A GAN-Based Data Augmentation Method for Multimodal Emotion Recognition

Yun Luo[1], Li-Zhen Zhu[1], and Bao-Liang Lu[1,2,3](✉)

[1] Center for Brain-Like Computing and Machine Intelligence, Department of Computer Science and Engineering, Shanghai Jiao Tong University, 800 Dong Chuan Road, Shanghai 200240, China

[2] Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognition Engineering, Shanghai Jiao Tong University, 800 Dong Chuan Road, Shanghai 200240, China

[3] Brain Science and Technology Research Center, Shanghai Jiao Tong University, 800 Dong Chuan Road, Shanghai 200240, China

{angeleader,zhulz98,bllu}@sjtu.edu.cn

Abstract. The lack of training data is an obstacle to build satisfactory multimodal emotion recognition models. Generative adversarial network (GAN) has recently shown great successes in generating realistic-like data. In this paper, we propose a GAN-based data augmentation method for enhancing the performance of multimodal emotion recognition models. We adopt conditional Boundary Equilibrium GAN (cBEGAN) to generate artificial differential entropy features of electroencephalography signal, eye movement data and their direct concatenations. The main advantage of cBEGAN is that it can overcome the instability of conventional GAN and has very quick converge speed. We evaluate our proposed method on two multimodal emotion datasets. The experimental results demonstrate that our proposed method achieves 4.6% and 8.9% improvements of mean accuracies on classifying three and five emotions, respectively.

Keywords: EEG · Eye movement · Emotion recognition · Generative adversarial network · Data augmentation

1 Introduction

Affective computing [12], which aims to equip machines with the ability to recognize, interpret, process, and simulate human affects, has drawn increasing attention in recent years. In the framework of affective computing, emotion recognition is the first critical phase since machines can never process human moods without precise emotion recognition. Researchers have made great progress in recognizing emotions from different signals, such as facial expressions, speeches, and some physiologoical signals including EEG and eye movement signals.

H. Lu et al. (Eds.): ISNN 2019, LNCS 11554, pp. 141–150, 2019.
https://doi.org/10.1007/978-3-030-22796-8_16

In recent years, researchers focused on studying multimodal emotion recognition methods to leverage the complementary property among different kinds of signals. Lu *et al.* introduced a multimodal emotion recognition framework for three emotions by combining EEG and eye movement signals [9]. By taking advantages of the deep neural networks, Liu *et al.* further improved the performance of multimodal framework [7]. Zhao *et al.* also adopted multimodal framework and extended it for recognizing five emotions [15]. Although these studies have developed various promising approaches for multimodal emotion recognition, the performance of emotion recognition models is unsatisfactory due to the lack of training data.

The popular multimodal emotion datasets contain physiological signals such as EEG and eye movement signals, which are difficult to collect. The high prices of EEG and eye movement acquisition devices and the high cost of performing multimodal emotion experiments limit the scale of the datasets. As a result, the training set is very small in size in comparison with image dataset such as ImageNet.

Data augmentation is a promising approach to dealing with the problem of lack of training data mentioned above. It enlarges the dataset by applying a transformation to the real data and generating realistic-like data [3]. Lotte generated artificial EEG data by relevant combinations and distortions of the original trials [8], and this approach increased the recognition accuracy when the training set is small. Krell *et al.* proposed to generate EEG data by rotational distortions [6]. Wang *et al.* improved the performance of the emotion recognition models by adding Gaussian noise to EEG features to generate artificial data [14]. However, the basic idea behind these methods is to generate more data by using geometric transformation and it is difficult to capture the deep information inside data.

By taking advantages of deep neural networks and adversarial training, GAN could learn information about data probability distribution and generate artificial data from real data distribution. In the field of computer vision, GAN has demonstrated its ability of generating realistic-like images by playing a zero-sum non-cooperative game [4,13]. Inspired by GAN, Hartmann *et al.* proposed EEG-GAN to generate EEG signals [5]. However, they did not use the generated data for classification. Luo and Lu generated EEG data in DE feature form by adopting cWGAN and enlarged the training dataset [10]. Their experimental results indicated that the accuracies of EEG-based emotion recognition models could be improved by adding training data generated by cWGAN.

In this paper, we propose a GAN-based data augmentation method for enhancing the performance of multimodel emotion recognition models. Since the original GAN suffers from instability and non-convergence problems, we implement cBEGAN to generate training data [1]. The main advantage of cBEGAN is that it has quick convergence speed and has an indicator for the training process. Meanwhile, we can control the category of the generated data by adding auxiliary conditional label information [11]. In this paper, we generate EEG signals and eye movement data in DE (differential entropy) feature form instead of raw data, because our previous studies have shown that the DE features of EEG and eye movement data are more effective for emotion recognition [2,9]. The multimodal data, which is the direct concatenation of EEG and eye movement data, are also generated in DE feature form with cBEGAN and cWGAN.

We evaluate our method on three-category and five-category multimodal emotion datasets. To the best of our knowledge, this is the first research work regarding GAN-based data augmentation for multimodal emotion recognition. Our experimental results demonstrate that cBEGAN has a better performance than cWGAN and significantly improves the accuracies of multimodal emotion recognition models.

2 Methods

2.1 GAN

A standard GAN consists of two competing parts which are both parameterized as deep neural networks. A generator G produces synthetic data given a noise variable input while a discriminator D tries to identify whether a sample comes from the real data distribution X_r or the generated data distribution X_g. In other words, the discriminator is trained to estimate the probability of a given sample coming from the real data distribution. And the generator is optimized to trick the discriminator to offer a high probability for the generated data. The two parts are optimized simultaneously to find a Nash equilibrium. More formally, the procedure can be expressed as a minimax function:

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

where θ_g and θ_d represent the parameters of generator and discriminator, respectively, and Z can be a Uniform noise distribution or a Gaussian noise distribution.

This function is optimized in two steps; (i) maximize it by fixing G and X_g, and get the optimum of D; and (ii) minimize the function by the optional D, and then it equals to minimizing the Jensen-Shannon divergence between X_r and X_g. The game will achieve equilibrium if and only if $X_r = X_g$.

Although GAN has shown great successes in realistic data generation, it suffers from some major problems such as non-convergence, mode collapse and diminished gradient. Researchers believed that the Jensen-Shannon divergence could lead to vanishing gradients, which was the main reason of the GAN's instability. In real world tasks such as image generation, the distribution of real images always lies in low dimensional manifolds, and the distribution of generated images also rests in low dimensional manifolds. The two distributions are almost certainly disjoint and have no overlaps. In this situation, Jensen-Shannon divergence between the two distributions is a fixed number, which can not provide useful gradients for the GAN's training.

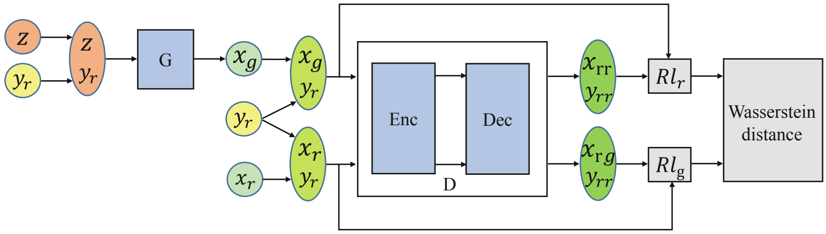

Fig. 1. Illustration of cBEGAN. cBEGAN adopts the auto-encoder to handle the difference between two reconstruction losses distributions. Here, G, D, Enc, Dec, Rl_r and Rl_g represent generator, discriminator, encoder, decoder and the two reconstruction losses, respectively.

2.2 cBEGAN

The discriminator in BEGAN adopts an auto-encoder which uses an encoder to extract the latent features from the input data and applies a decoder to reconstruct the data from the latent representations as shown in Fig. 1. And now the discriminator aims to matching the reconstruction loss distribution of real data and generated data. Berthelot *et al.* believe that matching auto-encoder loss could lead to the matching of the data distribution of real data and generated data directly [1], which is adopted in typical GANs. In other words, the generated data will have the similar data distribution when their reconstruction loss distributions are similar. In this way, BEGAN avoids the instability problem of conventional GAN.

BEGAN chooses Wasserstein distance to measure the difference between the two reconstruction loss distributions. The Wasserstein distance is also called Earth Mover's distance (EM distance). The distance formula for continuous probability domain is:

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

where $\Pi(X_r, X_g)$ is the set of all possible joint probability distributions between X_r and X_g. The reconstruction loss is defined as the pixel-wise L_1 or L_2 distance between input data and reconstructed data, which can be formulated as:

$$L_r(x) = |x - D(x)|^{\eta} \tag{3}$$

where D is the discriminator (auto-encoder) function, and $\eta \in 1, 2$, and x can be a sample of real data distribution or generated data distribution.

Let μ_{rr} and μ_{rg} be the real and generated reconstruction loss distributions, respectively, and let $m_{rr}, m_{rg} \in \mathbb{R}$ be their respective means, and $\Pi(\mu_{rr}, \mu_{rg})$ is the set of all possible joint probability distributions between two distributions. By using Jensens inequality, the formula can be expressed as:

$$\begin{aligned} W(\mu_{rr}, \mu_{rg}) &= \inf_{\gamma \sim \Pi(\mu_{rr}, \mu_{rg})} \mathbb{E}_{(x_{rr}, x_{rg}) \sim \gamma}[||x_{rr} - x_{rg}||] \\ &\geq \inf |\mathbb{E}[x_{rr} - x_{rg}]| = |m_{rr} - m_{rg}| \end{aligned} \tag{4}$$

so now we are aiming to optimize a lower bound of the Wasserstein distance between the two reconstruction losses. Then the loss of BEGAN is:

$$\begin{aligned}\min_{\theta_G}\max_{\theta_D} L(X_r, X_g) &= -\mathbb{E}_{x_r\sim X_r}[(L_r(x_r)] + \mathbb{E}_{z\sim Z}[L_r(G(z))] \\ &= -\mathbb{E}_{x_r\sim X_r}[(L_r(x_r)] + \mathbb{E}_{x_g\sim X_g}[L_r(x_g)]\end{aligned} \tag{5}$$

where θ_G and θ_D represent the respective parameters of the generator and the discriminator in BEGAN.

In BEGAN, the discriminator has two goals: auto-encode real data and discriminate real data from generated ones. In order to maintain the balance between the generator and discriminator losses, we can apply a hyper-parameter $\gamma \in [0, 1]$ defined as:

$$\gamma = \frac{\mathbb{E}[L_r(G(z))]}{\mathbb{E}[L_r(x_r)]} \tag{6}$$

To maintain the equilibrium $\mathbb{E}[L_r(G(z))] = \gamma\mathbb{E}[L_r(x_r)]$, we use Proportional Control Theory by adopting an extra variable $k_t \in [0, 1]$ to control the proportion of $L_r(G(z))$ during gradient descent. Similar with cWGAN, we add an extra label information to control the generated categories. The cBEGAN can be formulated as:

$$\begin{aligned}&\max_{\theta_D} L(X_r, X_g, Y_r) = \\ &\quad -\mathbb{E}_{x_r\sim X_r, y_r\sim Y_r}[(L_r(x_r|y_r)] + k_t\mathbb{E}_{x_g\sim X_g, y_r\sim Y_r}[L_r(x_g|y_r)]\end{aligned} \tag{7}$$

$$\min_{\theta_G} L(X_g, Y_r) = \mathbb{E}_{x_g\sim X_g, y_r\sim Y_r}[L_r(x_g|y_r)] \tag{8}$$

$$k_{t+1} = k_t + \lambda_k(\gamma L_r(x_r) - L(G(z))) \tag{9}$$

where Y_r is the label distribution. We initialize $k_0 = 0$ and set $\lambda_k = 0.001, \gamma = 0.75$ in this paper. Now we can define a convergence measure as:

$$Mglobal = L_r(x_r) + |\gamma L_r(x_r) - L_r(G(z)| \tag{10}$$

Mglobal can be used as an indicator for the convergence of the network.

In this paper, we also extend cWGAN, used in our previous work [10], to multimodal emotion recognition. For cWGAN and cBEGAN, the losses of generator and discriminator are optimized in an alternating procedure. The distribution of the generated data is similar with the real data when the networks converge.

3 Experiment Settings

3.1 EEG Datasets

We evaluate our proposed method on two multimodal emotion datasets SEED [1] [16] and SEED-V.

[1] http://bcmi.sjtu.edu.cn/~seed/index.html.

SEED dataset contains 62-channel EEG signals and eye movement signals of three different emotions (happy, sad, and neutral). The original EEG signals were recorded at a sampling rate of 1000 Hz with ESI NeuroScan System and the eye movement signals were collected with SMI ETG eye tracking glass, which contained information about blink, saccade, fixation and so on. In this dataset, 15 participants watched 15 emotional film clips for 3 times. In this work, 9 subjects' data (27 experiments) are used because they have completed multimodal data.

SEED-V dataset were also formed with 62-channel EEG signals and eye movement signals. 16 participants watched 15 emotional film clips to elicit five emotions: happy, sad, neutral, fear, and disgust. They took part in the experiments for three times, so there were totally 48 experiments. The EEG and eye movement signals were collected by the same device used in SEED.

3.2 Feature Extraction

We use a band pass filter (1–50 Hz) to eliminate low-frequency noise and high-frequency noise in the EEG signals. Then we extract DE features by adopting a 4s-length non-overlapping Hanning window for five frequency bands: δ: 1–3 Hz, θ: 4–7 Hz, α: 8–13 Hz, β: 14–30 Hz, and γ: 31–50 Hz. In order to eliminate the rapid changes of the DE features, we also adopt a linear dynamic system. Each EEG sample has 310 dimensions since there are 62 channels for each band.

As for eye movement signals, we extracted the same features as in [9,15]. The features include blink, saccade, fixation and so on. Notably, each eye movement sample has 41 dimensions in SEED dataset and it has 33 dimensions in SEED-V dataset since we simplify the eye movement features in SEED-V.

3.3 Evaluation Details

In order to demonstrate the effectiveness of the proposed method, we conduct cross validation on both datasets. Since, each experiment of the two datasets has 15 trials, so there are 5 trials for each emotion category in SEED dataset and 3 trails for each emotion category in SEED-V dataset. As for SEED dataset, we use 5-fold cross validation for each experiment. And as for SEED-V dataset, we adopt 3-fold cross validation for each experiment to make sure each fold has 5 emotion categories. We normalize both DE and eye movement features by min-max normalization before feeding them to the networks.

We perform grid search on the number of network layers and hidden nodes to optimize the network structure of cWGAN and cBEGAN. The numbers of layers are searched from 2 to 4 for both generator and discriminator. The input dimension is decided by the dimension of the corresponding input feature and the dimension of auxiliary label is 3 for SEED and 5 for SEED-V. The output dimension of cWGAN's discriminator is 1 while the output dimension is the same with its input for cBEGAN's discriminator.

The numbers of hidden nodes for each layer are randomly searched from 50 to 600. For cBEGAN, the hidden nodes of encoder and decoder are the same. The outputs of the two generators have the same dimension as the input data.

And ReLU activation function is used for all hidden layers. The batch size is set to 128. Adam with initial learning rate 0.0001 is used as the optimizer. The noises are sampled from a uniform distribution $U[-1, 1]$.

As for the classifier, we apply an SVM with linear kernel. The parameter c is searched from $2^{-10} \sim 2^{10}$ to find the optimal value.

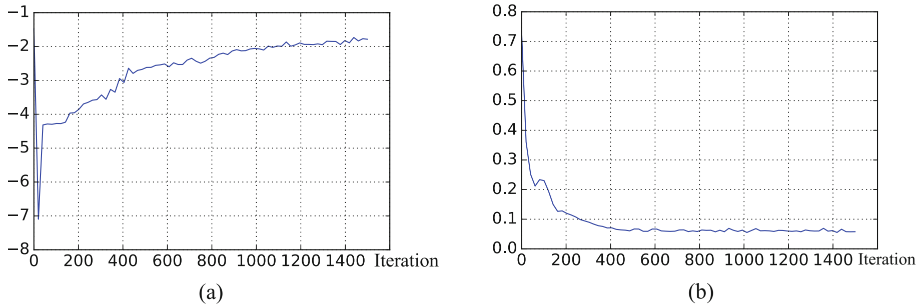

Fig. 2. Dloss for cWGAN (a), and Mglobal for cBEGAN (b) tendency along with training steps of SEED dataset.

4 Experimental Results

To evaluate the performance of the proposed method, we generate different features for both datasets. We generate DE features of EEG signals and eye movement data. As for multimodal data augmentation, we directly concatenate DE features of EEG signals and DE features of eye movement data, and generate realistic-like multimodal feature from the concatenated features. The number of the generated features for each emotion category is the same. In this section, we will first compare the convergence speed between cBEGAN and cWGAN, then discuss the performance of data augmentation for the two datasets.

4.1 Convergence Performance

As mentioned above, cWGAN and cBEGAN can overcome the instability problem of conventional GANs and both of them have an indicator for training procedure. Figure 2(a) shows the convergence curve of cWGAN. Dloss rises to -2 after 1000 iterations, which indicates the network have a good convergence performance. Besides, as the Wasserstein distance between real data distribution and generated data distribution, Dloss converging to a small value means the two data distributions are similar. As shown in Fig. 2(b), Mglobal decreases to about 0.6 and also has a stable convergence trend. cBEGAN has a better convergence performance since it converges after 500 iterations.

Table 1. Mean accuracies and standard deviations of the models trained on SEED dataset and appending datasets generated by cBEGAN and cWGAN

No. of appended training data	EEG		Eye movement		Multimodal	
	cWGAN	cBEGAN	cWGAN	cBEGAN	cWGAN	cBEGAN
0	0.8190	0.8190	0.7715	0.7715	0.8573	0.8573
	0.1074	0.1074	0.1327	0.1327	0.0879	0.0879
50	0.8331	0.8423	0.7881	0.7938	0.8606	0.8814
	0.1014	0.1020	0.1241	0.1202	0.0864	0.0906
200	0.8392	0.8557	0.7924	0.8043	0.8621	0.8878
	0.1028	0.0941	0.1249	0.1262	0.0877	0.0888
600	0.8372	0.8601	0.7956	0.8100	0.8539	0.9021
	0.1045	0.0876	0.1225	0.1228	0.0913	0.0858
700	0.8373	0.8641	0.7907	0.8063	0.8589	**0.9033**
	0.1086	0.0894	0.1262	0.1241	0.0883	0.0837
800	0.8377	0.8651	0.7929	0.8093	0.8558	0.9033
	0.1084	0.0914	0.1213	0.1139	0.0887	0.0837
2000	0.8338	**0.8756**	0.7958	**0.8160**	0.8586	0.9000
	0.1030	**0.0852**	0.1276	**0.1042**	0.0874	**0.0776**

4.2 SEED Results

For SEED dataset, the number of samples for each experiment is 842. And we generate 50, 200, 600, 700, 800, and 2000 artificial samples of the three features and add them to their respective original training datasets. Table 1 illustrates the performance at different number of augmented training data. 0 means the model is trained by original training dataset. As for single modality, cBEGAN achieves the best mean accuracies of 87.56% and 81.60% when we add 2000 samples of generated EEG and eye movement data, respectively. For multimodal data augmentation, cBEGAN reaches the best mean accuracy of 90.33% when adding 700 generated multimodal data.

4.3 SEED-V Results

For each subject, the number of sample for each experiment is 681, 541 and 601 since they watched different movie clips for each time. Considering these numbers are approximate, we neglect the difference and generate 50, 200, 400, 500, 1000 and 2000 samples of the three data and enlarge their respective original dataset for each experiment. As shown in Table 2, cBEGAN achieves the best mean accuracies of 62.87%, 60.19%, and 68.32% when we add 2000, 2000 and 1000 samples to the training datasets of EEG, eye movement and multimodal data.

Table 2. Mean accuracies and standard deviations of the models trained on SEED-V dataset and appending datasets generated by cBEGAN and cWGAN

No. of appended training data	EEG		Eye movement		Multimodal	
	cWGAN	cBEGAN	cWGAN	cBEGAN	cWGAN	cBEGAN
0	0.5434	0.5434	0.4862	0.4862	0.5946	0.5946
	0.1525	0.1525	0.1432	0.1432	0.1603	0.1603
50	0.5793	0.6064	0.5207	0.5533	0.6260	0.6485
	0.1534	0.1655	0.1381	0.1388	0.1599	0.1595
100	0.5846	0.6124	0.5336	0.5555	0.6279	0.6568
	0.1546	0.1616	0.1345	0.1382	0.1626	0.1559
200	0.5901	0.6181	0.5457	0.5609	0.6294	0.6674
	0.1536	0.1592	0.1437	**0.1369**	0.1594	0.1598
400	0.5946	0.6207	0.5446	0.5816	0.6366	0.6775
	0.1580	0.1558	0.1417	0.1434	0.1582	0.1584
500	0.5954	0.6225	0.5349	0.5815	0.6330	0.6810
	0.1571	0.1544	0.1400	0.1430	0.1606	0.1548
1000	0.5965	**0.6287**	0.5486	0.5892	0.6326	**0.6832**
	0.1594	0.1526	0.1456	0.1388	0.1590	0.1549
2000	0.5912	0.6278	0.5518	**0.6019**	0.6325	0.6831
	0.1593	**0.1442**	0.1470	0.1399	0.1620	**0.1504**

Compared with cWGAN, cBEGAN has higher accuracies for single and multimodal data augmentation on the two datasets. Besides, cBEGAN also has a better convergence performance. By measuring the difference between the two reconstruction loss distributions instead of two data distributions, cBEGAN can capture deeper information of the real data distribution than cWGAN, and generate artificial samples with rich information and diverse distribution, which leads better margins for the recognition models. Although cWGAN-based data augmentation has a poorer performance in terms of accuracy than cBEGAN-based data augmentation, the mean accuracies also has improvements on the two datasets, which demonstrates the multimodal emotion recognition models are more robust when adopting the proposed GAN-based data augmentation method.

5 Conclusion and Future Work

In this paper, we have proposed a GAN-based data augmentation method for improving the accuracy of multimodal emotion recognition models. We have generated realistic-like EEG, eye movement and their direct concentration data with cBEGAN and cWGAN. Our experimental results on two multimodal emotion datasets indicate the effectiveness of the proposed method and cBEGAN achieves the biggest improvements of mean accuracies on classifying three and five emotions with a better convergence speed. In the future, we will evaluate our method on more multimodal emotion recognition tasks and employ recurrent neural networks to consider temporal information of EEG and eye movement signals.

Acknowledgments. This work was supported in part by grants from the National Key Research and Development Program of China (Grant No. 2017YFB1002501), the National Natural Science Foundation of China (Grant No. 61673266), and the Fundamental Research Funds for the Central Universities.

References

1. Berthelot, D., Schumm, T., Metz, L.: BEGAN: Boundary equilibrium generative adversarial networks. arXiv preprint arXiv:1703.10717 (2017)
2. Duan, R.N., Zhu, J.Y., Lu, B.L.: Differential entropy feature for EEG-based emotion classification. In: IEEE EMBS NER 2013, pp. 81–84. IEEE (2013)
3. Dyk, D.A.V., Meng, X.L.: The art of data augmentation. J. Comput. Graph. Stat. **10**(1), 1–50 (2001)
4. Goodfellow, I., et al.: Generative adversarial nets. In: NIPS 2014, pp. 2672–2680 (2014)
5. Hartmann, K.G., Schirrmeister, R.T., Ball, T.: EEG-GAN: generative adversarial networks for electroencephalograhic (EEG) brain signals (2018)
6. Krell, M.M., Su, K.K.: Rotational data augmentation for electroencephalographic data. In: Engineering in Medicine and Biology Society, pp. 471–474 (2017)
7. Liu, W., Zheng, W.-L., Lu, B.-L.: Emotion recognition using multimodal deep learning. In: Hirose, A., Ozawa, S., Doya, K., Ikeda, K., Lee, M., Liu, D. (eds.) ICONIP 2016. LNCS, vol. 9948, pp. 521–529. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46672-9_58
8. Lotte, F.: Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain-computer interfaces. Proc. IEEE **103**(6), 871–890 (2015)
9. Lu, Y., Zheng, W.L., Li, B., Lu, B.L.: Combining eye movements and EEG to enhance emotion recognition. In: IJCAI 2015, pp. 1170–1176 (2015)
10. Luo, Y., Lu, B.L.: EEG data augmentation for emotion recognition using a conditional Wasserstein GAN. In: IEEE EMBS 2018, pp. 2535–2538. IEEE (2018)
11. Mirza, M., Osindero, S.: Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784 (2014)
12. Picard, R.W.: Affective Computing. MIT Press, Cambridge (2000)
13. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. In: ICLR 2016 (2016)
14. Wang, F., Zhong, S., Peng, J., Jiang, J., Liu, Y.: Data augmentation for EEG-based emotion recognition with deep convolutional neural networks. In: Schoeffmann, K., et al. (eds.) MMM 2018. LNCS, vol. 10705, pp. 82–93. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-73600-6_8
15. Zhao, L.M., Li, R., Zheng, W.L., Lu, B.L.: Classification of five emotions from EEG and eye movement signals: complementary representation properties. In: IEEE NER 2019. IEEE (2019)
16. Zheng, W.L., Lu, B.L.: Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. IEEE Transact. Auton. Ment. Dev. **7**(3), 162–175 (2015)