iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VI **Month of publication:** June 2022

DOI: https://doi.org/10.22214/ijraset.2022.44868

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

A Review on Deepfake Attack Detection of User

Prachi Chhajed[1], Dhanshree Phalke[2]
[1, 2]*Department of Computer Engineering, Y Patil College of Engineering, Akurdi Pune, Savitribai Phule Pune University (SPPU), India*

Abstract: ***The Deepfake attacks are common these days. It is hard to detect the faces of users are real or fake. Deepfake is mainly used in different fields as having positive points regarding virtual presence of any user but this also has drawback which leads to misuse of the replacing the user's facial data with someone else facial data. These changing of user's original identity to fake identity is done by the deepfake attackers who replace its own identity or human imposters to hide the originality of the user. These attacked images or videos need to be detected. There are various face detectors among which few detect the real or fake deepfake images or videos. There is procedure also which used for authentication purpose such as visual speaker authentication. Also, convolutional neural network is used for facial feature extraction purpose and recurrent neural network is used for detecting the video is manipulated or not.***

Keywords: CNN, RNN, Deep Learning, Facial recognition, lip movement

I. INTRODUCTION

Deep Learning is the successive representation of layers in which the number of layers contribute to the model of data is the depth of the model. This approach is also known as hierarchical representation learning or layered representation learning. This name is suggested because of its number of layers into the model. As compared to machine learning, deep learning is easier to solve the problem because it's automating. This was difficult task in machine learning.

Deep learning is used in many fields where one of technology is used for creating deepfake images and videos. Deepfake is the technology consisting of combination of deep learning and fake identity. [18] Deepfake is used for creating fake identity of videos. It is used in education sector, as delivering innovative lessons; in arts field, it is used for creation of scenarios like any artist who is no more living can be present by virtual mode using deepfake. Mainly it can be used for creating the virtual appearance of the any person or thing. Likewise, it also has some drawbacks like for blackmailing purpose, fraud, politics like a politician's speech can be replaced with false information which he/she has never spoken just to spread false awareness among the people. This leads to failure of trust among people which the politician has made. These are nothing but deepfake attacks. The images and videos are manipulated mainly for nefarious purpose. To detect such kind of deepfake attack, it is very important to detect these images or videos which are not real. There are different algorithms used for creating and detecting the deepfake generated images and videos. For generation of fake images and videos, generative adversarial network is used. GAN is used for swapping of images. This used for manipulating the videos and images which are never real. This is not known to the original user that his/her image or video are transformed to another which looks exactly like original one.

To prevent attacks, visual speaker authentication approach is used. It is used to authenticate the user information which is already stored prior to testing. These user data is trained by storing the user data into database. This data is stored prior as like for biometric[2] we store the data for a person prior and then while authentication any of the biometric features are used such as face, finger, iris, palm, and any behavioral features recognition are considered.[3][7] So here VSA is used to extract the lip movement of user by considering each letter that the user has spoken. Lets consider, if user is speaking some vowels such as 'a,e,i,o,u', the system will store each letter spoken by user into database. The data spoken by user are accepted in the form of frames. Each letter spoken by user is stored accordingly. It is retrieved in the form of text format.[1] The VSA is used while showing the data stored by the user or registered by the user matches with the input video/image or not. It will authenticate on bases of the data stored. This is done by extracting the facial features considering lip movement of the user. Lip movements are stored as data of user for capturing the unique features of the user which discriminates from other users observing the video or image are fake or real.

The VSA approach can also be based on joint spatiotemporal sparse coding and hierarchical pooling. Here, the lip sequence is divided into series of subsequences with the temporal dimension [4]. With this each subsequence is measured as a spatiotemporal cube. This cube is later partitioned into multiple densely overlying 3D spatiotemporal cells. Sparse coding is used to describe these cell contents, which bring about in a 3D SC matrix which is generated for each subsequence. To get the final feature for the sequence, Max-pooling with a preset hierarchical structure is performed on the 3D SC matrix, and a set of features which corresponds to the sequences having the whole lip sequence is representation.[4]

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

II. RELATED WORK

Deep Learning is used in applications specially for image recognition and self-driving vehicles. Deep learning involves highly sophisticated models that can find patterns of huge data.[20] Deep Learning has different models such as Multilayer Perceptron's (MLPs), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTMs) and Generative Adversarial Networks (GANs), are developed to train neural network which makes classification and prediction easy.

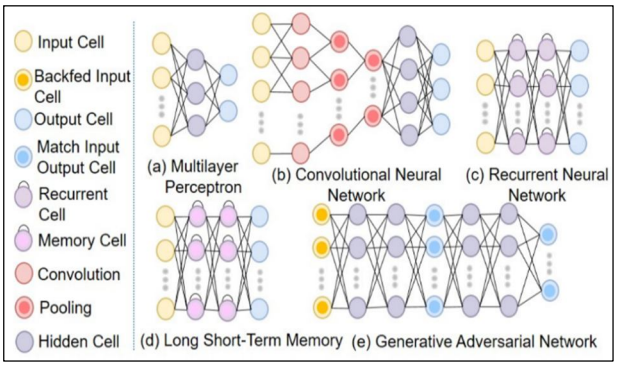


Fig. 1 Basic Structure of Deep Learning [20]

The Fig 1. shows the basic structure of Deep Learning. The Fig 1 (a) shows the Multilayer Perceptron (MLP), consisting of input layer, hidden layer and output layer. The input cell is feed forward to hidden cell which are then activated following to further cells, later transmitted to output cells. Error detected in between other layers are then adjusted by internal weights between each layer. MLP is accurate in terms of healthcare applications.

The Fig. 1 (b) shows the Convolutional Neural Network (CNN), [1][18][20] which is used to find something from the messy data. It is the correlation of high level extracted features and adjacent data blocks with the use of convolutional and pooling operations. It extracts features with minimum parameters and computation time [15][16].

The Fig. 1 (c) shows the Recurrent Neural Network (RNN), [1][20] which tracks the data such as memory cells. Each cell consists of information from upper layer and information from other channels. The neuron is furnished with feedback loop which provides the output and acts the input for next steps. [15] RNN are used to predict the information and restore the missing data.

The Fig 1 (d) shows Long Short-Term Memory (LSTM), [20] consists of extension of RNN. A memory cell is well defined and it is actively in control unit state.[15] LSTM outperforms RNN when dealing with the features of data for a long dependency over time.

The Fig 1 (e) shows Generative Adversarial Network (GAN), [20][12] which contains a generator and discriminator where generator is used to create the new image from the input layer and discriminator decides whether the data received is real input data or generated from generator. Here discriminator has multiple hidden layers, but it has 2 areas for input and output, 1 for real image and 0 for fake. Fake images are returned to the algorithm through backpropagation. This continues till new image is not created.[15]

According to [15], LSTM and CNN are used for detecting the deepfake videos or images. CNN is used for frame feature extraction and LSTM for temporal sequence analysis. An input frames is provided in sequence to the system where set of features are obtained with the help of CNN. The multiple consecutive frames of multiple features are then concatenated which are later pass-through LSTM for analysis. This analysis shows whether the video or image is manipulated or not.

It is necessary to identify the user information by its unique features to prevent the deepfake attacks. So [10], shows the speaker identification by person's lip-reading habits. Here, the visual features of person are extracted with the help of Hidden Markov Model (HMM) and mixtures of gaussian. It extracts the features of user from an image sequence of talking habit which consist of shape parameters having describing of lip boundary and intensity parameters. It converts images to grey level images of mouth area. Extracted features are both dependent speech as well as speaker information. It finds the probabilistic characteristics of the features extracted. It is used for user speaker classification method for handcrafted features.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

Visual speaker authentication is an approach used for authenticating the deepfake attacks. [1][9] This system works by storing the user's data prior into the database while doing the registration of the user same as the procedure in biometric devices used in government and private sectors. In prior, the data is registered and stored, later while accessing data this authentication verifies whether the person is real or fake to access the data. Likewise, the VSA approach is used for authentication purpose. The main work is, an approach is introduced for having four folds, first, it extracts the information which has talking habits of user; it consists of high-level lip feature which has high discriminative power to protect it from human imposters.[8]

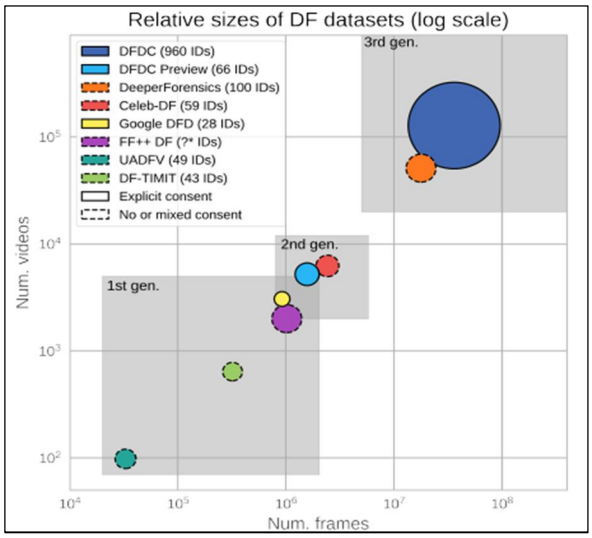


Fig 2: Comparison of current Deepfake datasets [17]

As per [17], the Fig 2, shows the comparison of current datasets used in [17]. Both the axis shows the log scale of deepfake detection challenge[11] is over an order of magnitude bigger than any other dataset which is represented in the form of x and y axis having number of frames on x axis and number of videos on y axis. Rough outlines of dataset generation are shown where circles size represents visualization of number of fake identities in that particular dataset.[5]

The first and second generation videos do not generalize to real deepfake videos and also do not contain enough identities which shows the sufficient detection generalization. Whereas, in third generation most recent deepfake datasets, DeeperForensics-1.0 and the DFDC Dataset, contain tens of thousands of videos and tens of millions of frames. These datasets are trained and tested to get the results in efficient manner.

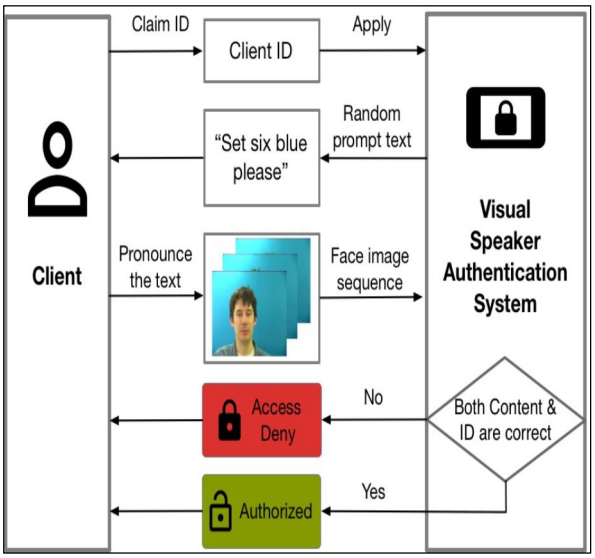


Fig 3. A visual speaker authentication system under the random password scenario.[1]

The Fig 3. shows the Visual Speaker Authentication (VSA) flowchart sharing the random password to the user while the id of user is claimed to the VSA system. This authentication system [6][19] will randomly create the password as a prompt text and the system will only accept if the text provided by system is valid and correctly pronounced by the user. With this prompt text the user's liveliness can be ensured as the attackers or the imposters cannot attack the original content of the user when the user's data is prerecorded. The deepfake attacks are increasing behalf of having VSA approach. The imposters can swap the face [7][14] on the spot and speak the random text without knowing the system that they are the imposters. For this, new deep learning based VSA is used to capture the unique talking habit of the users.

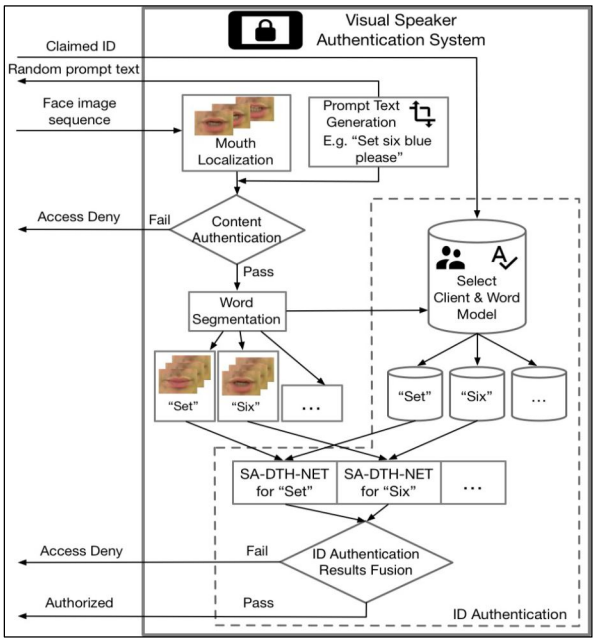


Fig 4. The overall architecture of authentication system [1]

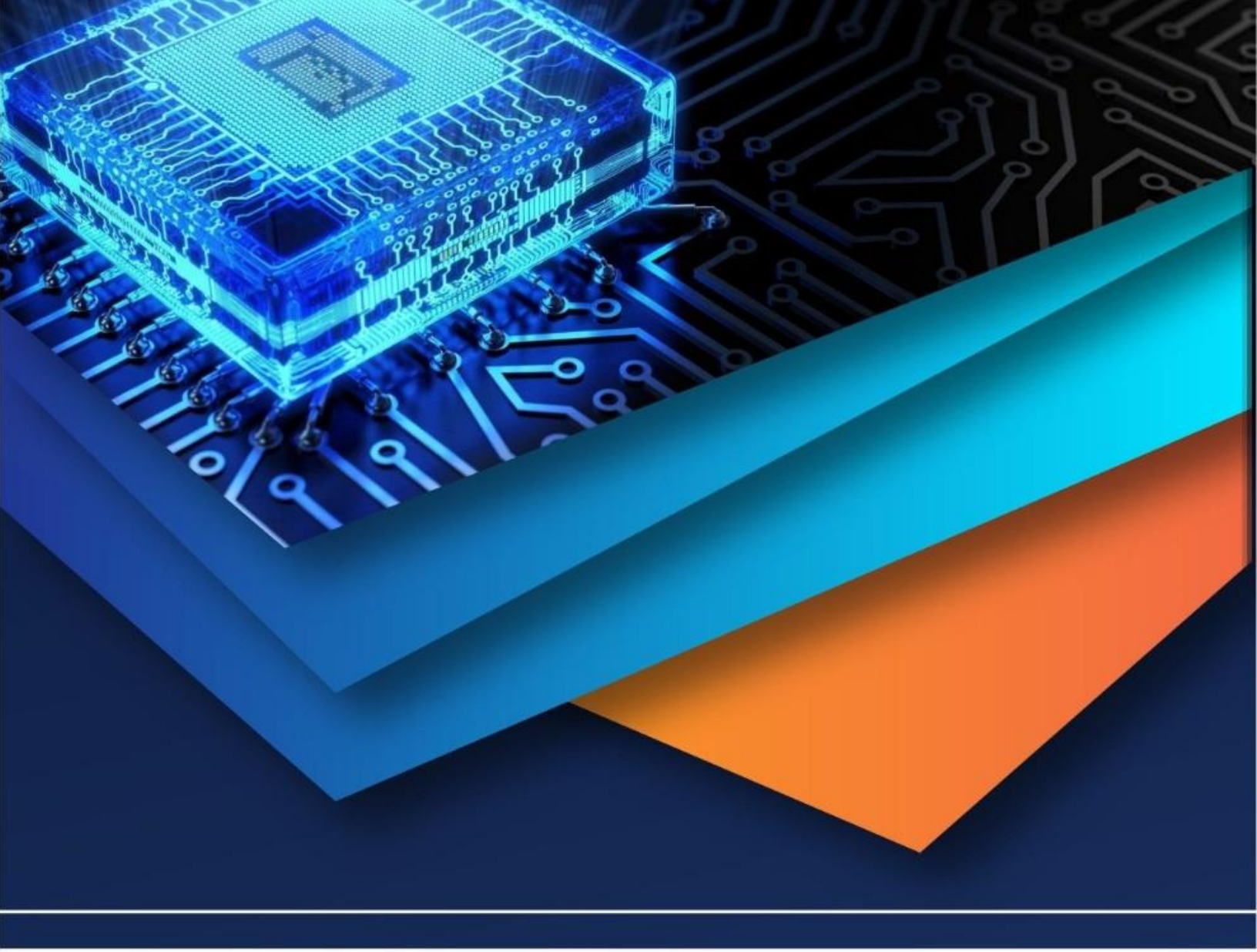
The Fig 4. shows the overall architecture of authentication system proposed by [1]. The information of user is stored and words are isolated based on Connectionist Temporal Classification (CTC) [13] output. The user authentication is performed at the word level in order to reduce the word variations resulting by the pronouncing various kind of contents. After word segmentation, Speaker Authentication network based on Dynamic Talking Habit (SA-DTH) was introduced which tests whether the lip sequences relate to the contents of the users' information about pronouncing the specific words. After word level, the lip sequence of user is analyzed giving authentication results. The end result shows how much number of the data i.e., word segments match with the data of the original user. The SA-DTH [1] is the main component of authentication system that can fight the deepfake attacks. In SA-DTH network structure, it consists of two parts, Fundamental lip Feature Extraction (FFE), which aims to focus on lip motion characteristics, it influence the static and dynamic lip shape considering its appearance, texture, size and pattern; second is Representative lip feature extraction and Classification subnet (RC-Net), which aims to extract lip features having high-level representation where it is looks over its authentication of the user. It not only authenticates the user's lip features but also looks over the representation on the user's talking habit.

III. CONCLUSION

The survey on deep learning and deepfake attacks shows the methods that are involved to check deepfake attacks. The brief description about the deep learning introduces the models which can help in different ways for detecting the facial features of the users. Also, it gives an overview of the extraction of the features which relates to deepfake attacks. Some of the procedures and methods are described to detect the deepfake attacks. The prevention measures and approaches which will help to find out the real and fake manipulated images and videos. This is a challenging task to detect the images are real or fake as more the accuracy of models is proposed, the more imposters and attackers find way to match the originality of the video to create the fraud appearances. As per review, this can be managed only if the videos or images are compared with unique talking habits of user with sufficient amount of data of user, because a large database is required for capturing the user's identity which makes him/her different from other users, imposters, or attackers.

REFERENCES

[1] Chen-Zhao Yang, Jun Ma, Shilin Wang and Alan Wee-Chung Liew, "Preventing DeepFake Attacks on Speaker Authentication by Dynamic Lip Movement Analysis", IEEE Transactions on Information Forensics and Security (Volume: 16), Page(s): 1841 - 1854, Digital Object Identifier 10.1109/TIFS.2020.3045937
[2] J. Liu-Jimenez, R. Sanchez-Reillo, and C. Sanchez-Avila, "Biometric coprocessor for an authentication system using iris biometrics," in Proc. 38th Annu. Int. Carnahan Conf. Secur. Technol., Oct. 2004, pp. 131–135.
[3] Jingxiao Zheng , Rajeev Ranjan, Ching-Hui Chen, Jun-Cheng Chen, Carlos D. Castillo, and Rama Chellappa, "An Automatic System for Unconstrained Video-Based Face Recognition", Digital Object Identifier 10.1109/TBIOM.2020.2973504, IEEE Transactions on Biometrics, Behavior, and Identity Science (Volume: 2, Issue: 3, July 2020) Page(s): 194 – 209
[4] J.-Y. Lai, S.-L. Wang, A. W.-C. Liew, and X.-J. Shi, "Visual speaker identification and authentication by joint spatiotemporal sparse coding and hierarchical pooling," Inf. Sci., vol. 373, pp. 219–232, Dec. 2016.
[5] C. Whitelam et al., "IARPA janus benchmark-B face dataset," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), 2017, pp. 592–600.
[6] Deepfacelab. Accessed: May 3, 2020. [Online]. Available: https://github.com/iperov/DeepFaceLab
[7] Article: Robert Mungovan, Aware "Face recognition: fighting the fakes"
[8] Hanqing Zhao1 Wenbo Zhou1,y DongdongChen2 Tianyi Wei1 Weiming Zhang1,y Nenghai Yu1, "Multi-attentional Deepfake Detection", 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), DOI 10.1109/CVPR46437.2021.00222
[9] Azat Davletshin. https://github.com/NTech-Lab/deepfake-detection-challenge.
[10] J. Luettin, N. A. Thacker, and S. W. Beet, "Speaker identification by lipreading," in Proc. 4th Int. Conf. Spoken Lang. Processing. ICSLP, Oct. 1996, pp. 62–65
[11] P. Korshunov and S. Marcel, "Deepfake detection: humans vs machines", *arXiv* preprint arXiv:2009.03155, 2020.
[12] S. Singh, R. Sharma and A.F. Smeaton, "Using GANs to Synthesise Minimum Training Data for Deepfake Generation", arXiv preprint arXiv:2011.05421, 2020.
[13] A. Graves, S. Fernández, F. Gomez, and J. Schmidhuber, "Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks," in Proc. 23rd Int. Conf. Mach. Learn. ICML, 2006, pp. 369–376.
[14] Deepfakes github, 03 2020, [online] Available: https://github.com/deepfakes/faceswap/.
[15] D. Güera and E. J. Delp, "Deepfake video detection using recurrent neural networks", 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 1-6, 2018.
[16] W. Zhang, C. Zhao and Y. Li, "A Novel Counterfeit Feature Extraction Technique for Exposing Face-Swap Images Based on Deep Learning and Error Level Analysis", Entropy, vol. 22, no. 2, pp. 249, 2020.
[17] Dolhansky B. Dolhansky, J. Bitton, B. Pflaum, J. Lu, R. Howes, M. Wang, et al., "The DeepFake Detection Challenge Dataset", arXiv preprint arXiv:2006.07397, 2020.
[18] Deng Pan, Lixian Sun, Rui Wang, Xingjian Zhang, Richard O. Sinnott, "Deepfake Detection through Deep Learning", 2020 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT), DOI: 10.1109/BDCAT50828.2020.00001
[19] P. Korshunov and S. Marcel, "Vulnerability assessment and detection of deepfake videos", The 12th IAPR International Conference on Biometrics (ICB), pp. 1-6, 2019.
[20] Zhuoqing Chang, Shubo Liu, Xingxing Xiong, Zhaohui Cai, and Guoqing Tu, " A Survey of Recent Advances in Edge-Computing-Powered Artificial Intelligence of Things", Journal of IEEE Internet Of Things Class Files, Vol. 14, No. 8, August 2021, DOI : 10.1109/Jiot.2021.3088875, Page(S): 13849 - 13875



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY