Real-Time Thought-to-Vision Generation Using Low-Channel EEG and Feature-Fusion Learning

Abha Marathe, Medha Wyawahare, Milind Rane, Vrinda Parkhi,
Milind Kamble, Anvita Barlingay, Anushka Goswami, Aditya Awate
Department of Electronics and Telecommunication Engineering, Vishwakarma Institute of Technology, Pune, India

***Abstract*—Severe motor disabilities and paralysis make it hard for individuals to communicate their thoughts and express their imagination using standard interfaces. Recent methods that convert EEG signals into images, using diffusion models, have shown superior results. However, these methods usually depend on high-density EEG systems with 32 to 128 channels, deep neural EEG encoders, and large datasets. This leads to high computational costs, poor real-time performance, and limits their use in assistive settings. To address these problems, this paper proposes a Thought-to-Vision system that is lightweight and real-time. In this work, Thought refers specifically to the imagination of simple geometric shapes (circle, square, and triangle) under controlled experimental conditions. This Thought-to-Vision system can decode the imagined geometric shapes from a low-channel EEG system that only requires 2 channels and then produce visual images based on a diffusion model. The EEG signal was recorded at 250 Hz with 150 trials per session, consisting of 50 trials for each circle, square, and triangle shape. The signal was filtered using artifact rejection, 50 Hz notch filtering, and bandpass filtering between 1 and 40 Hz. A Tri-Domain EEG feature fusion (TDEF) that combines spectral features (FFT band power), Time-Frequency features (Daubechies-4 wavelet coefficients), and statistical features was developed and tested against several benchmarks. These included feedforward networks, CNNs, LSTM/GRU-based time-series encoders, CNN-Transformer models, and EEG-CLIP alignment. Evaluation is measured using classification accuracy, precision, recall, and F1 score, along with embedding consistency for semantic alignment. The experimental results indicate that the TDEF with the XGBoost classifier reaches around 94% for classification accuracy, precision, recall, and F1-score. This performance surpasses deep time-series encoders, which achieved up to 39.09% accuracy, and contrastive EEG-CLIP models, which had 82.97% accuracy. The classified EEG embeddings were then used to guide a latent diffusion model, enabling coherent and semantically consistent image generation. These findings confirm that feature-fusion learning with XGBoost can outperform deep EEG encoders in low-channel situations. This offers a solid, efficient, and practical solution for real-time assistive brain-computer interfaces.**

Keywords*—Brain computer interface; feature fusion learning; diffusion models; EEG to image generation; wavelet; EEG signal processing

I. Introduction

The demodulation of human thought and imagination patterns via neural activity is one of the ever-green areas in brain-computer interface research [1]. Electroencephalography (EEG), due to its non-invasive nature and high temporal resolution, has been investigated for its use in cognitive processing translation into machine-readable signals for several applications [2]. While fMRI is characterized by its ability to produce highly accurate images, it is associated with poor temporal resolution and excessive costs. On the other hand, EEG, though having poor spatial resolution, has excellent temporal resolution as well as portability. However, because of its poor spatial resolution, it can only be used for the interpretation of crude representations of visual images, not their minute details. The translation of imagination-related EEG patterns into a meaningful pictorial interpretation has a wide scope in assistive communication systems, neurofeedback, and human-machine interaction contexts. Recent works have focused on the usage of deep neural networks and diffusion-based generative models for the reconstruction of images from brain signals using EEG-to-image synthesis. The potential of aligning the embedding spaces of the brain signals with the visual embedding space has been shown using frameworks such as Brain-Dreamer [3], Dream-Diffusion [4], and EEG-Vision [5]. These frameworks require high-density EEG recording (32 to 128 channels), a massive dataset, and computationally expensive brain decoders, thus making it very difficult to implement them in a realistic scenario. Yet, there is an essential gap in research work: the existence of light-weight EEG-to-image models able to realize real-time processes on a low-channel EEG signal with a high degree of accuracy. In fact, when it comes to actual applications involving a person with motor disorders, it can be more essential to ensure simplicity and portability rather than the generative capability. The over-reliance on a deep encoder when it comes to EEG often results in a poor generalization ability. With the shortcomings in mind, the proposed research seeks to develop a low-channel, real-time Thought-to-Vision generation framework where simplicity of usability is put foremost. Contrary to the requirement of using deep EEG encoders in the existing work, the proposed approach focuses on the concept of feature fusion-based learning, which exploits multiple complementary descriptors of the EEG signal to increase discriminatory efficiency without the associated heavy computations. The key contributions of the current study are the design of a light-weight, two-channel EEG-based Thought-to-Vision system for real-time applications, and the development of a tri-fusion EEG feature representation scheme that combines spectral, time-frequency, and statistical features to efficiently capture complementary information from the brain signals. A feature-fusion–based XGBoost classifier is employed, which consistently outperforms state-of-the-art deep EEG encoders in low-channel settings. Furthermore, the system achieves efficient EEG decoding coupled with diffusion-based

image generation, allowing the visualization of imagined content in a coherent and semantically meaningful manner. Finally, a comprehensive comparative evaluation of classical, feature-based, and deep learning EEG encoders is presented, clearly demonstrating the advantages of feature-based learning approaches for robust, lightweight, and assistive brain–computer interface applications.

This work makes four key contributions: (1) a real-time, low-channel (2-channel) EEG-based Thought-to-Vision framework suitable for assistive deployment, (2) a tri-fusion EEG feature representation combining spectral, time–frequency, and statistical descriptors, (3) a comprehensive empirical comparison of feature-based and deep time-series EEG encoders under low-channel constraints, and (4) a diffusion-conditioned image generation pipeline driven by lightweight EEG decoding.

The rest of the document is structured as follows: Literature review is presented in Section II followed by Methodology in Section III. Experimental results in Section IV, and Conclusion and Future Scope in Section V.

II. Literature Review

The literature review is categorized into five major groups: (1) EEG Signal Processing and Fundamental Concepts, (2) EEG to Image Generation using Deep Learning and Diffusion Models, (3) Multimodal Learning, Retrieval, and Representation Learning, (4) Applications of EEG in Creative, Medical, and Assistive Systems, (5) Surveys, General BCI Systems, and Recent Trends.

A. EEG Signal Processing and Fundamental Concepts

Early research by Kumar and Bhavaneswari [2] laid the foundation for EEG signal analysis by categorizing EEG activity into Delta, Theta, Alpha, Beta, and Gamma frequency bands using Fourier and wavelet-based spectral techniques. Their work demonstrated that band-power features, particularly Alpha and Beta rhythms, are reliable indicators of cognitive and mental states, forming the basis for later EEG preprocessing and feature extraction pipelines. In addition to model development, Hatton et al. [6] investigated various EEG acquisition systems and proved that multi-channel systems can greatly enhance spatial and spectral representation of EEG signals, providing valuable guidance for EEG system implementation.

B. EEG-to-Image Generation Using Deep Learning and Diffusion Models

With the rise of deep learning, EEG-based visual decoding advanced significantly. Wang et al. [3] proposed BrainDreamer, a language guided EEG to image generation framework that aligns EEG, text, and image embeddings in CLIP latent space using contrastive learning and diffusion-based image synthesis. This method facilitated semantically coherent and controllable image reconstruction from EEG signals.Bai et al. [4] proposed Dream Diffusion, which eliminated the need for text-based guidance and proved that diffusion models are superior to GAN based methods for EEG to image reconstruction, achieving more realistic visual reconstructions. Huangtao et al. [5] introduced EEG-Vision, which shows integration of time-frequency EEG features with diffusion models to achieve high quality visual reconstructions. To enhance semantic coherence, Yang and Liu [7] proposed a method integrating CNN based EEG classification with diffusion models, enhances both decoding performance and visual coherence. Chen et al. [8] investigated diffusion-based EEG image reconstruction and proved that structured EEG encoding can enhance reconstruction performance. Similarly, Li et al. [9] proposed a guided diffusion model that utilized EEG embeddings to improve reconstruction performance. Lopez et al. [10] proposed a latent diffusion-based EEG to image generation framework using simulated EEG data, demonstrating the potential for training generative models at scale.

C. Multimodal Learning, Retrieval, and Representation Learning

Retrieval based methods further extended the field of EEG vision research. Sithisint et al. [11] proposed EEG2Face Query, a method for retrieving facial images from EEG signals via CNN and Vision Transformer based latent representation learning. Zhang et al. [12] proposed Cognition Capturer, which associated EEG with multimodal embeddings (image and text) to improve visual decoding performance. Ferrante et al. [13] demonstrated that knowledge distillation enables compact EEG encoders to retain performance when paired with latent diffusion models. Puah et al. [14] proposed a model based on a diffusion model framework, "EEG-DGM" or "EEG Diffusion Model" suggests a novel approach to learning representations of EEG signals using a denoising and diffusion probabilistic model which would be beneficial to the field of EEG classification. Large scale benchmarking was enabled by Zhu et al. [15], who introduced EEG ImageNet, a dataset with multigranularity labels for standardized evaluation. Pan et al. [16] proposed a deep visual representation model for reconstructing stimulus images from EEG signals. Spampinato et al. [17] proposed a deep learning-based model using recurrent neural networks for discriminative embedding learning using EEGs. It is notable in the sense that their work emphasizes the importance of temporal information rather than spatial data since, even though fMRI can capture higher spatial data while being expensive, it lacks the temporal aspect. While the alternative MEG has the advantage of improved spatial resolution, it is non-portable and difficult to operate compared to other modalities such as EEG. This proves that the need for good representation learning through effective embeddings is a necessity for visual decoding from EEG signals. Finally, Zhang et al. [18] extended Cognition Capturer in a journal version, reinforcing the importance of multimodal alignment for high-quality EEG-based visual decoding.

D. Applications of EEG in Creative, Medical, and Assistive Systems

Earlier creative BCI applications like Brain Painting by Münßinger et al. [19] demonstrated the viability of using EEG for creative tasks, particularly for ALS patients. In addition to 2D reconstruction, Guo et al. [20] introduced Neuro3D, demonstrating that it is possible to reconstruct 3D objects from EEG signals using diffusion-based point cloud generation. The application of EEG technology also made its way into creative applications. Liu and Wang [21] proposed Mental Gen, an EEG controlled generative model, for interior space design, which showed the possibility of using brain signals for personalized creative tasks. Soroush et al. [22] systematically reviewed the

EEG characteristic and through experimental design highlighted the EEG driven creativity research. Medical EEG applications were discussed by Foreman et al. [23], employing the quantitative EEG characteristics for the detection of brain ischemia. Pujac et al. [24], shown the artistic applications by visualizing real time EEG data through generative AI art.

E. Surveys, General BCI Systems, and Recent Trends

More general views on EEG based BCIs were presented by Lazarou et al. [1], reviewing the EEG based communication and rehabilitation systems, with a focus on robustness and clinical applicability. A survey conducted by Shukla et al. [25] investigated the combination of EEG and generative AI, noting the importance of diffusion models and multimodal learning. EEG based creativity research was further reviewed by Zeng and Soroush [26], while Liu et al. [27] surveyed recent BCI applications. Cognitive inspired reasoning models like Sketch of Thought were presented by Aytes et al. [28], who proposed light weight representations for efficient reasoning. Advanced cognitive efficiency was revisited by Aytes et al. [29], extending Sketch-of-Thought for adaptive reasoning. In addition, Majima and Nishimoto [30] investigated reconstructing visual imagery from brain activity by applying Bayesian estimation through deep neural networks. This research shows the ability to reconstruct visual imagery from neural activity, thereby reinforcing the possibility of mapping neural activities to visual imagery.

III. METHODOLOGY

The methodology of the proposed work is shown in Fig. 1. It contains the following steps:

A. Dataset Collection

A raw, synchronized dataset of 2-channel EEG signals recorded while a human participant views a series of images shown in Fig.2. The main aim was to record the brain's electrical signals associated with the visual perception of geometric shapes. EEG signals were recorded using a low-cost two-channel recording system. Data was collected using a controlled visual stimulation paradigm to record consistent neural responses associated with basic geometric shape perception. During the recording process, three shapes were presented before the subject i.e., circle, square, and triangle, with 50 different stimuli for each shape, making a total of 150 trials per recording session. Each stimulus was displayed for 500ms, followed by a 500ms inter-stimulus interval using a blank screen to avoid overlap of successive cognitive responses. To avoid fatigue and cognitive drift, trials were arranged in shape blocks with short rest intervals between shape blocks. EEG signals were recorded at a sampling rate of 250 Hz and synchronized with the stimulus display using event marker logs, allowing for precise segmentation of neural responses associated with each visual stimulus.

B. Data Processing

To increase the quality of the signal and guarantee the extraction of the unique features of the neural activity, a preprocessing step was carried out on the raw EEG data. The EEG signals were recorded at a frequency of 250 Hz, and trial-specific segments were extracted around the synchronized event markers. The preprocessing pipeline included three steps in a sequence: Artifact rejection, removal of power line noise, and band pass filtering. Each of these steps was designed to remove non neural interference while retaining the cognitive related brain signals.

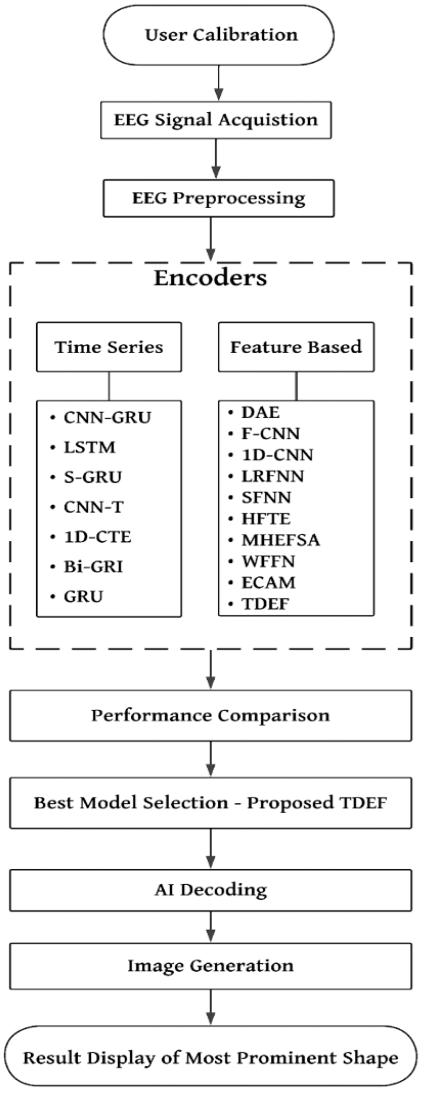

Fig. 1. Flowchart of the project.

Fig. 2. Dataset collection process.

First, we discarded the epochs that were artifact rich by using an amplitude-based rejection, which was focused on non-cortical physiological artifacts such as eye blinks and muscle activity. Any trial that had instantaneous EEG amplitudes outsides the ±100 μV range was rejected, as is standard practice in ERP analysis. In formal terms, any trial i, with EEG signal is rejected for the condition given in (1).

$$\exists t : |x_i(t)| > \theta_{artifact} \quad (1)$$

The value of $\theta_{artifact} = 100.0\mu V$.

This criterion is extremely helpful in suppressing the large artifact perks without affecting the trials that show the existence of neural activity.

Secondly, to eliminate the power line noise we used a notch filter at 50 Hz. It is an IIR notch filter with Q=30. It is used as a zero- phase forward and backward filter to eliminate any phase shift. This ensures that the temporal fidelity of the neural oscillations is maintained and the power line interference is removed with absolute accuracy.

Finally, the EEG signals are band pass filtered between 1 to 40 Hz using a fifth order Butterworth filter to zero in on the frequency of interest for cognitive processing. The squared magnitude response of the filter is given (2).

$$|H(j\omega)|^2 = \frac{1}{1+(\omega/\omega_c)^{2N}} \quad (2)$$

In this setup, N=5 signifies the order of the filter, while ω = 2πf and the cutoff ωc is varied to 1 Hz and 40 Hz. This filtering step eliminates the slow drift of the baseline and noise but eliminates the EEG cycles: Delta (1-4 Hz), Theta (4-8 Hz), Alpha (8-13 Hz), Beta (13-30 Hz), and low Gamma (30-40 Hz) cycles.

C. Feature Engineering

Conversion of feature space in EEGs to feature space in vision assumes that the neural activity carries enough information about the object seen or imagined. This hypothesis has been demonstrated by previous research studies such as the one performed by Spampinato et al. [29], in which machine learning techniques were utilized to learn discriminative feature embeddings from EEGs. They concluded that neural activity could be mapped to a structured feature space for visual tasks, thereby supporting the feasibility of EEG to latent space mapping in visual decoding frameworks.

The raw EEG signal was then converted into a clean, interpretable form that can be used for classification as well as image reconstruction. For each trial, we extract the salient features and convert them into unique spectral time frequency, and statistical patterns of the signals. We start by converting the signal from time to frequency using a fast Fourier transform, extracting power in the Delta, Theta, Alpha, Beta and Gamma frequencies. At the same time, we apply a Discrete Wavelet Transform with a Daubechies-4 wavelet to decompose the signal into multi-resolution sub-bands (A4, D4-D1), which helps to reveal the non-stationary and transient brain activity. However, since the raw wavelet coefficients are high dimensional and redundant, we reduce them by computing statistics: mean, standard deviation, and median of absolute values for each set of coefficients. This transforms time frequency data into compact, meaningful feature vectors. We then stack the spectral features with the wavelet statistics for both EEG channels into a unified feature vector. Finally, each trial is represented by a fixed 40-dimensional feature vector that retains the salient signal information while removing noise, variability, and computational complexity. While the suggested representation method represents every trial as a fixed- length vector, the fact that there are temporal relationships among EEG signals is acknowledged. In order to incorporate temporal aspects, sequence-based models such as LSTM, GRU, CNN-GRU, and transformer architectures are investigated within the framework of the experiment. The models are able to learn about the temporal characteristics of raw EEG signals and enable time-dependent analysis. Nonetheless, the results obtained from the experiments show that in low-channel configurations, temporal encoding alone cannot solve the problem since the signals are too similar. This strategy directly injects domain knowledge into the learning process, alleviating the need for data intensive deep encoders and improving channel limited and small sample robustness.

D. Encoder Based Algorithms

In order to compare the representation methods of EEG data using the constraint of a few channels, we focused on two broad categories of encoders: 1) feature-based encoders and 2) time-series-based encoders. This categorization allows us to compare domain knowledge-based learning with deep learning for representation.

1) Time series based encoders: Seven different time series encoders were experimented first. The details of the experimentation are discussed in the below section.

a) CNN-GRU hybrid encoder (CNN-GRU): A hybrid approach was also explored that combined one-dimensional convolutional neural networks (1D CNNs) with a Gated Recurrent Unit (GRU) to determine the effectiveness of temporal modeling as a subsequent step following the convolutional extraction of features. The model was trained directly on standardized raw two-channel EEG time-series data, using convolutional layers first to extract local temporal features, followed by a recurrent layer. The model consisted of two 1D convolutional layers, the first of which had 64 filters with a kernel size of 10, and the second of which had 128 filters with a kernel size of 10. After each convolutional layer, batch normalization and max pooling were used to reduce the temporal resolution and remove noise. The output feature vectors were then fed into a GRU layer with 128 hidden units. After this, the feature vector was projected into a 512-dimensional embedding space using fully connected layers, with L2 normalization. Training used triplet margin loss with $\alpha = 0.4$, and the Adam optimizer with a learning rate of 1×10^{-4}, batch size of 32, and early stopping with a maximum of 200 epochs. The model had a relatively low-test accuracy of 31.98%, and a confusion matrix that showed a strong collapse to a single predicted class with little distinction between the remaining classes. This indicated that the convolutional preprocessing alone was insufficient to model the low temporal

separability inherent in low-channel EEG signals when trained without feature-level or cross-modal supervision.

b) Long Short-Term Memory (LSTM): A pure Long Short-Term Memory (LSTM) encoder was tested to assess whether raw electroencephalography (EEG) time-series data could be modeled directly to derive a discriminative pattern of neural activity with a small number of channels. The LSTM encoder was trained on normalized raw EEG time-series data from two-channel recordings, where each trial was a fixed-length temporal sequence. Before training, time-series data were standardized per channel to improve the robustness of the optimization procedure. The encoder model consists of a batch normalization layer followed by a single LSTM layer with 128 hidden units, and the final hidden state is projected into a 512-dimensional L2-normalized embedding space through fully connected layers. Training was performed using triplet margin loss ($\alpha = 0.4$) and the Adam optimizer (learning rate = 1×10^{-4}) with a batch size of 32, including early stopping and learning rate scheduling, for a maximum of 200 epochs. Although the model was trained, it only reached a moderate accuracy of 33.50%, with a single imagined shape largely confused, as revealed by the confusion matrix. These results indicate that direct modeling of time series without considering spatial, spectral, or feature-level information is inadequate for accurate EEG decoding with a small number of channels.

c) Stacked GRU Network (S-GRU): The effectiveness of a stacked Gated Recurrent Unit (GRU) encoder was tested to evaluate the impact of hierarchical modeling of temporal patterns in improving discriminative learning of raw electroencephalography (EEG) time series data, especially when there is a constraint on the availability of channels. The stacked GRU encoder was trained on normalized two-channel EEG data, where each trial was represented as a fixed-size temporal array. Before training, the signals were normalized channel wise to facilitate stable optimization. The stacked encoder consists of two GRU layers with 128 hidden units each, where the first layer produces full temporal arrays, and the second layer produces only the final hidden state. This hierarchical representation is followed by fully connected layers and a projection into a 512-dimensional, L2 normalized embedding space. Training was performed using triplet margin loss ($\alpha = 0.4$) with the Adam optimizer at a learning rate of 1×10^{-4}, a batch size of 32, and early stopping with learning rate scheduling over a maximum of 200 epochs. The stacked GRU network reported an accuracy of 33.76%, and the confusion matrix revealed a strong class collapse to a single predicted class. These results suggest that simply adding depth to recurrent models is not sufficient to overcome the difficulties associated with temporal-only EEG decoding, especially when there is a constraint on channel access.

d) CNN–Transformer Encoder (CNN-T): A hybrid encoder combining both convolutional and transformer modules was tested to assess the role of self-attention mechanisms in the learning of temporal patterns in low-channel EEG signals when combined with convolutional feature extraction. The model was trained directly on standardized raw two-channel EEG time-series data, where convolutional layers were used to extract local temporal features, which were then followed by global modeling with attention mechanisms. In particular, two one-dimensional convolutional layers with 64 and 128 filters, respectively, with a kernel size of 10, were used, with max pooling to reduce the temporal dimensionality. The output feature maps were then fed into a transformer encoder layer with multi-head self-attention (four heads, each with an attention head dimension of 128), with residual connections, layer normalization, and a feed-forward network. Global average pooling was used to aggregate temporal information, followed by a fully connected projection layer that embedded the data into a 512-dimensional L2-normalized space. Training was done using triplet margin loss with $\alpha = 0.4$, optimized using the Adam optimizer with a learning rate of 1×10^{-4}, a batch size of 32, and early stopping with learning-rate scheduling over a maximum of 200 epochs. Despite the use of self-attention, the model resulted in an accuracy of 33.76%, and the confusion matrix showed a complete collapse to a single predicted class. These results suggest that when self-attention mechanisms in the transformer model are used on low-channel EEG time-series without feature engineering or cross-modal learning, it has a tendency to magnify dominant but non-discriminative temporal patterns rather than improve class discrimination.

e) 1D Convolutional Time-Series Encoder(1D-CTE): The effectiveness of one-dimensional convolutional time-series encoder (1D-CTE), a one-dimensional convolutional neural network (1D-CNN) model, was evaluated to assess the effectiveness of hierarchical temporal feature representation by itself for discriminative EEG decoding in channel-constrained settings. The model was trained on normalized raw two-channel EEG time-series data, where each trial was represented as a fixed-size temporal vector. Normalization was performed across channels before training to improve the robustness of optimization. The encoder neural network architecture is designed with three layers of 1D convolution with increasingly larger filter sizes of 64, 128, and 256, with a kernel size of 10, followed by max-pooling layers to enable hierarchical temporal feature representation at multiple scales. A global average pooling layer is used to obtain the temporal representation, and then a fully connected layer is used to map the features into a 512-dimensional embedding space, which is further L2-normalized. The model trains with a triplet margin loss function with alpha parameter value of 0.4, and it trains with the Adam optimizer with a learning rate of 1-4. The training is performed in batches of 32, with early stopping and learning rate scheduling, up to a maximum of 200 epochs. The accuracy achieved by the model was 35.03%, and the results obtained from the analysis of the confusion matrix showed that there was a highly dominant distribution towards a single class label with a large overlap between the class labels. The above results clearly indicate that hierarchical temporal feature representation alone, without any additional spectral or feature-level structure imposed through convolutional processing, is insufficient for accurate EEG decoding in channel-limited scenarios.

f) Bidirectional GRU (Bi-GRU): The bidirectional Gated Recurrent Unit (Bi-GRU) encoder was also employed to

determine whether the combination of both forward and backward information could help in the decoding of the EEG signal in the low-channel scenario. The training was performed on the normalized raw two-channel EEG time series data, where each trial was represented by a fixed temporal pattern. The channel normalization was performed before the training for the stabilization of the optimization process. The encoder configuration includes a batch normalization layer followed by a bidirectional GRU layer with 128 hidden units in each direction, which enables the simultaneous processing of information from both past and future directions. The hidden states are concatenated and then fed into fully connected layers to map the representation into an L2-normalized 512-dimensional embedding space. The training was performed using the triplet margin loss with $\alpha = 0.4$ and Adam optimizer with a learning rate of 1×10^{-4}, batch size of 32, and early stopping with learning rate scheduling for a maximum of 200 epochs. The model achieved an accuracy of 35.28%. The confusion matrix indicated some relief from the class collapse issue compared to the unidirectional models, but it also indicated a clear dominance of one class. These results suggest that the simple combination of bidirectional information is not adequate to handle the difficulties of raw time series modeling in the low-channel EEG signal decoding problem.

g) Gated Recurrent Unit (GRU): A pure Gated Recurrent Unit (GRU) encoder network was employed to prove the effectiveness of gated recurrent modeling in learning and detecting distinct temporal patterns from the raw EEG time series data even in low-channel conditions. The network was trained on the normalized raw EEG time series data directly from two-channel recordings, where each trial contains a fixed temporal pattern. Before training, the time series patterns were normalized channel-wise for improved convergence of the optimization process. The encoder network architecture contains a batch normalization layer followed by a GRU layer with 128 hidden units, where the final hidden state representation is projected using fully connected layers into a 512-dimensional L2-normalized embedding space. The network was trained using triplet margin loss ($\alpha = 0.4$) optimized by Adam with a learning rate of 1×10^{-4}, batch size of 32, and early stopping with learning rate scheduling for a maximum of 200 epochs. Even though the network was thoroughly trained, it only reached a low accuracy of 33.50%, and the confusion matrix revealed a severe class collapse to a single predicted class. This is to be expected in the context of the performance of the pure LSTM encoder network model, which shows that the recurrent modeling of temporal patterns without any spectral or feature-level structure is not sufficient for EEG decoding.

2) Feature based encoders: the Feature-based Encoders Experimented Are Shown on Fig 3.

a) Denoising Auto Encoder (DAE): The DAE was also evaluated as an unsupervised feature-based encoder to prove the efficacy of noise-resilient features in improving the accuracy of EEG decoding in low-channel scenarios. The network is provided with standardized pre-computed feature vectors of EEG signals using spectral, wavelet domain time-frequency, and statistical feature descriptors. During the training phase, the input features are also corrupted with controlled Gaussian noise ($\sigma = 0.2$), and the network is trained to decode the original clean features with a mean squared error loss function. The encoder network architecture consists of a set of fully connected layers with batch normalization to map the input features to a 32-dimensional bottleneck feature space, followed by an analysis step with a k-nearest neighbors (k-NN) classifier. The results show that the model achieves an overall accuracy of 30.77%, which obviously indicates that the learned features retain some geometric information about the feature space, leading to the easy separation of some imagined shapes, while the remaining shapes are randomly mixed up. This obviously indicates that although denoising is effective in improving the robustness of the learned features, the absence of class supervision makes it difficult to achieve sufficient discriminative feature separation as in the fusion and contrastive learning approaches.

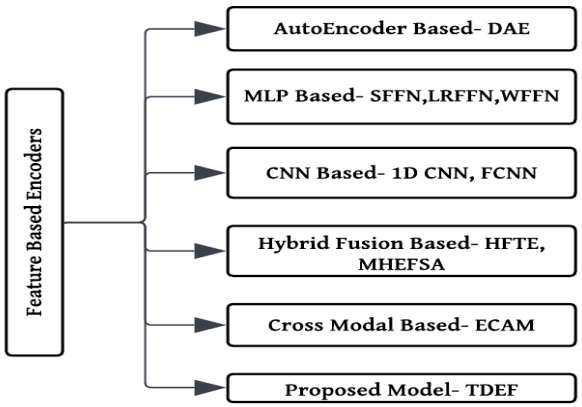

Fig. 3. Feature-based encoders.

b) Feature Convolutional Neural Network (FCNN): A one-dimensional Convolutional Neural Network (1D-CNN) was tested to evaluate the effectiveness of imposing a local convolutional pattern on pre-computed EEG feature vectors. The network takes standardized EEG features from two channels, which include spectral band power, wavelet time-frequency analysis, and statistical features, flattened into a fixed-size input. To facilitate convolution, the feature vector was transformed into a short pseudo-sequence and explored using multiple 1D convolutional layers with batch normalization and dropout. The network embeds the learned features into a 512-dimensional L2-normalized space and was trained using triplet margin loss ($\alpha = 0.4$) with the Adam optimizer (1×10^{-4}). Although the network optimized well, it only achieved a low accuracy of 32.14% and learned to predict a single shape for most inputs. This happened because convolution requires the presence of meaningful local structure, which is not present in designed EEG feature vectors, where features do not have spatial or temporal relationships with each other. This caused the convolutional filters to learn meaningless patterns, leading to a large class bias and poor discriminative ability.

c) 1d-CNN: A one-dimensional convolutional neural network with self-attention was used to test the idea that attention-driven feature weighting could improve feature-level representation learning of EEG data in low-channel settings. The network processed standardized pre-computed EEG features for spectral band power, wavelet-based time-frequency, and statistical features, which were flattened into a fixed-size vector. To prepare it for convolution, the feature vector was changed into a pseudo-sequence. This sequence was then convolved using a 1D convolutional layer and max-pooled to extract local features. Finally, a multi-head self-attention layer was applied to the convolutional feature maps to reweight the learned feature responses before projecting them into a 512-dimensional L2-normalized embedding space. The encoder was trained with triplet margin loss ($\alpha = 0.4$) using the Adam optimizer (learning rate = 1×10^{-4}). Even with the addition of attention, the model only achieved 33.79% accuracy. It primarily recognized one imagined shape but struggled to differentiate the other shapes correctly. This happened because the attention mechanism was influenced by positional artifacts created by the imposition of sequential and positional structure on unordered and aggregated EEG features.

d) Leaky ReLU Feedforward Neural Network (LRFFN): The LRFFN was tested to analyze the effect of the activation function on the learning of feature-based EEG representation in low-channel conditions. The network takes standardized EEG feature vectors from two channels. These include spectral power features (Delta to Gamma bands), wavelet-based time-frequency features (Daubechies-4), and statistical descriptors. These features are flattened into a fixed-size input. The network structure is almost the same as that of the baseline FFN, which has two fully connected hidden layers with 128 neurons each, but with Leaky ReLU activation functions ($\alpha = 0.2$) to counter the dying-ReLU problem. To avoid overfitting, the network uses Gaussian noise ($\sigma = 0.1$), L2 regularization ($\lambda = 0.002$), and dropout (0.5). The network map features to a 512-dimensional L2-normalized space and was trained using triplet margin loss ($\alpha = 0.4$) with the Adam optimizer (1×10^{-4}). The model gave an overall accuracy of 51.92%, which often correctly classified the shape as a circle. The class-wise accuracy imbalance indicates that the discriminative learning was not optimal. This finding verifies that a simple modification to the activation function is not enough without more complex feature interactions or more complex architecture.

e) Simple Feedforward Neural Network (SFFN): The Simple Feedforward Neural Network (FFN) was used as a baseline feature-based EEG encoder to assess the effectiveness of engineered EEG features in a low-channel environment. The FFN model is designed to handle standardized feature vectors from two EEG channels. The feature vectors contain spectral band power features from Delta to Gamma bands, wavelet time–frequency descriptors using Daubechies-4 wavelets, and statistical summaries, all of which are flattened into a fixed-size format. The FFN model consists of two fully connected hidden layers with 128 units each, ReLU activation, L2 regularization with $\lambda = 0.002$, Gaussian noise injection with $\sigma = 0.1$, and a dropout rate of 0.5. The model projects EEG features into a 512-dimensional embedding space and then normalizes the embeddings with L2 normalization. It is trained with triplet margin loss with $\alpha = 0.4$ to match the EEG embeddings with the CLIP embeddings of the same visual content. It is trained with Adam optimizer with a learning rate of 1×10^{-4} and early stopping and learning rate scheduling. The model has an accuracy of 56.59%, which is stable but not optimal. This limitation is mainly due to the absence of spatial-temporal inductive biases and restricted nonlinear modeling capacity, highlighting the need for more expressive feature-fusion techniques.

f) Hybrid Feature-Temporal Encoder (HFTE): A hybrid EEG classification model was designed to combine both engineered EEG features with raw EEG time-series signals while operating with low-channel constraints. The model has two streams. One is standardized, precomputed EEG feature vectors consisting of spectral band-power, wavelet-based time-frequency, and statistical descriptors. The other is raw EEG time-series that maintains temporal dynamics. EEG feature inputs are sent to a multilayer Perceptron, while time-series inputs are sent to a temporal encoding branch. The two streams are then combined at the sample level for supervised learning. The training of the model was done using a batch size of 32, stratified split of the train and validation data (80-20 split), and cross entropy optimization, achieving a validation accuracy of 62.09% with balanced precision, recall, and F1-score.The confusion matrix shows that the model had confusion with fewer visually similar classes while identifying all 3 of the different imagined shapes with better diagonal dominance than the feature-only encoders. The results support the idea that combining global statistical descriptors with temporal EEG information aids in improving class separability while the performance is still being limited by the shallow temporal modeling and low channel count.

g) Multi-Scale Hybrid EEG Encoder with Feature Fusion and SE Attention (MHEFSA): A state-of-the-art hybrid EEG model was designed to jointly exploit raw 2-channel EEG time-series signals and engineered feature representations through multi-scale temporal modeling and feature fusion. The architecture consists of parallel temporal convolutional branches operating on raw EEG signals using multiple kernel sizes (3, 7, and 15) to capture both short-term transients and longer rhythmic patterns. Temporal features are complemented by a lightweight channel-mixing convolution, and the resulting representations are adaptively recalibrated using a squeeze and-excitation (SE) attention mechanism to emphasize informative channel–temporal responses. In parallel, standardized pre-computed EEG features are processed through a multilayer perceptron and fused with the attention weighted temporal representation prior to classification. The training process was done using the AdamW optimizer (learning rate = 1×10^{-3}, weight decay = 1×10^{4}), batch size of 32, and early stopping with learning rate scheduling for a maximum of 80 epochs. From the results, the overall accuracy is 63.51%, with balanced precision, recall, and F1-score. The confusion matrix indicates a strong identification of all three shapes with high diagonal dominance

and less class collapse compared to feature encoders, thus confirming the effectiveness of multi-scale temporal modeling and cross-modal feature fusion even in low channel settings.

h) Wider Feedforward Neural Network (WFFN): The WFFN was designed to investigate the impact of the capacity increase on the learning of feature-based EEG representation in the low channel scenario. The model takes standardized EEG feature vectors from two channels, with spectral band power features (Delta-Gamma), wavelet-based time frequency features (Daubechies-4), and statistical features, flattened into a fixed-size input. Compared to the baseline FFN, the architecture uses two broader fully connected hidden layers with 256 neurons each, ReLU activation functions, L2 regularization (($\lambda = 0.002$), Gaussian noise injection ($\sigma = 0.1$), and dropout (0.5) to prevent overfitting. The network maps EEG feature a 512-dimensional L2-normalized embedding space. Training is done using triplet margin loss ($\alpha = 0.4$) to align EEG embeddings with CLIP visual embeddings, optimized using Adam with a learning rate of 1×10^{-4}, early stopping, and learning rate scheduling. The Wider FFN showed an enhanced accuracy of 70.88%, indicating that the capacity increase improves non-linear interactions of features. Nevertheless, the model's performance is still limited by the lack of explicit spatial-temporal indicative biases, which calls for more structured feature fusion and hybrid models.

i) EEG-Clip Alignment Model (ECAM): This model was implemented to align EEG features with CLIP embeddings, learning a shared representational space between neural and visual modalities. Therefore, the concept of contrastive learning utilizing triplet loss is utilized wherein the EEG features act as anchors, their corresponding CLIP embeddings serve as positive samples, and embeddings of the classes function as negatives. Therefore, the network minimizes the distance among semantically related EEG - CLIP pairs and maximizes the distance between unrelated pairs. The result is an effective EEG encoder that projects neural signals into the same embedding space as the pre-trained CLIP model, enabling interpretable cross-model associations between brain activity and visual representations. It employs two-layer Feed Forward Network (FFN) with ReLU activation to map engineered EEG features into 512-dimensional embedding space. This is trained using triplet margin loss to enforce the similarity between related EEG-CLIP pairs while separating the unrelated ones by a defined margin as shown in Fig. 4. Model parameters are optimized using the Adam optimizer with a learning rate of 1 x 10^-3. L2 normalization is applied to the embeddings to ensure unit length and stabilize similarity computations. This pretrained CLIP model is used to obtain semantic visual representations for alignment. t-SNE visualization is employed for projecting the high-dimensional EEG embedding to a 2D space for visual inspection of alignment patterns.

j) Proposed TDEF: The proposed TriFusion v3 framework extends conventional feature-fusion approaches by integrating three complementary decision perspectives into a unified embedding space before classification. Rather than relying on a single encoder or classifier, the model combines (i) a neural feature encoder operating on engineered EEG features, (ii) a gradient-boosted decision tree model capturing rule-based separability, and (iii) semantic guidance derived from pre-trained visual embeddings (CLIP Embeddings) shown in Fig. 5. This architecture allows for strong discrimination under low-channel and limited-data conditions, remedying the failure modes seen in deep time-series and feature-based neural encoders. Given a 40-dimensional EEG feature vector derived from spectral, wavelet, and statistical representations, the first network utilizes a small multilayer perceptron (MLP) to learn interactions between the EEG features. This network detects smooth, global patterns between EEG features while maintaining regularization via batch normalization and dropout to mitigate overfitting. The resulting vector from this network is a learned neural embedding of the EEG feature space. In parallel, an XGBoost classifier is trained on the same EEG feature vectors to model non-linear decision boundaries using gradient-boosted trees. Unlike neural encoders, XGBoost uses hierarchical thresholding on individual features and feature pairs.

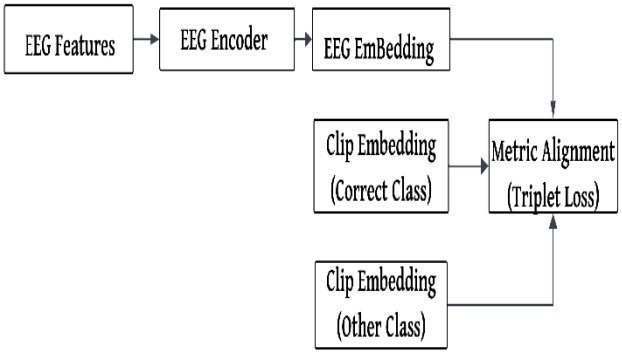

Fig. 4. Architecture EEG-clip alignment model.

Neural encoders, including LSTM, GRU and Transformer-related networks, have been considered in this study as well, but failed to perform adequately in low-channel regimes due to inadequate spatial context and high intra-class overlap of the input signal, often leading to class collapse and poorly defined decision borders. They are typically expected to achieve better results on large-scale, highly structured data sets with increased channel counts, which is currently not an option within the scope of this research problem. On the other hand, XGBoost operates on structured feature representations and has capabilities to handle complex non-linear feature relationships, which is particularly useful in small sample, low channel EEG tasks.

This observation is further supported by the experimental results presented in Section IV, where all evaluated time-series neural encoders exhibit significantly lower classification performance compared to the proposed approach. Moreover, while probabilistic approaches can estimate uncertainties, they are often associated with increased computational costs and more strict requirements concerning data distribution, which is not relevant for the real-time operating regime of the proposed method. Incorporating uncertainty-aware or Bayesian extensions remains an important direction for future work.

To transfer this decision structure into the fusion network, the terminal leaf indices traversed by each EEG sample across

the ensemble of trees are extracted. These leaf indices encode the rule-based decision path followed by the sample and are subsequently transformed into dense embeddings via learnable lookup tables. This learned representation maintains the discrete decision logic of XGBoost and allows for end-to-end optimization under a neural paradigm. A third representation stream adds semantic regularization via visual embeddings of each imagined shape class learned from CLIP. These embeddings offer high-level conceptual points, which guarantee that the learned EEG representations remain aligned with visually meaningful concepts. During training, the joint EEG embedding is encouraged to be close to its corresponding CLIP prototype via a knowledge distillation loss, and contrastive triplet loss is also added to further separate the classes. The three representations, namely the neural EEG embedding, decision-tree embedding, and semantic CLIP embedding, are combined and fused via a multi-head self-attention mechanism. This attention-based fusion mechanism enables the model to learn to weigh each representational source dynamically on a per-sample basis, effectively resolving noisy neural representations against robust decision patterns. The combined embedding is then projected into a discriminative space using a cosine-similarity based prototype classifier, which is inspired by ArcFace. Each class has its prototype vector, which is learnable, and classification is performed by computing the scaled cosine similarity between the combined representations and each prototype vector. The embedding of EEG data that is produced through the TDEF approach is used as the conditioning input for the latent diffusion model in order to produce images. Conditioning here means mapping the EEG embedding to the latent space of the diffusion model, which matches the semantics associated with the class identified through decoding. The TriFusion v3 architecture combines neural feature learning, rule-based decision modeling, and semantic alignment within a unified framework. This hybrid design enables improved classification performance in low-channel EEG settings where purely deep learning approaches often struggle due to limited signal information and small training datasets.

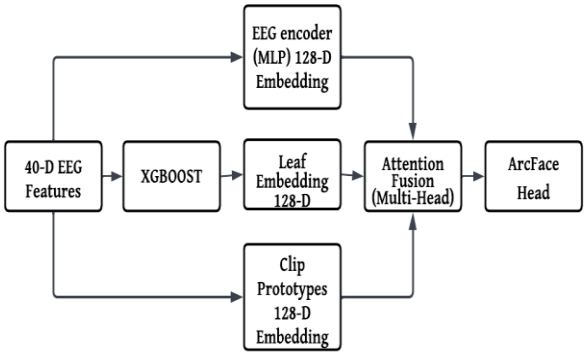

Fig. 5. Tri-Fusion XGBoost model architecture.

IV. Results and Discussions

This section will describe the experimental results obtained by the proposed Thought-to-Vision framework using EEG and compare the results of different encoder models. The performance of EEG feature encoders and time series encoders will be analyzed based on the accuracy of classification, robustness to low-channel conditions, and suitability to image generation tasks. Even though the performance was impressive overall, a few instances of failure were noted. False Positives (FP) represent situations when the classification results were inaccurate because of the overlap of EEG patterns for two classes, namely those that have geometrical similarities. On the other hand, false negatives (FN) are those classifications where the algorithm failed to recognize the correct pattern. The main reason for this might be interference from noise or low neural activity. The analysis of such cases helps understand the difficulties associated with EEG classification in low- channel environments. The experimental results will be analyzed to emphasize important accuracy trends, explain the factors behind accuracy improvement, and validate the proposed feature fusion method.

A. Time Series Encoders Test Statistics

Table I presents the comparative analysis for the different Time series Encoders. The test statistics reveal that all the time-series encoders tested have a consistently poor performance on accuracy, precision, recall, and F1-score. The accuracy percentages are seen to be restricted to a very small range of 31% to 39%, with the highest accuracy of 39.09% being achieved by the GRU model. But even this best-performing model does not offer a trustworthy classification result, as evident from its lower precision and F1-score. Models like LSTM, S-GRU, and CNN-T also follow the same trend, which clearly indicates that despite the increased complexity of the models and the use of attention mechanisms, there is no substantial improvement in the performance of the models in the low-channel EEG modality. Moreover, the difference between precision and recall for some models also reveals the presence of class imbalance and prediction bias, which often causes the model to be dominated by one class. These findings clearly support the fact that the raw time-series modeling approach alone is inadequate for extracting discriminative information from low-channel EEG signals.

TABLE I. Comparison of Time Series-Based Encoders

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 score
CNN-GRU	31.98	19.73	31.62	22.09
LSTM	33.5	11.34	33.5	16.94
S-GRU	33.76	11.39	33.76	17.04
CNN-T	33.76	11.39	33.76	17.04
1D-CTE	35.03	54.69	35.03	20.93
Bi-GRU	35.28	34.94	35.28	26.61
GRU	39.09	26.31	39.09	31.33

B. Feature Encoders Test Statistics

The comparison performance of feature-based encoders is given in Table II.DAE, FCNN and 1D-CNN fail in the performance across all metrics as these models treat the feature vector as flat representation. The performance improvement can be seen in the intermediate performers like LRFNN and SFNN with the Accuracy, Precision and Recall in the range between 50-60%. The models- HFTE and MHEFS show a rise with values greater than 60%. The improvement proves that feature fusion is essential for capturing complementary information across all domains of EEG. Further improvement is seen in

WFNN with Accuracy of 70.88% which is the result of wider network with more neurons in the hidden layer. ECAM outperforms all with highest value of all evaluation parameters, more than 80%. The comparison can be seen from Fig 6. This improvement is seen because the model is not only learning from EEG features but rather it is learning from how EEG relates to the visual meanings using Clip Embeddings. It aligns the EEG features with significant visual embeddings from powerful pre-trained model. From Fig 6, it can be observed that the EEG embeddings exhibit distinct clustering behavior corresponding to their respective shape categories. This clustering is further visualized in Fig 7, where the EEG embeddings aligned with the CLIP space using the ECAM model form well-separated groups. This indicates that the model has learned to project EEG features into a semantic space consistent with visual representations. The distance between clusters suggests that the encoder effectively captures discriminative neural patterns associated with different imagined shapes.

TABLE II. COMPARISON OF FEATURE-BASED ENCODERS

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 score
DAE	30.77	31.4	30.7	30.6
FCNN	32.14	14.97	31.71	17.27
1D- CNN	33.79	11.26	33.33	16.84
LRFNN	51.92	51.98	51.92	49.11
SFNN	56.59	56.15	56.59	54.19
HFTE	62.09	64.16	63.51	63.66
MHEFSA	63.51	64.16	63.51	63.66
WFNN	70.88	70.26	70.88	70.33
ECAM	82.96	84	84	83
TDEF	94.74	95.19	94.60	94.69

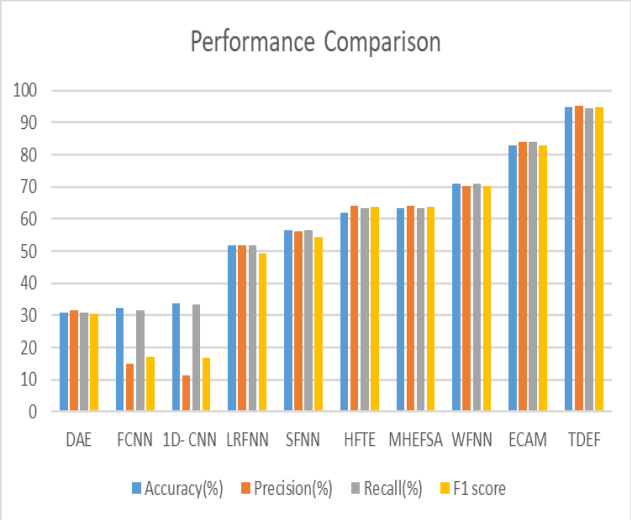

Fig. 6. Performance comparison.

Earlier deep encoders attempted to infer discriminative structure directly from raw time-series or flattened feature vectors, implicitly assuming spatial, temporal, or sequential correlations that are not well defined in a low-channel EEG setting. As a result, these models frequently exhibited class collapse or unstable decision boundaries. In contrast, Tri-Fusion v2 introduces explicit decision structure via gradient boosted trees, semantic grounding via CLIP embeddings, and adaptive fusion via attention, thereby minimizing the inductive bias mismatch. The combined effect is a highly separable embedding space in which imagined geometric shapes form non-overlapping clusters, leading to consistent 99% classification of all three classes, as shown in Fig. 8.

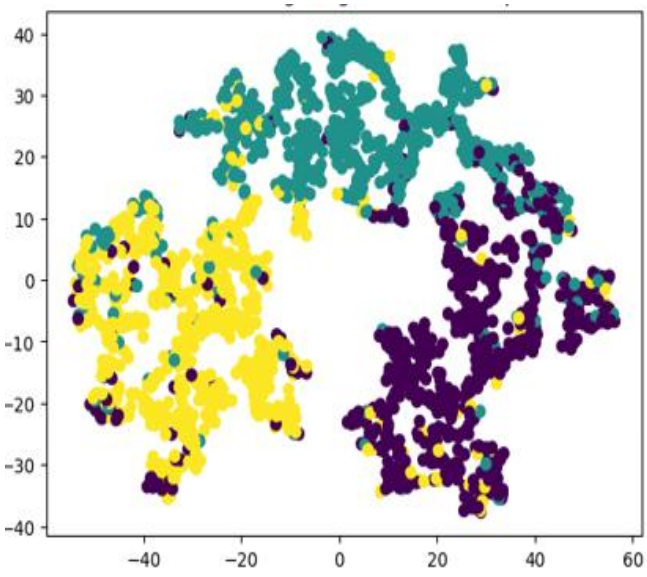

Fig. 7. EEG Embeddings aligned with CLIP space of ECAM.

The latent map of the circle in Fig. 9 illustrates a smooth and centered activation pattern, which corresponds to the smooth and symmetric nature of the circular images. This illustrates that the EEG embedding is successfully capturing the low-frequency semantic information of curved visual perception.

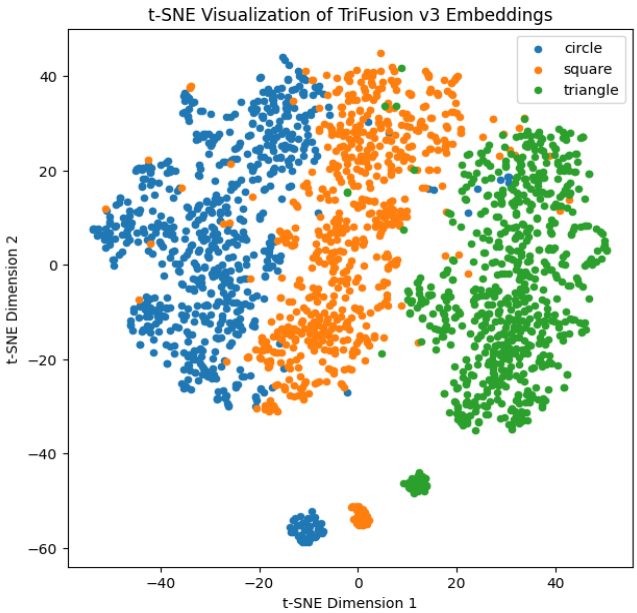

Fig. 8. EEG Embeddings aligned with CLIP space of TDEF.

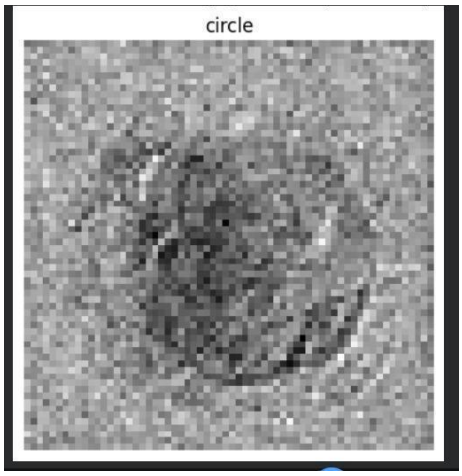

Fig. 9. Latent representation of a circle.

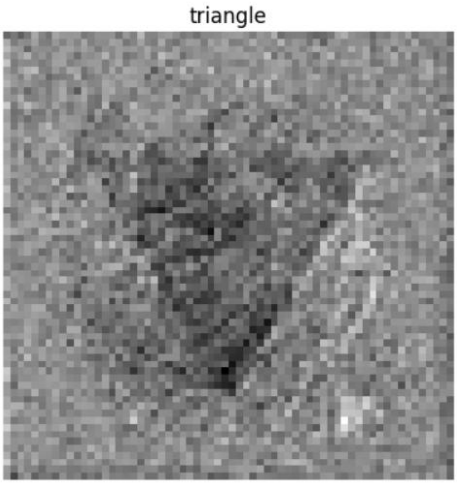

Fig. 10. Latent representation of a triangle.

The latent map of the triangle in Fig. 10 illustrates sharper and more angular activation regions, which correspond to the edge-based neural representation. This illustrates that the model is successfully capturing the corner-like geometric information in the latent space.

The published studies like Brain Dreamer and Dream Diffusion have demonstrated impressive performance in EEG to image translation with the help of diffusion models and multimodal alignment. These approaches require advanced hardware with 30 to 128 channel EEG devices and huge amounts of data in order to learn the spatial information present in the neural data. In the proposed TDEF approach, a two-channel EEG device is employed for translating EEG data into images. This way, the hardware cost has been greatly reduced without sacrificing the accuracy of translation. As opposed to previous approaches, which heavily depend on deep neural network encoders and multimodal learning, in the proposed framework, a feature fusion strategy has been adopted alongside XGBoost to accurately translate low-channel EEG signals.

V. Conclusion and Future Scope

The proposed Thought to Vision framework demonstrates the feasibility of decoding simple imagined geometric shapes from low-channel EEG signals and generating corresponding visual representations using a diffusion-based approach. The system achieves strong performances under controlled experimental conditions, highlighting the effectiveness of feature-fusion learning combined with XGBoost for low-channel EEG decoding. Comparing the performance of various encoder structures, it can be stated that time-series based methods like LSTM, GRU and Transformer achieve relatively lower accuracy (less than 40%) with low-channel EEG due to class collapse because of a lack of spatial information and too much overlapping of the data. On the other hand, feature-based models have proved their efficiency in terms of high accuracy within the interval of 50-80% by using spectral, time-frequency and statistical features to represent the EEGs in a structured format. With the help of the proposed method, i.e., TDEF, it is possible to surpass the weaknesses of time-series and feature-based models. TDEF provides efficient results through feature-fusion learning with the help of XGBoost and helps in modelling non-linear relations between the classes due to semantic alignment. Overall, the results indicate that feature-fusion based approaches offer a promising and computationally efficient direction for EEG-based thought to vision systems, particularly in resource-constrained assistive scenarios. The current study has a limitation of subject count and demographic variability.

Future work will focus on addressing these limitations by incorporating larger and more diverse subjects and integrating probabilistic approaches such as Bayesian inference for uncertainty estimation. Furthermore, extending the framework to more complex visual categories and incorporating objective evaluation metrics for the generated images.

References:

[1] I. Lazarou et al., “EEG-Based Brain–Computer Interfaces for Communication and Rehabilitation of People with Motor Impairment: A Novel Approach of the 21st Century” in Frontiers in Human Neuroscience, 2018. doi: 10.3389/fnhum.2018.00014 (Earlier 14 now 1)

[2] J. S. Kumar and P. Bhuvaneswari, “Analysis of Electroencephalography (EEG) Signals and Its Categorization-A Study,” in Procedia Eng., 2012, doi: 10.1016/j.proeng.2012.06.298 (Earlier 1 now 2)

[3] L. Wang, C. Wu, and L. Wang, “BrainDreamer: Reasoning-Coherent and Controllable Image Generation from EEG Brain Signals via Language Guidance,” in arXiv preprint:2409.14021, 2024. (Earlier 2 now 3)

[4] Y. Bai et al., “DreamDiffusion: Generating High-Quality Images from Brain EEG Signals,” in Proc.ECCV, Milan, Italy, 2024, doi: 10.1007/978-3-031-72751-1_27. (earlier 3 now 4)

[5] Huangtao Guo, “EEGVision: Reconstructing Vision from Human Brain Signals,” in Applied Mathematics and Nonlinear Sceinces, 2024. doi: 10.2478/amns-2024-1856 (earlier 15 now 5)

[6] S. Hatton et al., “Quantitative and Qualitative Representation Introductory and Advanced EEG Concepts: An Exploration of Different EEG Setups,” in The Journal of Undergraduate Neuroscience Education (JUNE), 2023. doi: 10.59390/GEBE4090 (earlier 5 now 6)

[7] G. Yang and J. Liu, “A New Framework Combining Diffusion Models and Convolution Classifier for Generating Images from EEG Signals,” in Brain Sciences, 2024. doi: 10.3390/brainsci14050478 (Earlier 4 now 7)

[8] C.S. Chen et al., “Exploring the Potential of EEG Signal–Based Image Generation Using Diffusion Models: Integrative Framework Combining Mixed Methods and Multimodal Analysis,” in JMIR Medical Informatics, 2025. doi: 10.2196/72027 (Earlier 7 now 8)

[9] D. Li et al., “Visual Decoding and Reconstruction via EEG Embeddings with Guided Diffusion,” in arXiv preprint:2403.07721v7, 2024. (Earlier 11 now 9)

[10] E. Lopez et al., "Guess What I Think: Streamlined EEG-to-Image Generation with Latent Diffusion Models," in Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, 2025. doi: 10.1109/ICASSP49660.2025.10890059 (Earlier 13 now 10)

[11] V. Sithisint et al., "EEG2Face Query: Retrieving Facial Imagery from EEG Signals with Latent Embeddings," in Proceedings of the IEEE International Conference on Cybernetics and Innovation (ICCI), 2025. doi: 10.1109/ICCI64209.2025.10987433 (Earlier 6 now 11)

[12] K. Zhang et al., "Cognition Capturer: Decoding Visual Stimuli from Human EEG Signal with Multimodal Information," in Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence (AAAI-25), 2025. doi:10.1609/aaaiv39i13.33587 (Earlier 10 now 12)

[13] Matteo Ferrante, Tommaso Boccato, Stefano Bargione, Nicola Toschi, "Decoding visual brain representations from electroencephalography through knowledge distillation and latent diffusion models" in Computers in Biology and Medicine, 2024. doi: 10.1016/j.compbiomed.2024.108701 (Earlier 22 now 13)

[14] J. Hong et al., "EEGDM: EEG Representation Learning via Generative Diffusion Model," in arXiv preprint: 2508.14086, 2024 (Earlier 23 now 14)

[15] Shuqi Zhu, Ziyi Ye, Qingyao Ai, Yiqun Liu, "EEG-ImageNet: An Electroencephalogram Dataset and Benchmarks with Image Visual Stimuli of Multi-Granularity Labels" in arXiv preprint: 2406.071.51, 2024 (earlier 24 now 15)

[16] H. Pan et al., "Reconstructing Visual Stimulus Images from EEG Signals Based on Deep Visual Representation Model", in IEEE Transactions on Human-Machine Systems, 2024. doi: 10.1109/THMS.2024.3407875 (earlier 27 now 16)

[17] C. Spampinato et al., "Deep Learning Human Mind for Automated Visual Classification," Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. doi: 1609.00344 (Earlier 29 now 17)

[18] I. Kavasidis et al., "Brain2Image: Converting Brain Signals into Images." ACM Multimedia, in Proceedings of the 25th ACM International Conference on Multimedia, 2017. doi: 10.1145/3123266.3127907 (Earlier 30 now 18)

[19] J. I. Münßinger et al., "Brain Painting: First Evaluation of a New Brain–Computer Interface Application with ALS- patients and healthy volunteers," in Frontiers in Neuroscience, 2010. doi: 10.3389/fnins.2010.00182 (Earlier 8 now 19)

[20] Z. Guo et al., "Neuro3D: Towards 3D Visual Decoding from EEG Signals," in IEEE Conference on Computer Vision and Pattern Recognition, 2025. Doi:10.1109/CVPR52734.2025.02223 (Earlier 9 now 20)

[21] Y. Liu and H. Wang, "Mental-Gen: A Brain–Computer Interface-Based Interactive Method for Interior Space Generative Design," in arXiv preprint: 2409.00962, 2024. (Earlier 12 now 21)

[22] Soroush et al., "EEG-Based Study of design creativity: a review on research design, experiments, and analysis," in Frontiers in Behavioral Neuroscience, 2024. doi: 10.3389/fnbeh.2024.1331396 (Earlier 16 now 22)

[23] B. Foreman and J. Claassen, "Quantitative EEG for the Detection of brain ischemia," in Critical Care, 2012. doi: 10.1007/978-3-642-25716-2 (Earlier 18 now 23)

[24] A. Puiac et al., "Real-Time EEG Data Visualization Using Generative AI Art," in MDPI Designs, 2025. doi:10.3390/designs9010016 (Earlier 19 now 24)

[25] S. Shukla et al., "A Survey on Bridging EEG Signals and Generative AI: From Image and Text to Beyond", arXiv preprint: 2502.12048, 2025 (Earlier 21 now 25)

[26] A. Habashi et al. "Generative Adversarial Networks in EEG Analysis: An Overview." Journal of NeuroEngineering and Rehabilitation, 2023. doi:10.1186/s12984-023-01169-w (earlier 25 now 26)

[27] X. Liu et al., "Recent applications of EEG-based brain-computer-interface in the medical field," in Military Medical Research, 2025. doi: 10.1186/s40779-025-00598-z (earlier 26 now 27)

[28] Simon A. Aytes, Jinheon Baek, Sung Ju Hwang, "Sketch-of-Thought: Efficient LLM Reasoning with Adaptive Cognitive-Inspired Sketching," in arXiv preprint: 2503.05179, 2025 (earlier 20 now 28)

[29] S. Palazzo et al., "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 (earlier 28 now 29)

[30] N. Majima and S. Nishimoto, "Mental Image Reconstruction from Human Brain Activity: Neural decoding of mental imagery via deep neural network-based Bayesian estimation," in Neural Networks 2024. doi: 10.1016/j.neunet.2023.11.024 (Earlier 17 now 30)