International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
OPEN ACCESS
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST251222688

Deepfake Detection Using XceptionNet

Muskan Kumari, Sanya Gupta, Neha Singh
UG Scholar B. Tech, Indira Gandhi Delhi Technical University for Women, Delhi, India

ARTICLEINFO

Article History:

Accepted : 01 May 2025
Published: 04 May 2025

Publication Issue :
Volume 12, Issue 3
May-June-2025

Page Number :
55-63

ABSTRACT

The rapid rise of synthetic media, especially deepfakes, has sparked major concerns around misinformation, identity fraud, and diminishing public confidence in visual content. As these altered videos grow increasingly realistic, there is a pressing demand for reliable and scalable detection methods. This paper explores the use of the XceptionNet convolutional neural network architecture for deepfake detection. The analysis is based on the FaceForensics++ dataset, which comprises more than 1.8 million manipulated images created with four sophisticated face manipulation methods: NeuralTextures, FaceSwap, Face2Face, and DeepFakes. Cropped facial images are used for binary classification which is a process of differentiating between authentic and fraudulent content. Experimental results; with an accuracy of over 95% on unprocessed, and high-quality videos; over 80% accuracy even when heavily compressed; demonstrate that XceptionNet significantly outperforms both human observers and traditional detection methods, particularly under conditions of image compression. These findings highlight the robustness of deep learning-based models and the critical role of domain-specific preprocessing in improving detection accuracy.

I. INTRODUCTION

The manipulation of visual content, particularly facial imagery, has become increasingly widespread and raises significant concerns in today's digital society. One notable example is DeepFakes [1], which demonstrates how advances in computer graphics and visualization can be misused—for instance, replacing an individual's face with another to fabricate events or statements. Human faces are especially susceptible to such manipulation techniques due to two main factors: firstly, the domain of facial tracking and reconstruction is well-established within computer vision [11]; secondly, the human face is central to communication, often serving as a standalone medium for conveying emotions and intentions [18].

Facial manipulation techniques generally fall into two main categories: **expression manipulation** and **identity manipulation**. Expression manipulation involves altering facial movements. For example, Thies et al. introduced Face2Face [59], which allows real-time transfer of one individual's facial expressions to another using basic consumer hardware. Similar works such as "Synthesizing Obama" [55] take this further by generating realistic facial animations from audio inputs alone.

Identity manipulation, on the other hand, typically involves replacing one person's face with another's—a process known as face swapping. This has gained mainstream popularity due to the rise of user-friendly applications like Snapchat and has been enhanced further by deep learning techniques such as DeepFakes [1]. Unlike real-time face swapping using conventional graphics, DeepFakes require extensive training for each pair of subjects, making the process computationally intensive.

These methods often rely on image-based rendering and have been extended to other domains. For instance, they can be applied in virtual reality with eye-tracking and reenactment [60] or even extended to full-body motion transfer [61]. In the deep learning space, Kim et al. [39] proposed an image-to-image translation model to convert rendered face images into photorealistic outputs. An improved approach, *NeuralTextures* [57], jointly optimizes a learned texture and rendering network, achieving particularly sharp results in areas like the mouth compared to *Deep Video Portraits* [39].

Other advancements include audio-driven reenactment, such as the work by Suwajanakorn et al. [55], who learned to synthesize lip movements from audio, using blending techniques similar to *Face2Face*. Averbuch-Elor et al. [8] introduced *Bringing Portraits to Life*, a method that deforms still images to match a source actor's expressions using 2D warping.

Recently, deep learning has enabled a variety of face image synthesis methods. A detailed overview is presented by Lu et al. [47]. Generative Adversarial Networks (GANs) have been employed for tasks like face aging [7], view synthesis [34], and changing facial attributes such as skin tone [46]. Techniques such as Deep Feature Interpolation [62] and Fader Networks [43] demonstrate impressive results in modifying features like age, facial hair, or expressions. However, early GAN-based models were limited by low image resolutions. This has been significantly improved by Karras et al. [37], who introduced progressively growing GANs capable of generating high-resolution, photorealistic faces.

In this paper, we apply XceptionNet model to the FaceForensics++ dataset a large scale collection of manipulated facial videos. We preprocess input images by extracting and enlarging the facial region using a robust face tracking algorithm, ensuring the model focuses on relevant areas. We evaluate XceptionNet across various manipulation techniques and compression levels, comparing it against baseline models and human performance.

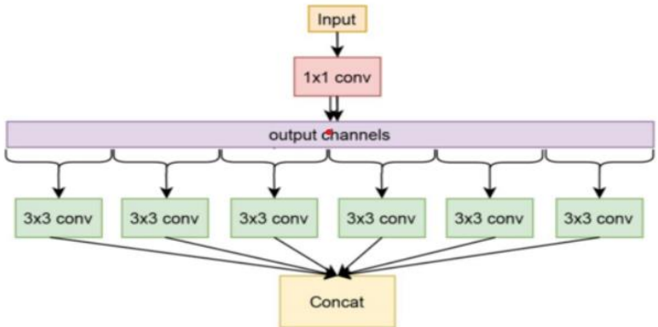

Fig 1: Block Diagram of XceptionNet Model

II. RELATED WORK

Traditional multimedia forensics and contemporary CNN-based classifiers have both been used in previous deepfake detection research. Statistical patterns are the foundation of hand-crafted feature techniques like steganalysis in conjunction with SVMs, however they suffer under compression. Convolutional neural networks (CNNs), as opposed to traditional models, have been used to detect facial manipulations in videos. Researchers like MesoNet, a compact CNN model, have shown greater robustness by automatically learning manipulation patterns from the data itself, while XceptionNet, which was first trained on ImageNet, provides a solid baseline that can be further refined using forensic datasets. To extract particular image features or modification types, the majority of forensic datasets were traditionally produced under extremely controlled conditions, which frequently required a significant amount of manual labor.

The following are the most significant linked papers:

Face Manipulation Method

Interest in virtual face alteration has increased dramatically during the last 20 years. Zollhöfer et al. provide a thorough overview of the field [68]. Video Rewrite by Bregler et al. [13], which presented an image-based technique for automatically producing lip movements to create fresh video footage, is one of the early milestones. One of the first automatic face-swapping techniques, Video Face Replacement, was developed by Dale et al. [20]. Their method warped the source face onto the target by reconstructing 3D models from monocular recordings.

Building on this foundation, Garrido et al. [29] developed a system that replaces an actor's face while retaining their original expressions. Similarly, VDub [30] utilized high-quality 3D face capturing to photo-realistically sync an actor's face with a dubber's voice. Thies et al. made major contributions with their real-time expression transfer methods. In [58], they used consumer-grade RGB-D cameras to track and transfer 3D facial deformations between individuals. Later, *Face2Face* [59] allowed real-time manipulation of facial expressions in online videos by blending modified 3D models into the original footage.

These methods often rely on image-based rendering and have been extended to other domains. For instance, they can be applied in virtual reality with eye-tracking and reenactment [60] or even extended to full-body motion transfer [61]. In the deep learning space, Kim et al. [39] proposed an image-to-image translation model to convert rendered face images into photorealistic outputs. An improved approach, *NeuralTextures* [57], jointly optimizes a learned texture and rendering network, achieving particularly sharp results in areas like the mouth compared to *Deep Video Portraits* [39].

Other advancements include audio-driven reenactment, such as the work by Suwajanakorn et al. [55], who learned to synthesize lip movements from audio, using blending techniques similar to *Face2Face*. Averbuch-Elor et al. [8] introduced *Bringing Portraits to Life*, a method that deforms still images to match a source actor's expressions using 2D warping.

Recently, deep learning has enabled a variety of face image synthesis methods. A detailed overview is presented by Lu et al. [47]. Generative Adversarial Networks (GANs) have been employed for tasks like face aging [7], view synthesis [34], and changing facial attributes such as skin tone [46]. Techniques such as Deep Feature Interpolation [62] and Fader Networks [43] demonstrate impressive results in modifying features like age, facial hair, or expressions. However, early GAN-based models were limited by low image resolutions. This has been significantly improved by Karras et al. [37], who introduced progressively growing GANs capable of generating high-resolution, photorealistic faces.

Multimedia Forensics

The goal of multimedia forensics is to verify the authenticity, origin, and integrity of images or videos without relying on embedded security mechanisms. Traditional approaches leveraged handcrafted features to detect statistical or physical inconsistencies during media formation [26, 53]. With the rise of deep learning, CNN-based methods—both supervised and unsupervised—have become prevalent in the field [9, 10, 12, 17, 35, 67].

In video forensics, the focus has been on detecting tampering operations that require minimal technical skill, such as frame duplication or deletion [31, 45, 63], interpolation artifacts [25], copy-move manipulations [11, 21], and chroma keying [48].

Particular attention has been given to face-specific forgeries. Techniques exist to identify computer-generated faces [15, 22, 51], morphing attacks [50], splicing [23, 24], and face swapping [38, 66]. Some detection approaches exploit visual inconsistencies from synthesis pipelines, like irregular eye blinking [44], or subtle differences in color, texture, and shape

[23, 24]. Others train deep networks to detect both low-level and semantic anomalies [5, 33, 38, 50, 66].

While many of these approaches show promising results, a key limitation is **robustness**. Common media operations—like compression, resizing, and re-encoding—can obscure forensic traces, making detection difficult in practical scenarios. These operations are especially common on social media platforms, where manipulated content is widely shared.

Forensic Analysis Datasets

Historically, most forensic datasets were created under highly controlled conditions, often requiring substantial manual effort to isolate specific image properties or manipulation types. While several datasets exist for image-level forgery detection, few adequately represent video-based manipulations. For instance, the MICC F2000 dataset [6] includes 700 forged images involving copy-move attacks, while the IEEE Forensics Challenge dataset contains 1,176 forged images. The Wild Web Dataset [64] and the Realistic Tampering dataset [42] offer small collections of real-world manipulations (90 and 220 images, respectively).

For facial forgeries, Zhou et al. [66] released a dataset of 2,010 images generated using FaceSwap and SwapMe. Korshunov and Marcel [41] contributed a collection of 620 DeepFake videos involving 43 subjects. The most extensive dataset to date was published by the National Institute of Standards and Technology (NIST), containing around 50,000 manipulated images and 500 forged videos [32].

In contrast to these efforts, our work introduces a **vast and diverse dataset** comprising more than **1.8 million images extracted from 4,000 manipulated videos**, surpassing prior datasets by an order of magnitude. This scale is essential for training robust deep learning models and evaluating their performance under realistic conditions—including varied resolutions and compression levels, as explored in Section 3 and further analyzed in Section 4. The availability of such a dataset supports the development of more effective and resilient facial forgery detectors suited for real-world forensic applications.

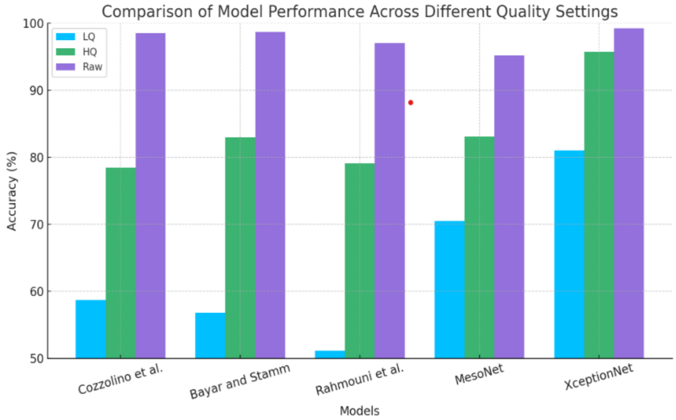

Fig 2: Comparison of model performance across different video quality settings (LQ: Low Quality, HQ: High Quality, Raw: Uncompressed). As observed, XceptionNet achieves the highest accuracy across all quality levels, with especially strong performance on uncompressed (Raw) videos.

Advancements in this field have been supported by datasets like DF-TIMIT and the DeepFake Detection Challenge, but FaceForensics++ remains a key benchmark due to its diverse manipulations, large scale, and realistic compression settings. The FaceForensics++ dataset includes both graphics-based and learning-based manipulation techniques, providing a comprehensive basis for model evaluation.

III.METHODOLOGY

We adopt XceptionNet for binary classification of facial images as real or fake. Our training pipeline uses face tracking to crop and enlarge facial regions by a factor of 1.3 to preserve contextual features. The model's final layer is replaced with a binary classifier, and the network is fine-tuned on the FaceForensics++ training set.

In order to simulate normal social media uploads, experiments are conducted under three distinct video quality levels: raw (uncompressed), HQ (lightly compressed), and LQ (heavily compressed). For each quality setting, we evaluate accuracy, precision, and robustness across various manipulation techniques.

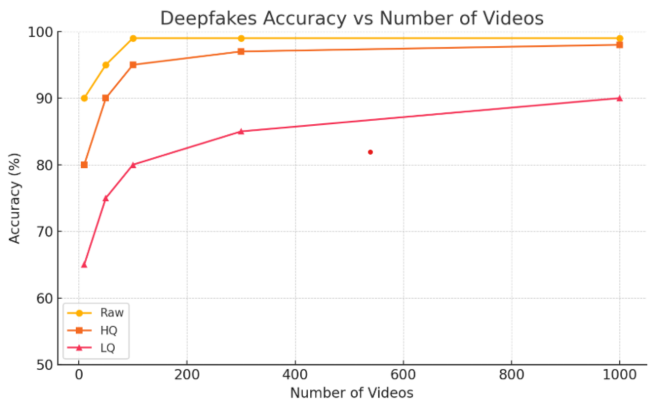

Fig 3: Performance evaluation of XceptionNet on Deepfake videos across different video quality settings: Raw (uncompressed), HQ (light compression), and LQ (heavy compression).

IV. RESULTS

With accuracy rates above 95%, XceptionNet performs exceptionally well in both raw and high definition videos. It outperforms shallow CNNs and steganalysis techniques, maintaining over 80% accuracy even under severe compression. The model also surpasses human detection capabilities, especially on NeuralTextures manipulations, which were notably difficult for human participants.

Our ablation study shows the importance of face-centric preprocessing; using the full image results in a performance drop. Additionally, increasing training data size improves detection in compressed scenarios, reinforcing the value of large-scale datasets.

V. CONCLUSIONS

The legitimacy of visual media is being threatened by the growing complexity of deepfake technologies. Using the XceptionNet architecture, which was refined on the large FaceForensics++ dataset, this work demonstrated an efficient method for deepfake detection.

Our method, which leverages face-centric preprocessing and binary classification, achieved over 95% accuracy on high-quality videos and maintained robustness even under heavy compression—conditions common in real-world scenarios. The experimental results highlight how important deep learning and domain-specific preparation are to surpassing both human perception and conventional forensics. Our ablation work also demonstrates how preprocessing and extensive training data can improve detection reliability to a great extent.

In the future, investigating ensemble models and including temporal characteristics from video sequences may provide even more advancements. However, this study establishes a solid basis for accurate and scalable deepfake detection systems, which are necessary for the verification of digital information.

REFERENCES

[1]. Deepfakes github. https://github.com/deepfakes/faceswap. Accessed: 2018-10-29. 1, 2, 4, 14

[2]. Darius Afchar, Vincent Nozick, Junichi Yamagishi, and Isao Echizen. Mesonet: a compact facial video forgery detection network. arXiv preprint arXiv:1809.00888, 2018. 3, 6, 7, 13, 14

[3]. Fakeapp. https://www.fakeapp.com/. Accessed: 2018-09-01. 4

[4]. Sami Abu-El-Haija, Nisarg Kothari, Joonseok Lee, Paul Natsev, George Toderici, Balakrishnan Varadarajan, and Sudheendra Vijayanarasimhan. YouTube-8m: A largescale video classification benchmark. arXiv preprint arXiv:1609.08675, 2016. 12

[5]. Faceswap. https://github.com/MarekKowalski/FaceSwap/. Accessed: 2018-10-29. 1, 2

[6]. Haiying Guan, Mark Kozak, Eric Robertson, Yooyoung Lee, Amy N. Yates, Andrew Delgado, Daniel Zhou, Timothee Kheyrkhah, Jeff Smith, and Jonathan Fiscus. Mfc datasets: Large-scale

benchmark datasets for media forensic challenge evaluation. In IEEE Winter Applications of Computer Vision Workshops, pages 63–72, Jan 2019. 3 2`21W

[7]. Grigory Antipov, Moez Baccouche, and Jean-Luc Dugelay. Face aging with conditional generative adversarial networks. In IEEE International Conference on Image Processing, 2017. 3

[8]. Hadar Averbuch-Elor, Daniel Cohen-Or, Johannes Kopf, and Michael F. Cohen. Bringing portraits to life. ACM Transactions on Graphics (Proceeding of SIGGRAPH Asia 2017), 36(4):to appear, 2017. 3

[9]. Jawadul H. Bappy, Amit K. Roy-Chowdhury, Jason Bunk, Lakshmanan Nataraj, and B.S. Manjunath. Exploiting spatial structure for localizing manipulated image regions. In IEEE International Conference on Computer Vision, pages 4970– 4979, 2017. 3

[10]. Belhassen Bayar and Matthew C. Stamm. A deep learning approach to universal image manipulation detection using a new convolutional layer. In ACM Workshop on Information Hiding and Multimedia Security, pages 5–10, 2016. 3, 6, 7, 13, 14

[11]. Michael Zollhofer, Justus Thies, Darek Bradley, Pablo ¨ Garrido, Thabo Beeler, Patrick Peerez, Marc Stamminger, ´ Matthias Nießner, and Christian Theobalt. State of the art on monocular 3d face reconstruction, tracking, and applications. Computer Graphics Forum, 37(2):523–550, 2018. 1, 2

[12]. Luca Bondi, Silvia Lameri, David Guera, Paolo Bestagini, ¨ Edward J. Delp, and Stefano Tubaro. Tampering Detection and Localization through Clustering of Camera-Based CNN Features. In IEEE Computer Vision and Pattern Recognition Workshops, 2017. 3

[13]. Christoph Bregler, Michele Covell, and Malcolm Slaney. Video rewrite: Driving visual speech with audio. In 24th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '97, pages 353–360, 1997. 2

[14]. Francois Chollet. Xception: Deep Learning with Depthwise Separable Convolutions. In IEEE Conference on Computer Vision and Pattern Recognition, 2017. 6, 7, 13, 14

[15]. Valentina Conotter, Ecaterina Bodnari, Giulia Boato, and Hany Farid. Physiologically-based detection of computer generated faces in video. In IEEE International Conference on Image Processing, pages 1–5, Oct 2014. 3

[16]. Davide Cozzolino, Diego Gragnaniello, and Luisa Verdoliva. Image forgery detection through residual-based local descriptors and block-matching. In IEEE International Conference on Image Processing, pages 5297–5301, October 2014. 6

[17]. Davide Cozzolino, Giovanni Poggi, and Luisa Verdoliva. Recasting residual-based local descriptors as convolutional neural networks: an application to image forgery detection. In ACM Workshop on Information Hiding and Multimedia Security, pages 1–6, 2017. 3, 6, 7, 13, 14

[18]. Chris Frith. Role of facial expressions in social interactions. Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1535), Dec. 2009. 1

[19]. Angela Dai, Angel X. Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. ScanNet: Richly-annotated 3D Reconstructions of Indoor Scenes. In IEEE Computer Vision and Pattern Recognition, 2017. 8

[20]. Kevin Dale, Kalyan Sunkavalli, Micah K. Johnson, Daniel Vlasic, Wojciech Matusik, and Hanspeter Pfister. Video face replacement. ACM Trans. Graph., 30(6):130:1–130:10, Dec. 2011. 2

[21]. Luca D'Amiano, Davide Cozzolino, Giovanni Poggi, and Luisa Verdoliva. A PatchMatch-

based Dense-field Algorithm for Video Copy-Move Detection and Localization. IEEE Transactions on Circuits and Systems for Video Technology, in press, 2018. 3

[22]. Duc-Tien Dang-Nguyen, Giulia Boato, and Francesco De Natale. Identify computer generated characters by analysing facial expressions variation. In IEEE International Workshop on Information Forensics and Security, pages 252–257, 2012. 3

[23]. Tiago de Carvalho, Fabio A. Faria, Helio Pedrini, Ricardo da S. Torres, and Anderson Rocha. Illuminant-Based Transformed Spaces for Image Forensics. IEEE Transactions on Information Forensics and Security, 11(4):720–733, 2016. 3

[24]. Tiago de Carvalho, Christian Riess, Elli Angelopoulou, Helio Pedrini, and Anderson Rocha. Exposing digital image forgeries by illumination color classification. IEEE Transactions on Information Forensics and Security, 8(7):1182– 1194, 2013. 3

[25]. Xiangling Ding, Gaobo Yang, Ran Li, Lebing Zhang, Yue Li, and Xingming Sun. Identification of Motion-Compensated Frame Rate Up-Conversion Based on Residual Signal. IEEE Transactions on Circuits and Systems for Video Technology, in press, 2017. 3

[26]. Hany Farid. Photo Forensics. The MIT Press, 2016. 3

[27]. Jessica Fridrich and Jan Kodovsky. Rich Models for Ste- ´ ganalysis of Digital Images. IEEE Transactions on Information Forensics and Security, 7(3):868–882, June 2012. 6, 7, 13

[28]. Davide Cozzolino, Justus Thies, Andreas Rossler, Chris- ¨ tian Riess, Matthias Nießner, and Luisa Verdoliva. ForensicTransfer: Weakly-supervised Domain Adaptation for Forgery Detection. arXiv preprint arXiv:1812.02510, 2018. 8

[29]. Pablo Garrido, Levi Valgaerts, Ole Rehmsen, Thorsten Thormaehlen, Patrick Perez, and Christian Theobalt. Automatic ´ face reenactment. In IEEE Conference on Computer Vision and Pattern Recognition, pages 4217–4224, 2014. 2

[30]. Pablo Garrido, Levi Valgaerts, Hamid Sarmadi, Ingmar Steiner, Kiran Varanasi, Patrick Perez, and Christian ´ Theobalt. Vdub: Modifying face video of actors for plausible visual alignment to a dubbed audio track. Computer Graphics Forum, 34(2):193–204, 2015. 2

[31]. A. Gironi, Marco Fontani, Tiziano Bianchi, Alessandro Piva, and Mauro Barni. A video forensic technique for detection frame deletion and insertion. In IEEE International Conference on Acoustics, Speech and Signal Processing, pages 6226–6230, 2014. 3 [

[32]. Irene Amerini, Lamberto Ballan, Roberto Caldelli, Alberto Del Bimbo, and Giuseppe Serra. A SIFT-based forensic method for copy-move attack detection and transformation recovery. IEEE Transactions on Information Forensics and Security, 6(3):1099–1110, Mar. 2011. 3

[33]. David Guera and Edward J. Delp. Deepfake video detection ¨ using recurrent neural networks. In IEEE International Conference on Advanced Video and Signal Based Surveillance, 2018. 3

[34]. Rui Huang, Shu Zhang, Tianyu Li, and Ran He. Beyond face rotation: Global and local perception GAN for photorealistic and identity preserving frontal view synthesis. In IEEE International Conference on Computer Vision, 2017. 3

[35]. Minyoung Huh, Andrew Liu, Andrew Owens, and Alexei A. Efros. Fighting fake news: Image splice detection via learned self-consistency. In European Conference on Computer Vision, 2018. 3

[36]. Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation

with conditional adversarial networks. CVPR, 2017. 5, 14

[37]. Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive Growing of GANs for Improved Quality, Stability, and Variation. In International Conference on Learning Representations, 2018. 3

[38]. Ali Khodabakhsh, Raghavendra Ramachandra, Kiran Raja, Pankaj Wasnik, and Christoph Busch. Fake face detection methods: Can they be generalized? In International Conference of the Biometrics Special Interest Group, 2018. 3

[39]. Hyeongwoo Kim, Pablo Garrido, Ayush Tewari, Weipeng Xu, Justus Thies, Matthias Nießner, Patrick Perez, Chris- ´ tian Richardt, Michael Zollhofer, and Christian Theobalt. ¨ Deep Video Portraits. ACM Transactions on Graphics 2018 (TOG), 2018. 3, 5

[40]. Davis E. King. Dlib-ml: A machine learning toolkit. Journal of Machine Learning Research, 10:1755–1758, 2009. 12

[41]. Pavel Korshunov and Sebastien Marcel. Deepfakes: a new threat to face recognition? assessment and detection. arXiv preprint arXiv:1812.08685, 2018. 3

[42]. Pawel Korus and Jiwu Huang. Multi-scale Analysis Strategies in PRNU-based Tampering Localization. IEEE Transactions on Information Forensics and Security, 12(4):809–824, Apr. 2017. 3

[43]. Guillaume Lample, Neil Zeghidour, Nicolas Usunier, Antoine Bordes, and Marc'Aurelio Ranzato Ludovic Denoyer. Fader networks: Manipulating images by sliding attributes. CoRR, abs/1706.00409, 2017. 3

[44]. Yuezun Li, Ming-Ching Chang, and Siwei Lyu. In Ictu Oculi: Exposing AI Created Fake Videos by Detecting Eye Blinking. In IEEE WIFS, 2018. 3

[45]. Chengjiang Long, Eric Smith, Arslan Basharat, and Anthony Hoogs. A C3D-based Convolutional Neural Network for Frame Dropping Detection in a Single Video Shot. In IEEE Computer Vision and Pattern Recognition Workshops, pages 1898–1906, 2017. 3

[46]. Yongyi Lu, Yu-Wing Tai, and Chi-Keung Tang. Conditional cyclegan for attribute guided face image generation. In European Conference on Computer Vision, 2018. 3

[47]. Zhihe Lu, Zhihang Li, Jie Cao, Ran He, and Zhenan Sun. Recent progress of face image synthesis. In IAPR Asian Conference on Pattern Recognition, 2017. 3

[48]. Patrick Mullan, Davide Cozzolino, Luisa Verdoliva, and Christian Riess. Residual-based forensic comparison of video sequences. In IEEE International Conference on Image Processing, 2017. 3

[49]. Patrick Perez, Michel Gangnet, and Andrew Blake. Pois- ´ son image editing. ACM Transactions on graphics (TOG), 22(3):313–318, 2003. 4, 14

[50]. Ramachandra Raghavendra, Kiran B. Raja, Sushma Venkatesh, and Christoph Busch. Transferable Deep-CNN features for detecting digital and print-scanned morphed face images. In IEEE Computer Vision and Pattern Recognition Workshops, 2017. 3

[51]. Nicolas Rahmouni, Vincent Nozick, Junichi Yamagishi, and Isao Echizen. Distinguishing computer graphics from natural images using convolution neural networks. In IEEE Workshop on Information Forensics and Security, pages 1–6, 2017. 3, 6, 7, 13, 14

[52]. Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Chris- ¨ tian Riess, Justus Thies, and Matthias Nießner. FaceForensics: A large-scale video dataset for forgery detection in human faces. arXiv, 2018. 3

[53]. Husrev T. Sencar and Nasir Memon. Digital Image Forensics — There is More to a Picture than Meets the Eye. Springer, 2013. 3

[54]. Wenzhe Shi, Jose Caballero, Ferenc Huszar, Johannes Totz, ´ Andrew P Aitken, Rob Bishop,

Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In IEEE Conference on Computer Vision and Pattern Recognition, pages 1874–1883, 2016. 14

[55]. Supasorn Suwajanakorn, Steven M. Seitz, and Ira Kemelmacher-Shlizerman. Synthesizing Obama: learning lip sync from audio. ACM Transactions on Graphics (TOG), 36(4), 2017. 1, 3

[56]. Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. 2017. 6

[57]. Justus Thies, Michael Zollhofer, and Matthias Nießner. De- ¨ ferred neural rendering: Image synthesis using neural textures. ACM Transactions on Graphics 2019 (TOG), 2019. 1, 2, 3, 4, 14

[58]. Justus Thies, Michael Zollhofer, Matthias Nießner, Levi Val- ¨ gaerts, Marc Stamminger, and Christian Theobalt. Real-time expression transfer for facial reenactment. ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH Asia 2015, 34(6):Art. No. 183, 2015. 2

[59]. Justus Thies, Michael Zollhofer, Marc Stamminger, Chris- ¨ tian Theobalt, and Matthias Nießner. Face2Face: Real-Time Face Capture and Reenactment of RGB Videos. In IEEE Conference on Computer Vision and Pattern Recognition, pages 2387–2395, June 2016. 1, 2, 3, 4, 5, 12

[60]. Justus Thies, Michael Zollhofer, Marc Stamminger, Chris- ¨ tian Theobalt, and Matthias Nießner. FaceVR: Real-Time Gaze-Aware Facial Reenactment in Virtual Reality. ACM Transactions on Graphics (TOG), 2018. 3

[61]. Justus Thies, Michael Zollhofer, Christian Theobalt, Marc ¨ Stamminger, and Matthias Nießner. Headon: Real-time reenactment of human portrait videos. arXiv preprint arXiv:1805.11729, 2018. 3

[62]. Paul Upchurch, Jacob Gardner, Geoff Pleiss, Robert Pless, Noah Snavely, Kavita Bala, and Kilian Weinberger. Deep feature interpolation for image content changes. In IEEE Conference on Computer Vision and Pattern Recognition, 2017. 3

[63]. Weihong Wang and Hany Farid. Exposing Digital Forgeries in Interlaced and Deinterlaced Video. IEEE Transactions on Information Forensics and Security, 2(3):438–449, 2007. 3

[64]. Markos Zampoglou, Symeon Papadopoulos, , and Yiannis Kompatsiaris. Detecting image splicing in the wild (Web). In IEEE International Conference on Multimedia & Expo Workshops (ICMEW), 2015. 3

[65]. Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters, 23(10):1499–1503, Oct 2016. 14

[66]. Peng Zhou, Xintong Han, Vlad I. Morariu, and Larry S. Davis. Two-stream neural networks for tampered face detection. In IEEE Computer Vision and Pattern Recognition Workshops, pages 1831–1839, 2017. 3

[67]. Peng Zhou, Xintong Han, Vlad I. Morariu, and Larry S. Davis. Learning rich features for image manipulation detection. In CVPR, 2018. 3

[68]. Paolo Bestagini, Simone Milani, Marco Tagliasacchi, and Stefano Tubaro. Local tampering detection in video sequences. In IEEE International Workshop on Multimedia Signal Processing, pages 488–493, October 2013. 3