DEEPFAKE DETECTION USING DEEP LEARNING

Kotagedara Liyanage Isuru Nadeeshan Perera

A dissertation submitted in partial fulfilment of the requirement for MSc in Advanced Software Engineering

School of Computing

Informatics Institute of Technology, Sri Lanka

in collaboration with

University of Westminster, UK

2022

Abstract

Deep learning models may generate hyper-realistic images, audio, and even videos such as the called "Deepfake" mainly created by Generative Adversarial Networks with altered audio and/or video samples, which are so perfect that they are undetectable from the genuine ones in human perception. Deep learning models are now accessible for use in a wide range of innovative applications thanks to recent technological breakthroughs. Apart from creative and acceptable uses, there are various malicious or illegal methods to exploit such fake materials in propaganda, political campaigns, cybercrime, blackmail, and other similar activities. To address the issues raised by Deepfake contents, deepfake detection is proposed.

The existing works are done in many ways using many technologies but still, there are new technologies that have not been widely experimented with. To enhance accuracy, stack additional layers or apply wider layers however, this introduces more parameters. More processing power will eventually be required. As a result, a novel design called ResNeXt is invented that improves accuracy while lowering network complexity and parameter number. Recently invented ResNeXt is put into an experiment and in deep learning ResNeXt and Long Short Term Memory network (LSTM) are not used together. ResNeXt reveals an additional dimension called cardinality when compared to a ResNet (the number of transformations in the set) as an important factor in addition to depth and width.

This research is evaluated by comparing acquired outcomes for popular datasets to those of current systems in the domain such as FaceForensics++, Celeb-DF, and Celeb-V2. To evaluate the effectiveness of the suggested approach, quantitative analysis was done using the Accuracy, F1 score, and Area Under Curve (AUC) metrics. Extensive experiments were performed on the FaceFornsics++ dataset, obtaining an accuracy of 93.16%, F1 Score of 0.92, and AUC of 0.93. The proposed approach was assessed for generalizability using the Celeb-DF (Accuracy of 93.86%, F1 score of 0.93, and AUC of 0.94) and Celeb-V2 (Accuracy of 97.74%, F1 score of 0.97, and AUC of 0.98) datasets. Furthermore, this research has identified several potential topics for future work.

Keywords: Deepfake detection, Image Processing, Convolutional Neural Networks, Recurrent Neural Networks, Deep Learning