

EEG2IMAGE: IMAGE RECONSTRUCTION FROM EEG BRAIN SIGNALS

Prajwal Singh*

Pankaj Pandey†

Krishna Miyapuram†

Shanmuganathan Raman*

* CVIG Lab, † Brain Lab

*† Indian Institute of Technology Gandhinagar, India
{ singh_prajwal, pankaj.p, kprasad, shanmuga}@iitgn.ac.in

ABSTRACT

Reconstructing images using brain signals of imagined visuals may provide an augmented vision to the disabled, leading to the advancement of Brain-Computer Interface (BCI) technology. The recent progress in deep learning has boosted the study area of synthesizing images from brain signals using Generative Adversarial Networks (GAN). In this work, we have proposed a framework for synthesizing the images from the brain activity recorded by an electroencephalogram (EEG) using small-size EEG datasets. This brain activity is recorded from the subject's head scalp using EEG when they ask to visualize certain classes of Objects and English characters. We use a contrastive learning method in the proposed framework to extract features from EEG signals and synthesize the images from extracted features using conditional GAN. We modify the loss function to train the GAN, which enables it to synthesize 128×128 images using a small number of images. Further, we conduct ablation studies and experiments to show the effectiveness of our proposed framework over other state-of-the-art methods using the small EEG dataset.

Index Terms— Deep Learning, EEG, GAN

1. INTRODUCTION

Human visual system is considered a highly advanced intelligent information processor that generates rich 3D visuals with semantic construction. The most challenging problem is to train artificial machines to construct images from brain activity directly to semantic categories [1]. The possibility of brain-to-image construction significantly contributes to advancing the Brain-Computer Interface (BCI) technology. The core purpose of BCI using invasive or non-invasive techniques is to provide communication and control of external devices by thought alone or using minimal muscular activity. This effort would be highly relevant in neuro-rehabilitation, i.e. to support patients with disabilities to have better everyday communication in their lives. Decoding brain responses to imagination/visual stimuli would greatly benefit the communication exchange for the disabled people. The most widely employed brain imaging modality with high temporal precision is Electroencephalography (EEG) due to its relatively lower cost and portability.

EEG is a non-invasive technique which makes it the most practical methodology to record the electrophysiological dynamics of the brain. EEG signals have been used to analyze a wide spectrum of research from aspects of cognition to clinical aspects [2]. For decades,

This work is supported by Prime Minister Research Fellowship (PMRF-2122-2557) to PS and thanks to SERB and PlayPowerLabs for PMRF to PP. We thank FICCI for facilitating the PMRF of PP.

EEG signals have been widely employed for classifying several disorders or understanding brain dynamics. The successful implication of the past results can be seen in the BCI. A previous effort made by the research community has shown promising results to augment healthy individuals with additional sensory or motor capabilities [3]. The most intriguing task is to decode the content of the mind using brain signals and draw a link between them. Two most challenging endeavors in this space are to reconstruct the visualized images [4] and decode imagined speech-to-text [5] based solely on recorded brain signals. Vision neuroscientists made the initial attempts [6, 7, 8] to provide an evidence of visual stimuli features represented in recorded brain activity. These attempts initiated the classification of image categories using brain signals using deep learning and further led to the reconstruction and generation of the images [9].

Our contributions are as follows: 1) A framework that can synthesize images using a small EEG dataset, 2) Use of semi-hard triplet loss [10] to learn features from EEG signals that show better k-means accuracy than the softmax counterpart, as shown in Figs. [2,3] and 3) Use of mode seeking regularization [11] and data augmentation [12] based modification to GAN for synthesizing high-quality images using conditional GAN as shown in Fig. 4(b).¹

2. RELATED WORKS

The development of advanced deep generative architectures in recent times has made it possible to see images from brain signals. The initial study by Kavasidis *et al.* implemented long short-term memory (LSTM) stacked with generative techniques to generate seen images from 40 Image Net classes [4]. Thoughtviz [13] encouraged the design of conditional GAN (cGAN) to decode EEG signals using a small dataset consisting of imagination tasks comprising digits, characters, and objects. Several architectures have been developed using CNN and LSTM on the time-series data of most biological areas. The capability of LSTM for identifying the sequential pattern and CNN to locate the neighborhood features was recently combined with spectral normalization generative adversarial network (SNGAN) to yield seen images from EEG encodings [14]. Researchers are putting effort into reconstructing geometrical shapes from brain activities, primarily in generating precise edges and other low-level details. Further advancement in GAN leads to synthesizing natural geometrical shapes, which enforces semantic alignment constraints to construct natural shapes at pixel-level [15, 16]. Recently, a siamese network was utilized to maximize the relationship between extracted manifold brain feature representation and visual features [17]. The obtained representation demonstrated better image classification and saliency detection performance on the learned

¹<https://github.com/prajwalsingh/EEG2Image>

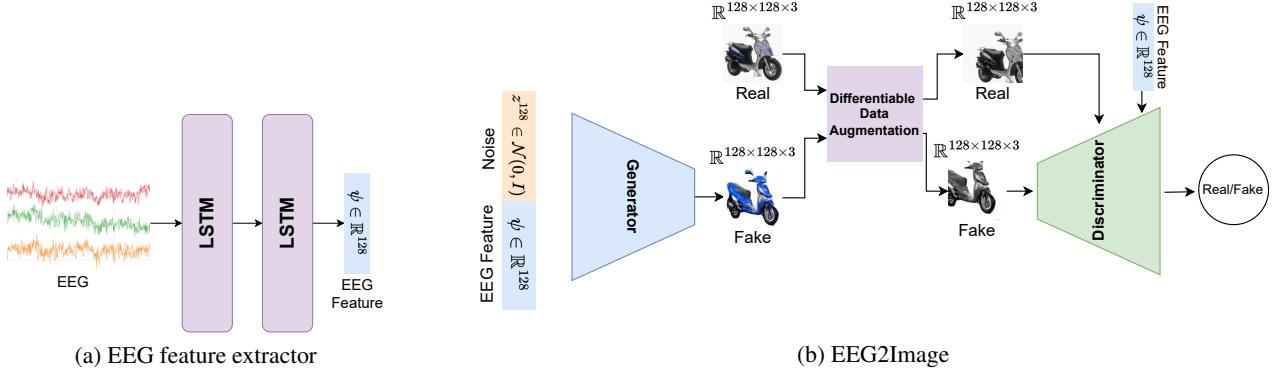


Fig. 1. This figure illustrates the proposed framework for EEG feature extraction and image generation. a) Shows the LSTM network with 128 hidden units that transforms EEG signal into 128D feature vector. b) Shows the GAN network with a data augmentation block that prevents the discriminator from memorizing the small dataset and helps the generator synthesize high-quality images.

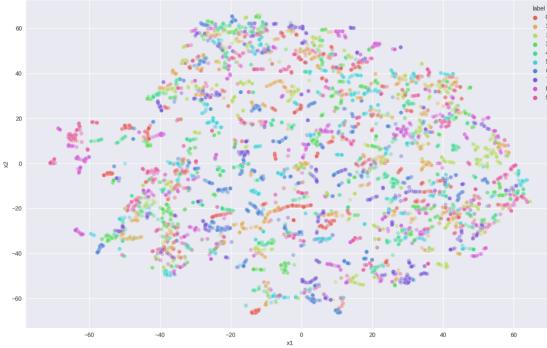


Fig. 2. t-SNE [20] visualization of Object test dataset [21] EEG feature space which is learned using label supervision with test classification accuracy 0.75 and k-means accuracy 0.18.

manifold. Khare *et al.* proposed conditional progressive growing of GANs (CProGAN) to develop perceived images [18] and showed higher inception than previous related work. Recent work on contrastive self-supervised approach has been shown to maximize the mutual information between visual stimulus and corresponding EEG latent representations [19]. They proposed an approach that employed cross-modal alignment enforcing image retrieval at the instance level rather than pixel-level generation.

3. PROPOSED METHOD

In this work, we proposed a framework shown in Fig.1, for visualizing the brain activity EEG signals. The framework consists of a two-phase approach: 1) extracting good features from the EEG signals with a contrastive learning approach and 2) a conditional data-efficient GAN to transform the extracted EEG features to image. In our case, a good feature implies useful information about an image that can help GAN to reconstruct that image.

Feature Extraction. Recent works [17] have shown that the contrastive learning-based approach outperforms the supervised setting in the case of generalized feature learning for downstream tasks such as object detection, classification, saliency map from EEG signals, etc. Building on this, we have used a triplet loss-based contrastive learning [10] approach in the proposed framework for EEG feature learning. Triplet loss aims to minimize the distance between the two data with the same labels and maximize the distance between

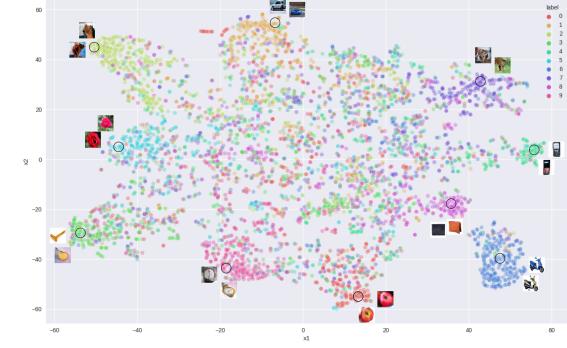


Fig. 3. t-SNE [20] visualization of Object test dataset [21] EEG feature space which is learned using triplet loss with test k-means accuracy 0.53. Each cluster's equivalent EEG-based generated images are also visualized in this plot.

the two data with different labels. To prevent the feature extraction network from squashing the representation of each data into a small cluster, a margin term is used in triplet loss. It ensures that the distance between the feature of the same label data is close to zero and greater than the margin for different label data. The formulation of triplet loss is as follows:

$$\min_{\theta} \mathbb{E} [\|f_{\theta}(x^a) - f_{\theta}(x^p)\|_2^2 - \|f_{\theta}(x^a) - f_{\theta}(x^n)\|_2^2 + \beta] \quad (1)$$

where, f is parameterized function on θ that maps EEG signals to a feature space i.e. $f_{\theta} : \mathbb{R}^{C \times T} \rightarrow \mathbb{R}^{128}$. The goal of Eqn. 1 is to minimize the distance between anchor (a) EEG signal and positive (p) EEG signal of the same class as the anchor and maximize the distance between anchor EEG signal and negative (n) EEG signal of different class with margin distance. This formulation is also known as metric learning or contrastive learning. The idea behind using the formulation is to ensure that the EEG signals generated by the brain activity for similar images should be close to each other in the learned feature space [22]. For learning better features, we have used semi-hard hard triplets, where the distance of the negative sample is more than positive but less than the margin, and also used an online hard-triplet mining strategy similar to [10].

Image Generation. In the proposed framework, we have used a Generative Adversarial Network (GAN) [23] to synthesize the image from the extracted EEG feature. A GAN architecture consists

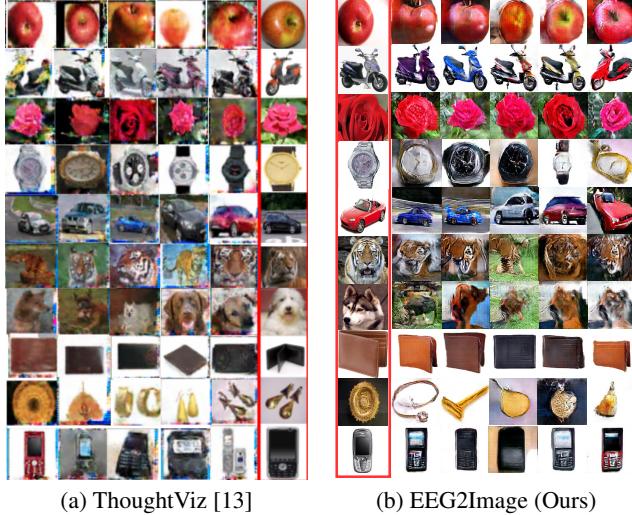


Fig. 4. Qualitative comparison between the images generated by EEG signals using the ThoughtViz method (left) and our proposed framework (right). Images in the red bounding box are the sample images from the Object test dataset [21]. These images are visualized by the participants, and the respective brain activity EEG signals is used here for reconstruction.

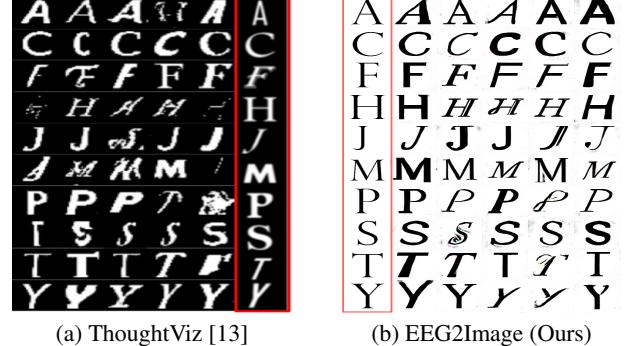
Method	Inception Score
AC-GAN [24]	4.93
ThoughtViz [13]	5.43
EEG2Image (Ours)	6.78

Table 1. Comparison of Inception Score values (on all classes of Object dataset [21]).

of two sub-networks: Generator (G) and Discriminator (D). The purpose of a Generator is to learn the transformation between a latent distribution (p_z) and real-world data distribution (p_{data}). In our case, we assume latent distribution as an isotropic Gaussian $\mathcal{N}(0, I)$ from which we sample a noise vector $z \in \mathbb{R}^{128}$. The discriminator learns to distinguish real images from synthesized images. The complete GAN architecture is trained in a min-max optimization setting. Where the discriminator tries to maximize the score for real images $D(x)$ and minimize the score for generated images $D(G(z))$, in contrast to the discriminator, the generator tries to minimize $(1 - D(G(z)))$ and the minimizing of the term is only possible if generator synthesizes photorealistic images. The complete GAN optimization process can also be represented below:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log(D(x))] + \mathbb{E}_{x \sim p_Z(z)} [\log(1 - D(G(z)))] \quad (2)$$

Similar to [13], we aim to develop a framework that can utilize a small-size EEG dataset for generating images from EEG signals. To overcome the problem of small dataset [13] has used the trainable weighted Gaussian layer [25], which learns the mean (μ) and variance (σ) for the encoded EEG signal. In this work, we follow a different strategy than [13]. Instead, we have used a Conditional DCGAN [26] architecture with the following modification 1) following the work of [27], we have used hinge loss for stable training of GAN, 2) we have added a differentiable data augmentation block between generator and discriminator which helps the network in learning from small datasize [12], and 3) to ensure the mode di-



(a) ThoughtViz [13] (b) EEG2Image (Ours)

Fig. 5. Qualitative comparison between the images generated by the ThoughtViz method (left) and our proposed framework (right). Images in the red bounding box are the sample images from the Charater test dataset [21], these images are visualized by the participants, and the respective brain activity EEG signals is used here for reconstruction.

versity of GAN we have also used a mode seeking regularization as proposed in work [11]. In work [27], Lim *et al.* have shown that the vanilla formulation of GAN suffers from mode collapse and unstable training problems. To solve these issues, they formulated an SVM separating hyperplane approach, which is known as GAN Hinge Loss as below:

$$\mathcal{L}_D = \mathbb{E}_{x \sim p_{data}(x)} [\max(0, 1 - D(x))] + \mathbb{E}_{x \sim p_Z(z)} [\max(0, 1 + D(G(z)))] \quad (3)$$

$$\mathcal{L}_G = -\mathbb{E}_{x \sim p_Z(z)} [D(G(z))] \quad (4)$$

Training a GAN for synthesizing photorealistic images requires a large number of data [12], and other deep learning approaches also face the same data scarcity issues. Zhao et al. [12] in their work have shown that the problem of sparse data for training a GAN can be resolved by adding a Differentiable Data Augmentation (DiffAug) block between the generator and discriminator, which is illustrated in Fig.1(b). The issue with sparse data is discriminator can easily memorize the data, which causes the vanishing gradient problem for the generator. The data augmentations we have used for our GAN network are translation and color jittering. The final loss term we aim to optimize for the proposed EEG2Image is given below:

$$\mathcal{L}_D = \mathbb{E}_{(x, \psi) \sim p_{data}(x)} [\max(0, 1 - D(\textcolor{red}{T}(x), \psi))] + \mathbb{E}_{x \sim p_Z(z), \psi \sim p_{data}(x)} [\max(0, 1 + D(\textcolor{red}{T}(G(z, \psi)), \psi))] \quad (5)$$

$$\mathcal{L}_{ms} = \min_G \left(\frac{d_I(G(\psi, z_1), G(\psi, z_2))}{d_z(z_1, z_2)} \right)^{-1} \quad (6)$$

$$\mathcal{L}_G = -\mathbb{E}_{x \sim p_Z(z), \psi \sim p_{data}(x)} [D(\textcolor{red}{T}(G(z, \psi)), \psi)] + \alpha * \mathcal{L}_{ms} \quad (7)$$

where \mathcal{L}_D is discriminator loss, \mathcal{L}_G is generator loss, \mathcal{L}_{ms} is mode seeking regularizer term [11], $\textcolor{red}{T}$ is DiffAugment [12] function, ψ is EEG feature vector and α is regularizer weight term which kept as 1.0 for all the experiments.

4. EXPERIMENTS AND RESULTS

In the first part of this section, we will discuss the experimental setup we used to train the feature extraction and generative network, including the dataset. Later in this section, we will discuss all the ablation studies done to justify choices for the proposed framework.

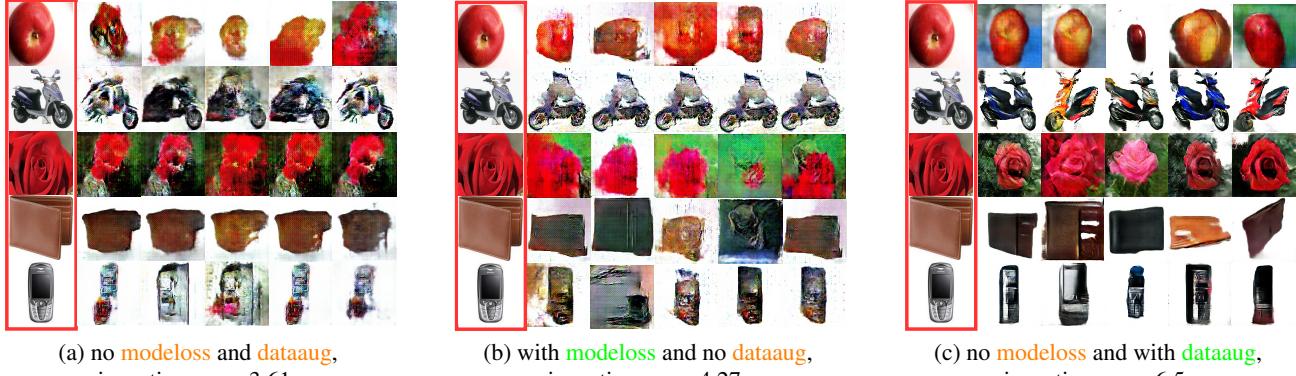


Fig. 6. Ablation study showing the qualitative result on Object dataset [21] using different loss combinations for training the GAN network.

Object Class	Apple (n07739125)	Car (n02958343)	Dog (n02084071)	Gold (n03445326)	Mobile (n02992529)	Rose (n12620196)	Scooter (n03791053)	Tiger (n02129604)	Wallet (n04548362)	Watch (n04555897)	All
Mean	6.09	6.15	6.99	6.98	7.33	5.44	5.81	5.67	6.48	6.67	6.78
SD	0.05	0.084	0.031	0.082	0.030	0.089	0.077	0.057	0.086	0.037	0.086

Table 2. Mean and standard deviation (SD) of Inception scores for each class of Objects dataset [21].

Datasets. We have used the EEG data from [21]. This dataset consists of EEG signals for 3 different subjects: Digits, Characters, and Objects. In our study, we have only used Characters and Object data because these are more diverse and complex data to show the effectiveness of the proposed framework. The Characters dataset consists of ten English alphabet classes and the subset of Chars74K [28]. Similarly, the Objects dataset consists of ten different object classes and the subset of ImageNet [29]. While collecting brain activity EEG signals of the participants, they were asked to think about one of these characters/objects at a time. To record the EEG signals, Emotiv EPOC+ [30] device is used, which has 14 channels with a sampling rate of 128 Hz per channel. For each dataset, 23 participants were asked to visualize every ten classes. Thus we have 230 EEG samples per dataset. For our work, we have used the EEG data provided by the authors with train-test splits [13]. We would like to thank the authors for making it publicly available.

EEG2Feature. The first stage of our proposed framework is to convert EEG signals into useful features for image generation. For this, we design two regimes. In the first regime, we train a classification network for extracting EEG features as done in [13]. The classifier is a LSTM [31] network with 128 hidden units using softmax cross-entropy loss. We use k-means clustering [32] as a metric for the learned EEG feature, i.e., higher k-means accuracy implies better learned representation [22]. The first regime gives us 74.3% & 75.4% classification accuracy on test data of Object dataset & Character dataset [21] and k-means accuracy of 17.8% and 16.3%, further we plot t-SNE map [20] to visualize the clustering of test data features from Object dataset in Fig. 2. For the second regime, we used a contrastive learning approach to learning the feature of an EEG signal. As discussed in the Sec. 3 we used semi-hard triplet loss for training the LSTM [31] network with 128 hidden units. The goal of triplet loss is to structured the feature space in such a way that positive pairs are in close proximity to each other while negative pairs are positioned far apart. The k-means accuracy we got on the test data of the Object dataset is 53%, and the Character dataset is 49%. Further, we plot t-SNE map [20] to visualize the clustering of test data features from Object dataset Fig. 3. We can see that the k-means accuracy and t-SNE plot are better for the second regime. Therefore we decided to use the contrastive learning method as an

EEG feature extractor for our proposed framework.

Feature2Image. The second stage of our proposed framework is to synthesize photorealistic images from extracted EEG features using the first stage. For synthesizing the image, we have used Conditional DCGAN [26] with modification as discussed in Sec. 3. We have used Inception Score (IS) [33] as a metric for image quality comparison with other methods. Table 1 shows our proposed GAN method performed better in synthesizing the images from less number of EEG data. In Table 2 we have shown per class inception score for test data of Object dataset [21]. We also performed the qualitative analysis of synthesized images for both the dataset Object and Character, which are shown in Figs. [4, 5]. We have performed several ablation studies to verify the importance of each loss in training the GAN network for the proposed framework. For that, we trained the proposed conditional GAN (cGAN) for three different regimes on the Object dataset [21]. In the first regime, we train the cGAN without mode seeking regularization and DiffAugment, shown in Fig. 6(a). which has an inception score of 3.61. In the second regime, we have added a mode seeking regularization term only and trained the cGAN from scratch, Fig. 6(b) shows improvement with an inception score of 4.27. In the third and last regime, we train the cGAN with the DiffAugment block, showing a large improvement in the synthesized image as shown in Fig. 6(c) with an inception score of 6.5. Based on these experiments, we used both the mode-seeking regularization term and the DiffAugment block in the proposed framework.

5. CONCLUSION

This work proposes a framework that uses a small-sized dataset for generating images from brain activity EEG signals. Our proposed framework has a better inception score than the previously proposed method for the small-sized EEG dataset and synthesized images of size 128×128 . The framework consists of a contrastive learning approach to learn the good features of EEG data, which is empirically shown to perform better than the softmax-based supervised learning method. We have performed several ablation studies to demonstrate the effectiveness of modified GAN loss function in synthesizing high-quality images. As future work, we plan to tackle large-size EEG datasets and approach for complete self/un-supervised learning for extracting features from EEG data and image synthesis.

6. REFERENCES

- [1] Yoichi Miyawaki, Hajime Uchida, Okito Yamashita, Masa-aki Sato, Yusuke Morito, Hiroki C Tanabe, Norihiro Sadato, and Yukiyasu Kamitani, “Visual image reconstruction from human brain activity using a combination of multiscale local image decoders,” *Neuron*, vol. 60, no. 5, pp. 915–929, 2008.
- [2] Vangelis Sakkalis, “Applied strategies towards eeg/meg biomarker identification in clinical and cognitive research,” *Biomarkers in medicine*, vol. 5, no. 1, pp. 93–105, 2011.
- [3] Mahdi Bamdad, Homayoon Zarshenas, and Mohammad A Auais, “Application of bci systems in neurorehabilitation: a scoping review,” *Disability and Rehabilitation: Assistive Technology*, vol. 10, no. 5, pp. 355–364, 2015.
- [4] Isaak Kavasidis, Simone Palazzo, Concetto Spampinato, Daniela Giordano, and Mubarak Shah, “Brain2image: Converting brain signals into images,” in *25th ACM MM*, 2017.
- [5] Zhenhailong Wang and Heng Ji, “Open vocabulary electroencephalography-to-text decoding and zero-shot sentiment classification,” in *AAAI*, 2022, vol. 36, pp. 5350–5358.
- [6] Thomas Carlson, David A Tovar, Arjen Alink, and Nikolaus Kriegeskorte, “Representational dynamics of object vision: the first 1000 ms,” *Journal of vision*, vol. 13, no. 10, pp. 1–1, 2013.
- [7] Thomas A Carlson, Hinze Hogendoorn, Ryota Kanai, Juraj Mesik, and Jeremy Turret, “High temporal resolution decoding of object position and category,” *Journal of vision*, vol. 11, no. 10, pp. 9–9, 2011.
- [8] Koel Das, Barry Giesbrecht, and Miguel P Eckstein, “Predicting variations of perceptual performance across individuals from neural activity using pattern classifiers,” *Neuroimage*, vol. 51, no. 4, pp. 1425–1437, 2010.
- [9] Concetto Spampinato, Simone Palazzo, Isaak Kavasidis, Daniela Giordano, Nasim Souly, and Mubarak Shah, “Deep learning human mind for automated visual classification,” in *IEEE CVPR*, 2017, pp. 6809–6817.
- [10] Florian Schroff, Dmitry Kalenichenko, and James Philbin, “Facenet: A unified embedding for face recognition and clustering,” in *IEEE CVPR*, 2015, pp. 815–823.
- [11] Qi Mao, Hsin-Ying Lee, Hung-Yu Tseng, Siwei Ma, and Ming-Hsuan Yang, “Mode seeking generative adversarial networks for diverse image synthesis,” in *IEEE/CVF CVPR*, 2019.
- [12] Shengyu Zhao, Zhijian Liu, Ji Lin, Jun-Yan Zhu, and Song Han, “Differentiable augmentation for data-efficient gan training,” *NeurIPS*, vol. 33, pp. 7559–7570, 2020.
- [13] Praveen Tirupattur, Yogesh Singh Rawat, Concetto Spampinato, and Mubarak Shah, “Thoughtviz: Visualizing human thoughts using generative adversarial network,” in *Proceedings of the 26th ACM International Conference on Multimedia*, New York, NY, USA, 2018, MM ’18, p. 950–958, ACM.
- [14] Xiao Zheng, Wanzhong Chen, Mingyang Li, Tao Zhang, Yang You, and Yun Jiang, “Decoding human brain activity with deep learning,” *Biomedical Signal Processing and Control*, vol. 56, pp. 101730, 2020.
- [15] Xiang Zhang, Xiaocong Chen, Manqing Dong, Huan Liu, Chang Ge, and Lina Yao, “Multi-task generative adversarial learning on geometrical shape reconstruction from eeg brain signals,” *arXiv preprint arXiv:1907.13351*, 2019.
- [16] Ahmed Fares, Sheng-hua Zhong, and Jianmin Jiang, “Brainmedia: A dual conditioned and lateralization supported gan (dcls-gan) towards visualization of image-evoked brain activities,” in *28th ACM MM*, 2020, pp. 1764–1772.
- [17] Simone Palazzo, Concetto Spampinato, Isaak Kavasidis, Daniela Giordano, Joseph Schmidt, and Mubarak Shah, “Decoding brain representations by multimodal learning of neural activity and visual features,” *IEEE PAMI*, vol. 43, no. 11, pp. 3833–3849, 2020.
- [18] Sanchita Khare, Rajiv Nayan Choubey, Loveleen Amar, and Venkanna Uddalapalli, “Neurovision: perceived image regeneration using crogram,” *Neural Computing and Applications*, vol. 34, no. 8, pp. 5979–5991, 2022.
- [19] Zesheng Ye, Lina Yao, Yu Zhang, and Silvia Gustin, “See what you see: Self-supervised cross-modal retrieval of visual stimuli from brain activity,” *arXiv preprint arXiv:2208.03666*, 2022.
- [20] Laurens Van der Maaten and Geoffrey Hinton, “Visualizing data using t-sne..,” *JMLR*, vol. 9, no. 11, 2008.
- [21] Pradeep Kumar, Rajkumar Saini, Partha Pratim Roy, Pawan Kumar Sahu, and Debi Prosad Dogra, “Envisioned speech recognition using eeg sensors,” *Personal and Ubiquitous Computing*, vol. 22, no. 1, pp. 185–199, Feb 2018.
- [22] Yaling Tao, Kentaro Takagi, and Kouta Nakata, “Clustering-friendly representation learning via instance discrimination and feature decorrelation,” *preprint arXiv:2106.00131*, 2021.
- [23] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, “Generative adversarial networks,” *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.
- [24] Augustus Odena, Christopher Olah, and Jonathon Shlens, “Conditional image synthesis with auxiliary classifier gans,” in *ICML*. PMLR, 2017, pp. 2642–2651.
- [25] Swaminathan Gurumurthy, Ravi Kiran Sarvadevabhatla, and R Venkatesh Babu, “Deligan: Generative adversarial networks for diverse and limited data,” in *IEEE CVPR*, 2017.
- [26] Alec Radford, Luke Metz, and Soumith Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” *preprint arXiv:1511.06434*, 2015.
- [27] Jae Hyun Lim and Jong Chul Ye, “Geometric gan,” *arXiv preprint arXiv:1705.02894*, 2017.
- [28] Teófilo Emídio De Campos, Bodla Rakesh Babu, Manik Varma, et al., “Character recognition in natural images..,” *VISAPP (2)*, vol. 7, no. 2, 2009.
- [29] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *IEEE CVPR*. Ieee, 2009, pp. 248–255.
- [30] “EPOC+ - 14 Channel EEG — emotiv.com,” <https://www.emotiv.com/epoch/>, [Accessed 14-Oct-2022].
- [31] Sepp Hochreiter and Jürgen Schmidhuber, “Long short-term memory,” *Neural computation*, vol. 9, no. 8, 1997.
- [32] Xin Jin and Jiawei Han, *K-Means Clustering*, pp. 563–564, Springer US, Boston, MA, 2010.
- [33] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen, “Improved techniques for training gans,” in *30th NeurIPS*, Red Hook, NY, USA, 2016, NIPS’16, p. 2234–2242, Curran Associates Inc.