BrainDiffNet: Unified Semantic Encoders for Diffusion-based EEG-to-Image Generation

Shreyas Bellary Manjunath, Sreyasee Das Bhattacharjee
Department of Computer Science & Engineering
State University of New York at Buffalo, USA
sbellary@buffalo.edu, sreyasee@buffalo.edu

***Abstract*—Gaining insight into the brain's visual representation through reconstructing what we see from brain activity is of immense importance and interest. Though fMRI and MEG achieve high-quality image reconstruction and classification, their cost and size restrict broader real-world applications, particularly outside clinical settings. In contrast, although Electroencephalography (EEG) is a cost-effective, non-invasive tool producing high temporal resolution signals, it remains less explored primarily due to its susceptibility to noise and complex spatio-temporal characteristics. To address these, we propose *BrainDiffNet*, an effective EEG-to-Image generation model that leverages a subject's contextual and EEG spatio-temporal information to guide a fine-tuned Stable Diffusion model, resulting in high-quality, semantically relevant images from brain activity. A robust Temporal Masked Autoencoder, designed for high-resolution EEG, enables the model to effectively extract features and manage noisy or incomplete EEG query representations. In-depth evaluation using the large-scale EEG-ImageNet dataset demonstrates the outperformance of *BrainDiffNet* in both tasks: Object Classification and Image Reconstruction. In fact, the model significantly outperforms state-of-the-art baseline methods, achieving a $15-20\%$ higher accuracy in classification across all granularity levels and a $7-12\%$ improvement in all feature-specific two-way identification metrics for image reconstruction.**

***Index Terms*—Diffusion, Masked Auto-encoders, EEG decoder, Image reconstruction**

I. Introduction

The inherently multidimensional structure of Electroencephalography (EEG) data provides a distinct means for understanding diverse neurological phenomena [1]. In fact, deciphering human brain activity through EEG signals has consistently remained a central focus within neuroscience, for its substantial potential in understanding cognitive processes, mental states, and various spatio-temporal aspects of brain function [2]. The advent of deep neural networks, especially diffusion-based and transformer-based models, has even made it possible to reconstruct human brain activities directly from fMRI or MEG recordings [3]–[5]. However, the necessity of costly, bulky machinery and specialized clinical expertise for data acquisition severely restrict these techniques from broader applications. Comparably EEG signals can be obtained non-invasively by placing electrodes on the head, making it a gentle way to monitor the brain activity. Further, exceptional temporal resolutions make EEG ideal for studying rapid cognitive processes like perception, attention, and event-related potentials (ERPs). Despite these, only limited research [6], [7], has addressed EEG-to-image reconstruction, and the fundamental challenges are two-fold: (1) Analyzing EEG's multidimensional data requires a comprehensive look at both its spatial and temporal aspects, unified within the subject's specific context during signal generation; (2) Noise sensitivity and missing segments of EEG signals, stemming from interference like muscle movements or eye blinks (noise sensitivity), and sometimes parts of the data being incomplete or lost (missing segments).

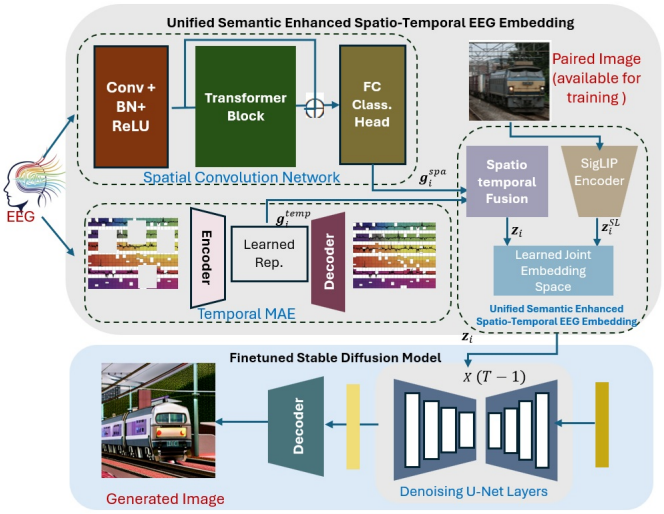

Fig. 1. Method Overview of *BrainDiffNet*

To address these, we propose *BrainDiffNet*, an effective EEG-to-Image generation model that utilizes a subject's contextual information alongside the spatial and temporal patterns within their EEG signals. This comprehensive understanding then guides the fine-tuning of a Stable Diffusion model, enabling the generation of semantically relevant images directly from brain activity. Its primary contributions include:

1) *A Unified Semantic Enhanced Spatio-Temporal EEG Embedding* that leverages both spatial and temporal EEG signal patterns, augmented by the subject's specific context to design a comprehensive and compact EEG embedding.
2) *An effective Temporal Masked Auto-encoder (MAE) based asymmetric feature extraction module*, specifically designed for high-resolution EEG signals. This module is expertly tailored to handle incomplete mode representations in query times by accurately approximating missing data points, ensuring robust feature extraction

even with imperfect EEG signals.

3) *Qualitative and Quantitative evaluation* demonstrating outperformance of the proposed *BrainDiffNet* against state-of-the-art models in two tasks using only a subject's EEG signals: Object Classification and Image Reconstruction.

II. Proposed Method

Given a multichannel EEG signal (comprising of M channel-specific 1D time sequences) x_i obtained from a subject S_i, our objective aims to reconstruct the image stimulus I the subject is exposed to. Each image I corresponds to a category c from a predefined set of object classes $\mathcal{C}$. In other words, we have $x_i = \{\mathbf{e}_i^1, \mathbf{e}_i^2, \ldots, \mathbf{e}_i^M\}$, where $\mathbf{e}_i^m \in \mathbb{R}^{N\times 1}$ and N is the signal length and $m \in \{1, \ldots, M\}$. Based on this input, *BrainDiffNet* aims to generate an image $\mathbf{I}^* \in \mathbb{R}^{n_1\times n_2}$ that closely approximates $\mathbf{I} \in \mathbb{R}^{n_1\times n_2}$. Figure 1 gives an overview of the proposed multisensor fusion network *BrainDiffNet*, comprised of two core task components: *Unified Semantic Enhanced Spatio-Temporal EEG Embedding* and *Finetuning pretrained Stable Diffusion Model with EEG encoder-image pairs*.

A. Unified Semantic Enhanced Spatio-Temporal EEG Embedding

EEG, which measures brain activity non-invasively, faces challenges in precisely localizing activity sources due to varying spatiotemporal resolutions that are often determined by a subject's demographics and cognitive state specifics. In this section, we describe a unified spatiotemporal embedding module that leverages two different types of encoders (*Spatial Convolution Network* [8] and *Temporal Masked auto-encoder* [6]) to form a *fused spatio-temporal EEG representation*.

1) Preprocessing: To ensure high-quality EEG data, we implement a preprocessing pipeline addressing noise and artifacts. This involves applying a 0.5-80 Hz band-pass filter to remove drifts, high-frequency noise, and 50 Hz line noise; and systematically eliminating artifacts like eye blinks, muscle artifacts, and head movement artifacts. Now on, unless specifically mentioned otherwise, each reference to $\mathbf{e}_i^m$ the paper will refer to a preprocessed EEG signal generated from channel m.

2) Spatial Convolution Network: Drawing inspiration from Zhao et al. [9], the model uses a convolutional transformer model that is particularly trained for EEG classification task and is comprised of three components applied in sequence: a Convolution block; a Transformer Encoder; and a Fully Connected Classification head.

The Convolution block has a sequence of three convolution layers and serves as the initial feature extractor, focusing on capturing local and spatial features from the raw EEG time series. It transforms the initial raw multichannel EEG signal into a more compact, higher-level representation. Specifically, it produces l_0-dimensional features ($\mathbf{f}_i^m \in \mathbb{R}^{l_0}, m \in \{1, \ldots, M\}$) for every approximately 1-second channel-specific segment of the original EEG. All convolutional layers are followed by batch normalization (BN), which makes the training smoother and mitigates the risk of overfitting.

These extracted features from the Convolution block are then sent to a Transformer Encoder to capture broader, global relationships. Multi-Head Attention (MHA) at the core of this Transformer Encoder allows the model to weigh the importance of different parts of the input feature sequence (different time points or channels) relative to each other. In our experiments, we use 3-layer Transformer Encoder, wherein each layer has three main parts: MHA Layer blocks; Residual Connections, and a Feed-Forward network. More precisely, following the MHA self-attention operation, the output features are additively combined with the original input features via a residual connection. This combined output then undergoes layer normalization to standardize its scale. The normalized result is subsequently passed through a position-wise fully connected feed-forward network, which applies the same transformation independently to each position. The entire Transformer Encoder processes the multichannel EEG representation $\{\mathbf{f}_i^1, \mathbf{f}_i^2, \ldots \mathbf{f}_i^M\}$ as input and produces a compact EEG descriptor $\mathbf{g}_i^{spa} \in \mathbb{R}^{l\times d}$.

The flattened output of the Transformer Encoder is used as the input to the Fully Connected Classification head comprising of a fully connected layer followed by the Softmax layer with $|\mathcal{C}|$ units, which is employed to map $\mathbf{g}_i^{spa}$ into a probabilistic classification decision. The multi-class cross entropy loss function that evaluates the model's prediction output $P(x_i|\theta)$ (where θ represents the classifier parameter) with ground truth label c_i as $\mathcal{L}_{CE} = CrossEntropyLoss\Big(P(x_i|\theta), c_i\Big)$, is used to train the model.

3) Temporal Masked auto-encoder: Self-supervised pre-training with masked modeling has proven highly effective in NLP [10] and visual representation learning [11]. This approach trains models to predict randomly masked input tokens based on the remaining available tokens. To process our data, we first divide each $\mathbf{e}_m^i$ into time-domain tokens of size p, i.e. each token represents p (in all experiments, we use $p = 3$) consecutive temporal samples and randomly hide a portion of them through masking. Next, these tokens are converted into embeddings using a 1D convolution layer. A Masked Auto-encoder (MAE) then leverages the temporal context from the unmasked tokens to predict the masked ones, allowing the MAE to build a robust understanding of the data's inherent temporal structure.

In particular, the MAE encoder architecture is adopted from EEGViT [12], a hybrid Vision Transformer (ViT) [13] style architecture designed for EEG data. It takes x_i (represented using M channel-specific signals $\{\mathbf{e}_i^1, \mathbf{e}_i^2, \ldots, \mathbf{e}_i^M\}$) as input and features a two-step Convolutional block [14] integrated with ViT layers followed by a projection layer to generate an encoder output $\mathbf{g}_i^{temp} \in \mathbb{R}^{l\times d}$. The ViT layers are initialized with weights pre-trained on the ImageNet dataset to facilitate downstream fine-tuning, while the projection layer learns its weights during this fine-tuning process. The MAE decoder consists of a series of Transformer blocks, which extract low-level semantic information regarding the underlying raw signal patterns. A reconstruction loss $\mathcal{L}_{rec}$ that evaluates the

alignment of the original $\mathbf{e}_i^m$ with the reconstructed signal $\mathbf{e}_i^{m,*}$, is used to train this module as:

$$\mathcal{L}_{rec} = 1 - \frac{1}{M}\sum_{m=1}^{M}\frac{\mathbf{e}_i^m \cdot \mathbf{e}_i^{m,*}}{||\mathbf{e}_i^m||||\mathbf{e}_i^{m,*}||} \quad (1)$$

4) Feature Fusion for Unified Semantic Enhanced Spatio-Temporal EEG Embedding : The proposed spatio-temporal representation of EEG signals specifically addresses the individual weaknesses of each pre-trained encoder (described above) by blending them via a cross-attention mixer with learnable queries. In particular, we use a randomly initialized query $\mathbf{Q} \in \mathbb{R}^{1\times d}$, keys as $\mathbf{K}_i = [\overline{\mathbf{g}}_i^{spa}\ \overline{\mathbf{g}}_i^{temp}] \in \mathbb{R}^{2\times d}$ where $\overline{\mathbf{g}}_i^{spa}, \overline{\mathbf{g}}_i^{temp} \in \mathcal{R}^d$ is each encoder's features averaged over the sequence dimension l for efficient computation, and values as $\mathbf{V}_i = [\mathbf{g}_i^{spa}\ \mathbf{g}_i^{temp}] \in \mathbb{R}^{2\times l\times d}$. The resulting unified EEG embedding $\mathbf{z}_i$ is then defined as $\mathbf{z}_i := CA(\mathbf{Q}, \mathbf{K}_i, \mathbf{V}_i) \in \mathcal{R}^{l\times d}$, where $CA(\mathbf{Q}, \mathbf{K}_i, \mathbf{V}_i) = Softmax\left(\frac{\mathbf{Q}\mathbf{K}_i^T}{\sqrt{d}}\right)\mathbf{V}_i$.

To incorporate a better understanding of the EEG encoder-image associations with underlying multimodal semantics, the model is trained using a sigmoid contrastive learning [15] on the $(\mathbf{z}_i, \mathbf{z}_i^{SL})$ pair, where $\mathbf{z}_i^{SL}$ represents the SigLIP encoder of image I. The resulting unified Semantic Enhanced Spatio-Temporal EEG Embedding is produced as a system output. Unless specifically mentioned otherwise, each reference to $\mathbf{z}_i$ in the next part of the paper will refer to the *Unified Semantic Enhanced Spatio-Temporal EEG Embedding* of the input x_i.

B. Finetuning pretrained Stable Diffusion Model with EEG encoder-image pairs

Given the unified embedding $\mathbf{z}_i$ of the input EEG x_i, we leverage a pretrained SD model [16] that utilizes a U-Net architecture [17] to approximate the visual stimulus I that the subject S_i was exposed to. More specifically, the U-Net starts with a random, noisy image and, guided by the fused EEG representative ($\mathbf{z}_i$) fed into its intermediate layers, gradually denoises the input. The process, further enhanced with a cross-attention mechanism, enables flexible and precise image generation, transforming the noise into an accurate visual representation $\mathbf{I}^*$ of the original stimulus $\mathbf{I}$. In particular, given $\mathbf{I}$ encoded by a VQ [18] encoder $\mathbf{z}_i^{img} = \mathcal{E}(\mathbf{I})$, the unified EEG embedding $\mathbf{z}_i$ is fed into each intermediate j^{th} finetunable layer of the U-Net to produce a compact cross-attentive layer output $\mathbf{u}_{i,t}^{(j)} := \mathbf{C}A(\mathbf{Q}^{diff}, \mathbf{K}_i^{diff}, \mathbf{V}_i^{diff})$. The terms $\mathbf{Q}^{diff} := W_Q^{(j)} \cdot \phi^{(j)}(\mathbf{z}_{i,t}^{img})$, $\mathbf{K}_i^{diff} := W_K^{(j)} \cdot \mathbf{z}_i$, and $\mathbf{V} := W_V^{(j)} \cdot \mathbf{z}_i$ form the learnable query, key and value respectively. The function $\phi^{(j)}(.)$ is the transformation function of the j^{th} layer within the U-Net. The parameter matrices $W_Q^{(j)} \in \mathbb{R}^{N_Q\times d^j}$, $W_K^{(j)} \in \mathbb{R}^{d_0\times d}$, and $W_V^{(j)} \in \mathbb{R}^{d_0\times d}$ represent the learnable weights. The stable diffusion loss function defined below is used for fine-tuning.

$$\mathcal{L}_{SD} = \mathbb{E}_{\mathbf{I},\epsilon\sim\mathcal{N}(0,1),t}\left[||\epsilon - \epsilon_{den}(\mathbf{I}_t, t, \mathbf{z}_i)||_2^2\right] \quad (2)$$

While the proposed image reconstruction module adopts a similar stable diffusion model architecture to [6], the core innovation lies in the process of conditioning the stable diffusion during fine-tuning. Instead of aligning EEG signals with their text-image semantics during fine-tuning, the proposed unified embedding module first develops a compact multimodal-semantic enhanced fused EEG encoder, which then conditions the stable diffusion model during fine-tuning, significantly simplifying the reconstruction process.

TABLE I
THE AVERAGE PERFORMANCE OF ALL PARTICIPANTS IN THE OBJECT CLASSIFICATION TASK AT DIFFERENT GRANULARITY LEVELS.

Method	Acc(All)	Acc(coarse)	Acc(fine)
Ridge Regression [7]	0.2859	0.3944	0.5833
Random Forest [7]	0.3489	0.4535	0.7288
Support Vector Machine [7]	0.3919	0.5057	0.7784
Multi-Layer Perceptron [19]	0.4037	0.5339	0.8163
EEGNet [14]	0.2604	0.3030	0.3645
RGNN [20]	0.4050	0.4703	0.7057
BrainDiffNet	**0.5522**	**0.6713**	**0.8913**

III. EXPERIMENTS

The proposed *BrainDiffNet* is evaluated using a recent and large-scale EEG-ImageNet dataset [7] featuring EEG recordings from 16 participants viewing $4,000$ images from ImageNet object categories. Following the same evaluation protocol followed by Zhu et al. [7], including its train-test data distributions, we investigate the performance of *BrainDiffNet* for two tasks: Object Classification and Image Reconstruction. In classification, the goal is to identify image stimuli from a subject's corresponding EEG, while in image reconstruction, the objective is to generate images that resemble the original visual stimuli based on the subject's concurrent EEG signal.

A. Results

1) Object Classification: Following the evaluation protocol proposed by the dataset authors in [7], experiments are conducted at three coarse-to-fine granularity levels: all, coarse (e.g., "animal," "vehicle"), and fine (e.g., "dog" within "animal," or "car" within "vehicle"). The "all" category represents the 80-class classification accuracy, the "coarse" category represents the 40-class classification accuracy, and the "fine" category represents the average performance of five 8-class classification tasks. The average accuracy scores of all participants in the object classification task are reported in Table 1. As observed in the table, *BrainDiffNet* outperforms the state-of-the-art methods, showing remarkable improvements of $15-20\%$ in accuracy across all granularity levels.

2) Image Reconstruction: As described by Zhu et al. [7], for the image reconstruction task, we adopt Alex(2), Alex(5), Inception Score, and CLIP ViT-L/14 as the comparison features to calculate two-way identification for performance evaluation. In particular, two-way identification is a method where a generated image is presented alongside its original and one distractor, with the task being to select the original correctly. This evaluates the reconstruction pipeline's ability to produce distinguishable and recognizable images. The second and fifth convolutional layers of AlexNet, the output before the linear layer of Inception, and the embeddings from CLIP ViT-L/14 are used as image features to compute Alex(2), Alex(5), Inception Score, and CLIP ViT-L/14 two-way identification scores, respectively. Table 2 reports the results that compare

TABLE II
THE AVERAGE RESULTS OF ALL PARTICIPANTS IN THE IMAGE RECONSTRUCTION TASK. THE TWO-WAY IDENTIFICATION METRIC [7] IS USED FOR PERFORMANCE EVALUATION.

Method	Alex(2)	Alex(5)	Inception Score	CLIP(ViT-L/14)
Baseline Rec. [7]	0.5605	0.6299	0.5675	0.6467
BrainDiffNet	**0.6323**	**0.7489**	**0.6608**	**0.7683**

Fig. 2. Some image reconstruction results. In each column, the top row shows the image visual stimuli, the middle row shows the generated image using the baseline reconstruction model [7], and the bottom row shows the reconstructed images by the proposed *BrainDiffNet*. Two example bad cases are shown in columns (6) and (7). As observed, while *BrainDiffNet* generated images are comparably better than the ones generated by the baseline model, some semantic details (e.g., color of the grapes, orientation of the ship) of the ground truths are still not preserved in the generated images.

the performance of the proposed *BrainDiffNet* against a baseline image reconstruction model [7] that uses a frozen Stable Diffusion 1.4 backbone, conditioned by prompt embeddings generated by a trained two-layer MLP encoder. Reconstructions are performed with 50 PNDM denoising timesteps, yielding 512x512 images, and trained using MSE loss. As reported in the table, *BrainDiffNet* outperforms the baseline by demonstrating improvements of $7-12\%$ across all feature-specific two-way identification metrics. Figure 2 shows some qualitative results. As seen, our results exhibit markedly higher visual quality and recognizability. The reconstructed images not only preserve finer details (e.g., cup shape) but also present a more coherent and semantically aligned representation (e.g., pool table angle of view with its surrounding environment patterns) of the original stimuli, making them significantly more distinguishable.

IV. CONCLUSION

We introduce *BrainDiffNet*, a powerful framework for generating high-fidelity images from EEG signals that employs a unified, semantic-enhanced spatio-temporal embedding to guide a fine-tuned stable diffusion model. This approach excels in handling noisy and incomplete data through semantic enhanced fused spatio-temporal features, achieving state-of-the-art results in both image reconstruction and object classification. Future research will focus on personalized calibration, using EEG signals to rapidly adapt the model to unique neural wave patterns for greater subject-specific precision.

V. ACKNOWLEDGEMENT

The project was partially funded by the National Science Foundation, Award ID: 2347251

REFERENCES

[1] Nathan Koome Murungi, Michael Vinh Pham, Xufeng Caesar Dai, and Xiaodong Qu, "Empowering computer science students in electroencephalography (eeg) analysis: A review of machine learning algorithms for eeg datasets," 2023.

[2] Qian Luo and Kalyani Vinayagam Sivasundari, "Whisper+ aasist for deepfake audio detection," in *International Conference on Human-Computer Interaction*. Springer, 2024, pp. 121–133.

[3] Yohann Benchetrit, Hubert Banville, and Jean-Rémi King, "Brain decoding: toward real-time reconstruction of visual perception," *arXiv preprint arXiv:2310.19812*, 2023.

[4] Yu Takagi and Shinji Nishimoto, "High-resolution image reconstruction with latent diffusion models from human brain activity," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 14453–14463.

[5] Paul Scotti, Atmadeep Banerjee, Jimmie Goode, Stepan Shabalin, Alex Nguyen, Aidan Dempster, Nathalie Verlinde, Elad Yundler, David Weisberg, Kenneth Norman, et al., "Reconstructing the mind's eye: fmri-to-image with contrastive learning and diffusion priors," *Advances in Neural Information Processing Systems*, vol. 36, pp. 24705–24728, 2023.

[6] Yunpeng Bai, Xintao Wang, Yan-Pei Cao, Yixiao Ge, Chun Yuan, and Ying Shan, "Dreamdiffusion: High-quality eeg-to-image generation with temporal masked signal modeling and clip alignment," in *European Conference on Computer Vision*. Springer, 2024, pp. 472–488.

[7] Shuqi Zhu, Ziyi Ye, Qingyao Ai, and Yiqun Liu, "Eeg-imagenet: An electroencephalogram dataset and benchmarks with image visual stimuli of multi-granularity labels," *arXiv preprint arXiv:2406.07151*, 2024.

[8] Wenhui Peng, Yao Zhang, and Michel Desmarais, "Spatial convolution neural network for efficient prediction of aerodynamic coefficients," in *AIAA Scitech 2021 Forum*, 2021, p. 0277.

[9] Wentao Zhao, Xin Jiang, Baojiang Zhang, Shenghua Xu, and Shuang Wang, "Ctnet: a convolutional transformer network for eeg-based motor imagery classification," *Scientific Reports*, vol. 14, pp. 20237, 2024.

[10] Xing Wu, Guangyuan Ma, Meng Lin, Zijia Lin, Zhongyuan Wang, and Songlin Hu, "Contextual masked auto-encoder for dense passage retrieval," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023, vol. 37, pp. 4738–4746.

[11] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick, "Masked autoencoders are scalable vision learners," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 16000–16009.

[12] Ruiqi Yang and Eric Modesitt, "Vit2eeg: leveraging hybrid pretrained vision transformers for eeg data," *arXiv preprint arXiv:2308.00454*, 2023.

[13] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al., "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[14] Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance, "Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces," *Journal of neural engineering*, vol. 15, no. 5, pp. 056013, 2018.

[15] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al., "Learning transferable visual models from natural language supervision," in *International conference on machine learning*. PmLR, 2021, pp. 8748–8763.

[16] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 10684–10695.

[17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241.

[18] Ze-bin Wu and Jun-qing Yu, "Vector quantization: a review," *Frontiers of Information Technology & Electronic Engineering*, vol. 20, no. 4, pp. 507–524, 2019.

[19] Luis B Almeida, "Multilayer perceptrons," in *Handbook of Neural Computation*, pp. C1–2. CRC Press, 2020.

[20] Peixiang Zhong, Di Wang, and Chunyan Miao, "Eeg-based emotion recognition using regularized graph neural networks," *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1290–1301, 2020.