Deepfake Detection Datasets

Subjects: Computer Science, Artificial Intelligence

Contributor: Liang Yu Gong , Xue Jun Li

Deepfakes are notorious for their unethical and malicious applications to achieve economic, political, and social reputation goals. Although deepfakes were initially associated with entertainment such as movie visual effects, camera filters, and digital avatars, they are defined as “believable generated media by Deep Neural Network” and have evolved into a mainstream tool for facial forgery. With the development of multiple forgery methods, deepfake data are increasing at a annual rate of ~300%. However, the data published online have different forgery qualities.

deepfake detection | deep learning methods | transformer | semi-supervised learning

1. Introduction

Although deepfakes were initially associated with entertainment such as movie visual effects, camera filters, and digital avatars [1], they are defined as “believable generated media by Deep Neural Network” and have evolved into a mainstream tool for facial forgery. Their illegal applications now pose serious threats to social stability, national security, and personal reputation [2]. Facial manipulation technologies started with 3D landmark face swap and auto-encoders [3] to generate fake media; however, the trend of deepfake generation nowadays involves more powerful generative models such as generative and adversarial networks (GANs) [4][5] and diffusion models (DMs) [6] for creating more realistic counterfeit media. As for the illegal application of this technique, one Reddit user first released generated pornographic videos of actress Gal Gadot as the protagonist of deepfakes, which caused a huge sensation and harmed the victim’s reputation at the end of 2017. In addition, Rana et al. [7] found that the top ten pornographic websites have released over 1790 deepfake videos to transfer celebrities’ faces to porn stars’ faces.

2. Deepfake Detection Datasets

Most online face forgery tools (such as DeepFaceLive [8] and Roop [9]) are open source and do not require sophisticated technical skills, so using open-source software such as Basic DeepFake maker [10] is the main method for creating deepfake datasets. Due to multiple forgery methods, deepfake data are increasing at a very high rate of approximately 300% every year [2], but the data published online have different forgery qualities. This section introduces several representative datasets and illustrates their advantages and disadvantages.

2.1. FaceForensics++

FaceForensics++ [11] is a pioneering large-scale dataset in the field of face manipulation detection. The main facial manipulations are representative, which include DeepFakes, Face2Face, FaceSwap, FaceShifter, and Neural Textures methods, and data are of random compression levels and sizes [12]. This database originates from YouTube videos with 1000 real videos and 4000 fake videos, the content of which contains 60% female videos and 40% male videos. In addition, there are three resolutions of videos: 480p (VGA), 720p (HD), and 1080p (FHD). As a pioneering dataset, it has different quality levels of data and equalized gender distributions. The deepfake algorithms include face alignment and Gauss–Newton optimization. However, this dataset suffers from low visual quality with high compression and visible boundaries of the fake mask. The main limitation of this dataset is the lack of advanced color-blending processing, resulting in some source facial colors being easily distinguishable from target facial colors. In addition, some target samples cannot effectively fit on the source faces because there exists facial landmark mismatch, which is shown in **Figure 2**.

Figure 2. Several FaceForensics++ samples. The manipulated methods are DeepFakes (Row 1), Face2Face (Row 2), FaceSwap (Row 3), and Neural Textures (Row 4). DeepFakes and FaceSwap methods usually create low-quality manipulated facial sequences with color, landmark, and boundary mismatch. Face2Face and Neural Textures methods can output slightly better-quality manipulated sequences but with different resolutions.

2.2. DFDC

From 2020 to 2023, Facebook, Microsoft, Amazon, and research institutions put efforts into this field and jointly launched a Deep Fake Detection Challenge (DFDC) [13] on Kaggle to solve the problem of deepfakes presenting realistic AI-generated videos of people performing illegal activities, with a strong impact on how people determine the legitimacy of online information. The DFDC dataset is currently the largest public facial forgery dataset, which contains 119,197 video clips of 10 s duration filmed by real actors. The manipulation data (See **Figure 3**) are

generated by deepfake, GAN-based, and non-learned techniques with resolutions ranging from 320 × 240 to 3840 × 2160 and frame rates from 15 fps to 30 fps. Compared with FaceForensics++, this database has a large-enough sample amount, different poses, and a rich diversity of human races. In addition, the original videos are from 66 paid actors instead of YouTube videos, and fake videos are generated with similar attributes to original real videos. However, the main drawback is that the quality level of data is different due to several deepfake generative abilities. Therefore, some samples have the problem of boundary mismatch, and source faces and target faces have different resolutions.

Figure 3. DFDC samples. Researchers manually utilized InsightFace facial detection model to extract human faces from the DFDC. Although some of the samples are without color blending and with obvious facial boundaries, the average quality is a little higher than the first-generation deepfake datasets.

2.3. Celeb-DF V2

Celeb-DF V2 is derived from 590 original YouTube celebrity videos and 5639 manipulated videos generated through FaceSwap [14] and DFaker as mainstream techniques. It consists of multiple age, race, and sex distributions with many visual improvements, making fake videos almost indistinguishable to the human eye [15]. The dataset exhibits a large variation of face sizes, orientations, and backgrounds. In addition, some post-processing work is added by increasing the high resolution of facial regions, applying color transfer algorithms and inaccurate face masks. However, the main limitation of this dataset is the low data amount with less sample

diversity because all original samples are downloaded from YouTube celebrity videos, and there is small ethnic diversity, especially for Asian faces. Here, a few samples of Celeb-DF V2 are presented (see **Figure 4**).

Figure 4. Celeb-DF V2 crop manipulated facial frames. Except for transgender and transracial fake samples (Row 3), it is hard to distinguish real and fake images with the human eye.

There are other higher-quality deepfake datasets created by extensive application of the GAN-based method; for example, DFFD [16], which was published in 2020, created an entire synthesis of faces by StyleGAN [17]. Comparing datasets published after 2020 with previous datasets, it can be observed the data amount is much larger with multiple forgery methods such as GAN and forgery tools. In addition, the original data sources are not limited to online videos such as YouTube and also consist of videos shot by real actors. Thus, researchers predict the trend of future DeepFakes datasets to be larger scale with various forgery methods, multiple shooting scenarios, and different human races. The advantages and disadvantages of several commonly used datasets are summarized in **Table 1**.

Table 1. The typical and commonly used datasets of facial forgery detection.

Datasets	Real/Fake	Data Source	Methods	Advantages	Limitations
UADF (2018) [18]	49/49	YouTube	FakeApp	Early release	Low data amount
FaceForensics++ (2019)	1000/5000	YouTube	FS, F2F, NT, DeepFakes, and FS	Multiple methods	Visible manipulated artifacts
DeepFake-Detection (2019)	363/3068	Actors	DeepFakes	Relatively good effects	Low data amount
Celeb-DF (2019)	590/5639	YouTube	Improved DeepFakes	Realistic manipulation	Less forgery methods
DFDC (2020)	19,197/1,000,000	Actors	DeepFakes and GAN	Various techniques	Different quality levels
DeeperForensics-1.0 (2020)	50,000/10,000	Actors	DeepFakes	Large-scale with different attributes	Less forgery methods
iFakeFaceDB (2020)	494,414/33,000	Previous dataset	GAN	Multi-scenarios	Unknown
DFFD (2020)	58,703/240,336	YouTube and previous datasets	GAN, DeepFakes, and FakeAPP	Large-scale and multi-techniques	Different quality levels
FFIW10K (2021)	12,000/10,000	YouTube	DeepFaceLab, FSGAN, and FS	Multi-face and scenarios	Unknown

FS: FaceSwap; F2F: Face2Face; NT: Neural Textures; FS: FaceShifter.

References

1. Abdulreda, A.S.; Obaid, A.J. A landscape view of deepfake techniques and detection methods. Int. J. Nonlinear Anal. Appl. 2022, 13, 745–755.
2. Zhang, L.; Lu, T. Overview of Facial Deepfake Video Detection Methods. J. Front. Comput. Sci. Technol. 2022, 17, 1–26.
3. FaceSwap-GAN. Available online: https://github.com/shaoanlu/faceswap-GAN (accessed on 15 December 2018).
4. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Wared-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Network. Proceeding Commun. ACM 2018, 63, 139–144.
5. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. In Proceedings of the Computer Vision and Pattern Recognition,

Boston, MA, USA, 7–12 June 2015.

6. Ho, J.; Jain, A.; Abbeel, P. Denoising Diffusion Probabilistic Models. In Proceedings of the 34th International Conference on Neural Information Processing System, Red Hook, NY, USA, 6–12 December 2020; pp. 6840–6851.
7. Rana, M.S.; Nobi, M.N.; Murali, B.; Sung, A.H. Deepfake Detection: A Systematic Literature Review. IEEE Access 2022, 10, 25494–25513.
8. DeepFaceLive. Available online: https://github.com/iperov/DeepFaceLive (accessed on 9 November 2023).
9. Roop. Available online: https://github.com/s0md3v/roop (accessed on 11 October 2023).
10. Li, Y.Z.; Yang, X.; Sun, P.; Qi, H.G.; Lyu, S. Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics. In Proceedings of the Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020.
11. Zhou, T.F.; Wang, W.G.; Liang, Z.Y.; Shen, J.B. Face Forensics in the Wild. In Proceedings of the Computer Vision and Pattern Recognition (CVPR), Virtual, 19–25 June 2021.
12. Guo, J.; Deng, J.; Lattas, A.; Zafeirioul, S. Sample and Computation Redistribution for Efficient Face Detection. In Proceedings of the Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020.
13. Kaggle. Available online: https://www.kaggle.com/c/deepfake-detection-challenge/overview (accessed on 12 December 2023).
14. FaceSwap. Available online: https://github.com/deepfakes/faceswap (accessed on 10 November 2020).
15. Tolosana, R.; Romero-Tapiador, S. DeepFakes Evolution: Analysis of Facial Regions and Fake Detection Performance. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020.
16. Dang, H.; Liu, F.; Stehouwer, J.; Liu, X.; Jain, A. On the Detection of Digital Face Manipulation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020.
17. Karras, T.; Laine, S.; Aila, A. A Style-Based Generator Architecture for Generative Adversarial Networks. In Proceedings of the Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019.
18. Li, Y.; Chang, M.C.; Lyu, S. In Ictu Oculi: Exposing AI Created Fake Videos by Detecting Eye Blinking. In Proceedings of the IEEE International Workshop on Information Forensics and Security, Hong Kong, China, 11–13 December 2018; pp. 1–7.