EZSL-GAN: EEG-based Zero-Shot Learning approach using a Generative Adversarial Network

Sunhee Hwang[1], Kibeom Hong[1], Guiyoung Son[2], Hyeran Byun[1,2,*]
[1]*Department of Computer Science, Yonsei University, Seoul, Korea*
[2]*The Graduate Program in Cognitive Science, Yonsei University, Seoul, Korea*
sunny16@yonsei.ac.kr, cha2068@yonsei.ac.kr, sgy1017@yonsei.ac.kr, hrbyun@yonsei.ac.kr

***Abstract*—Recent studies show that deep neural network can be effective for learning EEG-based classification network. In particular, Recurrent Neural Networks (RNN) show competitive performance to learn the sequential information of the EEG signals. However, none of the previous approaches considers recognizing the unknown EEG signals which have never been seen in the training dataset. In this paper, we first propose a new scheme for Zero-Shot EEG signal classification. Our EZSL-GAN has three parts. The first part is an EEG encoder network that generates 128-dim of EEG features using a Gated Recurrent Unit (GRU). The second part is a Generative Adversarial Network (GAN) that can tackle the problem for recognizing unknown EEG labels with a knowledge base. The third part is a simple classification network to learn unseen EEG signals from the fake EEG features which are generated from the learned Generator. We evaluate our method on the EEG dataset evoked from 40 classes visual object stimuli. The experimental results show that our EEG encoder achieves an accuracy of 95.89%. Furthermore, our Zero-Shot EEG classification method reached an accuracy of 39.65% for the ten untrained EEG classes. Our experiments demonstrate that unseen EEG labels can be recognized by the knowledge base.**

***Index Terms*—EEG, Zero-Shot Learning, Recurrent Neural Network, Conditional Wasserstein Generative Adversarial Network**

I. Introduction

Electroencephalography (EEG) classification has aroused research interest in Brain Computer Interaction (BCI) and Human Machine Interaction (HMI) [1]–[3] for several decades. EEG signals are useful for recognizing human inside, since the signals are noninvasive and immediately responsive to external stimuli. Besides, EEG signals are relatively easy to collect than functional Magnetic Resonance Imaging (fMRI), Positron Emission Tomography (PET), or Magnetoencephalography (MEG).

Recently, several studies for EEG classification have shown superior performance using Deep Learning Approach as Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNN) [7]–[11]. While, deep network requires large amount of the dataset to train the model. To address this problem there are few attempts to generate EEG dataset by data augmentation methods [4]–[6].

This work was supported by Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korea government (No. 2017-0-00451, Development of BCI based Brain and Cognitive Computing Technology for Recognizing Users Intentions using Deep Learning).

[*]Corresponding author.

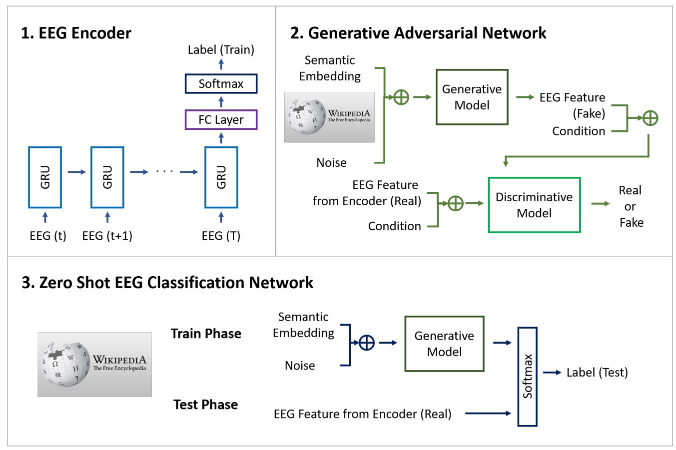

Fig. 1. Overview of the proposed approach. The proposed EZSL-GAN include three parts: 1) GRU-based EEG Encoder; 2) CWGAN; 3) Zero-Shot Classification Network. 1) and 2) networks are trained by seen dataset. The network 3) is trained by test labels which are not used in the 1) or 2). For the unseen data classification, the learned network 3) is used.

However, none of the prior works have attempted to classify unseen EEG labels which means that the labels are not included in the training dataset. In this paper, we first propose a new method for EEG Zero-Shot classification by learning Generative Adversarial Network (GAN) [12]. Our method is composed with three models as follows. 1) Gated Recurrent Unit (GRU) based EEG encoder [14]. 2) Conditioned Wasserstein GAN (CWAGAN) [13] for EEG feature Generation from knowledge base. 3) Zero-Shot classification network. (See figure 1.)

We briefly summerize our main contributions in this paper as follows:

- We achieved the state-of-the art result of 40 class EEG classification using a Gated Recurrent Unit (GRU).
- We first propose a Zero-Shot EEG Classification framework using Generative Adversarial Network.
- Our Zero-Shot Learning (ZSL) model can be learned without attribute dataset.

II. Related Work

In this section we briefly review recent works for EEG classification, Generative Adversarial Network and Zero-Shot Learning.

TABLE I
VARIOUS STUDIES FOR EEG CLASSIFICATION. SECOND TO FORTH ROW INDICATE TYPES OF USED DEEP NETWORKS, NUMBER OF EEG CLASSES, AND BEST ACCURACY.

Method	Deep Network	Number of Labels	Best Accuracy
[7]	CNN, LSTM	4 Cognitive Levels	0.91
[8]	CNN	12 Songs	0.28
[9]	LSTM	40 Visual Stimuli	0.93
[10]	RNN	3 Auditory Stimuli	0.83
[11]	LSTM	5 Motor Imagery Tasks	0.96

A. EEG Classification

Most of the previous EEG classification methods depends on Support Vector Machine (SVM), Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), or etc [1], [2]. However, several approaches for EEG classification is based on deep learning as shown Table I. In [7], they recognize four cognitive levels of EEG dataset containing 64 electrodes using a Convolutional Neural Network (CNN) and Long-Short Term Memory (LSTM) network [15]. [7] also learned CNN to recognize EEG dataset evoked by twelve different class music stimuli. More recently, several works show meaningful result using RNN [7], [9]. Above all, [9] has attempted to recognize large labels and achieved 92.9%. They also provide the EEG dataset consist of 40-classes.

To design our EEG encoder, we follow the previous work for EEG classification [9]. Also, we use a GRU which is a cell structure that reduces computational complexity while maintaining the advantages of LSTM.

B. Zero-Shot Learning

Zero-Shot Learning (ZSL) has received increasing attention for a decade. Most of previous ZSL works are based on embedding the source data into target semantic space which includes both seen and unssen class information. However, the dataset of existing ZSL studies for human inside are limited only fMRI or ECoG. That saids, there is none of the prior work for EEG-based ZSL. Besides, the existing methods require attribute dataset for embedding the semantic information [16], [17].

Unlike previous work, we propose a ZSL method that uses only a knowledge base without attribute dataset. We take word2vec [18] as our semantic information.

C. Generative Adversarial Network

Deep generative models have shown promising results for computer vision tasks recently. Among the generative models, GAN is the most popular model in various research fields as image generation, audio generation, or language translation. Moreover, there are few works to generate EEG features [4], [5]. In EEG-GAN [4], naturalistic EEG signals are generated by the proposed GAN framework. Moreover, [5] proposed EEG data augmentation GAN. They also have shown significant improvements in the emotion recognition accuracy by adding the generated data. In the two works mentioned above for EEG data generation, Wasserstein GAN (WGAN) is adopted for their generative model.

Inspired by prior works [4], [5], we adopt WGAN framework which optimizes the Wasserstein distance to produce more realistic samples than the original GAN. We generate EEG features using WGAN with a knowledge base which means we take the knowledge base as the condition for our WGAN to generate unlabeled EEG features.

TABLE II
SYMBOL DEFINITION

Symbol	Description
h_t	hidden state value at time t
$\widetilde{h_t}$	candidate hidden state
b	bias
$\odot$	element-wise multiplication

III. PROPOSED METHOD

We now introduce the details of our EZSL-GAN. We first define the ZSL problem to solve. Let $X_{tr} = \{x_1, x_2, x_3, ..., x_N\}$ and $Y_{tr} = \{y_1, y_2, y_3, ..., y_N\}$ are the training samples and corresponding labels $L_{tr} = \{l_1, l_2, l_3, ..., l_M\}$, where the test samples X_{te} and Y_{te} containing the disjoint class label $L_{te} = \{l_1, l_2, l_3, ..., l_K\}$; $(L_{tr} \cap L_{te} = \emptyset)$. Given X and Y, the task of ZSL is to learn our EZSL-GAN using only X_{tr} and Y_{tr}. In the test phase, we evaluate our model using X_{te} and Y_{te}. Our method consists of following three process as figure 1.

A. EEG Encoder

The first step of our model is encoding EEG signals using the training dataset. We take a GRU for the encoder, which is a simplified structure of LSTM and it employs update gate $z_{(t)}$ and reset gate $r_{(t)}$. The update gate determines how much of the previous memory information is retained to calculate the new state and the reset gate determines how the input data is combined with the previous memory.

Where, the symbols above are described in Table II. The output of our GRU layer fed into a Fully Connected (FC) layer with a Leaky ReLU activation function. To learn the model, we optimize following loss function:

$$L_{ENC} = -\sum_{i=1}^{N} y_i log(Softmax(GRU(x_i))) \qquad (1)$$

B. Generative Adversarial Network

After the learning the Encoder module, the next is training the proposed Generative model using the training dataset. Inspired by [4], [5], we develop conditional WGAN with a Gradient Penalty (GP) term. While, word2vec [18] is conducted as the condition of our network as Figure 2. Both Generator and Discriminator are consisted of three FC layers with Leaky ReLU.

Given a feature vector and corresponding word vector, our GAN tries to generate fake features $F_f = \{f_1, f_2, f_3, ..., f_N\}$

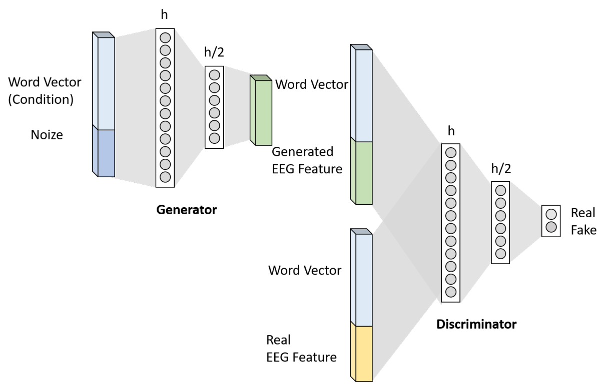

Fig. 2. The proposed CWGAN framework.

similar to real features $F_r = \{f_1, f_2, f_3, ..., f_N\}$. While F_r are extracted from the learned Encoder, the output of the last FC layer, and the word vector s_{y_n} indicate semantically embedded feature by the class name. The proposed GAN has two parts: Generator and Discriminator. We hope to generate EEG features as realistic data from word vector by adversarial training. Our CWGAN is learned by optimizing the following loss function:

$$\underset{G}{min}\underset{D}{max} L_{GAN} = E[D(f_{n(r)}, s_{y_n})] - E[D(f_{n(f)}, s_{y_n})] \\ -\lambda E[(|| \bigtriangledown_{f_{n(f)}} D(f_{n(f)}, s_{y_n})||_2 - 1)^2] \quad (2)$$

C. Zero-Shot Classification

The objective of EEG-based Zero-Shot classification is recognizing the labels of EEG features $F_r = \{f_1, f_2, f_3, ..., f_K\}$. We propose the method that training the classification network without training the real EEG features.

During training the proposed GAN, we also train the ZSL classification network at each iteration for unseen class classification. To build the training dataset of the ZSL classification network, embedded word features $S = \{s_1, s_2, s_3, ..., s_K\}$ of unseen classes are used to generate fake EEG features.

Given generated EEG feature $F_f = \{f_1, f_2, f_3, ..., f_K\}$, we train the network to classify the labels within the test set never have been used in the Encoder and GAN. The ZSL classification network is a simple single FC layer. For the unseen class classification, real EEG features F_r in the test dataset are used. We optimize the ZSL model using following loss function:

$$L_{ZSL} = -\sum_{i=1}^{K} y_i log(Softmax(FC(f_i))) \quad (3)$$

IV. Experimental Setting

A. Dataset

Our method is designed to classify subject-independant unseen EEG labels. As described above, [9] released well-designed large EEG dataset with the a lot of labels. We use the dataset for our experiment. Those EEG data are recorded from six subjects. During recording the dataset, participants watched 50 images of 40 different ImageNet classes. The dataset includes 128-Channels of EEG signals. We presented the details of the dataset in the Table III. The samples of visual stimuli are presented in figure 3.

TABLE III

Details of the EEG Dataset. We use the provided split 0

Stimuli	Visual Object
Number of Class	40
Number of Subject	6
Number of Images per Class	50
Watching Time for each Image	0.5s
Total Number of Available Samples	11,965
Number of Training Data	30
Number of Test Data	10

Fig. 3. Visual stimuli for EEG recording. Top left: mailbag, Top right: espresso maker, Bottom left: grand piano, Bottom right: banana

To learn the proposed network, we randomly split the training and test dataset. The selected test labels are as follows: espresso maker, revolver, german shepherd, broom, pizza, capuchin, banana, digital watch, airliner, radio telescope. We train our model without the test dataset.

B. Implementation Details

For the training three different networks of our work, we use the adam gradient descent as the optimization method. The learning rate of the Encoder, Generator, Discriminator, and Zero-Shot classification network is initialized to 0.001, 0.002, 0.001, and 0.001. Batch size of all the networks is 128. Hidden dimension of the Encoder is [64,128], while hidden dimension of both Generator and Discriminator are set to 128. We implemented our method using Pytorch v0.4.1 [20].

V. Experimental Results

A. GRU-based EEG Encoder

We evaluate our Encoder using the training and test dataset provided by [9]. For the evaluation of our Encoder, we used all the labels during training and test. The number of train, validataion and test dataset is 7,970, 1,998, and 1,997. While we split the labels for the Zero-Shot Learning framework. As [9] described in their work, we compared the result of two visualization time. One is the whole time with zero padding as [0-512], and another visualization time of EEG signals is selected as [320-480] which has shown the best accuracy in [9]. We also compared our GRU-based method with LSTM

TABLE IV
THE RESULT OF OUR PROPOSED EEG ENCODER. HD INDICATE THE SIZE OF HIDDEN DIMENSION. VA AND TA INDICATE VALIDATION ACCURACY AND TEST ACCURACY.

sequence	HD	RNN	Max VA	TA at Max VA
320-480	64	LSTM	92.94	92.14
320-480	64	GRU	92.99	92.49
320-480	128	LSTM	95.65	95.54
320-480	128	GRU	**96.30**	**95.89**
1-512 (w zero-padding)	64	LSTM	83.13	82.52
1-512 (w zero-padding)	64	GRU	83.18	83.93

TABLE V
THE RESULT OF OUR PROPOSED EEG ENCODER. HD INDICATE THE SIZE OF HIDDEN DIMENSION. VA AND TA INDICATE VALIDATION ACCURACY AND TEST ACCURACY.

Method	Accuracy
SAE (EEG to Word Vector) [19]	17.30
SAE (Word Vector to EEG) [19]	16.90
Proposed Method (CWGAN)	**39.65**

unit. The experimental result is shown in Table IV. Our method achieved the accuracy of 95.89% (split 0) for 40 classes EEG classification which is the state-of-the-art result over the prior work.

B. Zero-Shot Classification

We also evaluate our Zero-Shot Classification performance after split the dataset. Training data is used to train only Encoder and our Generative model. For the test, the labels are used to generate EEG features by learned generative model, and we train the Zero-Shot Classification network using the fake EEG features. The test accuracy of ZSL is measured by following equation:

$$Accuracy_{zsl} = \frac{\#Correct}{\#TestEEGSamples} \tag{4}$$

The result of our ZSL accuracy is presented in Table V. We compared with one of the most popular method for Zero-Shot Image classification [19]. As shown in the table, our method achieved relatively promising result.

VI. DISCUSSION AND CONCLUSION

In this paper, we first propose EEG-based Zero-Shot classification approach which is based on Conditional Wasserstein Generative Adversarial Network. Our method consists of three parts: (1) EEG Encoder (2) GAN (3) ZSL. The ZSL model can be trained attribute dataset is not available and each part of the method demonstrates that we achieved meaningful results on public benchmark dataset.

Although we focus on EEG-based classification in our work, the proposed approach can be generalized to various research fields of BCI. For the classification any brain mapping in humans using deep learning, large amount of dataset is required. However, the proposed methods can be learned by existing dataset, and test dataset can be used only for the test phase.

REFERENCES

[1] A. Stewart, A. Nuthmann, and G. Sanguinetti, "Single-trial classification of EEG in a visual object task using ICA and machine learning.", Journal of neuroscience methods, 228, 1-14., 2014.

[2] B. Kaneshiro, M. Guimaraes, H. Kim, A. Norcia, and P. Suppes, "A representational similarity analysis of the dynamics of object processing using single-trial EEG classification.", Plos one, 10(8), e0135697. 2015.

[3] S. Walter, C. Wendt, J. Bhnke, S. Crawcour, J. Tan, A. Chan, ... , and H. Traue., "Similarities and differences of emotions in humanmachine and humanhuman interactions: what kind of emotions are relevant for future companion systems?." Ergonomics 57.3 : 374-386., 2014.

[4] K. Gregor Hartmann, R. Tibor Schirrmeister, and T. Ball., "EEG-GAN: Generative adversarial networks for electroencephalograhic (EEG) brain signals.", arXiv preprint arXiv:1806.01875, 2018.

[5] Y. Luo, B. Lu, "EEG Data Augmentation for Emotion Recognition Using a Conditional Wasserstein GAN", 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2018.

[6] M. Krell, and S. Kim., (2017, July). "Rotational data augmentation for electroencephalographic data.", In Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE (pp. 471-474). IEEE. 2017.

[7] P. Bashivan, I. Rish, M. Yeasin, and N. Codella., "Learning representations from EEG with deep recurrent-convolutional neural networks.", International Conference on Learning Representations (ICLR), 2016.

[8] S. Stober, A. Sternin, A. Owen, and J. Grahn, "Deep feature learning for EEG recordings.", International Conference on Learning Representations (ICLR), 2016.

[9] C. Spampinato, S. Palazzo, I. Kavasidis, D. Giordano, N. Souly, and M. Shah. (2017, July). "Deep learning human mind for automated visual classification.", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 6809-6817). 2017.

[10] M. Moinnereau, T. Brienne, S. Brodeur, J. Rouat, K. Whittingstall, and E. Plourde., (2018). "Classification of auditory stimuli from EEG signals with a regulated recurrent neural network reservoir.", arXiv preprint arXiv:1804.10322. 2018.

[11] X. Zhang, L. Yao, Q. Sheng, S. Kanhere, T. Gu, and D. Zhang., (2018, March). "Converting your thoughts to texts: Enabling brain typing via deep feature learning of eeg signals.", In 2018 IEEE International Conference on Pervasive Computing and Communications (PerCom) (pp. 1-10). IEEE. 2018.

[12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, ... , and Y. Bengio, (2014). "Generative adversarial nets.", In Advances in neural information processing systems (pp. 2672-2680). 2014.

[13] M. Arjovsky, S. Chintala, and L. Bottou, (2017, July). "Wasserstein generative adversarial networks.", In International Conference on Machine Learning (pp. 214-223). 2017.

[14] K. Cho, B. Van Merrinboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio., (2014). "Learning phrase representations using RNN encoder-decoder for statistical machine translation.", arXiv preprint arXiv:1406.1078. 2014.

[15] S. Hochreiter, and J. Schmidhuber, (1997). "Long short-term memory.", Neural computation, 9(8), 1735-1780. 1997.

[16] C. Caceres, M. Roos, K. Rupp, G. Milsap, N. Crone, M. Wolmetz, and C. Ratto, (2017). "Feature Selection Methods for Zero-Shot Learning of Neural Activity.", Frontiers in neuroinformatics, 11, 41. 2017.

[17] K. Rupp, M. Roos, G. Milsap, C. Caceres, C. Ratto, M. Chevillet, ... , and M. Wolmetz. (2017). "Semantic attributes are encoded in human electrocorticographic signals during visual object recognition." NeuroImage, 148, 318-329. 2017.

[18] T. Mikolov, K. Chen, G. Corrado, and J. Dean. (2013). "Efficient estimation of word representations in vector space.", arXiv preprint arXiv:1301.3781. 2013.

[19] E. Kodirov, T. Xiang, and S. Gong., (2017). "Semantic autoencoder for zero-shot learning.", arXiv preprint arXiv:1704.08345. 2017.

[20] P. Adam, G. Sam, C. Soumith, C. Gregory, Y. Edward, D. Zachary, L. Zeming, D. Alban, A. Luca and L. Adam., (2017). "Automatic differentiation in PyTorch.", NIPS-W. 2017.