Sanjana S.
S7 CSE B.

2: Methods of Deepfake detection based on Machine learning

Deepfake videos are AI generated videos that look real but are actually fake. Using such videos and images, it is easy for malicious abuses to create arbitrary fake news and fool and mislead the public. Here finding of technology to decide whether photo was changed with deepfake technology or not with good accuracy.

Deepfake videos are created using autoencoders and GANs. Autoencoders contain Encoder part and decoder part. GAN contain 2 neural networks - generator and discriminator. Generator looks similar to autoencoder net, but better results can be achieved. Usual indicators used to detect deepfake videos are too smooth skin, colour mismatch between synthesized face and original face, Head position, Artifacts on small moving parts, eye blinking rate, Face warping artifacts and person's patterns of behaviour.

Celeb-DF dataset is used to evaluate model. It contains thousand videos and they are good quality synthesized videos with almost none artifacts of original face, small moving parts and other indicators. Videos are splited frame by frame, and each frame is analyzed. (Gaussian Blur, exponential blur or RedBrightBlur's applied)

If 1% of frames are classified as modified, we should mark the whole video as Deepfake.

Model is evaluated using 3 indicators and Face warping artifacts outperformed other indicators. That conclusion bases on fact that great part of present deepfake algorithms can synthesize only low resolution faces and then they should affine them to make picture whole and smooth.

Our project is also to detect deepfake videos using Recurrent Neural Network. Inorder to detect deepfake videos, we need knowledge about how they are created and that can be learned from this paper. It also helps our project by giving knowledge about best indicators that we should use to detect deepfakes.