The 38th Annual AAAI Conference on Artificial Intelligence
FEBRUARY 20-27, 2024 | VANCOUVER, CANADA
VANCOUVER CONVENTION CENTRE – WEST BUILDING

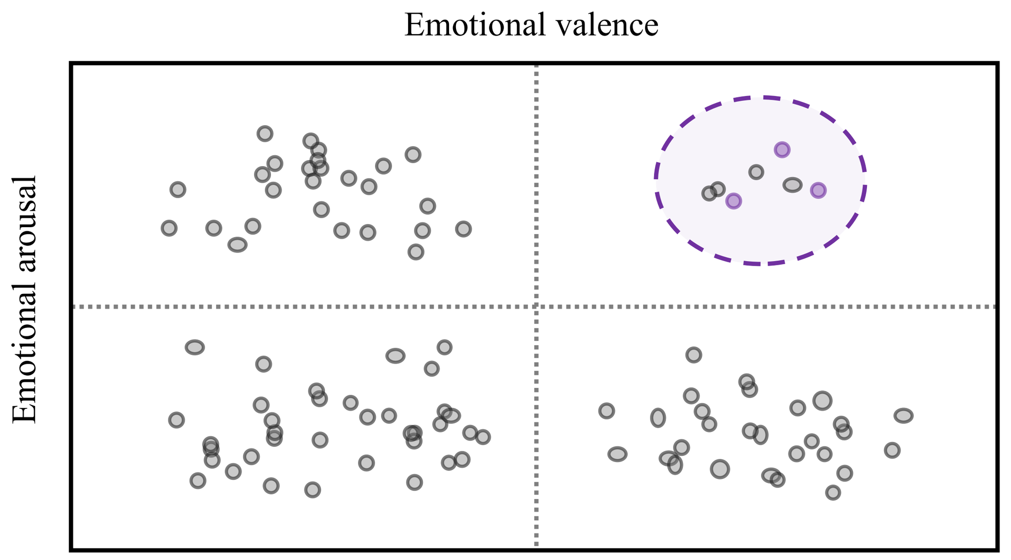
Beyond Mimicking Under-Represented Emotions: Deep Data Augmentation with Emotional Subspace Constraints for EEG-Based Emotion Recognition

Zhi Zhang [1,2], Shenghua Zhong [1,*], Yan Liu [2]

[1] Shenzhen University
[2] Hong Kong Polytechnic University

Abstract

In recent years, using Electroencephalography (EEG) to recognize emotions has garnered considerable attention. Despite advancements, limited EEG data restricts its potential. Thus, Generative Adversarial Networks (GANs) are proposed to mimic the observed distributions and generate EEG data. However, for imbalanced datasets, **GANs struggle to produce reliable augmentations for under-represented minority emotions by merely mimicking them.**



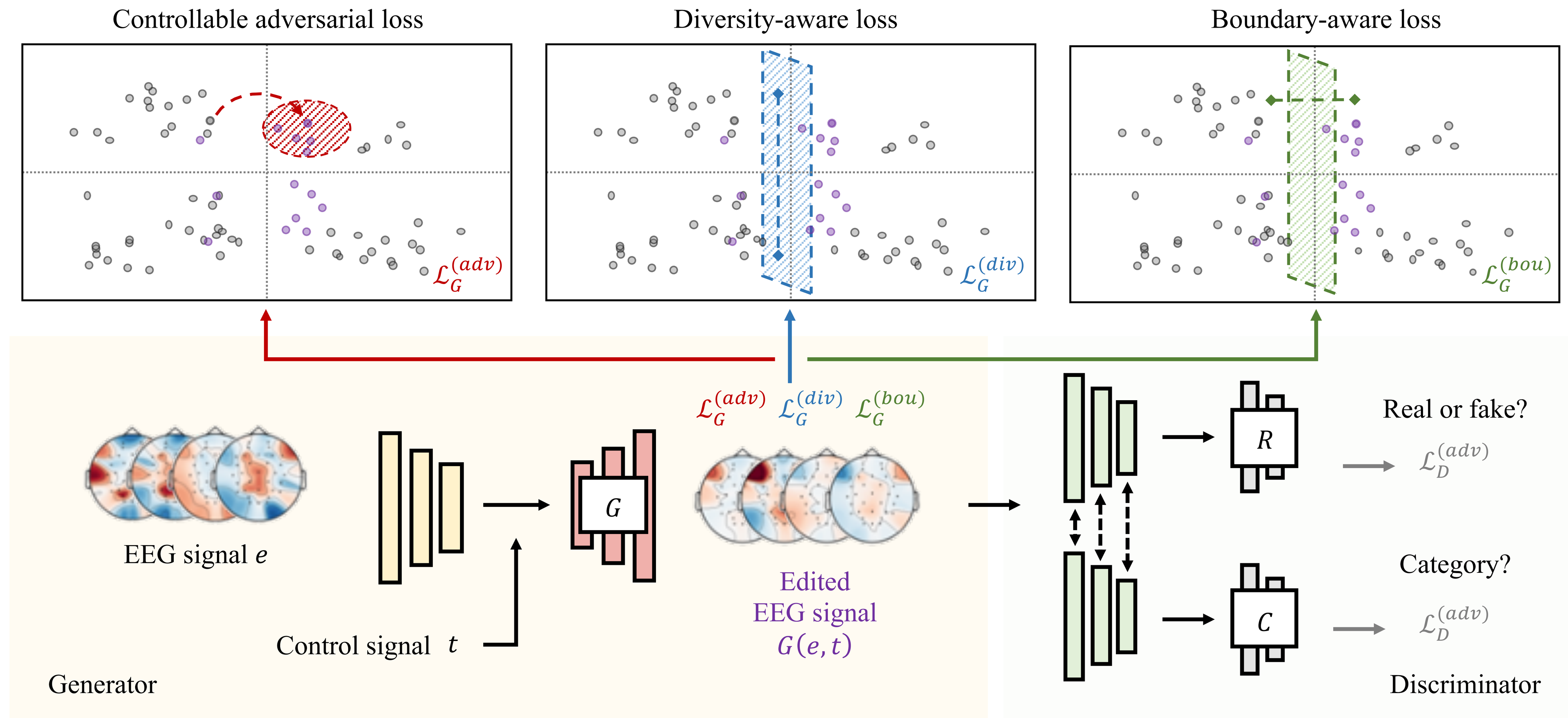
Approximate the observed distribution via existing GANs

To tackle this bottleneck, we introduce emotional subspace-constrained generative adversarial networks (ESC-GAN).

1. We first propose the **EEG editing paradigm**, editing reference EEG signals from well-represented to under-represented emotional subspaces.

2. We introduce **diversity-aware and boundary-aware losses** to constrain the augmented subspace. Diversity-aware loss encourages a diverse emotional subspace by enlarging the sample difference, while boundary-aware loss constrains the augmented subspace near the decision boundary, where recognition models can be vulnerable.

Emotional Subspace Constrained Generative Adversarial Networks



EEG editing paradigm: ESC-GAN consists of a generator and two discriminators. The generator is the core module of ESC-GAN. It **edits the reference EEG signal** e **to a new EEG signal** with the user-specific emotional category t, producing augmented samples.

For well-represented emotions, we optimize the editor to learn the distribution and produce real-like. Here, we introduce controllable adversarial loss to **approximate the real distribution via adversarial training** while learning a mapping to the target emotion category.

$$\mathcal{L}_D^{adv} = \mathbb{E}_{e\sim\mathbb{P}_e}[\mathcal{H}_{\text{cross}}(R(e),0)] + \mathbb{E}_{e\sim\mathbb{P}_e}[\mathcal{H}_{\text{cross}}(R(G(e,t)),\tau)] + \mathbb{E}_{e\sim\mathbb{P}_e}[\mathcal{H}_{\text{cross}}(C(e),y)]$$

$$\mathcal{L}_G^{adv} = \mathbb{E}_{e\sim\mathbb{P}_e} \parallel e - G(e,y) \parallel^2 + \mathbb{E}_{e\sim\mathbb{P}_e}[\mathcal{H}_{\text{cross}}(R(G(e,t)),0)] + \mathbb{E}_{e\sim\mathbb{P}_e}[\mathcal{H}_{\text{cross}}(C(G(e,t)),t)]$$

1. fool the discriminator R into classifying the generated samples as real EEG signals **2.** force the discriminator C to classify the generated samples into the target emotional category.

Diversity-aware loss: For under-represented emotions involving inadequate EEG signals, to mimic their observed distributions may result in similar samples near the limited observed samples. We propose to **encourage a diverse emotional subspace to enhance the diversity of under-represented emotional distributions**.

$$\mathcal{L}_G^{div} = \mathbb{E}_{e_i,e_j\sim\mathbb{P}_e,t_i,t_j\sim\mathbb{P}_t} e^{\frac{-|G(e_i,t_i)-G(e_j,t_j)|}{|F(G(e_i,t_i))-F(G(e_j,t_j))|}}$$

We utilize an RBF-like measurement to compute the difference between the generated EEG signals mapped in an infinite-dimensional space.

Boundary-aware loss: DNNs are vulnerable to erroneous instances near their decision boundaries, prompting us to **constrain the augmented subspace near the decision boundaries.**

$$\mathcal{L}_G^{bou} = \mathbb{E}_{e\sim\mathbb{P}_e,t\sim\mathbb{P}_t} KL\left(U(t) \parallel C(G(e,t))\right)$$

where $KL(.)$ is the KL divergence and $U(t)$ is a uniform distribution. The boundary-aware loss can prevent the model from producing the generated EEG signal, which can be classified into a single category with a higher probability, thereby nudging the generated samples toward the classification boundary.

To ensure the quality of the augmented samples, we use the discriminators R and C to **filter the generated samples**:

1. not classify the generated sample as fake **2.** classify the generated samples into the target category

Condition (1) : $R(G(e,t)) < \delta$ Condition (2): $\operatorname{argmax} C(G(e,t)) == t$

Experiments

Method		DEAP		AMIGOS		SEED
		Valence	Arousal	Valence	Arousal	
Recent state-of-the-art	Santamaria *et al.* (Santamaria-Granados et al. 2018)	-	-	75.00	76.00	-
	Siddharth *et al.* (Jung, Sejnowski et al. 2019)	71.09	72.58	83.02	79.13	-
	Topic *et al.* (Topic and Russo 2021)	76.61	77.72	87.39	90.54	88.45
	Hu *et al.* (Hu et al. 2021)	71.8	71.88	74.06	69.52	-
	Zhang *et al.* (Zhang, Zhang, and Wang 2022)	85.86	84.27	-	-	92.47
	Yang *et al.* (Yang et al. 2018)	90.26	90.98	-	-	-
	Tao *et al.* (Tao et al. 2020)	93.72	93.38	-	-	-
	Shen *et al.* (Shen et al. 2020)	94.22	94.58	-	-	94.74
	Feng *et al.* (Feng et al. 2022)	95.04	95.52	-	-	96.72
Imbalanced learning	Chawla *et al.* (Chawla et al. 2002)	94.41	94.65	93.77	94.38	95.85
	He *et al.* (He et al. 2008)	94.25	94.6	93.75	94.42	95.76
GAN-based augmentation	Luo *et al.* (Luo and Lu 2018)	73.89	78.17	-	-	86.96
	Liu *et al.* (Liu, Hao, and Guo 2023)	92.75	93.52	-	-	-
	Zhang *et al.* (Zhang, Zhong, and Liu 2022)	93.52	94.21	-	-	**97.7**
	Proposed	**96.33**	**96.68**	**94.40**	**95.23**	97.14

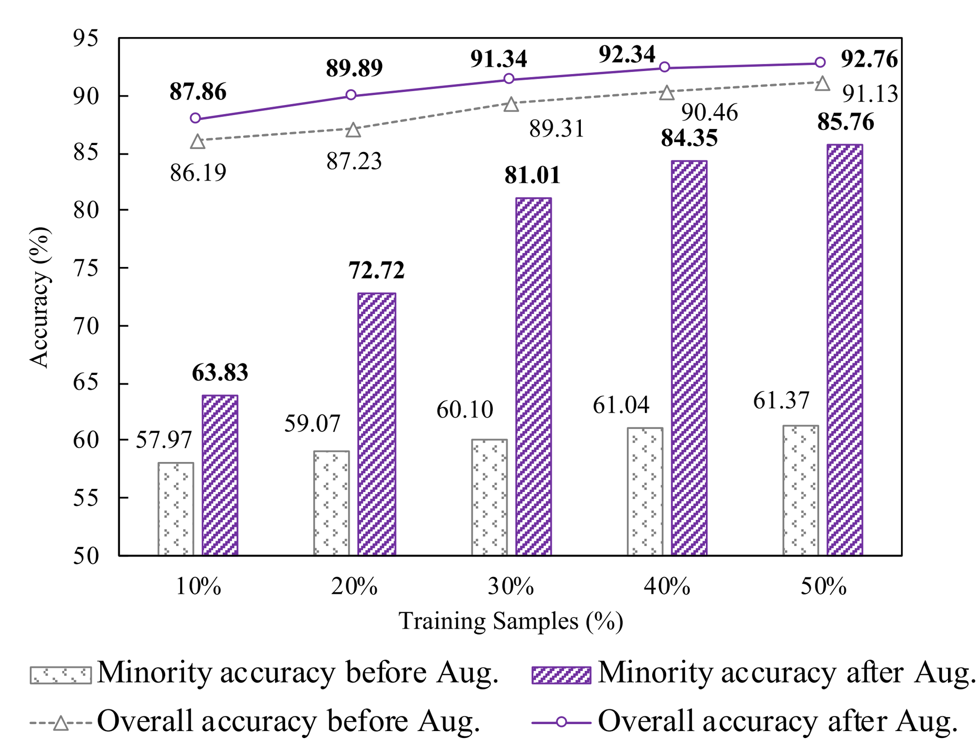


Fig. 1 The ***hard-case experiments*** *on the DEAP dataset by removing parts of training samples and keeping 10% to 50% training samples in the control quadrant.*

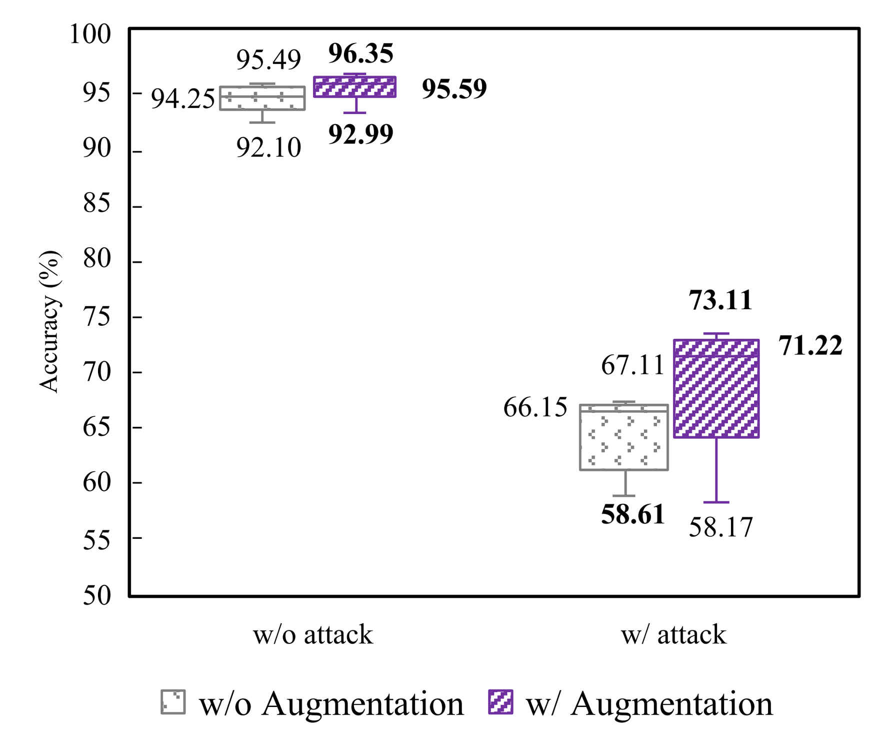


Fig. 2 Performance of classifying emotional arousal on the AMIGOS dataset before and ***after adversarial sample attacks*** *using Fast Gradient Sign Method.*

Adversarial Attacks	w/o Aug.	w/ Aug.
w/o attack	94.25	**95.59**
w/ FGSM attack	66.15	**71.22**
w/ PGD attack	63.3	**68.44**
w/ Deep Fool attack	1.24	**4.80**

1. Comparison experiments on DEAP, AMIGOS, and SEED datasets (Table on the left) demonstrate the superior performance of our proposed method compared to **state-of-the-art** approaches.

2. Hard-case experiments (Fig. 1) demonstrate the effectiveness of the ESC-GAN on **imbalanced problems**.

3. In **adversarial attack experiments** (Fig. 2, Table above), ESC-GAN shows robustness in defending **against adversarial samples**.

The proposed method opens new avenues for editing EEG signals under emotional subspace constraints, facilitating unbiased and secure EEG data augmentation.

zhi271.zhang@connect.polyu.hk, csshzhong@szu.edu.cn, yan.liu@polyu.edu.hk