Generative Adversarial Network for Facial Emotion Recognition: A Feasibility Study

Herag Arabian and Knut Moeller

Institute of Technical Medicine (ITeM) – Furtwangen University
H.Arabian@hs-furtwangen.de

Abstract. Integration of artificial intelligence into different domains has been a trending topic over the past few years. A closed-loop feedback system which immerses the subject in a virtual reality environment with a novel reward platform is being developed to help people suffering from autism spectrum disorder. In this work, the feasibility of using generative adversarial networks to generate synthetic images by restructuring unseen input data to match that of the training set for the recognition of human emotions is being studied. System performance was based on true positive predictions from the different classification models developed in previous work. Preliminary results showed that the proposed system was able to improve class predictions, but lacked in the ability to generate different class sets. The performance highlights the feasibility of this method and its practical applications in generating more data and improving model robustness.

Keywords: Facial Emotion Recognition; Generative Adversarial Networks; Therapeutic Application.

1 Introduction

Integration of artificial intelligence (AI) into different domains has been a trending topic over the past few years. One of the most popular forms of AI is the use of Deep learning techniques i.e. neural networks for classification tasks, as they show better performance over traditional machine learning methods [1]. Applications such as voice-command recognition and text transcription have become second nature in our daily lives. In contrast to the popularity and acceptance of this technology, it is important to note that they are still considered “Black Boxes” and vulnerable to even the smallest of disturbances [2], [3]. The robustness and reliability of deep learning algorithms is a key topic in research, as ever progressing studies highlight the need for AI to be incorporated in the medical domain.

The standard for image classification in deep learning has been the use of convolution neural networks (CNN) due to their ability to identify relevant features from data. However, several studies [3], [4] have shown that they are vulnerable to slight pixel changes and are strongly dependent on the training data. In order to improve the robustness of these models a different approach is proposed in this work, whereby generative adversarial networks (GAN) are used to generate synthetic images by restructuring the unseen input data to match that of the training set. GAN [5] is a system composed of a generator and a discriminator network, where synthetic data is generated then compared to the real data to see if they match.

In this work, the feasibility of the proposed approach is studied for the recognition of human emotions. Facial emotion recognition (FER) is currently being considered as a method to help treat patients with autism spectrum disorder (ASD) a developmental brain disorder that affects the social interactions and communications of individuals [6]. Facial expressions have shown to convey 55% of a person’s feelings and attitudes [7]. A closedloop feedback system with a novel reward system is being developed which immerses the subject in a virtual reality environment i.e., a game, in which the user is subjected to different activities i.e. social interactions as well as emotional stimuli [8]. In [9]–[12] the use of Conditional GANs were studied and the results showed the possibilities of utilizing such systems in reducing the impact of variations of new unseen data on trained FER models.

Three facial emotion databases were selected for this study the OULU-CASIA [13], FACES [14], and Japanese female facial expressions (JAFFE) [15]. Images are first pre-processed to highlight the face of the subject by removing background noise according to the technique described in [16]. The proposed model is then trained to generate synthetic images, from the FACES and JAFFE datasets, that resemble part of the data from the OULU-CASIA database. The performance of the model is based on the outcome of the generated images and its similarity to the original data.

The aim of this study is to determine the feasibility of such a proposed model in improving CNN model robustness for FER.

2 Methods

Image pre-processing, developed in previous work [16], was implemented on the datasets of OULU-CASIA, FACES, and JAFFE to focus on the face of the subject and remove background noise. Three subjects from the OULU-CASIA were chosen as the input for real images and the images of the FACES and JAFFE were set as input into to system to generate synthetic images that match the real inputs. **2.1 Model Development**

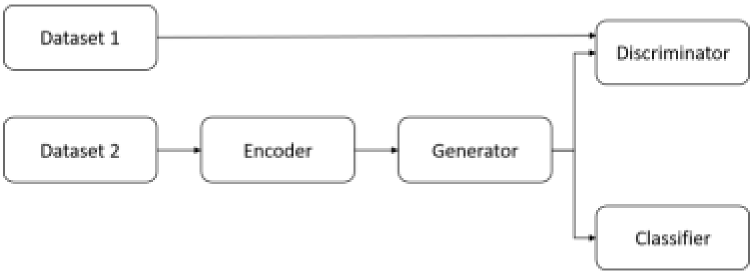

Fig. 1. Proposed GAN model

Figure 1 represents the flowchart of the proposed GAN system. The data first pass through an Encoder block (Enet) where the relevant features are extracted, the output is then taken as input to the Generator block (Gnet) which generates the synthetic images. After which the output is passed into the Discriminator block (Dnet) along with the labels which are embedded into the image, the Dnet distinguishes between the real and synthetic images. The outputs from the Gnet are also passed to the Classification block (Cnet) which classifies the inputs into the different emotion classes.

The Enet represents a shallow CNN with an architecture similar to that of Alexnet [17] coupled with normalization blocks after each convolution block. The Enet extracts different features from the input images of size 96x96x3 from Dataset 2 which refers to the FACES and JAFFE combined and outputs a 1x4096 feature vector. The Gnet also represents a shallow decoder network that takes in input from the Enet and outputs an image with the same size as the input images.

The Dnet takes in input the real images and labels of Dataset 1 which refers to the OULU-CASIA dataset, the generated images from Gnet and the labels from the Dataset 2. The labels are embedded into the image to make the input dimension 96x96x4. The Cnet represents a CNN that was trained on the OULU-CASIA data from previous work [16]. The system loss of Gnet is based on the logarithmic mean log likelihood function combined with the loss of the Cnet, which is computed by cross entropy loss. The Dnet loss is based on the logarithmic mean log likelihood function. The parameters of the Enet, Gnet and Dnet are updated after each iteration by means of the adaptive moment estimation (adam) optimization method. The Cnet parameters were not updated as the trained model showed high performance accuracies of 98% on its validation set [16].

2.2 Database Description

The Oulu-CASIA database is composed of 80 different subjects expressing six basic emotions of anger, disgust, fear, happiness, sadness and surprise. The database consists of image sequences beginning with Neutral expression and ending with strong emotion expression. Images of original RGB, visible light with strong illumination lighting were selected with a total of 10,379 images [13].

The Japanese female facial expressions (JAFFE) database is composed of 213 facial portrait images portrayed in grey scale from 10 different Japanese female students expressing seven emotions (six basic plus Neutral) [15]. The FACES dataset has a total of 2,052 images expressing six emotion classes of anger, disgust, fear, happiness, neutral and sadness of varying subject ages [14]. To analyze the performance of the model in generating synthetic images close to the real images, the classification performance was assessed by evaluating the true positive predictions from the different classification models developed in [16].

3 Results & Discussion

Table 1 shows the distribution of the images into each class for the FACES and JAFFE datasets. The image pre-processing algorithm excluded 8.77% and 1.41% of the images of FACES and JAFFE datasets respectively, from further processing due to the failure of the method to segment the prescribed regions [16]. The classes were distributed near equally with the Neutral class being removed from the analysis. Three subjects from the OULU-CASIA database were selected by random and set as the real images for training the system.

Table 1. Class Distribution before and after image pre-processing for the FACES and JAFFE databases.

	FACES			**JAFFE**		
Class	**Original**	**After Processing**	**% each Class***	**Original**	**After Processing**	**% each Class***
Anger	342	292	18.94	30	30	16.67
Disgust	342	292	18.94	29	29	16.11
Fear	342	303	19.65	32	32	17.78
Happiness	342	338	21.92	31	31	17.22
Sadness	342	317	20.55	31	31	17.22
Surprise	0	0	N/A	30	27	15.00
Total	**1710**	**1542**	**100.00**	**183**	**180**	**100.00**

*The percentage is calculated based on the Pre-processed image data

The mean performance of the classification models from [16] on the FACES and JAFFE datasets before and after synthetic image generation is represented in table 2. As seen from the results the GAN system was able to achieve a slightly better overall performance with a lower standard deviation. However, looking at the mean accuracies of each class it was seen that the GAN system was not able to generate the classes of Happiness, Sadness and Surprise correctly and lacked any predictive results. The performance of Anger, Disgust and Fear showed improvements of greater than 20% per class. This signifies that the GAN system was robust for these three classes especially taking into consideration the color variation between the datasets of FACES and JAFFE.

Table 2. Performance results of the true positive predictions on the testing set of FACES and JAFFE before and after synthetic image generation.

Mean ± SD %	**Before GAN**	**After GAN**
Anger	32.59 ± 16.38	72.41 ± 10.27
Disgust	73.83 ± 11.06	100.00 ± 0.00
Fear	73.87 ± 3.88	94.40 ± 1.97
Happiness	57.87 ± 13.05	0
Sadness	8.57 ± 1.86	0
Surprise	20.00 ± 17.21	0
Mean	**48.81 ± 4.31**	**49.39 ± 1.84**

4 Conclusion

In this study the feasibility of implementing a GAN system to improve FER robustness was analyzed. The preliminary results showed that the GAN system was able to improve the class predictions, however it lacked in the ability to generate the complete class set. The performance highlights the feasibility of this method and its practical applications in generating more data and improving the robustness of FER. More work is planned with better fine tuning of parameters and inclusion of more subjects for real images to improve on the existing results.

Author's Statement

Research funding: Partial support by a grant from the German Federal Ministry of Research and Education (BMBF) under project No. 13FH5I06IA – PersonaMed is gratefully acknowledged. Conflict of interest: Authors state no conflict of interest.

References

1. Y. LeCun, Y. Bengio, and G. Hinton, 'Deep learning', Nature, vol. 521, no. 7553, pp. 436–444, May 2015.
2. W. Samek, T. Wiegand, and K.-R. Müller, 'Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models', ArXiv170808296 Cs Stat, Aug. 2017
3. X. Yuan, P. He, Q. Zhu, and X. Li, 'Adversarial Examples: Attacks and Defenses for Deep Learning'. arXiv, Jul. 06, 2018.
4. N. Akhtar and A. Mian, 'Threat of Adversarial Attacks on Deep Learning in Computer Vision: A Survey'. arXiv, Feb. 26, 2018.
5. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.
6. C. on C. W. Disabilities, 'The Pediatrician's Role in the Diagnosis and Management of Autistic Spectrum Disorder in Children', Pediatrics, vol. 107, no. 5, pp. 1221–1226, May 2001.
7. A. Mehrabian, 'Communication without words', in Communication Theory, C. D. Mortensen, Ed. Routledge, 2017.
8. H. Arabian, V. Wagner-Hartl, J. Geoffrey Chase, and K. Moeller, 'Image Pre-processing Significance on Regions of Impact in a Trained Network for Facial Emotion Recognition', IFAC BMS 21, IFACPapersOnLine, Volume 54, Issue 15, 2021, Pages 299-303, ISSN 2405-8963, https://doi.org/10.1016/j.ifacol.2021.10.272.
9. J. Cai et al., 'Identity-Free Facial Expression Recognition Using Conditional Generative Adversarial Network', in 2021 IEEE International Conference on Image Processing (ICIP), Sep. 2021, pp. 1344–1348.
10. J. Chen, J. Konrad, and P. Ishwar, 'VGAN-Based Image Representation Learning for Privacy-Preserving Facial Expression Recognition'. arXiv, Sep. 07, 2018.
11. H. Yang, U. Ciftci, and L. Yin, 'Facial Expression Recognition by De-expression Residue Learning', in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, Jun. 2018, pp. 2168–2177.
12. H. Yang, Z. Zhang, and L. Yin, 'Identity-Adaptive Facial Expression Recognition through Expression Regeneration Using Conditional Generative Adversarial Networks', in 2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018), May 2018, pp. 294–301.
13. G. Zhao, X. Huang, M. Taini, S. Z. Li, and M. Pietikäinen, 'Facial expression recognition from nearinfrared videos', Image Vis. Comput., vol. 29, no. 9, pp. 607–619, Aug. 2011.
14. N. C. Ebner, M. Riediger, and U. Lindenberger, 'FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation', Behav. Res. Methods, vol. 42, no. 1, pp. 351–362, Feb. 2010.
15. M. J. Lyons, M. Kamachi, and J. Gyoba, 'Coding Facial Expressions with Gabor Wavelets (IVC Special Issue)', 2020.
16. H. Arabian, V. Wagner-Hartl, and K. Moeller, 'Image Pre-processing Effects on Attention Modules in Facial Emotion Recognition', IUPESM World Congress on Medical Physics and Biomedical Engineering (IUPESM WC2022), In Press.
17. A. Krizhevsky, I. Sutskever, and G. E. Hinton, 'ImageNet classification with deep convolutional neural networks', Commun. ACM, vol. 60, no. 6, pp. 84–90, May 2017.
18. H. Arabian, V. Wagner-Hartl, and K. Moeller, 'Traditional versus Neural Network Classification Methods for Facial Emotion Recognition', presented at the VDE BMT, 2021.
19. H. Arabian, V. Wagner-Hartl, and K. Möller, 'Facial emotion recognition based on localized region segmentation', Jun. 17, 2021. doi: 10.5281/zenodo.4922791.
20. H. Arabian, V. Wagner-Hartl, J. Geoffrey Chase, and K. Möller, 'Facial Emotion Recognition Focused on Descriptive Region Segmentation', in 2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC), Nov. 2021, pp. 3415–3418. doi: 10.1109/EMBC46164.2021.9629742.

21. H. Arabian, V. Wagner-Hartl, and K. Moeller, ‘Attention Modules for Facial Emotion Recognition Network Robustness Improvement’, In Press.
22. H. Arabian, V. Wagner-Hartl, J. Geoffrey Chase, and K. Möller, ‘Facial Emotion Recognition Focused on Descriptive Region Segmentation’, in 2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC), Nov. 2021, pp. 3415–3418. doi: 10.1109/EMBC46164.2021.9629742.
23. H. Arabian, T. Abdulbaki Alshirbaji, N. A. Jalal, N. Ding, B. Laufer, and K. Moeller, ‘Identifying User Adherence to Digital Health Apps’, presented at the IUPESM WORLD CONGRESS ON MEDICAL PHYSICS AND BIOMEDICAL ENGINEERING (IUPESM WC2022), in press.
24. H. Arabian, V. Wagner-Hartl, and K. Moeller, ‘Traditional versus Neural Network Classification Methods for Facial Emotion Recognition’, Current Directions in Biomedical Engineering, vol. 7, no. 2, pp. 203–206, Oct. 2021, doi: 10.1515/cdbme-2021-2052.
25. H. Arabian, V. Wagner-Hartl, and K. Moeller, ‘Network Architecture Influence on Facial Emotion Recognition’, In Press.
26. H. Arabian, V. Wagner-Hartl, and K. Moeller, ‘Transfer Learning in Facial Emotion Recognition: Useful or Misleading?’, In Press.