Facial Emotion Detection and Modification Using Generative Adversarial Networks

1st Sai Meghana Kolla
Department of Math and Computer Science
Penn State Harrisburg
Middletown, PA, United States
szk6163@psu.edu

2nd Dr. Hien Nguyen
Department of Math and Computer Science
Penn State Harrisburg
Middletown, PA, United States
nguyen.hien@psu.edu

***Abstract*—Facial emotion detection and facial emotion modification are applications of image processing, and with the use of Generative Adversarial Networks, these can be achieved. Rather than having multiple platforms, having a single platform hosting multiple applications of GAN would be a single stop for any user that wishes to take advantage of these features. In this paper, we aim to build an emotion detection GAN and an emotion modification GAN using the AffectNet data set and integrate both GANs into a single platform. Emotion detection GAN is made like a typical GAN with the addition of a classifier in the discriminator to classify the emotion of an input image. Emotion modification GAN is built using the StarGAN architecture. Both the GANs, after training, will be embedded into a single application that will be accessible to the user.**

***Index Terms*—GAN, Emotion Detection, Emotion Modification, Integration of GANs**

I. INTRODUCTION

With the massive increase in computer vision applications, the need for image processing is growing exponentially in various segments. Image processing is extracting information from an input image and processing it using an algorithm. An image is a group of pixels put together, and each pixel has a value associated with it to understand the opacity, color, and other features in it. By updating these values, we can modify an image.

Facial Emotion Detection and Modification(FEDM) is one of the applications where image processing comes into the picture. Facial emotion detection is the process of analyzing an image and identifying the facial emotion of the image. Facial emotion modification is the process of modifying the facial emotion of an image into a target emotion domain specified. In social media, people use FEDM as a filter on many platforms. There are several applications of emotion detection and modification in various fields, such as medical diagnostics to understand if the patient is suffering from depression or anxiety, intelligent automobiles to know how the driver feels while driving, security systems, and many more.

Although there are several deep neural networks for FEDM, the reason for using GAN here is to minimize the data loss during image processing[1]. GAN is made up of two deep neural networks called Generator and Discriminator. Both networks work together to get the job done. The job of the generator is to produce a fake image based on all the input images fed to it in such a way the discriminator should not identify that the image produced by the generator is fake. If the discriminator identifies the fake image, the image is fed back to the generator, and the iteration continues until the generator fools the discriminator. The generator and discriminator get better with every iteration, and thus, the continuous process of learning from the image helps in minimizing the data loss.

Many existing GANs, such as Exchange-GAN, StarGAN, and GANimation, are used for facial emotion detection and modification. In this project, StarGAN will be used, and a detailed explanation of why StarGAN is being used is in the literature review section.

The motivation behind the project is to come up with a user-interactive application where different applications of GAN can be put in a single platform without revealing the internal details. The project is being built on the hypothesis that facial emotion detection and modification can be achieved by a pipeline of GANs.

The data set used in the project is AffectNet[2]. The data set has images classified into eight expressions: neutral, happy, sad, surprise, fear, disgust, anger, contempt.

In this project, the idea is to create a pipeline of GANs so that the resulting model can perform both emotion detection and modification. The GANs built can be evaluated using the test data set from the AffectNet. We can calculate the model's accuracy and analyze the performance by running the trained GAN through the test data set. The detailed methodology will be explained in the following sections.

II. LITERATURE REVIEW

Various approaches for FEDM have been developed over the years. Traditional deep neural networks like Convolutional Neural Networks(CNN) are used in various image processing applications. In this section, we see how CNN is used for FEDM. Different GAN architectures, ExchangeGAN, GANimation, and StarGAN were discussed, which are used for FEDM.

A. Convolutional neural network:

A deep neural network is an artificial neural network with multiple hidden layers, and one of the popular deep neural networks is CNN. CNN has shown impressive results over the years in various applications from image processing to voice-recognition[3]. FEDM is one of the applications where CNN is widely used. However, there are more deep neural networks, such as GAN, developed which can also be used for FEDM. The deep paper CNN and deep GAN[12] give a list of applications where CNN is used and a list where GAN is used. After the analysis of the applications, it seems CNN can be used in applications when the image needs to be analyzed, and GAN is mostly used in applications when the data in the image should be generated or modified. In conclusion, GAN would fit best for FEDM.

B. Generative Adversarial Networks

Several GANs were developed for facial expression recognition and facial expression generation.

1) ExchangeGAN[11]: ExchangeGAN is a novel feature separation model used for facial emotion recognition. In this paper, the author builds a GAN to recognize the emotion of a facial image by separating the expression-independent features from expression-dependent features. Simply put, emotion recognition is done by ignoring the identity of facial features that do not account for expression. By doing this, the accuracy can be improved, and the training time can be reduced.

2) GANimation[4]: GANimation is one of the techniques used for facial emotion generation. In this paper, the author used action units to change the expression in a facial image. Action units (AUs) are the fundamental actions of individual muscles or groups of muscles. Ekman and Friesen analyzed the relationship between AU movement and facial expressions in the Emotion Facial Action Coding System (EMFACS)[5,6] and claimed that all AUs are external representations of muscle movements[7]. However, when the results of GANimation are compared to other GAN techniques, such as StarGAN, the results are not at par.

3) StarGAN[8]: StarGAN is an image-to-image translation technique that can support training the model on multiple datasets with different domains. It is trained on two datasets CelebA[9] and RaFD[10] parallelly. The attributes in CelebA and RaFD are different, but since StarGAN supports multiple datasets having different domains, the training is done with only one generator and one discriminator. Different datasets might have different attributes, and when the attributes that are not in the same dataset are needed to build an application, StarGAN will enable the model training on different datasets and produce a new facial image based on selected attributes. StarGAN also helps expand the model in the future by introducing additional attributes based on the requirements. Because of the flexibility and promising results, StarGAN would be a great fit for FEDM.

Comparing the results of GANimation and StarGAN, the results of GANimation changed the facial shape, whereas the results of StarGAN were better. In this project, StarGAN would best fit the requirement.

Although there are separate GANs for facial emotion detection and facial emotion modification, there is not any GAN that does both emotion generation and detection. In this paper, the aim is to build a single GAN or pipeline of GANs for FEDM. Using the StarGAN architecture mentioned above, a GAN for emotion modification is built. The details of the GAN for emotion modification and GAN for emotion detection are shown in the following methodology section.

III. METHODS

This section describes our approach to integrating various applications of GAN into a single platform. In this, we integrate facial emotion detection and facial emotion modification, two applications of GAN, into a single platform. The idea is to build GANs for each application and integrate them into the application. The UI will be built with Gradio[15] Package.

A. Emotion detection

For emotion detection, we built an Emotion Detection GAN (EDGAN). Figure 1 shows the architecture of an EDGAN.

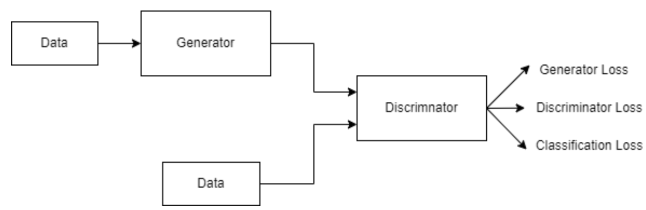

Fig. 1. Architecture of EDGAN

Like any regular GAN, the generator is fed with input data images. The generator reads the image, learns, and produces a fake image to fool the discriminator into believing that the fake image is an original image. The discriminator then reads the image produced by the generator and decides whether the image is fake or original.

If the discriminator identifies the image as fake, we call it generator loss. If the discriminator is fooled by the fake image, we call it discriminator loss. However, we introduce another loss here, classification loss. The discriminator is fitted with an extra classifier to detect the emotion of the input image. The classification loss on original images strengthens the discriminator, and the classification loss on fake images strengthens the generator.

1) Discriminator of EDGAN: The discriminator, along with identifying the fake images, also has to classify the emotion of the input image. For this purpose, the discriminator also has a classifier equipped. The discriminator uses CapsuleNet[14], a capsule network with capsules, to store information while processing the image.

2) Generator of EDGAN: The generator uses U-Net[13] architecture, where it has both an encoder and a decoder. The encoder extracts the features in the image and, through down-sampling, goes through the convolutional neural layers. The output from these layers is fed as input to the decoder, which by up-sampling, produces the image.

B. Emotion Modification

For emotion modification, we built a GAN using the StarGAN architecture mentioned in the literature review section. StarGAN is an existing architecture to build GAN mostly used for facial emotion modification. The EMGAN built can be trained on multiple domain data sets. However, in this project, the EMGAN will be trained with only one data set, AffectNet, which has all the eight emotion domains needed.

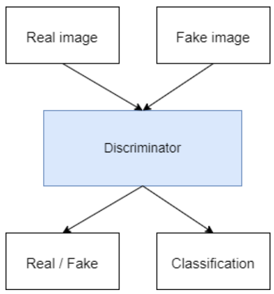

Fig. 2. Discriminator Training of EDGAN

1) Discriminator of EMGAN: Figure 2 shows the discriminator of the EMGAN. The real images are the original images in the data set, and the fake image is fed into the discriminator, a deep neural network. The discriminator processes the image and identifies if the image is real or fake. It also identifies the emotional domain of the image fed into it. To identify the emotion domain, the discriminator is equipped with a classifier similar to EDGAN. If the discriminator cannot identify the fake image, then it is called discriminator loss which helps to optimize the discriminator.

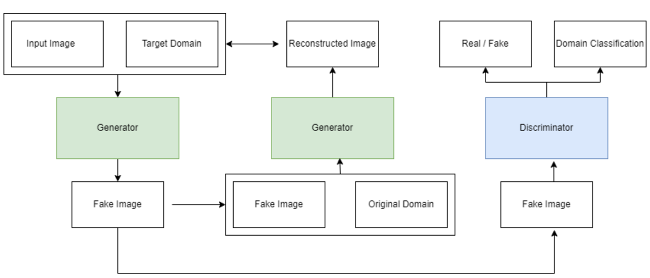

Fig. 3. Generator Training of EMGAN

2) Generator of EMGAN: Figure 3 shows the generator of the EMGAN. The input for the generator would be an image and a target domain, which is picked randomly from all the emotion domains present in the AffectNet data set. The generator then processes the input image with the target domain and produces an image with the target domain rather than the original domain. The image is then sent in again into the generator along with the original image to check if the initial input image is reconstructed. If the input image is not reconstructed, then the generator learns from it until it can generate a fake image which can reconstruct the original image when fed in with the original domain.

Once the fake image can reconstruct the original image, it is sent to the discriminator. The discriminator identifies whether the image is real or fake. If the generator is not able to fool the discriminator into believing that it is a real image, it is called generator loss which helps to optimize the generator.

Due to this iterative learning behavior of GAN, the data loss in an image while processing is reduced to its minimum.

C. Integration

In the integration phase of the project, both the GANs for emotion detection and modification are integrated as a single application for the user to use. The application will be a web application built using Gradio, a python package. Gradio provides a default user interface that can be used by any python or machine learning application. The application can then be hosted on gradio server for multiple users to access it. Figure 4 gives an outline of how the application works. The user will be able to access only the application and doesn't know any underground application mechanism. In this way, different applications of GAN can be put together into a single platform providing an abstraction of details from the user.

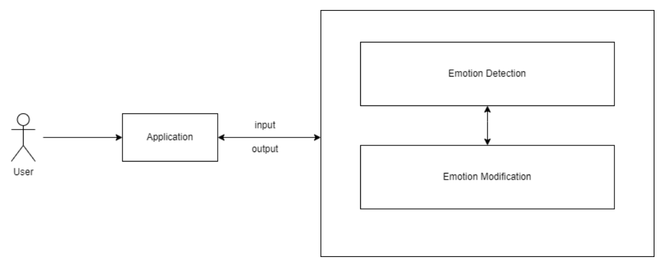

Fig. 4. Integration of EDGAN and EMGAN into a single application

After the training of EDGAN and EMGAN, testing the models is an essential step in evaluating the performance of the models trained. Initially, the AffectNet dataset is split into train data set and test data set. After the training, the test data set is run through EDGAN, and the results are compared with the actual results to calculate the accuracy of EDGAN. These accuracies give us the performance of EDGAN. Similarly, after the training, the test data set is run through EMGAN. For each image in the test data set, we modify the emotion and run these modified images through EDGAN and calculate the accuracies. Thus both the EMGAN and EDGAN performances can be evaluated. The performances of both EDGAN and EMGAN are displayed and discussed in the results section.

IV. Results

This section shows the evaluation results of facial emotion detection and facial emotion modification systems. A user interface is built integrating both EDGAN and EMGAN, where the user can upload an image and click any of the two buttons two either detect the emotion of the image or change the emotion to the specified emotion. Figure 5 shows the results of emotion modification in the user interface when an image has been uploaded. The user can either detect an emotion, modify the emotion in the image or do both operations simultaneously.

In this section,

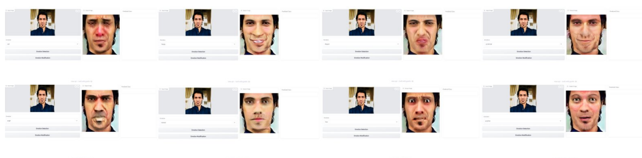

Fig. 5. User Interface

For training and testing the model, split the AffectNet data set in a 75-25 train-test ratio. 75% of the data set is used for training, and 25% is used for testing.

After training the model, to evaluate the performance of EDGAN, the model is run against the test data set. We then calculate the prediction accuracies for each emotion domain. Figure 7 shows a plot of the computed accuracies. The emotion domain sad has the lowest accuracy, and the emotion domain anger has the highest accuracy. However, there is little difference in the accuracies for all eight emotion domains.

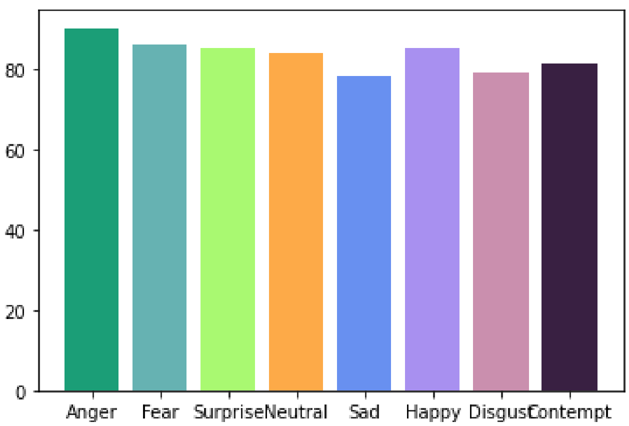

Fig. 6. Results of EDGAN

Similar to the EDGAN, to evaluate the performance of EMGAN, each image in the test data set is modified into eight emotions, and the resulting images are stored in different folders. These altered emotion images are now tested on the emotion detection model trained earlier, and the accuracies were calculated. Figure 6 shows a plot of the computed accuracies. The emotion domain fear has the lowest accuracy, and the emotion domain happy has the highest accuracy. However, there is little difference in the accuracies for all eight emotion domains.

V. Conclusion

In this paper, we built two GANs, one for facial emotion detection and one for facial emotion modification. In an attempt to integrate different applications of GAN, we built an application where the user can interact with the application and take advantage of the two GANs built. The user can upload an

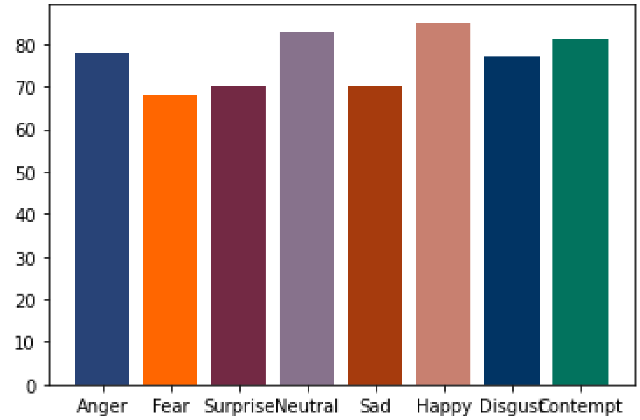

Fig. 7. Results of EMGAN

image and either detect the emotion of the image or modify the emotion of an image into a new emotion that can be specified in the user interface.

As discussed in the results section, EDGAN has shown a good performance overall. The best accuracy is obtained for anger emotion, and the least is for the sad domain. For EMGAN, the best accuracy is for happy emotion, and the least is for fear and sad domain. Although the performance of EDGAN and EMGAN is good, we might improve the performance by adding additional steps in the data preprocessing step.

As part of future work, we can improve the data preprocessing of the images in the data set by removing unnecessary features like background. The performance might also increase if we could add an inconclusive category while predicting the emotion domain. The inconclusive category decreases the False Negatives and False Positives. We can also extend the emotional palette of the GAN without limiting it to eight domains in the AffectNet data set.

References

[1] M. -Y. Liu, X. Huang, J. Yu, T. -C. Wang and A. Mallya, "Generative Adversarial Networks for Image and Video Synthesis: Algorithms and Applications," in Proceedings of the IEEE, vol. 109, no. 5, pp. 839-862, May 2021, doi: 10.1109/JPROC.2021.3049196.

[2] Mollahosseini, Ali & Hasani, Behzad Mahoor, Mohammad. (2017). AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild. IEEE Transactions on Affective Computing. PP. 10.1109/TAFFC.2017.2740923.

[3] S. Albawi, T. A. Mohammed and S. Al-Zawi, "Understanding of a convolutional neural network," 2017 International Conference on Engineering and Technology (ICET), 2017, pp. 1-6, doi: 10.1109/ICEngTechnol.2017.8308186.

[4] Pumarola, A., Agudo, A., Martinez, A.M., Sanfeliu, A., Moreno-Noguer, F. (2018). "GANimation: Anatomically-Aware Facial Animation from a Single Image". In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds) Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science(), vol 11214. Springer, Cham. https://doi.org/10.1007/978-3-030-01249-650.

[5] Benitez-Quiroz CF, Srinivasan R, Martinez AM (2016) EmotioNet: an accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild [C]// the IEEE conference on computer vision and pattern recognition (CVPR). IEEE:5562–5570

[6] Friesen W, Ekman P (1983) EMFACS-7: emotional facial action coding system. Unpublished manual, University of California, California

[7] Yao, L., Wan, Y., Ni, H. et al. Action unit classification for facial expression recognition using active learning and SVM. Multimed Tools Appl 80, 24287–24301 (2021).https://doi.org/10.1007/s11042-021-10836-w
[8] Y. Choi, M. Choi, M. Kim, J. -W. Ha, S. Kim and J. Choo, "StarGAN: Unified Generative Adversarial Networks for Multi-domain Image-to-Image Translation," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 8789-8797, doi: 10.1109/CVPR.2018.00916.
[9] Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015.
[10] O. Langner, R. Dotsch, G. Bijlstra, D. H. Wigboldus, S. T. Hawk, and A. Van Knippenberg. Presentation and validation of the radboud faces database. Cognition and Emotion, 24(8):1377–1388, 2010.
[11] Yang, Lie, et al. "A Novel Feature Separation Model Exchange-Gan for Facial Expression Recognition." Knowledge-Based Systems, vol. 204, 2020, p. 106217., https://doi.org/10.1016/j.knosys.2020.106217.
[12] Abirami, R. Vincent P M, Durai Srinivasan, Kathiravan Tariq, Usman Chang, Chuan-Yu. (2021). Deep CNN and Deep GAN in Computational Visual Perception-Driven Image Analysis. Complexity. 2021. 1-30. 10.1155/2021/5541134.
[13] Y. Wang, "Single Image Super-Resolution with U-Net Generative Adversarial Networks," 2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), 2021, pp. 1835-1840, doi: 10.1109/IMCEC51613.2021.9482317.
[14] Sabour, Sara, Nicholas Frosst, and Geoffrey E. Hinton. "Dynamic routing between capsules." Advances in neural information processing systems 30 (2017).
[15] Abid, Abubakar & Abdalla, Ali & Abid, Ali & Khan, Dawood & Alfozan, Abdulrahman & Zou, James. (2019). "Gradio: Hassle-Free Sharing and Testing of ML Models in the Wild."