A. Brömme, C. Busch, A. Dantcheva, K. Raja, C. Rathgeb and A. Uhl (Eds.): BIOSIG 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2020 1

Eyebrow Recognition for Identifying Deepfake Videos

Hoang (Mark) Nguyen[1], Reza Derakhshani[2]

Abstract: Deepfake imagery that contains altered faces has become a threat to online content. Current anti-deepfake approaches usually do so by detecting image anomalies, such as visible artifacts or inconsistencies. However, with deepfake advances, these visual artifacts are becoming harder to detect. In this paper, we show that one can use biometric eyebrow matching as a tool to detect manipulated faces. Our method could provide an 0.88 AUC and 20.7% EER for deepfake detection when applied to the highest quality deepfake dataset, Celeb-DF.

Keywords: Deepfake detection, eyebrow biometrics, biometric recognition.

1 Introduction

In recent years, digital media is playing an exceedingly influential role in different aspects of our lives, including shaping public opinion. More and more people are getting their information from social networks and video-sharing platforms. Unfortunately, technology has also allowed images and videos to be manipulated by nefarious actors to show misinformation and discord. This issue has become a public concern threatening information trustworthiness and even undermining democracies [Ci19]. The tools for manipulating imagery, such as those used for political misinformation, have become widely available [VC20, Ag19].

The now-famous term ”deepfake” refers to recent (deep-learning-based) techniques used to synthesize or otherwise alter imagery, mostly faces in videos, which is also the focus of this paper. Due to rapid advances in computer vision and with increasingly affordable and capable hardware, convincing fake visual contents are being created and distributed at an alarming rate. Recently we have seen deepfake videos seeding misinformation by depicting public figures uttering words they had never said, among many other egregious and vulgar applications. As a result, deepfake detection is quickly becoming a high priority topic for the research community, the industry, and the governments alike.

Current anti-deepfake algorithms heavily rely on detecting image or video abnormalities such as visible artifacts or lack of coordination between lip movements and spoken words. Some examples of the aforesaid facial artifact are shown in Fig 1. During facial synthesis, many deepfake generators extract facial landmarks from the videos to manipulate the facial areas of interest. After manipulating the targeted facial features, a series of post-processing

[1] Department of Computer Science and Electrical Engineering, University of Missouri at Kansas City, hdnf39@mail.umk.edu

[2] Department of Computer Science and Electrical Engineering, University of Missouri at Kansas City, derakhshanir@umkc.edu

methods such as resolution-enhancement and color correction are applied to render the manipulated visualizations more realistic. Facial manipulation methods may be applied to the entire face or just the parts needed for the facial expressions [To20]. However, as the deepfake technologies improve over time, the deepfake visualization has become more realistic, and fewer artifacts are visible in the altered images and videos. REFACE app[3] is a face swap mobile application and has successfully integrated user face into high-quality music videos. Celeb-DF [Li20] is the highest quality deepfake forensic dataset that is publicly available. Fig 2 shows examples from the dataset. The current state-of-the-art deepfake detection approaches have not performed very well on this dataset, given its high quality (table 1). Thus, instead of relying on the hard to find visible artifacts for such datasets, we utilize a biometric recognition model to distinguish between real and fake images from the same identity.

In this paper, our focus is on detecting face swap by matching the components of the swapped face. More specifically, we show the efficacy of matching the eyebrow area to counter deepfake attacks. One may add other components like lower periocular to such a system. As the deepfake algorithms improve over time, one can expect the altered image artifacts to vanish, and thus biometric comparison of the swapped components may be preferable to flag counterfeit imagery.

The main contribution of this work is establishing the utility of eyebrows for deepfake detection by way of biometric comparison. To the best of our knowledge, this is the first time that biometric comparison of the eyebrow region is proposed for deepfake detection. The eyebrow region is one of the most affected components in the synthesized images. Especially in high-resolution and high-quality deepfakes, we show that eyebrow alterations become more distinguishable if examined by a biometric comparison pipeline. In order to make this approach to work, the model needs to know the participant's identity beforehand (biometric enrollment is needed). Moreover, this will be applicable when the targets are well-known individuals are celebrities or politicians.

Fig. 1: Examples facial artifacts in deepfake database

2 Prior Work

[MRD19] is one of the most recent works in eyebrow recognition. Mohammad et al. investigated short term eyebrow recognition using VISOB and FERET datasets. The au-

[3] https://reface.app/

Fig. 2: Examples from celeb-DF dataset: (a) real images, (b) deepfake images. Images in each column belong to the same identity

thors proposed a fusion of GIST, HOG, and VGG16 features along with Support Vector Machine (SVM) classifiers for biometric comparison. 0.63% Equal Error Rate (or EER, lower is better) and 0.9942 Area Under the Curve (or AUC, higher is better) was their best-reported results when fusing three feature descriptors for both eyebrows. However, their evaluates followed a closed set protocol where there are overlaps between training and testing set identities.

[MRS19] exploit visual artifacts in images to detect deepfakes. The authors proposed various facial areas where their model could spot potential artifacts caused by manipulating facial imagery. Some examples of such artifacts are global inconsistencies, illumination mismatches, geometrical distortions (such as those observed over the teeth), and eye color issues. Their best-reported results is 0.866 AUC using their in-house dataset.

[St19] proposed using an attention mechanism to detect manipulated face images. The attention map guides a CNN to scrutinize the face region in the image. The attention map mask helps the elimination of irrelevant features and thus reduces the feature vector dimensionality. Therefore, only certain sub-region in the vicinity of the face make significant contributions to the CNN's decision. The proposed approach reportedly achieves a 0.984 AUC in the UADFV dataset and 0.712 AUC over the Celeb-DF dataset.

Table 1 summarizes reported deepfake detection results over the Celeb-DF dataset. As mentioned earlier, to the best of our knowledge, this paper's proposed method is the first work using an eyebrow biometric pipeline to counter deepfake attacks. It is also noteworthy that we did not train our biometric model on any deepfake datasets, saving them for eventual testing to demonstrate cross-dataset generalization.

Tab. 1: performance of recent deepfake detection on Celeb-DF dataset.

Ref	Detection Method Used	Classifiers	Best AUC
Zhi et al. (2018) [Zh17]	Image-related Steganalysis	CNN+SVM	538
Afchar et al. (2018) [Af18]	Mesoscopic Level	CNN	0.548
Yang et al. (2018) [YLL19]	Head Pose Estimation	CNN	0.546
Li et al. (2019) [LL18]	Face Wrapping Artifact	CNN	0.569
Matern et al. (2019) [St19]	Visual Artifact	Losgistic Regression MLP	0.551
Stehouwer et al. (2019) [MRS19]	Facial Forgery	Attention Mapping	0.712

3 Methods

We employed four deep learning models to evaluate our hypothesis: LightCNN, Resnet, DenseNet, and SquezeNet. They are widely used in biometric research publication. Therefore, we believe that they would achieve high performance in eyebrow matching task.

LightCNN LightCNN [WHS15] model heavily relies on Max-Feature-Map (MFM) operation which was proposed in place of ReLu activation function. The operation preserves element-wise maximum from two feature maps forcing only half of the features to reach the next layer. In other words, this acts as a filter allowing the only compact feature to pass through.

ResNet Resnet [He16] employs a shortcut connections to deal with the gradient degradation problem[GB10]. Such an issue happens when training very deep neural networks. The residual or shortcut connections introduced in ResNet allows for identity mappings to propagate to multiple nonlinear layers, preconditioning the optimization during training. In this paper, we used ResNet-50 consists of 49 convolution layers and a single fully connected layer.

DenseNet The unit's dense block was first introduced in Dense Convolutional Network or DenseNet [Hu17]. In each block, there are multiple convolution layers where each layer is a concatenation of feature maps from previous layers. 1x1 convolutions are also utilized to reduce a large number of feature maps and the computation complexity. In this work, the DenseNet-121 model is utilized.

SqueezeNet Iandola1 et al. proposed SqueezeNet[Ia16], an efficient model, which is 50 times smaller than AlexNet but achieved the same level of accuracy. The model employs many strategies to decrease the number of parameters, such as small filter size, reduced input channels, and squeezed layers. Fire module was also introduced in the paper consisting of two layers: a squeeze layer consisting of 1×1 convolution filters, and expand layer, which is a mix of 1×1 and 3×3 convolution filters. The module does not decrease only the number of 3×3 filters but also the input channel.

Matching After obtaining our models' feature vectors, we used cosine distance metric to measure the similarity between reference and probe eyebrows. This is a famous match score employed by many deep-learning-based biometric systems.

4 Experimental Evaluation

Training Data VISible light mobile Ocular Biometric (VISOB) [Ra16] is a publicly available dataset consisting of eye images of about 550 healthy adults captured by three different mobile phones in three different lighting conditions. The three smartphones used in data collection are OPPO N1, iPhone 5s, and Galaxy Note 4. During the data collection, the volunteers were asked to take selfie-like images during two visits (Visit 1 and Visit 2), 2-4 weeks apart. During each visit, images were taken in two sessions 10-15 minutes apart, and under three different illumination conditions: regular office light, dim indoors, and natural daylight. In this experiment, we used a subset of VISOB captured under office lighting using the OPPO device, which offers a better resolution than iPhone and Note 4 captures, for our model training. Our model was trained on a high-resolution subset of VISOB to tell apart identities by way of eyebrow matching.

Testing Data Celeb-DF is the large, high quality deepfake forensic dataset. This dataset consists of 590 real videos from 59 celebrities along with 5639 deepfake videos. Since we are not after visual artifact caused by image synthesis, we evaluated our model on the best quality deepfake dataset that provides the most realistic fake video. This is a challenging task that nonetheless can better demonstrate the advantages of our proposed method. Unlike the other datasets, Celeb-DF contains almost no splicing boundaries, color mismatch, and inconsistencies of face orientation, among other visible deepfake artifacts. As a result, several deepfake detection papers have reported low accuracy numbers on this dataset. As shown in table 1, the current detection methods peak around 75% AUC on this dataset.

Data processing and training setup : We divided the VISOB dataset into 80% for training and 20% for validation. The eyebrow images were resized to different sizes depending on the corresponding deep learning models' input requirements. Multiple augmentations such as random rotations and random cropping, were applied to the training set. We trained our models with an initial learning rate of $1e^{-3}$ and reduced it by ten if the validation loss

did not drop by ten consecutive epochs. We trained our model for a maximum of 200 epochs and ended with the weights from the epoch that yielded the best validation loss. The momentum and weight decay parameters were set to 0.9 and $10e^{-4}$, respectively.

Experimental Setup We chose two experiments, short term and long term evaluation, to evaluate our hypothesis. All genuine matches in the former came from different frames in the same real video, while genuine matching was performed across the real videos in the latter. For each celebrity, one video out of ten videos was chosen to perform genuine matching in short-term evaluation. On the other hand, for the long term experiment, we used all the real videos for evaluation. For both experimental setups, all the deepfake videos were included to perform imposter matching with the real videos. For both the experiments, we extracted one frame from each deepfake video, and 20 frames from each real video (10 for enrollment and 10 for verification). The genuine match score is calculated between two images from the real video, and the imposter match score is calculated between a frame from the real video and another frame from deepfake video. We only perform matching between the original video and synthesized video from the same identities. These experiments are completely open-set that the participants in the training set are not from the identities used in the testing set. Further, the fake vs. real evaluations are conducted within the same quality and samples enjoy comparable resolutions regardless of their class label. We used ROC's Equal Error Rate (EER%) and Area Under the Curve (AUC) metrics to convey accuracies.

Tab. 2: EER and AUC for short term eyebrows identification in real and deep fake imagery

	Model	lightCNN	ResNet	DenseNet	SqueezeNet
Left	AUC	0.729	0.762	0.700	0.832
	EER	31.8%	29.5%	35.7%	25.3%
Right	AUC	0.696	0.879	0.690	0.802
	EER	35.4%	20.7%	37.6%	28.0%

5 Results and Discussions

Table 2 shows the EER and AUC for our short term evaluation. The best-achieved accuracy is 20.7% EER and 0.879 AUC using ResNet on the right eyebrow. For the left eyebrow, SqueezeNet performed the best with 25.0% EER and 0.832 AUC (its corresponding results for the right eyebrow were 28.0% EER and 0.802 AUC). The worst performer was DenseNet with 0.690 AUC and 37.6% EER (right eyebrow).

The accuracies for our long term evaluation are summarized in table 3. As expected, these results are worse than the short term's results with AUCs from 0.548 to 0.589 and EERs around 45.0%. This indicates that eyebrow matching is not the best choice for long term comparisons.

Tab. 3: EER and AUC for long term eyebrows identification in real and deep fake imagery

	Model	lightCNN	ResNet	DenseNet	SqueezeNet
Left	AUC	0.597	0.567	0.563	0.573
	EER	44%	45.3%	45.1%	45.3%
Right	AUC	0.589	0.580	0.548	0.561
	EER	43.3%	43.4%	46.6%	45.3%

6 Conclusion and Future Work

With the rapid developments in image synthesis, the creation of convincing deepfake videos has become easier and readily available to almost many. Since most of the deepfake detection methods rely on visible structural artifacts or color inconsistencies, they do not perform well on high-quality deepfake datasets such as Celeb-DF. In this work, we showed the efficacy of a new approach to expose deepfake images or videos using eyebrow matching. Instead of detecting the visible signs of facial manipulation, we used eyebrow match scores between real versus fake images from the same identity. Our best-achieved accuracy was 20.7% EER and 0.879 AUC on Celeb-DF, which is significantly better than other recently reported results on this high-quality deepfake dataset. However, we also noted that our approach did not fare as well over long term evaluations. Another limitation of our method is the requirement for the subject's identity so that the biometric eyebrow matching can proceed. As a part of future work, we would like to utilize the more feature-rich continuous eyebrow band region (simultaneously presenting both eyebrows) with our approach. Lastly, although our evaluations were made on a dataset different from the development set, we wish to perform additional cross-dataset deepfake evaluations to further test the generalization capability of the proposed framework.

7 Acknowledgement

This work was made possible in part by a gift from ZOLOZ. Dr. Derakhshani is also a consultant for the company.

References

[Af18] Afchar, Darius; Nozick, Vincent; Yamagishi, Junichi; Echizen, Isao: Mesonet: a compact facial video forgery detection network. In: 2018 IEEE International Workshop on Information Forensics and Security (WIFS). IEEE, pp. 1–7, 2018.

[Ag19] Agarwal, Shruti; Farid, Hany; Gu, Yuming; He, Mingming; Nagano, Koki; Li, Hao: Protecting world leaders against deep fakes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 38–45, 2019.

[Ci19] Citron, Danielle: , How DeepFake Undermine Truth and Threaten Democracy, 2019.

[GB10] Glorot, Xavier; Bengio, Yoshua: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics. pp. 249–256, 2010.

[He16] He, Kaiming; Zhang, Xiangyu; Ren, Shaoqing; Sun, Jian: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778, 2016.

[Hu17] Huang, Gao; Liu, Zhuang; Van Der Maaten, Laurens; Weinberger, Kilian Q: Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4700–4708, 2017.

[Ia16] Iandola, Forrest N; Han, Song; Moskewicz, Matthew W; Ashraf, Khalid; Dally, William J; Keutzer, Kurt: SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and¡ 0.5 MB model size. arXiv preprint arXiv:1602.07360, 2016.

[Li20] Li, Yuezun; Sun, Pu; Qi, Honggang; Lyu, Siwei: Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics. In: IEEE Conference on Computer Vision and Patten Recognition (CVPR). Seattle, WA, United States, 2020.

[LL18] Li, Yuezun; Lyu, Siwei: Exposing deepfake videos by detecting face warping artifacts. arXiv preprint arXiv:1811.00656, 2018.

[MRD19] Mohammad, A. S.; Rattani, A.; Derakhshani, R.: Eyebrows and eyeglasses as soft biometrics using deep learning. IET Biometrics, 8(6):378–390, 2019.

[MRS19] Matern, Falko; Riess, Christian; Stamminger, Marc: Exploiting visual artifacts to expose deepfakes and face manipulations. In: 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW). IEEE, pp. 83–92, 2019.

[Ra16] Rattani, A.; Derakhshani, R.; Saripalle, S. K.; Gottemukkula, V.: ICIP 2016 competition on mobile ocular biometric recognition. In: 2016 IEEE International Conference on Image Processing (ICIP). pp. 320–324, Sept 2016.

[St19] Stehouwer, Joel; Dang, Hao; Liu, Feng; Liu, Xiaoming; Jain, Anil: On the detection of digital face manipulation. arXiv preprint arXiv:1910.01717, 2019.

[To20] Tolosana, Ruben; Vera-Rodriguez, Ruben; Fierrez, Julian; Morales, Aythami; Ortega-Garcia, Javier: DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection. arXiv preprint arXiv:2001.00179, 2020.

[VC20] Vaccari, Cristian; Chadwick, Andrew: Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. Social Media+ Society, 6(1):2056305120903408, 2020.

[WHS15] Wu, Xiang; He, Ran; Sun, Zhenan: A Lightened CNN for Deep Face Representation. CoRR, abs/1511.02683, 2015.

[YLL19] Yang, Xin; Li, Yuezun; Lyu, Siwei: Exposing deep fakes using inconsistent head poses. In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 8261–8265, 2019.

[Zh17] Zhou, Peng; Han, Xintong; Morariu, Vlad I; Davis, Larry S: Two-stream neural networks for tampered face detection. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, pp. 1831–1839, 2017.