An Approach of Fake Videos Detection Based on Haar Cascades and Convolutional Neural Network

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

Kaïs 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_ASET58101.2023.10150604

***Abstract*—Because deep fakes might skew our impression of the truth, we need to come up with a method for better spotting them. This paper proposes a new forensic technique to detect manipulated facial images from videos. It is based on CNNs architecture that can distinguish possible face manipulations in the "real-and-fake-face-detection" dataset offered by Kaggle. The results obtained highlight comparable performances with the state-of-the-art methods. It showed an accuracy of approximately 99 % for this binary classification into fake or real faces. Then to validate this model we added a human face detection technique using the Haar Cascade method to this model in order to detect the manipulated videos from Deep Fake Detection Challenge (DFDC) dataset and we achieve an accuracy of 91 correct predictions out of 100 total videos.**

keywords: CNN, Deepfakes Detection, Deep Learning, Haar-Cascade, Data Augmentation, Faces Manipulations, Fake and real videos

I. Introduction

Nowadays, the widespread adoption of cell phones and Digital photos have become the most widely used digital assets due to the expansion of social media since the invention of the first photograph in 1825 [1]. Many challenges still exist. Forensically addressing constantly advancing media acquisition and creation techniques and thwarting the rate of change in media manipulation will continue to necessitate sizable technical advances in defensive technology [2]

In particular, extracting real or fake content allows the appearance of certain types of computer applications. These applications were designed to detect the realism of digital manipulation data, which is presented in the literature as deepfakes or fakes-news [3]. The increase in the use of hypertrucage and the lack of legislation around this issue pose a cybersecurity threat [4]. Advances in artificial intelligence have facilitated the production of highly persuasive deepfakes. Detecting deepfakes is an increasingly important area of research in computer vision. As Adobe researcher Richard Zhang said," we live in a world where it is becoming increasingly difficult to trust the digital information we consume" [5]. This paper will present 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 introduce our proposed model's structure and the results obtained.

II. State of the art

To achieve good results, we spent a great deal of time reading and revising publications, articles, and books to see and understand the concepts and how to apply the deep learning models to our problem.

Digital manipulation has become a hot topic in recent years, especially 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

This modification produces entire nonexistent face representations. Typically, a robust Generative Adversarial Network (GAN), such as the recently described StyleGAN technique in [6].

These methods produce facial images with a high degree of realism and produce outstanding outcomes. Many industries, including those involved in video games and 3D modeling, could profit from this manipulation. But, it could also be employed negatively, such as fabricating plausible-looking phony profiles on social media sites to spread false information.

B. Identity swap

This manipulation involves replacing a person's face in a video or image with another person's face, as illustrated in Figure 2.

Two different approaches are generally considered:

- Classic infographic techniques such as FaceSwap [7].
- New deep learning techniques named DeepFakes, for example the recent ZAO mobile app.

Fig. 1. Examples of handling with expression change.

Fig. 3. Example of Attribute Manipulation.

C. Face morphing

One way to describe the morphing process is as a special effect that turns one image into another. The procedure of creating a single altered image by combining two facial photographs is shown in Fig. 3.
One of the many and completely free tools, such as MorphThing [8], 3D this Face Morph [9], Face Swap Online [10],FantaMorph [11], FaceMorpher [12], and MagicMorph [13] [14]. or Abrosoft, can be used to effortlessly morph objects.

Fig. 2. Examples of handling with identity change.

D. Attribute manipulation

This manipulation, also known as face editing or facial retouching, involves modifying certain facial attributes such as hair or skin colour, sex, age, adding glasses, etc [14], as illustrated in Figure 4. The FaceApp mobile app is an example of this type of manipulation. With the aid of this technology, customers may virtually try on a variety of things, including eyeglasses, cosmetics, and hairstyles.

E. Expression swap

This manipulation, also known as 'facial reconstruction', involves modifying the person's facial expression, as illustrated in Figure 5. However, different manipulation techniques are proposed in the literature, for example, at the image level through popular GAN architectures [15].

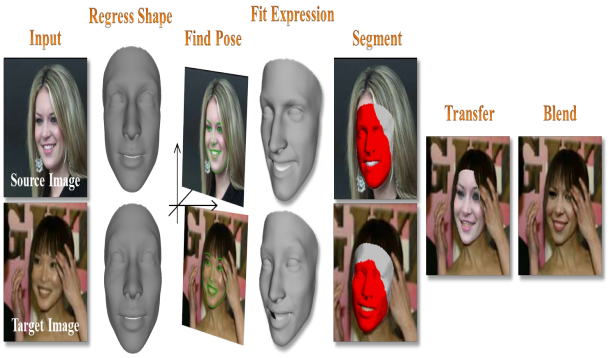

Fig. 4. Examples of handling with expression change.

III. Proposed CNN approach of fake face image detection

A. Convolutional Neural Network

In this section, we'll concentrate on Convolutional Neural Networks (or CNN), one of Deep Learning's most potent algorithms: Convolutive neural networks are potent programming models that enable image identification by automatically providing a label relating to each image's input class to each image.

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

This dataset contains about 2000 files divided into objective and fake face images [16]. The Yonsei University Department of Computer Science has made the benchmark deepfake dataset available to the public on Kaggle [17]. The deepfake dataset includes manipulated facial photographs created by professionals. Multiple faces, divided by nose, eyes, mouth, and complete face, are combined in the deepfake images that are produced. The dataset includes 1081 genuine faces and 960 synthetic faces [18].

Fig. 5. Real and fake faces.

The training set's quality and size greatly influence deep learning models' results.
Thus, many different data augmentation approaches have been applied to improve the training dataset.

C. Data Augmentation processing

Data quality is paramount since a representative data set is inseparable from successful training and an effective model. In data analysis," data augmentation" expands the amount of data by adding changed versions of either existing data or brand-new synthetic data derived from existing data. In this work, we have carried out a data transformation, which will be a matter of applying processing on each image to generate new images and variants of the initial image. For example, we can talk about rotation, contrast, and cropping, as shown in the following figure:

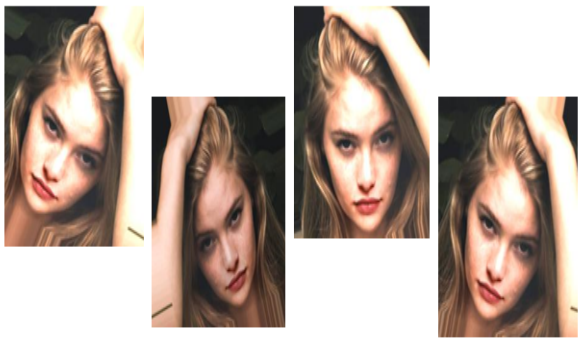

Fig. 6. Example of transformation applied in our dataset.

D. Proposed architecture of real and fake image classification

The target label organizes the fake and actual faces into a dataset. The train and test data sections of the structured deepfake dataset are separated.

The dataset's 80% train section is used to train the neural network algorithms that are being used. The best accuracy

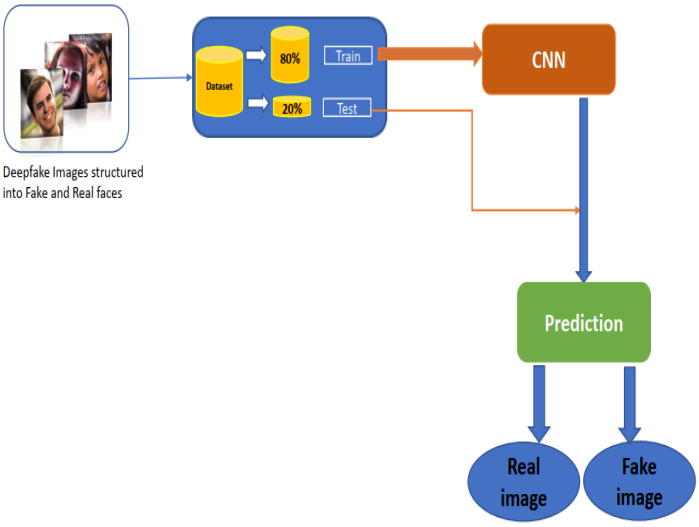

Fig. 7. The architectural analysis of the face image manipulation.

score in deepfake face identification comes from the outperforming innovative DFP technique, which has been fully hyper-parametrized.
The performance evaluation of neural network algorithms on unseen test data is determined by 20% of the dataset. With highly accurate results, the innovative proposed approach provides predictions on unobserved data. A sophisticated approach based on deep learning has been suggested, and it is now generalized and prepared for use in identifying phony and real 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 Rectified linear unit (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

The percentage of classifications a model successfully predicts divided by the total number of predictions is known as model accuracy.
On the train and test sets, the model's classification accuracy was reported to be about 98.46% and 99.44%, respectively.
Two line plots are constructed and displayed in a figure, one

for the classification of the train and test sets and the other for the learning curves of the loss on the train and test sets. The charts indicate that the model fits the issue well.

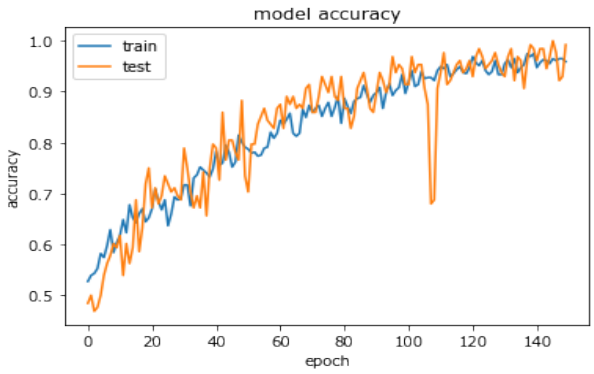

Fig. 8. The accuracy classification on the train and test sets.

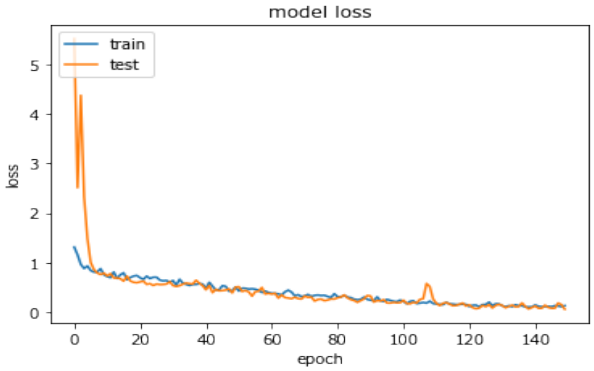

Fig. 9. The learning curves of the loss on the train and test sets.

IV. Validation with Real and Fake Videos Dataset

A. *Videos dataset description*

In this section, we'll make use of the Kaggle 2020 deep fake detection challenge data. The 470 GB of video files in this collection include training, test, and a metadata file for each video. To confirm the previous model, however, we only want to examine 100 videos. With both the picture dataset and the video dataset, we intend to create a model that is broadly applicable. The structure of this dataset is as follows.:
The video's filename and label mean the video is real or fake.

B. *Preprocessing dataset using Haar cascades*

The first step is the conversion of the videos to frames. Captured frames using VedioCapture class of the cv2 library from a video. Where each Video length (8 seconds) divides approximately into 300 Frames.
We know that faces are essential features in identifying fake and real images. Face detection is a more general case of face localization. The task is to find the locations and sizes of a known number of known faces (typically one) in arbitrary (digital) images. Face detection detects facial features and ignores anything else, such as buildings, trees, and bodies.
In our case, to obtain the faces from these frames we explored OpenCV's algorithm currently uses the Haar Cascade Classifier (HCC) [19]. Which are the input to the basic classifiers to detect faces in frames.
Applying each of the 6000 features individually would be preferable to applying them all at once on a window. (Initial stages often have a relatively small amount of features.) Discard a window if it doesn't pass the first test. The remaining features are not taken into account. Apply the second stage of features and carry on with the process if it succeeds. The face region is the window that completes all steps.

C. *Proposed architecture of real and fake image classification*

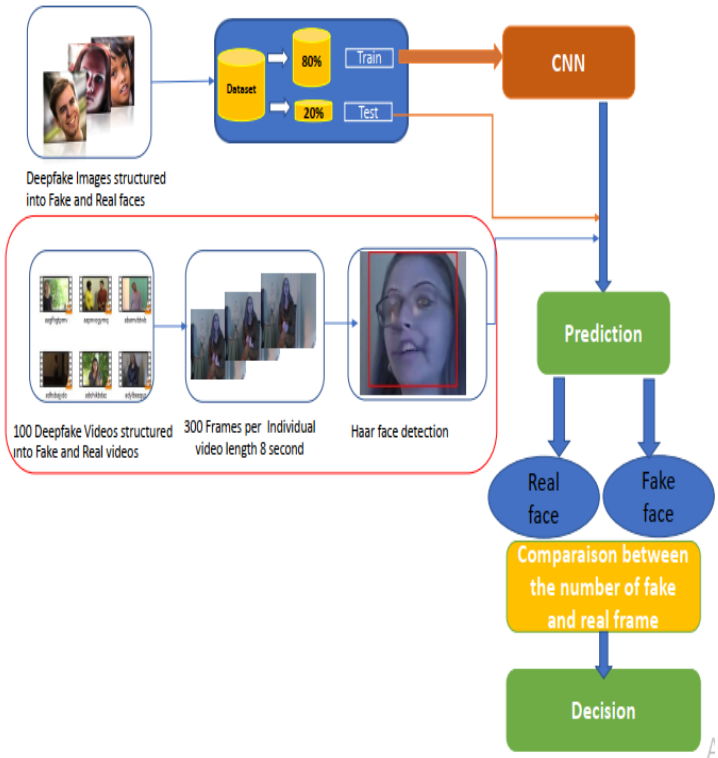

Fig. 10. The architectural classification of real and fake videos.

Generally, for the analysis of the realism videos we will follow the following algorithm:
- For each frame we capture the face with Haar Cascades technique and test it through the model that we created using the first dataset.
- Our algorithm is based on the detection of one fake frame.Which means, if one frame is fake of the 300 that we are testing is fake. Then the corresponding video is fake.

We were able to validate a best recognition rate using

the privious model with DFDC dataset.

D. Result and discussion

Figure 12 displays the outcomes of the prediction tests conducted on the DFDC dataset as two curve plots, one for the true labels (Blue one) and their predicted values (orange one).

Our model has a 91 percent accuracy rate with the videos dataset. Our results are interesting in relation to the other research works, but it can be more interesting when we improve the quality of our dataset.

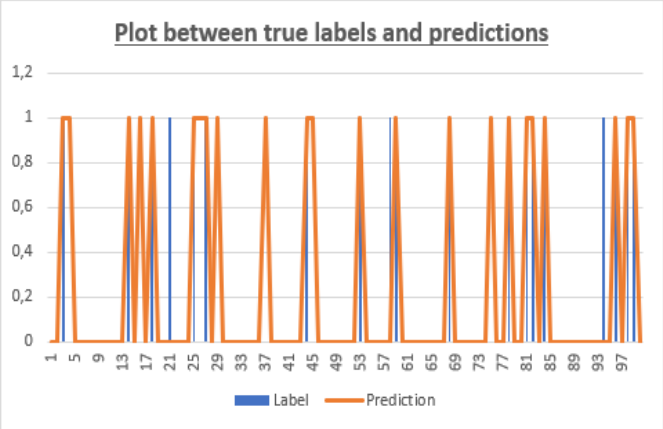

Fig. 11. Plot between true labels and predictions for the task concerning prediction of the fake videos.

V. Comparative Study

The most recent techniques for categorizing FAKE and REAL videos provide a number of tactics. The table that follows compares the findings of this study to a summary of other significant pieces of research in this area.

TABLE I
SOME RELATED WORKS.

Authors	Classifiers	Datasets	Performances
Tran, V. N et al. [20]	CNN	DeepFake Detection Dataset (DFDC)	Acc. = 95.8%
Saikia, P et al. [21]	CNN_LSTM	DFDC,	Acc= 66.26%
Wodajo, D et al. [22]	Convolutional Vision Transformer	DFDC	Acc= 91.5 %
Rahman, A et al. [23]	CNN	DFDC	Acc= 94.93%
Our work	CNN	DFDC	Acc= 91 %

VI. Conclusion

In this work, we have explored the area of deepfakes detection, which, like all the other fields of image processing, has achieved major evolution and great interest since the birth of deep learning. We tested tow types of datasets (image and video) with CNN architecture. Finally, before moving on to the perspectives, this work allowed us to put into practice our knowledge of convolutional neural networks and enrich them. We can cite perspectives to Develop the model to be more precise in detecting deepfakes videos. Then we plan to increase the size of the dataset to obtain better evaluation of the validation model.

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] Yaacoub, Jean-Paul A., et al. "Advanced digital forensics and anti-digital forensics for IoT systems: Techniques, limitations and recommendations." Internet of Things 19 (2022): 100544.
[3] Stieglitz, Stefan, et al. Communications Trend Radar 2022. Language awareness, closed communication, gigification, synthetic media cybersecurity. No. 14. Communication Insights, 2022.
[4] Harfoush, R., Basdevant, A., Hurstel, J., Bouarour, N. Récits et contre-récits.
[5] Zhang, R., Isola, P., Efros, A. A., Shechtman, E., Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 586-595).
[6] 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
[7] 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).
[8] Morph thing. https://www.morphthing.com/, 2020. Accessed: October 2020
[9] 3dthis face morph. https://3dthis.com/morph.htm, 2020. Accessed: October 2020
[10] Face swap online. https://faceswaponline.com/, 2020. Accessed: October 2020
[11] Abrosoft fantamorph. FantaMorph,Abrasoft:http://www.fantamorph.com/, 2020. Accessed: May 2020.
[12] Face morpher. http://www.facemorpher.com/, 2020. Accessed: October 2020
[13] Magic morph 1.95. https://downloads.tomsguide.com/magicmorph,0301-6817.html, 2020. Accessed: October 2020.
[14] Raja, Sushma Venkatesh Raghavendra Ramachandra Kiran and Busch, Christoph,2022, Face Morphing Attack Generation , Detection: A Comprehensive Survey.
[15] 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.
[16] 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.
[17] Raza, Ali and Munir, Kashif and Almutairi, Mubarak, 2022 ,A Novel Deep Learning Approach for Deepfake Image Detection,Applied Sciences, pages 9820, MDPI
[18] 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).
[19] Shikha Agrawal · Kamlesh Kumar Gupta · Jonathan H. Chan · Jitendra Agrawal · Manish Gupta, Machine Intelligence and Smart Systems, Proceedings of MISS, 2021, Springer
[20] 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.
[21] Saikia, P., Dholaria, D., Yadav, P., Patel, V., Roy, M. (2022, July). A hybrid CNN-LSTM model for video deepfake detection by leveraging optical flow features. In 2022 International Joint Conference on Neural Networks (IJCNN) (pp. 1-7). IEEE.

[22] Wodajo, D., Atnafu, S. (2021). Deepfake video detection using convolutional vision transformer. arXiv preprint arXiv:2102.11126.
[23] Rahman, A., Siddique, N., Moon, M. J., Tasnim, T., Islam, M., Shahiduzzaman, M., Ahmed, S. (2022, September). Short And Low Resolution Deepfake Video Detection Using CNN. In 2022 IEEE 10th Region 10 Humanitarian Technology Conference (R10-HTC) (pp. 259-264). IEEE.
[24] 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.