Data Augmentation for Convolutional Neural Network DeepFake Image Detection

Ameni Jellali
Research Laboratory
Smart Electricity & ICT,
SEICT, LR18ES44,
National Engineering School
of Carthage,
University of Carthage.
Tunis, Tunisia.
Email: jellaliameni@enicar.ucar.tn

Ines Ben Fredj
Research Laboratory
Smart Electricity & ICT,
SEICT, LR18ES44,
National Engineering School
of Carthage,
University of Carthage.
Tunis, Tunisia.
Email: ines_benfredj@yahoo.fr

Kais Ouni
Research Laboratory
Smart Electricity & ICT,
SEICT, LR18ES44,
National Engineering School
of Carthage,
University of Carthage.
Tunis, Tunisia.
Email: kais.ouni@enicar.ucar.tn

2023 IEEE International Conference on Advanced Systems and Emergent Technologies (IC_ASET) | 979-8-3503-2102-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/IC_ASET58101.2023.10150803

***Abstract*—We need to develop a technique for better identifying deepfakes because they can distort our perception of reality. This study offers a brand-new forensic technique for spotting falsified facial photos. We made advantage of the Kaggle-provided "real-and-fake-facial-detection" dataset. We are able to distinguish between probable facial alterations based on CNN's design. Thanks to data augmentation approaches, the results exhibit performances that are equivalent to those of previous works. The proposed approach fared better for this binary categorization into fake or real faces than the other cutting-edge studies. Our accuracy is close to 99 percent.**

keywords: CNN, Deepfakes Detection, Deep Learning, Data Augmentation, Faces Manipulations.

I. Introduction

Since the invention of the first photograph in 1825, images have been altered [1].
Specially with the growth of social media and the extensive use of cell phones and digital images as the most popular digital assets today.
Defensive technology will continue to require significant advancements in order to effectively counter the rate of change in media manipulation and forensically address ever-evolving media acquisition and generation tactics.
Extracting real or bogus content, in particular, enables the appearance of specific kinds of computer applications.
These programs were created to identify the authenticity of digitally altered data, often known as deepfakes or fake news in the literature. Cybersecurity is under risk because of the rise of hypertrucage use and the absence of relevant legislation. Cause of artificial intelligence developments have made it easier to create deepfakes that are incredibly convincing.
In this context, deepfake detection is a growingly crucial topic in computer vision research. According to Richard Zhang, an Adobe researcher," we live in a world where it is becoming increasingly difficult to trust the digital information we consume" [2].
This paper will resolve the problem of deepfakes detection, covering the background theory needed for this challenge on digital signal processing knowledge, its approaches, and related works. Then we will describe the dataset we used, and finally, we will presnet our proposed model's structure and the results obtained comparing with other works.

II. State of the art

Digital manipulation has received a lot of attention recently, particularly after the phrase "DeepFakes" gained popularity. This chapter introduces the primary digital changes focusing on facial content due to the large range of potential uses. We thoroughly go over the principles of five distinct digital facial image modifications [1]:

A. Face synthesis

As seen in Figure 1, this modification produces entire nonexistent face representations. Typically because of a potent Generative Adversarial Networks (GAN), such as the most recent StyleGAN technique proposed in [3].
These methods produce facial images with a high degree of realism and produce outstanding outcomes. Many companies, including the video game and 3D modeling industries, might profit from this manipulation. However, it could also be employed negatively, such as fabricating plausible-looking phony profiles on social media sites to spread false information.

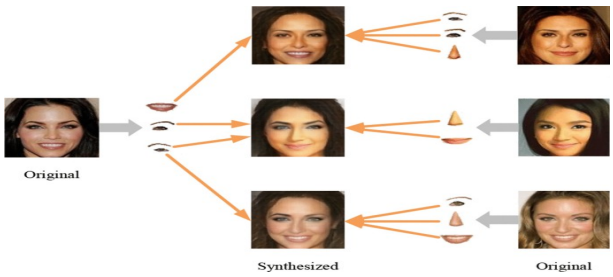

Fig. 1. Examples of manipulation with full facial synthesis.

B. Identity swap

As seen in Figure 2, this manipulation is swapping out a person's face in a video or photograph for another person's face.
Two distinct strategies are typically taken into account:
- Classic infographic techniques such as FaceSwap [4].
- New deep learning techniques known as DeepFakes, for example the recent ZAO mobile app [5].

Fig. 2. Examples of handling with expression change.

C. Face morphing

The transformation of one image into another during the morphing process can be described as a unique effect. Fig. 3 illustrates the process of integrating two facial images to create a single altered image.
One of the many and completely free tools, such as MorphThing [6], 3Dthis Face Morph [7], Face Swap Online [8],FantaMorph [9], FaceMorpher [10], and MagicMorph [11] [12]. or Abrosoft, can be used to effortlessly morph objects.

Fig. 3. Examples of handling with identity change.

D. Attribute manipulation

This alteration, often referred to as face editing or facial retouching, involves changing particular facial characteristics including the color of one's hair or complexion, their sex, their age, whether they have spectacles, etc [13], as illustrated in Figure 4. One example of this kind of manipulation is the mobile app FaceApp. With the aid of this technology, customers may virtually try on a variety of things, including eyeglasses, cosmetics, and hairstyles.

Fig. 4. Example of Attribute Manipulation.

E. Expression swap

The person's face expression is changed during this manipulation, commonly referred to as "facial reconstruction," as shown in Figure 5. However, many modification methods are put forth in the literature, like as at the image level using well-liked GAN structures. [14].

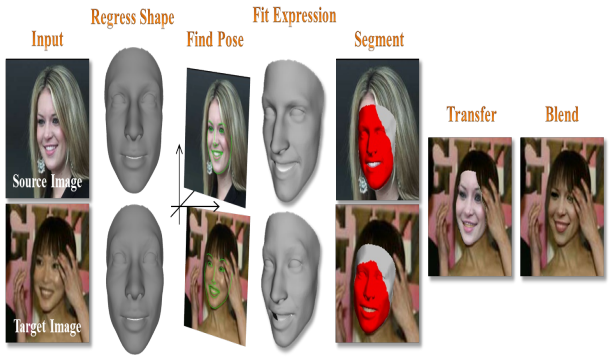

Fig. 5. Examples of handling with expression change.

III. Proposed CNN approach of fake face image detection

A. Convolutional Neural Network

The Convolutional Neural Network, also known as CNN, is one of the most potent Deep Learning algorithms. CNNs are networks of convolutive neurons, which are potent programming models that enable image recognition by automatically assigning to each provided input image a label corresponding to its class of membership.

B. Real-and-fake-face-detection dataset

This dataset contains about 2000 files divided into objective and fake face images [15]. The Yonsei University Department of Computer Science has made the benchmark deepfake dataset available to the public on Kaggle [16]. The deepfake dataset contains professionally edited facial photography. The resulting deepfake images combine many faces, split by the nose, eyes, mouth, and entire face. The collection contains 960 artificial faces and 1081 real faces. [17].

Fig. 6. Real and fake faces.

Results from deep learning models are heavily influenced by the amount and quality of the training data.
As a result, a variety of data augmentation techniques have been used to enhance the training dataset.

C. Data Augmentation processing

Data quality is crucial since successful training and an efficient model depend on a representative data collection.
Data augmentation, often known as ”data expansion,” is the process of adding modified copies of already-existing data or entirely new synthetic data that is derived from existing data to data analysis. There are geometric and color space augmentation options for images to add picture diversity to the model as shown in the figure 7. Numerous coding examples for these augmentation tweaks can easily be found in publications and open source libraries.

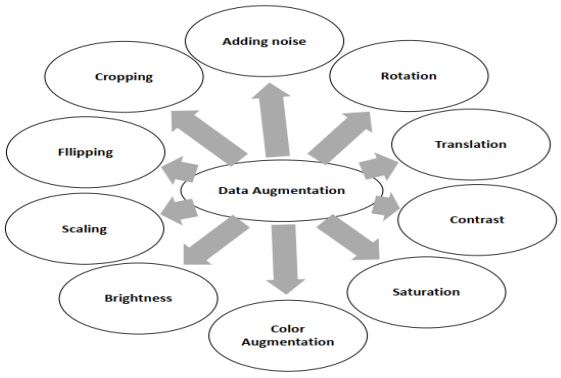

Fig. 7. Data augmentation techniques in computer vision.

The data transformation we performed in this work consisted of applying processing to each image to create new images and variations of the original image. For instance, as seen in the following figure, we can discuss rotation, contrast, and cropping:

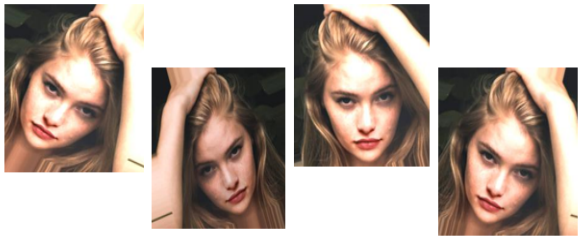

Fig. 8. Example of transformation applied in our dataset.

By lowering the possibility of overfitting, our dataset, which was obtained through data augmentation, contains about 2000 file, is valuable since it can increase the predicted accuracy and overall performance of our model.

D. Proposed architecture of real and fake image classification

The fake and real faces are arranged into a dataset by the goal label. The structured deepfake dataset is divided into train and test data portions as shown in the figure 9.

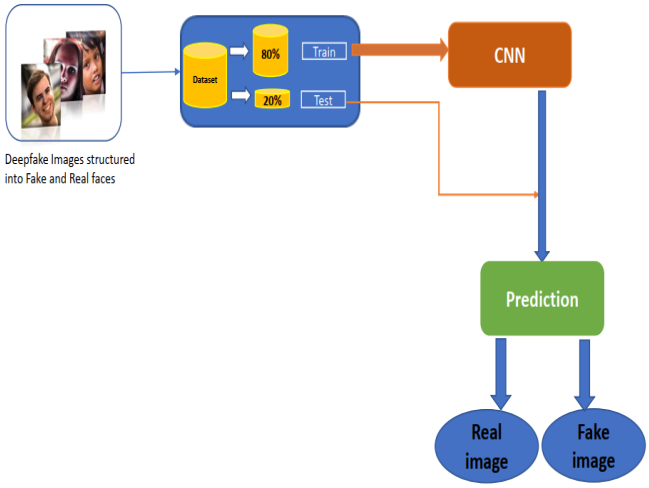

Fig. 9. The architectural analysis of the face image manipulation.

The neural network methods are trained using the dataset’s 80 percent train segment. The outperforming cutting-edge DFP technique, which has been fully hyper-parametrized, produces the highest accuracy score in deepfake face detection.
20% of the dataset is used to evaluate the performance of neural network algorithms on unlabeled test data. With highly accurate results, the innovative proposed approach provides predictions on unobserved data. A sophisticated deep learning-based system that has been proposed is generalized and prepared for use in detecting false and authentic faces.

The architecture of CNN model is built using 4 layers of Maxpooling and 4 layers of convolution. The input image, which has the dimensions 224*224*3, first goes through the first layer of convolution. Each layer is followed by a Relu activation function that compels neurons to provide positive results. Following this convolution, 32 features size maps (32*32) will be produced. This layer is made up of 32 size filters (3*3). The image size and the number of parameters and calculations are then reduced by using Maxpooling. 32 features with a size of 16x16 will be present after this layer is finished. While varying the number of filters, the approach is used four times. Following these convolution layers, we employ a network of neurons comprised of two Fully Connected layers. The activation function employed in the first layer's 128 neurons is the Relu. The distribution probability of the 2 classes is calculated in the last layer using the Softmax function.

E. Results and Discussion

Model accuracy is calculated as the proportion of classifications that a model correctly predicts over all predictions. On the train and test sets, the model's classification accuracy was reported to be about 98.46% and 99.44%, respectively. The classification of the train and test sets and the learning curves of the loss on the train and test sets are represented by two line plots that are built and displayed in the figure 10 and figure 11. According to the charts, the model satisfactorily addresses the problem.

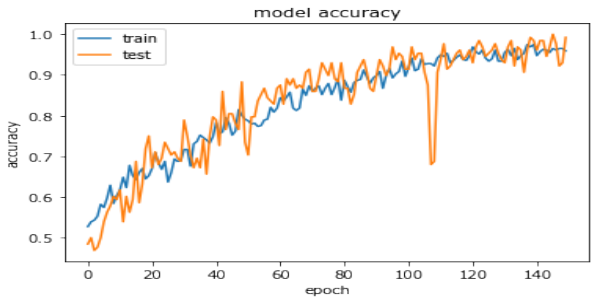

Fig. 10. The accuracy classification on the train and test sets.

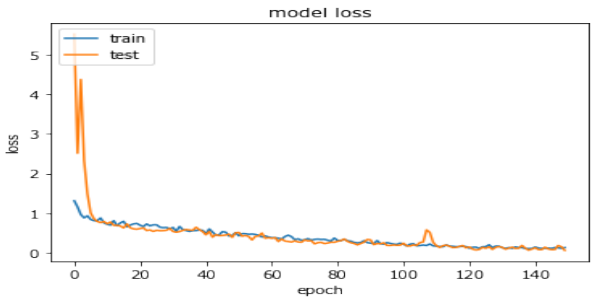

Fig. 11. The learning curves of the loss on the train and test sets.

While the validation accuracy varies with time, training accuracy increases in a predictable way. Yet, the accuracy of training and validation is remarkably similar. This indicates that the model was trained correctly and is capable of making additional predictions.

IV. Related Works

Table 1 examines the performance comparison of our proposed DFP with the other cutting-edge studies. For comparison, the most recent state-of-the-art methods from 2020 to 2022 are used. In our research investigation, the cutting-edge method we developed using our deepfake dataset augmented based on the conventional convolutional neural network. The analysis shows that our innovative approach fared better than existing state-of-the-art investigations.

TABLE I
SOME RELATED WORKS.

Reference	Technique	Accuracy
Cao.X et al [18]	CNN	89%
Ye.M et al [19]	VGG16	90%
Phiphiphatphaisit. S et al [20]	Mobile net	88%
Raza.A et al [21]	Novel DFP	94%
Proposed	CNN	99%

V. Conclusion

In this study, we investigated the subject of deepfakes detection, which, like all other areas of image processing, has seen significant development and attracted a lot of attention since the development of deep learning. With CNN architecture, we evaluated a limited dataset. Through this study, we were able to put data augmentation techniques into practice, allowing us to enhance our dataset and improve its accuracy. We can provide examples as we develop the model to better identify deepfake videos. Additionally, we use video datasets to train our model and verify our results.

References

[1] Rathgeb, Christian and Tolosana, Ruben and Vera-Rodriguez, Ruben and Busch, Christoph,2022,Handbook Of Digital Face Manipulation And Detection: From DeepFakes to Morphing Attacks,Springer Nature
[2] Korshunova, I., Shi, W., Dambre, J., Theis, L. (2017). Fast face-swap using convolutional neural networks. In Proceedings of the IEEE international conference on computer vision (pp. 3677-3685).
[3] T. Karras, S. Laine et T. Aila, 2019, A Style-Based Generator Architecture for Generative Adversarial Networks , IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, p. 4396-4405
[4] Priadana, A., Habibi, M. (2019, March). Face detection using haar cascades to filter selfie face image on instagram. In 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT) (pp. 6-9). IEEE.
[5] Morph thing. https://www.morphthing.com/, 2020. Accessed: October 2020
[6] 3dthis face morph. https://3dthis.com/morph.htm, 2020. Accessed: October 2020
[7] Face swap online. https://faceswaponline.com/, 2020. Accessed: October 2020
[8] Abrosoft fantamorph.FantaMorph,Abrasoft:http:/www.fantamorph.com/, 2020. Accessed: May 2020.
[9] Face morpher. http://www.facemorpher.com/, 2020. Accessed: October 2020

[10] Magic morph 1.95. https://downloads.tomsguide.com/magicmorph,0301-6817.html, 2020. Accessed: October 2020.
[11] Raja, Sushma Venkatesh Raghavendra Ramachandra Kiran and Busch, Christoph,2022, Face Morphing Attack Generation , Detection: A Comprehensive Survey.
[12] E. Gonzalez-Sosa, J. Fierrez, R. Vera-Rodriguez et F. Alonso-Fernandez, Facial Soft Biometrics for Recognition in the Wild : Recent Works, Annotation, and COTS Evaluation , IEEE Transactions on Information Forensics and Security, t. 13, no 8, p. 2001-2014, 2018.
[13] M. Liu, Y. Ding, M. Xia, X. Liu, E. Ding, W. Zuo et S. Wen, STGAN : A Unified Selective Transfer Network for Arbitrary Image Attribute Editing , in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, p. 3668-3677.
[14] Raza, Ali and Munir, Kashif and Almutairi, Mubarak, 2022 ,A Novel Deep Learning Approach for Deepfake Image Detection,Applied Sciences, pages 9820, MDPI
[15] YONSEI UNIVERSITY. Real and Fake Face Detection—Kaggle. Available online: https://www.kaggle.com/datasets/ ciplab/real-and-fake-face-detection (accessed on 14 July 2022).
[16] Shikha Agrawal · Kamlesh Kumar Gupta · Jonathan H. Chan · Jitendra Agrawal · Manish Gupta, Machine Intelligence and Smart Systems, Proceedings of MISS, 2021, Springer
[17] Tran, V. N., Lee, S. H., Le, H. S., Kwon, K. R. (2021). High Performance deepfake video detection on CNN-based with attention target-specific regions and manual distillation extraction. Applied Sciences, 11(16), 7678.
[18] Cao, X.; Yao, J.; Xu, Z.; Meng, D. Hyperspectral image classification with convolutional neural network and active learning. IEEE Trans. Geosci. Remote Sens. 2020, 58, 4604–4616.
[19] Ye, M.; Ruiwen, N.; Chang, Z.; He, G.; Tianli, H.; Shijun, L.; Yu, S.; Tong, Z.; Ying, G. A Lightweight Model of VGG-16 for Remote Sensing Image Classification. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2021, 14, 6916–6922.
[20] Phiphiphatphaisit, S.; Surinta, O. Food image classification with improved MobileNet architecture and data augmentation. In Proceedings of the 2020 The 3rd International Conference on Information Science and System, Cambridge, UK, 19–22 March 2020; pp. 51–56.
[21] Raza, A., Munir, K., Almutairi, M. (2022). A Novel Deep Learning Approach for Deepfake Image Detection. Applied Sciences, 12(19), 9820.