https://doi.org/10.2352/EI.2023.35.4.MWSF-378

Pros and cons of comparing and combining hand-crafted and neural network based DeepFake detection based on eye blinking behavior

***Dennis Siegel*[1], *Stefan Seidlitz*[1], *Christian Kraetzer*[1], *Jana Dittmann*[1]**

[1] ***Otto-von-Guericke University, Magdeburg, Germany***

Abstract

Temporal feature spaces are a promising approach for DeepFake detection, since DeepFake synthesis is most often done on a frame-by-frame basis. With the existing and upcoming regulations on European level, the EU General Data Protection Regulation (EU GDPR) and Artificial Intelligence Act (AIA) in particular, data minimization and decision transparency are of concern also for such media forensic methods. In order to bring these aspects together, this paper utilizes two different algorithms both analyzing the eye blinking in the videos. The first one is implemented using deep learning to predict blinking behavior. It shows challenges of hyper-parameter tuning for the training of such a model. The second detector uses an existing hand-crafted approach to identify a suitable number of frames (i.e., video duration) required to reliably detect DeepFakes. Considering GDPR concerns, an optimal trade-off between detection performance and data minimization is found in the range of 35 to 40 seconds of video, giving a detection accuracy of 96.88% for the DeepFakes tested.

Introduction and Motivation

DeepFakes present a recent advancement in technology enabling manipulations in digital media that focus on the replacement of a face in a video by another face. They have a wide area of use cases and their intent is not always clear, as they may also have positive aspects that need to be considered [19]. In particular the usage as a privacy enhancement technique (PET) has to be named here [6]. Regardless of their use case, DeepFakes should be identifiable, to detect and prevent their misuse, which requires suitable detection approaches. In general, these can be categorized according to temporal and spatial methods. This division goes hand in hand with image or video DeepFakes. Spatial methods utilize image manipulation detection techniques. In contrast, temporal methods have stricter requirements of inputting a video and potentially higher computational costs. Their suitability is given due to flaws / restrictions in current DeepFake synthesis methods. This is due to the fact that most DeepFake synthesis methods are working frame by frame, creating temporal anomalies in video streams. [24]

In this paper the focus is on temporal methods. It contains the following contributions: First, the evaluation of a deep learning based eye blinking predictor. Second, the identification of medical concerns regarding blinking and development of privacy enhancement strategies. Third, the identification of suitable video duration thresholds for DeepFake detection using eye blinking.

State of the art in DeepFake detection

A wide variety of different approaches for DeepFake detection has been introduced in literature. Mirsky and Lee [24] categorize detection approaches based on spatial and temporal features. Furthermore, the approaches are divided by them into hand-crafted and deep features. A similar survey overview can be found in Nguyen et al. [26], where the separation is done based on image- and video-based techniques, without further splitting based on the used machine learning techniques. In Yu et al. [41] the separation is solely done for DeepFake videos. Again, the categories are similar, including approaches for both spatial and temporal features. Although spatial approaches are also important (especially forensic approaches focusing on individual images), they are outside the scope of this paper. Instead, the following sections present selected approaches to DeepFake detection using hand-crafted and deep learning based temporal approaches.

DeepFake detection using hand-crafted feature spaces

In general, it is difficult to separate approaches based on the categories of 'hand-crafted' and 'deep learning'. There are various combinations of both modalities by introducing hand-crafted feature spaces, which are classified by deep learning [3, 7]. In terms of traditional machine learning classification, most hand-crafted detectors utilize support vector machines (SVM) [23, 39].

Agarwal et al. present DeepFake detection based on lip synchronization, by comparing the spoken word sounds (phonemes) with mouth movements in video (viseme) [3]. The evaluation is done both manually, by introducing a human operator labeling frames and automated using a convolutional neural network (CNN). In addition, the detection performance is evaluated based on video duration.

In [34] three hand-crafted detectors are proposed based on eye, mouth and the comparison of foreground and background to detect DeepFakes. While these detectors did not yield acceptable detection performances individually, a decision-level fusion increased the performance. In [19] both an hand-crafted and deep learning based feature extractor are used to detect DeepFakes based on inconsistencies in eye blinking behavior.

DeepFake detection using deep learning feature spaces

Established images based DeepFake detectors are by reason of the video compression not always applicable for video data, because video compression results in strong degradation within

the video frames [2]. Furthermore, most neural networks based detectors (e.g. [21], [25] or [32]) solely detect DeepFakes based on individual frames. In consequence, it is possible that contiguous frames of DeepFake videos results in inconsistencies between the frames which are in certain circumstances not visible by the human eye. In the area of neural networks those temporal artifacts are detectable by recurrent network structures. For example, Korshunov et al. [18] used a Long Short-Term Memory (LSTM) architecture to detect inconsistencies between the audio and video stream. For the audio stream they used Mel frequency cepstral coefficients (MFCC) and for the video stream they calculate 42 distances between mouth keypoints of the 68 landmarks dlib model [17]. Güera et al. [13] combine in a convolutional LSTM the spatial dimension using Convolutional Neural Networks (CNNs) and the temporal dimension using LSTM to analyze coherence inconsistencies between the frames. Another recurrent network structure is the Gated Recurrent Unit (GRU) network which is used by Sabir et al. [31]. They first cropped the frames to the facial area which are then compared by the GRU network to detect temporal discrepancies across the frames.

Motivated by the fact that the human blinking behavior was not or less present in first DeepFake videos, Li et al. [20] proposed a LSTM based blinking detector. They combined the LSTM layer with a convolutional layer to detect closed or opened eye states in the faces of all video frames. Newer DeepFake generation approaches solved the problem of missing blinking events within the video. The detector of Li et al., also known as In Ictu Oculi, is not able to differentiate between a real and a fake eye blink event. Only videos without blinking events allows the detector to classify those videos as fake. Further, the detector was not tested on DeepFakes which are not generated by the DeepFake tool used by its authors. An implementation of In Ictu Oculi is provided by its authors on GitHub[1] but this version only works as a blinking detector, not being able to differentiate between real and fake videos (it only returns open or closed eye states for the frames within a video with a probability between 0 and 1 but no indication of whether this implies a DeepFake or not).

Regulatory Requirements and their Impact to Feature Space Design

Additional requirements for AI applications (such as media forensics methods and frameworks) conditions are established by legislation at the European level. One such regulation was introduced with the EU General Data Protection Regulation (EU GDPR, [10]). It addresses general principles of data protection in terms of data collection and processing. In particular, the following three (out of seven) principles are of importance ([10]):

- Lawfulness, fairness and transparency: "*Processing must be lawful, fair, and transparent to the data subject*."
- Purpose limitation: "*You must process data for the legitimate purposes specified explicitly to the data subject when you collected it*."
- Data minimization: "*You should collect and process only as much data as absolutely necessary for the purposes specified*."

In addition, Article 9 of the GDPR states: "*Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited.*" [10]

Another regulation relevant in the context of this paper is the upcoming EU Artificial Intelligence Act (AIA) [11], addressing the usage of AI systems. One aspect of particular importance is the criterion of human oversight in using AI systems (Article 14). This is supposed to lead to a reduction of black-box algorithms and enforces human-in-the-loop and human-in-control aspects for AI systems. In addition, Article 52 Paragraph 3 of the current AIA draft states, that DeepFakes must be marked as such [11].

In this paper, as underlying forensic process model, the principles established in the best practice guidelines on IT forensics of the German BSI (German Federal Office for Information Security) [5] (German: "*Leitfaden IT-Forensik*") are used. This best practice document provides various means for modeling forensic processes, including the definition of a generic phase-driven investigation & reporting model, a basic data model and a classification of methods and tools. Like many other best practice documents in this field it covers basic investigation principles, process models, forensic data types, etc. but does not provide domain specific process models and guidelines for specific media forensic investigations such as DeepFake detection. Here, existing research, such as the latest extension to the BSI guidelines [5] described as the Data-Centric Examination Approach for Incident Response- and Forensics Process Modeling (DCEA) summarized in [16] and [35], is used as basis for extending the scope of these guidelines to achieve a higher degree of maturity for the state of the art in taylor-made models for media forensics (incl. DeepFake detection).

The core of DCEA has three main components: a model of the *phases* of a forensic process, a classification scheme for *forensic method classes* and *forensically relevant data types*. The six DCEA *phases* are briefly summarized as: Strategic preparation (SP), Operational preparation (OP), Data gathering (DG), Data investigation (DI), Data analysis (DA) and Documentation (DO). At this point only the importance of the SP has to be pointed out, since it is the phase that also includes all research and evaluation activities considered in this paper. For further details on the phase model as well as the method classes and data types, the reader is referred, e.g. to [16].

Privacy concerns in the evaluation of biometric data

The human face is an often used biometric trait, that besides the ID also reveals other information about the person. Even pictures of parts of the face allow to derive personal attributes like the gender, age, ethnical background, etc. as well as certain health issues [37]. The work presented in that paper indicates that it is possible to identify illnesses such as glaucoma and cataracts based even on single images. Furthermore, there are various studies addressing the aspect of spontaneous eye blinking. On average, a human blinks around 10 to 15 times a minute (i.e., once every 4 to 6 seconds [1]). In a study by Sforza et al. [33] it was identified,

[1] `https://github.com/yuezunli/WIFS2018_In_Ictu_Oculi`

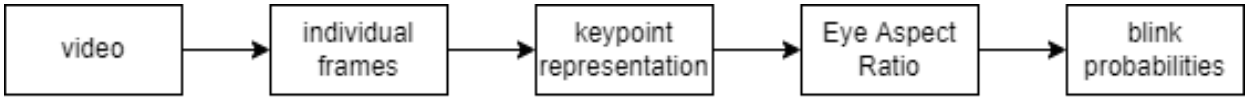

Figure 1. *Reduction steps taken for data acquisition*

that woman blink more frequently than men and it further differentiates based on age. In contrast, babies and children blink less frequent with around 2 times a minute. In addition, the blinking frequency can be affected by external influences, such as talking [36]. Another study by Jung et al. [15] identified a correlation between children frequently blinking and tic disorders.

Based on that, in conjunction with the previously discussed Article 9 of the GDPR, it is important to protect these personal attributes and prevent misuse or (unintended) information leakage. In general, there are three different possibilities to handle privacy concerns:

- all critical aspects are available to all
- features are overwritten by default parameters
- critical aspects are excluded, removed or overwritten

The first approach does not exclude personal attributes, instead it makes them available to all entities with access to the data set. This might require extensive labeling and also the agreement of the subjects of each sample. The privacy enhancement can be done either on feature or image level. On feature level, one possibility would be to overwrite features by default values. Relevant features have to be identified, that enable deriving personal attributes. As stated by Angwin et al. [4] personal attributes do not rely on individual features, but rather a correlation of multiple features. Lastly, critical aspects could be excluded, removed or overwritten. One possible approach for this is by using semantic image inpainting [40]. By now there are various existing privacy preserving methods, such as de-identification of facial images [8]. Othman and Ross [27] use morphing techniques to change the appearance of an face image. By using both a male and female image in the morph process they preserve the identity, but change the gender. Also DeepFake synthesis can be used for this purpose. In [6] its usage on social media is discussed, to anonymize faces in online media. For this purpose, the faces are replaced selectively based on the degree of acquantance, so to the user unknown faces are anonymized.

Although using image inpaiting or DeepFake to secure privacy in the video database seems most appropriate, it is not currently possible to use these techniques for the task of DeepFake detection. One reason for this is that the methods cause a change in the data and thus real training data might been changed by this method and then have to be regarded as DeepFake. To mitigate the downside for DeepFake detection, it is necessary to restore the original media of the synthesis. Based on a recent DeepFake challenge by Guarnera et al. [12], one question was to recreate the source image of DeepFake synthesis. Unfortunately no algorithms were submitted for this subtask.

In contrast to the possibilities discussed above, this paper presents an approach of information reduction based on a multi-level representation minimization. As shown in figure 1 a total of five different representations were considered for data extraction. Each reduction step also reduces the amount of information in the corresponding representation. So by changing from frames to keypoint representation for example, the requirement of storing the data as image is removed and replaced by keypoint graphs.

Development of a LSTM Network to predict blinking behavior

The development of the LSTM network based blinking predictor would occur within the strategical preparation (SP) phase of a forensic framework. The proposed forensic pipeline is illustrated in figure 2. The State-of-the-Art section above gives a small overview about existing LSTM approaches, but many more LSTM approaches exists. In consequence it is important to decide which approach is applicable for an eye blinking prediction which come with many different training iterations. A stacked LSTM network consisting of more than one LSTM layer seems the best strategy to train human eye blinking behavior.

For this paper, the training data for the LSTM network is the Celeb-real part of the Celeb-DF [22] data set. In preparation, the eye aspect ratio (EAR) for both eyes in each video is generated, according to the proposed method in [19]. All curves were normalized in the range of 0 and 1, calculated by the lowest and highest eye aspect ratio (EAR) value of all training samples. The prediction utilizes a sliding window approach, where two consecutive windows are taken, the first one for model training and the second for prediction. The window size is calculated as 5 seconds multiplied by 30 frames, which is the median frame rate of all Celeb-real videos. In other words the LSTM network was trained on 150 frames to create a prediction for the next 150 frames.

After the LSTM training the aim was to compare the predicted EAR curves with the calculated EAR curve from all videos of the Celeb-DF data set, divided into the three classes Celeb-real, YouTube-real and Celeb-synthesis. The difference between both the calculated and predicted curves is determined by $\sum_{x=s}^{n}(max(calc_x, pred_x) - min(calc_x, pred_x))/n$, with s being the index of the first predicted frame and n the total number of predicted frames. The calculated distance can then be used as feature for DeepFake detection.

Evaluation setup

As indicated above, for the training the Celeb-real part of the Celeb-DF [22] is used. The dlib face detector [17] analyzes all videos to detect faces in every frame of all included videos. In the training phase, a total of 56 videos had to be removed, because the face detection was not successful in several frames. Furthermore, the videos of Celeb-real do not have the same video length and some even had less than 300 frames. Due to the selected window size for the LSTM network of 300 frames, these videos were unusable. Additional 51 videos have been removed from the training data set because they were to short.

Addressing the hyper-parameter tuning for the LSTM network training, different training strategies were carried out. Different LSTM unit amounts were tested from ranging from 100 to 300, different counts of LSTM layers were tested from 2 to 4 layers and also dropout in different strengths from $p = 0.1$ to $p = 0.9$

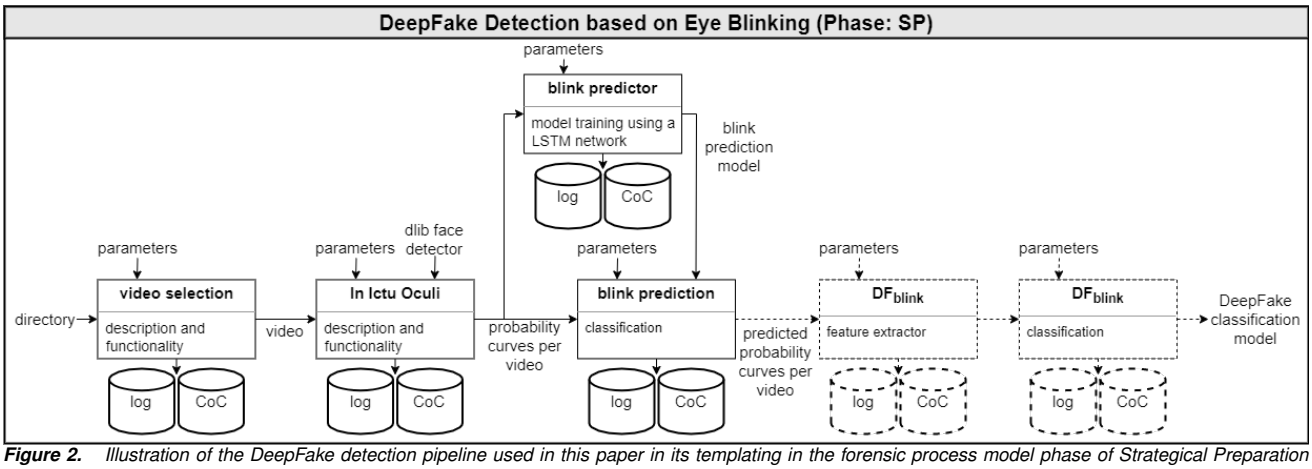

Figure 2. *Illustration of the DeepFake detection pipeline used in this paper in its templating in the forensic process model phase of Strategical Preparation (SP). Components outside the scope of this paper are marked by dashed lines.*

was inserted after every LSTM layer. The maximal training iterations was adjusted after full convergence of the training loss, which in most cases was after 500 epochs.

Evaluation results

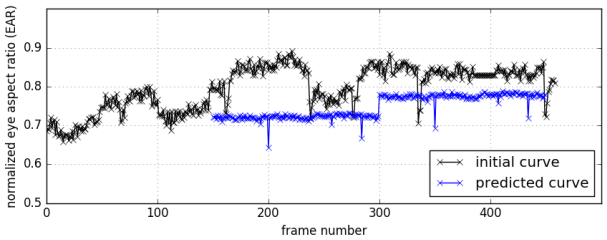

Figure 3. *Predicted eye aspect ratio and initial EAR curve on the example of the left eye for the video id13_0008 [22]*

Figure 3 shows an initial EAR curve calculated from a Celeb-real video of the Celeb-DF data set and the predicted blinking curve with estimated blinking events of a trained LSTM network. For the model a stacked LSTM network with two LSTM layers followed by a dropout layer with $p = 0.2$ and a final dense layer was trained for 500 epochs. The model by itself is not well trained, but a tendency of the predicted curve is visible and promising. The model predict approximately every 50 to 100 frames a blinking event, considering 30fps approximately every 1.66 to 3.33 seconds. Although the blinking appears to be slightly too often, it can be explained by the fact that the person in the video was talking (which results usually in a slightly increased blinking behavior). The sudden change in predicted values on frame 300 occurs because of a new segment, which is predicted with the real EAR data between frame number 150 and 300. The consequence of the results shown in figure 3 is the insight that further hyper-parameter tuning of the LSTM network is needed, which will also increase the computational cost that has to be invested into this detector in the strategical preparation phase. This highlights an important difference between hand-crafted and deep learning based approaches, namely the scope and depth of hyper-parameter tuning. At the current state of this blink predictor, further hyper-parameter tuning is required, to make the results more reliable.

The second evaluation goal is to identify a suitable video duration to detect DeepFakes based on eye blinking. For this purpose, the DeepFake detector DF_{eye} [34] is used on an in-house data set aggregating data from FaceForensics++ [29, 30], Celeb-DF [22], DFD [9] and HiFiFace [38] (2904 samples in total). The model is trained using the J48 [28] classification algorithm provided by WEKA [14] in its default parameterization and with 10-fold stratified cross-validation. Afterwards, the samples used are analyzed for the impact of the duration on the achieved accuracy. Due to the different frame rates, the optimal length is determined based on the video duration instead of the number of frames.

frames per second	15	18	24	25	29	29-30	30	60
# samples	6	2	413	852	5	19	1596	11

Framerate distribution in the considered data set.

Figure 4 shows the results categorized in 5 second video duration spans and the corresponding number of samples (=videos in the used set) per duration. The peak performance of 96./8% accuracy is achieved for video durations between 35 and 40 seconds (containing 96 samples in the used set). Longer samples first result in slight decrease in accuracy. A perfect classification is then again achieved for samples with an duration of at least 55 seconds, however the amount of samples of this duration (33) is too small to be relevant in the larger picture. The results obtained here suggest, that there is both a minimum (for accurate detection) and maximum (for privacy enhancement purposes) length for videos in the range of 35 to 40 seconds.

Summary, Conclusions and Future Work

This paper shows possibilities and challenges of deep learning approaches for the purpose of DeepFake detection, especially focusing on the relevance of suitable hyper-parameter tuning. The current state of the blinking predictor enables future work to extend the existing approach towards a full blown blinking-based DeepFake detector. This can be used to integrate both hand-crafted and neural network-based methods and evaluate and compare them against each other. Furthermore, the possibility to use both blinking probability curves generated by [20] as well as eye aspect ratios as baseline, allows to consider different representations of data. This enables the comparison of different training

duration (in s)	5-10	10-15	15-20	20-25	25-30	30-35	**35-40**	40-45	45-50	50-55	55-60	60-65	65-70	70-75
# samples	121	1179	707	343	158	107	**96**	49	57	18	17	8	7	1
accuracy (in %)	66.12	73.20	75.53	83.38	84.81	91.59	**96.88**	89.80	92.98	88.89	100	100	100	100

Evaluation results based on an inhouse data set for the detector DF_{eye}.

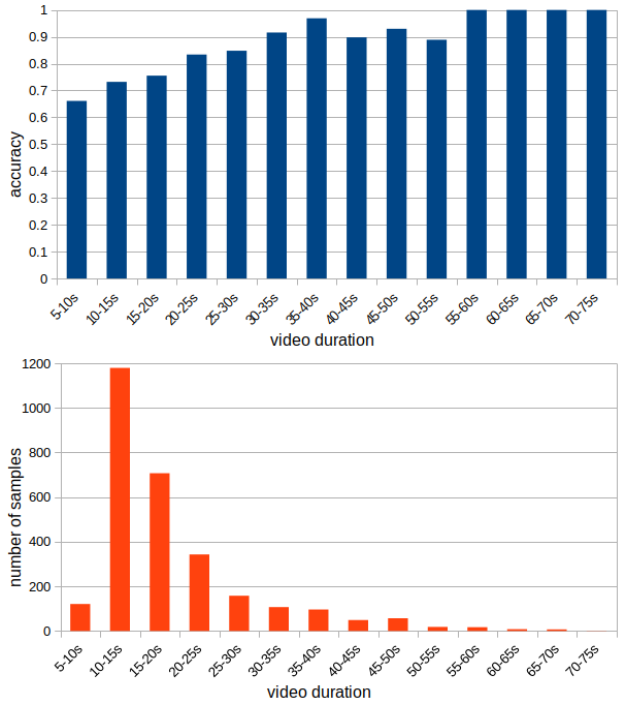

Figure 4. *Evaluation results based on an inhouse data set for the detector DF_{eye}.*

data representations, to evaluate the usage for privacy enhancement against the detection performance.

In addition to the work on blinking prediction, a video duration analysis based on this approach is possible. The experiments performed within this paper established an optimal minimum of 35 seconds and maximum of 40 seconds duration for this particular data set. In general, the human eye blinking and the evaluation itself is influenced by various external factors, such as distance of the person towards the camera and the fact that the person was talking in most samples used. Because of that, more training data is required to also increase the necessary diversity of training and testing material.

Acknowledgements

Funded in parts by the German Federal Ministry of Education and Research (BMBF) under grant number (FKZ): 13N15736 (project “Fake-ID”) and in particular for the privacy aspects on the example of medical concerns by the European Union Project “CyberSec LSA OVGU-AMSL” under grant number (FKZ): ZS/2018/12/96222.

Author Contributions: Initial idea & methodology: Jana Dittmann (JD), Stefan Seidlitz (StS) and Dennis Siegel (DS); Conceptualization: Stefan Seidlitz (StS), Dennis Siegel (DS) and Christian Kraetzer (CK); Empirical work on hand-crafted approaches: DS; Empirical work on deep learning driven approaches: StS; Writing – original draft: StS; Writing – review & editing: DS, CK and JD.

References

[1] F.H. Adler and R.A. Moses. *Adler’s Physiology of the Eye: Clinical Application*. C.V. Mosby Company, 1981.

[2] Darius Afchar, Vincent Nozick, Junichi Yamagishi, and Isao Echizen. MesoNet: a Compact Facial Video Forgery Detection Network. In *2018 IEEE International Workshop on Information Forensics and Security, WIFS 2018, Hong Kong, China, December 11-13, 2018*, pages 1–7. IEEE, 2018.

[3] Shruti Agarwal, Hany Farid, Ohad Fried, and Maneesh Agrawala. Detecting deep-fake videos from phoneme-viseme mismatches. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR Workshops 2020, Seattle, WA, USA, June 14-19, 2020*, pages 2814–2822. IEEE, 2020.

[4] Julia Angwin, Jeff Larson, Surya Mattu, Lauren Kirchner, and ProPublika. Machine Bias - ProPublica. `https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing`, May 23, 2016. Accessed: 23/06/2021.

[5] BSI. *Leitfaden IT-Forensik*. German Federal Office for Information Security, 2011.

[6] Umur A. Ciftci, Gokturk Yuksek, and Ilke Demir. My face my choice: Privacy enhancing deepfakes for social media anonymization. *CoRR*, abs/2211.01361, 2022.

[7] Umur Aybars Ciftci and Ilke Demir. FakeCatcher: Detection of Synthetic Portrait Videos using Biological Signals. January 2019.

[8] Benedikt Driessen and Markus Dürmuth. Achieving anonymity against major face recognition algorithms. In Bart De Decker, Jana Dittmann, Christian Kraetzer, and Claus Vielhauer, editors, *Communications and Multimedia Security*, pages 18–33, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[9] Nick Dufour and Andrew Gully. Contributing Data to Deepfake Detection Research. `https://ai.googleblog.com/2019/09/contributing-data-to-deepfake-detection.html`, September, 24 2019. Accessed: 09/09/2021.

[10] European Commission. REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). April, 27 2016. [Online]. Available at: `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02016R0679-20160504` [Last retrieved: 12.01.2023].

[11] European Commission. Proposal for a Regulation of the european parliament and of the council Laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts. *COM(2021) 206 final*, April, 21 2021. [Online]. Available at: `https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206` [Last retrieved: 14.09.2021].

[12] Luca Guarnera, Oliver Giudice, Francesco Guarnera, Alessandro Ortis, Giovanni Puglisi, Antonino Paratore, Linh MQ Bui, Marco

Fontani, Davide Alessandro Coccomini, Roberto Caldelli, et al. The face deepfake detection challenge. *Journal of Imaging*, 8(10):263, 2022.

[13] David Güera and Edward J. Delp. Deepfake video detection using recurrent neural networks. In *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–6, 2018.

[14] Mark A. Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The WEKA data mining software: an update. *SIGKDD Explor.*, 11(1):10–18, 2009.

[15] Hee-Yeon Jung, Sun-Ju Chung, and Jeong-Min Hwang. Tic disorders in children with frequent eye blinking. *J. AAPOS*, 8(2):171–174, April 2004.

[16] Stefan Kiltz. *Data-Centric Examination Approach (DCEA) for a qualitative determination of error, loss and uncertainty in digital and digitised forensics*. PhD thesis, Otto-von-Guericke-Universität Magdeburg, Fakultät für Informatik, 2020.

[17] Davis E. King. Dlib-ml: A machine learning toolkit. *J. Mach. Learn. Res.*, 10:1755–1758, 2009.

[18] Pavel Korshunov and Sébastien Marcel. Speaker inconsistency detection in tampered video. In *2018 26th European Signal Processing Conference (EUSIPCO)*, pages 2375–2379, 2018.

[19] Christian Kraetzer, Dennis Siegel, Stefan Seidlitz, and Jana Dittmann. Process-driven modelling of media forensic investigations-considerations on the example of deepfake detection. *Sensors*, 22(9), 2022.

[20] Yuezun Li, Ming-Ching Chang, and Siwei Lyu. In Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking. *CoRR*, abs/1806.02877, 2018.

[21] Yuezun Li and Siwei Lyu. Exposing DeepFake Videos By Detecting Face Warping Artifacts, 2019.

[22] Yuezun Li, Xin Yang, Pu Sun, Honggang Qi, and Siwei Lyu. Celeb-df: A large-scale challenging dataset for deepfake forensics. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, pages 3204–3213. IEEE, 2020.

[23] Scott McCloskey and Michael Albright. Detecting gan-generated imagery using color cues. *CoRR*, abs/1812.08247, 2018.

[24] Yisroel Mirsky and Wenke Lee. The creation and detection of deepfakes: A survey. *ACM Comput. Surv.*, 54(1), January 2021.

[25] Huy H. Nguyen, Junichi Yamagishi, and Isao Echizen. Capsule-forensics: Using capsule networks to detect forged images and videos. *CoRR*, abs/1810.11215, 2018.

[26] Thanh Nguyen, Quoc Viet Hung Nguyen, Dung Tien Nguyen, Duc Thanh Nguyen, Thien Huynh-The, Saeid Nahavandi, Thanh Tam Nguyen, Quoc-Viet Pham, and Cuong M. Nguyen. Deep learning for deepfakes creation and detection: A survey. *Comput. Vis. Image Underst.*, 223:103525, 2022.

[27] Asem Othman and Arun Ross. Privacy of facial soft biometrics: Suppressing gender but retaining identity. pages 682–696, 09 2014.

[28] J. Ross Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.

[29] Andreas Rössler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner. FaceForensics: A large-scale video dataset for forgery detection in human faces. *arXiv*, 2018.

[30] Andreas Rössler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner. Faceforensics++: Learning to detect manipulated facial images. In *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019*, pages 1–11. IEEE, 2019.

[31] Ekraam Sabir, Jiaxin Cheng, Ayush Jaiswal, Wael AbdAlmageed, Iacopo Masi, and Prem Natarajan. Recurrent convolutional strategies for face manipulation detection in videos. *CoRR*, abs/1905.00582, 2019.

[32] Selim Seferbekov. Deepfake detection (dfdc) solution by @selimsef, Jun 2029.

[33] Chiarella Sforza, Mario Rango, Domenico Galante, Nereo Bresolin, and Virgilio Ferrario. Spontaneous blinking in healthy persons: An optoelectronic study of eyelid motion. *Ophthalmic & physiological optics : the journal of the British College of Ophthalmic Opticians (Optometrists)*, 28:345–53, 08 2008.

[34] Dennis Siegel, Christian Kraetzer, Stefan Seidlitz, and Jana Dittmann. Media forensics considerations on deepfake detection with hand-crafted features. *Journal of Imaging*, 7(7), 2021.

[35] Dennis Siegel, Christian Krätzer, Stefan Seidlitz, and Jana Dittmann. Forensic data model for artificial intelligence based media forensics-illustrated on the example of deepfake detection. *Electronic Imaging*, 34:1–6, 2022.

[36] The Healthline Editorial Team. Excessive eye blinking: Causes, diagnosis, treatment and more, Aug 2019. [Online]. Available at: `https://www.healthline.com/health/eye-health/eye-blinking` [Last retrieved: 12.01.2023].

[37] Mateusz Trokielewicz, Adam Czajka, and Piotr Maciejewicz. *Iris Recognition in Cases of Eye Pathology*. Springer Singapore, Singapore, 2019.

[38] Yuhan Wang, Xu Chen, Junwei Zhu, Wenqing Chu, Ying Tai, Chengjie Wang, Jilin Li, Yongjian Wu, Feiyue Huang, and Rongrong Ji. Hififace: 3d shape and semantic prior guided high fidelity face swapping. In Zhi-Hua Zhou, editor, *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 1136–1142. International Joint Conferences on Artificial Intelligence Organization, 8 2021. Main Track.

[39] Xin Yang, Yuezun Li, and Siwei Lyu. Exposing Deep Fakes Using Inconsistent Head Poses. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2019, Brighton, United Kingdom, May 12-17, 2019*, pages 8261–8265. IEEE, 2019.

[40] Raymond A. Yeh, Chen Chen, Teck-Yian Lim, Alexander G. Schwing, Mark Hasegawa-Johnson, and Minh N. Do. Semantic image inpainting with deep generative models. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 6882–6890. IEEE Computer Society, 2017.

[41] Peipeng Yu, Zhihua Xia, Jianwei Fei, and Yujiang Lu. A survey on deepfake video detection. *IET Biometrics*, n/a(n/a), 02 2021.

Author Biography

Jana Dittmann *is a Professor on multimedia and security at the University of Otto-von-Guericke University Magdeburg (OvGU). She is the leader of the Advanced Multimedia and Security Lab (AMSL) at OvGU, which is partner in national and international research projects and has a wide variety of well recognized publications in IT security.* ***Christian Kraetzer*** *is a post-doc researcher and* ***Dennis Siegel*** *as well as* ***Stefan Seidlitz*** *are PhD students at AMSL.*