Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC Accredited with "A" Grade (CGPA : 3.18)

Department of Information Technology

Academic Year 2020 – 21

Project Title: DeepFake Detection

Project Guide: Vinaya Sawant; Project Members: Ankit Basrur, Devansh Mehta, Dhrumil Mehta

Abstract:
Deepfakes are increasingly exponentially which can be a threat to privacy of an individual. Deepfake creation is easily done by layman due to availability in various tools available, which can be used for the defamation of an individual/organization/community. Therefore, different methods have been introduced to address this issue. Early methods were based on the inconsistencies of fake video obtained by using handcrafted features. Recent methods, on the other hand, applies deep learning to automatically extract prominent and discriminatory features to detect deepfakes. This paper shows study related to existing deepfake detection and proposes a technique using diffirent algorithm that can be used in videos and images.

Introduction:
The number of fake images as well as videos are increasing due to high level of editing tools available on Internet. The knowledge of ML and computer vision decreases the manual steps required for image or video editing. This has led to an increase in spreading replicated images and videos (Deepfake). Deepfakes are synthetic media in which a person in an existing image or video is replaced with someone else's likeness. Convolutional Neural Networks (CNNs) have gained much popularity in recent years for vision-related applications and have the potential to achieve high accuracy. Different type of Generative Adversarial Network (GAN) algorithm is used for creation of fake images, which makes it difficult to differentiate between real and fake images. In this paper, we have used separate architecture for detection of Deepfake in images and for videos. We have used MTCNN for face detection [1] followed by face alignment using dlib. We trained models using different backbone networks, XceptionNet and Densenet.

Architecture:
In our proposed achitecture as per Fig.1, the video will be broken down into frames. Those frames will be passed to our preprocessing unit in which the face detection algorithm will be applied. After analyzing various algorithm, MTCNN gave the best results as it detects profile faces too. Depending on the face obtained bounding boxes will be created and cropped faces will be passed to Face alignment module. After alignment, those frames will be passed to DenseNet for videos and XceptionNet for images. Finally, the output for each frame will be given with fake probability and shown to user.

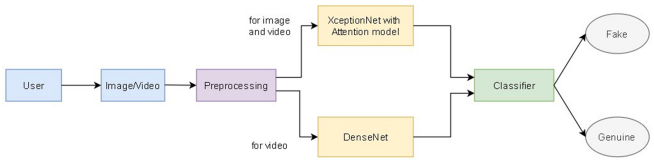

Fig. 1: System Architecture

Algorithm:
Pre-processing algorithms:
1.Multi-Task Cascaded Convolutional Neural Networks(MTCNN): MTCNN[1] is a neural network which detects faces and facial landmarks on images. It is relatively newer method of face detection.This technique consists of 3 neural networks which are connected in cascade, first detects the bounding boxes of faces in an image along with 5 features of facial landmarks. Each NN passes its inputs to a CNN. Bounding boxes with scores are returned. The input image is scaled down and is passed through CNN. Image extraction is performed for each bounding box and resize them (24x24 and 48x48 for phase 2 and phase 3, respectively) and are forwarded to CNN. Stage 3 additionally computes 5 features of facial landmarks along with the confidence value.MTCNN is highly accurate for detecting non-frontal faces making it ideal for our
use case.

2.OpenCV Face alignment: It is an important process to reduce irregularities so that our model is consistent and trains faster. We have used OpenCV and dLib for getting the coordinates of eyes and nose. It consists of following steps:Initially we have to detect the face and eyes after which centre of eyes is found. Connecting those two points and forming a right -angled triangle will help us in finding the angle. The last step is rotating the image by that angle using OpenCV [2].

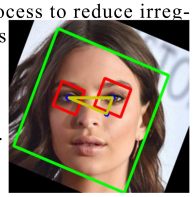

Fig. 2: Face Alignment

Detection Model:
1.XceptionNet with Attention Model mechanism: It is a traditional CNN trained on ImageNet based on separable convolutions with residual connections. It makes use of a pre-trained ResNet18 model on ImageNet, and move the weights is for learning features that are more beneficial for detection i.e., objects, skin colour alteration, distortion, etc. It makes the model to look for more useful features. Attention Model gives attention to several salient feature, which will be passed ahead in the network. It is the most suitable model since it will highlight the region that has discrepancies in it, rather than focusing on entire image[3]. This will help to learn salient features and it will eventually increase accuracy and speed. In this technique we transfer it to our task by replacing the final fully connected layer with two outputs. The model is trained using Adam optimizer with learning-rate set to 0.0001.

2.DenseNet: It extracts features at different layers in a network. In this model, each layer can access loss function of the previous layer and the original video signal which helps in improving efficiency. It then concatenates features learned from each layer[4]. It helps in decreasing the back-propagation time while training.

Results: This model gives an accuracy of 81% on training and 72% on validation for XceptionNet whereas, 76% on training and 71.6% on validation for DenseNet.

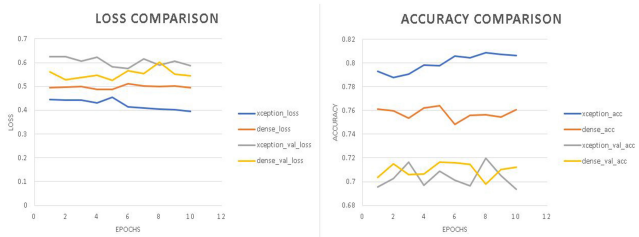

Fig. 3: Model training History

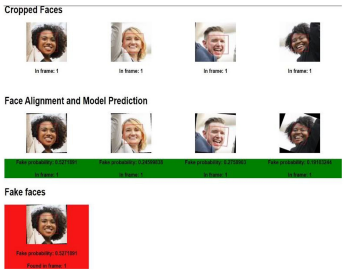

Fig. 4: Working Demo

Conclusion:
Based on our study we have used MTCNN for pre-processing as it is capable to detect non-frontal faces. For the detection purpose DenseNet and XceptionNet is used for video and image detection respectively as DenseNet can identify differences between two frames. The application provides high value of efficacy which could improve upon training over more dataset.

References:
[1] K. Zhang, Z. Zhang, Z. Li and Y. Qiao, "Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks," IEEE, vol. 1, 2016.
[2] V. Kovenko, "How to precisely align face in Python with OpenCv and Dlib," Towards Data Science, 11 August 2019. [Online]. Available:https://towardsdatascience.com/precise-face-alignment-with-opencv-dlib-e6c8acead262. [Accessed 26 AUgust 2020].
[3] H. Dang, F. Liu, J. Stehouwer, X. Liu and A. Jain, "On the Detection of Digital Face Manipulation," In Proceeding of IEEE Computer Vision and Pattern Recognition (CVPR 2020), vol. 4, 19 April 2020.
[4] G. Huang, Z. Liu, L. v. d. Maaten and K. Q. Weinberger, "Densely Connected Convolutional Networks," arXiv CVPR 2017, vol. 5, 2018.