SHIELD: A Specialized Dataset for Hybrid Blind Forensics of World Leaders

Qingran Lin, Xiang Li, Beilin Chu, Renying Wang, Xianhao Chen, Yuzhe Mao, Zhen Yang, Linna Zhou, and Weike You(✉)

Beijing University of Posts and Telecommunications, Beijing, China
{linqingran,xli,chuchad1998,wangry017,chenxianhao,
maoyuzhe,yangzhenyz,zhoulinna,ywk}@bupt.edu.cn

Abstract. The speech videos of public figures, such as movie celebrities and world leaders, have an extensive influence on the Internet. However, the authenticity of these videos is often difficult to ascertain. These videos may have been carefully imitated by comedians or manipulated using Deepfake methods, which creates significant obstacles for the video forensics of specific characters. Moreover, the vast amount of data on social networking platforms renders manual screening impractical. To specifically address this issue, we present SHIELD, which stands for **S**pecialized dataset for **H**ybrid bl**I**nd for**E**nsics of wor**L**d lea**D**ers. Unlike most previous public Deepfake datasets that only contain Deepfake samples, this dataset exquisitely includes a collection that can quickly test this issue, encompassing both impersonator and Deepfake videos. We provide a detailed dataset production process and conduct an elaborate experiment under the hybrid blind detection scenario. Our findings reveal the limitations of existing methods, demonstrate the potential of identity-based models, and illustrate the increased challenges posed by SHIELD.

Keywords: Video forensics · Deepfake detection · Dataset

1 Introduction

Compared to videos of ordinary people, videos of specific characters (such as world leaders and famous actors/actresses) may significantly impact political activities and public opinion. With the popularity and development of AI manipulation technology, Deepfake videos of specific characters are more likely to be maliciously exploited and lead to severe consequences [5,10]. Additionally, some specific characters have many impersonators who mimic their facial expressions and gestures for entertainment or other purposes [2,3]. An impersonation performance is costly for a given character due to its more realistic visual effects, but it may exert a more substantial deceptive impact.

Under the combined effect of the above two factors, the mass of online videos can become misleading for viewers. Therefore, in the realistic scenario, the video forensics

Q. Lin and X. Li—Equal contribution.

S. Goel and P. R. Nunes de Souza (Eds.): ICDF2C 2023, LNICST 571, pp. 79–91, 2024.
https://doi.org/10.1007/978-3-031-56583-0_6

task of specific characters can be regarded as a hybrid detection of both the Deepfake and impersonator videos (Fig. 1).

Fig. 1. SHIELD contains real and faked videos of four world leaders and provides a dataset dedicated to character-specific video forensics models. Shown from top to bottom are example frames of clips from the original, FaceSwap [36], Wav2Lip [29], FOMM [32], and impersonator videos.

Current methods for this task mainly focus on one kind of forgery data: the Deepfake detection methods (used to classify Deepfake videos) [6,13,25] or the face recognition methods (used to identify imitators) [1,4]. One mainstream idea of Deepfake detection pays attention to the inconsistencies between faces generated by Deepfake and natural faces [17,26,28], which can effectively detect deepfake videos and meet the expectations. As for face recognition models, they can obtain excellent results when identifying impersonators. However, in the hybrid blind detection scenario where the manipulation methods are unknown, this type of Deepfake detection models cannot achieve the ideal results on videos with real faces because they ignore the videos' identity information. As some Deepfake generation methods [29,32] do not change the identity information of the source video, face recognition models misjudge the samples. Therefore, these methods alone may not be effective for addressing the complexities of the hybrid blind detection scenario.

The absence of public dataset for studies on hybrid blind detection of specific characters prompted us to develop the dataset for specialized forensics of world leaders (SHIELD), which specifically addresses the challenges of detecting Deepfakes and impersonators in online videos featuring public figures with significant societal impact, such as celebrities and leaders. Differing from most public Deepfake datasets, SHIELD targets the protection of people with broad influence on society, such as movie

celebrities and leaders. The dataset is primarily used for fine-tuning the data on individual characters, and its lightweight scale makes it easier for the model to conduct experiments.

Our contributions are as follows: (1) We propose a specialized dataset for hybrid blind forensics of world leaders (SHIELD), which is a public dataset with impersonator and Deepfake videos, addressing video forensics techniques for specific characters. (2) We conduct a systematic experiment to evaluate the performance of proper methods for video forensics on the SHIELD and offer open issues for future research in the video forensics of specific individuals. (3) Our experimental results indicate that the identity-based model performs well on our dataset, thereby highlighting their potential in the hybrid blind detection task.

2 Related Works

The generation of Deepfake forensics methods requires specific supervised datasets, which gives rise to the need for suitable, well-crafted datasets. The early appearing datasets like UADFV [37] and DeepFakeTIMIT [22] dataset have laid the foundation for deepfake detection models. After that, many outstanding public Deepfake detection datasets have emerged in recent years [16,20,21,23,24,27,31,38], which have given a significant boost to the development of this field. Here are some representative datasets.

In 2019, the FaceForensics++ dataset [31] was presented by Rössler *et al*. Until now, it is still one of the most popular and influential Deepfake video datasets. It introduced an automated benchmark and a large-scale dataset where the original real dataset contains 1000 real videos and 509,914 images. FaceSwap [36], DeepFakes [36], Face2Face [35], and NeuralTextures [34] are used to generate Deepfake videos. Because of its huge scale and the diversity of manipulation methods, FF++ has become a baseline dataset for many Deepfake detections.

The DFDC dataset [16] was published in a competition and is widely used in numerous works. It is a large dataset with 48,190 videos and over 25TB of raw data, the videos were mostly filmed at the resolution of 1080p. The dataset contains videos of real-world scenarios, individuals were filmed indoors and outdoors. In addition to image tampering, audio swapping was performed on partial clips.

In 2020, Li *et al*. introduced the Celeb-DF dataset [27], which was specifically designed for deepfake detection and comprises 590 authentic videos and 5,639 Deepfake videos. The dataset includes videos of 59 celebrities delivering speeches, which were sourced from YouTube, as well as Deepfake videos generated using an enhanced face-swapping technique. Notably, the Celeb-DF dataset features significant improvements in visual quality, such as higher resolution and the application of color conversion. Despite these advancements, the dataset remains a difficult challenge for deepfake detection models, and it continues to serve as a benchmark for evaluating the efficacy of various detection techniques.

In 2019, Agarwal *et al*. presented the World Leaders dataset (WLDR) [8], which marks a significant advancement in the field of protecting specific individuals from both deepfake and impersonator videos. This dataset introduces a forensics technique that is customized for specific characters, utilizing biometric features unique to each individual. The WLDR dataset includes authentic videos of five U.S. political figures and their corresponding impersonator videos and deepfake videos generated by applying face-swap deepfake methods. After that, the clips were extracted by sliding a window across the segment five frames at a time. The proposed technique is tailored to address the challenges posed by deepfake and impersonator videos, thereby enabling the effective detection of these malicious videos.

Our study draws inspiration from the WLDR and is motivated by the need to develop a more "omnipotent" approach for detecting forged videos. Our lightweight dataset contains sufficient video hours with more comprehensive data, where every character has its corresponding authentic and forged videos. Specifically, we adopted a rigorous evaluation using both Deepfake detection methods and face recognition methods in a systematic manner. SHIELD utilized updated forgery methods to generate Deepfake videos, which enriched the amount of data. In order to enhance the diversity and complexity of our dataset, the 10-second clips are partitioned from the videos directly, making it more suitable for fine-tuning and evaluating the advanced forensics algorithms.

3 SHIELD Dataset

This section specifies the specific production process and information about the dataset.

SHIELD, proposed by us, is a dataset specified for protecting specific characters' videos, consisting of real and forgery data of four world leaders (Barak Obama, Donald Trump, Hillary Clinton, and Joe Biden). It is a lightweight and diverse dataset, where four leaders of different skin colors and genders are selected. The dataset contains 255 Deepfake videos of four leaders, 297 real videos, and 40 impersonator videos. The Deepfake videos are generated with three different and typical synthesis models. We introduce the forgery methods and detailed production process in the following subsections (Table 1).

Table 1. Specific information of SHIELD dataset.

Video Type		Leaders - video duration (hours)/count of clips			
		Barak Obama	Donald Trump	Hillary Clinton	Joe Biden
Authentic Videos		7.25 h/2,647	7.66 h/3,188	5.19 h/2,028	4.79 h/1,956
Deepfake Videos	FaceSwap	10.02 h/3,696	13.44 h/4,993	13.12 h/4,885	11.10 h/4,124
	Wav2Lip	4.08 h/1,205	2.69 h/807	4.15 h/1.247	2.87 h/860
	FOMM	3.34 h/1,209	3.35 h/1,216	3.35 h/1,212	3.34 h/1,213
Impersonator Videos		0.71 h/256	1.71 h/620	1.33 h/487	1.12 h/406

3.1 Real Data

We first gather the videos of four world leaders delivering speeches on official occasions, like press conferences, public debates, and television broadcasts. All the source videos were downloaded from public sources online, which are highly trusted. Besides, these videos are carefully selected in which the leaders constantly talk and face straight toward the camera most of the time. Lastly, qualified videos are cut into non-overlapping 10-second clips for standardization.

3.2 Forgery Data

Forgery data consists of videos forged by Deepfake methods and videos of impersonators. The details of data collection and processing are shown as follows.

Impersonator Videos. We collect and download the videos of impersonators from YouTube. In these videos, comedians imitate our target world leaders' physical appearance, facial expression, and postures. Unlike Deepfake videos, these videos contain genuine footage of real individuals whose expressions and movements exhibit natural features. Also, they do not contain any artifacts that may indicate a synthetic source.

Deepfake Videos. To generate manipulation videos with diversity, we adapted three state-of-the-art Deepfake techniques (Fig. 2).

FaceSwap. FaceSwap [36] is one of the most popular methods in graphical Deepfake forgery. It generates Deepfake videos by transferring the face area from a source video to a target one. The model first extracts the landmarks of the face area, encodes the source and target images into the same latent space, and the two decoders are used to reconstruct different facial features.

We employ the use of FaceSwap software to facilitate the automatic detection and extraction of the facial area belonging to a specific leader from the collected videos. This process involves the replacement of the identified leader's face with the faces of three other leaders. The resulting sequences of both the processed face images and the original videos serve as input data for the subsequent face-swapping operation.

Wav2Lip. Wav2Lip [29] is a way of tampering through audio-driven video. It manipulates the movements of one's mouth area according to the content of the source audio, resulting in the alteration of the original video content. The quality of the corresponding audio plays a vital role in this type of tampering method. In the meantime, audios from VOA were selected for forgery to avoid possible artifacts.

Therefore, two components are necessary for the Wav2Lip approach. In our implementation, we use the videos of world leaders giving speeches, while the audio comes from the TED-LIUM 3 dataset [19].

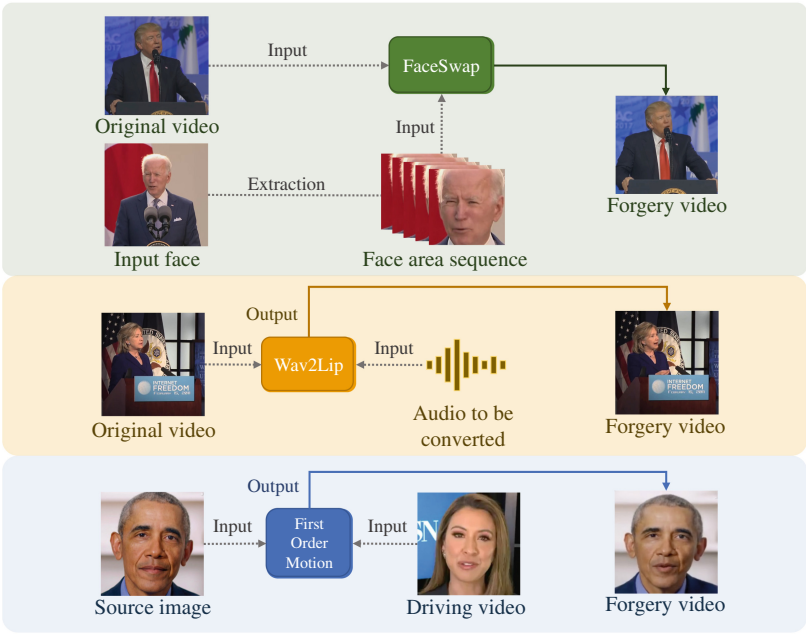

Fig. 2. Process illustration of the three deepfake methods.

FOMM. First order motion model [32] is a self-supervised network that transfers a series of motions from a driving video to a target image. Both the video and the image should contain objects of the same category. The model decouples appearance and motion by modeling the movement around the key points using affine transformations in a self-supervised manner. We use this method to represent the face-reenactment strategy.

A source image and a driving video are needed when generating a FOMM Deepfake video. In order to ensure the quality of the tampered video and reduce artifacts as much as possible, we chose the videos of the news anchor as driving videos, in which their head basically facing towards the camera with a slight angle of rotations during continuous talking.

3.3 Postprocessing

The postprocessing phase of our dataset creation comprises three key steps: segmentation, quality assurance, and audio track retention. First, all videos are sliced into non-overlapping 10-second segments to facilitate downstream usage. Next, to ensure the integrity and fidelity of the forged videos, manual quality checks are performed by qualified personnel. Videos deemed to be of low quality or containing significant artifacts are subsequently excluded from the dataset. Notably, as the FaceSwap and FOMM deepfake techniques do not involve audio manipulation, we opt to retain the original audio tracks of the source videos.

4 Evaluation

The objective of this performance evaluation is to demonstrate that although a relatively adequate amount of data is available for a particular world leader, the Deepfake detection models fail to accurately differentiate between genuine videos and those created by impersonators. Additionally, face recognition models are incapable of identifying the individuals appearing in the Deepfake videos. Thus, a hybrid blind detection approach is required to effectively identify the authenticity of these videos. In this section, we present our experimental setups, including detailed training procedures, and conduct an elaborate evaluation of various state-of-the-art Deepfake detection and face recognition models using our dataset. We analyze the obtained results and highlight their respective limitations.

4.1 Preprocessing of Evaluation

To implement frame-level evaluation, we first preprocess our dataset. We cut the video clips to be evaluated into frames and used RetinaFace [15] to extract the face area within them. Finally, we pass the crops of faces to the detection model as input. In order to make an independent analysis of different world leaders, we repeat the above steps and store their outputs separately.

4.2 Deepfake Detection Methods

With the development of Deepfake technology and the rise of comedic impersonator videos, ensuring massive online videos' authenticity has become challenging. To solve this task, most of the literature deals with Deepfake videos and focuses on various features of the video, such as learned features and identity-based features [12]. To achieve improvement, one category of works utilizes learned features in the video for detection, which involves constructing the novel network architecture [6], designing the data augmentation method [13], proposing new loss fuction [25], etc. The other category of works emphasizes the high-level semantics that exhibits impressive performance in terms of generalizability, such as biometrical features [7] and identity. Nowadays, some identity-based models have come to light. ID-Reveal [12] is a method that only requires real videos for training but exhibits high generalizability. In the detection method proposed by Dong *et al*. [17], photos of known identities are brought in as a reference set, which improves the detection effect to a certain extent. Bohácek *et al* [9]. constructed a model leveraging Zelenskyy's distinctive features of facial and gestural behaviors.

In order to ensure the diversity of evaluation, we choose different Deepfake detection models to evaluate our dataset, and the categories are as follows (Table 2).

Learned Features

XceptionNet. The XceptionNet [11] has been the baseline model by many well-known Deepfake datasets, such as FF++ [31], DF-TIMIT [22], Deepfake MNIST+ [20], etc. The model has achieved high accuracy in detecting deepfake videos: 99.26% on FF++

Table 2. Detection accuracy of selected models evaluated on SHIELD. Bolded are the best results on each forgery method of different world leaders.

Methods	Leaders	Evaluating Indicator (%)			
		Impersonator	FaceSwap	Wav2Lip	FOMM
XceptionNet-c40 [11] (w/ finetune)	Joe Biden	69.34	**97.68**	93.30	**97.12**
	Hillary Clinton	63.01	**98.82**	93.20	80.23
	Barak Obama	71.65	88.86	**92.19**	87.89
	Donald Trump	63.77	86.66	72.55	79.98
F^3-Net [30] (w/ finetune)	Joe Biden	69.14	92.60	88.84	92.99
	Hillary Clinton	69.63	94.73	**94.49**	90.71
	Barak Obama	64.43	**90.01**	70.18	73.34
	Donald Trump	67.30	87.33	78.29	69.04
EfficientNet-B3 [33] (w/ finetune)	Joe Biden	**77.44**	94.77	**94.27**	93.67
	Hillary Clinton	67.60	91.20	90.87	**93.75**
	Barak Obama	**72.92**	89.27	90.69	**88.28**
	Donald Trump	74.75	**87.50**	**87.33**	**86.40**
(a) Detection accuracy of Deepfake detectors based on learned features.					
ICT-Ref [17] (w/o finetune)	Joe Biden	65.47	91.77	51.01	61.77
	Hillary Clinton	**73.43**	93.61	60.50	64.25
	Barak Obama	63.22	88.35	66.78	53.91
	Donald Trump	**78.19**	87.26	50.22	68.16
(b) Detection accuracy of Deepfake detectors based on identity features					
Methods	Leaders	Evaluating Indicator (%)			
		Impersonator	FaceSwap	Wav2Lip	FOMM
ArcFace [14] + ResNet50 [18] (w/ finetune)	Joe Biden	84.68	69.37	41.36	64.14
	Hillary Clinton	88.17	59.03	58.49	59.49
	Barak Obama	87.50	57.56	33.50	45.42
	Donald Trump	69.58	66.50	54.95	56.53
(c) Detection accuracy of face recognition model					

(RAW), 99.91% on DF-TIMIT(HQ), and 92.38% on Deepfake MNIST+(Raw). It is an image-level method for Deepfake detection, which is a CNN model inspired by Inception. In our evaluation, we used the XceptionNet-c40 model pertained on ImageNet.

F^3-Net F^3-Net [30] is a novel face forgery detection method, which takes advantage of two frequency-aware clues, 1) frequency-aware decomposed image components and 2) local frequency statistics. Two clues are mixed in a two-stream collaborative learning framework, realizing Deepfake detection in the frequency domain.

EfficientNet. EfficientNet [33] is a convolutional neural network architecture and scaling method that uses a compound coefficient to scale all depth, width, and resolution

dimensions uniformly. Considering the promising results of the EfficientNet in DFDC public competition, we use EfficientNet-B3 as a convolutional extractor for processing the input faces.

Identity-Based Features

Identity Consistency Transformer. ICT [17] achieves the state-of-the-art performance over many benchmark datasets. It is a novel face forgery detection method that focuses on using high-level semantics: identity information. It detects suspicious faces by finding identity inconsistencies between the inner and outer face regions. Particularly, ICT-Ref leverages the real face available to build a reference set. Due to the addition of general identity information for enhancement and the availability of authentic videos, it is well suited for video forensics scenarios of world leaders.

4.3 Face Recognition Methods

Unlike most Deepfake datasets, SHIELD contains videos of impersonators of the four leaders. However, the actors imitate the leaders by their appearance and mannerisms. Rather than crafted employing Deepfake techniques, such videos are usually obtained by direct filming. Under this circumstance, facial recognition technology is selected to identify the face in the video. Accordingly, out of the above considerations, we add a face recognition model to evaluate our dataset.

Arcface. ArcFace [14] is a widely adopted loss function due to its easiness in implementation and state-of-the-art performance on a number of benchmarks. It improves the conventional softmax loss by optimizing the feature embedding on a hypersphere manifold where the learned face representation is more discriminative. We use ResNet-50 [18] pretrained on CASIA Webface as the backbone network.

4.4 Experimental Settings

There are four world leaders in SHIELD where leader $p \in \mathcal{P} = \{BarakObama, DonaldTrump, JoeBiden, HillaryClinton\}$. Videos in SHIELD can be divided into three categories: authentic videos $a_p \in \mathcal{A}_p$, Deepfake videos $f_p \in \mathcal{F}_p$, impersonator videos $i_p \in \mathcal{I}_p$. In that way, SHIELD dataset can be described as $\mathcal{D}_p = \{(v_p, l_p)\}$, where $v_p \in \mathcal{V}_p$ and $l_p \in \mathcal{L}_p = \{0, 1\}$. All the video samples in SHIELD can be formulated as $\mathcal{V}_p = \{\mathcal{A}_p, \mathcal{F}_p, \mathcal{I}_p\}$. In most of the Deepfake detection tasks, the positive samples are $\mathcal{A}_p$, while the negative samples are $\mathcal{F}_p$. However, in the hybrid blind detection task, $\mathcal{F}_p$ and $\mathcal{I}_p$ are negative samples of p. The label of real videos is $l_p = 0$, and the label of forged videos is $l_p = 1$. What we want is that there exist classification models f, which can map the video space to the class space: $f : \mathcal{V}_p \rightarrow \mathcal{L}_p$. To achieve this, the classification error of f can be minimized on training set $\mathcal{T}_p$:

$$\arg\min_{\theta} \mathbb{E}_{(v_p, l_p) \in \mathcal{T}_p^{l(f(v_p), l_p)}}, \tag{1}$$

where l is the loss function and θ is the parameters of model f for training.

We conduct two different comparison experiments on each leader separately. In the first experiment, we use the training set $\mathcal{T}_p = \{\mathcal{A}_p, \mathcal{F}_p\}$ of a specific leader p to train the detection models, which are then used to discriminate all the forgery videos $\{\mathcal{F}_p, \mathcal{I}_p\}$ of this very leader. We also use faces from $\mathcal{A}_p$ to train the face recognition model. Results of the first experiment are shown in Table 2.

In the second experiment, to see how Deepfake detection models perform when dealing with the type of forgery data outside the training set, we compared the detection results of $\mathcal{I}_p$ where detectors are respectively trained on $\mathcal{I}_p$ and $\{\mathcal{I}_p, \mathcal{F}_p\}$.

We construct our training set with balanced real and fake samples. All the models are initialized by their pretrained weights and trained using cross-entropy loss and Adam optimizer with batch size 64. We set the initial learning rate at 0.0001 and trained for 10,000 iterations. The model with the best test accuracy was chosen as the final model.

Table 3. Detection accuracy of detectors trained on impersonators videos $\mathcal{I}_p$ or mixed data $\{\mathcal{I}_p, \mathcal{F}_p\}$.

Methods	Evaluating Indicator (%)			
	Joe Biden	Hillary Clinton	Barak Obama	Donald Trump
XceptionNet-c40 [11]	81.74/**90.14**	87.02/**90.01**	76.22/**84.55**	76.94/**79.65**
F^3-Net [30]	**78.32**/76.07	69.50 / **82.74**	73.78/**80.56**	**71.03**/68.92
EfficientNet-B3 [33]	90.62/**92.87**	75.20/**80.03**	84.67/**88.99**	76.90 / **77.80**

4.5 Results and Analysis

The ultimate goal of SHIELD would be to help develop effective forensics techniques that take full advantage of the identity information in the hybrid blind detection task of face forgery.

To evaluate the performance of each detection method on four world leaders person-specifically, we employ frame-level accuracy rate as the performance metric. As presented in Table 2a, our experimental results demonstrate that all three models based on learned features exhibit satisfactory performance in detecting Deepfake videos. Notably, the detection accuracy rates are slightly higher with the FaceSwap method, ranging from 86.66% to 98.82%. This phenomenon could be attributed to the widespread use of the FaceSwap method in the pretraining dataset. Furthermore, a notable variance in detection accuracy is observed across the four leaders. In particular, the detection accuracy rate of Biden and Trump on Wav2Lip using the XceptionNet model demonstrates a 20.75% difference.

There is a noticeable drop in accuracy rate when using the Deepfake classifier to detect impersonator videos. After training on the data of specific characters, the DeepFake classifier cannot reach the ideal result. In contrast, the identity-based model, such as ICT-Ref shown in Table 2b, exhibits similar levels of accuracy in detecting impersonator videos as the other models, despite not having undergone fine-tuning on our dataset.

The notable difference in performance could potentially be attributed to the potential for further enhancement through additional identity information in the reference set.

In the case of detecting impersonator videos, the face recognition model produces satisfactory results as in Table 2c but is unable to accurately detect all three categories of Deepfake videos. While the aforementioned deepfake detection methods can detect some forged samples, there remains considerable room for improvement in their results.

As presented in Table 2, the performance of classifiers in detecting impersonator videos exhibits a noticeable improvement after adding Deepfake videos to the training data. This outcome suggests that the inclusion of identity information can enhance the model's ability to portray specific characters comprehensively and capture more features in the forensic task.

In conclusion, despite having adequate data on a specific world leader, the Deepfake detection model falls short in discriminating the videos of impersonators, while face recognition models cannot identify the person appearing in the Deepfake videos. The face recognition model can achieve satisfying results in detecting impersonator videos but cannot deal with hybrid blind scenarios in Deepfake detection. From the results of the identity-based model, the inclusion of identity information can improve the performance of classifiers when detecting impersonator videos, indicating that the identity information in Deepfake videos should not be ignored. The study suggests that incorporating identity information can enhance the model's ability to depict specific characters from multiple dimensions well-roundedly in the hybrid blind detection task of video forensics.

5 Conclusion

In this study, we introduce a compact hybrid dataset named SHIELD for video forensics of specific characters, which expands the current literature by addressing the challenging hybrid blind detection task. Our experimental results highlight the potential for further improvements in this area, particularly in leveraging identity information for enhanced classification performance. Furthermore, we emphasize the importance of considering the identity information present in deepfake videos when developing detection methods. While our dataset has some limitations, we believe that SHIELD can provide a valuable foundation for future research in this field.

Acknowledgments. Research was supported by the National Natural Science Foundation of China (No. 62172053), the National Key Research and Development Program of China (No. 2021YFC3340700, No. 2022YFC3303300, No. 2021YFC3340600, No. 2022YFC3300800), and the Fundamental Research Funds for the Central Universities (No. 2023RC30).

References

1. Amazon rekognition: Automate your image and video analysis with machine learning. https://aws.amazon.com/cn/rekognition/. Accessed 11 Oct 2022
2. Hillary Clinton impersonator teresa barnwell on vibe with sinbad. https://www.youtube.com/watch?v=r-lKbeOg0ro. Accessed 11 Oct 2022

3. Life as donald trump, hillary clinton impersonators. https://www.youtube.com/watch?v=bBl67bYAJm8&t=113s. Accessed 11 Oct 2022
4. Microsoft azure cognitive services: an AI service that analyzes content in images and video. https://azure.microsoft.com/en-us/products/cognitive-services/computer-vision/#overview. Accessed 11 Oct 2022
5. A nixon deepfake, a 'moon disaster' speech and an information ecosystem at risk. https://www.scientificamerican.com/article/a-nixon-deepfake-a-moon-disaster-speech-and-an-information-ecosystem-at-risk1/. Accessed 11 Oct 2022
6. Afchar, D., Nozick, V., Yamagishi, J., Echizen, I.: Mesonet: a compact facial video forgery detection network. In: WIFS, pp. 1–7. IEEE (2018)
7. Agarwal, S., Farid, H., El-Gaaly, T., Lim, S.: Detecting deep-fake videos from appearance and behavior. In: WIFS, pp. 1–6. IEEE (2020)
8. Agarwal, S., Farid, H., Gu, Y., He, M., Nagano, K., Li, H.: Protecting world leaders against deep fakes. In: CVPR Workshops, pp. 38–45. Computer Vision Foundation/IEEE (2019)
9. Bohácek, M., Farid, H.: Protecting president zelenskyy against deep fakes. CoRR abs/2206.12043 (2022)
10. Castillo, M.: Fake video news is coming, and this clip of Obama 'insulting' Trump shows how dangerous it could be. https://www.cnbc.com/2018/04/17/jordan-peele-buzzfeed-psa-edits-obama-saying-things-he-never-said.html. Accessed 11 Oct 2022
11. Chollet, F.: Xception: deep learning with depthwise separable convolutions. In: CVPR, pp. 1800–1807. IEEE Computer Society (2017)
12. Cozzolino, D., Rössler, A., Thies, J., Nießner, M., Verdoliva, L.: Id-reveal: identity-aware deepfake video detection. In: ICCV, pp. 15088–15097. IEEE (2021)
13. Das, S., Seferbekov, S.S., Datta, A., Islam, M.S., Amin, M.R.: Towards solving the deepfake problem : an analysis on improving deepfake detection using dynamic face augmentation. In: ICCVW, pp. 3769–3778. IEEE (2021)
14. Deng, J., Guo, J., Xue, N., Zafeiriou, S.: Arcface: additive angular margin loss for deep face recognition. In: CVPR, pp. 4690–4699. Computer Vision Foundation/IEEE (2019)
15. Deng, J., Guo, J., Zhou, Y., Yu, J., Kotsia, I., Zafeiriou, S.: Retinaface: Single-stage dense face localisation in the wild. CoRR abs/1905.00641 (2019)
16. Dolhansky, B., Howes, R., Pflaum, B., Baram, N., Canton-Ferrer, C.: The deepfake detection challenge (DFDC) preview dataset. CoRR abs/1910.08854 (2019)
17. Dong, X., et al.: Protecting celebrities from deepfake with identity consistency transformer. In: CVPR, pp. 9458–9468. IEEE (2022)
18. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR, pp. 770–778. IEEE Computer Society (2016)
19. Hernandez, F., Nguyen, V., Ghannay, S., Tomashenko, N., Estève, Y.: TED-LIUM 3: twice as much data and corpus repartition for experiments on speaker adaptation. In: Karpov, A., Jokisch, O., Potapova, R. (eds.) SPECOM 2018. LNCS (LNAI), vol. 11096, pp. 198–208. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-99579-3_21
20. Huang, J., Wang, X., Du, B., Du, P., Xu, C.: Deepfake MNIST+: a deepfake facial animation dataset. In: ICCVW, pp. 1973–1982. IEEE (2021)
21. Khalid, H., Tariq, S., Kim, M., Woo, S.S.: Fakeavceleb: a novel audio-video multimodal deepfake dataset. In: NeurIPS Datasets and Benchmarks (2021)
22. Korshunov, P., Marcel, S.: Deepfakes: a new threat to face recognition? assessment and detection. CoRR abs/1812.08685 (2018)
23. Kwon, P., You, J., Nam, G., Park, S., Chae, G.: Kodf: a large-scale korean deepfake detection dataset. In: ICCV, pp. 10724–10733. IEEE (2021)
24. Le, T., Nguyen, H.H., Yamagishi, J., Echizen, I.: Openforensics: large-scale challenging dataset for multi-face forgery detection and segmentation in-the-wild. In: ICCV, pp. 10097–10107. IEEE (2021)

25. Li, J., Xie, H., Li, J., Wang, Z., Zhang, Y.: Frequency-aware discriminative feature learning supervised by single-center loss for face forgery detection. In: CVPR, pp. 6458–6467, Computer Vision Foundation/IEEE (2021)
26. Li, Y., Chang, M., Lyu, S.: In ictu oculi: exposing AI generated fake face videos by detecting eye blinking. CoRR abs/1806.02877 (2018)
27. Li, Y., Yang, X., Sun, P., Qi, H., Lyu, S.: Celeb-df: a large-scale challenging dataset for deepfake forensics. In: CVPR, pp. 3204–3213. Computer Vision Foundation/IEEE (2020)
28. Matern, F., Riess, C., Stamminger, M.: Exploiting visual artifacts to expose deepfakes and face manipulations. In: WACV Workshops, pp. 83–92. IEEE (2019)
29. Prajwal, K.R., Mukhopadhyay, R., Namboodiri, V.P., Jawahar, C.V.: A lip sync expert is all you need for speech to lip generation in the wild. In: ACM Multimedia, pp. 484–492. ACM (2020)
30. Qian, Y., Yin, G., Sheng, L., Chen, Z., Shao, J.: Thinking in frequency: face forgery detection by mining frequency-aware clues. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12357, pp. 86–103. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58610-2_6
31. Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., Nießner, M.: Faceforensics++: learning to detect manipulated facial images. In: ICCV, pp. 1–11. IEEE (2019)
32. Siarohin, A., Lathuilière, S., Tulyakov, S., Ricci, E., Sebe, N.: First order motion model for image animation. In: NeurIPS, pp. 7135–7145 (2019)
33. Tan, M., Le, Q.V.: Efficientnet: rethinking model scaling for convolutional neural networks. In: ICML. Proceedings of Machine Learning Research, vol. 97, pp. 6105–6114. PMLR (2019)
34. Thies, J., Zollhöfer, M., Nießner, M.: Deferred neural rendering: image synthesis using neural textures. ACM Trans. Graph. **38**(4), 66:1–66:12 (2019)
35. Thies, J., Zollhöfer, M., Stamminger, M., Theobalt, C., Nießner, M.: Face2face: real-time face capture and reenactment of RGB videos. Commun. ACM **62**(1), 96–104 (2019)
36. torzdf: Faceswap. https://github.com/deepfakes/faceswap. Accessed 11 Oct 2022
37. Yang, X., Li, Y., Lyu, S.: Exposing deep fakes using inconsistent head poses. In: ICASSP, pp. 8261–8265. IEEE (2019)
38. Zi, B., Chang, M., Chen, J., Ma, X., Jiang, Y.: Wilddeepfake: a challenging real-world dataset for deepfake detection. In: ACM Multimedia, pp. 2382–2390. ACM (2020)