Performance comparison of synthetic face databases on the Xception model for recognizing genuine and generated facial images

Filip Dobrev[1], Anna Lukina[1]

[1]Technical University of Delft

1. Introduction

Text-to-image generation models may be useful and interesting to use, but they can pose serious threats. The creation of fake images of people can be used to spread slander, create fake news, or scam people. The wide accessibility of this technology necessitates countermeasures for detecting synthetic images.

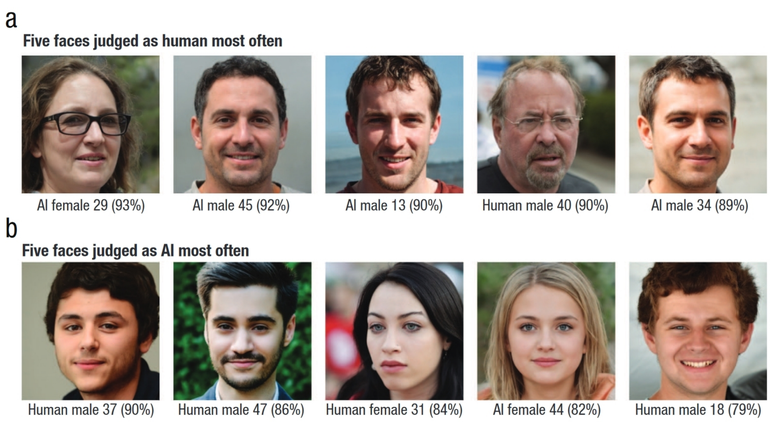

Figure 1. The results from a study asking people to classify synthetically generated and real faces. Top 5 faces most often regarded as human and as AI. Courtesy of [1]

2. Research Question & Methodology

Research Question: Performance comparison of synthetic face databases on the Xception model for recognizing genuine and generated images

- **Literature review** Conducting a comprehensive review of the literature on synthetic image classification and synthetic face databases
- **Analysis & Experimentation** Analyzing the different facial databases and experimenting with them on the Xception model
- **Results** Summarizing my findings for the database along with the results of the experiments.

3. Model

Xception has shown promising results when it comes to detecting facial manipulations compared to other models, as can be seen in Figure 2

Compression	Raw	HQ	LQ
[14] XceptionNet Full Image	82.01	74.78	70.52
[27] Steg. Features + SVM	97.63	70.97	55.98
[17] Cozzolino *et al.*	98.57	78.45	58.69
[10] Bayar and Stamm	98.74	82.97	66.84
[51] Rahmouni *et al.*	97.03	79.08	61.18
[5] MesoNet	95.23	83.10	70.47
[14] XceptionNet	**99.26**	**95.73**	**81.00**

Figure 2. Results of multiple models on facial image classification. Courtesy of [2]

4. Databases

iFakeFaceDB

iFakeFaceDB [3] is a database compiled from other already existing databases of real and synthetic facial images. Its contribution is the GANprintR autoencoder, which removes the GAN "fingerprints" while not changing the RGB representation of the final image.

Figure 3. Examples of fake facial images from iFFDB

Diverse Fake Face Dataset (DFFD)

Diverse Fake Face Dataset (DFFD) [4] is the biggest dataset I have looked at - 2.6 million facial images. It is very varied as it includes facial manipulations such as identity and expression swap, attribute manipulation, and entire facial synthesis. Examples can be seen in Figure 4

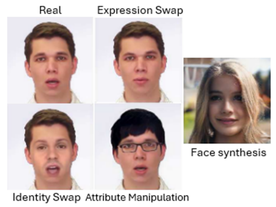

Figure 4. Different types of facial manipulation present in the Diverse Fake Face Dataset - Courtesy of [4]

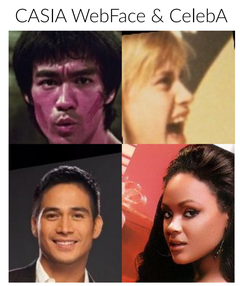

Figure 5. Examples of real facial images from Casia Webface(top two) and CelebA(bottom two)

Figure 6. Example of our transformation of the images from CelebA and CASIA

5. Results

Training & Validation Data		Testing Data		Accuracy	
Real	Fake	Real	Fake		
DFFD(FFHQ) - 10,999	DFFD - 10,999	DFFD - 9,000	DFFD - 8,997	97.11%	
DFFD(FFHQ) - 10,999	DFFD - 10,999	CelebA - 3,000	iFFDB - 20,974	98.07%	73.15%
DFFD(FFHQ) - 10,999	DFFD - 10,999	CASIA - 3,000		96.87%	
CASIA - 16,680	iFFDB - 20,000	CASIA - 3,000	iFFDB - 3,000	99.96%	
CASIA - 16,680	iFFDB - 20,000	DFFD - 9,000	DFFD - 8,997	50.01%	
CelebA - 19,995	iFFDB - 20,000	CelebA - 20,006	iFFDB - 20,974	100.00%	
CelebA - 19,995	iFFDB - 20,000	DFFD - 9,000	DFFD - 8,997	52.30%	

Figure 7. The results of our experiments with different databases. The numbers next to the databases are the number of images used for training and testing.

6. Limitations

- Different models use different generation approaches, complicating detection.
- Training data quality varies, affecting how convincing the images are.
- Data generation is time-consuming and demanding.
- Longer training times didn't allow us to experiment with more hyperparameters and databases

Figure 8. Different facial conditions that might bypass classifiers

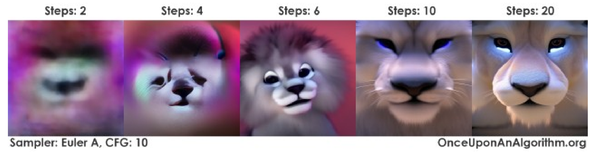

Figure 9. Diffusion generation process

7. Conclusions

- Xception model shows good performance in detecting AI-generated
- Reliable performance across multiple tests, indicating model reliability.
- Need for diversified datasets including underrepresented groups to address dataset biases.
- Investigation of dataset-specific fingerprints.
- Expansion to general human detection, beyond faces.
- Real-time detection applications in social media.

8. References

[1] E. J. Miller, B. A. Steward, Z. Witkower, C. A. M. Sutherland, E. G. Krumhuber, and A. Dawel, "Ai hyperrealism: Why ai faces are perceived as more real than human ones," *Psychological Science*, vol. 34, no. 12, pp. 1390–1403, 2023, pMID: 37955384. [Online]. Available: https://doi.org/10.1177/09567976231207095

[2] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "Faceforensics++: Learning to detect manipulated facial images," 2019.

[3] J. C. Neves, R. Tolosana, R. Vera-Rodriguez, V. Lopes, H. Proenca, and J. Fierrez, "Ganprintr: Improved fakes and evaluation of the state of the art in face manipulation detection," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 5, p. 1038–1048, Aug. 2020. [Online]. Available: http://dx.doi.org/10.1109/JSTSP.2020.3007250

[4] H. Dang, F. Liu, J. Stehouwer, X. Liu, and A. Jain, "On the detection of digital face manipulation," 2020.

https://www.linkedin.com/in/filipdobrev/

f.dobrev@student.tudelft.nl