OPEN ACCESS

International
Journal
Of Modern Engineering Research (IJMER)

Deepfake Image and video detection using deep learning method

Rana Amer Sattar ,Dhuha Mohammed Noori Mohammed Raouf[2],
Rasha Thamer shawi[3]

[1]*Directorate of Education Rusafa First as an assistant teacher at Al-Rafidain Secondary School for Girls Ministry ofEducation, Baghdad, Iraq.*
[2]*AL- Sayidiyah Vocational Preparatory School for Girls, Al- Karkh Second Vocational Education Directorate, Iraqi Ministry of Education, Baghdad, Iraq.*
[3]*Department of Computer Science (College of Science) Al-Mustansiriyah University Baghdad, Iraq*

Abstract- *Deep fake is considered a form of artificial intelligence that relies on a database of images and video clips with the aim of creating "imaginary masked" images or clips with audio and video, with no obstacles to the creation of texts, audio and video clips by artificial intelligence, the possibility of misuse Impersonation, financial fraud or defamation is of global concern. and that Artificial intelligence technological advancements will undermine public trust, strengthen populists and authoritarians, and destabilize markets and enterprises. Also, progress in deepfakes, facial recognition, and voice synthesis will make a person's control over his image a thing of the past, and the spread of deepfakes may lead to the informational end of the world war, according to a scenario in which many people are unable to distinguish between reality and imagination and know what they are Reliable news sources. This paper presents a comprehensive review of the most recent studies of deepfake detection using deep learning-based approaches. We aim to broaden the scope of modern research by systematically reviewing the different categories to detect dummy content. And make a comparison between them.*

Keywords: *fake content, machine learning, deep learning.*

Date of Submission: 15-01-2024 Date of acceptance: 28-01-2024

I. Introduction

Deepfake technologies have turned AI into a primary vehicle for spreading disinformation at breakneck speed: from chatbots that spread lies, to face-swapping apps used to fake porn videos, to voice-spoofing programs, Controlling this technology has become a major goal for many countrieswith a torrent of electronically generated images and videos showing people saying things they have never said, artificial intelligence has called the adage "don't believe before you see" into question and led more people to distrust online content. And that most countries are trying to keep up with this quickly growing technology, amid concerns that establishing restrictions in this field may stifle innovation or be abused to limit free expression [1]. using no boundaries to synthesizing text, audio, and video using AI, the potential for misappropriation for impersonation, financial fraud, or defamation is a global concern. Artificial intelligence breakthroughs will erode social trust, empower populists and authoritarians, and disrupt businesses and markets, and advances in deepfakes, facial recognition, and voice synthesis will make controlling a person's image obsolete[2].Tools (detecting a fake from the original) are evolving quickly, but deepfake technology is possibly going even faster. As a result, we will not be able to find a solution to cyber security, and all we can do is try to keep abreast of its developments. The process of using artificial intelligence to manipulate a video clip is based on a technique known as the A machine learning model called a generative adversarial network pits two neural networks against one another in an effort to produce more accurate results. This technique seeks to enhance outcomes[3].We may find it difficult to understand what is meant by the competition of two networks, but how it works is as follows: one computer will tell the other computer if the digital reproduction, whether video or audio, which it has made to any person is convincing and is sufficiently identical to that person's original; Meaning, is the movement of the lips in the cloned version identical to the original version, as well as are the facial expressions identical. Using the Generative Adversarial Network, the system as a whole is optimized until a satisfactory result is reached. Although this technology is constantly developing and improving, anyone can detect manipulation, whether visual or audio, in other words, catching the "deep fake" if the thing we are looking for is identified[4].

In (2020). Karandikar, A., Deshpande, V., Singh, S., Nagbhidkar, S., & Agrawal, S. presented convolutional neural network for detecting deep fake videos. A deep-learning system may create a convincing copy by examining several photos and videos of a target individual, then imitating their behaviour and voice patterns. Because more realistic deepfake production technologies are always being developed, it is extremely difficult to detect these videos. By putting out a model that uses deep learning to analyse video frames in order to find disparities in facial feature consistency, compression rate, and other issues presented during video creation, the goal is to address this issue. The model, which can detect these embedded faults in the deepfakes, is trained using a convolutional neural network and transfer learning. The neural network is trained using the differences that were introduced in the area of the face during deepfake development. The model is trained on a dataset titled "Celeb-DF: A New Dataset for Deep Fake Forensics," and it goes into great detail about the methods that could be used to enhance this model's learning[5].

In (2021). Ramachandran, S., Nadimpalli, A. V., &Rattani, A.presented an experimental evaluation on deepfake detection using deep face recognition.The majority of the current deepfake detection techniques use two-class convolutional neural networks (CNNs) to solve the binary classification problem of identifying real images or videos from fake ones. These techniques work by identifying visual glitches, temporal irregularities, or colour discrepancies brought on by deep generative models. For model training, these methods need a lot of genuine and fake data, however when compared to samples produced using sophisticated deepfake creation techniques, their performance suffers noticeably. Produced carefully evaluates the efficiency of deep facial recognition in this study's attempt to identify deepfakes by making use of a variety of loss functions and deepfake generating techniques. Deep face recognition is more effective at spotting deepfakes than other methods, according to experimental studies on the difficult Celeb-DF and FaceForensics++ deepfake datasets over the ocular modality and two-class CNNs. According to published studies, using face recognition on the Celeb-DF dataset, it is possible to detect deep fakes with an Area Under Curve (AUC) of up to 0:98 and an Equal Error Rate (EER) of up to 7:1%. Compared to the EER achieved for the two-class CNN and the ocular modality on the Celeb-DF dataset, this EER is reduced by 16:6%. On the FaceForensics++ dataset, further analysis yielded an AUC of 0:99 and EER of 2:04%. Bypassing the requirement for a significant amount of fake data for model training and achieving improved generalizability to developing deepfake production procedures are two benefits of using biometric facial recognition technology [6].

in (2022). Sabah, H. A. Deep Fake Detection in Face Images Using Deep Learning was given. Finding bogus images is a serious issue that needs to be managed and prevented from having many negative impacts. Convolution Neural Network, the most well-liked deep learning algorithm, is suggested for use in identifying fraudulent photos. Pre-processing is one of the first phases, which entails first converting images from RGB to YCbCrcolor format, then doing gamma correction. Finally, add the Canny filter to them to extract edge detection. Utilizing both convolution neural networks with and without principal component analysis. there are two forms of neural networks: However, convolution neural networks without principal component analysis the latter is considered as a classifier, two alternative detection methods are then used. The findings show that this research's application of CNN and PCA yields satisfactoryaccuracy. Using CNN, on the other hand, only provided the maximum level of accuracy in detecting manipulated photos[7].

in (2021) Shad, H. S., Rizvee, M., etal.Comparative examination of deepfake image detection method utilizing convolutional neural network was reported. Neuroscience and Computational Intelligence. The major goal is to accurately distinguish deepfake photos from real ones. In this study, multiple approaches were used to detect deepfake photos and compare them. A total of 70,000 photos from the Flickr dataset and 70,000 images created using style GAN were included in the Kaggle datasets used to train the model. For this comparative study on using convolutional neural networks (CNN) to distinguish between real and deep fake images, eight different CNN models were created. These models were trained using a variety of architectures, including the DenseNet architecture (DenseNet121, DenseNet169, and DenseNet201), the VGG Net architecture (VGG16, VGG19), the ResNet50 architecture, the VGGFace architecture, the ResNet50 architecture, and a bespoke CNN architecture, Additionally, a unique model was built that contains approaches such as dropout and padding that aid in determining whether the other models meet their objectives Five evaluation measures were used to summarize the results: accuracy, precision, recall, F1-score, and area under the ROC (receiver operating characteristic) curve.With 99% accuracy, VGGFace outperformed all other models. Additionally, we obtained 90% from the custom model, 97% from ResNet50, 96% from DenseNet201, 95% from DenseNet169, 94% from VGG19, 92% from VGG16, and 97% from DenseNet121 [8].

In (2018). Korshunov, P., & Marcel, S.Deepfakes: A New Threat to Face Recognition? This is the first group of Deepfake films from the VidTIMIT database that are publicly accessible. GAN-based open-source software was used to construct the Deepfakes, and we emphasize that the training and mixing parameters can significantly affect the calibre of the resulting videos.By changing the parameter sets, we produced videos with low and high visual quality (320 videos each) to illustrate this influence. that showed how sensitive modern facial recognition systems are to Deepfake videos, with erroneous acceptance rates of 85.62% and 95.00%, respectively (on high quality versions). This implies that techniques for identifying Deepfake videos are required.We discovered proved it was impossible to discern Deepfake videos using an audio-visual technique based on lip sync inconsistency detection. On high grade Deepfakes, the highest performing technique, which is based on visual quality measurements and is frequently employed in the detection of presentation attacks, produced an equivalent error rate of 8.97%. The findings demonstrate that GAN-generated Deepfake films are challenging to distinguish using both face recognition systems and current detection techniques, and that the development of face swapping technologies will only make this task more challenging. Deepfake films, face swapping, video databases, tamper detection, and face recognition are all index terms [9].

in (2020), Pu, J., Mangaokar, N., Wang, B., & etal.Deep-fake picture detection using noise scope in a blind environment. This article presents a blind detection technique. dubbed Noise Scope for finding GAN images among other genuine photos in the Annual Computer Security Applications Conference. A blind approach generalizes better than supervised detection methods and doesn't require prior knowledge of GAN images for training. Our most important discovery is that GAN images also contain distinctive patterns in the spacenoisy just as images from cameras. To distinguish GAN images, -we extract these patterns in an unsupervised way. We test Noise Scope on 11 different datasets that contain GAN images and are able to recognize GAN images with an F1 score of up to 99.68%. We test Noise Scope's restrictions against a range of countermeasures and find that Noise Scope holds up well or is flexible [10].

in (2020). Singh, S., Sharma, R., & Smeaton, A. F. synthesis of deep fake generation's minimum training data using GANs. This study examined how deepfake films of non-celebrity persons might be produced with little training data, or only a few training photos. The range of face expressions among the few photographs utilized, in addition to the small number of images used, is of particular interest. In order to test this, they created a model of each individual face using a big number of photographs. From this model, they created a few lifelike but artificial image that they used to develop a deepfake. Although it may seem counterintuitive to create a small number of synthetic images of a celebrity from a large number of images of that celebrity, doing so enables the synthetic images to include a variety of facial expressions that would be challenging to obtain if we were to use a real collection as the small number of deepfake training images [11].

in (2022). Mitra, A., Mohanty, S. P. & et al.An a reliable detection method that is IoT-friendly for deep-fake images produced by GANs on social media. Regarding the Internet of Things. technologies and their uses. The development of deep learning technology, in particular, has ushered in a new era of multimedia forgeries. Deepfake elevates everything to a new level. With features learned from a previous collection of photos, this deep learning-based approach creates brand-new images. Deepfakes are a realistic choice because to the Generative Adversarial Networks (GANs) rapid growth. Deep learning is used to create extremely complex and realistic images, and image-to-image translation is utilized to implement deepfake. This research proposes a novel, lightweight deepfake detection method based on machine learning that has been successfully used in an IoT platform. An API for detection is also suggested in addition to the detection technique. To the authors' knowledge, this is the first attempt at the cutting edge for identifying very sophisticated GAN-generated deepfake images. The work is innovative in that it achieves a high level of accuracy with minimal training time and edge device inference. The entire process of sending the image to the edge, detecting it, and showing the outcomes via the API appears to be working well. There is also some discussion of how to speed up inference and improve accuracy. The Horlicks texture attributes of contrast, dissimilarity, homogeneity, and correlation are some of the three steps of a three-stage textual analysis that is used in comparison research. The other phase is computing Shannon's entropy. The findings demonstrate that the created false images differ significantly in entropy, contrast, dissimilarity, homogeneity, and correlation even when they appear to be identical to the matched genuine images [12].

in (2020). Mehra, A.Deepfake detection employing extended short-term memory networks in capsule networks. Frame-level representation and long short-term memory (LSTM) networks using a Capsule Network (Capsule Net) to build a spatiotemporal hybrid model, a method for identifying deepfake movies is proposed in this research. analysed the influence of frame sequence choice on fraudulent video detection as well. By illustrating how capsules are activated, you may further clarify how the Capsule Network determines whether a sample is legitimate or fraudulent. and contrast the outcomes produced by the suggested model with those produced by cutting-edge techniques.The effectiveness of the false video identification algorithm is crucial given the enormous volume of video data that is available on the Internet. With roughly 1/5 the number of parameters and performance on par with cutting-edge models, the combined Capsule and LSTM network has lower computational costs[13].

in (2021) Liu, M. Y., Huang, X., Yu, J., Wang, T. C., & Mallya, A. Algorithms and applications for generative adversarial networks for Synthesis of images and videos. For a range of image and video synthesis tasks, the Generative Adversarial Network (GAN) framework has proven to be an efficient tool for these applications, enabling conditional or conditional tuning of visual content. It made it feasible to create photorealistic films and high-resolution photos, a task that would have been challenging or impossible using earlier techniques. Numerous new applications for content generation have also resulted from it. With an emphasis on optical texturing algorithms and applications, they provide a general review of GANs.in this study. They discuss a number of critical stabilization strategies for famously challenging to train GANs. They also talk about how it can be used for neural displays, image processing, video synthesis, and image translation[14].

Conclusion

We have provided a quick overview of a few studies that discuss several techniques for identifying deepfake movies and images in this document. How those procedures can be merged or adjusted in a new project to produce outcomes that are more accurate than those produced by the practices already in use, Fake technologies based on deep learning have been expanding at an unprecedented rate in recent years. The potential of deepfake algorithms to produce malevolent face-altered movies allows for their rapid dissemination, putting social stability and individual privacy at risk. Commercial enterprises and academic institutions throughout the world are performing pertinent studies to this goal in an effort to lessen the harmful effects that deepfake videos have on viewers. In this essay, we explain detection algorithms and underline the importance of promising research.in this analysis we expect that by reducing the negative effects of deepfake films, this article will be helpful to academics working on deepfake detection research.

SUMMARY OF REMARKABLY RELATED ARTICLES ON DEEPFAKE

Serial	Paper name	Objective	Dataset	Methodology	Conclusion
1		Using a deep learning approach, create a model that analyzes the video frames.	an image pre-processing approach and a convolutional neural network (CNN),	1-A sizable collection of genuine and false videos that are converted to frames. 2. A collection of facial features that have been extracted and aligned, on which a convolutional neural network (VGG16 or OxfordNet) that has been previously trained extracts hidden features is used. 3. A model that, after post-processing for videos, learns from the transformed dataset and divides the data into real and fraudulent.	obtained that can tell a real image from a phony one. This is accomplished by using a dataset to train the model.
2		evaluate how well deep facial recognition 1, which was trained using different loss functions, performs at spotting deep fakes produced in a variety of ways.	CNNs, FaceForensics++, the ocular modality, and Celeb-DF.	1-FaceSwap: An identity swapping technique that employs a graphics-based strategy based on recognized facial landmarks to move the face region from a source video to a target video. It makes use of face alignment, Gauss-Newton optimization, and picture blending to switch the source person's face for the target person's face. 2-FaceShifter: This identity swapping technique employs a new In order to adaptably combine the identity and the attributes for face synthesis, a face synthesis generator with Adaptive Attentional Denormalization (AAD) layers and an attributes encoder with the ability to extract multi-level target face attributes are used.	used the idea of detecting corrupted facial features rather than image anomalies to test the performance of deep face recognition in detecting high-quality deepfake photos or videos from the actual ones of the same identity. Experimental results showed that face recognition technology was more effective than two-class CNNs on the same datasets at detecting deepfake-based identity swapping techniques. The best deepfake detection rate was achieved using a combination of margin and CosFace loss algorithms.
3		detect the fake images using Convolution Neural Network	CNN with PCA CNN without PCA	1- changing the color space of images from RGB to YCbCr and applying the gamma adjustment.	Convolution Neural Network, the most widely used deep

				2. two distinct ways of detection are used. Where Extraction of edge detection through application of the Canny filter. Then, as a classifier, we used both Convolution Neural Network with Principal Component Analysis and Convolution Neural Network without Principal Component Analysis.	learning technique, is used in a detection model to find fraudulent images.
4		detect deepfake images from real ones accurately. Using eight different CNN models	convolutional neural networks (CNN)	The dataset was assembled before it began. In order to achieve the optimal outcome, the features have been extracted and implemented using a variety of CNN architectures. Finally, the accuracy, precision, recall, and F1-score metrics were used to evaluate each model. Last but not least, another criterion for evaluating the effectiveness of the models was the area under the ROC curve.	Identify whether the images are authentic or false. People will continue to be cautious because they can spot a deep phony image with 99% accuracy.
5		detection of the Deepfakes using setsof videos from VidTIMIT	sets of videos from VidTIMIT	1-Treating Deepfake videos as digital presentation attacks, using methods such as linear discriminant analysis (LDA), simple principal component analysis (PCA), and the technique based on image quality measurements (IQM) and support vector machines (SVM). 2- To make the database of Deepfake videos7, face recognition systems, and Deepfake detection systems with associated scores available as an open-source Python program, allowing researchers to check, replicate, and expand the work.	provided an initial database of 620 Deepfake videos from the Vid-TIMIT database with an accuracy rate of 97.8% for 16 pairs of subjects.
6		building adeepfake detection scheme.	NoiseScope	1-One discovery is that fraudulent images have distinctive Low-level noise patterns that are related to the GAN model that generated them. These patterns are related to the deconvolution layers of GANs. 2 describe the development of NoiseScope, a blind detection method that makes use of distinctive patterns in bogus images. In order to recognize a GAN, NoiseScope collects any model fingerprints or patterns that are present, then utilizes the fingerprint to identify bogus images in the set. 3 tested NoiseScope on 4 excellent GAN models that were used to produce 11 different deep fake picture datasets. NoiseScope has an F1 score of up to 99.68% for detecting bogus images.	offer Noise Scope, a technique for blindly identifying deep fakes.
7		how to create deepfake videos of non-celebrity persons with a small amount of training photos and training data. We are particularly interested in the variability in face expressions among the few photographs utilized, in addition to the small number of	StyleGAN2	1-To determine the level of variability among the generated faces, Open Facewas applied to a set of images produced by StyleGAN2 and the mean and variance of inter-image scores were calculated. 2- The Varied" dataset, the second dataset, contains a variety of face expressions that were recorded. This necessitates 4,950 comparisons for each dataset from which we derive mean and variance. The Monotone	A concept was developed to combine different methods for creating video dialogue of person speaking to the camera using a small dataset of the subject's images.

		images used.		dataset has a lower mean and variance score, which indicates that the person in the dataset is the same but has less variation in their facial expressions than in the other dataset, which has more variation.	
8		They suggest a machine learning (ML) based deepfake detection system that can be installed in an IoT end device to allow users to verify the validity of any image at any time and from any location. This will assist in stopping the spread of rumors or false information.	StarGAN and CycleGAN datasets	1. The classifier learns to recognize bogus photos during the training phase. On a PC, the first training was completed. Retraining is possible in the cloud if necessary. 2. Testing Phase: The untested samples are evaluated during this phase. At the edge device, this is put into practice. Through the suggested API, the testing process is carried out.	As little computation as feasible. they used 30 attributes each image in order to accurately deduce at an IoT device with limited resources.
9		create machine learning models that can be used to distinguish between authentic and fake media material.	Capsule and LSTM network	1-account a measure of the structural similarity (SSIM) between two successive frames by choosing 10 frames from a video at one-second intervals. choose the pair of frames with the least SSIM out of the 10 chosen frames; 2choose ten evenly spaced frames, including the start and terminus, from the chosen interval.	The suggested CapsuleNet+LSTM model is better able to generalize on data that has been changed invisible and on fraudulent videos made using novel tampering methods.
10		Differentiate between authentic and false data.	GANs	1-A training environment where photos of the same individual in various stances are provided. 2: The synthesized image was back-rendered to its initial position and a cycle-consistency constraint was applied.to automatically separate the shape and look from images, use a two-stream auto-encoding architecture..	GAN network design development and GAN stabilization techniques

Reference

[1]. Ebermann, A. (2021). The effects of deepfakes and synthetic media on communication professionals.

[2]. Westerlund, M. (2019). The emergence of deepfake technology: A review. Technology innovation management review, 9(11).

[3]. Van Huijstee, M., Van Boheemen, P., & Das, D. (2021). Tackling deepfakes in European policy.

[4]. Van der Sloot, B., &Wagensveld, Y. (2022). Deepfakes: regulatory challenges for the synthetic society. Computer Law & Security Review, 46, 105716.

[5]. Karandikar, A., Deshpande, V., Singh, S., Nagbhidkar, S., & Agrawal, S. (2020). Deepfake video detection using convolutional neural network. International Journal of Advanced Trends in Computer Science and Engineering, 9(2), 1311-1315.

[6]. Ramachandran, S., Nadimpalli, A. V., &Rattani, A. (2021, October). An experimental evaluation on deepfake detection using deep face recognition. In 2021 International Carnahan Conference on Security Technology (ICCST) (pp. 1-6). IEEE.

[7]. Sabah, H. (2022). A Detection of Deep Fake in Face Images Using Deep Learning. Wasit Journal of Computer and Mathematics Sciences, 1(4), 94-111.

[8]. Shad, H. S., Rizvee, M. M., Roza, N. T., Hoq, S. M., Monirujjaman Khan, M., Singh, A., ... & Bourouis, S. (2021). Comparative analysis of deepfake image detection method using convolutional neural network. Computational Intelligence and Neuroscience, 2021.

[9]. Korshunov, P., & Marcel, S. (2018). Deepfakes: a new threat to face recognition? assessment and detection. arXiv preprint arXiv:1812.08685.

[10]. Pu, J., Mangaokar, N., Wang, B., Reddy, C. K., & Viswanath, B. (2020, December). Noisescope: Detecting deepfake images in a blind setting. In Annual computer security applications conference (pp. 913-927).

[11]. Singh, S., Sharma, R., & Smeaton, A. F. (2020). Using GANs to synthesise minimum training data for deepfake generation. arXiv preprint arXiv:2011.05421.

[12]. Mitra, A., Mohanty, S. P., Corcoran, P., &Kougianos, E. (2021, November). EasyDeep: An IoT friendly robust detection method for GAN generated deepfake images in social media. In IFIP International Internet of Things Conference (pp. 217-236).

[13]. Mehra, A. (2020). Deepfake detection using capsule networks with long short-term memory networks (Master's thesis, University of Twente).

[14]. Liu, M. Y., Huang, X., Yu, J., Wang, T. C., & Mallya, A. (2021). Generative adversarial networks for image and video synthesis: Algorithms and applications. Proceedings of the IEEE, 109(5), 839-862.