CHIEFS: Corneal-Specular Highlights Imaging for Enhancing Fake-Face Spotter

Muhammad Mohzary[1,2(✉)], Khalid Almalki[3], Baek-Young Choi[1], and Sejun Song[1]

[1] School of Science and Engineering, University of Missouri-Kansas City, Kansas City, MO, USA
{mm3qz,choiby,songsej}@umsystem.edu
[2] Department of Computer Science, Jazan University, Jazan, Saudi Arabia
mmohzary@jazanu.edu.sa
[3] College of Computing and Informatics, Saudi Electronic University, Riyadh, Saudi Arabia
k.almalki@seu.edu.sa

Abstract. This paper presents a novel Machine Learning (ML)-based DeepFake detection technology named CHIEFS (Corneal-Specular Highlights Imaging for Enhancing Fake-Face Spotter). We focus on the most reflective area of a human face, the eyes, upon the hypothesis that the existing DeepFake creation methods fail to coordinate their counterfeits with the reflective components. In addition to the traditional checking of the reflection shape similarity (RSS), we detect various corneal-specular highlights features, such as color components and textures, to find corneal-specular highlights consistency (CHC). Furthermore, we inspect the ensemble of the highlights with the surrounding environmental factors (SEF), including the light settings, directions, and strength. We designed and built them as modular features and have conducted extensive experiments with different combinations of the components using various input parameters and Deep Neural Network (DNN) architectures on Generative Adversarial Network (GAN)-based DeepFake datasets. The empirical results show that CHIEFS with three modules improves the accuracy from 86.05% (with the RSS alone) to 99.00% with the ResNet-50-V2 architecture.

Keywords: DeepFake · DeepFake Detection · Media Manipulation · Digital Media Forensics · Corneal-Specular Highlights

1 Introduction

The AI-fueled production and manipulation techniques of fictitious human facial images, DeepFake, have accomplished notable advancement. Due to the sophisticated DeepFake generation technologies [15,16,26], it is getting harder to distinguish the forged images by eye. Despite many benign applications such as fun memes, visual effects, and realistic avatars, the generated fake media can be

G.-V. Jourdan et al. (Eds.): FPS 2022, LNCS 13877, pp. 158–172, 2023.
https://doi.org/10.1007/978-3-031-30122-3_10

malignantly used by spreading misinformation on social media, creating deception for identity theft, and causing manipulation on election security. Hence, DeepFake has become a pandemic risk to authenticity, privacy, and security for our society. DeepFake detection technologies have become essential vaccines to mitigate the possible malignant risks.

There has been a large number of research works to detect DeepFakes. For example, [33] proposed an attention-based DeepFake detection distiller by applying frequency domain learning and optimal transport theory in knowledge distillation to improve the detection of low-quality DeepFake images. Le et al. [17] explored the asynchronous frequency spectra of color channels to train unsupervised and supervised learning models to identify GAN-based synthetic facial images. [31] extracted deep features from facial images using a Convolutional Neural Network (CNN). Another technique [19] checked eye blinking motions, which tended to be missing in DeepFake videos using the Long-Term Recurrent Convolutional Network (LRCN). Sun et al. [30] also detected DeepFake using facial geometric characteristics. However, previous methods lacked detection generalization on unseen data because they were trained on datasets containing few low-quality video frames generated with a single model and fewer subjects. In addition, eye-based DeepFake detection techniques in [7,9,19], and [22] only focused on a single artifact of eyes, either iris color, blinks, or similarity of corneal reflections on both eyes. Hence, they failed to detect sophisticated DeepFake media.

This paper presents a novel ML-based DeepFake detection technology named **CHIEFS** (**C**orneal-Specular **H**ighlights **I**maging for **E**nhancing **F**ake-Face **S**potter). As shown in Fig. 1, we focus on the most reflective area of a human face, eyes, upon the hypothesis that DeepFake technologies, such as replacement and synthesis, are hard to coordinate their counterfeits with the reflective components. We seek similarity and consistency of corneal-specular highlights (CSH) with multiple surrounding semantics, such as illumination and environmental conditions that are hard to forge. Thus, instead of checking a single aspect of the eyes, we extract multiple features, including *CSH*s' color components, shapes, and textures. In addition, we extract facial images surrounding environmental factors (SEF) to check the ensemble of the reflectance with the *SEF* such as indoor/outdoor, bright/dark, backgrounds, and light strength. CHIEFS embeds the *SEF* into the feature extraction and classification process to detect the symmetricity and consistency in both eyes' color components and reflection patterns.

As illustrated in Fig. 2, CHIEFS consists of a couple of ML components, including Training Data Collection and Annotation (TDCA), Highlights and Environmental Factors Detection (HEFD), and Feature Extraction, Embedding, and Classification (FEEC). The TDCA involves creating and annotating a new dataset named CHIEFS DeepFake Detection (CHIEFS-DFD). The CHIEFS-DFD dataset includes real and GAN-generated DeepFake facial images annotated with various *CSH* and environmental information. The HEFD detects right and left *CSH*, as well as identifies the *SEF* features. The FEEC extracts features

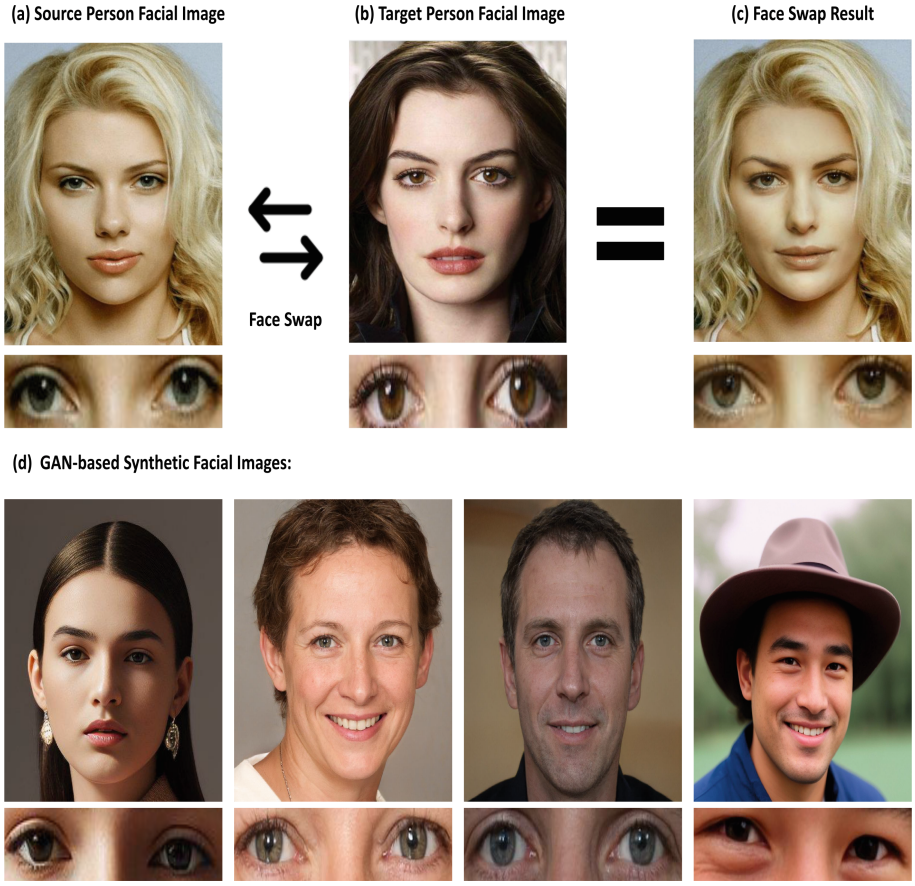

Fig. 1. Samples of Real and DeepFake Facial Images with their Reflective Elements. (a) and (b) are both Real, (c) is a DeepFake Face Generated Using the Face Swapper Online Tool [11], and Facial Images in (d) are GAN-based Synthetic Faces From [2,15].

from the *CSH* images, measures the right and left corneal highlights consistency (CHC), embeds additional *SEF* features, and classifies the input facial images as fake or real. We use Siamese Convolutional Neural Networks (SCNN) with various configurable neural network backbones, including ResNet-50-V2 [8], VGG-16 [29], Xception [3], and DenseNet-201 [10], for the feature extraction. We have conducted experiments with various GAN-generated DeepFake datasets to validate the accuracy of CHIEFS. The results show that CHIEFS achieves 99.00% accuracy in detecting highly realistic DeepFake facial images. Further, the modular design of CHIEFS renders itself as a complementary DeepFake detection module for any existing tools to limit the potential harm from DeepFake.

The main contributions of this work include:

- A new facial images dataset is collected and annotated for corneal reflection segmentation and DeepFake detection applications.
- A ML method is proposed to build an ensemble with various facial reflection features instead of a single feature.
- We study the impact of environmental factors on reflectance by collecting various parameters such as color and illumination conditions.
- We made modular designs for feature extraction and embedding to make it portable to other existing tools as a complementary solution module.

The remainder of this paper is organized as follows. Section 2 describes the existing DeepFake detection methods. Section 3 explains the design of CHIEFS. Section 4 discusses the experiment setups and results. Section 5 concludes the paper.

2 Related Work

This section discusses the current GAN-generated DeepFake detection methods and their limitations. Recently, several works have been proposed for DeepFake images detection. For instance, [21] presented a shallow learning method that fused spatial and spectrum features from an image to capture the up-sampling artifacts of DeepFake faces. [21] achieved 87% average accuracy on the FaceForensics++ dataset and AUC rates of 76.88% and 66.16% on the Celeb-DF and DFDC datasets, respectively. Mo et al. [23] proposed a CNN-based DeepFake images detection method that transformed the input image into residuals and fed the resulting residuals into three-layer groups where each group was composed of a convolutional layer with rectified linear activation function and a max-pooling layer. Next, the last group's output feature maps were aggregated and fed into two fully connected layers. Finally, the softmax layer was used to produce the output probability. The proposed method achieved 99.4% accuracy in detecting real facial images from CELEBA- HQ dataset [12], and DeepFake images from the fake face images database generated by [12]. Nguyen et al. [24] also developed a multi-task DeepFake images detection approach which performed classification and segmentation using an autoencoder model containing an encoder and a Y-shaped decoder. The activation of the encoded features was used for classification. The output of one branch of the decoder was used for segmentation, and the output of the second branch was used to reconstruct the input data. Their model achieved 92.60% average accuracy on the FaceForensics dataset and 68% average accuracy on the FaceForensics++ dataset.

Furthermore, several methods have exploited the eyes' visual features for DeepFake image detection. For example, [22] identified GAN-synthesised faces through the eyes' inconsistent iris colors or missing corneal specular reflections. However, such artifacts have been improved in the recent DeepFake generation models. Similarly, Hu et al. [9] also proposed a GAN-synthesized faces detection method that used the inconsistency of the corneal specular highlights between

the two synthesized eyes, assuming that two eyes looking at the same scene, their corneal specular highlights should show high similarities. This method can distinguish between the real and GAN-synthesized faces when light sources are visible to both eyes, and the eyes are distant from the light source. However, when these two conditions are defied, [9] will raise many false positives. [7] presented a DeepFake detection method based on irregular pupil shapes. This method can be effective on a specific dataset, but it will result in wrong predictions when the pupil shapes are non-elliptical in the real faces or there are occlusions on the pupil.

CHIEFS is designed to efficiently detect sophisticated DeepFakes using similarity and consistency of corneal-specular reflections with multiple surrounding semantics, such as illumination and environmental conditions, that are hard to counterfeit. It also coordinates various features (e.g., colors, edge, textures, etc.) of *CSH* images. It embeds surrounding environmental factors, such as indoor/outdoor, bright/dark, and light strength, and checks the ensemble with the reflectance.

3 CHIEFS Architecture

CHIEFS is an ML-based DeepFake detection technology that analyzes facial images' corneal-specular highlights consistency (CHC) and checks the ensemble of the highlights with multiple surrounding environmental factors (SEF). CHIEFS is designed in a hierarchical structure, and its components are separated into three modules. Training Data Collection and Annotation (TDCA), Highlights and Environmental Factors Detection (HEFD), and Feature Extraction, Embedding, and Classification (FEEC) modules in Fig. 2. The modular structure of CHIEFS allows agile updates of every module, like adding new features and enhancements according to specific use cases, as well as making CHIEFS available as a complementary DeepFake detection module for other existing tools.

3.1 Training Data Collection and Annotation (TDCA)

Current DeepFake detection datasets, such as UADFV [34], FaceForensics++ [27], Celeb-DF [20], and DFDC [5] do not contain the *CSH* annotation or facial image environmental factors information. Therefore, the main responsibility of ***the TDCA module in Fig. 2(a)*** is to create CHIEFS-DFD dataset [1] by collecting and annotating real and GAN-generated DeepFake facial images. We manually label the right and left *CSH* and provide the facial image-specific *SEF* information using the VGG Image Annotator (VIA) software [6]. The CHIEFS-DFD dataset contains 1,285 facial images in high resolution. 716 real facial images were collected from different datasets, including Flickr Faces HQ (FFHQ) dataset [14], Celeb-DF dataset, FaceForensics++ dataset, and DFDC dataset. Additionally, 569 GAN-generated DeepFake facial images were acquired

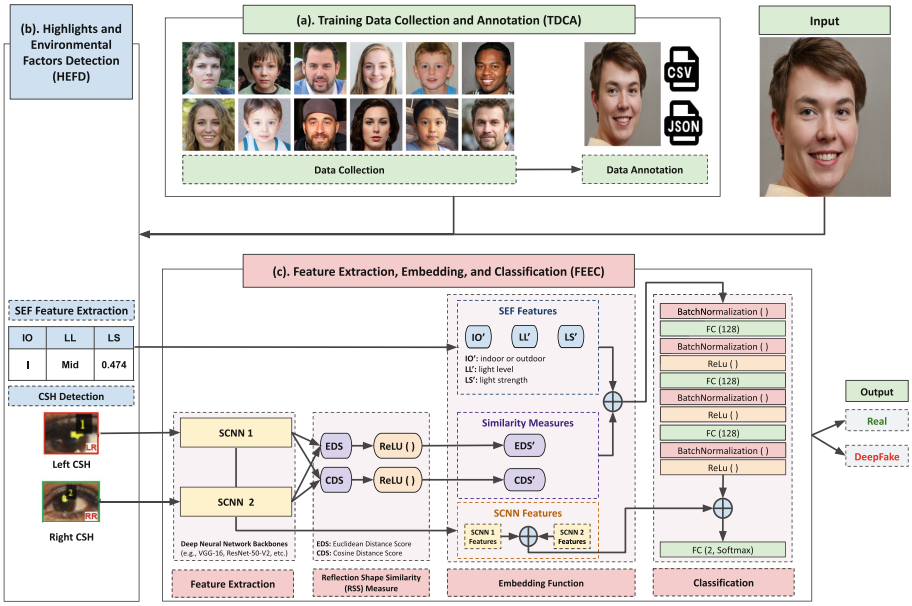

Fig. 2. The CHIEFS Architecture Block-diagram.

from various DeepFake detection datasets and human visual DeepFake generation tools, such as StyleGAN2 [15], StyleGAN3 [13], FSGAN [25], DeepFaceLab [26], and FaceShifter [18].

As illustrated in Fig. 3 (a), the CHIEFS-DFD dataset contains DeepFake and real facial images in high resolutions with different environmental parameters, including illumination conditions, background colors, indoor or outdoor settings, face pose orientations, age, ethnicity, and appearances (e.g., wearing makeup and accessories). As demonstrated in Fig. 3 (b) and Fig. 3 (c), the CHIEFS-DFD-dataset contains two types of annotations. The *CSH* region annotation in Fig. 3 (b) defines the shapes and locations of *CSH* and classifies them into right-reflection and left-reflection classes. The *Image Annotation* in Fig. 3 (c) identifies the image label (either Real or DeepFake), along with *SEF*, including indoor or outdoor (IO), light level (LL), and light strength (LS). The CHIEFS-DFD dataset contains the 2,570 annotated *CSH* segmentation masks for 1,285 facial images (two eyes per facial image). In addition, 959 images (74.63%) are labeled as indoor, and 362 images (28.17%) are labeled as outdoor. Furthermore, collecting and analyzing the distribution of CHIEFS-DFD dataset facial images' *LS* values (explained in Subsect. 3.2) results in different LL classes (806 mid images (62.72%), 258 low images (20.07%), and 221 high images (17.19%)).

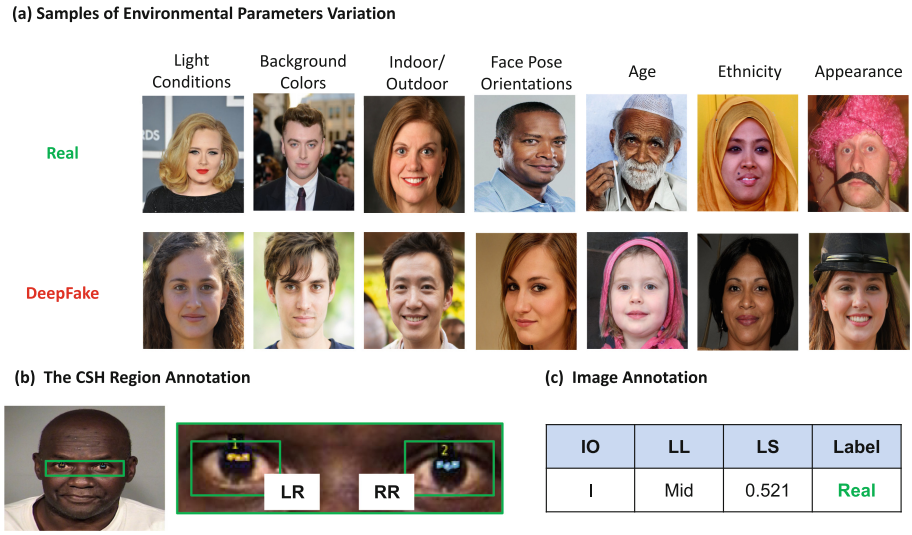

Fig. 3. Environmental Parameter Samples and Annotations in CHIEFS-DFD Dataset.

3.2 Highlights and Environmental Factors Detection (HEFD)

The HEFD module in Fig. 2(b) performs two major tasks, including *SEF* feature extraction and *CSH* detection. The *SEF* parameters include *IO*, *LS*, and *LL*. We train a MobileNet-V2 model on the Dense Indoor and Outdoor Depth (DIODE) dataset [32] and labeled facial images from the CHIEFS-DFD dataset (total 20,420 images) to classify the *IO* of an input image. To calculate the *LS*, we convert the input image's color space into a LAB format. The *L* channel is independent of color information in the LAB color space and only encodes intensity. The other two channels *A* and *B* encode color. Then, we extract the *L* channel and normalize it by dividing all pixel values by the maximum pixel value to have an *LS* value of the input image. Using the *LS* value, we identify an *LL* into the low, mid, and high classes (e.g., according to the *LS* distribution, the *LL* is a low if *LS* is less than 0.419, high if *LS* is greater than 0.637, and a mid if it is in between). To detect the right and left reflections, we train the *CSH* detection model using the MobileNetV2-SSDLite [28] to detect the bounding boxes of right and left *CSH* regions and class labels.

3.3 Feature Extraction, Embedding, and Classification (FEEC)

Using the right and left *CSH* images and the *SEF* extracted from the HEFD module Sect. (3.2), ***the FEEC module in Fig. 2(c)*** performs four primary functions, including deep hierarchical feature extraction using Siamese Convolutional Neural Network (SCNN) model with configurable neural network backbones, reflection shape similarity (RSS) measure, similarity measures (*RSS*), environmental factors (*SEF*), and CSH features embedding, and classification.

Feature Extraction: As shown in Fig. 2 (c), two SCNN models with the same weights and network architecture receive the right and left *CSH* images in parallel. Various configurable neural network backbones can be used for feature extraction, including VGG-16, Xception, ResNet-50-V2, and DenseNet-201. The two SCNN models use feedforwards to extract features using a global max-pooling layer by removing the fully-connected layer at the top of every network ($include_top$ = False). We do not need activation and classes because we only use the backbone models for feature extraction. Then, we use the right and left *CSH* features to measure *RSS* using euclidean and cosine distance scores.

Reflection Shape Similarity (RSS) Measure: *CSH* can be detected in various shapes, which can be deformed in different colors according to illumination conditions and blended into the background. Furthermore, *CSH* can be occluded by glasses, eyelids, or eyelashes, and only a tiny portion of the reflection can be visible. Hence, the similarity measures of a single factor, such as the shape or color of the *CSH* alone, cannot be a strong indicator for classifying DeepFake or real images. We measure the similarity scores using the extracted feature vectors, which contain multiple features, including color, edge, and the texture of the *CSH* images. We measure both Euclidean distance scores (EDS) and cosine distance scores (CDS) to statistically compare the similarity between two extracted feature vectors and find the geometric differences between right and left *CSH* images. The EDS is defined as:

$$d\left(A, B\right) = \sqrt{\sum_{i=1}^{n} \left(A_i - B_i\right)^2} \tag{1}$$

where n is the number of elements of the feature vectors, A and B are the corresponding *CSH* image vectors. d is a numerical value representing the Euclidean distance between A and B. The more similar *CSH* images, the EDS converges to 0. We also compute CDS, which is defined as:

$$\cos(A, B) = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} \left(A_i\right)^2} \sqrt{\sum_{i=1}^{n} \left(B_i\right)^2}} \tag{2}$$

If A and B are identical, the $\cos(A, B) = 1$. Otherwise, if they are completely different $\cos(A, B) = -1$. Thus, numbers between 0 and 1 indicate a similarity score, and numbers between -1 and 0 indicate a dissimilarity score. We applied the ReLU activation function to the EDS and CDS to avoid vanishing gradient problems while training our classifiers. The output [CDS, EDS] represents the semantic similarity between the projected representations of the two input *CSH* images.

Embedding Similarity Measures and Environmental Factors: In addition to the reflection shape similarity (RSS) measure, we have designed similarity measures and environmental factors embedding function, which takes similarity

measures [CDS, EDS], *SEF* features, and extracted (right and left) CSH features. Taking [IO, LL, LS] values from the input and annotated *SEF* values from the TDCA during training or from HEFD during testing, the similarity measures and environmental factors embedding function creates adjusted *SEF* values such as [IO', LL', LS']. Merging them with the similarity measures [CDS', EDS'] creates a row of mixed values [CDS', EDS', IO', LL', LS'] as an output. Finally, it takes vectors of (right and left) *CSH* images features and combines them in a vector for classification.

Classification: As illustrated in Fig. 2, the classification module classifies the input image, either real or DeepFake, by taking features from the embedding facility. We defined the classification network with a sequence of five blocks. The first block consists of a single BatchNormalization layer that normalizes its inputs ([CDS', EDS', IO', LL', LS']) by applying a transformation that maintains the mean output close to 0 and the output standard deviation close to 1. The following three blocks are similar. Every block consists of a sequence of a fully connected (*fc*) layer with 128 nodes, a single BatchNormalization layer followed by a ReLU activation function. The BatchNormalization layer centers the learned features from the fully connected layer on 0, while the ReLU activation uses 0 as a pivot to keep or drop the activated channels [4]. The fifth block consists of a concatenate layer and a fully connected layer. The concatenate layer merges the fourth block's output tensor with the CSH features vector. The fully connected layer (predication layer) returns a probability distribution with two nodes and a softmax activation function for binary classification. A binary cross-entropy probabilistic loss function was used to compute the cross-entropy loss between actual and predicted labels and to measure the model's accuracy during training and testing. Eventually, it creates a binary classification result (either real or DeepFake).

4 Evaluations

We conducted extensive experiments using CHIEFS-DFD datasets to evaluate the performance under real-world scenarios and compare the accuracy with current state-of-the-art (SOTA) DeepFake detection methods. We demonstrate one of the environmental parameter classification results (indoor or outdoor (IO)) and evaluate *CSH* regions detection. Finally, we present the classification performances with the CHIEFS-DFD datasets using different feature extraction backbone models and various similarity measures and environmental factors.

4.1 Evaluation of Indoor/Outdoor Classification

The primary purpose of this experiment is to assess the CHIEFS accuracy in classifying input facial images to either indoor or outdoor environments. We combined the CHIEFS and DIODE datasets with training the indoor/outdoor classifier. Among the 20,420 images, we labeled indoor (50%) and outdoor (50%) images equally and divided 16,336 images (80%) for the training set and 4,084 images (20%) for validation and testing sets. We used MobileNetV2 inverted residuals and linear bottlenecks neural network with binary cross-entropy loss function, dense layer of two nodes, and softmax activation at the top of the network to train the indoor/outdoor classifier. All images were pre-processed and scaled between −1 and 1. We used the Glorot normal initializer from the Keras library for the default weight initialization. We trained the model on the GPU environment for 18 h using the Google Colab Compute Engine (GCE) VM backend with (NVIDIA Tesla-P100-PCIE-16 GB) model for 512 iterations with an RMSprop optimizer, batch size of 32, and learning rate of 0.001. The early stopping criterion was used with patience set to 32 to stop training when a monitored metric (validation loss) stopped improving. The indoor/outdoor classifier achieves a 94.00% success rate in predicting indoor and outdoor images. The result indicates that CHIEFS can efficiently classify input facial images into indoor or outdoor categories.

4.2 Evaluation of CSH Regions Detection

We evaluated the CHIEFS accuracy in detecting *CSH* regions from the facial images. We split the CHIEFS dataset (1,285 facial images containing 2,570 annotated *CSH* segmentation masks) into 1,028 images (80%) for the training set and 257 images (20%) for validation and testing sets. We used the MobileNet-V2 feature extractor model and the Single Shot Detector (SSD) to detect and return the bounding boxes of right and left *CSH* regions and class labels. We trained the *CSH* detection model on the GPU environment for 6 h using the Google Colab Compute Engine (GCE) VM backend with (NVIDIA Tesla-P100-PCIE-16GB) model for 1,028 iterations. We use the standard RMSprop optimizer by configuring decay and momentum to 0.9, the standard weight decay to 0.00004, an initial learning rate of 0.045, a learning rate of 0.98 per epoch, and a batch size of 32. The result demonstrates that the overall mean average precision (mAP) of detecting right and left *CSH* regions is 90.53%, the right-reflection average precision (AP) is (90.81%), and the left-reflection AP is (90.26%), both are high enough for the *CSH* detection task.

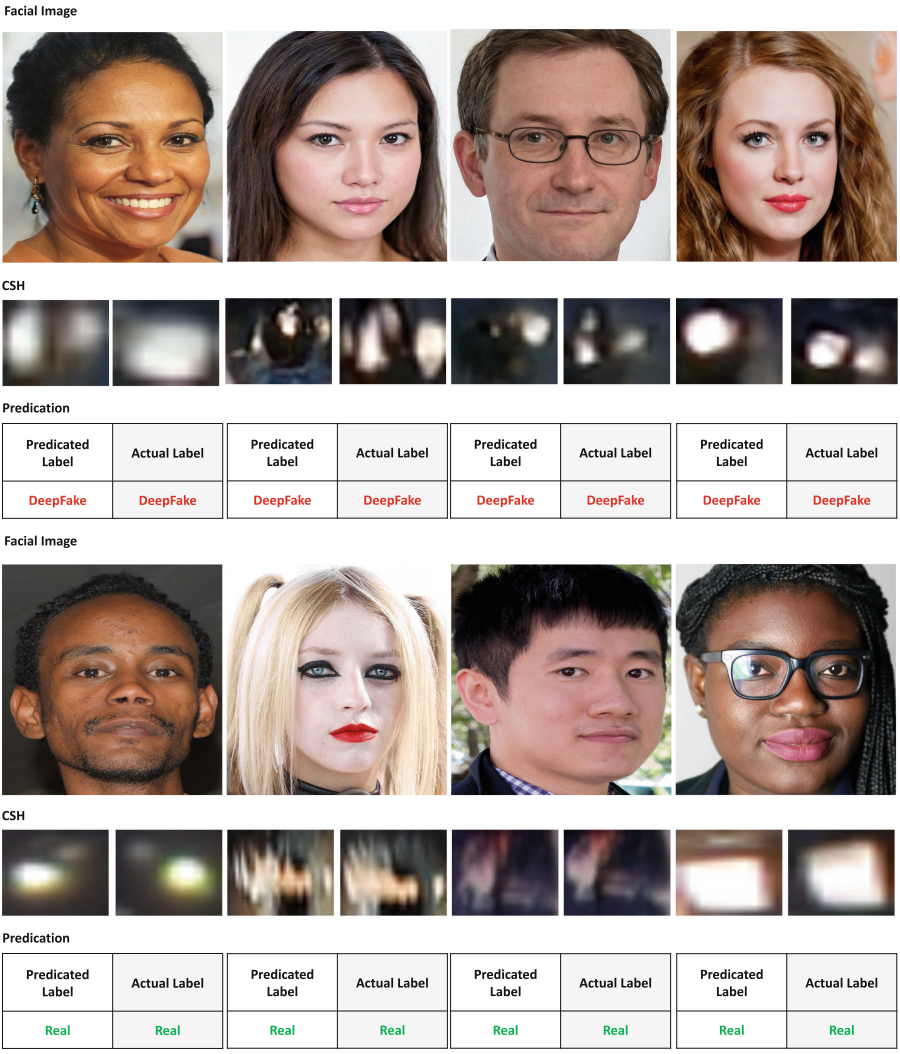

Fig. 4. Sample of the CHIEFS-DFD Testing Dataset Classification Result.

Table 1. Classification Performance Comparison on CHIEFS-DFD Dataset with Different Backbone Models for Feature Extraction.

Backbone	**Accuracy**	**Loss**
CHIEFS (DenseNet-201)	96.00%	0.592
CHIEFS (Xception)	98.00%	0.242
CHIEFS (VGG-16)	98.75%	0.203
CHIEFS (ResNet-50-V2)	**99.00%**	**0.160**

4.3 Classification Using Different Backbone Models for Feature Extraction

We evaluated the CHIEFS method with four different neural network backbones for feature extraction, including ResNet-50-V2, VGG-16, Xception, and DenseNet-201, using the CHIEFS-DFD dataset. After splitting the dataset with an 80:20 (training vs. validation) ratio. We trained the models on the GPU environment using the Google Colab Compute Engine (GCE) VM backend with (NVIDIA Tesla-P100-PCIE-16GB) model for 1,024 iterations with RMSprop optimizer, batch size of 8, and a learning rate of 1e−5. The early stopping criterion was used with patience set to 64 epochs to stop training when a monitored metric (validation loss) stopped improving. The results in Table 1 show the classification accuracy and loss of the CHIEFS method with different backbone models for feature extraction on the CHIEFS-DFD testing datasets. Overall, CHIEFS performs well with different feature extractors. For example, CHIEFS (ResNet-50-V2) is the best in both accuracy (99.00%) and loss (0.160). CHIEFS (VGG-16) is the second-best in both accuracy (98.75%) and loss (0.203). CHIEFS (Xception) is the third-best with accuracy (98.00%) and loss (0.242). Finally, CHIEFS (DenseNet-201)'s accuracy is the least (96.00%), and its loss is the highest (0.592). Figure 4 presents samples of the CHIEFS-DFD testing dataset classification results. CHIEFS detects DeepFake images with various face pose orientations, age, ethnicity, and appearances, such as makeup and accessories. Results indicate that CHIEFS performs well on realistic human visual DeepFake images.

4.4 Classification Using Different Feature Classifiers

Using the CHIEFS-DFD dataset, we assess different feature classifiers for CHIEFS (ResNet-50-V2). Table 2 shows that using all features, including right and left *CSH*, *RSS* ([CDS', EDS']), and *SEF* ([IO', LL', LS']) for classification achieves the best performance for CHIEFS (ResNet-50-V2) (99.00%) in accuracy. However, using a single *RSS* feature alone, such as [CDS'] or [EDS'], results in low accuracy (around 89.92%) with [CDS'] and (86.05%) with [EDS']. It also demonstrates that using right and left *CSH* features achieves high accuracy (93.00%) compared with other single components such as [CDS'] and [EDS']. When *SEF* features are used with the *CSH* features, the accuracy improves to (97.00%). Similarly, when *SEF* features are used with [CDS'] and [EDS'], the accuracy also improves to (94.00%) and (96.00%), respectively. The results indicate that using a single feature alone is not a good idea, and combining various features can improve performance greatly. In addition, the *SEF* features significantly impact accuracy improvement.

Table 2. Classification Performance Comparison with CHIEFS-DFD Dataset Using Different Feature Classifiers (i.e., CSH, CDS', EDS', IO', LL', LS') for CHIEFS (ResNet-50-V2).

Feature Classifiers	**Accuracy**
[CDS']	89.92%
[CDS', IO', LL', LS']	94.00%
[EDS']	86.05%
[EDS', IO', LL', LS']	96.00%
[CDS', EDS']	91.47%
[CSH]	93.00%
[CSH, IO', LL', LS']	97.00%
[CSH, CDS', EDS', IO', LL', LS']	**99.00%**

5 Conclusions

We proposed a novel ML-based DeepFake detection technology named CHIEFS (Corneal-Specular Highlights Imaging for Enhancing Fake-Face Spotter). We focus on the most reflective area of a human face, eyes, using *CSH* images. We verified the hypothesis that DeepFake technologies struggle to fake reflective components in their counterfeits by using various classifiers with environmental factors embedding. We designed and implemented feature extractors, classifiers, and embedding functions using advanced DNN architectures and tested them with different GAN-generated DeepFake datasets. The experimental results show that CHIEFS achieved high accuracy 99.00% in detecting sophisticated GAN-generated DeepFake images. Note that the modular design of CHIEFS renders itself as a complementary DeepFake detection module for any existing tools.

References

1. Chiefs-DFD dataset (2022). https://github.com/READFake/CHIEFS-DFD-Dataset. Accessed 26 Nov 2022
2. Lexica stable diffusion search engine (2022). https://lexica.art/. Accessed 24 Nov 2022
3. Chollet, F.: Xception: deep learning with depthwise separable convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1251–1258 (2017)
4. Chollet, F.: Deep Learning with Python. Simon and Schuster, New York (2021)
5. Dolhansky, B., et al.: The DeepFake detection challenge (DFDC) dataset. arXiv preprint arXiv:2006.07397 (2020)
6. Dutta, A., Zisserman, A.: The VIA annotation software for images, audio and video. In: Proceedings of the 27th ACM International Conference on Multimedia. MM 2019, ACM, New York, NY, USA (2019). https://doi.org/10.1145/3343031.3350535

7. Guo, H., Hu, S., Wang, X., Chang, M.C., Lyu, S.: Eyes tell all: irregular pupil shapes reveal GAN-generated faces (2021)
8. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9908, pp. 630–645. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46493-0_38
9. Hu, S., Li, Y., Lyu, S.: Exposing GAN-generated faces using inconsistent corneal specular highlights. arXiv preprint arXiv:2009.11924 (2020)
10. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks (2016). https://doi.org/10.48550/ARXIV.1608.06993, https://arxiv.org/abs/1608.06993
11. Icons8: Face Swapper (2022). https://icons8.com/swapper
12. Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of GANs for improved quality, stability, and variation. CoRR abs/1710.10196 (2017). http://arxiv.org/abs/1710.10196
13. Karras, T., et al.: Alias-free generative adversarial networks. In: Proceedings of the NeurIPS (2021)
14. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4401–4410 (2019)
15. Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T.: Analyzing and improving the image quality of StyleGAN. In: Proceedings of the CVPR (2020)
16. Korshunova, I., Shi, W., Dambre, J., Theis, L.: Fast face-swap using convolutional neural networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3677–3685 (2017)
17. Le, B.M., Woo, S.S.: Exploring the asynchronous of the frequency spectra of GAN-generated facial images. CoRR abs/2112.08050 (2021). https://arxiv.org/abs/2112.08050
18. Li, L., Bao, J., Yang, H., Chen, D., Wen, F.: FaceShifter: towards high fidelity and occlusion aware face swapping (2020)
19. Li, Y., Chang, M.C., Lyu, S.: In Ictu oculi: Exposing AI generated fake face videos by detecting eye blinking. arXiv preprint arXiv:1806.02877 (2018)
20. Li, Y., Yang, X., Sun, P., Qi, H., Lyu, S.: Celeb-DF: A large-scale challenging dataset for DeepFake forensics. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3204–3213 (2020). https://doi.org/10.1109/CVPR42600.2020.00327
21. Liu, H., et al.: Spatial-phase shallow learning: Rethinking face forgery detection in frequency domain (2021)
22. Matern, F., Riess, C., Stamminger, M.: Exploiting visual artifacts to expose DeepFakes and face manipulations. In: 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW), pp. 83–92. IEEE (2019)
23. Mo, H., Chen, B., Luo, W.: Fake faces identification via convolutional neural network. In: Proceedings of the 6th ACM Workshop on Information Hiding and Multimedia Security, p. 43–47. IH& MMSec 2018, Association for Computing Machinery, New York, NY, USA (2018). https://doi.org/10.1145/3206004.3206009
24. Nguyen, H.H., Fang, F., Yamagishi, J., Echizen, I.: Multi-task learning for detecting and segmenting manipulated facial images and videos. In: 2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS), pp. 1–8 (2019). https://doi.org/10.1109/BTAS46853.2019.9185974
25. Nirkin, Y., Keller, Y., Hassner, T.: FSGAN: subject agnostic face swapping and reenactment. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 7184–7193 (2019)

26. Petrov, I., et al.: DeepFaceLab: a simple, flexible and extensible face swapping framework. arXiv preprint arXiv:2005.05535 (2020)
27. Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., Nießner, M.: FaceForensics++: learning to detect manipulated facial images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1–11 (2019)
28. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: MobileNetV2: inverted residuals and linear bottlenecks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4510–4520 (2018)
29. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
30. Sun, Z., Han, Y., Hua, Z., Ruan, N., Jia, W.: Improving the efficiency and robustness of DeepFakes detection through precise geometric features (2021)
31. Tariq, S., Lee, S., Kim, H., Shin, Y., Woo, S.S.: Detecting both machine and human created fake face images in the wild. In: Proceedings of the 2nd International Workshop on Multimedia Privacy and Security, pp. 81–87 (2018)
32. Vasiljevic, I., et al.: DIODE: a dense indoor and outdoor depth dataset. CoRR abs/1908.00463 (2019). http://arxiv.org/abs/1908.00463
33. Woo, S., et al.: Add: frequency attention and multi-view based knowledge distillation to detect low-quality compressed DeepFake images. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, pp. 122–130 (2022)
34. Yang, X., Li, Y., Lyu, S.: Exposing deep fakes using inconsistent head poses. In: 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). ICASSP 2019, pp. 8261–8265. IEEE (2019)