes, followed by a Softmax activation function to identify the desired class. The Sparse Categorical Cross-Entropy loss function is used for the classifier head. The architecture of the discriminator is shown in Table 2.

Table 2 The architecture of the discriminator.

operation	kernel	strides	padding	Number of filters	Output shape
input					(None, 64, 64, 3)
Convolution 2D, Spectral Normalization, activation LeakyReLU (alpha = 0.2), Dropout (0.5)	4 × 4	2	yes/same	32	(None, 32, 32, 32)
Convolution 2D, Spectral Normalization, activation LeakyReLU (alpha = 0.2), Dropout (0.5)	4 × 4	2	yes/same	64	(None, 16, 16, 64)
Convolution 2D, Spectral Normalization, activation LeakyReLU (alpha = 0.2), Dropout (0.5)	4 × 4	2	yes/same	128	(None, 8, 8, 128)
Convolution 2D, Spectral Normalization, activation LeakyReLU (alpha = 0.2), Dropout (0.5)	4 × 4	2	yes/same	256	(None, 4, 4, 256)
Convolution 2D, Spectral Normalization, activation LeakyReLU (alpha = 0.2), Dropout (0.5)	4 × 4	2	yes/same	512	(None, 2, 2, 512)
Concatenate with minibatch_stdev					(None, 2, 2, 513)
Flatten					(None, 2052)
Dense, Dropout (0.5), activation LeakyReLU (alpha = 0.2)					(None, 512)
Dense, activation sigmoid					(None, 1)
Dense, activation softmax					(None, 5)

The training of the image generation network is performed in two stages. In the first stage, 7500 images (1500 for each category) are augmented and used to train the proposed model. This image data set was collected manually from ImageNet and the internet. Data augmentation techniques

for this stage include flip, contrast, and rotation. The conditional vector for this stage is the average of the extracted Feature vectors for each class. This stage of training is conducted for 400 epochs.

In the second stage, the pre-trained network from the first stage is used, and the network is further trained using images and feature vectors extracted from the EEG signal recording stage for an additional 100 epochs. The use of recorded signals and the assignment of an image to each signal trial is illustrated in Figure 5. Each image was observed for two seconds at a sampling rate of 250 Hz. The 500 recorded samples for each image are divided into 5 parts, with 100 samples each. One part is set aside as a test dataset, and for the remaining 4 parts, one sample from the augmented data for the original image is assigned. The optimization function is the Adam optimizer with learning rates of 0.002 for the generator and 0.008 for the discriminator.

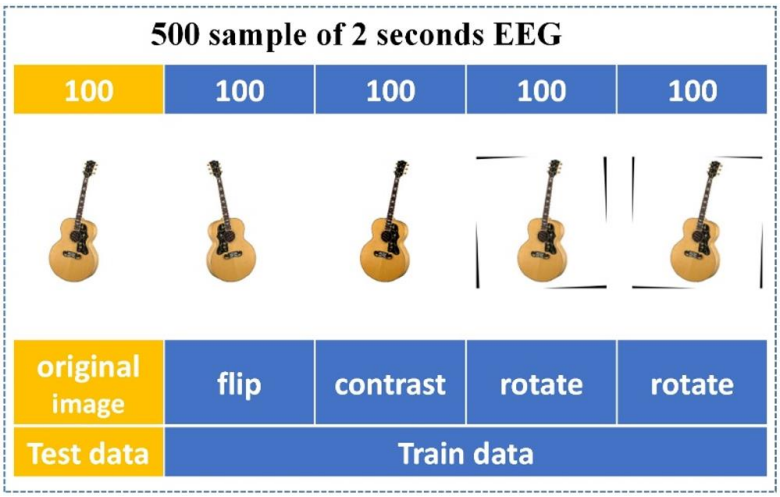

Figure 5 Separating recorded signals and assigning an image to each signal trial to train the generator model in the second stage of training.

2.5 Ethics Statement

The proposed EEG-to-image system is intended solely as an assistive tool for enhancing communication between users and designers, not as a covert "mind-reading" technology. All participants provided informed consent, and ethical approval was obtained from the Research Ethics Committee of the University of Tabriz (IR.TABRIZ.REC.1402.084). Any future real-world deployment of such neuromarketing tools must prioritize transparency, voluntary participation, privacy, and responsible data usage.

3. Evaluation Metrics

The evaluation of the proposed model is conducted in two stages. The first stage involves evaluating the EEG encoder, and the second stage focuses on assessing the GAN network for generating synthesized images.

3.1 EEG Encoder Evaluation

The confusion matrix is employed to compute the accuracy and F1-score. These metrics provide a comprehensive evaluation of the EEG encoders as classifiers. Equations (2) and (3) are used to

calculate Accuracy and F1-score, respectively. The elements of the confusion matrix include True Negative (TN), True Positive (TP), False Positive (FP), and False Negative (FN).

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (2)$$

$$F1 - score = 2TP/(2TP + FP + FN) \quad (3)$$

The Receiver Operating Characteristic (ROC) curve is a key metric for assessing classifier performance. The Area under the ROC Curve (AUC) indicates how much the model can distinguish between classes. The closer the AUC value is to one, the more it indicates the accurate performance of the classification system [41, 42].

3.2 Evaluation of Generated Images Using GANs

To evaluate the GAN network and the generated images, the following metrics can be used.

3.2.1 Inception Score (IS)

The Inception Score is a widely recognized metric for evaluating the quality of generated images. The Inception Score was introduced by Salimans et al. [38] as a metric for evaluating the quality of generated images, especially synthetic images produced by GAN models. Initially, using a pre-trained model, the conditional class probabilities for each generated image, $p(y \mid \boldsymbol{x})$ are calculated. Images with a meaningful object should have a low conditional entropy. Additionally, the model is expected to generate diverse images; therefore, the marginal integral $p\big(\, y \mid \boldsymbol{x} = G(z)\,\big)dz$ should have high entropy. By combining these two requirements, the Inception Score is defined as Equation (4). KL is the Kullback-Leibler divergence.

$$IS = \exp\Big(\mathrm{E}_{\boldsymbol{x}}\mathrm{KL}\big(p(\, y \mid \boldsymbol{x}\,) \parallel p(y)\big)\Big) \quad (4)$$

3.2.2 Classification Accuracy of Generated Images

The accuracy obtained when classifying generated images with a pre-trained classifier serves as a metric for evaluating the generator's performance. The goal of this method is to determine if the generated images from the generator have sufficient quality to be classified [25].

4. Results

The results of the proposed model will be presented in two separate sections for the EEG coder as a classifier and image generation using the proposed ACGAN-based model.

4.1 EEG Encoder

The objective of the EEG encoder is to build a model for extracting EEG features that are as relevant as possible to image-related features. In this stage, 80% of the EEG data is used for training the model, 10% for validation, and the remaining 10% for testing and the final evaluation of the employed model.

The results of the proposed TS-Transformer classifier, as indicated by Accuracy, F1-score, and AUC, are shown in Table 3. All the presented results are based on the test dataset, where a 0.928 F1-score and 92.8% accuracy are obtained. The highest accuracies on the training and validation datasets are 95.5% and 93.9%, respectively.

Table 3 Accuracy, F1-score, and AUC of the TS-Transformer.

Classifier	Accuracy (%)	AUC	F1-score
TS-Transformer	92.8	0.99	0.928

Figure 6A displays the confusion matrix. The analysis of the confusion matrix reveals that the chair class exhibits the highest accuracy, correctly identifying 191 out of 200 images in the test dataset. Conversely, the guitar class demonstrates the least satisfactory performance, misclassifying 19 items as mug, 1 as headphones, and 2 as a backpack. Figure 6B illustrates the ROC curves for each of the five classes in the test dataset using a one-vs-rest approach. As shown, the model achieves exceptionally high AUC scores across all classes (chair = 1.00, headphone = 1.00, mug = 0.99, guitar = 0.98, backpack = 1.00), suggesting excellent separability between positive and negative instances for each category. These results highlight the robustness of the proposed model in handling multi-class prediction tasks, with near-perfect discrimination for most object categories.

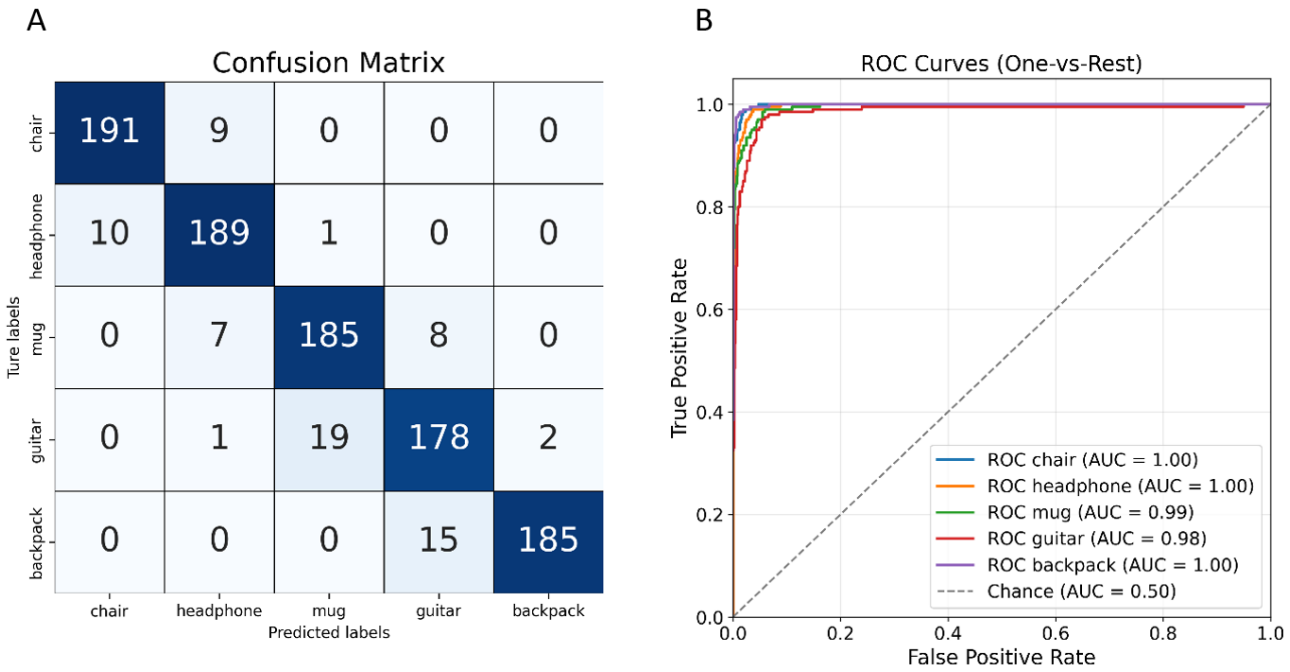

Figure 6 ROC curve and Confusion matrix using the test dataset. (A) Confusion matrix; (B) ROC curve.

The t-SNE observations, as shown in Figure 7, reveal that the transform-based models successfully increase inter-category separations while reducing intra-category feature distances. In contrast, the raw EEG signal categories exhibit significant overlap, making them difficult to distinguish. These results validate the robustness of the feature extractor, as the extracted features are distinguishable across multiple categories.

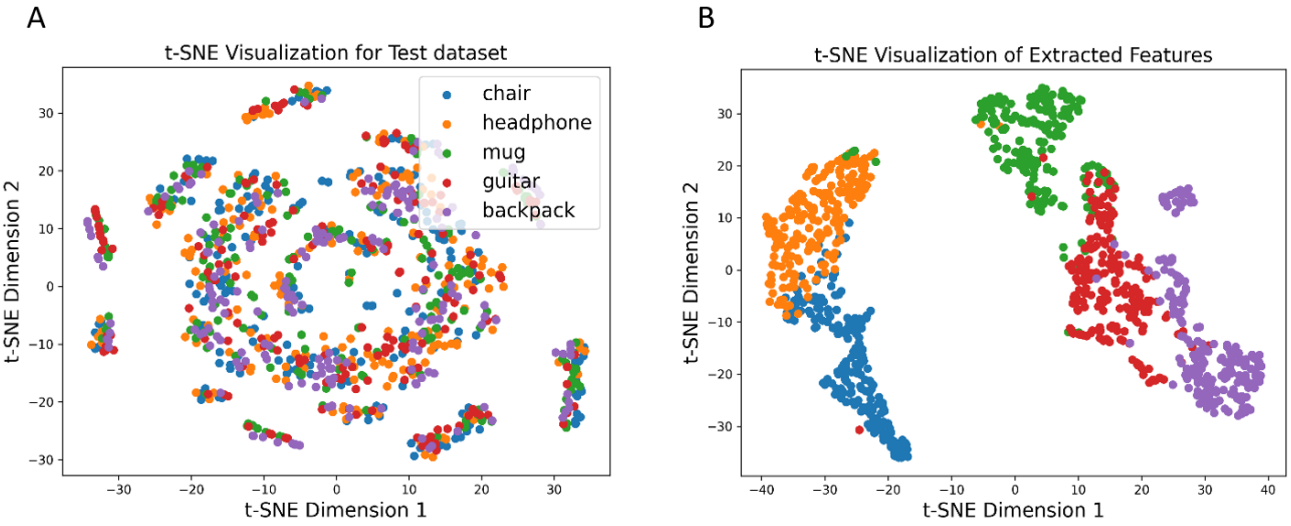

Figure 7 t-SNE visualization of (A) the raw test dataset, (B) the extracted features from the test dataset.

4.2 Evaluation of the Generated Images

The evaluation of the generated images based on IS and classification accuracy is presented in Table 4. In this paper, to evaluate the classification accuracy of the generated image, the ViT_B16 network has been employed. This network is one of the pre-trained ViT-Keras configurations on the ImageNet dataset. To use this classifier for evaluating generated images, a series of changes has been made to the number of parameters in this network. The accuracy of this model, trained on images from five products, is 98%. Figure 8 provides samples of presented images and well-generated images from the generator, while Figure 9 shows samples of poorly generated images by the generator. The results are obtained using the test dataset, which was not used during model training.

Table 4 Results obtained from the images generated by the generator.

Evaluation metrics	IS	Classification accuracy of generated images
For the test data set	5.22	78%

Figure 8 Samples of presented images from the first stage of signal recording and good results of generated images using these recorded signals.

Figure 9 Bad results of generated images using recorded signals from the first stage of signal recording.

4.3 Using the Model

After training the encoder and generator, the trained model is used for the second stage. During model testing, data collected in the second data recording stage is fed into the trained generator for image generation. Figure 10 illustrates the result of the second stage - testing the model.

Figure 10 Sample images observed (left) and generated (right) from the second stage of data recording.

5. Discussion

Multi-channel EEG signals contain information within and between channels and also show long-term relationships in the frequency and time domains. Consequently, it is possible to extract the temporal, spectral, and spatial dependencies of EEG signals for classification applications. To address the limitations of RNN and LSTM in handling very long sequences due to their sequential training approach, increasing the time of convergence can be mentioned. Various deep learning methods for extracting temporal features have been explored. These include CNN-based models, which excel at capturing local interactions, and RNN-based models, which are effective at extracting long-term dependencies. Despite their strengths, these models still face challenges in terms of convergence time [19].

Transformer-based models leverage the attention mechanism to process the entire sequence simultaneously, enabling them to capture long-term dependencies without being constrained by sequence length. This stands in contrast to RNN or LSTM, where the sequential processing can lead to slower convergence times. Transformers offer faster computation speeds, thanks to their ability to perform parallel computations.

Since the accurate generation of images for each product depends entirely on the encoder's accuracy, the discriminative power of the encoder becomes a crucial factor; Therefore, the classification results for the encoder are presented in the form of a classifier.

While the proposed system decodes visual features from EEG signals, it still faces a significant gap with the concept of "mind-reading" and generating designs perfectly aligned with the user's taste. The presented method attempts to map EEG signals into feature vectors corresponding to image features. Some presented image features, such as color and size, are observable in specific generated images.

Table 5 provides a general comparison based on metrics like IS and classification accuracy for generated images from the generator in various proposed methods. The IS metric is employed to assess realism (whether the generated images resemble a specific object) and diversity (whether a broad spectrum of objects is generated).

Table 5 Overall comparison based on IS and classification accuracy for image reconstruction using EEG signals.

	IS	Classification accuracy of generated images (%)	Number of classes
Palazzo et al. [25] and Kavasidis et al. [16]	5.07	43	40
Wang et al. [8]	4.9	-	5
Deng et al. [29]	2.18	-	5
Khare et al. [30]	5.07	-	40
Proposed model	5.22	78	5

A dataset with a larger number of classes typically leads to a higher Inception Score (IS) if the generated images capture this diversity. The IS rewards images that cover a wide range of courses and are confidently classified into distinct categories. However, when the dataset contains fewer classes, even if the GAN produces diverse images within those classes, the score may be limited due to the inherent lack of class variety.

According to Table 5, the IS obtained from the generated images using the proposed method exceeds that of other approaches with the same number of classes. The number of classes in the dataset used in this study is significantly lower than that of the databases used in the studies by Khare et al. [30], Palazzo et al. [25], and Kavasidis et al. [16], all of which employed 40 different classes.

From a neurophysiological perspective, our EEG features encompass the 0.5-50 Hz frequency range, which includes delta, theta, alpha, beta, and low-gamma bands known to contribute to visual and attentional processing. The 64-channel montage, covering occipital, parietal, and temporal sites, captures activity from cortical regions associated with object perception and category recognition. Although the present study focuses on algorithmic feasibility, future work may analyze Transformer attention weights to identify the most informative cortical regions and frequency bands contributing to image reconstruction. Additionally, future work can incorporate leave-one-subject-out validation to directly quantify between-subject variability and further assess the robustness of the proposed framework.

Age-related factors and neurodegenerative conditions such as Alzheimer's disease are known to alter neural representations and the way brain signals encode perceptual or cognitive information. Recent machine-learning studies have demonstrated that computational models can sensitively capture these neural changes [43, 44]. Although the present study focuses on EEG-based visual decoding in a neuromarketing context, future work could examine how aging or early neurocognitive changes might modulate the EEG representations learned by Transformer-based models. Such extensions would help determine whether the proposed framework remains robust across different age groups or neurocognitive profiles.

6. Conclusions

In this paper, a framework is introduced for generating product designs inspired by brain signals. To extract image-related features, a Transformer-based method is developed to extract temporal and spectral features. These features are then fed into ACGAN-based networks to generate product images. To enhance the performance of Generative Adversarial Networks, various methods recommended by different papers have been applied to improve the proposed network's performance. The results indicate that the extracted features from Transformer-based models, when input to Generative Adversarial Networks, can produce images that align with individual preferences, particularly in aspects such as shape and color.

The main limitation of this paper and the presented methods is the insufficiency of data for deep learning-based approaches. Since recording EEG signals is time-consuming and costly, recording a large number of signals was not feasible. For practical applications and to enhance the model's generalization, more data and a wider range of diverse products will be needed.

Author Contributions

Mahsa Zeynali: Conceptualization, Methodology, Data curation; Software, Formal analysis, Writing - Original Draft, Investigation. Hadi Seyedarabi: Conceptualization, Supervision, Writing - Review Editing, Project administration, Validation, Resources. Reza Afrouzian: Conceptualization, Methodology, Investigation.

Funding

This research received no external funding.

Competing Interests

The authors have declared that no competing interests exist.

Data Availability Statement

The raw/processed data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study.

AI-Assisted Technologies Statement

Artificial intelligence (AI) tools were used solely for basic grammar correction and language refinement in the preparation of this manuscript. Specifically, OpenAI's ChatGPT was employed to improve the readability and linguistic clarity of the English text. All scientific content, data interpretation, and conclusions were developed independently by the authors. The authors have thoroughly reviewed and edited the AI-assisted text to ensure its accuracy and accept full responsibility for the content of the manuscript.

References

1. Balconi M, Stumpo B, Leanza F. Advertising, brand and neuromarketing or how consumer brain works. Neuropsychol Trends. 2014; 16: 15-21.
2. Morin C. Neuromarketing: The new science of consumer behavior. Society. 2011; 48: 131-135.
3. Agarwal S, Dutta T. Neuromarketing and consumer neuroscience: Current understanding and the way forward. Decision. 2015; 42: 457-462.
4. Lawhern V, Hairston WD, McDowell K, Westerfield M, Robbins K. Detection and classification of subject-generated artifacts in EEG signals using autoregressive models. J Neurosci Methods. 2012; 208: 181-189.
5. Khurana V, Kumar P, Saini R, Roy PP. EEG based word familiarity using features and frequency bands combination. Cognit Syst Res. 2018; 49: 33-48.
6. Khurana V, Gahalawat M, Kumar P, Roy PP, Dogra DP, Scheme E, et al. A survey on neuromarketing using EEG signals. IEEE Trans Cogn Dev Syst. 2021; 13: 732-749.
7. Yu S, Dong H, Wang P, Wu C, Guo Y. Generative creativity: Adversarial learning for bionic design. In: Artificial Neural Networks and Machine Learning-ICANN 2019: Image Processing. Cham: Springer International Publishing; 2019. pp. 525-536.
8. Wang P, Wang S, Peng D, Chen L, Wu C, Wei Z, et al. Neurocognition-inspired design with machine learning. Des Sci. 2020; 6: e33.
9. Kourtzi Z, Kanwisher N. Cortical regions involved in perceiving object shape. J Neurosci. 2000; 20: 3310-3318.
10. de Beeck HP, Torfs K, Wagemans J. Perceived shape similarity among unfamiliar objects and the organization of the human object vision pathway. J Neurosci. 2008; 28: 10111-10123.
11. Peelen MV, Downing PE. The neural basis of visual body perception. Nat Rev Neurosci. 2007; 8: 636-648.
12. Kepes G. Language of vision. North Chelmsford, MA: Courier Corporation; 1995.
13. Won S, Westland S. Colour meaning and context. Color Res Appl. 2017; 42: 450-459.
14. Machajdik J, Hanbury A. Affective image classification using features inspired by psychology and art theory. In: MM'10: Proceedings of the 18th ACM International Conference on Multimedia; 2010 October 25-29; Firenze, Italy. New York, NY: Association for Computing Machinery; 2010. pp. 83-92.
15. O'Connor Z. Colour, contrast and gestalt theories of perception: The impact in contemporary visual communications design. Color Res Appl. 2015; 40: 85-92.
16. Kavasidis I, Palazzo S, Spampinato C, Giordano D, Shah M. *Brain2image*: Converting brain signals into images. In: MM'17: Proceedings of the 25th ACM International Conference on Multimedia;

2017 October 23-27; Mountain View, CA, USA. New York, NY: Association for Computing Machinery; 2017. pp. 1809-1817.

17. Bashivan P, Rish I, Yeasin M, Codella N. Learning representations from EEG with deep recurrent-convolutional neural networks. Proceedings of the 4th International Conference on Learning Representations, ICLR 2016; 2016 May 02-04; San Juan, Puerto Rico. Ithaca, NY: arXiv; 2016; arXiv:1511.06448. Available from: https://arxiv.org/abs/1511.06448.
18. Spampinato C, Palazzo S, Kavasidis I, Giordano D, Souly N, Shah M. Deep learning human mind for automated visual classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017 July 21-26; Honolulu, HI, USA. Washington, D.C.: IEEE Computer Society; 2017. pp. 6809-6817.
19. Zeynali M, Seyedarabi H, Afrouzian R. Classification of EEG signals using transformer based deep learning and ensemble models. Biomed Signal Process Control. 2023; 86: 105130.
20. Zeynali M, Narimani H, Seyedarabi H. EEG-based identification and cryptographic key generation system using extracted features from transformer-based models. Signal Image Video Process. 2024; 18: 9331-9346.
21. Bagchi S, Bathula DR. EEG-ConvTransformer for single-trial EEG-based visual stimulus classification. Pattern Recognit. 2022; 129: 108757.
22. Mishra R, Bhavsar A. EEG classification based on visual stimuli via adversarial learning. Cogn Neurodyn. 2024; 18: 1135-1151.
23. Deng Y, Ding S, Li W, Lai Q, Cao L. EEG-based visual stimuli classification via reusable LSTM. Biomed Signal Process Control. 2023; 82: 104588.
24. Abbasi H, Seyedarabi H, Razavi SN. A combinational deep learning approach for automated visual classification using EEG signals. Signal Image Video Process. 2024; 18: 2453-2464.
25. Palazzo S, Spampinato C, Kavasidis I, Giordano D, Shah M. Generative adversarial networks conditioned by brain signals. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV); 2017 October 22-29; Venice, Italy. Washington, D.C.: IEEE Computer Society; 2017. pp. 3410-3418.
26. Tirupattur P, Rawat YS, Spampinato C, Shah M. Thoughtviz: Visualizing human thoughts using generative adversarial network. In: MM'18: Proceedings of the 26th ACM international conference on Multimedia; 2018 October 22-26; Seoul, Republic of Korea. New York, NY: Association for Computing Machinery; 2018. pp. 950-958.
27. Wang P, Zhou R, Wang S, Li L, Bai W, Fan J, et al. A general framework for revealing human mind with auto-encoding GANs. Ithaca, NY: arXiv; 2021; arXiv:2102.05236. Available from: https://arxiv.org/abs/2102.05236.
28. Wang P, Peng D, Yu S, Wu C, Wang X, Childs P, et al. Verifying design through generative visualization of neural activity. In: Design Computing and Cognition'20. Cham: Springer International Publishing; 2022. pp. 555-573.
29. Deng X, Wang Z, Liu K, Xiang X. A GAN model encoded by CapsEEGNet for visual EEG encoding and image reproduction. J Neurosci Methods. 2023; 384: 109747.
30. Khare S, Choubey RN, Amar L, Udutalapalli V. Neurovision: Perceived image regeneration using cprogan. Neural Comput Appl. 2022; 34: 5979-5991.
31. Mishra R, Sharma K, Jha RR, Bhavsar A. NeuroGAN: Image reconstruction from EEG signals via an attention-based GAN. Neural Comput Appl. 2023; 35: 9181-9192.

32. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Advances in neural information processing systems. Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017); 2017 December 04-09; Long Beach, CA, USA. Red Hook, NY: Curran Associates Inc.; 2017; arXiv:1706.03762. Available from: https://arxiv.org/abs/1706.03762.
33. Tao Y, Sun T, Muhamed A, Genc S, Jackson D, Arsanjani A, et al. Gated transformer for decoding human brain EEG signals. In: Proceedings of the 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); 2021 November 01-05; Mexico. Piscataway, NJ: IEEE; 2021. pp. 125-130.
34. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: NIPS'14: Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2; 2014 December 08-13; Montreal, Canada. Cambridge, MA, USA: MIT Press; 2014. pp. 2672-2680.
35. Yadav A, Shah S, Xu Z, Jacobs D, Goldstein T. Stabilizing adversarial nets with prediction methods. Proceedings of the 6th International Conference on Learning Representations, ICLR 2018; 2018 April 30-May 3; Vancouver, BC, Canada. Ithaca, NY: arXiv; 2018; arXiv:1705.07364. Available from: https://arxiv.org/abs/1705.07364.
36. Karras T, Aila T, Laine S, Lehtinen J. Progressive growing of GANs for improved quality, stability, and variation. Proceedings of the 6th International Conference on Learning Representations, ICLR 2018; 2018 April 30-May 3; Vancouver, BC, Canada. Ithaca, NY: arXiv; 2018; arXiv:1710.10196. Available from: https://arxiv.org/abs/1710.10196.
37. Miyato T, Kataoka T, Koyama M, Yoshida Y. Spectral normalization for generative adversarial networks. Proceedings of the 6th International Conference on Learning Representations, ICLR 2018; 2018 April 30-May 3; Vancouver, BC, Canada. Ithaca, NY: arXiv; 2018; arXiv:1802.05957. Available from: https://arxiv.org/abs/1802.05957.
38. Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A, Chen X. Improved techniques for training GANs. In: NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems; 2016 December 05-10; Barcelona, Spain. Red Hook, NY: Curran Associates Inc.; 2016. pp. 2234-2242.
39. Odena A, Olah C, Shlens J. Conditional image synthesis with auxiliary classifier GANs. In: Proceedings of the 34th International Conference on Machine Learning, ICML 2017; 2017 August 06-11; Sydney, Australia. Ithaca, NY: arXiv; 2017; arXiv: 1610.09585. pp. 2642-2651. Available from: https://arxiv.org/abs/1610.09585.
40. Liao C, Dong M. ACWGAN: An auxiliary classifier wasserstein GAN-based oversampling approach for multi-class imbalanced learning. Int J Innov Comput Inf Control. 2022; 18: 703-721.
41. Sammut C, Webb GI. Encyclopedia of machine learning and data mining. New York, NY: Springer; 2017.
42. Feizi H, Sattari MT, Mosaferi M, Apaydin H. An image-based deep learning model for water turbidity estimation in laboratory conditions. Int J Environ Sci Technol. 2023; 20: 149-160.
43. Karami V, Ricci G, Pesel G, Nittari G. A computer approach to assess age-related changes of the brain white matter in Alzheimer's disease. Heliyon. 2024; 10: e37836.

44. Karami V, Nittari G, Traini E, Amenta F. An optimized decision tree with genetic algorithm rule-based approach to reveal the brain's changes during Alzheimer's disease dementia. J Alzheimers Dis. 2021; 84: 1577-1584.