Enhancing the Classification of EEG Signals using Wasserstein Generative Adversarial Networks

Vlad Mihai Petruţiu
Technical University of Cluj-Napoca
Cluj-Napoca, Romania
Email: petrutiuvlad@yahoo.ro

Liana Daniela Palcu
Technical University of Cluj-Napoca
Cluj-Napoca, Romania
Email: ldpalcu@gmail.com

Camelia Lemnaru
Technical University of Cluj-Napoca
Cluj-Napoca, Romania
Email: Camelia.Lemnaru@cs.utcluj.ro

Mihaela Dînşoreanu
Technical University of Cluj-Napoca
Cluj-Napoca, Romania
Email: Mihaela.Dinsoreanu@cs.utcluj.ro

Rodica Potolea
Technical University of Cluj-Napoca
Cluj-Napoca, Romania
Email: Rodica.Potolea@cs.utcluj.ro

Raul Mureşan
Transylvanian Institute of Neuroscience
Cluj-Napoca, Romania
Email: muresan@tins.ro

Vlad Vasile Moca
Transylvanian Institute of Neuroscience
Cluj-Napoca, Romania
Email: moca@tins.ro

***Abstract*—Collecting EEG signal data during a human visual recognition task is a costly and time-consuming process. However, training good classification models usually requires a large amount of quality data. We propose a data augmentation method based on Generative Adversarial Networks (GANs) to generate artificial EEG signals from existing data, in order to improve classification performance on data collected during a visual recognition task. We evaluate the quality of the artificially generated signal in terms of the accuracy of a Convolutional Neural Network-based classifier that uses both real and augmented data to classify the outcome of the cognitive task. The preliminary results suggest that the introduction of artificially generated signals have a positive effect on the performance of the classifier. Moreover, we provide a method to quantify the level of information which indicates that the generated signals indeed follow the properties of the real ones.**

I. Introduction and Problem Statement

It is widely known that Machine Learning models require a large quantity of high-quality data in order to produce state of the art results. *Neuroscience* is the domain which constitutes both a source of inspiration, but also a very challenging, yet attractive application domain for Deep Learning methods. One downside when applying Deep Learning strategies on Neuroscience data is that the data collection process is very costly and time consuming. Also, depending on the technique used during collection, the signals might be affected by external factors and thus they might contain unwanted noise. Additionally, the ability of the algorithm to generalize on this domain is generally limited because the variability of neural activity and interaction is very high among individuals. For addressing all these problems, this paper proposes the use of a data augmentation technique.

978-1-7281-9080-8/20/$31.00 ©2020 IEEE

Data augmentation has proved to be extremely useful in adding variety to the original data and multiplying the number of examples which were originally captured. The Generative Adversarial Networks (GANs) [1] were first developed in the Computer Vision domain for producing new examples which respected an initial distribution of the features, but were extremely different from the original ones. Following its initial success, the method started to be expanded and applied in other domains too.

One of the goals when analysing brain signals is to gain insight into how neurons communicate while performing a cognitive task, such as object recognition. One step in this process is to build models which try to capture the real interaction between neurons or neural areas. However, it is difficult to evaluate whether the models encode indeed the neural interaction. An indirect solution to this is to further use that information in a classification task. Additionally, the hypothesis is that - by appropriately analysing the knowledge encoded in those models - one might achieve a better understanding of how the brain processes visual information. To that end, this paper explores the possibility of using Wasserstein GANs (WGANs) [2] for generating synthetic signals for enhancing the classification results on a visual recogition task, using data from ElectroEncephaloGram (EEG) recordings.

II. State of the Art

Initially explored in the Computer Vision domain, Generative Adversarial Networks [1] have quickly replaced **Variational Auto-Encoders (VAEs) [3]** as data augmentation strategies, as they were found to provide better variety in the training data, thus boosting the performance of learning algorithms. Initially, VAEs managed to capture relevant features for low-sized images, but did not manage to focus on details, this being the main reason why an update to this method was needed. The instability that GANs initially provided was their main disadvantage, as they exhibited vanishing gradients during training, resulting in blurred and heavily noised outputs.

For solving these problems, different techniques for shortening the distance between the probability distributions of the generated against the real data were proposed. **Deep Convolutional Generative Adversarial Networks (DCGANs) [4]** managed to increase the quality of the resulting images by introducing Convolutional layers into the Discriminator's and Generator's architectures. Two variants of GANs have evolved relying mostly on the distance between the generated and real data: **MMD GAN [5]** which uses the Maximum Mean Discrepancy as the distance metric and **Wasserstein GANs (WGANs) [2]** which uses the Wasserstein distance. All of these GAN's variations have managed to overcome the instability problems and produced high-quality results in the Computer Vision domain. An increased focus started to be shifted towards generating synthetic data in the medical domain. [6] managed to improve the accuracy of a liver lesions classifier by introducing generated images using GANs. [7] explored further the potential of using Generative Adversarial Networks as data augmentation techniques concluding that a Decision Tree classifier obtained equal or even better accuracy using only generated data compared to the same classifier trained on the original data.

Regarding the generation of EEG signals, to the best of our knowledge, there have been two attempts – [8] and [9] – which tackled the problem of signals generation. [8] has used a mixture of DCGAN and WGAN, managing to generate well-shaped signals which respected the original signals' properties, while [9] has used only WGANs and obtained promising results from a visual point of view. One problem is that the generated signals were evaluated only from a visual perspective and it was not quantified whether they managed to capture the truly relevant information. From the classification perspective, to the best of our knowledge, there is no up-to-date work which has proved that artificial signals would enhance the classification results on a cognitive task.

III. Materials and Methods Used

A. Dataset Description

High-density (128 electrodes) electroencephalographic (EEG) signals were recorded @1024 samples/s from healthy human volunteers with a Biosemi ActiveTwo system. Human subjects were freely exploring visual stimuli consisting of deformed lattices of dots representing objects [visual stimuli were presented on a 22" monitor 1680x1050@120fps; distance 1.12m]. The task of the subjects was to indicate a perceptual decision by pressing one of three buttons congruent with their perception ("nothing", "uncertain", "seen"). A similar protocol was described elsewhere [10]. The protocol was approved by the Local Ethics Committee (approval 1/CE/08.01.2018). Data was collected in accordance with relevant legislation: Directive (EU) 2016/680 and Romanian Law 190/2018.

The main electrodes which have been studied were the ones corresponding to the *Occipital Lobe*. In this paper we used segments having duration of 500ms – for speed and memory efficiency. We have divided the frequency range of brain rhythms into three categories, spanning the low range (**Delta and Theta**), the intermediate range (**Alpha and Beta**), and the high range (**Gamma**). While delta-theta and alpha-beta may be expressed differently throughout the visual task, we grouped these rhythms together due to computational constraints. Nevertheless, such rhythms can coexist independently of each other in the EEG signal [11] and therefore keeping them together is not expected to harm the ability of the GAN to learn or extract information from each of these components.

Each subject has 210 trials and each trial has one of three possible labels: *Unseen*, *Uncertain* and *Seen*. The focus is towards the first and third class – as they represent the majority of the examples. The number of examples per class varies from subject to subject.

B. Data Preprocessing

As the brain connections when performing a specific task most certainly vary between subjects, the data recorded is also extremely different from subject to subject. Blinking artifacts might affect the quality of the recorded signals. For avoiding the inclusion of noisy signals, which might affect the classification process, **Principal Component Analysis (PCA)** was used. In order to not have to make a distinction between the signals as we do not know which are worth keeping and which contain noise, we will filter out all the signals for a subject. Because the sampling frequency is 1024Hz, for taking 1 second of data a vector of 1024 elements will be used, so for representing 500ms a 512 elements sized vector is taken. By applying PCA on this vector, its size will be reduced in a 2D space for being able to both compute the distance between the signals and visualize their representation. For computing the signal variability per subject, the following formula was used:

$$V = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(dist(signal(i), signal(j))), \qquad (1)$$

where **N** represents the total number of signals for a given subject and **dist** is the Euclidean Distance.

By computing the mean of all subjects' variety a threshold **T** was obtained and it will be used to eliminate the subjects which have the data too variable for training.

C. Generative Adversarial Networks

The Generative Adversarial Networks [1] consist of two Neural Networks [12] which have a similar layered architecture and are competing against each other, each one having a different goal. The first one is called the **Discriminator** and has the goal of recognizing correctly the fake against the real EEG signals. The second one is called the **Generator** and aims to produce from a randomly distributed noise a high-quality signal (i.e. the fake signal) as similar as possible to

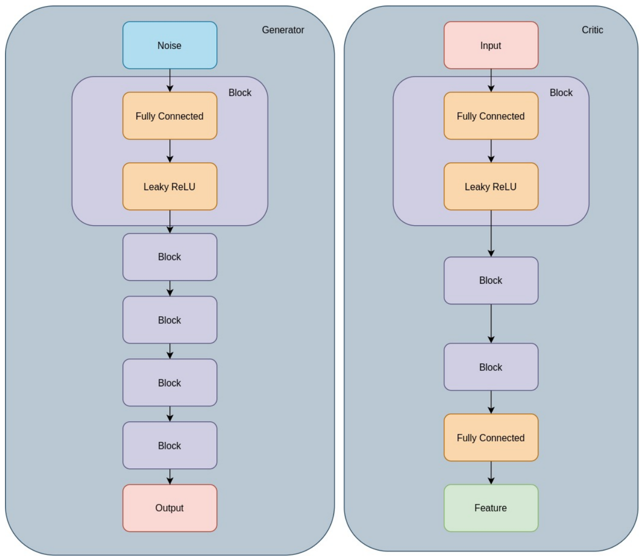

Fig. 1. Generator and Critic Architectures

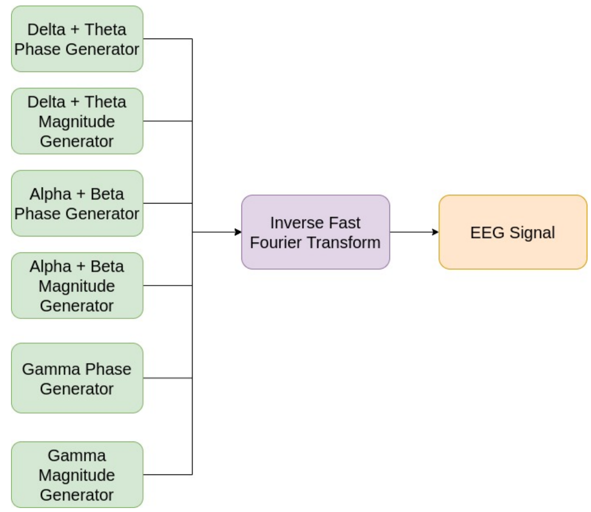

Fig. 2. Process of Reconstructing a Generated EEG Signal

TABLE I
CLASSIFIER LAYER SIZES

Layer	Output Shape	Kernel Size	Stride
Convolutional Layer 1	1 x 16	3	1
Convolutional Layer 2	16 x 32	2	2
Convolutional Layer 3	32 x 32	2	1
Convolutional Layer 4	32 x 32	2	2
Linear Layer 1	32 * 127 x 256	-	-
Linear Layer 2	256 x 1	-	-

the real signal. This results in a **minimax** game between the two Neural Networks.

Due to the Discriminator's likeliness to either predict from the beginning the correct class extremely well or not being able to recognize anything at all, early day GANs tended to collapse due to vanishing gradients for both the Generator and the Discriminator. This problem is solved in **Wasserstein GANs [2]**, providing enhanced stability during training. The main focus of this enhanced method was oriented towards bringing the probability distribution of the fake signals as close as possible to that of the real signals. The minimax problem can be stated as below:

$$\min_{G} \max_{D \in \mathcal{D}} \mathbb{E}_{x \sim \mathbb{P}_r}[D(x)] - \mathbb{E}_{\tilde{x} \sim \mathbb{P}_g}[D(\tilde{x})], \tag{2}$$

where $\mathcal{D}$ denotes a set of 1-Lipschitz functions and $\tilde{x}$ represents generator's distribution.

The architecture used in this paper for both networks consists in a chaining of *Fully Connected* and *Leaky ReLU* layers. The Discriminator has five blocks with these layers and it will output a value representing the learnt distance between the two provided signals. The lower this distance, the more the Generator will have to improve for enhancing the generated signals. On the other side, the Generator is provided with a **400** sized normally distributed noise from which it will learn to produce a new signal of size **512**. It consists of three different blocks connected at the end by a *Fully Connected* layer for producing the final signal. A visual representation of both the Discriminator and Generator architectures is represented in Fig 1.

For being able to analyze each band in particular, six different Generators were used, three for generating the phases of those bands and three for the magnitudes. By using **Inverse Fast Fourier Transform** the original signals could be reconstructed based on both generated phases and magnitudes. The whole process of generating an EEG Signal is represented in Fig. 2.

D. Classification

The classification of the signals is done by a Deep Neural Network, consisting of 4 Convolutional blocks, each block having a *Convolutional layer*, followed by a *Batch Normalization, Leaky ReLU* and a *Dropout* layer. Because we aim at classifying signals which are represented as a vector, 1D Convolutional layers were used for learning to extract the high-level features of the EEG signals and also diminishing the noise which they might possess. Dropout layers have been used to avoid overfitting by randomly deactivating neurons, forcing the important features for the classification task to be learned in a distributed manner. In this way, there will not be neurons which play a crucial role in the classification process as all of them have equally learnt important features. At the end of the pipeline there are two *Fully Connected* layers which need to combine the features learnt by the Convolutional layers in order to produce a prediction based on the given signal. The architecture of the classifier is presented in Fig. 3. The number of layers and neurons were obtained via empirical evaluations and are presented in Table I.

IV. EMPIRICAL EVALUATIONS

This section focuses on assessing whether the generated EEG signals help a classifier to better predict the real signals. Moreover, the generated signals are visually inspected in the

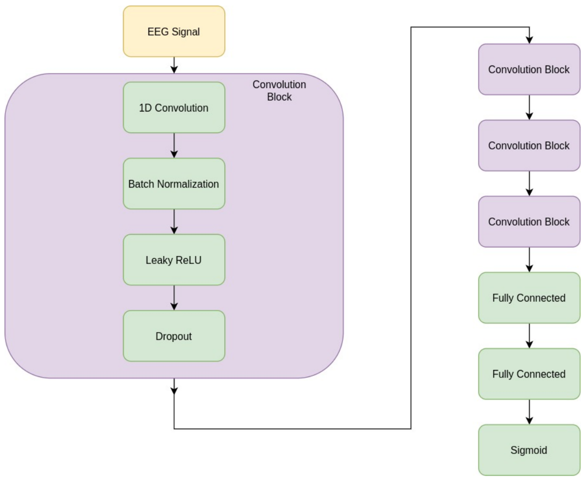

Fig. 3. Classifier Architecture

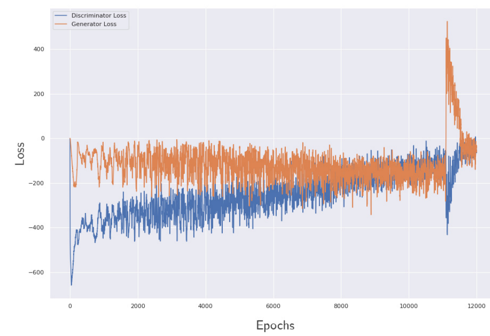

Fig. 4. Loss evolution during training on Alpha and Beta bands for Electrode A23, Seen state.

spectral domain for analyzing whether they respect the power frequencies of the real signals. Two electrodes, **A23 (which is Oz)** and **A24 (just below electrode A23)**, will be analyzed, both placed in the occipital lobe, as this is where the cognitive task is rooted. The training of the WGAN is done by using all the signals of a single recognition outcome from the filtered subjects. For evaluating the classifier, one subject out of the filtered ones will be inserted into the test set and the classifier will be trained using the rest of them. In this way, by having each subject in the test set, the ability of the classifier to generalize will be tested. This process is repeated until each subject has been in the test set. For proving the robustness of the algorithm the presented steps were repeated twelve times and the mean of the results was taken into account for comparing the results.

A. Training Considerations

The training of the Generator and Discriminator was done in 12000 epochs, all the signals for a recognition outcome being in the train dataset. The batch size used for training is 64. Every fifth batch the generator is being trained, in this way the discriminator is allowed to learn the distance metric between the fake and the real signals. The penalty term $\mathcal{E}$ was selected to be 0.001 for maintaining the Discriminator centered at 0. The optimizer which was used is **Adam** for all the discriminators and generators with a learning rate of 0.0001.

For ensuring that the networks are training properly, we plotted the losses for both the Discriminator and the Generator. As depicted in Fig. 4, initially the Discriminator is learning to separate the fake signals from the real ones and at some point the losses begin to converge, meaning that the competition between the two networks has started to balance.

On the Classification part, the Dropout layer is set at **0.5** to avoid overfitting by dropping half of each layer's neurons. At the end of these Convolutional blocks a **127** element vector will result after applying the convolutions. For predicting between the two possible signal's states, *Seen* or *Unseen*, the network ends with two *Linear* layers, the first one having a Leaky ReLU activation and the second one having *Sigmoid* for outputting a prediction.

The training is done in a *Supervised* manner with a balanced dataset for training. The used loss function is **Binary Cross-Entropy** as there are only two possible classes to differentiate between. The optimizer is **Adam** with a learning rate of 0.00001. The training is done for **200** epochs to avoid overfitting. Also, it is mandatory to keep the number of epochs constant after introducing the generated data in order to be able to make a comparison whether the generated signals helped in increasing the classifier's accuracy. The number of signals per batch is **16**.

B. Assessing the Quality of the Generated Signals

The first method for ensuring that the Generator has managed to capture relevant features of the original signal is to compare the power of each generated frequency with the power of the real frequency, this comparison being done in the spectral domain. In Fig. 5 all the Seen signals for the remained subjects were taken. The mean of the power of the frequencies was computed per subject and plotted. Fig. 5 depicts how diverse the powers are across the subjects. After that, on the same state, the mean of all the power values for each frequency of both generated and real signals were computed and compared. It can be observed in Fig. 6 that some powers of the generated signals are close to the values of the real ones, but are not overlapping as the GAN has managed to mediate between the variation which is present in the real data. The frequencies where the powers are overlapped should gain a greater focus as they were better reproduced by the GANs and might contain a higher level of information.

C. Classification Evaluation

One problem in evaluating the quality of the signals by the classification performance is that the test data set is imbalanced, so the accuracy might not be the best metric to evaluate the performance of the classifier. However, we argue that an increased value in the accuracy will prove the assumption that synthetic data helps the classifier.

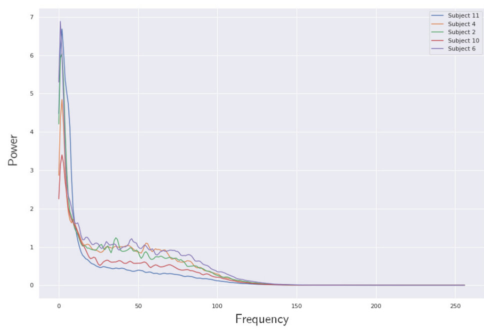

Fig. 5. Comparison of the Mean of Frequencies' Power on the Seen state for Electrode A24 for all subjects with only Real Data.

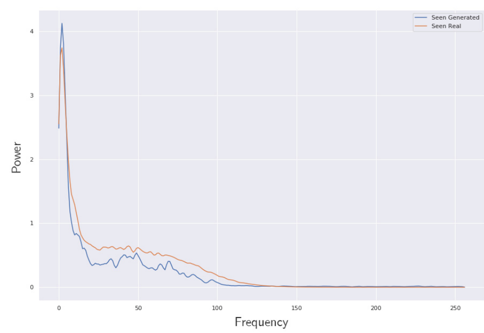

Fig. 6. Comparison of the Mean of Frequencies' Power on the Seen state for Electrode A24 for both Generated Signals and Real Ones.

Five different experiments were conducted building on top of each other:

- Training only on raw signals.
- Inserting generated signals which contain only **Delta + Theta** bands.
- Inserting generated signals which contain only **Alpha + Beta** bands.
- Inserting generated signals which contain only **Gamma** band.
- Inserting generated signals which contain all bands.

Table II and Table III show the results of the classification process on two different electrodes placed in the Occipital Lobe. The rows of the tables represent the five experiments which were previously presented and each column shows the result of the classification when that subject was in the test set. It can be seen that by adding generated signals the classifiers' predictions are more accurate. The biggest increase in the accuracy occurs when only *Gamma* band was used for reconstructing the signals which might indicate that the highest quantity of information is present in this band. For some subjects there is no distinguishable separation between the two states and this emerges in the discrepancy between the accuracies from subject to subject. The two electrodes were chosen because they reflect areas that play an important role in

TABLE II
ACCURACIES ON ELECTRODE A23

	S 2	S 4	S 6	S 9	S 10	S 11
Raw	66.5%	82.5%	45%	40%	65%	**55%**
Theta + Delta	63%	81%	**50%**	36.5%	65%	53.5%
Alpha + Beta	70%	81%	45%	40%	**67.5%**	**55%**
Gamma	**74%**	**90.5%**	47.5%	**44%**	55%	**55%**
Entire Signal	67%	81%	48%	42%	61.5%	**55%**

TABLE III
ACCURACIES ON ELECTRODE A24

	S 2	S 4	S 6	S 10	S 11
Raw	70%	80%	49%	52%	**57.5%**
Theta + Delta	62.5%	81%	**62%**	60%	56%
Alpha + Beta	70%	77.5%	55%	**64%**	**57.5%**
Gamma	**75%**	**87.5%**	55%	55%	56.5%
Entire Signal	62.5%	80%	60%	55%	55%

the human image processing process, being part of the *primary visual area*. Improvements in the classification process were only seen on electrodes which are known to be important in the cognitive task performed by the human brain.

V. Discussion

In this paper we explored the possibility of generating artificial signals which capture the relevant properties of the real ones. The quality of the generated signals was tested on a classification task, observing whether the introduction of EEG signals, produced by a Generative Adversarial Network model, boosts the performance of the classifier. By using a simple WGAN architecture, we have shown significant improvements in the classification accuracy. From a Neuroscience perspective, the next step in evaluating the benefits which GANs might bring in this domain could be focused around building functional networks from the synthetic signals.

The evaluations performed aimed to assess the ability of the classifier to generalize from subject to subject and its robustness. The high variations in the results are due to low activation for some subjects of the electrodes while performing the recognition task, not providing a clear enough separation between the different class instances. Also, the simple architecture used for classification might also affect the classification results, but the goal was to study the performance increase when introducing generated signals, and this increase is clearly visible in the results.

VI. Conclusion

Data augmentation has become an essential practice when training deep models, due to its role in improving model generalization by increasing the training data variability and volume. One application domain which would greatly benefit from an effective augmentation strategy is *Neuroscience*, where data collection is slow and expensive, and the available recording means are prone to noise. In this paper we explored the idea of using WGANs to generate artificial EEG signals with the aim of improving classification accuracy. We measured the quality of generated signals in the spectral domain and have shown that they help a classifier in better predicting between two classes.

We assessed the effectiveness of the method by evaluating the gain in classification performance and estimating the properties of the generated signals by evaluating them in the spectral domain. The results indicate that WGAN represents a reliable source for producing high-quality EEG signals and the Neuroscience domain might benefit from this augmentation strategy.

ACKNOWLEDGEMENT

This work was supported by a grant from the Romanian National Authority for Scientific Research and Innovation CNCS-UEFISCDI (project COFUND-NEURON-NMDARPSY) and one National Science Foundation grant (number NSF-IOS-1656830) funded by the US Government.

REFERENCES

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” *Advances in Neural Information Processing Systems*, vol. 3, 06 2014.

[2] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein gan,” 01 2017.

[3] D. Kingma and M. Welling, “Auto-encoding variational bayes,” 12 2014.

[4] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” 11 2016.

[5] C.-L. Li, W.-C. Chang, Y. Cheng, Y. Yang, and B. Poczos, “Mmd gan: Towards deeper understanding of moment matching network,” 05 2017.

[6] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, “Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification,” *Neurocomputing*, vol. 321, pp. 321 – 331, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231218310749

[7] F. H. K. dos Santos Tanaka and C. Aranha, “Data augmentation using gans,” *CoRR*, vol. abs/1904.09135, 2019. [Online]. Available: http://arxiv.org/abs/1904.09135

[8] K. Hartmann, R. Schirrmeister, and T. Ball, “Eeg-gan: Generative adversarial networks for electroencephalograhic (eeg) brain signals,” 06 2018.

[9] F. Fahimi, Z. Zhang, W. Goh, K. Ang, and C. Guan, “Towards eeg generation using gans for bci applications,” 05 2019, pp. 1–4.

[10] V. Moca, I. Tincas, L. Melloni, and R. Mureşan, “Visual exploration and object recognition by lattice deformation,” *PloS one*, vol. 6, p. e22831, 07 2011.

[11] G. Buzsáki, *Rhythms of The Brain*, 01 2009, pp. xiv, 448 p.

[12] L. K. Hansen and P. Salamon, “Neural network ensembles,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 10, pp. 993–1001, 1990.