Identifying Fake and Real images by using Masked Face Periocular Region

Udayasri Nannuri
North Carolina A&T State University
Greensboro, NC
unannuri@aggies.ncat.edu

Kaushik Roy
North Carolina A&T State University
Greensboro, NC
kroy@ncat.edu

Jinsheng Xu
North Carolina A&T State University
Greensboro, NC
jxu@ncat.edu

Tony Gwyn
North Carolina A&T State University
Greensboro, NC
tgwyn@aggies.ncat.edu

Bianca T Govan
North Carolina A&T State University
Greensboro, NC
btgovan@ncat.edu

I. Abstract

In this paper, we focus on the face spoofing of masked images to determine whether a masked person is real or fake. We developed a dataset of spoofed masked images generated using the DALL.E 2 tool, performed the ROI extraction using the CNN-DLib detector, and extracted the features using BoVW-SIFT. We applied deep learning and machine learning algorithms. XGBoost and Xception achieved the highest accuracy of 92% and 94% to determine whether the images were real or fake. The approach was tested on the real-world masked face recognition dataset (RMFRD). This shows that periocular information can predict whether the masked image is real or fake.

***Index Terms*—periocular region, DLib detector, VGG16, Xception, Spoofing, Presentation attack.**

II. Introduction

In the recent past, COVID-19 was considered at the forefront as a highly contagious virus that can lead to symptoms such as high body temperature, coughing, headache, tiredness, difficulty breathing, and loss of smell and taste. Viruses can be spread through the air when we breathe. Viruses can be spread through the air when we breathe. Since the spread of COVID-19, it has been mandatory to wear a face mask in all public places. This includes restaurants, bars, entertainment venues, offices, and airlines. The years 2020 and beyond are accompanied by the terms pandemic, self-distance, masking, and unmasking. As reported by the World Health Organization (WHO), COVID-19 is a global pandemic that has affected all 220 countries [1] around the world. The right to use a face mask is the only defense against COVID-19.

The term" Biometrics" [2] refers to the automatic recognition of individuals by means of physical and behavioral characteristics. Biometric authentication and identification can be used for facial attendance, airport security, building access, the banking sector, IoT for the home, and some other applications. The physical characteristics of individuals, such iris, faces, and fingerprints, have been used extensively for authentication [3].

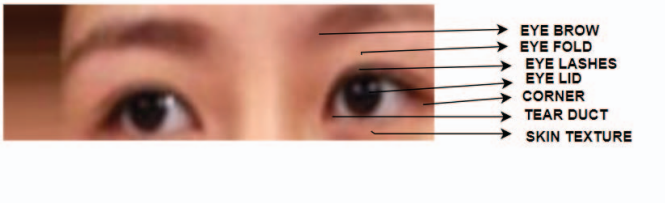

Fig. 1. Periocular Region.

Image recognition plays a big role in our daily lives and is employed in many areas, like the military, intelligence, journalism, criminal investigation, the news, the courtroom, and so on. Since image alteration apps are easily available to use, manipulation of the images has become easier and more widespread [4]. It is hard to tell if the images are real or fake. Because images are used as legal evidence in many fields, it is important to know if an image is fake or not. Many spoofing algorithms are developed and used to deceive biometric-based authentication techniques.

Biometric authentication has evolved into a component that is now necessary for day-to-day activities [5]. Human eyes are categorized into two parts. The upper eye consists of the eyebrow, upper eye fold, upper eyelid, eye, tear duct, and eyelashes. The lower eye consists of the lower eyelid, lower eye-fold, eye socket, skin texture, and outer corner as shown in the Fig 1.

Due to COVID-19, people wear face masks to avoid contacting the virus. A periocular region, the area around the eye, can be used to identify an individual when one wears a mask. The lower part of the face, like the mouth, cheeks, and chin, is covered by a mask. Therefore, a periocular region can be used to authenticate an individual. Recently, an AI-based image manipulation tool has been extensively used to alter the biometric data. In this research, we aim to identify the

979-8-3503-2759-5/23/$31.00 ©2023 IEEE
DOI 10.1109/CSCE60160.2023.00121

Fig. 2. RMFRD Dataset.

real and fake images in an attempt to improve periocular-based authentication.

The key characteristics, including the shape and position of the eyes and eyebrows in the periocular region, are important for identification.

The main contributions of this paper are as follows:

1) We created a dataset by utilizing DALL.E 2, an open AI tool, to generate 5000 fake photos from the real-world masked face dataset.
2) We applied DLib [6], a deep learning-based face detector, to segment the periocular region from the masked faces.
3) We deployed Bag of Visual Words (BoVW) SIFT-based feature extraction techniques [7] to elicit the most important periocular features.
4) We employed two deep learning (DL) models, Xception and VGG16, for spoof detection. We also applied traditional machine learning (ML) algorithms such as support vector machines (SVM), random forests (RF), XGBoost, gradient boosting, and decision trees to identify whether the images are real or fake.

The remainder of this paper is as follows: Section III, discusses the related work briefly. Section IV, describes the dataset generation, extraction of the periocular region, and feature extraction. Section V, discusses the results we obtained using the ML and DL models. Section VI, summarizes our research findings and concludes with possible future directions.

III. Related work

Researchers in [8], trained a convolutional neural network (CNN) to distinguish between real and fake faces in a single, full-size image or a batch of five reduced photos using ImageNet requires biomarkers for image alignment. A support vector machine performed binary classification (fake/real) on the CNN output (SVM). A two-stream CNN is proposed in [9], The first stream is employed to scan local face segments and assign spoofing estimates. While the second stream is trained to estimate scene depth from 3D data. In [10], authors proposed a CNN variant with a more intricate design; they referred to it as CNN deep part features. To categories data, a second, fine-tuned VGG (Visual Geometry Group) CNN was given a subset of the extracted features from the first CNN. In [11], the authors used CNN to separate a photograph into a real face and a spoofing noise. In this study's photo categorization system, noise was used to implement authenticity checking. Because they depend on deep learning models, adversarial perturbation attacks are a relatively recent

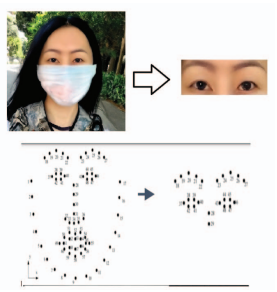

Fig. 3. Feature Extraction [5].

development. The adversarial modification is reduced to a little change in the brightness or contrast of the input image, which is unnoticeable by human vision but leads the deep neural network to make an incorrect classification. In [12], the authors proposed analyzing the responses of filters in convoluted layers and deleting the most troublesome filters in order to detect such hidden attacks. In [13], the authors proposed the Smart Box software instrument for measuring the effectiveness of adversarial and countermeasure techniques in face recognition systems. The Smart Box software package supports many methods, including Deep Fool and Elastic-Net, as well as tools against gradient attacks and L2 attacks. Although there has been some success in limiting adversarial disturbance attempts, the sophistication of these attacks continues to increase, demanding progressively more refined responses. The Stealing of detailed face templates for the purpose of manipulation by third parties is a further, more specific form of attack that must be recognized. NB-Net, a de-convolutional neural network was presented in [14], as a protective measure against such risks. The issue is that generative adversarial networks can be employed in digital manipulation attacks to generate fully or partially altered photorealistic facial images by altering an expression, modifying features, or synthesizing a whole new face. Hence, adversarial perturbation attacks are aimed at deep neural networks that have demonstrated efficacy in face recognition.

IV. Methodology

In this section, we describe the data collection process, feature extraction and ML and DL models used.

A. Data collection.

In this paper, we use the Real World Masked Face Recognition dataset (RMFRD) [15]. Fake dataset created from RMFRD using the DALL.E 2, an Open AI tool [16]. 5000 fake images are generated from the real world masked face dataset(RMFRD). Fig 2 and 4 shows the samples of real images. Fig 5 and 6 shows the generated fake images.

B. Feature Engineering.

The feature engineering process consists of three phases: periocular region extraction, preprocessing, and feature extraction.

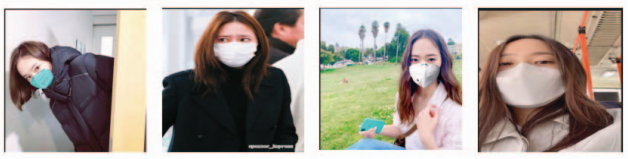

Fig. 4. Real Image.

Fig. 5. Generated image 1.

1) periocular extraction : In order to extract the periocular region, we apply an OpenCV library known as dlib [6]. The tool finds the 68 facial landmarks in masked facial images. This research uses the facial landmarks from the leftmost part of the left eyebrow to the bottom part of the right eyeball. The facial landmarks 18 through 30 are considered to capture the periocular region well, as shown in Fig 3. The extracted periocular regions of the fake and real masked images, as shown in Fig 7, and 8.

2) Preprocessing : For preprocessing, we resized all the images to a uniform size of 128x128 pixels. Furthermore, to reduce the complexity of the image data, we converted all the images into grayscale.

3) Feature extraction using (BoVW) SIFT : The Bag of Visual Words Scale-Invariant Feature Transform BoVW-SIFT approach [17] is used to extract the textural features from the periocular region. The steps for the BoVW-SIFT based feature extraction process reported in are given below:

- Scale-space extrema detection : The image feature points are identified by computing the Difference of Gaussian (DoG) at multiple scales and looking for local extrema in the scale space of an image.
- Keypoint localization : After detecting potential feature points, a 3D quadratic function is used to fit the DoG scale space around each extrema and to localize the key points.
- Orientation assignment : After locating the key points, orientation is determined. Neighborhood gradient directions determine the key point orientation.
- Descriptor generation : The final step is to generate the descriptors for each key point. Gradient orientations surrounding the key point are used to generate histograms.
- Feature Quantization: The quantization process turns the SIFT features into visual words. Clustering the SIFT descriptors using the K-means algorithm accomplishes this process.
- Encoding the features : After getting the visual words, SIFT characteristics are encoded into a histogram. Each SIFT descriptor is assigned to the closest visual word, and its histogram bin is incremented.

Fig. 6. Generated image 2.

Fig. 7. Fake Image.

Fig. 8. Real Image.

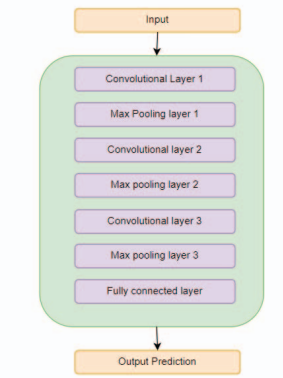

Fig. 9. VGG16 Model.

- Normalization : The final step normalizes the histogram to account for image size and illumination.

C. Machine Learning and Deep Learning.

In this research, we apply traditional ML models such as Support Vector Machines (SVM) [18], XGBoost [19], Random Forest (RF) [20], Decision Tree [21] and Gradient Boosting [22] to classify the images as real or fake. We also apply DL models such as VGG16 and Xception for real and fake image classification.

VGG16. It has sixteen layers [23] including thirteen convolutional layers and three fully connected layers as shown in Fig. 9. One of the most important aspects of the VGG16 network is that it makes use of very small convolutional filters (3x3) across the network. Downsampling the feature maps and

TABLE I
MACHINE LEARNING MODELS EVALUATION

Model	Accuracy	Precision	Recall	F1
Decision Tree	81.28%	83.61%	82.40%	84.63%
XG-Boost	92.42%	91.18%	93.50%	94.22%
SVM	81.55%	83.12%	82.41%	83.43%
Gradient Boosting	88.85%	89.32%	89.31%	87.30%
Random Forest	90.45%	93.12%	92.31%	92.53%

TABLE II
DEEP LEARNING MODELS

Model	Accuracy	Precision	Recall	F1
VGG16	92.32%	93.38%	94.98%	93.89%
Xception	94.12%	96.04%	95.38%	95.46%

lowering their dimensionality of these maps are accomplished within the pooling layers. The VGG16 makes use of skip connections between the convolutional layers.

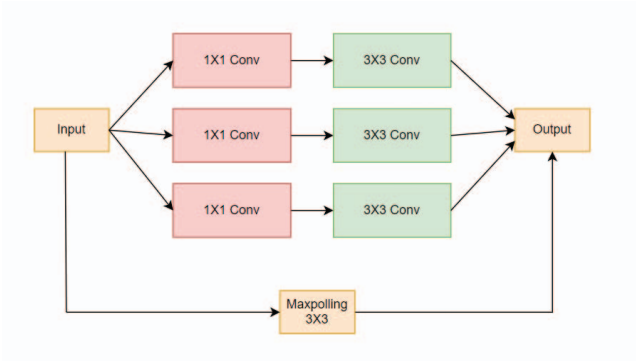

Fig. 10. Xception Model.

Xception. It is an extension of the Inception architecture [24], which captures features at several resolutions by using a mix of parallel convolutional filters operating at different scales as shown in Fig. 10. The model distinguishes the cross-channel correlations from spatial correlations. The network is able to obtain information about the space at a finer scale while still maintaining its channel-wise interactions.

V. RESULTS

In this section, we demonstrate the results obtained through the ML and DL models. The model is evaluated based on four different measures: Accuracy, Precision, Recall and F1-Score. Fig.11, 12 show the samples of images predicted by the ML and DL algorithms. Table I shows the performance metrics of ML models. XGBoost was the most accurate at 92%, followed by RF at 90%. Gradient Boosting attained 88% accuracy. Both Decision Tree and SVM have lower accuracies, 81.28% and 81.55%, respectively. RF and Gradient Boost have F1-scores of 92.53% and 87.30%, respectively. The SVM and Decision Tree models have F1-scores of 84.63% and 83.43%, respectively.

Among the ML models, XGBoost achieved the highest accuracy and F1-score of 92.42% and 94.22% respectively.

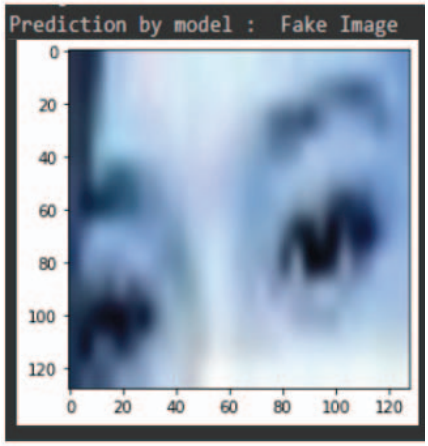

Fig. 11. Fake image predicted by our model.

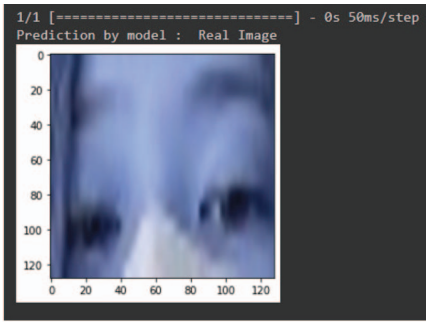

Fig. 12. Real Image predicted by our model.

Table II shows the DL-based performance scores. While VGG16 achieved 92.32% accuracy, Xception model achieved the highest accuracy of 94.12%. However, Xception obtained a high F1-score of 95.46% whereas VGG16 received 93.89% F1-score. The Xception model shows the highest accuracy among all the models. The results indicate that the Xception network can be used to classify real and fake images.

VI. CONCLUSION

In this paper, we developed a spoofed image dataset by using DALL.E 2, an AI tool, from the RWMFD face dataset. We performed periocular region extraction using dlib, a CNN-based face detector. The feature extraction was accomplished using BoVG-SIFT. We applied different traditional ML models to classify the images as fake and real. The DL models were also applied for image classification. The results indicate DL-based models outperformed ML models in terms of accuracy. In future, we would like to use a visual transformer model to improve the performance. We will apply various explainable techniques [25] to analyze the most important features contributing to the possible outcome.

ACKNOWLEDGEMENT

This research is supported by National Science Foundation (NSF). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of NSF.

REFERENCES

[1] W. Coronavirus, "Dashboard [https://covid19. who. int/]," *Accessed March*, vol. 15, 2021.

[2] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on circuits and systems for video technology*, vol. 14, no. 1, pp. 4–20, 2004.

[3] A. K. Jain, "Biometric recognition," *Nature*, vol. 449, no. 7158, pp. 38–40, 2007.

[4] T. J. Jayan and R. Aneesh, "Image quality measures based face spoofing detection algorithm for online social media," in *2018 International CET Conference on Control, Communication, and Computing (IC4)*. IEEE, 2018, pp. 245–249.

[5] P. Kumari and K. Seeja, "Periocular biometrics: A survey," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 4, pp. 1086–1097, 2022.

[6] D. E. King, "Dlib-ml: A machine learning toolkit," *The Journal of Machine Learning Research*, vol. 10, pp. 1755–1758, 2009.

[7] V. D. Sachdeva, J. Baber, M. Bakhtyar, I. Ullah, W. Noor, and A. Basit, "Performance evaluation of sift and convolutional neural network for image retrieval," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 12, 2017.

[8] J. Yang, Z. Lei, and S. Z. Li, "Learn convolutional neural network for face anti-spoofing," *arXiv preprint arXiv:1408.5601*, 2014.

[9] Y. Atoum, Y. Liu, A. Jourabloo, and X. Liu, "Face anti-spoofing using patch and depth-based cnns," in *2017 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 2017, pp. 319–328.

[10] L. Li, X. Feng, Z. Boulkenafet, Z. Xia, M. Li, and A. Hadid, "An original face anti-spoofing approach using partial convolutional neural network," in *2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)*. IEEE, 2016, pp. 1–6.

[11] A. Jourabloo, Y. Liu, and X. Liu, "Face de-spoofing: Anti-spoofing via noise modeling," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 290–306.

[12] G. Goswami, N. Ratha, A. Agarwal, R. Singh, and M. Vatsa, "Unravelling robustness of deep learning based face recognition against adversarial attacks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.

[13] A. Goel, A. Singh, A. Agarwal, M. Vatsa, and R. Singh, "Smartbox: Benchmarking adversarial detection and mitigation algorithms for face recognition," in *2018 IEEE 9th international conference on biometrics theory, applications and systems (BTAS)*. IEEE, 2018, pp. 1–7.

[14] G. Mai, K. Cao, P. C. Yuen, and A. K. Jain, "On the reconstruction of face images from deep face templates," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 5, pp. 1188–1202, 2018.

[15] Z. Wang, G. Wang, B. Huang, Z. Xiong, Q. Hong, H. Wu, P. Yi, K. Jiang, N. Wang, Y. Pei *et al.*, "Masked face recognition dataset and application," *arXiv preprint arXiv:2003.09093*, 2020.

[16] F. Zhang, B. Du, and L. Zhang, "Scene classification via a gradient boosting random convolutional network framework," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 3, pp. 1793–1802, 2015.

[17] F. B. Silva, S. Goldenstein, S. Tabbone, and R. d. S. Torres, "Image classification based on bag of visual graphs," in *2013 IEEE International Conference on Image Processing*. IEEE, 2013, pp. 4312–4316.

[18] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[19] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[20] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[21] C. Agarwal and A. Sharma, "Image understanding using decision tree based machine learning," in *ICIMU 2011: Proceedings of the 5th international Conference on Information Technology & Multimedia*. IEEE, 2011, pp. 1–8.

[22] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[23] X. Xuan, B. Peng, W. Wang, and J. Dong, "On the generalization of gan image forensics," in *Biometric Recognition: 14th Chinese Conference, CCBR 2019, Zhuzhou, China, October 12–13, 2019, Proceedings*. Springer, 2019, pp. 134–141.

[24] E. Westphal and H. Seitz, "A machine learning method for defect detection and visualization in selective laser sintering based on convolutional neural networks," *Additive Manufacturing*, vol. 41, p. 101965, 2021.

[25] H. Manthena, "Explainable machine learning based malware analysis," Ph.D. dissertation, North Carolina Agricultural and Technical State University, 2022.