Combining Deep Learning and Super-Resolution Algorithms for Deep Fake Detection

Baran Deniz Korkmaz

August 2020

1 Introduction

Frequently, solution methods have used visible artifacts, that are common among most of Deep Fakes. The most successful methods are based on eye blinking [1], mismatched color profiles [2] and face warping artifacts [3]. Such artifacts provide good accuracy on big part of deep fakes, especially on old ones. On the other side, issue is considered as more complex, requiring other correlations besides visual ones. Due to this, we have attempts that are directly classifying faked content using actual algorithms of machine and deep learning [4].

Another look to this problem gave the different approach to Deep Fake detection. Method was designed to expose Deep Fakes basing on mismatch between directions of different face regions [5]. It gave good accuracy on more accurate fakes, but still have problems with low resolution video fakes.

2 Research Materials and Methods

Basing on the results of previous research we suggest the new method for detecting Deep Faked video content. Our method is inherently an alliance of 2 methods: Exposing Deep Fakes using inconsistent head poses [5] and detecting Deep Fake pictures using CNN Resnet50 [6] model.

Pipeline of the system **(see Figure 1)** consists of 4 blocks:

1. Dataset Preprocessing: Original videos fed to Face Recognition module. There they are separated for frames; face location is estimated on each frame and 68 face landmarks **(See Figure 2)** are estimated for each frame.

2. ResNet Classification: All preprocessed frames are going through 2 classifications.

3. Inconsistent Head Pose estimator: In addition to ResNet, the super-resolution algorithm was applied to solve this problem. It aims to increase the accuracy of predictions on low-resolution videos: the quality of

Deep Fakes often artificially decreased in order to hide artifacts and make pictures closer to the real one.

4. Arbitrage (Decision Maker): Results from both classifications are transferred to Decision maker that is announcing the final judgement.

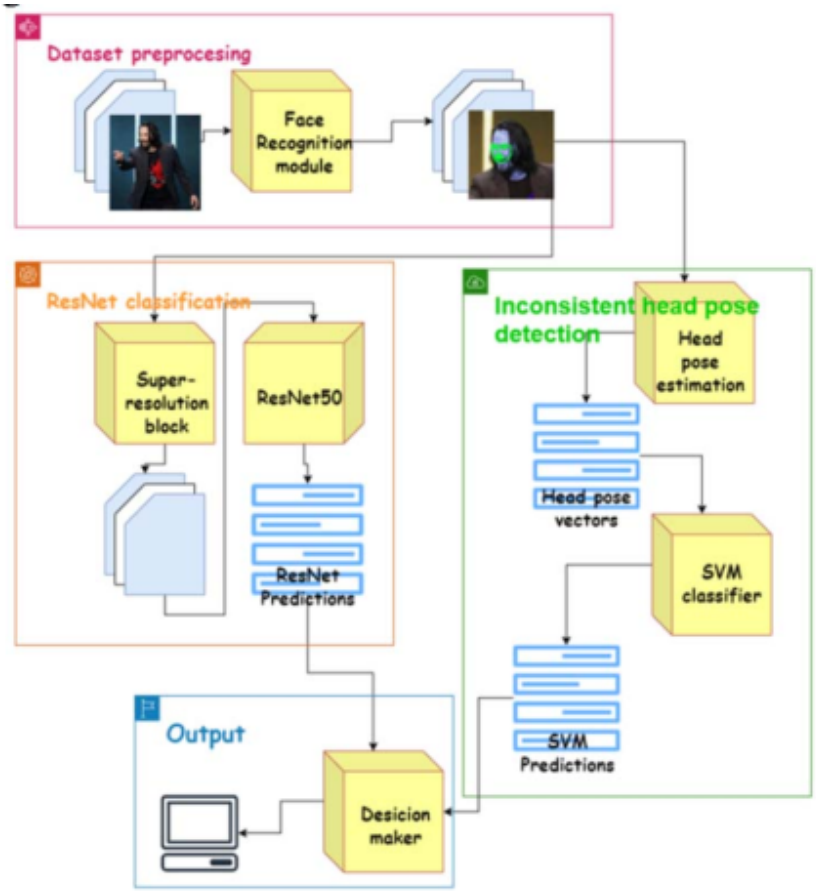

Figure 1: Detection system pipeline

Figure 2: 68 face landmarks got using Dlib package

2.1 ResNet Classifier

The hole research could be classified for several sub- objectives. Firstly, single image face detection methods for video content were applied: video is considered as a list of frames. For each frame we applied face detection methods realized in Dlib software library [7] and face recognition python module [8].

Basing on these landmarks the face rectangle is built which is used to crop all frames. Cropped frames, containing only face regions, are further used for training and evaluating classifying network.

In our research we implemented Resnet50v2 model which was finetuned for binary classification purposes. Weights of pretrained on ImageNet dataset were used as a base of our learning process.

To finetune model instructions from "Exposing DeepFake Videos By Detecting Face Warping Artifacts." were followed (we took learning rate for 0.001, momentum equal to 0.9, used Cross Entropy loss function and set SGD optimizer), although the expected results of near 95 accuracy weren't achieved [3]. So, it was decided to increase the learning rate of the model from 0.001 to 0.01, owing to that the accuracy about 94.9% **on evaluation dataset was achieved. The model was trained for 20 epochs, evaluated after each epoch and its parameters were saved. The best were picked for further research.**

2.2 Inconsistent Head Pose Estimator

Another big part of our study was reproducing method used in Li Yuezun, Lyu Siwei research [5]. The main idea consists in difference of estimated face direction vectors: the one based on outer landmarks and the one based on inner landmarks. During mask reconstruction Deep Fake algorithms inevitably make invisible difference in face directions between outer and inner parts. Thanks to that vector analyze gives program an opportunity to expose Deep Fakes even when sophisticated eye can't recognize it. For head pose estimation one may need to find 2 vectors (rotation vector and translation vector), which relates world coordinates of facial landmarks and their locations on the image. Together these 2 vectors let the one to build head pose vector.

In the paper, the introduction of optimization problem to formulate the objective discussed above is also given.

2.3 Super-Resolution Preprocessing

The last part of our research was applying super-resolution algorithms, so it was decided to implement Fast Super- resolution CNN model [11], which would be able to upsample data from single image [12].

3 Source

- Combining Deep Learning and Super-Resolution Algorithms for Deep Fake Detection; Nikita S. Ivanov, Anton V. Arzhskov, Vitaliy G. Ivanenko

4 Referenced Resources

Deepfake Detection Techniques

- **Eye Blinking - [1]** - Y. Li, M. C. Chang, and S. Lyu, "In Ictu Oculi: Exposing AI created fake videos by detecting eye blinking," in 10th IEEE International Workshop on Information Forensics and Security, WIFS 2018, 2019.

- **Mismatched Color Profiles - [2]** - H. Li, B. Li, S. Tan, and J. Huang, "Detection of Deep Network Generated Images Using Disparities in Color Components," Aug. 2018.
- **Face Warping Artifacts - [3]** - Y. Li and S. Lyu, "Exposing DeepFake Videos By Detecting Face Warping Artifacts."
- **Deep Learning (RNN) - [4]** - D. Guera and E. J. Delp, "Deepfake Video Detection Using Recurrent Neural Networks," in Proceedings of AVSS 2018 - 2018 15th IEEE International Conference on Advanced Video and Signal-Based Surveillance, 2019.
- **Inconsistent Head Poses - [5]** - X. Yang, Y. Li, and S. Lyu, "Exposing Deep Fakes Using Inconsistent Head Poses," in ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2019.

Useful Auxiliary Resources (Libraries,Modules, etc.)

- **CNN ResNet 50 - [6]** - K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016, vol. 2016-Decem, pp. 770–778.
- **Dlib software library - [7]** - D. E. King, "Dlib-ml: A machine learning toolkit," J. Mach. Learn. Res., 2009.
- **Face Recognition Python Module - [8]** - Adam Geitgey, "Machine Learning is Fun! Part 4: Modern Face Recognition with Deep Learning," Medium, 2016.

Super-Resolution Preprocessing

- **Fast Super-Resolution CNN - [11]** - C. Dong, C. C. Loy, and X. Tang, "Accelerating the Super-Resolution Convolutional Neural Network," Aug. 2016.
- **Multiple-Image Super-Resolution - [12]** - M. Kawulok, P. Benecki, S. Piechaczek, K. Hrynczenko, D. Kostrzewa, and J. Nalepa, "Deep Learning for Multiple-Image Super-Resolution."