Dataset Details and References

ANAND M K
242CS008

Dataset Details

Table 1: Summary of Deepfake Datasets

Dataset	Description	Reference
DFDC	Largest and most diverse deepfake video dataset with over 100,000 clips. Includes multiple generation methods like **DFAE (Deepfake Autoencoder)**, **MM/NN (Morphable Mask/Neural Network)**, **NTH (Neural Talking Heads)**, and **FSGAN (Face Swapping GAN)**.	[1]
FaceForensics++	Standardized benchmark for facial manipulation detection. Contains 5,000 videos with real and manipulated content using FaceSwap, DeepFakes, Face2Face, and NeuralTextures.	[2]
DeeperForensics-1.0	Largest face forgery detection dataset with 60,000 videos. Uses **DF-VAE (DeepFake Variational Auto-Encoder)** for high-quality face swapping and includes diverse real-world perturbations.	[3]
DF-TIMIT	Created using GANs on VidTIMIT database. Contains 640 videos (low and high quality) with blending techniques like CNN-based segmentation and landmark alignment.	[4]
UADFV	Comprises 49 real and 49 Deepfake videos. Detection based on 3D head pose inconsistencies using DLib and OpenFace2.	[5]

References

References

[1] Brian Dolhansky, Joanna Bitton, Ben Pflaum, Jikuo Lu, Russ Howes, Menglin Wang, Cristian Canton Ferrer, *The DeepFake Detection Challenge (DFDC) Dataset*, 2020.

[2] Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, Matthias Niessner, *FaceForensics++: Learning to Detect Manipulated Facial Images*, 2019.

[3] 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.

[4] Pavel Korshunov and Sébastien Marcel, *DeepFakes: a New Threat to Face Recognition? Assessment and Detection*, 2018.

[5] Xin Yang, Yuezun Li, and Siwei Lyu, *Exposing Deep Fakes Using Inconsistent Head Poses*, 2018.