How Do Deepfakes Move?
Motion Magnification for Deepfake Source Detection
Supplemental Material

İlke Demir
Intel Labs
Santa Clara, CA
ilke.demir@intel.com

Umur Aybars Çiftçi
Binghamton University
Binghamton, NY
uciftci@binghamton.edu

A. Network Architecture

In Table 1 we document our C3D [3] based network architecture. Our input with $\omega = 16$ contains 16 consecutive frames and 4 channels for the dual representation.

Layer	BN3D	Act.	MaxPool3D	Output
Input	-	-	-	112x112x4x16
Conv3D	Yes	Relu	Yes	56x56x64x16
Conv3D	Yes	Relu	Yes	28x28x128x8
Conv3D	Yes	Relu	No	28x28x256x8
Conv3D	Yes	Relu	Yes	14x14x256x4
Conv3D	Yes	Relu	No	14x14x512x4
Conv3D	Yes	Relu	Yes	7x7x512x2
Conv3D	Yes	Relu	No	7x7x512x2
Conv3D	Yes	Relu	Yes	4x4x512x1
Flat	No	No	No	8192
Linear	No	Relu	No	4096
Linear	No	Relu	No	4096
Linear	No	No	No	2

Table 1. **Network Architecture.** Layer details and sizes are documented for the source detection network.

B. Source Detection vs. Deepfake Detection vs. Real Class Accuracy

We emphasize that, although these are related tasks, their construction is different. Source detection classifies videos into real or *several* fake classes, computing accuracy among all of these classes, including the accuracy on the real class. Deepfake detection classifies videos into real and fake classes as a binary choice, independent of the number of generators contributed to each fake class, computing accuracy by correctly classified videos over all videos, which can be the mean of real and fake class accuracies in case the dataset is balanced. Source detection is a harder task in terms of categorizing the data into different classes and deepfake detection is a harder task in case different generators create significantly different samples with varying realism. In Table 2, we list and compare these accuracies for FF [2] and FAVC [1] datasets. (These results are listed throughout the main text but not in this format).

Accuracy	Level	FF	FAVC
Source Det.	Sample	95.92	89.67
Source Det.	Video	97.77	94.03
Real Class	Sample	89.17	90.71
Real Class	Video	91.00	91.43
Deepfake Det.	Sample	92.37	95.10
Deepfake Det.	Video	94.80	95.12

Table 2. **Source vs. Deepfake Detection.** Sample/video accuracies for different detection tasks are compiled on FF and FAVC.

References

[1] Hasam Khalid, Shahroz Tariq, Minha Kim, and Simon S. Woo. Fakeavceleb: A novel audio-video multimodal deepfake dataset. *arXiv preprint arXiv:2108.05080*, 2021. 1

[2] Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Niessner. Faceforensics++: Learning to detect manipulated facial images. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019. 1

[3] Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri. Learning spatiotemporal features with 3d convolutional networks. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 4489–4497, 2015. 1