Brain-media: A Dual Conditioned and Lateralization Supported GAN (DCLS-GAN) towards Visualization of Image-evoked Brain Activities

Ahmed Fares*
ahmd.fares@szu.edu.cn
Research Institute for Future Media Computing
College of Computer Science and Software Engineering
Shenzhen University
Shenzhen, China

Sheng-hua Zhong†
Jianmin Jiang†‡
csshzhong@szu.edu.cn
jianmin.jiang@szu.edu.cn
Research Institute for Future Media Computing
College of Computer Science and Software Engineering
Shenzhen University
Shenzhen, China

ABSTRACT

Essentially, the current concept of multimedia is limited to presenting what people see in their eyes. What people think inside brains, however, remains a rich source of multimedia, such as imaginations of paradise and memories of good old days etc. In this paper, we propose a dual conditioned and lateralization supported GAN (DCLS-GAN) framework to learn and visualize the brain thoughts evoked by stimulating images and hence enable multimedia to reflect not only what people see but also what people think. To reveal such a new world of multimedia inside human brains, we coin such an attempt as "brain-media". By examining the relevance between the visualized image and the stimulation image, we are able to measure the efficiency of our proposed deep framework regarding the quality of such visualization and also the feasibility of exploring the concept of "brain-media". To ensure that such extracted multimedia elements remain meaningful, we introduce a dually conditioned learning technique in the proposed deep framework, where one condition is analyzing EEGs through deep learning to extract a class-dependent and more compact brain feature space utilizing the distinctive characteristics of hemispheric lateralization and brain stimulation, and the other is to extract expressive visual features assisting our automated analysis of brain activities as well as their visualizations aided by artificial intelligence. To support the proposed GAN framework, we create a combined-conditional space by merging the brain feature space with the visual feature space provoked by the stimuli. Extensive experiments are carried out and the results show that our proposed deep framework significantly outperforms the representative existing state-of-the-arts under several settings, especially in terms of both visualization and classification of brain responses to the evoked images. For the convenience of research dissemination, we make the source code openly accessible for downloading at GitHub.[1]

CCS CONCEPTS

• **Computing methodologies → Unsupervised learning**; **Neural networks**.

KEYWORDS

EEG, image generation, brain-media, deep learning, GAN, variant bi-directional LSTM, regional attention gate

ACM Reference Format:
Ahmed Fares, Sheng-hua Zhong, and Jianmin Jiang. 2020. Brain-media: A Dual Conditioned and Lateralization Supported GAN (DCLS-GAN) towards Visualization of Image-evoked Brain Activities. In *Proceedings of the 28th ACM International Conference on Multimedia (MM '20), October 12–16, 2020, Seattle, WA, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3394171.3413858

1 INTRODUCTION

In general, the existing concept of multimedia can be deemed as external since it is fundamentally established by recording what we view instead of what we think. While not all thoughts can be visualized into multimedia, some of their elements, such as imaginations, emotional memories, and especially those responding to the external stimuli images, could be envisioned, and procreated into a novel custom of multimedia. For the expediency of identification, we coin such a novel custom of multimedia, reflecting the human brain interior, as "brain-media". As a matter of fact, study on human brain intelligence has been researched across a number of areas, knowingly, neuroscience, brain science and computer science, in which a number of studies on human brain intelligence has been conducted mainly based upon brain EEG interface [2, 13, 31]. While artificial intelligence is considered the most contemporary tackled theme in computer vision, exploitation of brain intelligence could

*Ahmed Fares is also with the Department of Electrical Engineering, the Computer Engineering branch, Faculty of Engineering at Shoubra, Benha University, Cairo.
†Sheng-hua Zhong and Jianmin Jiang are also with Guangdong Laboratory of Artificial Intelligence & Digital Economy (SZ), Shenzhen University, Shenzhen, China
‡Jianmin Jiang, the corresponding author.

ACM ISBN 978-1-4503-7988-5/20/10…$15.00
https://doi.org/10.1145/3394171.3413858

[1]https://github.com/aneeg/LS-GAN

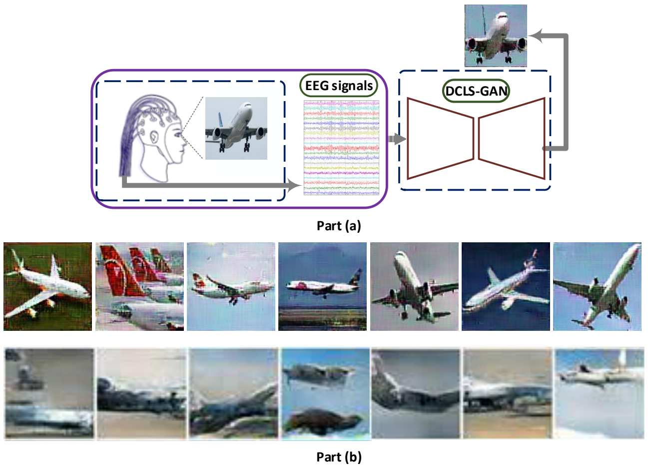

Figure 1: Part (a)- overview of our proposed deep DCLS-GAN on visualization of brain activities and part (b)- the samples of our accomplished experimental results.

boost the potential for more evolving AI tactics as well as their in-field applications. For the preceding years, unceasing research was seeking to comprehend brain manners through EEGs aroused by purposefully designed stimuli for brain-computer interfacing (BCI) [6, 18, 32], and conclusions in neuropsychology disclose numeral distinct categories can be identified by event-related potential (ERP) recorded via EEG data [2, 13]. Additionally, a variety of machine learning paradigms [10, 16, 29] have also been established prioritizing the enigma of multimedia-evoked brain grasping via attempts of pattern recognition and categorizations, with enhanced results. In this paper, we expand the stride of the current EEG-based brain research and uphold it toward the novel custom of "brain-media", and hence delve into the possibility of enabling people to foresee what we thought rather than what we see. To direct such a striving notion into an achievable research focus, we introduce a GAN-based deep framework to envision those brain events evoked by natural images.

To overview our proposed research, we illustrate the conceptual structure of our proposed DCLS-GAN framework as well as some samples of the visualized "brain-media" in Fig. 1. As shown in part (a), the brain is aroused by a "airliner" image and the generated EEG signals are fed into our proposed DCLS-GAN. Via deep adversarial learning of the EEG sequences by our proposed DCLS-GAN, to illustrate the cognitive events within the human brain, an image is created as the output that parallels the brain retorts to the provoking image at the input. The consistency between the input stimuli and output images was thoroughly comparatively scrutinized, enabling us to assess the visualization efficiency of the deep framework regarding the precision of conception and the quality of the reconstructed output image. Part (b) of Fig. 1 demonstrates two series of sample images generated as the output. Hence the first line of samples are the "airliner" images generated by the proposed DCLS-GAN, the second line denotes the samples generated by the current benchmark brain2image GAN [11]. As illustrated, our output images show notably superior quality to that of the benchmark, denoting the enhancement provided by our proposed DCLS-GAN. This paper also provides the validated surpassed accuracy attained by our proposed DCLS-GAN, as concluded by the analogous classification performances.

In summary, the work presented in this paper has double-up impacts, which can be emphasized as: (i) our proposed dual-conditioned GAN utilizes the interfacing amid brain domain and visual domain to promote the adversarial learning and consequently attain notable improvements on both visualization quality and accuracy. Whereas the brain domain affords essential brace for our proposed deep framework to portray the thoughts-allied activities, the visual domain denotes the attributes in visual content to aid with superior quality visualization of those brain thoughts; and (ii) enthused by the marvelous mechanism of hemispheric lateralization and attention, we are the first, to the best of our knowledge, to add a new regional attention gate to the well-known LSTM, and therefore accentuation of the dual variances between the hemispheres, and raising the value of diverse EEG trails. In this manner, the proposed deep framework reinforces our competencies in brain interpretation on the road to its enhanced visualizations.

2 RELATED WORK

While EEG-based brain activity analysis has been extensively researched across several areas, knowingly brain science, bioengineering, psychology, and machine learning over the past decades, little work is reported for EEG-based studies of human brains evoked by images, where multimedia content such as natural images are directly applied as stimuli to generate EEG sequences for human-brain interfacing. Prior to the ubiquity of deep learning, Kaneshiro et al. [10] integrated linear discriminant analysis (LDA) with principal component analysis (PCA) to classify visually-evoked EEG dataset representing by 12 different object classes, and the authors reported an accuracy of 28.87% on their proposed dataset, namely ObjectCategory-EEG dataset. Since the in-depth promotion of deep learning models, various new research trials have been published to leverage its quality in creating ambitious algorithms to accomplish a more real comprehension of brain behaviors and reproducing a perceived visual stimulus through EEGs. Instances of such endeavors can be delineated by the classification of brain behaviors by mining of EEGs. Kulasingham et al. [13] utilized deep automatic encoders and deep belief networks (DBN) to interpret EEGs for identifying unique patterns on their dataset. Yin and Zhang [35] introduced a single-channel EEG classification approach with a deep belief network, extracting mental loads from EEG data, an average precision rate of 71% was reported based on the non-overlapped training and testing of experiments on the EEG data. Lu et al. [14] suggested a frequential DBN (FDBN) to classify the motor imagery. The proposed FDBN network depends on three restricted Boltzmann machines (RBMs) accumulated with a SoftMax regression, in which fast Fourier transform (FFT) and wavelet packet decomposition (WPD) are utilized to acquire the frequency domain description of the EEG data. During the process of training, the FDBN is fine-tuned employing conjugate gradient and backpropagation techniques, and the assessments are completed by utilizing public benchmark datasets. Stober et al. [26] utilized convolutional neural networks (CNNs) and an autoencoder to categorize EEG recordings evoked by audio stimuli, and an accuracy of 28% over 12 vocals. Ogawa et al. [19] used a recurrent neural network (RNN) to simultaneously input video features and video viewers' EEG signals to achieve video classification dependent on user preferences. Spampinato et al. [25], utilized a special type of RNN network, i.e long short term memory (LSTM) network to study an EEG feature description dependent on visual stimuli and created a mapping relationship from deep learned visual characteristics to EEG feature descriptions. Eventually, they utilized their proposed description of EEG signals for categorizing natural images. Compared with other existing methods, these deep learning-based approaches have achieved outstanding classification results on their dataset, namely ImageNet-EEG.

On the other hand, research on generative models is also very active, including generative adversarial networks (GANs) [5], deep convolutional generative adversarial network (DCGAN) [22] and variational autoencoder (VAE) [12]. While these generative models show great performances in generating realistic images from the data distributions, training these models is problematic since it required a huge amount of training data, which are often difficult and inaccessible for EEG-related experiments.

In addition to the recent efforts and studies related to the visualization of brain activities by mining EEG data, Gogna, Majumdar, and Ward [4] proposed deep learning methods to solve the problem of reproducing and classification of EEG data. These deep learning methods are based on an autoencoder for reconstruction. The reported semi-supervised stacked autoencoder used the Split Bregman technique to solve the nonlinear optimization problem. Schirrmeister et al. [24] announced the impacts of CNNs on unraveling and visualizing EEG data. They assessed an enormous number of convolutional neural networks on an EEG decoding task, and exhibited that progresses from the domain of deep learning, including exponential linear units and batch normalization, are essential for accomplishing high exactnesses accuracies. Tirupattur et al. [27] introduced an EEG-based deep learning approach, namely ThoughtViz, for visualizing human thoughts using a GAN-based framework. Kavasidis et al. [11] suggested a structure for producing the visual stimuli content data through EEG data. By employing variable-valued autoencoder (VAE) and generative adversarial networks (GANs), Kavasidis et al. observed that EEG data include patterns associated to visual content, and the content can be used to produce images that are semantically compatible with the input visual stimuli. Palazzo et al. [21] utilized conditional GAN-based structure [11] to produce visual stimuli through EEG data.

Despite the fact that the aforementioned approaches have shown the strength of utilizing deep learning models for brain imagination visualization and categorization, those approaches experience the streaming impediments (i) the raw EEG signals or the obtained time-frequency characteristics based on signal analysis techniques are often utilized as the input, and some features of human brains have not been genuinely examined, such as hemispheric lateralization; (ii) the significance of channel-based spatial information have not been jointly studied with hemispheric lateralization information; (iii) the spatial and dynamic relationships embedded inside the EEG sequences have not been jointly utilized; and eventually (iv) the state-of-the-art precision rate and Inception score (IS) accomplished by Spampinato et al. [25] and Kavasidis et al. are only 82.9% and 5.07, respectively, giving some extension for additional exploration and improvement.

To draw the current work nearer to the practical investigation of the "brain-media" concept, we need to solve two primary issues, which can be featured as: (i) significantly enhance the precision of distinguishing those components that can be visualized as "brain-media" out of the brain thoughts that react to stimulating images; (ii) significantly enhance the quality of such visualizations and thus the visualized images can be relished as any other natural images we generally meet in the existing multimedia. To this end, we present a dual-conditioned and lateralization-supported GAN framework, where the brain feature space is combined with representation learning of visual feature spaces to provide additional support for deep learning of brain EEG sequences and enhancing the visualization achievements. To accomplish seamless combination with EEG representations of brain cognitive responses to the external stimulation via natural images, we propose a new concept of regional attention gate to strengthen the existing LSTM and hence provide the first condition to support the proposed deep learning framework, where the regional information is employed to maintain the hemispheric lateralization information [34]. Hemispheric lateralization refers to the tendency for some neural functions or cognitive processes to be specialized to the right or the left hemispheres of human brains [1]. In addition, the attention mechanism, which allows a deep network to pay attention to only part of the input information, becomes one of the common robust and influential ideas in deep learning [30]. As EEGs are channel-based temporal-spatial signal sequences, some components of human brains are more deeply involved than others, leading to the oscillations in EEG signals and hence generating further spaces for development. To the best of our knowledge, no research has been tried to integrate jointly the hemispheric lateralization and the attention mechanism into a gated structure of the recurrent deep learning model to extract the region-level information from brain signals.

3 THE PROPOSED DCLS-GAN FRAMEWORK

A structure illustration of our proposed DCLS-GAN deep framework is outlined in Fig. 2 for envisioning the "brain-media" components of the human thoughts relating to the natural image stimulation at the input. As seen, to create the first condition (top-left of Fig. 2), we utilize a lateralization-inspired LSTM to decode EEG descriptions from raw brain signals, and to create the second condition (top-right of Fig. 2), we use an auto-encoder to learn data representations and decode visual features across all the candidate images for stimuli.

Structurally, GAN comprises of two sub-networks, knowingly generator (G) and discriminator (D). The generator attempts to generate a sample from a random noise input ($\mathbf{z}$); however, the discriminator examines whether the generator is really generating a fake sample or real sample. While the generator is supposed to catch the overall training data distributions and produce realistic-looking samples, the discriminator would be commonly unsure of whether its inputs are real or fake images. To envision the natural image stimulation and guarantee that such automated visualization can defeat the vagueness and variation of the stimulation images led to the brain activations, we demand to maximize the possibilities from

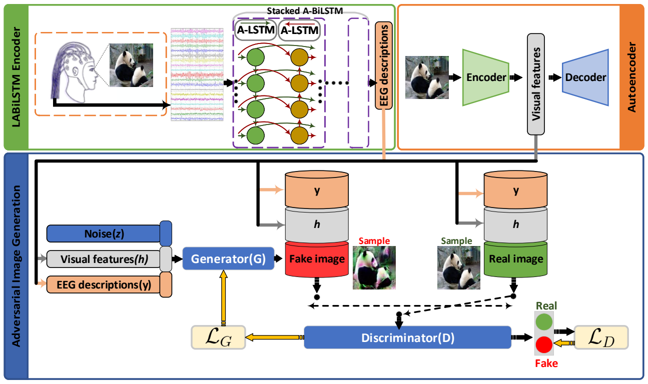

Figure 2: Overview of the proposed deep framework architecture.

both the brain side and the content side of natural images. To this end, our proposed DCLS-GAN produces image samples through both the arbitrary noise vector and the combined description vector, incorporated from the two conditions (see Fig. 2), so as to empower the proposed "brain-media" framework to accomplish the most reliable possible visualization of the caught brain activities regarding both quality and precision.

In order to create the first condition and mine the EEG description of brain activities in the process of adversarial learning, we structure a stack of n lateralization-inspired and bi-directional LSTM layers as seen in Fig. 2. Given the input $\mathbf{e}$ from all channels at time t, explicitly, an extra gate, alluded to as regional attention gate, is made to cooperate with the current 3 gates, and thus their state values , i.e. the regional attention gate Γ^t_{ra}, the update gate Γ^t_u, the forget gate Γ^t_f, and the output gate Γ^t_o, which are demonstrated by colorful boxes in the attention-gated LSTM cell in Fig. 3, can be determined from the raw EEG brain signals $\mathbf{E} = [\mathbf{e}_i]_{i=1}^{l_{ch}}$, where $i \in [1,\ l_{ch} = 128]$ is the index for EEG channels, l_{ch} is the number of EEG channels, and the previous layer output $\mathbf{a}^{t-1}$ as per the accompanying equation:

$$\begin{pmatrix}\Gamma_{ra}^t\\ \Gamma_f^t\\ \Gamma_u^t\\ \Gamma_o^t\end{pmatrix} = g\begin{pmatrix}\mathbf{W}_{ra} & \mathbf{U}_a & 0\\ 0 & \mathbf{U}_f & \mathbf{W}_f\\ 0&\mathbf{U}_u&\mathbf{W}_u\\ 0&\mathbf{U}_o&\mathbf{W}_o\end{pmatrix}\begin{pmatrix}\left[\left(\mathbf{E}^t_{[l],j}-\mathbf{E}^t_{[r],j}\right)\ \mathbf{E}^t_{[m]}\right]^\top\\ \mathbf{a}^{t-1}\\ \Gamma^t_{ra}\end{pmatrix} + g\begin{pmatrix}\mathbf{b}_{ra}\\ \mathbf{b}_f\\ \mathbf{b}_u\\ \mathbf{b}_o\end{pmatrix} \quad (1)$$

where, for $m \in \{ra, u, f, o\}$, $\mathbf{W}_m$ is the weight matrix interfacing the layer input with the 4 gates, $\mathbf{U}_m$ is the weight matrix joining the past cell output state to the 4 gates, and $\mathbf{b}_m$ is the bias vector. The function $g(.)$ is structured as ReLU activation function for Γ^t_{ra} and element-wise sigmoid for Γ^t_u, Γ^t_f, and Γ^t_o, respectively. To accomplish the demanded lateralization impact, the state of the regional attention gate Γ^t_{ra} is maintained through the three gates, and also Γ_{ra} parts the EEG data into three groups, including the right hemisphere group, the left hemisphere group, and the central

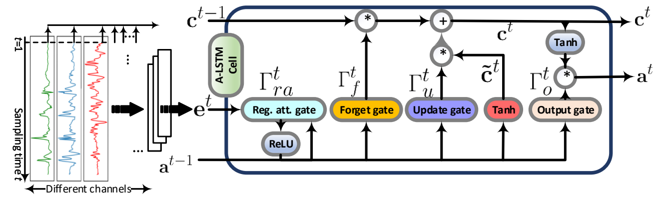

Figure 3: Overview of the proposed attention-gated LSTM cell architecture.

part. By indicating the right hemisphere group, the left hemisphere group, and the central group as, $\mathbf{E}_{[r]}$, $\mathbf{E}_{[l]}$, and $\mathbf{E}_{[m]}$, respectively, each channel $\mathbf{e}_i$ can be connected to one group according to its corresponding electrode physical location. Besides, each channel in the right hemisphere group has an associating channel in the left hemisphere group. Γ_{ra} consolidates the difference, $(\mathbf{E}^t_{[l],k} - \mathbf{E}^t_{[r],k})$, and the central group, $\mathbf{E}_{[m]}$, into one variable, and then transfers it to the attention part as an input, where $k \in \left[1,\ l_g\right]$ is the index for the right hemisphere, the left hemisphere, and l_g is the number of channels associated to the left hemisphere or the right hemisphere. To advance the way toward including the regional attention-driven mechanism, we introduce the soft regional attention gate, where the input EEG signals of various channels are fully connected with the nodes in the gate. Accordingly, the size of $\mathbf{W}_{ra}$ relies upon the number of channels and the number of nodes in the regional attention gate. In light of the results of (1), the cell output state $\mathbf{c}^t$ and the layer output $\mathbf{a}^t$ (both the forward and the backward outputs) can be determined from the state of the regional attention gate Γ^t_{ra} and the previous layer output $\mathbf{a}^{t-1}$, details of which are given below:

$$\mathbf{c}^t = \Gamma^t_f * \mathbf{c}^{t-1} + \Gamma^t_u * \Big(\overbrace{\tanh(\mathbf{U}_c\mathbf{a}^{t-1} + \mathbf{W}_c\Gamma^t_{ra} + \mathbf{b}_c)}^{\text{Nominee for replacing the memory cell}} \Big) \quad (2)$$

$$\mathbf{a}^t = \Gamma^t_o * \tanh(\mathbf{c}^t) \quad (3)$$

where $\mathbf{W}_c$ is the weight matrix, interfacing the layer input with the nominee for replacing the memory cell. While $\mathbf{U}_c$ is the weight matrix interfacing the past cell output state with the nominee for replacing the memory cell, $\mathbf{b}_c$ is a bias vector controlling the balance of strength between ($\mathbf{U}_c$ and ($\mathbf{W}_c$, and the function tanh(.) indicates a hyperbolic tangent. At the point when other layers are available, the output of the first layer is fed as an input to the subsequent layer and so on. The final output of the deepest LSTM layer is a vector of all outputs, denoted by $\mathbf{Y} = \left[\mathbf{y}^t\right]_{t=1}^{l_s}$. At each time of iteration t, $\mathbf{y}^t$ can be determined according to the standard LSTM equation [8]. For our proposed DCLS-GAN framework; however, only the last component of the output vector, $\mathbf{y}^{l_s}$, is considering as a candidate to represent the first condition.

In order to visualize the "brain-media", we train the generator network $G(\mathbf{z}|\mathbf{y}, \mathbf{h})$ in a conditional GAN framework to map the random inputs from a $p_z(\mathbf{z})$ noise distribution and the combined-conditional vector, including EEG descriptions ($\mathbf{y}$) and the visual features ($\mathbf{h}$), to a target image distribution $p_{data}(\mathbf{x})$ as seen in Fig. 2. The minimax strategy is utilized to train both the generator and the discriminator simultaneously. The generator tries to maximize the likelihood of

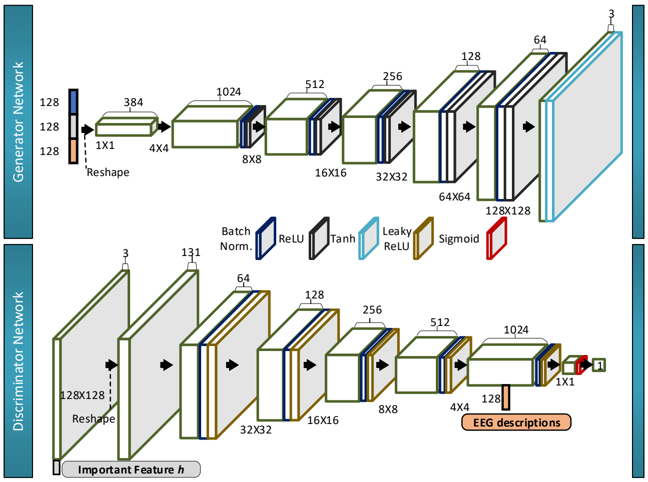

Figure 4: Overview of the proposed DCLS-GAN architecture.

committing the discriminator errors its inputs $p_G(\mathbf{z}|\mathbf{y},\mathbf{h})$ as real; however, the discriminator tries to amplify the likelihood of partner the correct labels, i.e. real samples to $p_{data}(\mathbf{x})$ and fake samples to $p_G(\mathbf{z}|\mathbf{y},\mathbf{h})$. The overall objective function $\mathbf{V}(D,G)$ can be determined by the accompanying equation:

$$\min_G \max_D V(D,G) = \mathbb{E}_{\mathbf{x}\in p_{data}}[\log D(\mathbf{x}|\mathbf{y},\mathbf{h})] + \mathbb{E}_{\mathbf{z}\in p_z}[\log(1 - D(G(\mathbf{z}|\mathbf{y},\mathbf{h})|\mathbf{y},\mathbf{h}))] \quad (4)$$

Practically, the generator loss function $\mathcal{L}_G$ and the discriminator loss function $\mathcal{L}_D$ are designed using a hinge loss, where we change the generator loss function by compiling the constrictive loss to the adversarial loss. To legitimize our decision, we additionally explored utilizing Wasserstein GAN loss via empirical studies. The inception score accomplished by WGAN loss is 6.59; however, the IS value grows to 6.64 by utilizing hinge loss. Additional analyses expose that, by utilizing either the hinge loss or the WGAN loss, our proposed framework still beats the compared benchmarks. Specifically, an image produced by a generator network is transferred as an input to a supplemented constructive loss that assesses the disparity between the generation results and the ground truth. Details of the generator and the discriminator loss functions, $\mathcal{L}_G$ and $\mathcal{L}_D$, can be represented as follows:

$$\mathcal{L}_G = -\alpha\, \mathbb{E}[D(\mathbf{z}|\mathbf{y},\mathbf{h})] + \beta\, \ell_1(p_G(\mathbf{z}|\mathbf{y},\mathbf{h}),\ p_{data}(\mathbf{x})) \quad (5)$$

$$\mathcal{L}_D = -\,\mathbb{E}[\min(0,-1+D(\mathbf{x}|\mathbf{y},\mathbf{h}))] - \mathbb{E}[\min(0,-1-D(G(\mathbf{z})|\mathbf{y},\mathbf{h}))] \quad (6)$$

where α and β are the two weighting coefficients adjusting the contribution of the adversarial and constrictive losses.

A structural diagram of the proposed adversarial image generation sub-framework, including the generator G and discriminator D, which are designed as 2 CNNs inspired by the DCGAN [22] is shown in Fig.4. As seen in the generator network G, the joined-conditional vector $(\mathbf{y},\mathbf{h})$ is connected with the random noise vector $\mathbf{z}$ and a progression of deconvolutions are intended to upsample the concatenated vector to an output image. To the discriminator network D, it begins by getting an image, either real or generated image, linked with the condition vector $\mathbf{h}$ related to the input image. As objected to the G network, the D network plays a progression of convolutions, each of which reduces the size of the feature map spatial dimensions, and afterward adds the conditional vector $\mathbf{y}$, associated with the input image, to the last convolutional layer. Finally, the discriminator additionally computes the output probabilities.

4 EXPERIMENTS

To assess our proposed "brain-media" based deep framework, extensive experiments are carried out and evaluated the visualization performances in terms of both accuracy and quality. For visualization accuracy, we utilize a part of our proposed deep framework, the attention-gated LSTM encoder, to classify the EEG descriptions of brain signals into a class of candidate images for stimuli and see if the classified category resides the same as that of the input image or not. To assess the quality of the proposed "brain-media" visualization, we primarily investigate the generated output images and evaluate their quality on subjective perception basis in the same way as that for the existing multimedia (images), although some quantified testing is also carried out via Inception score (IS) measurements.

4.1 Experimental Settings Details

As illustrated in Fig. Fig. 2, the size of all layers in the attention-gated LSTM is primed to 68, involving the consequent non-linear layers, and with double layers in the stacked BiLSTM ($n = 2$). The iteration limit is set to 2500, and the batch size is set to 440. Regarding the autoencoder, 4 deconvolutional layers are included in the encoder network, which receipts an input image and yields the image representations, i.e. the visual features. Conversely, the decoder network entails of 5 convolutional layers, which receives the image representation as input and attempts to restore the same image as an output. The number of epochs is set to 200, and the batch size is set to 128.

Concerning to the networks inside the GAN, as shown in Fig. 4, the generator network takes a 384-concatenated-dimensional vector as the input, embracing 128-dimensional random noise vector z and a combined-conditional vector consists of 128-dimensional visual features h and 128-dimensional propped encoder output y. It reformats this input vector to a 4-dimensional vector and then serves it to a series of 5 layers, each entails of 3 procedures, incorporating deconvolutions, Batch Normalization, and ReLU operation. This chain of operations doubles the spatial dimensions of the input vector while bisects its number of channels before the final one, which outputs a $128 \times 128 \times 3$ RGB colored image squashed between values of -1 and 1 through the `tanh` function. Furthermore, the discriminator network yields a concatenated input involving $128 \times 128 \times 3$ images and their linked 128-dimensional autoencoder output h. Similarly, the input is also redesigned into a 4-dimensional vector and then served into a series of 5 layers, each one entails of three procedures, including convolutions, Batch Normalization, and Leaky ReLU operation. This series of operations bisects the spatial dimensions of the input while doubles its number of channels, and the output of the final convolutional layer is concatenated with its

Table 1: Summary of experimental parameters

Parameters/Dataset	ImageNet-EEG	ObjectCategory-EEG
No. of classes	40	6
No. of stimuli per class	50	12
Total No. of stimuli	2000	72
No. of trails per subject per stimuli	1	72
No. of subjects	6	10
Time for each stimuli	500ms	500ms

associated 128-dimensional EEG descriptions y. The last layer is levelled and then served to a single sigmoid output.

We were facing two major encounters: balancing the generator and discriminator, and overfitting due to the size of the dataset. To overcome the first challenge, we have considered the two-timescale update rule (TTUR) technique [7] which is inferred by Goodfellow et al [33] as more successful for unbalancing the learning rate between the generator and discriminator than other methods as spectral normalization. So, we recommended it as grants various learning rates to balance for the slow learning rate of the discriminator. Meanwhile the other approaches may have its benefits, our choice provides further pros: (i) adequate appraisal with the current benchmarks could be clearly applied by deactivating the self-attention layer; and (ii) SAGAN is basically derived from DCGAN, which tops most of the state-of-the-art research on GANs, affording extensive compatibility with research semantics on GANs. Specifically, we set the discriminator learning rate as 0.0004, and the generator learning rate at 0.0001, providing lesser generator steps for every single discriminator step. For the second challenge, we utilized the chief dataset, ImageNet-EEG, to provide us with appropriate room to explore the overfitting problem. The limited number of images with the allied EEG recordings (50 recording per class) renders either the generator or the discriminator to overfit if we hastly train these two networks on it. Consequently, we train the proposed GAN in double steps. First, we train the proposed DCLS-GAN using solely images from ImageNet [3] without EEGs for 100 epochs. At this step, the attention-gated LSTM conditional vector is set to 0. Secondly, we re-trained the models on the images with EEGs for 50 more epochs. During training process, data is augmented by resizing images at 143×134 pixels, extracting random 128×128, and flipping images horizontally with a chance of 50%. Our deep framework is implemented on a Tesla® P100 GPU.

For standardization goals, the proposed DCLS-GAN framework is weighed against the EEG-based image generation methods [11, 21], which are the latest deep learning methods conditioned by brain signals on the ImageNet-EEG dataset. In this research, we exploit the popular Inception score (IS) [23] for evaluating the GANs output images by examining two conditions simultaneously, ensuring that the images have variety, and each image distinctly represent something. If both conditions are true, the floating-point score will be high. If either or both are false, the floating-point score will be small.

4.2 Assessment of Visualization Accuracy

In order to assess our proposed DCLS-GAN deep framework regarding the visualization accuracy, experiments are carried out to test the EEG-based classification precisions of our proposed framework as opposed to the existing research. 2 publicly available datasets are utilized, namely ImageNet-EEG [25] and ObjectCategory-EEG [10], so as to diminish the risk of overfitting to any specific dataset and restricting the generality of our proposed research. Table 1 summarizes the experimental parameters, and Fig. 5 reports the experimental results in terms of the classification accuracies for our proposed attention-gated LSTM encoder and 5 benchmarks representing the existing state-of-the-arts, including the RNN-based method [25], Siamese network [20], multimodal network [9], and CogniNet [17], and the RS-LDA method [10]. As observed, while the classification accuracy accomplished by our proposed encoder network is 98.4%, the RNN-based method, Siamese network, multimodal network, CogniNet, and RS-LDA compared are 82.9%, 91.4%, 94.1%, 89.6% and 13.0%, respectively.

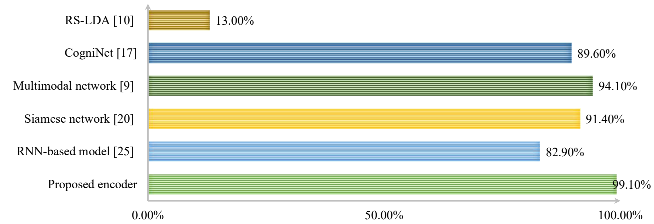

Figure 5: Comparative classification accuracies between the existing benchmarks and the proposed encoder.

Table 2: Comparative assessment between RS-LDA and our proposed encoder

System/No. of classes	6 classes	2 classes Faces vs objects
RS-LDA [10]	40.68%	81.06%
Proposed encoder	61.10%	89.06%

For the convenience of comparative examinations and result analysis, further experiments are carried out to verify the effectiveness of our encoder network for EEG-based object categorization on ObjectCategory-EEG dataset [10]. In these experiments, we utilized the same experiment set up as the existing work [10].

The experimental results in terms of the classification accuracies for our proposed encoder and the RS-LDA method [10] are summarized in Table 2. As seen, the classification accuracies accomplished by our proposed encoder are 61.10% and 89.06%; however, the RS-LDA compared are 40.68% and 81.06% on 6-classes and 2-classes, respectively. From these outcomes, we claim: (i) the generality of our proposed attention-gated LSTM encoder is approved; (ii) the generalization capability of our proposed encoder is better than that of RS-LDA [10], supported by the better results achieved upon both datasets, ObjectCategory-EEG and ImageNet-EEG; (iii) the improvement accomplished by our proposed has almost the same ratio upon both data sets, which is about 20%.

For the convenience of further comparative investigation and result analysis, Fig.6 demonstrates the confusion matrix of each class for ObjectCategory-EEG. As the diagonal of the confusion matrix expresses the highest value in each row, labels predicted by the proposed encoder were most often the correct class labels.

	HB	HF	AB	AF	FV	IO
HB	50.02	6.60	6.60	15.60	6.60	14.60
HF	0.88	95.49	0.88	1.88	0.00	0.88
AB	4.87	4.87	60.63	14.87	2.87	11.87
AF	9.77	7.77	7.77	62.15	6.77	5.77
FV	10.01	7.01	13.01	7.01	43.96	19.01
IO	9.98	4.98	5.98	4.98	14.98	59.11

Figure 6: Proposed encoder confusion matrix.

Table 3: IS comparative assessment .

Model	IS
DCLS-GAN	6.64
EEG-based GAN [21]	5.07
brain2image GAN [11]	5.07
brain2image VAE [11]	4.49

As observed, while the precision rate of the "human face (HF)" category is better than others, with 95.49% of trials being labeled correctly, the lower right portion of the confusion matrix, such as the "fruit vegetable (FV)" and "inanimate object (IO)", show notable confusion. Those findings are consistent with what is reported in the existing work [10].

4.3 Qualitative and Quantitative Assessment on Visualization

To assess the quality of our proposed representations and visualizations, we essentially study our DCLS-GAN framework to ImageNet-EEG dataset and examine our visualized generated images with those produced by the existing benchmarks [11, 21] with equivalent experimental settings.

In order to quantify the contributions of our DCLS-GAN deep framework to the quality of brain activity visualizations, we we additionally figured the IS on 50000 generated sample images, i.e, each class creates a sample of 1250 images. The experimental results in terms of the ISs for our proposed DCLS-GAN deep framework and the existing state-of-the-arts, including the EEG-based GAN [21], brain2image GAN [11], and brain2image VAE [11] are summarized in Table 3. The IS accomplished by our proposed DCLS-GAN deep framework is 5.89; however, the IS achieved by both brain2image GAN and EEG-based GAN is 5.07, and the IS accomplished by brain2image VAE is 4.49.

In order to qualify the contribution of our DCLS-GAN deep framework, a further examination is carried. In previous research [11, 21], high-quality outcomes for 3 of the ImageNet-EEG visual categories, including "jack-o′-lantern", "panda", and "airliner", and low-quality outcomes for other 3 of the ImageNet-EEG visual categories, including "banana", "capuchin", and "bolete", are summarized and represented for visual inspections and subjective evaluations. We follow the same approach and show 2 sets of generated samples by our proposed and the 3 existing benchmarks in Fig. 7 and Fig. 8, in which the 2 benchmarks, EEG-based GAN and brain2image GAN , are from the same authors and the samples

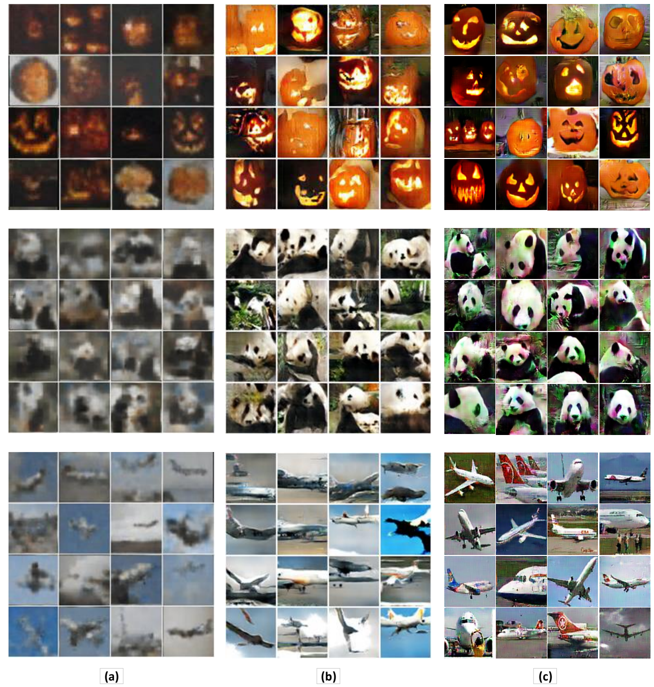

Figure 7: Qualitative comparison among the visualization samples for the categories of "jack-o′-lantern", "panda", and "airliner" achieved by: (a) brain2image VAE, (b) brain2image GAN, and (c) our proposed DCLS-GAN deep framework.

generated by these two benchmarks are the same too. Thus, the accompanying observations can be addressed: (i) the quality of the visual content across the compared systems is different; (ii) our proposed DCLS-GAN deep framework can turn the EEG descriptions into a significant and class-dependent images than other compared systems; and (iii) the visual quality of our generated images on all compared categories are better that of the compared 3 benchmarks. While the first and the third observations validate our contribution to the quality improvement upon the proposed GAN model in generating images, the second observation confirms our contribution in introducing the new concept of "brain-media" and outlines the potential for forming it into plausible examination headings, where semantic components of the brain thoughts can be obtained and visualized into enjoyable multimedia content by EEG-based deep learning models.

As shown by the existing work over areas of psychology, brain science, and neural computations etc, eventually, the essential methods for interfacing with human brains at present are two approaches, including EEG or fMRI [15, 28]. In other words, brain reactions to the external stimuli are basically described either by EEGs or by fMRIs. To this end, our proposed framework is basically evolved to deeply learn and extract, out of those EEG data, the brain responding activities to the external stimulations of the input image. When the EEGs are classified through our proposed attention-gated LSTM encoder, for instance, we are basically attempting to interpret which class of the input image the brain activity is reacting to. When the EEGs are visualized through the proposed GAN-based

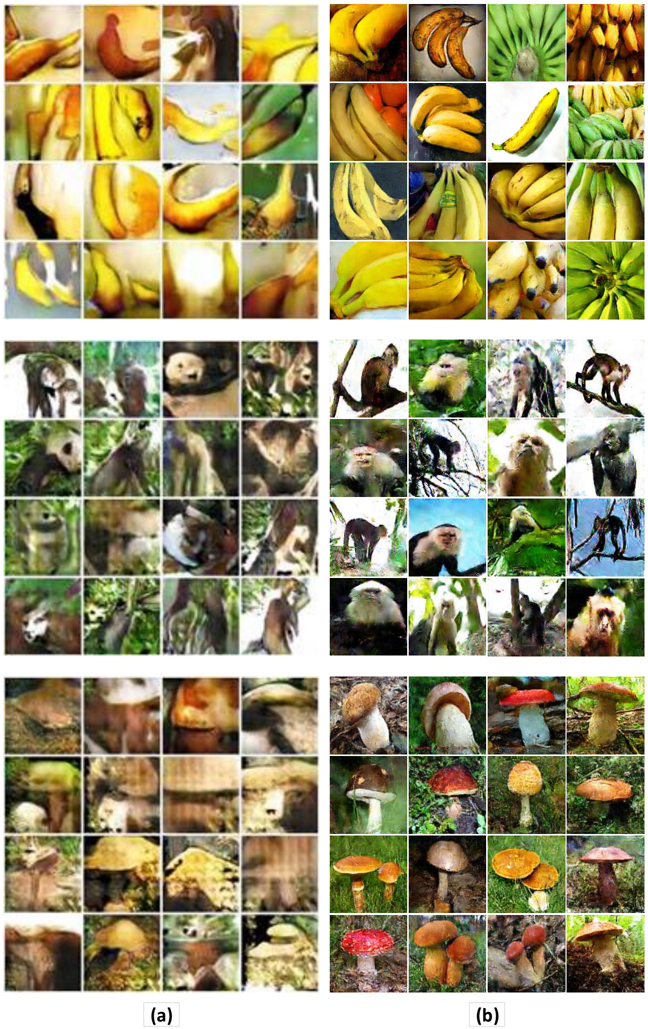

Figure 8: Qualitative comparison among visualization samples for the categories of "banana", "capuchin", and "bolete" achieved by: (a) EEG-based GAN and brain2image GAN, and (b) DCLS-GAN deep framework.

model, equally, we are attempting to visualize the brain activities reacting to the input stimuli image. Thusly, the visualized image indicates a form of our introduced "brain-media" as long as: (a) the visualized image quality is good enough to be enjoyed in the same way as that of any natural image, and (b) the visualized image content is meaningful to human understandings and perceptions. To support the statement further, we complete extra ablation studies and summarize the results in Table 4, from which it can be observed that EEG sequences provide principal support for learning and analyzing brain activities reacting to natural image stimulations. The extra condition included from the visual feature space just offers minimal support for brain activity analysis as inferred by the small difference in ISs; however, the visualization results demonstrate that the extra condition plays a significant role in enhancing the quality of the visualized output image.

Table 4: Assessment under different configurations.

Training of DCLS-GAN	Testing of DCLS-GAN	IS
Brain+visual feature space	Brain+visual feature space	6.64
Brain+visual feature space	Brain feature space only	6.39
Brain feature space only	Brain feature space only	6.34

5 CONCLUSIONS

In this paper, we have introduced an innovative dual-conditioned and lateralization supported GAN framework aiming to appraise the novel concept of "brain-media" through the incorporation of the brain and visual feature spaces, offering an inventive partner to the multimedia empire. At different locations, the sensitivity of the produced EEG signals is instantly dissimilar, hence we introduce a regional attention gate to the LSTM and permit it to unearth the region-level information to uphold and accentuate the hemispheric lateralization or cognitive means of human brains. Moreover, the adjoined regional attention gate also assesses and grasps the impact of various EEG channels via the integrated channel-level attention mechanism, and consequently lead the proposed DCLS-GAN to seize the dynamic veiled associations in the EEGs. Extensive experiments on ImageNet-EEG and Object Category-EEG, prove that our framework surpasses the current state-of-the-arts under numerous settings. Besides, our experiments have provided solid evidence to reinforce that data derived directly from human brains has the potential to: (i) authorize the upgraded machine learning models to make proficient comprehension of human brain cognitive means; (ii) express vision-related information into multimedia translation of brain retorts to the external stimuli; and (iii) generate the brain-perceived multimedia content via EEG demonstrations and their deep learnings.

Whilst the concept of "brain-media" we proposed in this paper has the aptitude of computers empowerment to comprehend brain processes, it is not the equivalent to "mind reading". Principally, we are seeking to seize those semantic brain activities that can be transformed into joyful multimedia content. Obviously, not all human thoughts can be visualized into "brain-media". Human brain activities are vastly complex; however, current research has to be restricted to brain activities aroused via external images or graphic designs in order to render the procedure handy and controllable. The future research based upon the novel concept of "brain-media" can be foreseen as human brains can be observed by computers to catch those meaningful semantics and envisaged into "brain-media" devoid of any external stimuli. Consequently, the present research with external stimulation can be regarded as standards provider for the mentioned objective similar to the setting of training deep learning frameworks.

ACKNOWLEDGMENTS

The authors wish to acknowledge the financial support from: (i) Natural Science Foundation China (NSFC) under the Grant No. 61620106008; and (ii) The second round of Shenzhen University Research Foundation funding (2018-2020).

REFERENCES

[1] Nawfal Al-Hadithi, Ahmed Al-Imam, Manolia Irfan, Mohammed Khalaf, and Sara Al-Khafaji. 2016. The relation between cerebral dominance and visual analytic skills in Iraqi medical students, a cross sectional analysis. *Asian Journal of Medical Sciences* 7, 6 (Oct. 2016), 47–52. https://doi.org/10.3126/ajms.v7i6.15205

[2] Koel Das, Barry Giesbrecht, and Miguel P Eckstein. 2010. Predicting variations of perceptual performance across individuals from neural activity using pattern classifiers. *Neuroimage* 51, 4 (2010), 1425–1437.

[3] Jia Deng, Wei Dong, Richard Socher, Li Jia Li, Kai Li, and Fei Fei Li. 2009. ImageNet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. 248–255.

[4] Anupriya Gogna, Angshul Majumdar, and Rabab Ward. 2017. Semi-supervised Stacked Label Consistent Autoencoder for Reconstruction and Analysis of Biomedical Signals. *IEEE Transactions on Biomedical Engineering* 64, 9 (2017), 2196–2205.

[5] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative Adversarial Nets. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2* (Montreal, Canada) *(NIPS'14)*. MIT Press, Cambridge, MA, USA, 2672–2680. http://dl.acm.org/citation.cfm?id=2969033.2969125

[6] Andrea M Green and John F Kalaska. 2011. Learning to move machines with the mind. *Trends in neurosciences* 34, 2 (2011), 61–75.

[7] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. 2017. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. In *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.). Curran Associates, Inc., 6626–6637. http://papers.nips.cc/paper/7240-gans-trained-by-a-two-time-scale-update-rule-converge-to-a-local-nash-equilibrium.pdf

[8] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[9] J. Jiang, A. Fares, and S. Zhong. 2019. A Context-Supported Deep Learning Framework for Multimodal Brain Imaging Classification. *IEEE Transactions on Human-Machine Systems* (2019), 1–12. https://doi.org/10.1109/THMS.2019.2904615

[10] Blair Kaneshiro, Marcos Perreau Guimaraes, Hyung-Suk Kim, Anthony M Norcia, and Patrick Suppes. 2015. A representational similarity analysis of the dynamics of object processing using single-trial EEG classification. *Plos one* 10, 8 (2015), e0135697.

[11] Isaak Kavasidis, Simone Palazzo, Concetto Spampinato, Daniela Giordano, and Mubarak Shah. 2017. Brain2Image: Converting Brain Signals into Images. In *Proceedings of the 2017 ACM on Multimedia Conference*. ACM, 1809–1817.

[12] Diederik P. Kingma and Max Welling. 2014. Auto-Encoding Variational Bayes. In *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*. http://arxiv.org/abs/1312.6114

[13] JP Kulasingham, V Vibujithan, and AC De Silva. 2016. Deep belief networks and stacked autoencoders for the P300 Guilty Knowledge Test. In *Biomedical Engineering and Sciences (IECBES), 2016 IEEE EMBS Conference on*. IEEE, 127–132.

[14] Na Lu, Tengfei Li, Xiaodong Ren, and Hongyu Miao. 2017. A deep learning scheme for motor imagery classification based on restricted boltzmann machines. *IEEE transactions on neural systems and rehabilitation engineering* 25, 6 (2017), 566–576.

[15] S Makeig, M Westerfield, TP Jung, S Enghoff, J Townsend, E Courchesne, and TJ Sejnowski. 2002. Electroencephalographic sources of visual evoked responses. *Science* 295 (2002), 690–694.

[16] Jinyoung Moon, Yongjin Kwon, Kyuchang Kang, Changseok Bae, and Wan Chul Yoon. 2015. Recognition of Meaningful Human Actions for Video Annotation Using EEG Based User Responses. In *International Conference on Multimedia Modeling*. Springer, 447–457.

[17] Pranay Mukherjee, Abhirup Das, Ayan Kumar Bhunia, and Partha Pratim Roy. 2018. Cogni-Net: Cognitive Feature Learning through Deep Visual Perception. *CoRR* abs/1811.00201 (2018). arXiv:1811.00201 http://arxiv.org/abs/1811.00201

[18] Gernot R Muller-Putz and Gert Pfurtscheller. 2008. Control of an electrical prosthesis with an SSVEP-based BCI. *IEEE Transactions on Biomedical Engineering* 55, 1 (2008), 361–364.

[19] T. Ogawa, Y. Sasaka, K. Maeda, and M. Haseyama. 2018. Favorite Video Classification Based on Multimodal Bidirectional LSTM. *IEEE Access* 6 (2018), 61401–61409. https://doi.org/10.1109/ACCESS.2018.2876710

[20] S. Palazzo, C. Spampinato, I. Kavasidis, D. Giordano, J. Schmidt, and M. Shah. 2020. Decoding Brain Representations by Multimodal Learning of Neural Activity and Visual Features. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2020), 1–1.

[21] Simone Palazzo, Concetto Spampinato, Isaak Kavasidis, Daniela Giordano, and Mubarak Shah. 2017. Generative Adversarial Networks Conditioned by Brain Signals. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*. 3430–3438. https://doi.org/10.1109/ICCV.2017.369

[22] Alec Radford, Luke Metz, and Soumith Chintala. 2016. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. In *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*. http://arxiv.org/abs/1511.06434

[23] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. 2016. Improved Techniques for Training GANs. In *Proceedings of the 30th International Conference on Neural Information Processing Systems* (Barcelona, Spain) *(NIPS'16)*. Curran Associates Inc., USA, 2234–2242. http://dl.acm.org/citation.cfm?id=3157096.3157346

[24] Robin Tibor Schirrmeister, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. 2017. Deep learning with convolutional neural networks for EEG decoding and visualization. *Human brain mapping* 38, 11 (2017), 5391–5420.

[25] C. Spampinato, S. Palazzo, I. Kavasidis, D. Giordano, N. Souly, and M. Shah. 2017. Deep Learning Human Mind for Automated Visual Classification. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 4503–4511. https://doi.org/10.1109/CVPR.2017.479

[26] Sebastian Stober, Avital Sternin, Adrian M Owen, and Jessica A Grahn. 2015. Deep feature learning for EEG recordings. *arXiv preprint arXiv:1511.04306* (2015).

[27] Praveen Tirupattur, Yogesh Singh Rawat, Concetto Spampinato, and Mubarak Shah. 2018. ThoughtViz: Visualizing Human Thoughts Using Generative Adversarial Network. In *Proceedings of the 26th ACM International Conference on Multimedia* (Seoul, Republic of Korea) *(MM '18)*. ACM, New York, NY, USA, 950–958. https://doi.org/10.1145/3240508.3240641

[28] Jing Wang, Vladimir L. Cherkassky, Ying Yang, Kai min Kevin Chang, Robert Vargas, Nicholas Diana, and Marcel Adam Just. 2016. Identifying thematic roles from neural representations measured by functional magnetic resonance imaging. *Cognitive Neuropsychology* 33, 3-4 (2016), 257–264. https://doi.org/10.1080/02643294.2016.1182480 arXiv:https://doi.org/10.1080/02643294.2016.1182480 PMID: 27314175.

[29] Jun Wang, Eric Pohlmeyer, Barbara Hanna, Yu-Gang Jiang, Paul Sajda, and Shih-Fu Chang. 2009. Brain state decoding for rapid image retrieval. In *Proceedings of the 17th ACM international conference on Multimedia*. ACM, 945–954.

[30] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. 2015. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. In *ICML*. 2048–2057. http://proceedings.mlr.press/v37/xuc15.html

[31] Longhao Yuan and Jianting Cao. 2018. Patients' EEG Data Analysis via Spectrogram Image with a Convolution Neural Network. In *Intelligent Decision Technologies 2017*, Ireneusz Czarnowski, Robert J. Howlett, and Lakhmi C. Jain (Eds.). Springer International Publishing, Cham, 13–21.

[32] Dalin Zhang, Lina Yao, Xiang Zhang, Sen Wang, Weitong Chen, Robert Boots, and Boualem Benatallah. 2018. Cascade and Parallel Convolutional Recurrent Neural Networks on EEG-based Intention Recognition for Brain Computer Interface. In *AAAI Conference on Artificial Intelligence*. https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16107

[33] Han Zhang, Ian Goodfellow, Dimitris Metaxas, and Augustus Odena. 2019. Self-Attention Generative Adversarial Networks. In *Thirty-sixth International Conference on Machine Learning (ICML)*.

[34] Sheng-hua Zhong, Ahmed Fares, and Jianmin Jiang. 2019. An Attentional-LSTM for Improved Classification of Brain Activities Evoked by Images. In *Proceedings of the 27th ACM International Conference on Multimedia* (Nice, France) *(MM '19)*. ACM, New York, NY, USA, 1295–1303. https://doi.org/10.1145/3343031.3350886

[35] Yin Zhong and Zhang Jianhua. 2017. Cross-subject classification of mental fatigue by neurophysiological signals and ensemble deep belief networks. In *Control Conference (CCC), 2017 36th Chinese*. IEEE, 10966–10971.