EEG Data Augmentation for Emotion Recognition Using a Conditional Wasserstein GAN

Yun Luo, Bao-Liang Lu* *Senior Member, IEEE*

***Abstract*— Due to the lack of electroencephalography (EEG) data, it is hard to build an emotion recognition model with high accuracy from EEG signals using machine learning approach. Inspired by generative adversarial networks (GANs), we introduce a Conditional Wasserstein GAN (CWGAN) framework for EEG data augmentation to enhance EEG-based emotion recognition. A Wasserstein GAN with gradient penalty is adopted to generate realistic-like EEG data in differential entropy (DE) form. Three indicators are used to judge the qualities of the generated high-dimensional EEG data, and only high quality data are appended to supplement the data manifold, which leads to better classification of different emotions. We evaluate the proposed CWGAN framework on two public EEG datasets for emotion recognition, namely SEED and DEAP. The experimental results demonstrate that using the EEG data generated by CWGAN significantly improves the accuracies of emotion recognition models.**

I. INTRODUCTION

Affective brain-computer interfaces (aBCIs) that aim to equip machines with emotional intelligence, have recently attracted widespread attention [1]. Many researchers have studied EEG-based emotion recognition methods and have made significant progresses [2][3]. However, compared with images and speech signals, the high price of EEG acquisition devices makes it hard to collect large-scale EEG data. Besides, it is difficult to acquire sufficient labeled data from subjects in EEG-based emotion recognition experiments. The highly cost nature of these experiments directly results in the lack of EEG data, which hinders the classifier performance.

Data augmentation is one of promising ways to solve the data scarcity problem. This approach generates the realistic-like data by applying a transformation to the real data [4]. It is common to generate artificial images through geometric transformation (translation, rotation, scaling, horizontal shearing) [5] in the field of computer vision. The similar technique has also been applied to EEG-based tasks. Krell and Su proposed rotational distortions, which were similar to affine/rotational distortions of images, to generate EEG data [6]. In [7], artificial EEG trials were generated by the relevant combinations and distortions of the original trials.

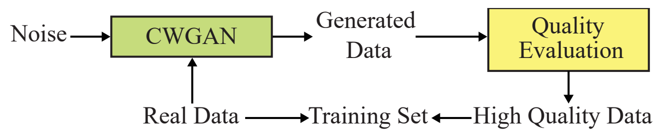

Fig. 1. The framework for EEG data augmentation.

Instead of generating data through geometric transformation, very recent generative adversarial networks (GANs) [8] have revealed their potential in generating realistic-like data such as images [9] by adopting an adversarial training. Typical GANs are formulated as a two-player game which consists of two adversarial networks called generator and discriminator. The distribution of the generated data is approximate to the real data when the game achieves its equilibrium. In computer vision, GANs-based methods have been successfully adopted to generate realistic-like images, and the performance of classifiers has been enhanced significantly by appending the generated images [10][11]. However, to the best of our knowledge, this technique has not been investigated in EEG-based emotion recognition.

In this work, we introduce a Conditional Wasserstein GAN (CWGAN) framework for EEG data augmentation to enhance EEG-based emotion recognition. A gradient-penalty version of Wasserstein GAN is adopted to generate artificial EEG data in differential entropy (DE) form from noise distribution [12][13]. Note that it is common for classifiers to handle the high-level features of EEG data, and DE features have been demonstrated to be suited for EEG-based emotion recognition tasks [14]. Therefore, CWGAN is used to generate DE data instead of raw EEG data. Besides, an auxiliary conditional information is appended to generate multiple emotion categories. The qualities of generated data are evaluated by using three indicators and only high quality data are appended to the training set. The framework consists of two parts, namely Conditional Wasserstein GAN and quality evaluation as illustrated in Fig. 1.

Compared with signal-level transformation through distortions [6][7], CWGAN learns the representation of the real distribution in a deeper level, which leads to better classification of different emotions. However, due to the instability problems of traditional GANs' training procedure, the generated data sometimes have poor qualities. Unlike visible images, it is intractable for human to judge the qualities of the generated high-dimensional EEG data in DE

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFB1002501, the National Natural Science Foundation of China under Grant 61673266, the Major Basic Research Program of Shanghai Science and Technology Committee under Grant 15JC1400103, the ZBYYMOE Joint Funding under Grant 6141A02022604, and the Fundamental Research Funds for the Central Universities. Yun Luo and Bao-Liang Lu are with the Center for Brain-Like Computing and Machine Intelligence, Department of Computer Science and Engineering, the Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering, and the Brain Science and Technology Research Center, Shanghai Jiao Tong University, 800 Dong Chuan Road, Shanghai 200240, China. Corresponding author: Bao-Liang Lu (bllu@sjtu.edu.cn).

form directly. To overcome the instability problem of training traditional GAN, Wasserstein GAN is used in this work. To evaluate the qualities of the generated data for user's reference, three indicators are used to evaluate the qualities as well. Our experimental results on two public EEG datasets indicate that the proposed CWGAN framework significantly improves the accuracies of emotion recognition models by appending generated EEG data to the training sets.

II. EXPERIMENTS

A. EEG Datasets

The data in SEED[1] [14] dataset were formed with 62-channel EEG signals. 15 participants watched 15 emotional film clips to elicit three emotions: positive, neutral, and negative. They took part in the experiments for three times with an interval of about one week. So there were totally 45 experiments. The original EEG signals were recorded at a sampling rate of 1000 Hz with ESI NeuroScan System. The experimental procedures involving human subjects described in this experiment were approved by the Ethics Committee.

The DEAP[2] [15] dataset consisted of 32 participants. Each of them was required to watch 40 music videos and the level of each video was rated 1-9 by the participants in terms of arousal, valence, like, and dislike. So there were totally 32 experiments where 32-channel EEG signals were recorded by an international 10-20 system. Besides, 8-channel peripheral physiological signals were recorded as well.

B. Data Preprocessing

The EEG data of both datasets are preprocessed and the DE features are extracted. For SEED dataset, DE features are extracted per second from five frequency bands: δ: 1-3 Hz, θ: 4-7 Hz, α: 8-13 Hz, β: 14-30 Hz, and γ: 31-50 Hz [14]. The input data of each subject thus have 310 dimensions (62 channels $\times$ 5 frequency bands). The number of sample with label for each subject is 3394. For DEAP dataset, DE features are extracted per second except for δ frequency due to the fact that the low frequency band is filtered in this dataset. The input data of each subject thus have 128 dimensions (32 channels $\times$ 4 frequency bands). The number of samples with label for each subject is 2400. Here, we adopt two emotion models using valence value (high valence: level $>$ 5, low valence: level $\leq$ 5) and arousal value (high arousal: level $>$ 5, low arousal: level $\leq$ 5), respectively.

III. METHODS

A. Conditional Wasserstein GAN

GANs consist of two competing components which are both parameterized as deep neural networks. Given real data distribution X_r and generated data distribution X_g, the generator G produces realistic-like X_g to confuse the discriminator D, while the discriminator D tries to distinguish whether a sample comes from X_r or X_g. The adversarial training procedure can be formulated as a minimax problem:

$$\begin{aligned}\min_{\theta_G}\max_{\theta_D} L(X_r, X_g) = \mathbb{E}_{x_r\sim X_r}[\log(D(x_r))] \\ + \mathbb{E}_{x_g\sim X_g}[\log(1-D(x_g))]\end{aligned} \tag{1}$$

where θ_g and θ_d represent the parameters of the generator and discriminator, respectively. And X_g is implicitly defined by $x_g = G(x_z)$, where x_z is sampled from uniform or Gaussian noise distribution.

The adversarial training of traditional GANs can be formalized with minimizing the Jensen-Shannon divergence between the probability distribution of real data and generated data. However, the discontinuity of Jensen-Shannon divergence makes it hard to provide useful gradient for optimizing generator, which is one of the main causes of GANs instability. To eliminate this issue, Jensen-Shannon divergence is replaced with the Earth-Mover distance (EMD, also called Wasserstein-1) in Wasserstein GANs [12]:

$$W(X_r, X_g) = \inf_{\gamma\sim\Pi(X_r,X_g)} \mathbb{E}_{(x_r,x_g)\sim\gamma}[||x_r - x_g||] \tag{2}$$

where $\Pi(X_r, X_g)$ denotes all possible joint distributions of real distribution X_r and generated distribution X_g defined in traditional GANs.

The EMD is continuous and differentiable almost everywhere, and thus can provide meaningful gradients for optimizing generator, which ensures the convergence of the GANs. Since it is difficult to implement the infimum of Eq.(2) in reality, an alternative approach is to apply Kantorovich-Rubinstein duality of EMD:

$$\begin{aligned}W(X_r, X_g) = \frac{1}{K}\sup_{||f||_L\leq K} \mathbb{E}_{x_r\sim X_r}[f(x_r)] \\ -\mathbb{E}_{x_g\sim X_g}[f(x_g)]\end{aligned} \tag{3}$$

where f denotes the set of 1-Lipschitz functions.

In realistic implementations, f is replaced by discriminator D and $||f||_L \leq K$ is replaced by $||D||_L \leq 1$. In order to make the training procedure more stable and make convergence faster, Gulrajani *et al.* enforced Lipschitz constraint with gradient penalty instead of weight clipping to directly constrain the gradient norm [13]. In their approach, an extra penalty term is appended to the loss function:

$$\begin{aligned}\min_{\theta_G}\max_{\theta_D} L(X_r, X_g) = \mathbb{E}_{x_r\sim X_r}[D(x_r)] \\ - \mathbb{E}_{x_g\sim X_g}[D(x_g)] \\ - \lambda\mathbb{E}_{\hat{x}\sim\hat{X}}[(||\nabla_{\hat{x}}D(\hat{x})||_2 - 1)^2]\end{aligned} \tag{4}$$

where λ is a hyperparameter controlling the trade-off between original objective and gradient penalty, and $\hat{x}$ denotes the data points sampled from the straight line between real distribution X_r and generator distribution X_g:

$$\hat{x} = \alpha x_r + (1-\alpha)x_g, \alpha\sim U[0,1], x_r\sim X_r, x_g\sim X_g \tag{5}$$

In order to generate data with multiple categories, an auxiliary label Y_r is feed into both discriminator and generator. In the generator, we concatenate X_z with Y_r. And in the discriminator, we concatenate both X_r and X_g with

[1]http://bcmi.sjtu.edu.cn/∼seed/index.html
[2]http://www.eecs.qmul.ac.uk/mmv/datasets/deap/

Y_r to construct a hidden representation, which controls the categories of generated data. Then the proposed CWGAN can be formulated by:

$$\begin{aligned}&\min_{\theta_G}\max_{\theta_D} L(X_r, X_g, Y_r) = \\ &\quad \mathbb{E}_{x_r\sim X_r, y_r\sim Y_r}[D(x_r|y_r)] - \mathbb{E}_{x_g\sim X_g, y_r\sim Y_r}[D(x_g|y_r)] \\ &\quad -\lambda \mathbb{E}_{\hat{x}\sim \hat{X}, y_r\sim Y_r}[(||\nabla_{\hat{x}|y_r} D(\hat{x}|y_r)||_2 - 1)^2]\end{aligned} \tag{6}$$

The losses of discriminator (maximum term) and generator (minimum term) are optimized in an alternating procedure. In practice, the generator loss only contains the second term of Eq.(6). And the discriminator loss is updated for *critic* times in each adversarial training iteration. The structure of the CWGAN is shown in Fig. 2.

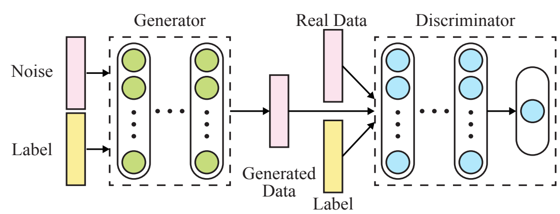

Fig. 2. Illustration of the proposed CWGAN, which consists of two parts: the generator generates realistic-like data and the discriminator distinguishes real and generated data. An auxiliary label information is concatenated to control the categories of the generated data.

B. Quality Evaluations

In computer vision, the qualities of generated images can be assessed directly by users. However, the qualities of the generated high-dimensional EEG data are impossible to be visually evaluated. Therefore, we use three indicators to evaluate the qualities of generated EEG data in this work. The generated EEG data are considered to have high qualities if their distributions are approximate to real distributions from the following three aspects:

1) Discriminator loss: The loss of discriminator represents EMD between X_r and X_g when the network converges. That is, the generated data are high qualities if the loss is approximate to 0.

2) Maximum Mean Discrepancy (MMD): MMD [16] is frequently used as a measurement of the distance between two distributions [17]. Here we adopt it to evaluate the distance between X_r and X_g from another point of view.

3) Two-dimensional mapping: The high-dimensional X_g are mapped into two-dimensions by t-SNE [18]. The distribution of X_g then can be visualized directly.

IV. EVALUATION DETAILS

In order to compare experimental results with those in [14], we use the same strategy to divide the training set and test set for each subject. For SEED dataset, the first 9 trials (2010 samples) are set to original training set and the remaining 6 trials (1384 samples) are set to test set. And three emotions (positive, negative, and neutral) are classified. For DEAP dataset, the first 24 trials (1440 samples) are set to original training set and the remaining 16 trials (960 samples) are set to test set. The arousal model and the valence model are classified separately. The qualities of the generated EEG data are evaluated and only high quality data are appended to the original training set.

To optimize the network structure, we perform grid search on the number of network layers. The numbers of layers are searched from 3 to 5 for both generator and discriminator. Each hidden layer of both generator and discriminator network has 512 nodes for SEED dataset and 256 nodes for DEAP dataset. The dimensions of auxiliary labels are 3 for SEED and 2 for DEAP. The ReLU activation function is used for all hidden layers and the networks are optimized by an Adam optimizer. The *critic* value is set to 20 and the λ value is set to 10. The noises are sampled from a uniform distribution $U[-1, 1]$. We apply an SVM with linear kernel as the classifier. The parameter c is searched from $2^{-10} \sim 2^{10}$ to find the optimal value, which is also employed in [14].

V. RESULTS

In this section, we carry out experiments to evaluate the efficiency of the proposed CWGAN framework for EEG data augmentation. The qualities of the generated data are evaluated by the three indicators and only high quality data (discriminator loss and MMD are close to 0) are appended to the training set to enlarge the dataset.

In our experiments, we find the discriminator loss quickly converges to a small value (close to 0) along with the training epoch for each subject. As EMD between the distributions of real and generated data, discriminator loss converging to 0 indicates that the two distributions are similar to each other and the generated DE data have high qualities. Besides, the MMD curve of each subject decreases to a small value (close to 0) along with the training epoch, which also implies that the generated data have high qualities.

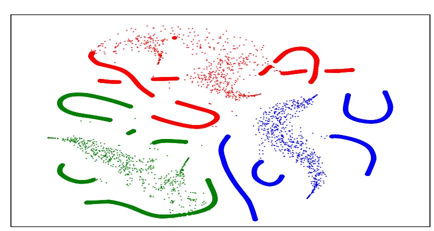

Fig. 3. Two-dimensional visualizations of the real and generated DE data of one subject in SEED dataset. Data points with red, green and blue colors represent three emotions of negative, neutral and positive, respectively. The lines represent the real data and the thin points represent the generated data.

The distributions of real and generated data are visualized by t-SNE as illustrated in Fig. 3. Data from each emotion can be clustered in the latent space and the generated data are close to the corresponding real data, which implies the generated data carry enough realistic information. Besides, the distribution of real data is sparse and the boundaries of different categories in the data manifold are not obvious. The generated data supplement the training data manifold, which leads to better margins for the classifier.

Data Appended	SEED		DEAP-Arousal		DEAP-Valence	
	Mean	Std.	Mean	Std.	Mean	Std.
0×Dataset	0.8399	**0.0972**	0.6902	0.1361	0.5376	0.1308
1×Dataset	**0.8696**	0.1272	**0.7817**	**0.0958**	**0.7389**	**0.1082**
2×Dataset	0.8659	0.1283	0.7437	0.1227	0.6543	0.1506
3×Dataset	0.8638	0.1276	0.7479	0.1133	0.6625	0.1477
4×Dataset	0.8623	0.1275	0.7450	0.1013	0.6496	0.1504
5×Dataset	0.8600	0.1294	0.7434	0.1151	0.6467	0.1372

TABLE I

PERFORMANCE OF APPENDING GENERATED EEG DATA

We evaluate the performance of SVMs when appending different numbers of the generated EEG data and show the average accuracies and standard deviations of the two datasets in Table 1. Since the performance of three indicators demonstrate that all of the generated data have high qualities, we append all generated data to the original dataset. The number of the appended data is denoted with the times of the number of real dataset. Namely, 0 (the baselines of different datasets) represents the original real dataset while 1 represents that the number of the appended data is the same with the real dataset. The baseline of SEED (no generated data appended) is referenced in [14]. We can see that appending the same number of generated EEG data to the training sets achieves the best performance, in terms of accuracy, on SEED, DEAP-Arousal, and DEAP-Valence tasks. With the increment of generated data, the accuracies of the enlarged training sets are still higher than the original datasets. This phenomenon demonstrates that appending generated data to training set improves the performance of the model.

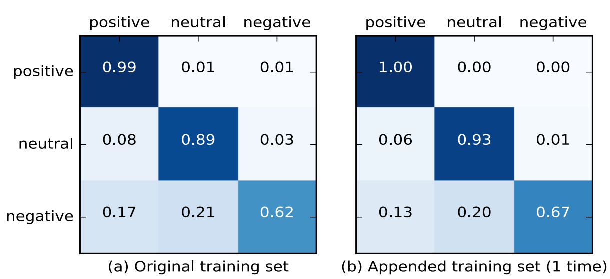

Fig. 4. The confusion matrixes (SEED dataset) trained by (a) original training set and (b) appended training set (1 time).

The confusion matrixes (SEED dataset) trained by different training sets are shown in Fig. 4. The rows of the matrixes represent real emotions while the columns represent predicted emotions. Compared with original training set, the recognition accuracies of three emotions achieve 1%, 4%, 5% improvements when appending the generated data to original training set. This phenomenon also indicates the proposed framework is able to enhance EEG-based emotion recognition by appending generated data.

VI. CONCLUSION AND FUTURE WORK

In this paper, we have proposed a Conditional Wasserstein GAN framework for EEG data augmentation to enhance EEG-based emotion recognition. The proposed framework generates realistic-like EEG data by using Wasserstein GAN gradient penalty version. An auxiliary label information is appended to Wasserstein GAN to generate different categories. The performance of our framework has been evaluated on two public EEG datasets for emotion recognition. By using three evaluation indicators, we see that high-quality EEG data are generated. The emotion recognition models trained on appending datasets achieve 2.97%, 9.15% and 20.13% improvements on SEED dataset and DEAP dataset for arousal and valence classifications, respectively. In the future, we will study more quantifiable methods to evaluate the qualities of the generated EEG data.

REFERENCES

[1] C. Mühl, B. Allison, A. Nijholt, and G. Chanel, "A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges," *Brain-Computer Interfaces*, vol. 1, no. 2, pp. 66–84, 2014.

[2] Y.-P. Lin, Y.-H. Yang, and T.-P. Jung, "Fusion of electroencephalographic dynamics and musical contents for estimating emotional responses in music listening," *Frontiers in Neuroscience*, vol. 8, 2014.

[3] Y. Lu, W.-L. Zheng, B. Li, and B.-L. Lu, "Combining eye movements and EEG to enhance emotion recognition," in *IJCAI'15*, 2015, pp. 1170–1176.

[4] D. A. V. Dyk and X. L. Meng, "The art of data augmentation," *Journal of Computational & Graphical Statistics*, vol. 10, no. 1, pp. 1–50, 2001.

[5] S. Dieleman, K. W. Willett, and J. Dambre, "Rotation-invariant convolutional neural networks for galaxy morphology prediction," *Monthly Notices of the Royal Astronomical Society*, vol. 450, no. 2, pp. 1441–1459, 2015.

[6] M. M. Krell and K. K. Su, "Rotational data augmentation for electroencephalographic data," in *Engineering in Medicine and Biology Society*, 2017, pp. 471–474.

[7] F. Lotte, "Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain-computer interfaces," *Proceedings of the IEEE*, vol. 103, no. 6, pp. 871–890, 2015.

[8] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *NIPS'14*, 2014, pp. 2672–2680.

[9] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *Computer Science*, 2015.

[10] X. Zhu, Y. Liu, Z. Qin, and J. Li, "Data augmentation in emotion classification using generative adversarial networks," *arXiv preprint arXiv:1711.00648*, 2017.

[11] A. Antoniou, A. Storkey, and H. Edwards, "Data augmentation generative adversarial networks," *arXiv preprint arXiv:1701.04340*, 2017.

[12] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," *arXiv preprint arXiv:1701.07875*, 2017.

[13] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in *NIPS'17*, 2017, pp. 5769–5779.

[14] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, pp. 162–175, 2015.

[15] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis; using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[16] A. Gretton, K. M. Borgwardt, M. Rasch, B. Schölkopf, and A. J. Smola, "A kernel method for the two-sample-problem," in *NIPS'07*, 2007, pp. 513–520.

[17] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, 2011.

[18] L. V. D. Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, no. Nov, pp. 2579–2605, 2008.