DEEP FAKE DETECTION OF IMAGES USING MACHINE LEARNING AND DEEP LEARNING TECHNIQUES

Mrs. Rahimunnisa Shaik[1,] Appalabathula Bhavana[2], Badiganti Lavanya[3], Bulusu Sreenidhi[4], Gorle Divya[5]

[1]Assistant Professor, Department of Computer Science and Engineering, Vignan ‘s Institute of Engineering for Women, Visakhapatnam, India

[2-5]Department of Computer Science and Engineering, Vignan ‘s Institute of Engineering for Women, Visakhapatnam, India

ABSTRACT

Big data analytics, computer vision, and human-level control are just a few of the complicated issues that deep learning has been effectively used to address. Yet, breakthroughs in deep learning have also been used to develop software that poses a threat to privacy. Both national security and democracy. Deep fake is one of those recently developed technology that uses deep learning. Deep fake algorithms can produce fake images that are so convincing that no one can tell them apart from real ones. Therefore, the development of systems that can instantly identify and evaluate the integrity of digital visual media is essential. This study provides an overview of the deep fake making algorithms and, more crucially, the deep fake-detection techniques that have been suggested in the literature to yet. We provide in-depth conversations about deep fake technology difficulties, research trends, and directions.

INTRODUCTION

Rapid advancements in AI, machine learning, and deep learning over the past few decades have produced new methods and a variety of tools for manipulating multimedia. Although technology has primarily been used for good for entertainment, education, etc. unlawful or malevolent people have also taken advantage of it.

While some deep fakes can be produced using conventional visual effects or computer graphics techniques, deep learning models like auto encoders and generative adversarial networks (GANs), which have seen extensive use in the computer vision field, currently serve as the common underlying mechanism for producing deep fakes. Deep neural networks in the form of generative adversarial networks (GANs) are frequently employed to produce deep fake. One benefit of GNAs is that they can develop samples by learning from a collection of training data.

Same qualities and attributes of data. GANs can be used, for instance, to compare a person's "actual" image to a phoney one. An encoder and a decoder are the two neural network components that make up the architecture of GANs. First, the model trains on a big data set using the encoder to generate fake data. The bogus data is then learned from genuine data using the decoder. These models are used to analyze the motions and expressions on a person's face and create synthetic facial images of different people with similar movements and expressions.

These models are used to analyze the motions and expressions on a person's face and create synthetic facial images of different people with similar movements and expressions.

To train models to produce photorealistic images using deep fake techniques, a lot of image and video data is typically needed. Platforms are also employed negatively by some parties, frequently for financial benefit, and occasionally for slanted opinion formation, mind control, or the dissemination of satire or absurdity. Fake news is a typical term for the phenomenon. The phenomenon is quickly encroaching on the film industry and endangering journalistic organizations.

LITERATURE SURVEY

Shruti Agarwal, Hany Farid, Yuming Gu, Mingming He, Koki Nagano, and Hao Li. Protect world leaders from deep fakes. computer vision and pattern recognition Workshop, Volume 1, pp. 38-45, 2019.Sophisticated fake video creation is mostly done by Hollywood studios and state actor. However, recent advances in deep learning have made this much easier. Create sophisticated and compelling fake videos. relatively small amounts of data and with computing power, even an average person can create a video of his leader in the world market. for example, acknowledging illegal actions leading to constitutional crisis, military leader said. Racially insensitive behavior that leads to civil unrest in areas of military activity; or The corporate giant claims weak earnings are leading to global stock manipulation. So-called deepfakes pose a grave threat to democracy, national security and society. To combat this growing threat, we describe forensic techniques for modeling faces. Facial expressions and movements that characterize a person's speech patterns. even if it's not visual Obviously, these correlations are often broken by how deepfake videos are created. Therefore, it can be used for authentication.

Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre Antoine Manzagol Using denoising auto encoders to extract and combine robust features. Pages 1096–1103 of the 2008 edition of Proceedings of the 25th International Conference on Machine Learning. This method can be used to train auto encoders, and these auto encoders can be utilised for denoising. stacking is used to launch deep architectures. The algorithm may be driven either by generative models or by theories of manifold learning and information. Comparative investigations on a pattern classification benchmark suite clearly demonstrate the unexpected benefit of corrupting the input of autoencoders.

Diederik P Kingma and Max Welling. Auto-encoding variational Bayes. arXiv preprint arXiv:1312.6114, 2013.In the presence of continuous latent variables with intractable posterior distributions, big datasets, and directed probabilistic models, they conduct effective inference and learning. A scalable stochastic variational inference and learning algorithm is introduced. Big datasets, and under some minor differentiability constraints, it even functions in the intractable scenario.

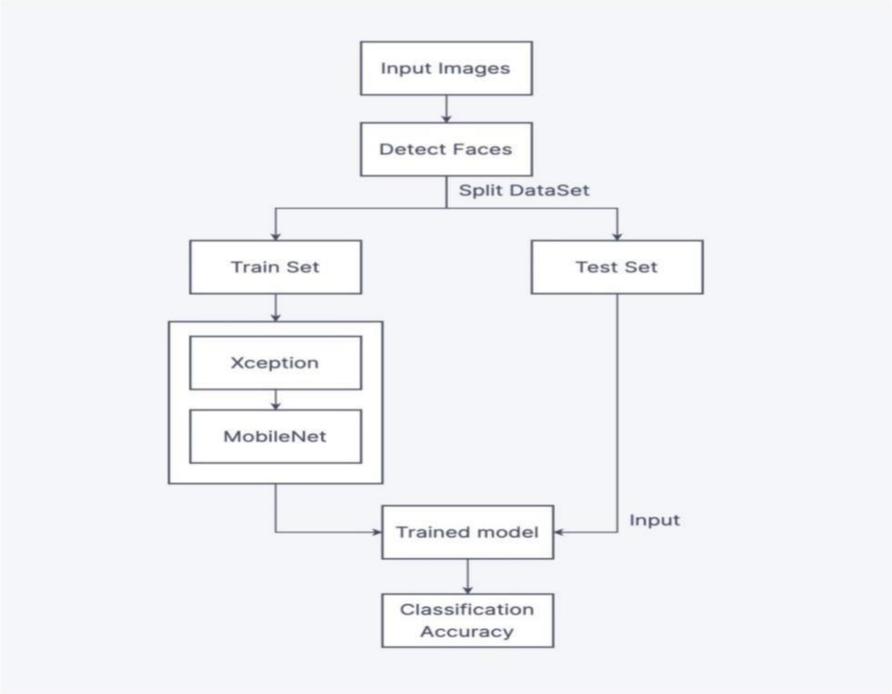
Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Adversarial generative networks. 2014; 27:2672–2680 in Advances in Neural Information Processing Systems. They suggest a brand-new method for estimating generative models through an adversarial process in which we simultaneously train two models: a generative model G that represents the data distribution and a discriminative model D that calculates the likelihood that a sample came from the training data in contrast to G. The goal of the training process for G is to increase the likelihood that D will make a mistake. This framework is equivalent to a two-player minimax game. There is just one solution in the space of arbitrary functions G and D, with G recovering the training data distribution and D being equal to 1/2 everywhere. The entire system can be improved if G and D are specified by multilayer perceptrons. Backpropagation is a technique for training the entire system. There is no requirement for any Markov chains or unrolled approximate inference networks for either training or generation of samples. By analyzing the created samples on a qualitative and quantitative level, experiments show the framework's potential.

Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, Ian Goodfellow, and Brendan Frey. Adversarial autoencoders. arXiv preprint arXiv:1511.05644, 2015The "adversarial auto encoder" (AAE), a probabilistic auto encoder that uses the recently proposed generative adversarial networks (GAN) to perform variation inference by matching the aggregated posterior of the auto encoder's hidden code vector with any given prior distribution, is what they have proposed. Generating from any region of the prior space produces relevant samples since the aggregated posterior and the prior are matched. To map the imposed prior to the data distribution, the adversarial auto encoder's decoder builds a deep generative model. They demonstrate how the adversarial auto encoder may be applied to tasks including semi-supervised classification, separating the content and style of images, unsupervised clustering, dimensionality reduction, and data visualization.

PROPOSED SYSTEM

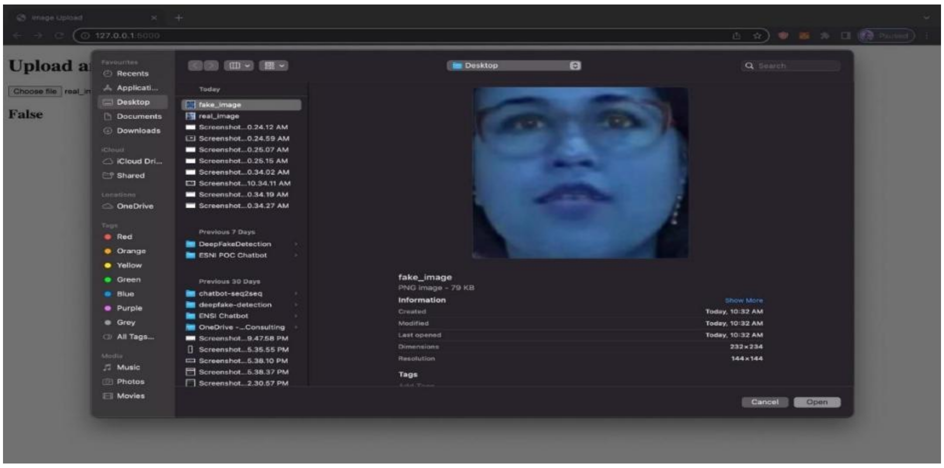
This research's main objective was to determine whether an image was real or had been created using deep fake technology. Identifying and labelling the faces that appeared in the photograph was the first stage. The cascade classifier offered by Open CV is utilized for this purpose. The detected area of the face is saved as a new image in the subsequent phase. The facial photos need to be uniformly scaled before being stored. The Xception model requires a picture with a size of 299*299, while the Mobile Net model requires a smaller image with a size of 224*224. We will layer Moblie Net over Xception in our model. Mobile Nets are effective neural network topologies that are suitable for mobile devices, as their name implies and low processing-power embedded vision-based applications. The foundation of this architecture is a depth-wise separable convolution. Standard convolutions are divided into 1*1 point wise convolutions in depth wise separable convolution, a type of convolution. The inputs are filtered using standard convolutions, which combine them to create a new set of outputs. Typical convolution procedures combine these functions to produce a new representation by filtering out functions depending on the convolution kernel. Two layers make up the depth-wise separable convolution: distinct layers for filtering and separate layers for merging. Two layers make up deep and separable convolution: depth wise convolution and point wise convolution. To apply a single filter to each input channel (input depth), depth wise convolution is utilized. Point wise convolution then produces a linear combination of depth layer output using a straightforward 1*1 convolution. The calculation and model size are greatly decreased as a result of this factorization.

PROPOSED SYSTEM ARCHITECTURE



RESULTS AND DISCUSSIONS

1.DEEP FAKE IMAGE



Upload a

Choose file real_im

False

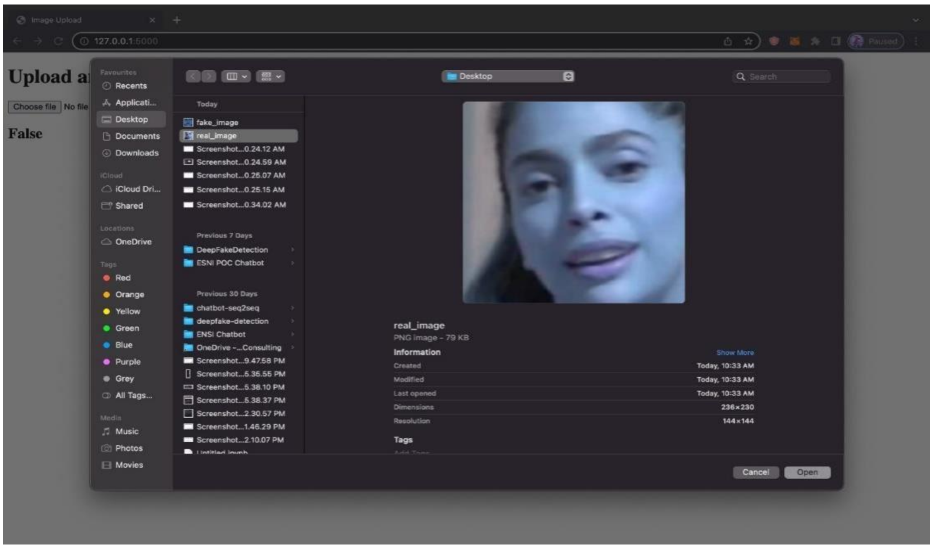
fake_image
PNG image - 79 KB

Information

Created Today, 10:32 AM
Modified Today, 10:32 AM
Last opened Today, 10:32 AM
Dimensions 232×234
Resolution 144×144

Fig1:Deep fake Image

2. NOT A DEEP FAKE IMAGE



Upload a

Choose file No file

False

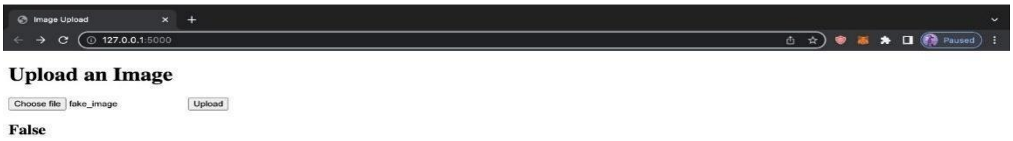
real_image
PNG image - 79 KB

Information

Created Today, 10:33 AM
Modified Today, 10:33 AM
Last opened Today, 10:33 AM
Dimensions 236×230
Resolution 144×144

Fig2:Real Image

3.DETECTING DEEPFAKES THROUGH FACIAL IMAGES



Upload an Image

False

Fig3:Detecting Facial Images

CONCLUSION

As viewing them no longer corresponds to believing in them, deepfakes have started to weaken people's faith in media material. They may distress individuals who are the target, exacerbate hate speech and disinformation, and even heighten political unrest, incite public unrest, violence, or even war. This is especially important today because social media platforms can swiftly propagate bogus news and deepfake creation technologies are becoming more accessible. This study offers a fresh viewpoint on the identification of deep fakes. So, this effort will be helpful for the artificial intelligence research community to provide new, efficient techniques for combating deepfakes.

FUTURE SCOPE

The feasibility of this project will increase the svm accuracy to 95% from its current 80% level**.**

REFERENCES

[1] Shruti Agarwal, Hany Farid, Yuming Gu, Mingming He, Koki Nagano, and Hao Li. Protecting world leaders against deep fakes. In Computer Vision and Pattern Recognition Workshops, volume

[2] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and PierreAntoine Manzagol. Extracting and composing robust features with denoising autoencoders. In Proceedings of the 25th International Conference on Machine learning.

[3] Diederik P Kingma and Max Welling. Auto-encoding variational Bayes. arXiv preprint arXiv:1312.6114, 2013.

[4] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. Advances in Neural Information Processing Systems, 27:2672–2680, 2014.

[5] Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, Ian Goodfellow, and Brendan Frey. Adversarial autoencoders. arXiv preprint arXiv:1511.05644, 2015.

[6] Ayush Tewari, Michael Zollhoefer, Florian Bernard, Pablo Garrido, Hyeongwoo Kim, Patrick Perez, and Christian Theobalt. High-fidelity monocular face reconstruction based on an unsupervised model-based face autoencoder. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42 (2):357–370, 2018.

[7] Jiacheng Lin, Yang Li, and Guanci Yang. FPGAN: Face deidentification method with generative adversarial networks for social robots. Neural Networks, 133:132–147, 2021.

[8] Ming-Yu Liu, Xun Huang, Jiahui Yu, Ting-Chun Wang, and Arun Mallya. Generative adversarial networks for image and video synthesis: Algorithms and applications. Proceedings of the IEEE, 109(5):839– 862, 2021.

[9] Siwei Lyu. Detecting 'deepfake' images in the blink of an eye. http://theconversation.com/detectingdeepfakevideos - in - the - blink - of - an - eye - 101072, August 2018.

[10] Bloomberg. How faking videos became easy and why that's so scary. https://fortune.com/2018/09/11/deepfakes-obama-video/, September 2018.