MSc Project - Reflective Essay

Project Title:	Deep Fake Detection Using Artificial Neural Network
Student Name:	**Bharath Radhakrishna**
Student Number:	**210667521**
Supervisor Name:	Eranjan Padumadasa
Programme of Study:	MSc Artificial Intelligence

Overview

My project is a Deep learning project which is a subbranch of Artificial Intelligence and deals with the human brain inspired neural network technology. Computer vision plays an important role in our project. It helps in processing the video and frames with the help of Open-CV. A PyTorch trained model is a classifier to classify the source video as deepfake or pristine. Here, I quickly explain the thinking behind the chosen strategy before going into the project's advantages and disadvantages in relation to earlier research, areas for future study, real-world difficulties and solutions, as well as the legal, moral, and environmental considerations. I finish by considering the project's overall success in the perspective of my personal and professional development.

Approach

I examined the problem description to determine whether it could be solved. I have read several study papers. after determining if the problem statement is feasible. The collecting and analysis of the dataset came next. I analysed the data set using several training approaches, such as negatively or positively trained, that is, I found that using only real or fake films to train the algorithm might generate additional bias., which would result in erroneous predictions. Therefore, after doing a thorough analysis, it was shown that balanced algorithm training is the best technique for reducing algorithm bias and variance while attaining great accuracy. I assessed the solution's cost, processing speed, usefulness, level of expertise, and equipment accessibility. Design I created the system architecture for the solution as described in the paper after conducting research and analysis. I choose the Model's basic architecture, which details the many layers and their densities. Development After consideration, it was chosen to programme using the Python 3 language and the PyTorch framework. The reason PyTorch was selected is because it can be customised and has strong CUDA, or Graphic Processing Unit (GPU), support. For training the end build over a vast array of data sources, use Google Cloud Platform. I used a sizable real-time dataset, including a collection of YouTube videos, to assess my model. The trained model's accuracy is assessed using the confusion matrix method. Users may assess if a new video is deepfake or legitimate using the solution's output, which comprises of trained deepfake detection models. Applications The user will utilise a web-based application to upload and submit the video for processing. If a video is uploaded, the model will pre-process it and determine if it is a deepfake or not. Resources and Equipment Required For this project, a machine with adequate processing power is needed. This project requires too much computing power since the photographs and videos must be processed in batches.

The application's user will be able to determine if the uploaded video is real or fraudulent, as well as the model's level of confidence in its forecast. The user will be able to watch the playing video combined with the model's assurance and output on the face. simple and convenient. Users appear to choose a Deep Fake video detection algorithm that is less complicated. A simple and user-friendly interface is therefore designed. To choose the video for processing, use the browse tab on the user interface. It simplifies things while also improving the user experience. Your top focus should be accessibility if you

have a growing target market. You may expand your reach across several platforms by turning on a cross-platform compatibility option. Being a server-side programme, it will function on any machine equipped with a web browser.

Practical Challenges and limitations

I applied a spiral model. The Software Development Model emphasises the individuals performing the task, their interactions, and risk management. Spiral's goal is to ensure that changes can be made more rapidly and during the developmental procedure to ensure up routine evaluations of the product while taking the desired outcomes into account.

Prior to the training, thousands of pictures for each person had to be created. We can easily get facial pictures from their films by employing a face detection library. Practice taking photos of faces for a long period. It materially affects your outcome. Remove any image frames with many people in them. Make certain you have a tonne of video. Face angles, poses, and facial emotions vary in extract facial images. Similarities between the two people, such as a similar facial shape, may be helpful. In Deepfakes, the produced face is given a mask to help it blend in with the target footage. In order to further remove the artefacts, to further disperse the mask boundary region, apply a Gaussian filter. Configure the programme to further enlarge or minimise the mask. Control the mask's form. Cannot explain the sounds using my method. Because of this, our method is unable to identify the audio deep fake. However, in the future, I would advocate for identifying audio deep fakes.

Legal, Ethical and Environmental Considerations

The fact that data is readily available and that datasets may be created, if necessary, as well as the fact that a lot of public movies and images are being made available online, are all advantages of this initiative. As a result, there was no need to conduct interviews or handle any sensitive personally identifiable information, or to include the proprietary research data of any institution or private research organisation in the study. Every effort was made to guarantee that all data came from publicly available sources. All reviewers' contributions were duly recognised, and the datasets were properly cited. Although no intellectual property was registered because of this work, there is certainly room for further development of these exploratory results into industrially useful applications, notably super-resolution. a source of increased anxiety. Large neural networks designed for multi-task learning from natural language text, like BERT (Devlin et al., 2019), have been predicted to burn roughly sixty times more carbon than the normal human lifespan (Strubell et al., 2019). However, employing proper initializations, such as weights learnt by pre-training on the ImageNet repository, as in the current work, the computing requirements of model training can be decreased. Fortunately, Google Cloud Compute Engine, whose website claims that 100% of its energy usage is offset with renewable sources and is supported by Luccioni et al., was used for most of the neural network training in this project (2020). However, future research should look at whether the same performance can be obtained with lighter versions.

Individual Development

Despite the difficulties of resource availability and the complexity of the subject, this study has both met and exceeded its goals, in addition to the potential to extend this work mentioned above. Being somewhat experienced with computer science, I am happy with

the advancements made in this MSc in Artificial Intelligence during the last year. I am now confident in the fundamental ideas of machine learning, the specifics of certain computational and computer-vision components, and the application of distributed deep learning. Without the help of my supervisor Professor Eranjan Padumadasa, none of the would have been feasible. I want to express my heartfelt appreciation to him. The tasks I have completed for my projects have provided me with a wealth of information, The dataset must be downloaded for this activity. Implementation entails dividing the movie into frames and cropping each frame to only contain faces once the dataset has been analysed and made suitable for pre-processing. Pre-processing involves generating a new dataset with just face-cropped videos in it. the use of a data loader to load the labels and video. Using a little quantity of data to train a basic model. The goal is to attain maximum accuracy by adjusting the learning rate, batch size, weight decay, and model architecture. The huge dataset's final model is trained using unit testing.

Conclusion and Future Work

With the confidence of the provided model, I have presented neural-network based method for classifying the movie real or deep-fake. The technique can accurately anticipate the result after analysing one frame of video (10 frames per second). In order to retrieve at the frame level characteristics and analyse the Spatial order to identify changes between frames t and t-1, I employed the learned ResNext Convolutional neural network. I employed the developed ResNext Convolutional neural network and the LSTM. Our model is capable of processing videos with 10, 20, 40, 60, 80, and 100 frames per second. Any built system may always use improvements, especially if the project was created utilising current technology and has promising long-term potential. For user convenience, a web-based platform can be upgraded to a browser plugin. Although the system may be improved to detect whole body deep fakes, it currently only detects face deep fakes.