Deepfakes Face image, Face Manipulation and Fake Detection

Amira Elsharaby
AI, Nile University
Giza, Egypt
A.elsharaby.nu.edu.eg

***Abstract*—fake images, also known as DeepFakes, are intended to distribute offensive content and false information to millions of people, a problem made worse by the reach of contemporary media and the content's inherent controversies. The detection and classification of DeepFake images using a learning model architecture are the main topics of this research. The first model we'll employ is ResNet50, along with CNN (convolution neural network), both of them are neural network models made for identifying features that can identify a fake or real. (*Abstract*)**

Keywords—deepfake, image classification, ResNet50, CNN (*key words****)***

I. INTRODUCTION

Deepfakes are modified media products created to disseminate false information, hoaxes, or other offensive material. Due to the proliferation of easy-to-use tools for editing and creating images, image faking has recently become a severe problem.

Deep fakes occur when it is easy to toy with an image's features or to transfer person X's facial expressions to person Y's face. We strive to find a solution to the issue of picture and video alteration because deep fakes are challenging to spot due to the change in the images being so convincing. Therefore, it is challenging to counteract false information and data because of current developments in architectural models that produce these fakes.

These are deficient However, Deepfakes leave behind observable visual artefacts that may be examined using different neural networks models. This paper constrains the problem of deepfakes images and how to solve this problem using different neural networks models.

II. BACKGROUND

The field of deepfake detection has seen significant research in recent years, with many algorithms, techniques, and tools developed to address the problem. One commonly used approach remains the use of Artificial neural networks (ANNs), which are deep learning models designed to solve different problems.Two ANN models are employed in this project to classify deep fake images.

1- ResNet50.
2- CNN.

The algorithms used in this project to analyze many aspects of an image, such as shape, and then identify the image as real or fake using this data. To increase the accuracy of models and make them more resilient to changes in the input data, techniques including transfer learning, data augmentation, and data image generator are used.

A) Dataset

This dataset was created to identify the fake or real face images. The dataset consist 3 folders on it (test- train- valid) each folder Contains more than 100,000 images which make the dataset size is to large (4 GB).This dataset from Kaggle “140k Real and Fake Face “ .

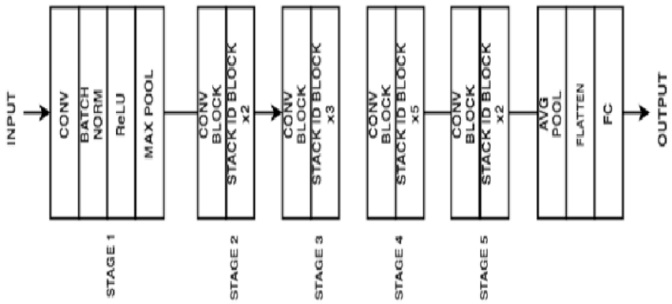
B) Tools and libraries

Python, TensorFlow, Keras, transformers, PIL, pandas and numpy are used to implement models and perform the necessary steps for data pre-processing and post-processing. To properly train and evaluate models, the project also makes use of a sizable collection of real images and deepfakes.

Overall, the project offers a thorough and current understanding of the fake image detection issue and suggests a workable solution using ANN models and related methodologies. The project expands on earlier research in this field and shows how deep learning techniques could be used to detect fake images

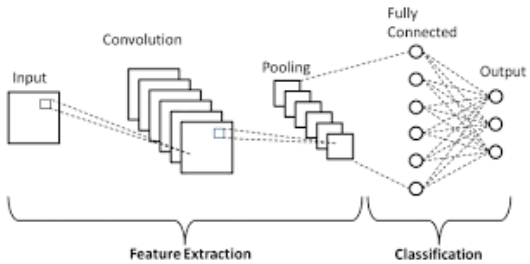
III. METHODOLOGY

1) A Residual Neural Network (ResNet) model is a simple 50-layer CNN model with residual blocks that workes to didn't forget the previous weights.



ResNet50 is a special type of ResNet. This "ResNet" architecture Know as a "bottleneck" that reduces the number of parameters and matrix multiplications. we used a fine-tuned ResNet 50. It is a pre-trained model that has been sparingly applied to image classification tasks, and it is then tuned or tweaked to help it do a second, related task in a large dataset of deepfake and real images. In other words, fine-tuning means using the trained weights of a residual neural network as initialization for a new trained model on data from the same domain.I try to add more layer but we found a over fitting so I only used 3 layers flatten, GlobalAveragePoolink2D(), and 2 dense layers with different activition functions. When I fitting this ResNet model it take more than 12 hours to run so I did a earlystoping and Checkpoint list to save all of output in csv file to call it when I run the notebook agin and didn't take a lot of time in runing. The model with a trained number of epochs equal to 30 and batch size equal to 16. Several loss functions were evaluated using our model algorithm, and the one offering the best performance and accuracy was chosen. On the test data set, the model had an accuracy of 0.9885200262%. For testing the model.which it amazing

2) A convolutional neural network (CNN or Convnet) model is a specific network architecture for deep learning algorithms used for image recognition because convolutional layers independently process each local image region while sharing related parameters across the whole image.



The model was built using a single input layer, four hidden layers, and a single output layer. The first layer consists of 16 input channels, which are connected to the input layer. The second layer is MaxPooling2D() using the highest value of a non-overlapping 2x2 patch, this max pooling layer down samples the output of the preceding layer. Then the third layer consists of 32 input channels. Then maxpooling

The second consists of 32 input channels. The first two hidden layers automatically enter the flatten layer to reshape the output to be a 1D vector. then the fully connected layer, which takes 64 input channels and the positive activation function Relu., then adding to the output layer with a Sigmoid activation function. For compiling the model, we used the Adam optimizer technique to train the model and tf.losses Binary Cross entropy loss for calculating the loss. metric was accuracy ". model with a trained number of epochs equal to 10 and batch size equal to 3. Several loss functions were evaluated using our model algorithm, and the one offering the best performance and accuracy was chosen. On the test data set, the model had an accuracy of 0.72%. For testing the mode which is not bad.

IV. Results and Analysis

The results of the two models, the fine-tuned ResNet50 and the simple CNN, show that both models are capable of detecting deepfake images with high accuracy. However, the fine-tuned ResNet50 model achieved higher accuracy, recall, and precision than the simple CNN model.

V. Conclusion

To sum up, this research used CNN and ResNet50, two neural network models, to discuss deep fake image detection. A collection of real and fake face images was used to train and evaluate the models. Both models successfully identified deep fake images, according to the data, but the ResNet50 model that had been adjusted performed better in terms of accuracy, and recall. To improve the accuracy and resilience of the models, the research project also used a number of strategies, including transfer learning, data augmentation, and early stopping.

References

[1] deepfakeinthewild.(n.d.).deepfakeinthewild/deepfake-in-the-wild. GitHub.Retrieved[insert date of access], from https://github.com/deepfakeinthewild/deepfake-in-the-wild

[2] Dolhansky, B., Howes, L., Sakaridis, C., & Pflaum, C. (2020). DeepFakeDetectionChallenge[Dataset].Zenodo. https://doi.org/10.5281/zenodo.4068245

[3] Li, Y., Chang, M.-C., Lyu, S., & Juan, C.-H. (2020). DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection. https://arxiv.org/pdf/2001.00179v3.pdf.

[4] Rahmouni, H., Jourabloo, A., & Liu, X. (2021). Deep Learning for Deepfakes Creation and Detection. https://arxiv.org/pdf/2103.02406v3.pdf