AC-WGAN-GP: Augmenting ECG and GSR Signals using Conditional Generative Models for Arousal Classification

Andrei Furdui
Centrum Wiskunde & Informatica
The Netherlands
andreifurdui.g@gmail.com

Tianyi Zhang
Centrum Wiskunde & Informatica
The Netherlands
tianyi.zhang@cwi.nl

Marcel Worring
University of Amsterdam
The Netherlands
m.worring@uva.nl

Pablo Cesar
Centrum Wiskunde & Informatica
The Netherlands
p.s.cesar@cwi.nl

Abdallah El Ali
Centrum Wiskunde & Informatica
The Netherlands
abdallah.el.ali@cwi.nl

ABSTRACT

Computational recognition of human emotion using Deep Learning techniques requires learning from large collections of data. However, the complex processes involved in collecting and annotating physiological data lead to datasets with small sample sizes. Models trained on such limited data often do not generalize well to real-world settings. To address the problem of data scarcity, we use an Auxiliary Conditioned Wasserstein Generative Adversarial Network with Gradient Penalty (AC-WGAN-GP) to generate synthetic data. We compare the recognition performance between real and synthetic signals as training data in the task of binary arousal classification. Experiments on GSR and ECG signals show that generative data augmentation significantly improves model performance (avg. 16.5%) for binary arousal classification in a subject-independent setting.

CCS CONCEPTS

• **Computing methodologies** → **Machine learning**; • **Human-centered computing** → *Human computer interaction (HCI).*

KEYWORDS

Data augmentation; Generative Adversarial Networks; Physiological signals; Arousal classification

ACM Reference Format:
Andrei Furdui, Tianyi Zhang, Marcel Worring, Pablo Cesar, and Abdallah El Ali. 2021. AC-WGAN-GP: Augmenting ECG and GSR Signals using Conditional Generative Models for Arousal Classification. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers (UbiComp-ISWC '21 Adjunct), September 21–26, 2021, Virtual, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3460418.3479301

1 INTRODUCTION

Accurate recognition of emotions plays a crucial role in understanding users' preference for media items such as images or video clips [8]. Emotional states are accompanied by certain involuntary responses in different parts of the body, such as in the brain, heart and skin. These responses can be measured using physiological sensors, thus giving us an objective window into the realm of emotions. In recent years, advances in Deep Learning (DL) techniques has enabled researchers to directly model the mappings between physiological signals (e.g., galvanic skin response (GSR) and electrocardiogram (ECG)) and human emotions [5, 6], without resorting to crafting features manually that require expert knowledge.

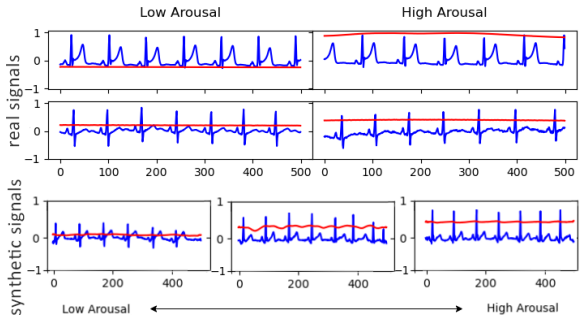

Figure 1: Real and synthetic signals from high/low arousal (ECG in blue, GSR in red)

The data-hungry nature of DL-based systems require a massive amount of information to harness their full potential. Although collecting large amounts of physiological measurements from sensors is trivial, reliably annotating these large datasets is not. The self-report annotation process is usually performed continuously over the course of an experiment [3]. This continuous annotation process, which requires significant time and resources, limits the potential size of the datasets. Thus, most widely-used datasets are collected from a small sample of people, typically less than 50 [3], which make it difficult to learn recognition models that generalize well to all subjects (i.e., subject-independent (SI) models).

To overcome the challenge of insufficient amounts of diverse physiological data, we propose the use of Generative Adversarial Networks (GANs) to sample new physiological signals (e.g., ECG and GSR) from a learned generative distribution. Specifically, we combine the Wasserstein GAN with Gradient Penalty (WGAN-GP) [1] with an Auxiliary Classifier network [2] (AC-WGAN-GP) in order to introduce conditioning information within the generative framework. We use the arousal information as the conditioning variable, which

UbiComp-ISWC '21 Adjunct, September 21–26, 2021, Virtual, USA

ACM ISBN 978-1-4503-8461-2/21/09.
https://doi.org/10.1145/3460418.3479301

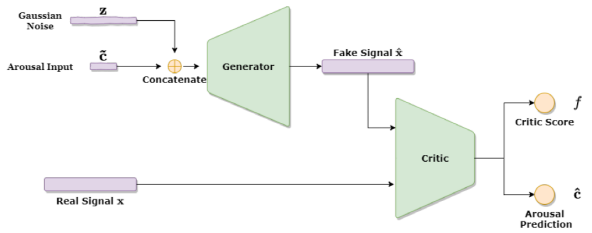

Figure 2: The architecture of AC-WGAN-GP

allows us to sample labeled synthetic signals to use in the process of supervised learning. Our work contributes an intelligent learning and sampling algorithm which can increase the size and variety of datasets for physiological signal-based emotion recognition. We test the proposed algorithm on ECG and GSR signals. Our experiment results from binary arousal classification show that synthetic data can significantly improve the classification performance (16% and 17% using CNN and LSTM, respectively) of real data by providing a balanced and densely distributed dataset for training DL algorithms.

2 METHOD: AC-WGAN-GP

To generate synthetic physiological signals, we combine WGAN-GP [1] and Auxiliary Classifier (AC) to create AC-WGAN-GP [7]. We fuse the two structures to generate physiological signals which correspond to the distribution of specific emotion categories (e.g., high/low arousal). Thus, the synthetic signals can be used for training emotion classification models. A schematic representation of the AC-WGAN-GP framework is shown in Figure 2. For the generator input, we concatenate the noise vector (Gaussian noise, initialization for the synthetic signals) z of length 128 with the one-hot encoded arousal class labels. The critic, parameterized by w takes as input a real or a synthetic data point and outputs two values: a scalar $D_{w:f(x)}$ which is the critic score corresponding to the 1-Lipschitz function, a probability distribution $P(Cjx)$ over the arousal class C (hereon denoted as $Dw:c(x)$). The whole network is trained end-to-end, using an objective function that combines the Wasserstein loss [1] with the Gradient Penalty and classification loss.

3 RESULTS AND DISCUSSION

We test AC-WGAN-GP for generating ECG and GSR signals using the CASE [3] dataset as it has continuous annotations (cf., [4]) taken during video watching. Continuous annotation is important, especially here where window segmentation (5 sec) is applied to the data, and arousal labels have to be assigned to each individual signal segment (instance). We generate synthetic signals using one-hot encoded conditional labels, which means synthetic instances correspond to low or high arousal categories. A visual comparison of real and synthetic signals is shown in Figure 1.

A binary classification task is implemented by training both with real signals and synthetic signals. A complete Leave-One-Subject-Out Cross-Validation (LOSOCV) would involve alternatively swapping out each subject to be used as test set data, amounting, in our case, to 120 individual models for all 30 subjects in the CASE dataset. For practical reasons, we restrict our experimentation to 20% of the search space. We randomly sampled 6 (subjects 4, 6, 11, 18, 22 and 25) out of the 30 subjects to be used as test subjects for LOSOCV. For the classification task, we use a Convolutional Neural Network (CNN) and Long Short-Term Memory network (LSTM) to compare the performance trained on real (-R) and synthetic (-S) data.

Table 1: The ACC and W-F1 score for the arousal classification using real (-R) and synthetic (-S) data.

		S4	S6	S11	S18	S22	S25	AVG
CNN-R	**ACC**	0.42	0.55	**0.73**	0.46	0.49	**0.50**	0.53
	W-F1	(0.37)	(0.49)	(0.46)	**(0.49)**	(0.38)	(0.34)	(0.42)
CNN-S	**ACC**	**0.68**	0.53	**0.73**	**0.52**	0.59	0.43	**0.58**
	W-F1	**(0.80)**	**(0.51)**	**(0.49)**	**(0.49)**	(0.70)	(0.49)	**(0.58)**
LSTM-R	**ACC**	0.41	**0.59**	0.73	0.50	0.49	0.47	0.53
	W-F1	(0.34)	(0.45)	(0.47)	(0.38)	(0.38)	(0.27)	(0.38)
LSTM-S	**ACC**	0.56	0.55	0.71	0.49	**0.62**	0.45	0.56
	W-F1	(0.65)	(0.49)	**(0.49)**	(0.37)	**(0.71)**	**(0.57)**	(0.55)

Table 1 shows the accuracy (ACC) and Weighted F1 (W-F1) score for all 6 subjects as testing data, respectively. While average (AVG) ACC improves by a small amount, we see a significant increase of W-F1 by 16% and 17% for the CNN and LSTM, respectively. This indicates that models trained on real data tend to over-classify one of the two classes. Using synthetic data, however, leads to more balanced classifiers. This fact appears even more clearly if we look at the performance for individual subjects. We see large improvements in W-F1 for subjects 4 and 22, which are the most imbalanced subsets at 27.5% and 37.5% minority class, respectively. For these subjects, W-F1 increases by upwards of 43%, which is a significant improvement. The increased ACC for SI testing shows that amount of data is sufficient for generating signals correspond to the distribution of high/low arousal.

4 CONCLUSION

We present AC-WGAN-GP to generate synthetic ECG and GSR signals which can be used as training data to enhance the generalizability of emotion recognition algorithms. The synthetic data generated by our method significantly improves the classification performance in subject-independent testing by providing more balanced classification results for different arousal categories. We provide our initial steps towards further investigating how generative models can be applied to diverse physiological signals for the task of physiologically-driven emotion recognition.

REFERENCES

[1] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron Courville. 2017. Improved training of wasserstein gans. *arXiv preprint arXiv:1704.00028* (2017).
[2] Augustus Odena, Christopher Olah, and Jonathon Shlens. 2017. Conditional image synthesis with auxiliary classifier gans. In *International conference on machine learning*. PMLR, 2642–2651.
[3] Karan Sharma, Claudio Castellini, Egon L van den Broek, Alin Albu-Schaeffer, and Friedhelm Schwenker. 2019. A dataset of continuous affect annotations and physiological signals for emotion analysis. *Scientific data* 6, 1 (2019), 1–13.
[4] Tianyi Zhang, Abdallah El Ali, Chen Wang, Alan Hanjalic, and Pablo Cesar. 2020. RCEA: Real-Time, Continuous Emotion Annotation for Collecting Precise Mobile Video Ground Truth Labels. In *Proc. CHI '20*. ACM, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376808
[5] Tianyi Zhang, Abdallah El Ali, Chen Wang, Alan Hanjalic, and Pablo Cesar. 2021. Cornet: Fine-grained emotion recognition for video watching using wearable physiological sensors. *Sensors* 21, 1 (2021), 52.
[6] Tianyi Zhang, Abdallah El Ali, Chen Wang, Xintong Zhu, and Pablo Cesar. 2019. CorrFeat: Correlation-based Feature Extraction Algorithm using Skin Conductance and Pupil Diameter for Emotion Recognition. In *2019 International Conference on Multimodal Interaction*. 404–408.
[7] Ming Zheng, Tong Li, Rui Zhu, Yahui Tang, Mingjing Tang, Leilei Lin, and Zifei Ma. 2020. Conditional Wasserstein generative adversarial network-gradient penalty-based approach to alleviating imbalanced data classification. *Information Sciences* 512 (2020), 1009–1023.
[8] Mircea Zloteanu and Eva G Krumhuber. 2021. Expression authenticity: The role of genuine and deliberate displays in emotion perception. *Frontiers in Psychology* 11 (2021), 4001.