How far Generated Data Can Impact Neural Networks Performance?

Sayeh Gholipour Picha[a], Dawood Al Chanti[b] and Alice Caplier[c]

Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France

Keywords: Facial Expression Recognition, Generative Adversarial Networks, Synthetic Data.

Abstract: The success of deep learning models depends on the size and quality of the dataset to solve certain tasks. Here, we explore how far generated data can aid real data in improving the performance of Neural Networks. In this work, we consider facial expression recognition since it requires challenging local data generation at the level of local regions such as mouth, eyebrows, etc, rather than simple augmentation. Generative Adversarial Networks (GANs) provide an alternative method for generating such local deformations but they need further validation. To answer our question, we consider noncomplex Convolutional Neural Networks (CNNs) based classifiers for recognizing Ekman emotions. For the data generation process, we consider generating facial expressions (FEs) by relying on two GANs. The first generates a random identity while the second imposes facial deformations on top of it. We consider training the CNN classifier using FEs from: real-faces, GANs-generated, and finally using a combination of real and GAN-generated faces. We determine an upper bound regarding the data generation quantity to be mixed with the real one which contributes the most to enhancing FER accuracy. In our experiments, we find out that 5-times more synthetic data to the real FEs dataset increases accuracy by 16%.

1 INTRODUCTION

Deep learning (DL) has achieved high accuracy performance in various complex tasks including recognition (Rakesh et al., 2022), detection (Zhou et al., 2022), localization (Grumiaux et al., 2022), etc. Yet despite its success, it requires large amounts of labeled data, especially if high performance is required. For instance, considering a Facial Expression Recognition (FER) model trained on a specific Facial Expressions (FEs) dataset, it would not perform as well when applied to a moderately different real-world dataset. This is due to the distribution shift coming from a lack of diversity and biases in the datasets against certain demographic changes (Drozdowski et al., 2020) such as race, gender, and age.

Biases in the training data prone trained models towards overfitting as they are optimized over the majority samples (e.g. certain age) represented in the dataset. Hence a low performance is expected over minor samples (e.g. certain races). To address this issue, we argue that having at disposal a diverse dataset would help in overcoming such biases and building a generalizable model. However, acquiring and la-

beling image and video data is a very expensive and time-consuming task and sometimes it is not even feasible. In this paper, we study the impact of synthetic data generation on the performance of neural networks. We propose to alleviate the bias issue by testing a data augmentation procedure able to generate balanced and diverse data samples.

Several works (da Silva and Pedrini, 2015), (Gu et al., 2012), (Hasani and Mahoor, 2017), and (Zavarez et al., 2017) routinely performed standard data augmentation using affine transformation (e.g., translation, scaling, rotation, reflection, shearing, cropping, etc.). Standard augmentation does not bring any new information to enrich the training dataset to solve the bias problem. On the contrary, Generative adversarial networks (GANs) (Goodfellow et al., 2014) offer the opportunity, to increase the amount of training samples, and to enrich the diversity of the final training set under certain experimental data generation process. In this paper, we consider an FER task and we address and evaluate the use of generated synthetic FEs via GANs to compensate the lack of diversity in FE training databases in an attempt to reduce the bias of the considered FER model and to increase its generalization ability.

Here, we consider a classical CNN classification scheme as it is not our intention to build a novel classi-

---

[a] https://orcid.org/0000-0003-2675-5463
[b] https://orcid.org/0000-0002-6258-6970
[c] https://orcid.org/0000-0002-5937-4627

fier. However we carefully design the data augmentation scheme based on combining multiple GANs that consider generating: i) new and diverse FEs with new identities and races, various genders, and different ages; ii) various FEs deformation intensities, which makes the generated facial expressions closer to spontaneous human behavior; and iii) balanced dataset where we guarantee that each identity gets the same amount of generated images per emotion class.

To this end, our contributions are:

- We design a method to generate diverse and balanced facial expression deformations.

- We empirically investigate the contribution of synthetic data and their role in improving DL performance.

- We perform a cross-database evaluation to estimate fairly the impact of generated data on the generalizability of the trained model.

The paper is structured as follows: Section 2 discusses related works; Section 3 presents the proposed procedure of building an FER system based on augmented data; Section 4 discusses the experimental results; Finally, Section 5 concludes the paper.

2 RELATED WORKS

In most traditional research in facial expression recognition, the combination of face appearance descriptors used to represent facial expressions with deep learning techniques is considered to overcome the difficult factors for FER. Regardless, due to the small size of public image-labeled databases, Data Augmentation (DA) techniques are often used to increase the size of the database. In addition to DA geometric transformations, more complex guided augmentation methods can be used for DA, such as GAN. In (Yi et al., 2018), a conditional GAN is used to generate images to augment the FER2013 dataset. A CNN is used to train the predictive model, and the average accuracy increased by 5% after applying the GAN DA technique. (Chu et al., 2019) proposed an FER method based on Contextual GAN. Chu's model uses a contextual loss function to enhance the facial expression image and a reconstruction loss function to retain the subject's identity information in the expression image. Experimental results with the extended CK+ database (Lucey et al., 2010) show that Chu's method improves recognition performance by 7%. However, neither Yi's nor Chu's studies perform cross-database evaluation nor consider the generation of balanced synthetic FEs classes. (Porcu et al., 2020) experimented with the combination of various data

augmentation approaches, such as using synthetic images, and discovered that a combination of synthetic data with horizontal reflection, and translation can increase the accuracy by approximately 30%. They performed cross-database evaluations by training their model on an "augmented" KDEF database (Lundqvist et al., 1998) and testing it on two different databases (CK+ and ExpW (Zhanpeng Zhang and Tang, 2016)). Unlike them, we design our method to consider a diverse but balanced generation of FE classes and create our experimental setup to resemble fair performance metrics.

3 DATA MODALITY

Generative Adversarial Networks are used to generate different FEs for training our FER algorithm. Our model design splits into three different compartments: the data generation stage, the CNN classifier training stage, and the inference stage.

3.1 Dataset Generation Process

Our data generation process relies on using two GANs on top of each other. One is for new identity generation while the other is used to impose the generation of local FEs. First, we generate new identities with new facial features using the StyleGAN model of (Karras et al., 2020) that randomly generates realistic human faces. Additionally, since we want to compare the performance of our FER model trained with both real or generated facial features, we build a database that resembles existing public databases. In those public datasets, subjects pose different expressions in front of a fixed-setting camera. For this reason, we build a novel method that jointly uses StyleGAN and StarGAN on top of each other as a way to reinforce the FEs generation process over new identities. However, due to the randomness of the StyleGAN model and the desirability of a balanced training set, we use the structure of a StarGAN model (Choi et al., 2017) for image-to-image translation with different settings to artificially synthesize the six Ekman emotions (anger, disgust, fear, happiness, sadness, and surprised) on a single generated identity. We train the StarGAN model with the spontaneous public database Affectnet-HQ (Mollahosseini et al., 2019) since this database captures images from various settings, and from lots of people through the internet. We use the trained model to generate facial expressions on both real face images and StyleGAN generated face images as shown in figure 1. The final result of using the image-to-image translation Star-

Table 1: Model summary of the considered CNN-based model for Facial Expression Recognition.

	Layer	input	number of filters	Pool size	Activation function
1	Conv2D	$64 \times 64 \times 1$	32		Relu
2	Conv2D	$64 \times 64 \times 32$	64		Relu
3	Max Pooling	$64 \times 64 \times 64$		2×2	
4	Drop out 25%				
5	Conv2D	$32 \times 32 \times 64$	128		Relu
6	Max Pooling	$32 \times 32 \times 128$		2×2	
7	Conv2D	$16 \times 16 \times 128$	128		Relu
8	Max Pooling	$16 \times 16 \times 128$		2×2	
9	Drop out 25%				
10	Flatten				
11	Dense **1024**				Relu
12	Drop out 50%				
13	Dense **6**				Softmax

GAN model to synthesize different expressions for a given real or generated identity is shown in figure 2. As we can see, there are several artifacts in the output images, which are mainly found on the outer part of the face. However, these artifacts are not important in our task since we only focus on facial features for facial expression recognition. During this process, we generated 100,000 identities and synthesized 6 basic emotions on each of them. Finally, with some preprocessing (face cropping, gray-scale, and resizing), we generated the balanced dataset illustrated in figure 3.
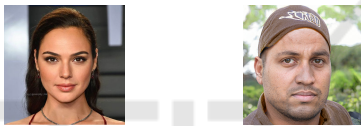

(a) Real human facial features. (b) StyleGAN Generated identity.

Figure 1: Two samples of facial features.

3.2 Convolutional Neural Network

In the second stage of our method, we design a CNN classifier whose architecture is summarized in table 1. Our purpose is to use a simple yet effective classifier in order to focus our attention on the contribution of GAN-generated images with respect to model enhancement. To avoid the overfitting effect, we use drop-out layers.

In the first experiment, the CNN classifier is trained on the two facial expression public databases (RaFD (Langner et al., 2010), and Oulu-CASIA (Zhao et al., 2011)) labeled with the 6 basic Ekman emotions. In the second experiment, the CNN model is trained again from scratch using only generated facial expressions. These two control experiments serve as baselines. Finally, we re-train the CNN model again by gradually augmenting the public databases with images of generated FEs with the same number of identities each time.

3.3 Testing Phase

To fairly evaluate the performance of our method, we split the real and generated database to create a test database. About 17% of the data that are created are used for testing purposes. Although these test datasets are necessary to assess the model's performance, the similarity in distribution between the test and train sets makes it difficult to determine the exact contribution of the generated data to the model's performance. To ensure a fair analysis of the results and prevent bias in model prediction, it is important for our settings to perform a cross-database evaluation in which our test data resemble zero correlation with training sets. Hence, to have a fixed reference test dataset to compare all the models, we use the MMI database (Pantic et al., 2005) which is completely blind to the training process.

4 RESULTS AND ANALYSIS

In the following, we present the results of three experiments with different settings and designs. RaFD (Langner et al., 2010) and Oulu-CASIA (Zhao et al., 2011) real databases have been used in experiments 1 and 3. Alternatively, in experiment 2 we use only synthetic data, and these synthetic data are also used in experiment 3 for further analysis.

4.1 Experiment 1 - Training with Real Data

In this first baseline experiment, the CNN classifier for the FER model has been trained with real face images coming from RaFD and Oulu-CASIA which in total have 144 subjects. Here, we use 109 subject images with 6 basic emotions for training, 10 for validation, and 25 for testing. From these data, we only consider the frontal faces that are associated with emotional labels.

Our trained CNN classifier achieves 69.6% of accuracy when it is tested over 25 subjects. However, we observe overfitting because our model achieves 84.5% in the training phase. By applying a cross-database evaluation using the MMI database (Pantic et al., 2005), the obtained accuracy drops to 45% which is expected due to the limited number of subjects we have in the training dataset. Next, we analyze each class separately to get an insight into the classes' separability. Figure 4 presents the associated confusion matrix. Based on this map, the "Happy" emotion has the best performance and the "Disgust" class is also showing a good performance compared to other

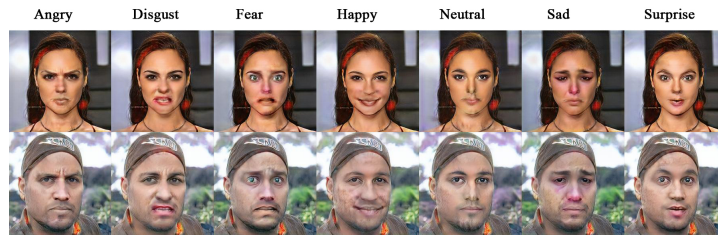| Angry | Disgust | Fear | Happy | Neutral | Sad | Surprise |

Figure 2: Examples of synthetic facial expressions. In the top row, the StarGAN model acts on an actual human face. In the bottom row, the model works with a generated identity.
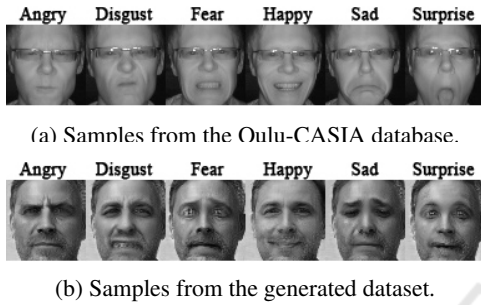

(a) Samples from the Oulu-CASIA database.

(b) Samples from the generated dataset.

Figure 3: Samples of generated FEs either on a real face or on a generated one.

Figure 4: Confusion Matrix on the MMI database in experiment one.

classes of emotions. While this confusion matrix provides a comprehensive overview of each class, it is unnormalized, making later comparisons difficult. To analyze the results of the cross-database evaluation on the MMI database in a more detailed manner, in a second step, we measure three metrics (precision, recall, and F1-score) to explore the model's prediction with the annotations provided for the MMI database. Figure 5 presents these metrics results for each class of emotions. It can be noticed that the model trained on real facial features has the most difficulty at recognizing the "Sad" class, while the other classes are not showing a good performance either. Despite having two real databases for training, this model fails to perform adequately. As a result, adding more data to the training is necessary and we argue that adding synthetic data might aid in overcoming the overfitting and also in lifting up the accuracy rate.

4.2 Experiment 2 - Training with Synthetic Generated Data

In the second experiment, we follow the same protocol. The only difference is that we are using synthetic facial images as the training dataset. We consider the same number of synthetic identities as in the real dataset in the first experiment (109 identities). These identities are generated with the process presented in section 3.1. For each identity, all six basic emotions exist in the dataset. Training our model with this synthetic dataset, the accuracy reaches 99.84% during
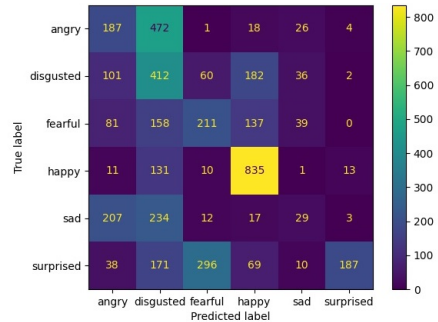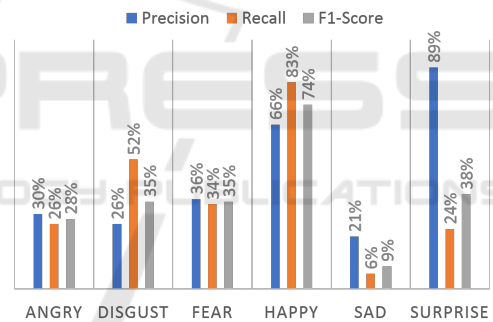

Figure 5: Precision, recall, and F1-score on the MMI database obtained on the CNN model trained with real faces only (cf. experiment 1).

the training process and 97.6% while testing on the synthetic dataset. Also, no overfitting is observed in this experiment. Although these results show a significant improvement, performing the cross-database evaluation on the MMI database is not that promising. On the MMI database, the obtained accuracy drops to 47% showing nearly the same performance as the model trained in experiment 1.

Figure 6 presents the confusion matrix of the model trained with the synthetic dataset for the MMI database to discover whether the model trained on synthetic data has similar classifying difficulties as the model trained on real data. Comparing the Confusion Matrix in figures 4 and 6 we notice:
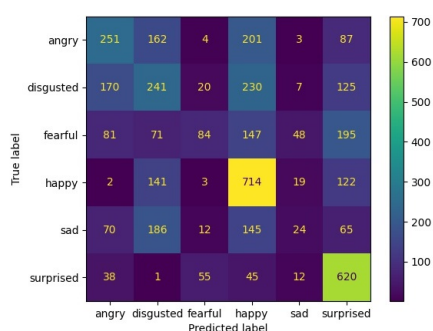
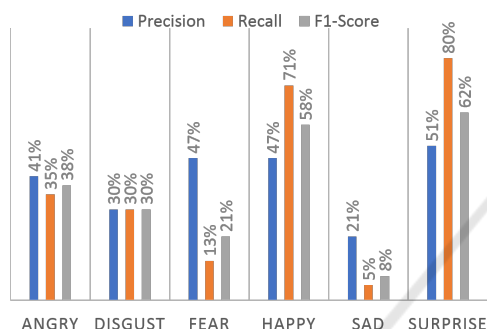Figure 6: Confusion Matrix on MMI database in experiment two.

Figure 7: Precision, recall, and the F1-score on the MMI database for the model trained in experiment two (section 4.2).

1. There has been a huge improvement in recognizing the class "Surprised", 620 samples instead of 187 samples.

2. We can observe improvements in the recognition of the "Angry" class, 251 samples instead of 187.

3. As compared to the CNN classifier trained on real faces, we see some drop in the "Disgust", "Fear", and "Sad" classes, but both models seem to have similar difficulties.

Also based on the presentation of precision, recall, and F1-score in figure 7, it appears the recall scores have decreased for most classes. We can therefore say that except for the "Surprised" class, the CNN model trained on synthetic data alone is unable to match the actual facial expressions annotations provided in the MMI database. The results of our current experiment prove that synthetic datasets can achieve similar performance as real datasets. Hence our final aim is to increase the dataset size to improve the performance at all class levels. In this case, we hope that the combination of these two databases will help solve such problems.

Table 2: Accuracy of the model trained on augmented datasets (Real Facial Expressions (RFEs) augmented by Generated Facial Expressions (GFEs)).

	Training accuracy	Testing accuracy
RFEs + GFEs	91%	85.3%
RFEs + 2 × GFEs	93.8%	89%
RFEs + 3 × GFEs	94.8%	92.7%
RFEs + 4 × GFEs	95.8%	92.5%
RFEs + 5 × GFEs	97.6%	94.3%
RFEs + 6 × GFEs	97.8%	94%
RFEs + 10 × GFEs	97.9%	95.1%
RFEs + 15 × GFEs	98.9%	95.5%
RFEs + 20 × GFEs	98.9%	97%

4.3 Experiment 3 - Training with Augmented Datasets

In the last experiment, we augment the Real Facial Expressions (RFEs) dataset of experiment 1 with Generated Facial Expressions (GFEs). The number of generated identities in each unit is the same as the number in the real database used for experiment 1 (109 identities for training, 10 identities for validation, and 25 identities for testing). As an example, RFEs + 2 × GFEs is the extension of the real FEs with two units of generated FEs (109 real identities + 218 generated identities for training). Each of the augmented datasets is split into training, validation, and test sets. And the CNN model is trained on each dataset individually. Each augmented dataset is represented in table 2 indicating the model accuracy during training and testing. The results demonstrate that adding more synthetic FEs to the training set results in constant improvement of training and testing accuracies. The study also reports no evidence of overfitting.

The cross-database evaluation on the MMI database is then performed for further validation and figure 8 shows the accuracy obtained from each trained model. Note that the first two points are the result of cross-database evaluation obtained in experiments 1 and 2 respectively. According to this figure, the highest accuracy corresponds to the model trained on the RFEs + 5 × GFEs dataset with 58.3%. This performance from the model trained on the 5th augmented dataset indicates a 13% gain in response to the model trained in experiment 1 (with real FEs). But beyond this point, the accuracy drops significantly due to a catastrophic forgetting mode caused by the large number of synthetic facial features in the training set overwhelming the facial features of the real face database.

To see the improvement of our best model (RFEs + 5 × GFEs) in each class separately, we present the confusion matrix and the calculated precision, recall, and F1-score metrics on the MMI database in figures 9 and 10 respectively. It appears that the "Sad" class performs significantly better than the two base-
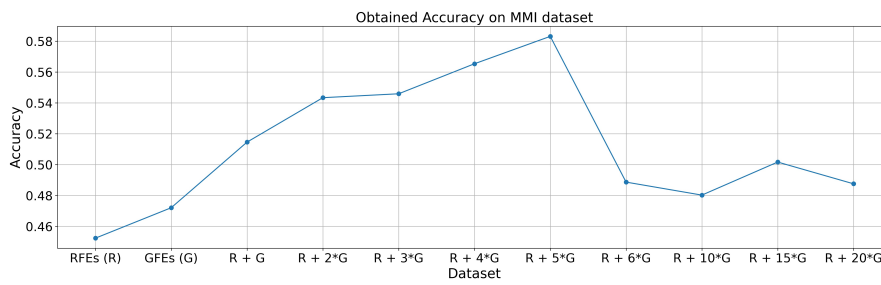
Figure 8: The result of cross-database evaluation on the MMI database. RFEs is referring to Real Facial Expressions and GFEs is referring to Generated Facial Expressions.
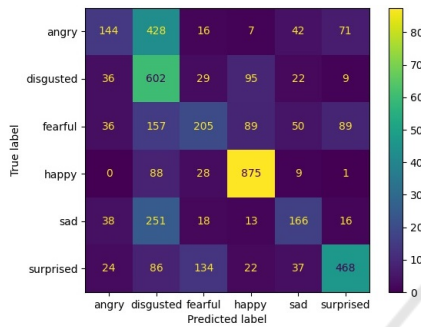

Figure 9: Confusion Matrix on MMI database in experiment three with the model trained on RFEs + 5 × GFEs database.
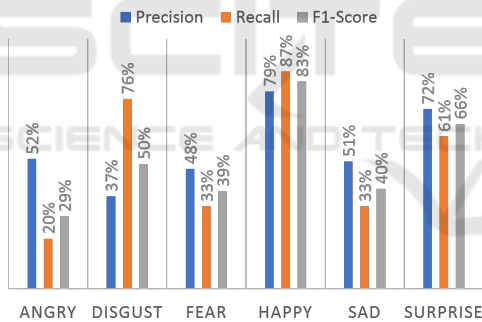

Figure 10: Precision, recall, and F1-score on the MMI database for the model trained on the RFEs + 5 × GFEs database in experiment three (section 4.3).

line experiments in all three metrics. Based on the recall scores in all classes, we can conclude that this trained model matches facial expressions in the MMI database to their actual annotations better than other trained models. In contrast with the model trained in experiment 1 (training set of real FEs), only the "Anger" class's performance decreases. In conclusion, based on our observation, we can say that generated data along with the real facial features is helping the model's recognition ability.

Furthermore, there is no limit to the number of identities we can generate. But there is a point beyond which adding new generated FEs no longer improves the results. We have observed experimentally

that there is an upper limit in augmentation in relation to the size of the real face database.

4.4 Comparison with the State-of-the-Art

As a final step in this study, we compare our results with state-of-the-art findings. We use the VGG16 tool to calculate the accuracy of the FER VGG16 model on the MMI database. With that model, we achieve 54.08% accuracy while our best CNN-based model reaches 58.3% in accuracy. Through the use of synthetic facial features and a simpler model, we enhance the accuracy by 4%.

Many state-of-the-art studies have reported their evaluation results on the CK+ database. Nevertheless, we did not use the CK+ database in our training or testing processes. Therefore, in order to perform the comparison, we evaluate our best model performance on the CK+ database and the result is presented in table 3. It can be seen that the approach proposed by (Zavarez et al., 2017) is the only one that outperforms our proposed CNN model. However, the difference is only 1.09% while they trained their model using 6 different public databases and some classical data augmentation techniques. Whilst our results are achieved with smaller training datasets using only two public databases and GAN images which makes our results more outstanding. In addition, compared to the study in (Porcu et al., 2020) that is explained in section 2, even though their model's accuracy increased by 30%, our model had more promising results based on this cross-database evaluation.

Figure 11 shows the result of our CNN-based model in the cross-database evaluation on the CK+ database. For this database, we achieve an accurate model for most of the classes even though there is no record of this public database in our training data.

During this study, we replaced the MMI database with the CK+ database for cross-database evaluation. As a result, model performance increased by 16% rather than the 13% gain we previously achieved.

Table 3: Comparison among state-of-the-art cross-database experiments tested on the CK+ database.

|	Method	Training Database	Accuracy
1	Proposed method	RaFD + Oulu-CASIA +GAN	**87.49%**
2	(Porcu et al., 2020)	KDEF	83.30%
3	(da Silva and Pedrini, 2015)	MUG	45.60%
4	(da Silva and Pedrini, 2015)	JAFFE	48.20%
5	(da Silva and Pedrini, 2015)	BOSPHOROUS	57.60%
6	(Lekdioui et al., 2017)	KDEF (Lundqvist et al., 1998)	78.85%
7	(Gu et al., 2012)	JAFFE	54.05%
8	(Hasani and Mahoor, 2017)	MMI+FERA	73.91%
9	(Mollahosseini et al., 2016)	MultiPIE (Gross et al., 2008), MMI, CK+, DISFA (Mavadati et al., 2013), FERA (Valstar et al., 2017), SFEW (Dhall et al., 2011), and FER2013	64.20%
10	(Zavarez et al., 2017)	CK+, JAFFE, MMI, RaFD, KDEF, BU-3DFE (Yin et al., 2006), and ARFace (Martinez and Benavente, 1998)	**88.58%**
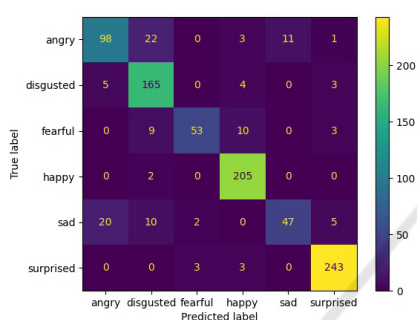
Figure 11: Confusion Matrix on CK+ database.

While it is undeniable that generated data is a cost-less method that can help improve FER model accuracy, the exact gain would be determined by the test databases in applications.

5 CONCLUSIONS

The purpose of this study was to investigate how generated data could be used to augment the data in a deep learning model to improve its performance. We chose a simple facial expression recognition model for this proposal. Our synthetic balanced dataset was created using two GAN models to test the potential improvement of the FER model performance. With real databases, synthetic datasets, and augmented datasets, we trained the CNN classifier multiple times for the FER task.

Our study confirms that enriching the training dataset with GAN images can improve CNN classifier performance. Training and cross-database evaluation performances were improved by augmenting real databases with synthetic facial features. In comparison to a model trained solely from real facial images, our best model shows a 16% increase in accuracy. On the same database, we also compared our results with the state-of-the-art and computed the accuracy of the

VGG16 model, achieving 4% higher accuracy.

For further study, we propose to first augment the training database with additional real facial expressions. This will enable us to improve the performance of the model, as it would also let us augment more GAN images. Secondly, we propose to enrich the VGG16 database with our generated dataset to see if we can improve the performance of the VGG16 model as well. And third, we would like to study the potential performance increase for other applications related to facial models.

Material and Codes. Results can be reproduced using the code available in the GitHub repository https://github.com/sayeh1994/synthesizin_facial_expression and https://github.com/sayeh1994/Facial-Expression-Recognition.

REFERENCES

Choi, Y., Choi, M., Kim, M., Ha, J.-W., Kim, S., and Choo, J. (2017). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation.

Chu, Q., Hu, M., Wang, X., Gu, Y., and Chen, T. (2019). Facial expression recognition based on contextual generative adversarial network. In *2019 IEEE 6th International CCIS*, pages 120–125.

da Silva, F. A. M. and Pedrini, H. (2015). Effects of cultural characteristics on building an emotion classifier through facial expression analysis. *Journal of Electronic Imaging*, 24(2):1 – 9.

Dhall, A., Goecke, R., Lucey, S., and Gedeon, T. (2011). Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark. In *2011 IEEE ICCV Workshops*, pages 2106–2112.

Drozdowski, P., Rathgeb, C., Dantcheva, A., Damer, N., and Busch, C. (2020). Demographic bias in biometrics: A survey on an emerging challenge. *IEEE SSIT*, 1(2):89–103.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., and Weinberger, K., editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc.

Gross, R., Matthews, I., Cohn, J., Kanade, T., and Baker, S. (2008). Multi-pie. In *8th IEEE FG 2008*, pages 1–8.

Grumiaux, P.-A., Kitić, S., Girin, L., and Guérin, A. (2022). A survey of sound source localization with deep learning methods. *ASA journal*, 152(1):107–151.

Gu, W., Xiang, C., Venkatesh, Y., Huang, D., and Lin, H. (2012). Facial expression recognition using radial encoding of local gabor features and classifier synthesis. *Pattern Recognition*, 45(1):80–91.

Hasani, B. and Mahoor, M. H. (2017). Spatio-temporal facial expression recognition using convolutional neural networks and conditional random fields. In *12th IEEE FG 2017*, pages 790–795.

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., and Aila, T. (2020). Analyzing and improving the image quality of StyleGAN. In *Proc. CVPR*.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., and van Knippenberg, A. (2010). Presentation and validation of the radboud faces database. *Cognition and Emotion*, 24(8):1377–1388.

Lekdioui, K., Messoussi, R., Ruichek, Y., Chaabi, Y., and Touahni, R. (2017). Facial decomposition for expression recognition using texture/shape descriptors and svm classifier. *Signal Processing: Image Communication*, 58:300–312.

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., and Matthews, I. (2010). The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *2010 IEEE CVPRW*, pages 94–101.

Lundqvist, D., Flykt, A., and Öhman, A. (1998). The karolinska directed emotional faces—kdef.

Martinez, A. and Benavente, R. (1998). The ar face database. *Tech. Rep. 24 CVC Technical Report*.

Mavadati, S., Mahoor, M., Bartlett, K., Trinh, P., and Cohn, J. (2013). Disfa: A spontaneous facial action intensity database. *Affective Computing, IEEE Transactions on*, 4:151–160.

Mollahosseini, A., Chan, D., and Mahoor, M. H. (2016). Going deeper in facial expression recognition using deep neural networks. In *2016 IEEE WACV*, pages 1–10.

Mollahosseini, A., Hasani, B., and Mahoor, M. H. (2019). AffectNet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing*, 10(1):18–31.

Pantic, M., Valstar, M., Rademaker, R., and Maat, L. (2005). Web-based database for facial expression analysis. In *2005 ICME*, pages 5 pp.–.

Porcu, S., Floris, A., and Atzori, L. (2020). Evaluation of data augmentation techniques for facial expression recognition systems. *Electronics*, 9.

Rakesh, R. K., Namita, G. R., and Kulkarni, R. (2022). Image recognition, classification and analysis using convolutional neural networks. In *2022 First ICEEICT*, pages 1–4.

Valstar, M. F., Sánchez-Lozano, E., Cohn, J. F., Jeni, L. A., Girard, J. M., Zhang, Z., Yin, L., and Pantic, M. (2017). FERA addressing head pose in the third facial expression recognition and analysis challenge. *Proc Int Conf Autom Face Gesture Recognit*, 2017:839–847.

Yi, W., Sun, Y., and He, S. (2018). Data augmentation using conditional gans for facial emotion recognition. In *2018 PIERS-Toyama*, pages 710–714.

Yin, L., Wei, X., Sun, Y., Wang, J., and Rosato, M. (2006). A 3d facial expression database for facial behavior research. In *7th International FGR 2006*, pages 211–216.

Zavarez, M. V., Berriel, R. F., and Oliveira-Santos, T. (2017). Cross-database facial expression recognition based on fine-tuned deep convolutional network. In *2017 30th SIBGRAPI*, pages 405–412.

Zhanpeng Zhang, Ping Luo, C. C. L. and Tang, X. (2016). From facial expression recognition to interpersonal relation prediction.

Zhao, G., Huang, X., Taini, M., Li, S. Z., and Pietikäinen, M. (2011). Facial expression recognition from near-infrared videos. *Image and Vision Computing*, 29(9):607–619.

Zhou, Z., Zhang, J., and Gong, C. (2022). Automatic detection method of tunnel lining multi-defects via an enhanced you only look once network. *Computer-Aided Civil and Infrastructure Engineering*, 37(6):762–780.