Deepfake Detection: Challenges and Solutions

Davide Alessandro Coccomini

ISTI-CNR, Pisa, Italy

Deepfakes can have a serious impact on the spread of fake news and on people's lives in general, becoming every day more dangerous. Moderation of online content and databases is vital to mitigate this phenomenon but the development of systems to distinguish between fake and genuine content comes with its own challenges: (a) The lack of generalization capabilities, due to the fact that most deepfake detection models are trained on a specific type of deepfake and struggle to detect deepfakes generated using different techniques. (b) When applying the deepfake detectors to the real world many peculiarities may occur; for example, the management of videos in which there are multiple people in the same scene or the recognition of the faces' movements towards or backwards the camera.

In the analysis work we conducted, we started focusing on the generalization problem, trying to understand whether a particular deep learning architecture was more capable of abstracting the concept of deepfake to such an extent that it could detect images or videos that had been manipulated even with novel techniques. In [2] and [5] we compared Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) of various kinds by putting them in a cross-forgery context revealing the superiority of the ViTs, which are less tied to the specific anomalies they see during training. After that, noting a scarcity of methods based on ViT and even more so those based on hybrid architectures, we developed our first real deepfake detector. In [1] we created a new architecture, combining an EfficientNet-B0 and Cross ViT, which we have named Convolutional Cross Vision Transformer. Thanks to the local-global attention mechanism within it and the exploitation of features extracted from the CNN, the model was able to effectively detect deepfake videos, achieving SOTA results on DFDC[6] and FaceForensics++[9] dataset, all while keeping the number of parameters low. The model was also used to participate in the competition presented in [7]. In [4] we designed a new type of Convolutional TimeSformer that take into account both the spatial position of faces in the frame and their temporal position in the video. It is also capable of managing multiple identities and being robust to face-size movements thanks to the introduction of a novel attention mechanism and positional embedding. Our method surpassed the SOTA on in-dataset tests on [8] and performed robustly in real-world situations. Future work will mainly focus on improving deepfake detectors in order to make them more robust to other real-world problems. We also want to make detectors capable of combining information also of a textual nature, context, and the reputation of the account disseminating it, to understand video veracity. Also, as we started doing in [3], we will work on the more generic problem of synthetic content detection.

SEBD 2023: 31st Symposium on Advanced Database System, July 02–05, 2023, Galzignano Terme, Padua, Italy
davidealessandro.coccomini@isti.cnr.it (D. A. Coccomini)
0000-0002-0755-6154 (D. A. Coccomini)

References

[1] Davide Alessandro Coccomini et al. "Combining EfficientNet and Vision Transformers for Video Deepfake Detection". In: *Image Analysis and Processing – ICIAP 2022*. Ed. by Stan Sclaroff et al. Cham: Springer International Publishing, 2022, pp. 219–229. ISBN: 978-3-031-06433-3.

[2] Davide Alessandro Coccomini et al. "Cross-Forgery Analysis of Vision Transformers and CNNs for Deepfake Image Detection". In: *Proceedings of the 1st International Workshop on Multimedia AI against Disinformation*. MAD '22. Newark, NJ, USA: Association for Computing Machinery, 2022, pp. 52–58. ISBN: 9781450392426. DOI: 10.1145/3512732.3533582. URL: https://doi.org/10.1145/3512732.3533582.

[3] Davide Alessandro Coccomini et al. *Detecting Images Generated by Diffusers*. 2023. DOI: 10.48550/ARXIV.2303.05275. URL: https://arxiv.org/abs/2303.05275.

[4] Davide Alessandro Coccomini et al. *MINTIME: Multi-Identity Size-Invariant Video Deepfake Detection*. 2022. DOI: 10.48550/ARXIV.2211.10996. URL: https://arxiv.org/abs/2211.10996.

[5] Davide Alessandro Coccomini et al. "On the Generalization of Deep Learning Models in Video Deepfake Detection". In: *Journal of Imaging* 9.5 (2023). ISSN: 2313-433X. URL: https://www.mdpi.com/2313-433X/9/5/89.

[6] Brian Dolhansky et al. "The deepfake detection challenge (dfdc) dataset". In: *arXiv preprint arXiv:2006.07397* (2020).

[7] Luca Guarnera et al. "The Face Deepfake Detection Challenge". In: *Journal of Imaging* 8.10 (2022). ISSN: 2313-433X. DOI: 10.3390/jimaging8100263. URL: https://www.mdpi.com/2313-433X/8/10/263.

[8] Yinan He et al. "ForgeryNet: A Versatile Benchmark for Comprehensive Forgery Analysis". In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2021, pp. 4358–4367. DOI: 10.1109/CVPR46437.2021.00434.

[9] Andreas Rossler et al. "Faceforensics++: Learning to detect manipulated facial images". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019, pp. 1–11.