DeepFakes Evolution: Analysis of Facial Regions and Fake Detection Performance

Ruben Tolosana, Sergio Romero-Tapiador, Julian Fierrez and Ruben Vera-Rodriguez
Biometrics and Data Pattern Analytics - BiDA Lab, Universidad Autonoma de Madrid
{ruben.tolosana, sergio.romerot, julian.fierrez, ruben.vera}@uam.es

Abstract—Media forensics has attracted a lot of attention in the last years in part due to the increasing concerns around DeepFakes. Since the initial DeepFake databases from the 1st generation such as UADFV and FaceForensics++ up to the latest databases of the 2nd generation such as Celeb-DF and DFDC, many visual improvements have been carried out, making fake videos almost indistinguishable to the human eye. This study provides an exhaustive analysis of both 1st and 2nd DeepFake generations in terms of facial regions and fake detection performance. Two different approaches are considered in our experimental framework: i) selecting the entire face as input to the fake detection system, and ii) selecting specific facial regions such as the eyes or nose, among others, as input to the fake detection system.

Among all the findings resulting from our experiments, we highlight the poor fake detection results achieved even by the strongest state-of-the-art fake detectors in the latest DeepFake databases of the 2nd generation, with Equal Error Rate results ranging from 15% to 30%. These results remark the necessity of further research in order to develop more sophisticated fake detection systems.

Index Terms—Fake News, DeepFakes, Media Forensics, Face Manipulation, Fake Detection, Benchmark

I. INTRODUCTION

Fake images and videos including facial information generated by digital manipulations, in particular with DeepFake methods [1], [2], have become a great public concern recently [3], [4]. The very popular term “DeepFake” is referred to a deep learning based technique able to create fake videos by swapping the face of a person by the face of another person. Open software and mobile applications such as ZAO [5] allow nowadays to automatically generate fake videos by anyone, without a prior knowledge of the task. But, how real are these fake videos compared with the authentic ones?

Since the initial publicly available fake databases, such as the UADFV database [5], up to the recent Celeb-DF and DeepFake Detection Challenge (DFDC) databases [6], [7], many visual improvements have been carried out, increasing the realism of fake videos. As a result, face swap databases can be divided into two different generations.

In general, fake videos from the 1st generation are characterised by: i) low-quality synthesised faces, ii) different colour contrast among the synthesised fake mask and the skin of the original face, iii) visible boundaries of the fake mask, iv) visible facial elements from the original video, v) low pose variations, and vi) strange artifacts among sequential frames. Also, they usually consider controlled scenarios in terms of camera position and light conditions. Many of these aspects have been successfully improved in databases of the 2nd generation. For example, the recent DFDC database considers different acquisition scenarios (i.e., indoors and outdoors), light conditions (i.e., day, night, etc.), distances from the person to the camera, and pose variations, among others. So, the question is, how easy is for a machine to automatically detect these kind of fakes?

Different fake detection approaches have been proposed based on the visual features existed in the 1st generation of fake videos. Yang et al. performed in [8] a study based on the differences between head poses using a full set of facial landmarks (68 extracted from DLib [9]) and those in the central face regions to differentiate fake videos from real videos. Once these features were extracted, Support Vector Machines (SVM) were considered for the final classification, achieving a final Area Under the Curve (AUC) of 89.0% for the UADFV database [5].

The same authors proposed in [10] another approach based on the detection of face warping artifacts. They proposed a detection system based on Convolutional Neural Networks (CNNs) in order to detect the presence of such artifacts from the detected face regions and the surrounding areas. Their proposed detection approach was tested using the UADFV and DeepfakeTIMIT databases [5], [11], outperforming the state of the art with 97.4% and 99.9% AUCs, respectively.

Agarwal et al. proposed in [12] a detection technique based on facial expressions and head movements. Their proposed approach achieved a final AUC of 96.3% over their own database, being robust against new contexts and manipulation techniques.

Finally, Sabir et al. [13] proposed a method to detect fake videos based on the temporal discrepancies across frames. They considered a recurrent convolutional network similar to [14], trained end-to-end instead of using a pre-trained model. Their proposed detection approach was tested through FaceForensics++ database [15], achieving AUC results of 96.9% and 96.3% for the DeepFake and FaceSwap methods, respectively.

Therefore, very good fake detection results are already achieved on databases from the 1st generation, being an almost solved problem. But, what is the performance achieved on current face swap databases of the 2nd generation?
The present study provides an exhaustive analysis of both 1\(^{\text{st}}\) and 2\(^{\text{nd}}\) DeepFake generations. Two different approaches are considered to detect fake videos: i) selecting the entire face as input to the fake detection system, and ii) selecting specific facial regions such as the eyes or nose, among others, as input to the fake detection system. This analysis intends to provide more details about the improvements achieved between 1\(^{\text{st}}\) and 2\(^{\text{nd}}\) generations, not only at performance level but also at visual level. The main contribution of this study is two-fold:

- In-depth comparison in terms of performance among fake video databases from the 1\(^{\text{st}}\) and 2\(^{\text{nd}}\) generation using a state-of-the-art fake detection system based on Xception network [16].
- Analysis of the different facial regions between the 1\(^{\text{st}}\) and 2\(^{\text{nd}}\) generations. This analysis can be very valuable for both fake detection and fake synthesis to improve the next generation of DeepFakes.

The remainder of the paper is organised as follows. Sec. II describes our proposed evaluation framework. Sec. III describes all databases considered in the experimental framework of this study. Sec. IV and V describe the experimental protocol and results achieved, respectively. Finally, Sec. VI draws the final conclusions and points out some future lines of research.

II. PROPOSED EVALUATION FRAMEWORK

Fig. 1 graphically summarises our evaluation framework for the analysis of both facial regions and fake detection performance of 1\(^{\text{st}}\) and 2\(^{\text{nd}}\) DeepFake generations. Two different fake detection approaches are considered: i) using the entire face as input to the fake detection system, and ii) using only specific facial regions.

Regarding the second approach, 4 different facial regions are selected: eyes, nose, mouth, and rest (i.e., the part of the face obtained after removing the eyes, nose, and mouth from the entire face). For the segmentation of each part, we consider the open-source toolbox OpenFace2 [17]. This toolbox extracts 68 total landmarks for each face. Fig. 2 represents an example of the 68 landmarks (blue circles) extracted by OpenFace2 over a frame of the Celeb-DF database. It is important to highlight that OpenFace2 is robust against pose variations, distance from the camera, and light conditions, extracting reliable landmarks even for challenging databases such as the DFDC database [7]. The specific key landmarks considered in our evaluation framework to extract each facial region are as follow:

- **Eyes**: using landmark points from 18 to 27 (top of the mask) and using landmarks 1, 2, 16, and 17 (bottom of the mask).
- **Nose**: using landmark points 22, 23 (top of the mask), from 28 to 36 (length and bottom of the nose), and 40, 43 (width of the middle-part of the nose).
- **Mouth**: using landmark points 49, 51-53, 55, and 57-59 to build a circular/elliptical mask.
- **Rest**: extracted after removing eyes, nose, and mouth masks from the entire face.

Each facial region is highlighted by yellow lines in Fig. 2. Once each facial region is segmented, the remaining part of the face is discarded (black background as depicted in Fig. 1). Also, for each facial region, we keep the same image size and resolution as the original face image to perform a fair evaluation among facial regions and the entire face, avoiding therefore the influence of other pre-processing aspects such as interpolation.

Finally, we train one specific fake detection system for each database and facial region as shown in Fig. 1. In particular, we consider the state-of-the-art Xception network proposed in [16]. This network has achieved very good fake detection performances in recent studies [7], [15], [18], [19]. Xception is a CNN architecture inspired by Inception [20], where Inception modules have been replaced with depthwise separable
convolutions. In our evaluation framework, we follow the same training approach considered in [15]: 

1. We first consider the Xception model pre-trained with ImageNet [21].
2. We change the last fully-connected layer of the ImageNet model by a new one (two classes, real or fake).
3. We fix all weights up to the final fully-connected layer and pre-train the network for few epochs, and finally
4. We train the whole network for 20 more epochs and choose the best performing model based on validation accuracy.

All experiments are implemented under Keras framework using Tensorflow as back-end, with a NVIDIA GeForce RTX 2080 Ti GPU. Adam optimiser is considered with default parameters (learning rate of 0.002) and a loss function based on binary cross-entropy.

III. DATABASES

Four different public databases are considered in the experimental framework of this study. In particular, two databases from the 1st generation (UADFV and FaceForensics++) and two recent databases from the 2nd generation (Celeb-DF and DFDC). Table I summarises their main features.

A. UADFV

The UADFV database [5] comprises 49 real videos and 49 fake videos. Real videos were downloaded from Youtube whereas fake videos were created using FakeApp [3] swapping in all of them the original face by the face of Nicolas Cage. The average length of the videos is approximately 11.14 seconds, with a typical resolution of 294×500 pixels.

B. FaceForensics++

The FaceForensics++ database [15] was introduced in 2019 as an extension of the original FaceForensics database [22], which was focused only on facial expression manipulation. FaceForensics++ contains 1000 real videos extracted from Youtube. Fake videos were generated using both computer graphics and deep learning approaches (1000 fake videos per approach). In this study we focus on the computer graphics approach where fake videos were created using the publicly available FaceSwap algorithm [4]. This algorithm consists of face alignment, Gauss Newton optimization and image blending to swap the face of the source person to the target person.

C. Celeb-DF

The Celeb-DF database [6] aims to generate fake videos of better visual quality. This database consists of 408 real videos extracted from Youtube, corresponding to interviews of 59 celebrities with a diverse distribution in terms of gender, age, and ethnic group. In addition, these videos exhibit large range of variations in aspects such as the face sizes (in pixels), orientations, lighting conditions, and backgrounds. Regarding fake videos, a total of 795 videos were created using DeepFake technology, swapping faces for each pair of the 59 subjects. The final videos are in MPEG4.0 format.

D. DFDC

The DFDC database [7] is one of the latest public DeepFake databases, released by Facebook in collaboration with other companies and academic institutions such as Microsoft, Amazon, and the MIT. In the present paper we consider the DFDC preview dataset consisting of 1131 real videos from 66 paid actors, ensuring realistic variability in gender, skin tone and age. It is important to remark that no publicly available data or data from social media sites was used to create this dataset, unlike other popular databases. Regarding fake videos, a total of 4119 videos were created using two different unknown fake approaches. Fake videos were generated by swapping subjects with similar appearances, i.e., similar facial attributes such as skin tone, facial hair, glasses, etc. After a given pairwise model

https://www.malavida.com/en/soft/fakeapp/
https://github.com/MarekKowalski/FaceSwap

| Database | Real Videos | Fake Videos |
|----------|-------------|-------------|
| 1st Generation |
| UADFV (2018) | 49 (Youtube) | 49 (FakeApp) |
| FaceForensics++ (2019) | 1000 (Youtube) | 1000 (FaceSwap) |
| 2nd Generation |
| Celeb-DF (2019) | 408 (Youtube) | 795 (DeepFake) |
| DFDC Preview (2019) | 1131 (Actors) | 4119 (Unknown) |
TABLE II
FAKE DETECTION PERFORMANCE RESULTS IN TERMS OF EER (%) AND AUC (%) OVER THE FINAL EVALUATION DATASETS. TWO APPROACHES ARE CONSIDERED AS INPUT TO THE FAKE DETECTION SYSTEMS: I) SELECTING THE ENTIRE FACE (Face), AND II) SELECTING SPECIFIC FACIAL REGIONS (Eyes, Nose, Mouth, Rest). 1ST GENERATION DATABASES: UADFV AND FACEFORENSICS++. 2ND GENERATION DATABASES: CELEB-DF AND DFDC. FOR EACH DATABASE, WE REMARK IN BOLD THE BEST FAKE DETECTION RESULTS, AND IN BLUE AND ORANGE THE FACIAL REGIONS THAT PROVIDE THE BEST AND WORST RESULTS.

|                | Face | Eyes | Nose | Mouth | Rest |
|----------------|------|------|------|-------|------|
|                | EER (%) | AUC (%) | EER (%) | AUC (%) | EER (%) | AUC (%) | EER (%) | AUC (%) | EER (%) | AUC (%) |
| UADFV (2018)   | 1.00  | 100.00 | 2.20  | 99.70 | 13.50 | 94.70 | 12.50 | 95.40 | 7.90  | 97.30  |
| FaceForensics++ (2019) | 3.31 | 99.40 | 14.23 | 92.70 | 21.97 | 86.30 | 13.77 | 93.90 | 22.37 | 85.50  |
| Celeb-DF (2019) | 28.55 | 83.60 | 29.40 | 77.30 | 38.46 | 64.90 | 39.37 | 65.10 | 43.55 | 60.10  |
| DFDC Preview (2019) | 17.55 | 91.00 | 23.82 | 83.90 | 26.80 | 81.50 | 27.59 | 79.50 | 29.94 | 76.50  |

was trained on two identities, they swapped each identity onto the others videos.

The DFDC database considers different acquisition scenarios (i.e., indoors and outdoors), light conditions (i.e., day, night, etc.), distances of the person to the camera, and pose variations, among others.

IV. EXPERIMENTAL PROTOCOL

All databases have been divided into non-overlapping datasets, development (≃ 80% of the identities) and evaluation (≃ 20% of the identities). It is important to remark that each dataset comprises videos from different identities (both real and fake), unlike some previous studies. This aspect is important in order to perform a fair evaluation and to predict the generalisation capacity of the fake detection systems against unseen identities. For example, for the UADFV database, all real and fake videos related to the identity of Donald Trump were considered only for the final evaluation of the models. For the DFDC database, we follow the same experimental protocol proposed in [7] as the authors already considered this concern. For the FaceForensics++ database, we consider 860 development videos and 140 evaluation videos per class (real/fake) as proposed in [15], selecting different identities in each dataset (one fake video is provided for each identity).

V. EXPERIMENTAL RESULTS

Table II shows the fake detection performance results achieved in terms of Equal Error Rate (EER) and AUC over the final evaluation datasets of both 1st and 2nd generation fake videos. Both approaches considered in this study are included in the table: i) selecting the entire face (Face), and ii) selecting specific facial regions such as Eyes, Nose, Mouth, and Rest. For each database, we remark in bold the best fake detection performance results achieved, and in blue and orange the facial regions that provide the best and worst results, respectively.

We first analyse the performance results achieved when feeding the systems with the entire face (Face). For the 1st generation, EER values of 1.00% and 3.31% are achieved for the UADFV and FaceForensics++ databases, respectively. Very good results close to 100% AUC are also obtained in both databases, proving how easy it is for the system to detect fake videos from the 1st generation, even when considering our proposed experimental protocol where different identities are used for training and evaluation. However, a high performance degradation is observed when using Celeb-DF and DFDC databases from the 2nd generation. In particular, EER values of 28.55% and 17.55% are achieved for Celeb-DF and DFDC databases, respectively. As a result, an average absolute worsening of 20.90% EER is produced compared with databases from the 1st generation. This degradation is specially substantial for the Celeb-DF database, with a 28.55% EER and 83.60% AUC. These results prove the higher realism achieved in the 2nd DeepFake generation.

Table II also includes the results achieved when using only specific facial regions. It is important to remark that a separate fake detection model is trained for each facial region and database. Finally, we also visualise in Fig. 3 which part of the image is more important for the final decision, for both real and fake examples. We consider the popular heatmap visualisation technique Grad-CAM [23].

In general, as shown in Table II, the facial region Eyes provides the best results whereas the Rest (i.e., the remaining part of the face after removing eyes, nose, and mouth) provides the worst results. For the UADFV database, the Eyes provides a 2.20% EER, close to the best result achieved using the entire Face (1.00% EER). This aspect was remarked by Matern et al. in [24], proposing features based on the missing reflection details of the eyes. Also, in this particular database, good results are achieved using the Rest of the face, a 7.90% EER. This is produced due to the different colour contrast among the synthesised fake mask and real skin, and also to the visible boundaries of the fake mask. These aspects can be noticed in the examples included in Fig. 3.

Regarding the FaceForensics++ database, the Mouth is the facial region that achieves the best result with a 13.77% EER. This is produced due to the lack of details in the teeth (blurred) and also the lip inconsistencies among the original face and the synthesised. Similar results are obtained when using the Eyes. It is interesting to see in Fig. 3 how the decision of the fake detection systems is mostly based on a single eye (the same
Fig. 3. Real and fake image examples of the DeepFake video databases evaluated in the present paper with their corresponding Grad-CAM heatmaps, representing the facial features most useful for each fake detector (i.e., *Face*, *Eyes*, *Nose*, *Mouth*, and *Rest*).
happens in other databases such as UADFV). Finally, the fake detection system based on the Rest of the face provides the worst result, a 22.37% EER. This may happen because both colour contrast and visible boundaries were further improved in FaceForensics++ compared with the UADFV database.

For the Celeb-DJ database, the best result is achieved when using the Eyes of the face, with a 29.40% EER. It is important to remark that this EER is 13 times higher than the original 2.20% EER achieved on the UADFV database. Similar poor fake detection results, around 40% EER, are obtained when using other facial regions, being one of the most challenging databases nowadays. Fig. 3 depicts some fake examples of the Celeb-DJ database, showing very realistic features such as the colour contrast, boundaries of the mask, quality of the eyes, teeth and nose, etc.

Regarding the DFDC database, better detection results are obtained compared with the Celeb-DJ database. In particular, the facial region Eyes provides the best results with a 23.82% EER, an absolute improvement of 5.58% EER compared with the Eyes facial region in Celeb-DJ database. Despite this performance improvement, the EER is still much worse compared with the databases from the 1st generation.

Finally, we would like to highlight the importance of selecting different identities (not only videos) for the development and final evaluation of fake detection systems, as we have done in our experimental evaluation. As an example of how relevant this aspect is, Table III shows the detection performance results in our experimental evaluation. As an example of how relevant and final evaluation of fake detection systems, as we have done in other databases such as UADFV and FaceForensics++, compared with the UADFV database.

Regarding the DFDC database, better detection results are obtained compared with the Celeb-DJ database. In particular, the facial region Eyes provides the best results with a 23.82% EER, an absolute improvement of 5.58% EER compared with the Eyes facial region in Celeb-DJ database. Despite this performance improvement, the EER is still much worse compared with the databases from the 1st generation.

In this study we have performed an exhaustive analysis of the DeepFakes evolution, focusing on both facial regions and fake detection performance. Popular databases such as UADFV and FaceForensics++ from the 1st generation, as well as the latest databases such as Celeb-DJ and DFDC from the 2nd generation, are considered in the analysis.

Two different approaches have been followed in our evaluation framework to detect fake videos: i) selecting the entire face as input to the fake detection system, and ii) selecting specific facial regions such as the eyes or nose, among others, as input to the fake detection system.

Regarding the fake detection performance, we highlight the poor results achieved in the latest databases of the 2nd generation, with results of 28.55% and 17.55% EER (83.60% and 91.00% AUC) for the Celeb-DJ and DFDC Preview databases, respectively. In addition, we remark the significant improvements achieved at image level in some facial regions such as the nose, mouth, and edge of the face in the databases of the 2nd generation, with fake detection results between 24% and 44% EERs.

The analysis carried out in this study provides useful insights about the improvements achieved between 1st and 2nd DeepFake generations, not only at performance level but also at visual level. This information can be very valuable for both the proposal of more robust fake detection systems and also the improvement of the next DeepFake generation.

**VI. CONCLUSIONS**

In this study we have performed an exhaustive analysis of the DeepFakes evolution, focusing on both facial regions and fake detection performance. Popular databases such as UADFV and FaceForensics++ from the 1st generation, as well as the latest databases such as Celeb-DJ and DFDC from the 2nd generation, are considered in the analysis.

Two different approaches have been followed in our evaluation framework to detect fake videos: i) selecting the entire face as input to the fake detection system, and ii) selecting specific facial regions such as the eyes or nose, among others, as input to the fake detection system.

Regarding the fake detection performance, we highlight the poor results achieved in the latest databases of the 2nd generation, with results of 28.55% and 17.55% EER (83.60% and 91.00% AUC) for the Celeb-DJ and DFDC Preview databases, respectively. In addition, we remark the significant improvements achieved at image level in some facial regions such as the nose, mouth, and edge of the face in the databases of the 2nd generation, with fake detection results between 24% and 44% EERs.

The analysis carried out in this study provides useful insights about the improvements achieved between 1st and 2nd DeepFake generations, not only at performance level but also at visual level. This information can be very valuable for both the proposal of more robust fake detection systems and also the improvement of the next DeepFake generation.

**ACKNOWLEDGMENTS**

This work has been supported by projects: PRIMA (H2020-MSCA-ITN-2019-860315), TRESPASS-ETN (H2020-MSCA-ITN-2019-860813) and BIBECA (MINECO/FEDER RTI2018-101248-B-I00). Ruben Tolosana is supported by Consejería de Educación, Juventud y Deporte de la Comunidad de Madrid y Fondo Social Europeo.

**REFERENCES**

[1] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, “DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection,” arXiv preprint arXiv:2001.00179, 2020.

[2] L. Verdoliva, “Media Forensics and DeepFakes: an Overview,” arXiv preprint arXiv:2001.00179, 2020.

[3] D. Citron, “How DeepFake Undermine Truth and Threaten Democracy,” https://www.ted.com, 2019.

[4] R. Cellan-Jones, “Deepfake Videos Double in Nine Months,” https://www.bbc.com/news/technology-49961089, 2019.

[5] Y. Li, M. Chang, and S. Lyu, “In Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking,” in Proc. IEEE International Workshop on Information Forensics and Security, 2018.

[6] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, “Celeb-DJ: A New Dataset for DeepFake Forensics,” arXiv preprint arXiv:1909.12962, 2019.

[7] B. Dollhansky, R. Howes, B. Pflaum, N. Baram, and C. Ferrer, “The Deepfake Detection Challenge (DFDC) Preview Dataset,” arXiv preprint arXiv:1910.08854, 2019.
[8] X. Yang, Y. Li, and S. Lyu, “Exposing Deep Fakes Using Inconsistent Head Poses,” in Proc. IEEE Int. Conference on Acoustics, Speech and Signal Processing, 2019.

[9] D. King, “DLib-ML: A Machine Learning Toolkit,” Journal of Machine Learning Research, vol. 10, pp. 1755–1758, 2009.

[10] Y. Li and S. Lyu, “Exposing DeepFake Videos By Detecting Face Warping Artifacts,” in Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops, 2019.

[11] P. Korshunov and S. Marcel, “Deepfakes: a New Threat to Face Recognition? Assessment and Detection,” arXiv preprint arXiv:1812.08685, 2018.

[12] S. Agarwal, H. Farid, Y. Gu, M. He, K. Nagano, and H. Li, “Protecting World Leaders Against Deep Fakes,” in Proc. Conf. Computer Vision and Pattern Recognition Workshops, 2019.

[13] E. Sahir, J. Cheng, A. Jaiswal, W. AbdAlmageed, J. Masi, and P. Natarajan, “Recurrent Convolutional Strategies for Face Manipulation Detection in Videos,” in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019.

[14] D. Güera and E. Delp, “Deepfake Video Detection Using Recurrent Neural Networks,” in Proc. IEEE Int. Conference on Advanced Video and Signal Based Surveillance, 2018.

[15] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, “FaceForensics++: Learning to Detect Manipulated Facial Images,” in Proc. IEEE/CVF International Conference on Computer Vision, 2019.

[16] F. Chollet, “Xception: Deep Learning with Depthwise Separable Convolutions,” in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2017.

[17] T. Baltrusaitis, A. Zadeh, Y.C. Lim, and L.P. Morency, “OpenFace 2.0: Facial Behavior Analysis Toolkit,” in Proc. IEEE International Conference on Automatic Face & Gesture Recognition, 2018.

[18] J. Neves, R. Tolosana, R. Vera-Rodriguez, V. Lopes, H. Proença, and J. P. Fierrez, “GANprintR: Improved Fakes and Evaluation of the State-of-the-Art in Face Manipulation Detection,” arXiv preprint arXiv:1911.05351, 2019.

[19] J. Stehouwer, H. Dang, F. Liu, X. Liu, and A. Jain, “On the Detection of Digital Face Manipulation,” arXiv preprint arXiv:1910.01717, 2019.

[20] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going Deeper with Convolutions,” in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2015.

[21] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, “Imagenet: A Large-Scale Hierarchical Image Database,” in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2009.

[22] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, “FaceForensics: A Large-Scale Video Dataset for Forgery Detection in Human Faces,” arXiv preprint arXiv:1803.09179, 2018.

[23] R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization,” in Proc. IEEE International Conference on Computer Vision, 2017.

[24] F. Matern, C. Riess, and M. Stamminger, “Exploiting Visual Artifacts to Expose DeepFakes and Face Manipulations,” in Proc. IEEE Winter Applications of Computer Vision Workshops, 2019.