A Comparative Study of Deepfake Video Detection Method

2020 3rd International Conference on Information and Communications Technology (ICOIACT) 978-1-7281-7356-6/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICOIACT50329.2020.9331963

Kurniawan Nur Ramadhani
School of Electrical Engineering and Informatics
Institut Teknologi Bandung,
School of Computing, Telkom University,
Bandung, Indonesia
kurniawannr@telkomuniversity.ac.id

Rinaldi Munir
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
rinaldi.munir@itb.ac.id

***Abstract*—Deepfake technology allows humans to manipulate images and videos using deep learning technology. The results from deepfakes are very difficult to distinguish using ordinary vision. Many algorithms are built to detect deepfake content in images and videos. There are several approaches in deepfake detection, including a visual feature-based approach, a local feature-based approach, a deep feature-based approach and a temporal feature-based approach. The main challenge in developing deepfake detection algorithms is the variety of existing deepfake models in both images and videos. Another challenge is that deepfake technology is still evolving, making deepfake images and videos look more realistic and harder to detect.**

***Keywords*—deepfake, Generative Adversarial Networks, autoencoder, deep learning**

I. Introduction

Deepfake is a technology used in producing certain videos that are manipulated using an artificial intelligence technique called deep learning[1]. Deepfake videos are generally videos that contain actions taken by certain people, but with other people's faces. Replacement of people's faces in videos using deep learning techniques.

Many people use deepfake technology for negative purposes. There are two deepfake cases that are widely circulating on social media. The first case is a deepfake case related to pornography and the second case is a deepfake case related to a black campaign on political contestation. In the first case, pornographic videos are processed with deepfake technology so that the faces of the actors in the videos are replaced with the faces of artists or other public figures, with the aim of bringing down the character's good name. Whereas in the second case, videos of certain figures making controversial statements but with the faces of the speakers being replaced by other political figures, usually those who are following a political contestation process, with the aim of bringing down the electability of that figure[2].

Research on deepfake video detection methods has been widely conducted since 2017. Zhou et al. [3] used two combined features of image convolution and steganalysis to detect deepfake content in video. Li et al. [4] in 2018 detected deepfake content in a video using a visual feature such as a wink. The same researchers in 2019 used another visual feature, namely head pose to detect deepfake content in images[5]. Akhtar and Dasgupta [6] analyzed several local feature extraction methods in detecting deepfake content in videos. Apart from using the feature extraction approach, several deep learning approaches are also used to detect deepfakes. using the Recurrent Neural Network (RNN) model in detecting deepfake video[7]. Amerini et al. [8] using a Convolutional Neural Network (CNN) with Optical Flow feature. Several other researchers have developed classification models such as DeepFD[9], MesoNet[10], Capsule Forensic[11], Multitask-learning[12] and other models based on the CNN architecture.

In this paper, we discuss about deepfake technology, deepfake detection methods and research opportunities that are still open in deepfake detection. With the development of information technology, the presence of technology to detect deepfake content in the future will be urgently needed.

II. Deepfake Technology

Deepfake is a process of manipulating images and videos to produce fake content that looks realistic to the eye. Before the emergence of deepfake technology, the usual image and video manipulation processes were image and video splicing. In image splicing, an image is modified by overwriting certain objects[13]. Overwritten objects can come from the same image or from a different image. The result of the overwriting process is an image with objects added or duplicated, shifted or removed[14]. As for video splicing, the manipulation process can be in the form of inserting frames or removing frames depending on the purpose of the splicing video. The process of image and video splicing is generally referred to as copy-move forgery. To detect image and video splicing, the usual technique is to look for duplicated parts of the image and video through a matching process using a feature extraction algorithm[15]–[19]. Another method that can be used to detect image and video splicing is to statistically analyze the content of the image and video to find areas that have anomalies[20]–[22].

Unlike image and video splicing, the process of forming deepfake content generally uses a neural network architecture known as Deep Autoencoder[23]. Deep autoencoder architecture is an artificial neural network architecture that has two parts, namely encoder and decoder. The encoder accepts input in the form of an image and transforms the image into a vector value in latent space, while the decoder works by carrying out the process of reconstructing the vector into the original image. The deep autoencoder is trained in such a way that the reconstructed image is as close as possible to the original image.

To generate deepfake content, two autoencoder architectures that share the same encoder are used. Suppose we want to replace face A with face B. So, we build two autoencoders A and B with the same encoder and different decoders, namely decoder A and decoder B. Each autoencoder is trained to reconstruct images A and B. Deepfake content is produced. by inputting image A into the

autoencoder B. So that the input of face image A will output face image B. An illustration of the deepfake generation process using the autoencoder can be seen in Fig. 1.

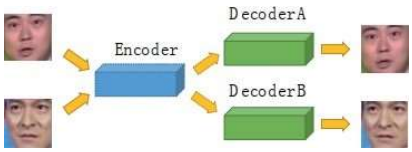

Fig. 1. Deepfake Generation Illustration Using Autoencoder [23]

Another method commonly used in producing deepfake content is Generative Adversarial Networks (GAN). GAN is a two-part neural network architecture called generator and discriminator. Generator is part of GAN which functions to produce fake content (fake) from a random vector. Discriminators are part of GAN which functions to detect whether content is original content or fake content produced by a generator[24]. GAN is trained to improve the performance of the generator so that the content produced is as natural as possible. Fig. 2 shows an illustration of the mechanism of the GAN.

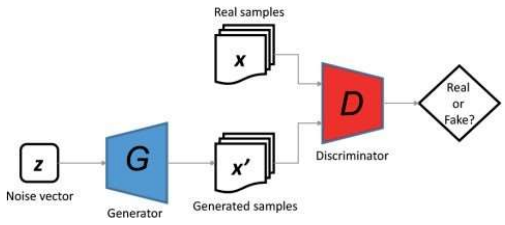

Fig. 2. Illustration of GAN[25]

Deep Autoencoder and GAN algorithms have been proven capable of producing natural looking deepfake content. However, these two methods have several loopholes that can be used in detecting the content of the resulting deepfake. This gap is mainly caused by the upsampling and affine transformation processes carried out in both methods so that the deepfake content can be found in several cases:

a. Resolution inconsistencies between the face image and other areas of the image.
b. The incompatibility of facial styles with other body parts.
c. Temporal discontinuity which can be detected in facial muscles movement[26], [27].

Fig. 3 shows an example of a deepfake image taken from a footage and a video clip with the artist's face replaced.

Fig. 3. Deepfake Example

III. Deepfake Detection Method

The deepfake detection method is used to detect whether an image or video is deepfake content or original content. The deepfake detection method can be thought of as a binary classification process labeled as deepfake or original class. Deepfake detection works by extracting features from the image or video that can be used to differentiate between deepfake content and original content. Based on the feature extraction approach used, deepfake detection methods can be grouped into four types as follows.

a. Visual feature-based deepfake detection. The visual feature-based deepfake detection method uses the features of the image that can be seen with the naked eye such as blinking, head position and also differences in the shapes of the facial organs.
b. Local feature-based deepfake detection. The local feature-based deepfake detection method uses a pixel-based feature extraction method that extracts the features of each pixel. The advantage of this method over visual feature-based detection is that local feature-based detection has higher reliability than visual feature-based detection.
c. Deep feature-based deepfake detection. Like the local feature-based approach, the deep feature-based deepfake detection method also performs the feature extraction process at the pixel level. The difference is, in deep features, the feature extraction process uses multiple layers, so it can extract more complex features than the local feature-based approach.
d. Temporal feature-based deepfake detection. Unlike the other three approaches, the temporal feature-based deepfake detection method extracts the features from several consecutive frames to get the temporal features of the video. This detection method can be used only on video.

A. Visual Feature-based Deepfake Detection

A visual feature-based deepfake detection method was first proposed in 2018 using blink detection[4]. The assumption used in this approach is that there is a difference in the eye blink pattern between the deepfake video and the original video. Another visual feature approach is done using inconsistency head poses[5]. This approach calculates the inconsistency between the pose of the face and parts of the body outside the face, such as the neck and shoulders. Figure 3 shows how to measure head pose inconsistencies. Inconsistency was measured by comparing pose directions based on 68 facial landmarks (blue landmarks) and compared with pose directions from 17 facial landmarks (red landmarks) which represented the pose direction from the center area of the face..

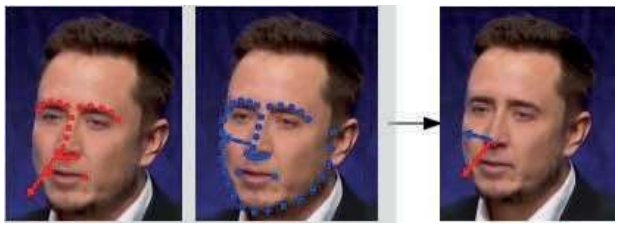

Fig. 4. Illustration of Measuring Head's Pose Inconsistency[5]

Another visual feature approach tries to extract some visual features in deepfake faces which are called visual artifacts. This visual artifact is obtained from imperfect deepfake content because the resources used in the deepfake generation process are limited[28]. These characteristics include differences in the color of the left and right eyes,

disproportionate shadows, details of light reflection that do not appear, and geometry that is not detailed..

We can see from Fig. 4 that several visual artifacts, such as the difference in color between the left eye and the right eye, disproportionate shadows on the nose area, reflected light that does not appear in the eye, and the geometry of the teeth that is not detailed. To extract these visual features, a geometric and color-based extraction approach is used from specific areas of the face, such as eyes, nose, eyebrows, lips and teeth.

Fig. 5. Some Visual Artifact from Deepfake Image[28]

These visual feature approaches can be used to detect deepfakes. However, with the development of deepfake content generation methods, this visual feature has become increasingly difficult to detect, so this visual feature-based deepfake detection method becomes less reliable.

B. Local Feature-based Deepfake Detection

The local feature-based deepfake detection method uses feature extraction methods that extract the image features at the pixel level. Research conducted in 2017 used a combination of two features of the image convolution and the steganalysis feature of the image to detect modified facial areas in the image[3]. This method is the basis of detection methods based on local feature and deep feature. Another approach is to use Photo Response Non Uniformity (PRNU) analysis with cross correlation operation[29]. However, the study only used 10 videos as the dataset.

Another feature extraction method used in the deepfake detection process is the Scale Invariant Feature Transform (SIFT) which detects the keypoint pixel in the image and extracts features from the keypoint [36]. The SIFT method and several other feature extraction methods such as Local Binary Pattern (LBP), Binary Gabor Pattern (BGP), Binarized Statistical Image Features (BSIF), Local Phase Quantization (LPQ), Pyramid of Histogram of Oriented Gradients (PHOG), Speeded Up Robust Feature (SURF) and Image Quality Metric (IQM) were analyzed to determine the best feature extraction method in detecting deepfakes[6]. The result is that the IQM method produces the best performance in detecting deepfakes. This result was confirmed in a study comparing the performance of IQM with PCA and LDA which showed the best performance on the IQM method [30].

C. Deep Feature-based Deepfake Detection

The local feature-based deepfake detection method has a fairly good performance in detecting deepfakes in some video data. However, as the deepfake algorithm develops, the resulting images and videos become more natural and increasingly difficult to detect as deepfakes. More complex features are needed to distinguish deepfake images and videos from original images and videos. The deepfake detection method based on deep feature is a method based on the deep layer neural network that can extract features that are more complex than the local feature extraction method. Research conducted in 2018 compared several CNN architectures such as InceptionNet, DenseNet and XceptionNet in detecting deepfake images[31]. This research shows the reliability of XceptionNet architecture in detecting deepfake images.

Another study proposes a 5-layer CNN architecture called Deep Forgery Discriminator (DeepFD) with a loss function in the form of contrastive loss which provides good performance for some deepfake generating GAN methods[9]. The DeepFD model was developed in subsequent studies using a pairwise learning approach that succeeded in increasing the generalizability of DeepFD[32]. This pairwise learning approach studies the features of two pairs of original and fake images and calculates the difference between the two using contrastive learning. From this pairwise learning approach, a new model called the Common Feature Fake Network (CFFN) is produced.

Another CNN model developed in deepfake detection is MesoNet which uses the inception module as the backbone of its architecture[10]. This model is able to detect deepfake videos with compressed video conditions as the conditions of videos uploaded on social media applications. Fig. 6 shows the architecture of MesoNet.

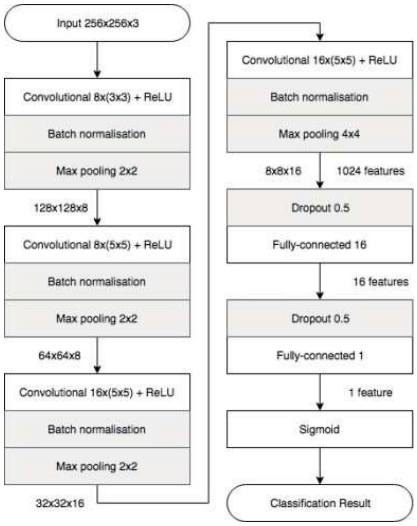

Fig. 6. MesoNet Architecture[10]

However, research using Capsule Network architecture is able to match the performance of MesoNet in the case of deepfake video detection, both for the per frame level and for the level of one whole video[11]. Capsule Network corrects the weakness of unequivariance convolutional blocks by using a routing by agreement mechanism[33]. Another deep neural network model developed in deepfake detection is a deep autoencoder that is able to reconstruct images and segment the deepfake area of the image[12]. The autoencoder uses Y-shaped decoder that has two branches, one branch to reconstruct the deepfake area segmentation and one branch to calculate the loss function. This autoencoder architecture can be seen in Fig. 7. With the segmentation in the image area, this method is able to provide a description of the detection results that is more than just detecting deepfake or original content.

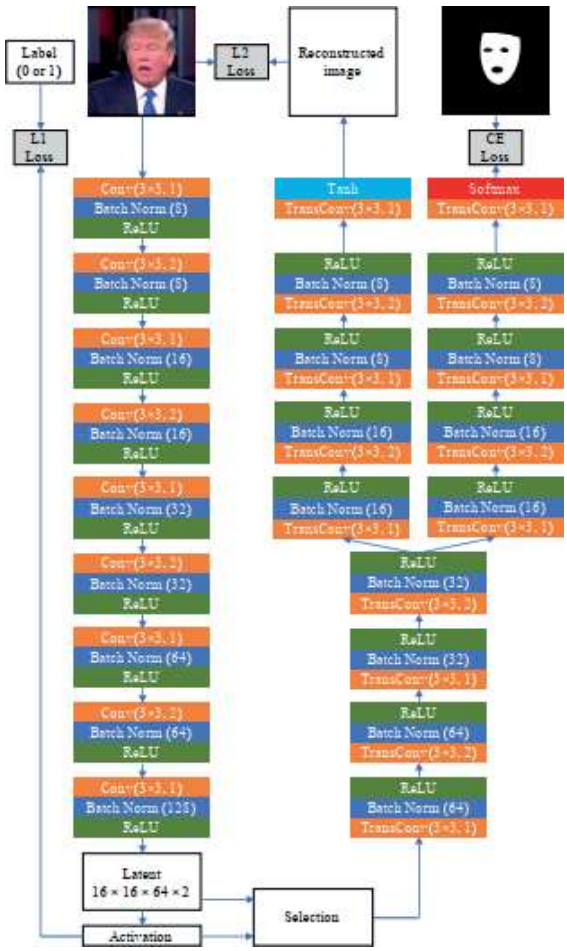

Fig. 7. Autoencoder using Y-shaped decoder[12]

D. *Temporal Feature-based Deepfake Detection*

Temporal feature-based deepfake detection performs the video detection process using temporal features obtained from a series of sequential frames from a video. Basically, a series of frames can be thought of as sequential data sets. A well-known method for processing sequential data is Reccurent Neural Network (RNN). This model was used in 2018 to detect deepfake content in videos[7]. Fig. 8 shows an illustration of the RNN architecture. x is the input data, s is the processing on RNN, o is the output, U is the input weight matrix, V is the output weight matrix and W is the output processing matrix to be added as input at the next time. It can be seen that data from one time, for example x_0, will be processed and the results of the processing will be used to process the next data x_1.

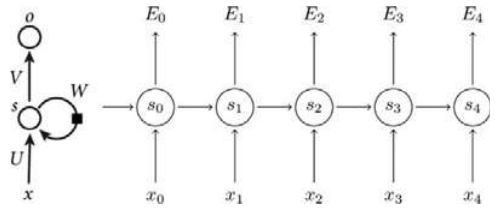

Fig. 8. RNN Illustration[34]

Apart from RNN, the Optical Flow method is also used in deepfake video detection research with CNN as a classification algorithm[8]. Deepfake video detection with RNN was re-investigated in 2019 using a bidirectional RNN model with feature extraction of DenseNet frames[35]. This model outperforms several deep neural network models such as ResNet50 and DenseNet.

Each of the deepfake detection approaches has advantages and disadvantages. The deepfake detection approach with visual features can in some cases detect deepfakes in images and videos. Another advantage of this approach is that it has good interpretability skills so that this approach is easy for humans to understand. However, with the development of deepfake content development algorithms, this visual feature approach is no longer effective. Visually, the new deepfake content looks more natural and closer to the original video. Characteristics such as different eye blinking frequency, different eye color and facial geometric imperfections cannot be used to detect deepfakes in the most recent deepfake dataset.

The local feature approach has better reliability than visual features because it extracts all areas of the face and the resulting features are intrinsic features that are not visible visually. The local feature approach is proven to have higher accuracy than the visual feature in detecting deepfakes with an EER difference of up to 38% in the DF-TIMIT dataset[30]. The deep feature approach works like a local feature but uses a deeper filter so that it can extract more complex features. This deep feature approach is the approach that currently has the highest level of accuracy, namely deepfake detection using the XceptionNet architecture which has achieved an accuracy rate of up to 99.7% on the FF-DF dataset[36]. The problem with this deep feature approach is its low level of interpretability because it cannot explain which pixel area is part of the deepfake. Meanwhile, the temporal feature-based detection method has the advantage of capturing temporal features that cannot be captured by other approaches. However, from the results of research that has been done, this method has not been able to outperform the performance of the deep feature-based method[37].

IV. Deepfake Dataset

The following are some of the datasets commonly used in research on deepfake detection in images and videos[36].

a. The UADFV dataset contains 49 original videos from YouTube and 49 deepfake videos built on the Deep Neural Network (DNN) model.
b. The DF-TIMIT dataset contains 640 deepfake videos built from faceswap-GAN and based on the Vid-TIMIT dataset. Videos are divided into two categories: low quality DF-TIMIT-LQ and high quality DF-TIMIT-HQ.
c. FaceForensics ++ (FF-DF) dataset containing 1000 original videos from YouTube and 1000 deepfake videos built using faceswap.
d. The Google / Jigsaw Deepfake detection (DFD) dataset contains 3068 deepfake videos built from 363 original videos with 28 individuals of various genders, ages and ethnicities.
e. The Facebook Deepfake Detection Challenge (DFDC) dataset contains 4113 deepfake videos built from 1131

original videos with 66 individuals of various genders, ages and ethnicities.

f. The Celeb-DF dataset containing 590 original videos and 5639 deepfake videos was built using faceswap.

Table 1 shows the results of the performance comparison of several deepfake detection models using this dataset.

TABLE 1. COMPARISON OF THE PERFORMANCE OF DEEPFAKE DETECTION METHODS ON SEVERAL DATASETS[36]

Metode	UADFV	DF-TIMIT		FF-DF	DFD	DFDC	Celeb-DF
		LQ	HQ				
Two-stream	85,1	83,5	73,5	70,1	**52,8**	**61,4**	**53,8**
Meso4	84,3	87,8	**68,4**	84,7	76	75,3	**54,8**
Face Warping *Artifact*	97,4	99,9	93,2	80,1	74,3	72,7	**56,9**
Head Pose	89	**55,1**	**53,2**	**47,3**	**56,1**	**55,9**	**54,6**
Visual Artifact	70,2	**61,4**	**62,1**	**66,4**	**69,1**	**61,9**	**55**
Xception	91,2	95,9	94,4	99,7	85,9	72,2	**65,3**
Multi-task	**65,8**	**62,2**	**55,3**	76,3	**54,1**	**53,6**	**54,3**
Capsule Network	**61,3**	78,4	74,4	96,6	**64**	**53,3**	**57,5**

From Table 2, we can see that the Celeb-DF dataset has the highest difficulty level in deepfake detection problem. We can see that for the Celeb-DF dataset, the performance of the deepfake detection methods that have been developed to date is still below 70%. The last three datasets (DFD, DFDC and Celeb-DF) have different difficulty levels compared to other datasets. It can be seen that the currently developed deepfake detection methods are still unable to handle a wider variety of deepfake video data.

In order for the classification algorithm to recognize varied data, it requires good generalizability. Generalization is the ability of a classification algorithm to recognize patterns that are different from previously studied data patterns[38]. For deepfake detection cases, Ranjan et al.[39] analyzed the generalizability of deepfake detection methods using the three datasets DFD, DFDC and Celeb-DF. The detection algorithm used is XceptionNet, which is one of the best accuracy deepfake detection methods. From the results of the tests carried out, it was found that the deepfake detection method has good accuracy in recognizing testing sets originating from the same dataset as the training set. However, when tested with a testing set derived from a different dataset from the training set, the accuracy of the deepfake detection method fell below 70%. This suggests a problem with the generalizability of existing deepfake detection methods.

V. CONCLUSION

Deepfake technology has the potential to be abused which can result in harm to many people. Therefore, a deepfake recognition system that has a high accuracy is needed. Until now, there are four types of approaches in detecting deepfakes, those are visual features, local features, deep features and temporal features. The main challenge in detecting deepfake content is how deepfake detection methods can recognize different deepfake content. This is strongly influenced by the generalizability of deepfake detection methods. Research must continue to be conducted to establish deepfake detection methods that have good detection capabilities and also have good generalizability in order to recognize more varied patterns of deepfake content in the future.

REFERENCES

[1] M. Westerlund, "The Emergence of Deepfake Technology: A Review," *Technol. Innov. Manag. Rev.*, vol. 9, no. 11, pp. 39–52, Jan. 2019, doi: 10.22215/timreview/1282.

[2] E. Meskys, A. Liaudanskas, J. Kalpokiene, and P. Jurcys, "Regulating deep fakes: legal and ethical considerations," *J. Intellect. Prop. Law Pract.*, vol. 15, no. 1, pp. 24–31, Jan. 2020, doi: 10.1093/jiplp/jpz167.

[3] P. Zhou, X. Han, V. I. Morariu, and L. S. Davis, "Two-Stream Neural Networks for Tampered Face Detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Mar. 2018, pp. 1831–1839, Accessed: Apr. 01, 2020. [Online]. Available: http://arxiv.org/abs/1803.11276.

[4] Y. Li, M.-C. Chang, and S. Lyu, "In Ictu Oculi: Exposing AI Created Fake Videos by Detecting Eye Blinking," in *2018 IEEE International Workshop on Information Forensics and Security (WIFS)*, Hong Kong, Hong Kong, Dec. 2018, pp. 1–7, doi: 10.1109/WIFS.2018.8630787.

[5] X. Yang, Y. Li, and S. Lyu, "Exposing Deep Fakes Using Inconsistent Head Poses," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, United Kingdom, May 2019, pp. 8261–8265, doi: 10.1109/ICASSP.2019.8683164.

[6] Z. Akhtar and D. Dasgupta, "A Comparative Evaluation of Local Feature Descriptors for DeepFakes Detection," in *2019 IEEE International Symposium on Technologies for Homeland Security (HST)*, Woburn, MA, USA, Nov. 2019, pp. 1–5, doi: 10.1109/HST47167.2019.9033005.

[7] D. Guera and E. J. Delp, "Deepfake Video Detection Using Recurrent Neural Networks," in *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Auckland, New Zealand, Nov. 2018, pp. 1–6, doi: 10.1109/AVSS.2018.8639163.

[8] I. Amerini, L. Galteri, R. Caldelli, and A. Del Bimbo, "Deepfake Video Detection through Optical Flow Based CNN," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, Seoul, Korea (South), Oct. 2019, pp. 1205–1207, doi: 10.1109/ICCVW.2019.00152.

[9] C.-C. Hsu, C.-Y. Lee, and Y.-X. Zhuang, "Learning to Detect Fake Face Images in the Wild," in *2018 International Symposium on Computer, Consumer and Control (IS3C)*, Taichung, Taiwan, Dec. 2018, pp. 388–391, doi: 10.1109/IS3C.2018.00104.

[10] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "MesoNet: a Compact Facial Video Forgery Detection Network," in *2018 IEEE International Workshop on Information Forensics and Security (WIFS)*, Hong Kong, Hong Kong, Dec. 2018, pp. 1–7, doi: 10.1109/WIFS.2018.8630761.

[11] H. H. Nguyen, J. Yamagishi, and I. Echizen, "Capsule-forensics: Using Capsule Networks to Detect Forged Images and Videos," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, United Kingdom, May 2019, pp. 2307–2311, doi: 10.1109/ICASSP.2019.8682602.

[12] H. H. Nguyen, F. Fang, J. Yamagishi, and I. Echizen, "Multi-task Learning For Detecting and Segmenting Manipulated Facial Images and Videos," presented at the 2019 IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS), Jun. 2019, Accessed: Apr. 01, 2020. [Online]. Available: http://arxiv.org/abs/1906.06876.

[13] J. H. Bappy, C. Simons, L. Nataraj, B. S. Manjunath, and A. K. Roy-Chowdhury, "Hybrid LSTM and Encoder–Decoder Architecture for Detection of Image Forgeries," *IEEE Trans. Image Process.*, vol. 28, no. 7, pp. 3286–3300, Jul. 2019, doi: 10.1109/TIP.2019.2895466.

[14] R. Thakur and R. Rohilla, "Copy-Move Forgery Detection using Residuals and Convolutional Neural Network Framework: A Novel Approach," in *2019 2nd International Conference on Power Energy, Environment and Intelligent Control (PEEIC)*, Greater Noida, India, Oct. 2019, pp. 561–564, doi: 10.1109/PEEIC47157.2019.8976868.

[15] L. Su, C. Li, Y. Lai, and J. Yang, "A Fast Forgery Detection Algorithm Based on Exponential-Fourier Moments for Video Region Duplication," *IEEE Trans. Multimed.*, vol. 20, no. 4, pp. 825–840, Apr. 2018, doi: 10.1109/TMM.2017.2760098.

[16] S. Jia, Z. Xu, H. Wang, C. Feng, and T. Wang, "Coarse-to-Fine Copy-Move Forgery Detection for Video Forensics," *IEEE Access*, vol. 6, pp. 25323–25335, 2018, doi: 10.1109/ACCESS.2018.2819624.

[17] L. Su and C. Li, "A novel passive forgery detection algorithm for video region duplication," *Multidimens. Syst. Signal Process.*, vol. 29, no. 3, pp. 1173–1190, Jul. 2018, doi: 10.1007/s11045-017-0496-6.

[18] D.-N. Zhao, R.-K. Wang, and Z.-M. Lu, "Inter-frame passive-blind forgery detection for video shot based on similarity analysis," *Multimed. Tools Appl.*, vol. 77, no. 19, pp. 25389–25408, Oct. 2018, doi: 10.1007/s11042-018-5791-1.

[19] G. Singh and K. Singh, "Video frame and region duplication forgery detection based on correlation coefficient and coefficient of variation," *Multimed. Tools Appl.*, vol. 78, no. 9, pp. 11527–11562, May 2019, doi: 10.1007/s11042-018-6585-1.

[20] M. Aloraini, M. Sharifzadeh, C. Agarwal, and D. Schonfeld, "Statistical Sequential Analysis for Object-based Video Forgery Detection," *Electron. Imaging*, vol. 2019, no. 5, pp. 543-1-543–7, Jan. 2019, doi: 10.2352/ISSN.2470-1173.2019.5.MWSF-543.

[21] M. Saddique, K. Asghar, U. I. Bajwa, M. Hussain, and Z. Habib, "Spatial Video Forgery Detection and Localization using Texture Analysis of Consecutive Frames," *Adv. Electr. Comput. Eng.*, vol. 19, no. 3, pp. 97–108, 2019, doi: 10.4316/AECE.2019.03012.

[22] C. C. Huang, Y. Zhang, and V. L. L. Thing, "Inter-frame video forgery detection based on multi-level subtraction approach for realistic video forensic applications," in *2017 IEEE 2nd International Conference on Signal and Image Processing (ICSIP)*, Singapore, Aug. 2017, pp. 20–24, doi: 10.1109/SIPROCESS.2017.8124498.

[23] Y. Guo, X. Ke, and J. Ma, "A Face Replacement Neural Network for Image and Video," in *Proceedings of the 2019 11th International Conference on Machine Learning and Computing - ICMLC '19*, Zhuhai, China, 2019, pp. 163–167, doi: 10.1145/3318299.3318311.

[24] I. Goodfellow *et al.*, "Generative Adversarial Nets," 2014, pp. 2672--2680.

[25] B. Dai, S. Fidler, R. Urtasun, and D. Lin, "Towards Diverse and Natural Image Descriptions via a Conditional GAN," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Oct. 2017, pp. 2989–2998, doi: 10.1109/ICCV.2017.323.

[26] Y. Li and S. Lyu, "Exposing DeepFake Videos By Detecting Face Warping Artifacts," in *2019 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, p. 7.

[27] L. Jiang, R. Li, W. Wu, C. Qian, and C. C. Loy, "DeeperForensics-1.0: A Large-Scale Dataset for Real-World Face Forgery Detection," 2020, pp. 2889–2898, Accessed: Aug. 24, 2020. [Online]. Available: https://openaccess.thecvf.com/content_CVPR_2020/html/Jiang_DeeperForensics-1.0_A_Large-Scale_Dataset_for_Real-World_Face_Forgery_Detection_CVPR_2020_paper.html.

[28] F. Matern, C. Riess, and M. Stamminger, "Exploiting Visual Artifacts to Expose Deepfakes and Face Manipulations," in *2019 IEEE Winter Applications of Computer Vision Workshops (WACVW)*, Waikoloa Village, HI, USA, Jan. 2019, pp. 83–92, doi: 10.1109/WACVW.2019.00020.

[29] M. Koopman, A. M. Rodriguez, and Z. Geradts, "Detection of Deepfake Video Manipulation," in *The 20th Irish Machine Vision and Image Processing Conference (IMVIP). 2018*, 2018, p. 4.

[30] P. Korshunov and S. Marcel, "Vulnerability assessment and detection of Deepfake videos," in *2019 International Conference on Biometrics (ICB)*, Crete, Greece, Jun. 2019, pp. 1–6, doi: 10.1109/ICB45273.2019.8987375.

[31] F. Marra, D. Gragnaniello, D. Cozzolino, and L. Verdoliva, "Detection of GAN-Generated Fake Images over Social Networks," in *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, Miami, FL, Apr. 2018, pp. 384–389, doi: 10.1109/MIPR.2018.00084.

[32] C.-C. Hsu, Y.-X. Zhuang, and C.-Y. Lee, "Deep Fake Image Detection Based on Pairwise Learning," *Appl. Sci.*, vol. 10, no. 1, p. 370, Jan. 2020, doi: 10.3390/app10010370.

[33] A. Jaiswal, W. AbdAlmageed, Y. Wu, and P. Natarajan, "CapsuleGAN: Generative Adversarial Capsule Network," in *Computer Vision – ECCV 2018 Workshops*, vol. 11131, L. Leal-Taixé and S. Roth, Eds. Cham: Springer International Publishing, 2019, pp. 526–535.

[34] L. Medsker and L. C. Jain, *Recurrent Neural Networks: Design and Applications*. CRC Press, 1999.

[35] E. Sabir, J. Cheng, A. Jaiswal, W. AbdAlmageed, I. Masi, and P. Natarajan, "Recurrent Convolutional Strategies for Face Manipulation Detection in Videos," in *2019 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 80–87.

[36] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics," *ArXiv190912962 Cs Eess*, Mar. 2020, Accessed: Apr. 01, 2020. [Online]. Available: http://arxiv.org/abs/1909.12962.

[37] U. A. Ciftci, I. Demir, and L. Yin, "FakeCatcher: Detection of Synthetic Portrait Videos using Biological Signals," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–1, 2020, doi: 10.1109/TPAMI.2020.3009287.

[38] X.-Z. Wang, R. Wang, and C. Xu, "Discovering the Relationship Between Generalization and Uncertainty by Incorporating Complexity of Classification," *IEEE Trans. Cybern.*, vol. 48, no. 2, pp. 703–715, Feb. 2018, doi: 10.1109/TCYB.2017.2653223.

[39] P. Ranjan, S. Patil, and F. Kazi, "Improved Generalizability of Deep-Fakes Detection using Transfer Learning Based CNN Framework," in *2020 3rd International Conference on Information and Computer Technologies (ICICT)*, San Jose, CA, USA, Mar. 2020, pp. 86–90, doi: 10.1109/ICICT50521.2020.00021.