**AI for Deep Fake Images Detection**

**Sneh Raj1, Ashok Veda2**

AI Intern, Rubixe, Bengaluru, India

sneh.raj[at]rubixe.com

CEO, Rubixe, Bengaluru, India

ashok[at]rubixe.com

**ABSTRACT:** Over the last few decades, rapid progress in AI, machine learning, and deep learning has resulted in new techniques and various tools for manipulating multimedia. To standardize the evaluation of detection methods, we propose an automated benchmark for facial manipulation detection .Though the technology has been mostly used in legitimate applications such as for entertainment and education, etc., malicious users have also exploited them for unlawful or nefarious purposes. For example, high-quality and realistic fake videos, images, or audios have been created to spread misinformation and propaganda, foment political discord and hate, or even harass and blackmail people. The manipulated, high-quality and realistic videos have become known recently as Deepfake. Various approaches have since been described in the literature to deal with the problems raised by Deepfake. To provide an updated overview of the research works in Deep fake detection, we conduct a Proof of Concept(POC) in this paper, summarizing a variety of methodologies. We analyze them by grouping them into different categories: deep fake images detection, facial recognition, edge detection, feature extraction, depth estimation, . We also evaluate the performance of the detection capability of the various methods with respect to different datasets and conclude that the deep learning-based methods outperform other methods in Deepfake detection. This article proposes a solution to track images whether they were fake or real before uploading to any social media site so that users can make an educated choice they can take.

**INDEX TERMS:** Deep fake detection, Image manipulation, Image augmentation, Forgery detection model, CNN, RestNet18, RestNet50, ImageNet, VGG16, Sequential model, Inceptionv3.

1. **INTRODUCTION:**

DeepFakes has shown how computer graphics and visualization techniques can be used to defame persons by replacing their face by the face of a different person.Deepfake image detection refers to the process of identifying and distinguishing manipulated images created using deep learning techniques. These techniques, such as Generative Adversarial Networks (GANs), can generate fake images that appear realistic to human eyes. The potential impact of deep fake image detection is significant. Faces are of special interest to current manipulation methods for various reasons: firstly, the reconstruction and tracking of human faces is a well-examined field in computer vision, which is the foundation of these editing approaches. Secondly, faces play a central role in human communication, as the face of a person can emphasize a message or it can even convey a message in its own right.By detecting these unnatural patterns, it is possible to identify deep fake images .

Nowadays we have also see that using prompt

anyone can create different images however they want and they can use it without the consent of that person in any social media platform and people are unaware of that. So introducing deep fake image detection we have built an application where these instances can be identified easily with the help of real face detection that we have used to train our model.

In late 2017, Motherboard reported on a video that had appeared on the Internet in which the face of Gal Gadot had been superimposed on an existing pornographic video to make it appear that the actress was engaged in the acts depicted. Despite being a fake, the video quality was good enough that a casual viewer might be convinced – or might not care. An anonymous user of the social media platform Reddit, who referred to himself as “Deepfakes,” claimed to be the creator of this video.

The term ‘‘Deepfake’’ is derived from ‘‘Deep Learning (DL)’’ and ‘‘Fake,’’ and it describes specific photo-realistic video or image contents created with DL’s support. This word was named after an anonymous Reddit user in late 2017, who applied deep learning methods for replacing a person’s face in pornographic videos using another person’s face and created photo-realistic fake videos. However, considering the threats and potential risks in privacy vulnerabilities, the study of Deep Fake emerged super fast. Rossler et al. introduced a vast video dataset to train the media forensic and Deep Fake detection tools called FaceForensic in March 2018.

We found the following additional information from this report [6]: •

* The top 10 pornographic platforms posted 1,790+ Deepfake videos, without concerning pornhub.com, which has removed ’Deepfakes’ searches.
* Adult pages post 6,174 Deepfake videos with fake video content.
* New platforms were devoted to distributing Deep Fake pornography.
* In 2018, 902 articles were published in arXiv, including the keyword GAN either in titles or abstracts.
* Apart from Deep Fake pornography, there are many other malicious or illegal uses of Deepfake, such as spreading misinformation, creating political instability, or various cybercrimes.

To address such threats, the field of Deepfake detection has attracted considerable attention from academics and experts during the last few years, resulting in many Deepfake detection techniques.

1. **Defining the problem:**

We have seen many incidents and cases related to deep fake image generation where many peoples are unaware of that and get troubled like someone uses their faces in pornography video and just superimposed them and the video even looks more realistic. We are also aware of the fact that we have seen even more cases than just pornography, people use it for blackmailing, use it for fun, cybercrimes, use it on social media platforms, for publicity and so on.

Deepfakes actually represent a subset of the general category of “synthetic media” or “synthetic content.” The synthetic media refers to the media that was created through Artificial Intelligence and with the help of computer vision and the modification part of the images generally contains algorithmic functions and it depends upon the content you want. On the other hand, synthetic content refers to the Images, audio, video, text that was generated through AI unlike the content that was created by humans.

**2.1. Where is this problem coming from:**

The problem of deep fake images comes when there is an advancement in technology and when there is an ease of generating new things from the preexisting images comes into picture. When people think that using these things to manipulate anyone and use it for their benefits.

They also want to use it as a false evidence purpose and communicate it as false information.

People also use deep fake images to defame any celebrities as they were using their face and change it with someone else. Using someone’s face leads to theft, fraud and scams as can be rolled easily with visual recognition system.

Many popular articles on this subject define synthetic media as any media which has been created or modified through the use of artificial intelligence/machine learning (AI/ML), especially if done in an automated fashion.