Analysis Survey on Deepfake detection and Recognition with Convolutional Neural Networks

Saadaldeen Rashid AHMED
Computer Engineering
Karabuk University
Karabuk, Turkey
saadaljanabi95@gmail.com

Emrullah SONUÇ
Computer Engineering
Karabuk University
Karabuk, Turkey
esonuc@karabuk.edu.tr

Mohammed Rashid AHMED
Computer Engineering
Karabuk University
Karabuk, Turkey
mohammed.r.aljanabi@gmail.com

Adil Deniz DURU
Physical Education and Sports
Marmara University
Istanbul, Turkey
deniz.duru@marmara.edu.tr

2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA) | 978-1-6654-6835-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/HORA55278.2022.9799858

***Abstract*—Deep Learning (DL) is the most efficient technique to handle a wide range of challenging problems such as data analytics, diagnosing diseases, detecting anomalies, etc. The development of DL has raised some privacy, justice, and national security issues. Deepfake is a DL-based application that has been very popular in recent years and is one of the reasons for these problems. Deepfake technology can create fake images and videos that are difficult for humans to recognize as real or not. Therefore, it needs to be proposed some automated methods for devices to detect and evaluate threats. In another word, digital and visual media must maintain their integrity. A set of rules used for Deepfake and some methods to detect the content created by Deepfake have been proposed in the literature. This paper summarizes what we have in the critical discussion about the problems, opportunities, and prospects of Deepfake technology. We aim for this work to be an alternative guide to getting knowledge of Deepfake detection methods. First, we cover Deepfake history and Deepfake techniques. Then, we present how a better and more robust Deepfake detection method can be designed to deal with fake content.**

Keywords— Deep-fakes, face exploitation, AI, DL, auto-encoders, generative adversarial network, forensics, review

I. Introduction

When it comes to Deep fakes, a technique that can overlay the facial image of a specific user onto the video of an original individual to create the video recorder to the goal making or telling the same things that the actual person does is called "deep fake." Face swap is a type of Deepfake that falls into this category. Deep fakes can also be lip-syncing or puppet-masters, depending on how the content is created by utilizing artificial intelligence. Deep fakes with lip sync are videos in which the mouth movements are synced to the audio. Videos of puppets keeping an eye on the face mask gestures, watching activities, and head movements of a grand in the head of a photographic camera are known as puppet-master deep fakes.

A few Deep fakes can be generated by utilizing traditional visual effects or computer graphics. Still, DL techniques like autoencoders and the area to computer vision (GAN) have been commonly used. Facial expressions and movements are studied using these models, generating other people's faces and body motions with comparable expressions and actions. Deepfake technologies often demand a vast quantity of picture and video info to train design of identify picture-realistic pictures and movies. Nation figures like personalities including candidates are the first to be targeted by deep fakes because they have many images as well as videos online. Deep fakes have been used in pornographic images and videos to replace the faces of celebrities and politicians and those of more people's bodies. Since then, many other celebrities have replaced their faces with those of porn actors in Deepfake videos.

II. Related Work

The face is the most prominent one in terms of distinguishing human characteristics. Face synthesis technology is advancing rapidly and posing an ever-greater threat to national security. Because of the plethora of sets of rules based on DL expertise, it's possible to replace the faces of real people with fake ones that look convincing. In deep fake, a face from one person is superimposed on the front of another, a new subfield in artificial intelligence. High-resolution Deepfake images can be generated by using generative adversarial networks (GANs) [1].

Deepfake information is spreading more quickly than ever before in the twenty-first century because of the increasing use of cell phones and the creation of multiple public net sites [2]. The pixel collapse phenomenon, which tends to cause strange visual anomalies in the membrane tone or face from photos, initially made Deepfake images visible to the human eye. Deep fakes can be created from audio as well as visual media. As technology has advanced, deep fakes have become nearly indistinguishable from real images [3]. As a result, individuals worldwide are dealing with a wide range of difficulties. Fashion and e-commerce businesses benefit from the usage of Deepfake technology, which allows consumers to make their purchases more rapidly.

The entertainment industry is also benefiting from this technology, as it provides artificial voices for musicians who can't label promptly. Deepfake technology also allows filmmakers to reproduce or employ special effects in many iconic sequences. Alzheimer's patients may be able to converse as well as a newer version of their correct utilization Deepfake technology, which might help them preserve recollections. The use of GANs to detect anomalies in X-ray pictures is also being researched [4]. If the witnesses are persuaded to trust the Deepfake approaches, they generally require many pictures, video, or audio data to make natural photographs. In addition to all the attention, there are some serious downsides. Celebrities, athletes, and politicians, to name just a few, are particularly vulnerable to Deep fakes

because of the sheer volume of publicly available media they generate. Aside from making fun of others, deep fake technologies are mostly utilized to make adulterous content. Famous people's faces have been pasted onto pornographic models, as well as the resulting photos can be found in abundance on the Internet [2]. It is possible to make humorous, pornographic, or political content utilization images as well as voices of individuals you know without their permission. It is now possible for anyone to create any artificial content that is completely undetectable from the original content [2].

Cyberbullying is a growing problem for today's youth. Numerous victims take their own lives in the worst-case scenario. A video of former American president Barack Obama stating things he has never said is circulating on the Internet these days. Joe Biden's film from the 2020 U.S. presidential election, which shows him with his tongue out, has already been altered utilization Deep-fakes. Along with Taylor Swift and Gal Gadot, Deepfake technology has also snared Emma Watson and Meghan Markle [5]. Many women in the United States, as well as Asia, are also victims of deep-fake technologies. On social media [6], the destructive usage of deep fakes can significantly impact our culture and enhance the spread of false information. Nevertheless, deep fakes need to pose a major concern of the existing creation due to the harm they do to a variety of individuals and businesses. As a result, academics have been doing tirelessly to detect Deep fakes to eliminate libel, swindles, deceit, and insecurity from the public. Because of the discovery of Deep-fakes, crimes will be reduced over the world. Researchers have focused on their validation process [2]. Some global corporations have taken measures in response to this tendency. Scholars can use a false video database provided by Google, Facebook, and Microsoft to develop new rules for detecting Deep fakes [7].

Machine learning classifiers such as the SVM Algorithm and the naive set of rules may also be used to identify G.A.N generated Deepfake pictures, along with many other techniques, like CNN and recurrent NNS. The work's key contribution is the use of CNN architecture to recognize as well as distinguish Deepfake images from normal ones. An array of convolutional neural network architectures, such as DenseNet169, Dense-Net121, DENSNET201, and VGG16, 19, VGG-Face, as well as ResNet50, were used to detect Deepfake images in this study. To perform a comparison, a new custom model was developed using the data set taken from Kaggle. The data was acquired prior to the start of the project. Since the features have been identified, implementing several CNN designs has yielded the best outcomes so far. Use these four criteria to evaluate each model's accomplishment: accuracy, precision, recall, and F1-score. And as an added metric of model performance, we looked at its "A.U.C." (area under the curve). The capacity of DL to handle large and complex datasets is well-known. Deep autoencoders have extensively used this feature for dimension reduction and picture compression [8–10], a deep CNN. Using an auto-encoder-decoder coupling, a Reddit user created. The Fake App was the first and remained a crucial part of Deepfake [11,12].

The auto-encoder captures underlying properties after face mask photographs, and the decoding re-builds of pictures in this fashion. Double codec couples are needed to swap looks among source and target photographs, encoders settings are shared via dual net pairings, and each couple is being utilized to learn at image capture. In both pairs, the encoder networks are identical. Deepfakes [13], Faker [14], or Deep Facet "TensorFlow-based Deep-fakes" [15] use this encoder-decoder architecture in their research, as does DF [14]. An upgraded form of Deep fakes built on the novel adaptive G.A.N [10], of face change G.A.N, has been proposed in [16] via combining of antagonistic failure as well as the lost perceptual opportunity of the encoder and decoder design like performed in VGG-Face [17].

A multi-task CNN is introduced in the Face Net execution [18] to enhance face recognition and orientation consistency. Generative networks can be built with the Cycle GAN [19]. There is an increasing breach of privacy, security, and democratic republic from Deep fakes [20]. Immediately after discovering the dangers of deep fakes, a strategy for monitoring them was devised. DL has been used in recent approaches to automatically identify deep fakes [21, 22]. Utilization of the open-source Face swap-GAN [19] code, Korshunov as well as Marcel [23, 24] created a deep fake data set including (620) records utilization of the G.A.N shape to address this trouble. The VidTIMIT database [25] was used to create low down and high ranking-quality Deepfake films that accurately mimicked facial scrub expressions, brim movement patterns, and look blink. The popular V.G.G. as well as Face net [18, 26]. Based on facial recognition set of rules is ineffective at detecting deep-fake faces in tests. Forensics models have become more challenging to use since DL methods such as CNN and G.A.N can enhance the readability, face representation, and illumination in pictures [27].

Self-interest is the driving force behind BIGGIN. G.A.N. and spectroscopic normalization G.A.N have created 128 pixel-sized fake pictures. [28,27,28]. G.A.N is also used for spectral normalization. Agarwal and Varshney [30] studied the authenticity of G.A.N-based Deep fakes [31] using statistical frameworks, even though they can be detected using G.A.Ns which are trained to find fakes. When used to identify Deepfake cinema in this recently founded dataset, other methodologies, like lip-syncing approaches [32–34] and photo quality of SVM [35], create highly high error rates. An SVM classifier uses the derived features to arrive at detection models. In Zhang et article .'s [36], classifiers such as SVM, R.F., and M.L.P. were used to extract a catalog of compressed characteristics fed into the classifiers. [37, 38, 39] By Hsu et al., DL could be used to recognize fake images. Dense Net and gated recurrent unit cells were added to make an R.C.N. that could benefit temporal differences among frames in the first phase of the feature extractor using the Siamese network configuration described in [41]. With a set of data of 1,000 videos, the proposed method was tested and found to be promising (Face Forensics++ [44]). According to Guera and Delp [45], interframe and temporal inconsistencies are prevalent in Deepfake videos. They proposed a time-aware CNN-based technique for identifying deep-fake films (LSTM). Regular videos blink much faster than deep fakes. Li et al. [46] used seven eye landmarks to deconstruct videos into images. These neatly trimmed eye groundbreaking scenes are dispersed into the long and complex recurrent convolutional nets (LRCN) [47] to complex nonlinear prediction. Nguyen et al. [48] recommended utilizing container nets to find fake images as well as videos. A Cylinder network has been introduced to overcome of CNNs if used for opposite special effects tasks [49]. Recent research found a cylinder net

established active routing to direct hierarchy cause intercommunication [50].

The Face to Face method data set [51-53] and the wholly computer-generated picture dataset [54] are examples of these datasets. Picture response nonuniformity (PRNU) assessment can tell legitimate deep fakes from counterfeits, according to [55]. The digital camera's thumbprint is called PRNU [56 - 60]. The did try to swap eyes is aimed to alter municipal PRNU trends in the face skin. To develop tools that automatically assess a "picture's or video's" reliability. The current study examined feature-based and CNN-based reliability assessment methods [61, 62]. Alternatively, in his paper, Z. [66] suggested utilizing 2 stream system for 2 different face exchanging processes [65]. One set of data by Rossler [67] includes 50,000 altered images that practitioners will find useful. The article is laid out: Second, we studied seminal papers on sighting deep fake images.

This is a primary target of the present article to classify deep fake pictures from real ones. Several studies have been conducted on the delicate subject of "Deepfake." Deepfake images were detected by many researchers' utilization of a CNN-based strategy, while others relied on feature-based techniques. Some of them utilize machine learning classifiers to identify Deepfake images. However, the contribution of this paper would be that it uses the VGG-Face model to accurately detect Deepfake pictures from normal images. We use more CNN models in our research than any other team, setting us apart from the competition. We have demonstrated a thorough analysis, and the results have outperformed previous efforts in our work.

III. METHODOLOGY

The basic diagram from several DL architectures is shown in Figure 1. The first step was to gather the data as well as then extract the relevant features from it. Thus, a total of eight neural network architectures were used, each assessed against five distinct evaluation metrics, such as accuracy, accuracy, F1-score, recall, as well as the region under the R.O.C. curve. This level extracts various qualities from the feedback pictures. Utilization filter of a specific size (P*P), convolution is performed between the input image and the filter. Slide the filter from across the image (P*P) to calculate the e mark produced in among the filter and the image section. Additional levels can then use this information to learn more about what they see in the image's corners and edges. After that, it goes through a level of pooling. By lowering the connections between levels as well as running autonomously on every element plan, this level achieves its primary goal of reducing the size to a convolved feature plan. Pooling can be accomplished in a variety of ways to get different results. The largest feature on the feature map is selected by max pooling. Image sections are averaged together to get a sense of how many items are included in each area. A fully connected level is then passed over. There is one level of fully connected neurons. It has neurons, weights and biases. The level receives the flattened input from previous layers. The categorization process begins at this stage.

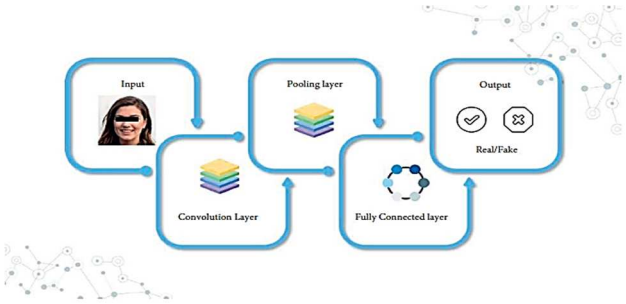

Fig 1: Employment Process Chart.

TABLE I: CATEGORIZATION OF PROCESS WITH INTERPRETATION OF EACH PROCESS

Categorization of Process	**Interpretation of Each Process**
Data.	Acquired data as of Kaggle contained 70 sides as of the Flickr data set that was compiled as well as maintained by Nvidia. Another 70,000 of one 000 fake faces produced by style GAN were fake. Images were downsized to 256 pixels after being combined from both datasets. We also created a training, validation, as well as test set for our dataset. In the training dataset, there were 100,000 images, half of which were real as well as the other 100,000 were fictitious. There were 20,000 pictures in the validation set, of which 10,000 were genuine as well as the rest were fictitious. After that, a test set of 20,000 images was created and divided equally between the real and fake ones. Deepfake assisted is a complex process that considers several variables. Identification of a proper classification scheme, training sample collection, image pre-processing, and feature extraction is all essential steps in the imaging classification process. Generic versions that understand exactly how as well as where to disseminate their information, not including supervision, are a key component of the deep fake framework. In this study, the Kaggle dataset "140k Real and Fake Faces" contains 70 fake faces created via style GAN. To conduct this relative investigation of the work to C.N.N nets of categorizing true and deep fake images, we have trained 8 CNN types. A Dense-Net model (DenseNet121, DenseNet169 as well as DenseNet201), two VGG-Net models (VGG16 as well as VGG19), one ResNet50, one VGG-Face, as well as a custom C.N.N style were all trained. In the further units, we'll go over each model in order.
Proposed Net.	It is possible to build a convolutional neural network by layering many smaller units of neurons together. The weight values of the training phase are updated every epoch utilization method, such as backpropagation, which connects the neurons together and gives the edges between them weight. The first part of CNN is used to obtain characteristics, as well as the second part is used for classification. Pretrained networks like Dense-Net, which is available in Kera's API, were utilized. Dense-Net's design is shown in Figure 2. DenseNet201, DenseNet169, as well as DenseNet121 are some of the models that we've used to improve the calculation effects. It's FF-CNN that connects each level. To keep the F.F. nature, each level receives different inputs from each prior level as well as goes them up on all subsequent levels [42].
Dense Blocks.	One of the most basic components of a neural network is the convolutional layer. The

	complicated features of a given data are extracted utilizing a fixed size. There are many dense blocks in a Dense-Net convolutional network. There are 169 levels in 4 dense blocks in Dense-Net169, for example. In addition, there are three transition layers, a classification layer, as well as a convolutional level in each dense block. After a convolution of 112, a max-pooling of 56 is applied to the architecture. A place that accepts 1 3 224 images in B.G.R. order is the model input.
DenseNet121.	The Residual CNN (Res-Net) architecture has been greatly expanded by the Dense Convolutional Network (Dense-Net). Each level of Dense-Net is directly connected to the next level of the network, unlike Res-Net as well as other convolutional neural networks [42]. We remembered that Dense-Net121 in the Keres model is correct with a bit of a dense utilization level like the last layers. B.N. and 3*3 turnaround were two of the many levels that made up the model's closely packed blocks. Every dense block had a transition level with an average pooling of 2 by two and a concentration of 1 by one between each dense block. After the last dense block, we inserted a dense custom level with sigmoid activation.
DenseNet201	The Dense-Net201 utilizes a compressed net to enable ease of train and parameter efficient models expected to the capacity to recycle features via consecutive levels. As a result, The next layer's input becomes more diverse, which improves performance.
DenseNet169.	One of the most efficient models for dealing with the problem of gradients disappearing is 169 levels deep, with just three parameters. Additionally, Res-Net50 has been used at this job to monitor the measurement assessment system. Res-Net50's design is depicted in Figure 3. For more complex issues, a neural network called Residual Network (short for Residual Network) employs multiple deep neural network levels to achieve greater accuracy and performance. In the belief that the more levels that are added, the more complex the characteristics of these levels will become.
ResNet50.	Forty-eight convolution levels, one maximum pool layer, as well as one typical pool level comprises ResNet50. Each of the floating-point processes in it totals 3.8*109.
VGG16.	The emphasis was on 3*3 filter convolution levels along with a pace of one rather than many hyperparameters. These levels constantly utilized similar padding and max pool-level like the 2*2 filter pace tow. VGG16's architecture is depicted in Figure 4. Convolution and max pool levels have been arranged in a similar fashion throughout the design. A SoftMax is used for output after two fully connected levels [71]. Thee 16 in VGG16 refers to the fact that it includes sixteen levels of various weights.
VGG-19.	Advanced CNN style outshines only one convolutional level in terms of performance. VGG19's layout is shown in Figure 4. For example, Max-pooling downsampling and a change to the Re-LU activation function are all made possible by the layer, which chooses the greatest values into a mental image as the mutual worth for an area. When reducing the number of variables while still preserving the primary characteristics of the sample, a down-sampling level is most often used.
V.G.G. Face.	Utilization of the standard datasets for face recognition developed by the Oxford Visual Geometry Group researchers is a mental image identification type that generates the most improved results. As a result of this method, we can build a large training dataset while utilizing only a tiny portion of footnote energy. Fig. 6 shows the VGG-Face structure. We have used the Tai Do as well as Kim [73]. VGG-Face architecture to build the model. Convolutional and max-pooling levels were included in all five lines of the layer. The first and 2nd blocks each had two 3x3 convolution levels accompanied by a pooling level. Re 3x3 convolution, levels are observed via max-pooling level, each composed of the Blocks three, four, and five. In all convolutional levels, we used the Re-LU activation function. Pretrained weights are a requirement for VGG-Face, so our approach has been modified. After all, we added additional dense levels to our five-level blocks, which tried to give us the facial characteristics we desired. Lastly, the net production along with activated sigmoidal remained added to the dense level set. A pattern has as well have been introduced to this study showing this same total difference, as illustrated in Fig 7.

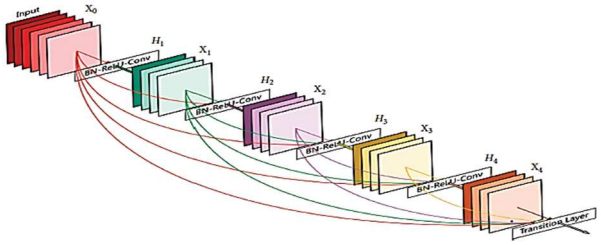

Fig 2: Design of Dense-Net [69].

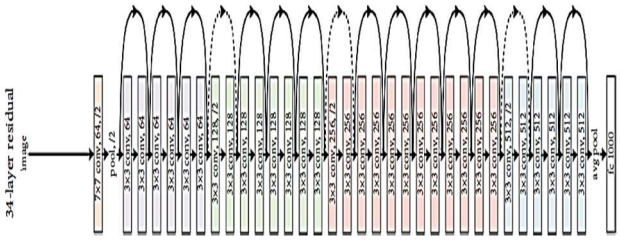

Fig 3: Design of Res-Net50 [70].

Classifier-associated confusion matrixes of the following sizes are available: n*n "no lines as well as rows), wherever n is the amount to classes. Real Optimistic (TO), True Pessimistic (T.P.), untrue Pessimistic (F.P.), as well as Untrue Optimistic (F.O.), have been calculated to utilize the subsequent calculations for n*n matrices [75].

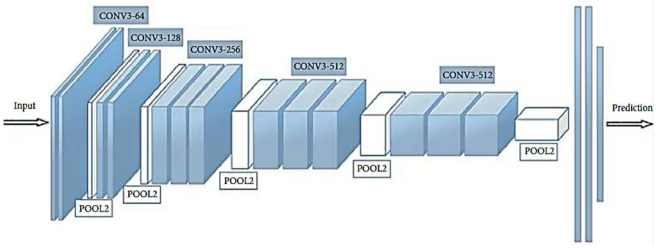

Fig 4: Design of VGG16 [72].

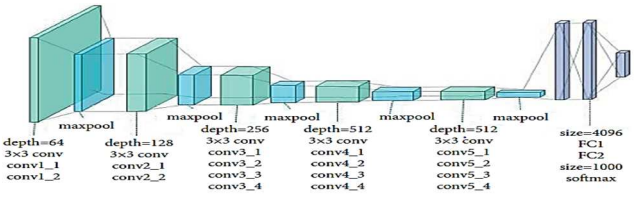

Fig 5: Design of VGG19 [73].

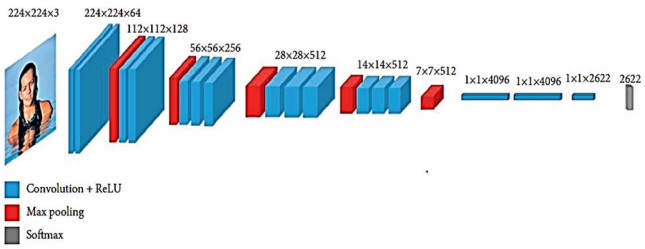

Fig 6: Design of VGG-face [74].

Projections can either remain accurate or incorrect in this situation. Figure 2 shows the DenseNet121 confusion matrix. 8. There are 9480 fictitious images in the confusion matrix. The network was able to classify 9926 real-world images correctly. However, 520 of the mythological images were determined to be genuine, while the remaining 74. Real photos were mistakenly labeled as fakes. DenseNet201's confusion matrix has been published.

Participating in conditions of defining false photos, that is the 9503. Although there remains non big variation, the standard differs. They misclassified 138 actual pictures as fake and 497 fake ones, making it impossible to identify real ones like the real thing.

TABLE II: ACHIEVED OUTCOMES WHEN IMPLEMENTING THE SIMULATIONS.

C.N.N design	Accuracy	Precision	Re -call	*F*1 score	**A.U.C**
Dense-Net169	0. 930	0. 970	0. 900	0. 930	0. 9760
Dense-Net201	0. 940	0. 940	0. 950	0. 930	0. 9740
Dense-Net121	0. 950	0. 970	0. 680	0. 800	0. 9510
VGG1	0. 920	0. 890	0. 950	0. 920	0. 9670
VGG-16	0. 900	0. 910	0. 900	0. 900	0. 9570
VGG-Face	0. 970	0. 970	0. 960	0. 970	0. 9780
Res-Net50	0. 950	0. 970	0. 920	0. 950	0. 9800

TABLE III: COMPARE IN DIFFERENCE GRAPH.

Reference work	**Design**	**Precision (percent)**	**Accuracy at this work (percent)**
[42]	Dense-Net169	93.1%	93%
[42]	Dense-Net201	83.6 %	94%
[42]	Dense-Net121	92.2%	95%
[80]	Donald Trump	90%	95.77%
[78]	VGG-19	80.2%	92%
[79]	VGG-16	81.60%	90%
[79]	Res-Net50	81.6%	95%

As mentioned in [80], work is created utilizing the rationale augmented convolutional neural network (CNN) on the Matlab R2019a platform by using the Kaggle DeepFake Video dataset with an accuracy of 95.77%. Hence, the task of real-time Deepfake facial reconstruction for security purposes is a difficult task to complete concerning limited hardware and efficiency. We have applied the "Donald Trump" filter to the women's video, and her face has been faked with it. This research paper investigates rational augmented CNN state-of-the-art technology that can potentially be utilized for Deepfake facial reconstruction via hardware such as webcams and security cameras in real-time.

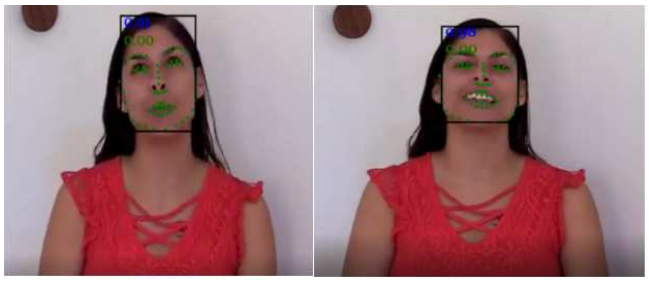

Fig 7: The input video sample images that are inserted to the Deepfake network for evaluation [80].

V. CONCLUSION and FUTURE WORK

Many people have been duped by Deepfake, a new method of deception. Since some Deepfake content is harmful to the world, it must be detected even if it is not malicious. Finding an effective way to identify Deepfake images was the primary objective of this research. Scholars have already been working very hard to identify Deepfake material utilization various detection techniques. The importance of this work lies in the effective use of CNN architecture. Deep fakes were detected utilizing eight CNN architectures on a large dataset. The findings are accurate as well as trustworthy. When it came to precision, accuracy, F1-score, and area under the curve, VGGFace came out on top. The study's specially made model surpassed the VGGFace in terms of recall. Results from models DenseNet169 and DenseNet201 were less impressive than those from the standard models. Means this determine and detection all images until that image does not have deepfake. In the long term, it will have a profound effect on the lives of many. Fake images can be quickly identified by utilizing this new technology. As a result of our efforts, people will continue to look for Deepfake images.

In the future, a video Deepfake set of data may be generated utilizing the CNN set of rules. Some other tests, as well as experiments, would have to wait. Classify Deepfake images utilization (CNN). Detection of Deepfake images can be improved by utilizing more efficient models, which will stop violence in our society and around the world. We aim to lessen the number of unintentional suicides as well as blackmail cases in our community.

REFERENCES

[1] I. J. Goodfellow, J. P. Abadie, M. Mirza et al., "Generative adversarial nets, "NIPS" 14," Proceedings of the 27th Inter- national Conference on Neural Information Processing Sys- tems, vol. 2, pp. 2672–2680, 2014.

[2] T. Nguyen, Q. Nguyen, C. M. Nguyen, D. Nguyen, D. Nguyen, as well as S. Nahavandi, "DL for Deep fakes creation as well as detection: a survey," pp. 1–17, 2019, https://arxiv.org/abs/ 1909.11573.

[3] T. Jung, S. Kim, as well as K. Kim, "DeepVision: Deep fakes detection utilization human eye blinking pattern," IEEE Access, vol. 8, pp. 83144–83154, 2020.

[4] M. Westerlund, "The emergence of Deepfake technology: a review," Technology Innovation Management Review, vol. 9, no. 11, pp. 39–52, 2019.

[5] M.-H. Maras as well as A. Alexandrou, "Determining authenticity of video evidence in the age of artificial intelligence as well as in the wake of Deepfake videos," International Journal of Evidence as well as Proof, vol. 23, no. 3, pp. 255–262, 2019.

[6] A. M. Almars, "Deep fakes detection techniques utilization deep learning: a survey," Journal of Computer as well as Communica- tions, vol. 9, no. 5, pp. 20–35, 2021.

[7] L. Guarnera, O. Giudice, as well as S. Battiato, "DeepFake detection by analyzing convolutional traces," in Proceedings of the 2020 IEEE/CVF Conference on Computer Vision as well as Pattern Recognition Workshops (CVPRW), pp. 2841–2850, Seattle, WA, U.S.A., 2020.

[8] A. Punnappurath as well as M. S. Brown, "Learning raw image reconstruction-aware deep image compressors," IEEE Transactions on Pattern Analysis as well as Machine Intelligence, vol. 42, no. 4, pp. 1013–1019, 2020.

[9] Z. Cheng, H. Sun, M. Takeuchi, as well as J. Katto, "Energy compaction-based image compression utilization convolutional AutoEncoder," IEEE Transactions on Multimedia, vol. 22, no. 4, pp. 860–873, 2020.

[10] J. Chorowski, R. J. Weiss, S. Bengio, as well as A. van den Oord, "Unsupervised speech representation learning utilization WaveNet autoencoders," IEEE/ACM Transactions on Audio, Speech, as well as Language Processing, vol. 27, no. 12, pp. 2041–2053, 2019.

[11] Faceswap, "Deep fakes software for all," https://github.com/ Deepfakes/faceswap.

[12] FakeApp 2.2.0, https://www.malavida.com/en/soft/fakeapp/.

[13] DeepFaketf, "Deepfake based on tensorflow," https://github.com/StromWine/DeepFake%20tf.

[14] DFaker, https://github.com/dfaker/df.

[15] DeepFaceLab, https://github.com/iperov/DeepFaceLab.

[16] Faceswap-GAN, https://github.com/shaoanlu/faceswap-GAN.

[17] Keras-VGGFace, "VGGFace implementation with Keras framework," https://github.com/rcmalli/keras-vggface.

[18] FaceNet, https://github.com/davidsandberg/facenet.

[19] CycleGAN, https://github.com/junyanz/pytorch-CycleGAN- and-pix2pix.

[20] K. Danielle Citron as well as R. Chesney, "Deep fakes: a looming challenge for privacy, democracy, as well as national security, 107 California law review," p. 1753, 2019, https://scholarship.law. bu.edu/faculty_scholarship/640.

[21] O. De Lima, S. Franklin, S. Basu, B. Karwoski, as well as A. George, "Deepfake detection utilization spatiotemporal convolutional networks," 2020, https://arxiv.org/abs/2006.14749.

[22] I. Amerini as well as R. Caldelli, "Exploiting pre- diction error inconsistencies through LSTM-based classifiers to detect Deepfake videos," in Proceedings of the 2020 A.C.M. Workshop on Information Hiding as well as Multimedia Security, pp. 97–102, Denver, CO, U.S.A., June 2020.

[23] P. Korshunov as well as S. Marcel, "Vulnerability assessment as well as detection of Deepfake videos," in Proceedings of the 12th IAPR International Conference on Biometrics (ICB), pp. 1–6, Crete, Greece, June 2019.

[24] VidTIMIT database, http://conradsanderson.id.au/vidtimit/.

[25] O. M. Parkhi, A. Vedaldi, as well as A. Zisserman, "Deep face recognition," in Proceedings of the British Machine Vision Con- ference (BMVC), pp. 41.1–41.12, Swansea, U.K., September 2015.

[26] F. Schroff, D. Kalenichenko, as well as J. Philbin, "Facenet: a unified embedding for face recognition as well as clustering," in Proceed- ings of the IEEE Conference on Computer Vision as well as Pattern Recognition, pp. 815–823, Boston, MA, U.S.A., June 2015.

[27] H. Zhang, I. Goodfellow, D. Metaxas, as well as A. Odena, "Self-attention generative adversarial networks," 2018, https://arxiv.org/abs/1805.08318.

[28] A. Brock, J. Donahue, as well as K. Simonyan, "Large scale GAN training for high fidelity natural image synthesis," 2018, https://arxiv.org/abs/1809.11096.

[29] T. Miyato, T. Kataoka, M. Koyama, as well as Y. Yoshida, "Spectral normalization for generative adversarial networks," 2018, https://arxiv.org/abs/1802.05957.

[30] S. Agarwal as well as L. R. Varshney, "Limits of deep- fake de- tection: a robust estimation viewpoint," 2019, https://arxiv. org/abs/1905.03493.

[31] U. M. Maurer, "Authentication theory as well as hypothesis test- ing," IEEE Transactions on Information Theory, vol. 46, no. 4, pp. 1350–1356, 2000.

[32] J. S. Chung, A. Senior, O. Vinyals, as well as A. Zisserman, "Lip reading sentences in the wild," in Proceedings of the IEEE Conference on Computer Vision as well as Pattern Recognition (CVPR), pp. 3444–3453, Honolulu, HI, U.S.A., July 2017.

[33] S. Suwajanakorn, S. M. Seitz, as well as I. Kemelmacher-Shlizer- man, "Synthesizing Obama," A.C.M. Transactions on Graphics, vol. 36, no. 4, pp. 1–13, 2017.

[34] P. Korshunov as well as S. Marcel, "Speaker inconsistency detection in tampered video," in Proceedings of the 26th European Signal Processing Conference (EUSIPCO), pp. 2375–2379, Rome, Italy, September 2018.

[35] J. Galbally as well as S. Marcel, "Face anti-spoofing based on general image quality assessment," in Proceedings of the 22nd Inter- national Conference on Pattern Recognition, pp. 1173–1178, Stockholm, Sweden, August 2014.

[36] Y. Zhang, L. Zheng, as well as V. L. Thing, "Automated face swapping as well as its detection," in Proceedings of the IEEE 2nd International Conference on Signal as well as Image Processing (ICSIP), Singapore, August 2017.

[37] X. Wang, N. Thome, as well as M. Cord, "Gaze latent support vector machine for image classification improved by weakly su- pervised region selection," Pattern Recognition, vol. 72, pp. 59–71, 2017.

[38] S. Bai, "Growing random forest on deep convolutional neural networks for scene categorization," Expert Systems with Ap- plications, vol. 71, pp. 279–287, 2017.

[39] L. Zheng, S. Duffner, K. Idrissi, C. Garcia, as well as A. Baskurt, "Siamese multi-levelperceptrons for dimensionality reduc- tion as well as face identification," Multimedia Tools as well as Appli- cations, vol. 75, no. 9, pp. 5055–5073, 2016.

[40] C.-C. Hsu, Y.-X. Zhuang, as well as C.-Y. Lee, "Deep fake image detection based on pairwise learning," Applied Sciences, vol. 10, no. 1, p. 370, 2020.

[41] S. Chopra, "Learning a similarity metric discriminatively, with application to face verification," in Proceedings of the IEEE Conference on Computer Vision as well as Pattern Recognition, pp. 539–546, San Diego, CA, U.S.A., September 2005.

[42] G. Huang, Z. Liu, L. Van Der Maaten, as well as K. Q. Weinberger, "Densely connected convolutional networks," in Proceedings of the IEEE Conference on Computer Vision as well as Pattern Recognition, pp. 4700–4708, Honolulu, HI, U.S.A., July 2017.

[43] K. Cho, B. van Merrienboer, C. Gulcehre et al., "Learning phrase representations utilization R.N.N. encoder–decoder for statistical machine translation," in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Pro- Cessing (EMNLP), pp. 1724–1734, Doha, Qatar, October 2014.

[44] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and

[45] M. Nießner, "Faceforensics++: learning to detect manipulated facial images," in Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1–11, Seoul, Republic of Korea, 2019.

[46] D. Guera as well as E. J. Delp, "Deepfake video detection utilization recurrent neural networks," in Proceedings of the 2018 15th IEEE International Conference on Advanced Video as well as Signal Based Surveillance (AVSS), pp. 1–6, Auckland, New Zealand, November 2018.

[47] Y. Li, M. C. Chang, as well as S. Lyu, "Ictu oculi: exposing A.I. created fake videos by detecting eye blinking," in Proceedings of the 2018 IEEE International Workshop on Information Forensics as well as Security (WIFS), pp. 1–7, Hong Kong, China, December 2018.

[48] J. Donahue, L. Anne Hendricks, S. Guadarrama et al., "Long- term recurrent convolutional networks for visual recognition as well as description," in Proceedings of the IEEE Conference on Computer Vision as well as Pattern Recognition, pp. 2625–2634, Boston, MA, U.S.A., June 2015.

[49] H. H. Nguyen, J. Yamagishi, as well as I. Echizen, "Capsule-forensics: utilization capsule networks to detect forged images as well as videos," in Proceedings of the 2019 IEEE International Con- ference on Acoustics, Speech as well as Signal Processing (ICASSP), pp. 2307–2311, Brighton, UK, May 2019.

[50] G. E. Hinton, A. Krizhevsky, as well as S. D. Wang, "Transforming auto-encoders," in Proceedings of the International Conference on Artificial Neural Networks, pp. 44–51, Espoo, Finland, June 2011.

[51] S. Sabour, N. Frosst, as well as G. E. Hinton, "Dynamic routing between capsules," in Advances in Neural Information Pro- cessing Systems, pp. 3856–3866, M.I.T. Press, Cambridge, MA, U.S.A., 2017.

[52] I. Chingovska, A. Anjos, as well as S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in Proceedings of the International Conference of Biometrics Special Interest Group (BIOSIG), pp. 1–7, Darmstadt, Germany, September 2012.

[53] D. Afchar, V. Nozick, J. Yamagishi, as well as I. Echizen, "MesoNet: a compact facial video forgery detection network," in Pro- ceedings of the 2018 IEEE International Workshop on Infor- mation Forensics as well as Security (WIFS), pp. 1–7, Darmstadt, Germany, December 2018.

[54] J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt, and M. Nießner, "Face2Face: real-time face capture as well as reen- actment of RGB videos," in Proceedings of the IEEE Conference on Computer Vision as well as Pattern Recognition, pp. 2387–2395, Las Vegas, NV, U.S.A., June 2016.

[55] N. Rahmouni, V. Nozick, J. Yamagishi, as well as I. Echizen, "Distinguishing computer graphics from natural images utilization convolution neural networks," in Proceedings of the 2017 IEEE Workshop on Information Forensics as well as Security (WIFS), pp. 1–6, Rennes, France, December 2017.

[56] M. Koopman, A. M. Rodriguez, as well as Z. Geradts, "Detection of Deepfake video manipulation," in Proceedings of the 20th Irish Machine Vision as well as Image Processing Conference (IMVIP), pp. 133–136, Belfast, Ireland, August 2018.

[57] K. Rosenfeld as well as H. T. Sencar, "A study of the robustness of PRNU-based camera identification," Media Forensics as well as Security International Society for Optics as well as Photonics, vol. 7254, Article ID 72540M, 2009.

[58] C. T. Li as well as Y. Li, "Color-decoupled photo response non-uniformity for digital image forensics," IEEE Transactions on Circuits as well as Systems for Video Technology, vol. 22, no. 2, pp. 260–271, 2012.

[59] X. Lin as well as C. T. Li, "Large-scale image clustering based on camera fingerprints," IEEE Transactions on InformaTion Forensics as well as security, vol. 12, no. 4, pp. 793–808, 2017.

[60] U. Scherhag, L. Debiasi, C. Rathgeb, C. Busch, as well as A. Uhl, "Detection of face morphing attacks based on PRNU anal- ysis," IEEE Transactions on Biometrics, Behavior, as well as Identity Science, vol. 1, no. 4, pp. 302–317, 2019.

[61] Q.-T. Phan, G. Boato, as well as F. G. B. De Natale, "Accurate as well as scalable image clustering based on sparse representation of camera fingerprint," IEEE Transactions on Information Fo- rensics as well as security, vol. 14, no. 7, pp. 1902–1916, 2019.

[62] H. T. Sencar as well as N. Memon, Digital Image Forensics, Springer, New York, NY, U.S.A., 2013.

[63] H. Farid, Photo Forensics, M.I.T. Press Ltd., Cambridge, MA, U.S.A., 2016.

[64] D. Gu¨era, Y. Wang, L. Bondi, P. Bestagini, S. Tubaro, as well as E. J. Delp, "A counter forensic method for CNN-based camera model identification," in Proceedings of the IEEE Conference on Computer Vision as well as Pattern Recognition Workshops, pp. 1840–1847, Honolulu, HI, U.S.A., July 2017.

[65] D. Gu¨era, S. K. Yarlagadda, P. Bestagini, F. Zhu, S. Tubaro, as well as E. J. Delp, "Reliability map estimation for cnn-based camera model attribution," in Proceedings of the IEEE Winter Conference on Applications of Computer Vision, Lake Tahoe, NV, U.S.A., March 2018.

[66] R. Raghavendra, K. B. Raja, S. Venkatesh, as well as C. Busch, "Transferable deep-cnn features for detecting digital as well as print-scanned morphed face images," in Proceedings of the IEEE Conference on Computer Vision as well as Pattern Recog- nition Workshops, pp. 1822–1830, Honolulu, HI, U.S.A., July 2017.

[67] P. Zhou, "Two-stream neural networks for tampered face detection," in Proceedings of the IEEE Conference on Computer Vision as well as Pattern Recognition Workshops, pp. 1831–1839, Honolulu, HI, U.S.A., July 2017.

[68] A. Ro¨ssler, "Faceforensics: a large-scale video dataset for forgery detection in human faces," 2018, https://arxiv.org/abs/ 1803.09179.

[69] 140K Real as well as Fake Faces, https://www.kaggle.com/xhlulu/ 140k-real-and-fake-faces.

[70] https://www.kaggle.com/keras/resnet50.

[71] https://towardsdatascience.com/step-by-step-vgg16-impleme ntation-in-keras-for-beginners-a833c686ae6c.

[72] https://www.kaggle.com/shivamb/cnn-architectures-vgg-resnet-inception-tl.

[73] I. N. Tai Do Nhu as well as S. H. Kim, "Forensics face detection from GANs utilization convolu ional neural network," pp. 1–8, 2018, https://arxiv.org/abs/1902.11153v2.

[74] https://sefiks.com/2018/08/06/deep-face-recognition-with- keras/.

[75] M. S. Junayed, "AcneNet-a deep CNN based classification approach for acne classes," in Proceedings of the 12th Inter- national Conference on Information & Communication Technology as well as System (ICTS), pp. 203–208, Surabaya, Indonesia, 2019.

[76] P. Accuracy, “Recall or F1,” https://towardsdatascience.com/ accuracy-precision-recall-or-f1-331fb37c5cb9.

[77] Classification, "R.O.C. curve as well as A.U.C.," https://developers. google.com/machine-learning/crash-course/classification/ roc-and-auc.

[78] D. Gong, Y. Jaya Kumar, O. S. Goh, Zi Ye, as well as W. Chi, "DeepfakeNet, an efficient Deepfake detection method," In- ternational Journal of Advanced Computer Science as well as Ap- plications (IJACSA), vol. 12, no. 6, 2021.

[79] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "Deep fakes as well as beyond: a Survey of face manipulation as well as fake detection," Information Fusion, vol. 64, pp. 131–148, 2020.

[80] S. R. A. Ahmed and E. Sonuç, "Deepfake detection using rationale-augmented convolutional neural network," Applied Nanoscience, Sep. 2021.